Proceedings

rf technology expo 88

February 10-12, 1988
Anaheim, California

Sponsored by
rfdesign Magazine

Proceedings

rf technology expo 88

February 10-12, 1988
Anaheim, California

RF Design
6300 S. Syracuse Way
Suite 650
Englewood, Colorado 80111

Table of Contents

The papers contained in these Proceedings represent the views of the authors and/or their employers, and have been presented by them for publication as original works. While every attempt has been made to assure accuracy, Cardiff Publishing Company and RF Design will not be responsible for any copyright, trademark or patent infringement that may be contained in these papers.

RF POWER AMPLIFIER INSTABILITY - CAUSES AND CURES

By
Peter A. Kwitkowski
Motorola Communications Sector
1301 E. Algonquin Road
Schaumburg, Illinois, 60196

ABSTRACT

Spurious oscillations are a common problem encountered during the design of RF power amplifiers. Typically the resolution has been by trial and error with little understanding of their causes. The analytical prediction of stability in RF power amplifiers is considerably behind the advances made in low level, linear amplifiers over the past 25 years. As a result of the inherent non-linearities in large signal solid state amplifiers mathematical analysis and modeling have been exceedingly difficult. However, techniques have been developed which allow the designer to evaluate tendencies toward instability in operating RF power amplifiers. This paper presents a method of stability analysis based upon small signal gain measurements of an operating RF power amplifier. Coupling these measurement methods with a proper selection of circuit topology greatly reduces the tendency for instability.

INTRODUCTION

While the examples in this paper deal specifically with 800 MHz bipolar class C amplifiers in a common emitter configuration, the principles can be generalized for application to other types and classes of operation. Instabilities in RF power amplifiers may be observed in several ways. They may be evidenced by erratic tuning, current being drawn when drive is removed, or actually seen on a spectrum analyzer. Whatever their manifestation, they can result in an amplifier that does not meet its specifications or one that destroys itself due to overdissipation. Instabilities can be traced to several potential causes:

1) External feedback around the amplifier stage(s)
2) Internal feedback within the active device
3) Varactor modes

The first two causes require the designer to identify the feedback mechanism(s) and eliminate it (them) or to reduce the gain at those frequencies to a level which will not support oscillation. Varactor modes of oscillation result from the non-linear capacitance vs. voltage characteristics of semiconductor junctions. While the circuitry external to the transistor affects varactor mode instabilities the transistor itself may be the inherent cause. Three basic rules can be applied in order to insure the stability of an RF power amplifier. They are the following:

1) Present RF power transistors with proper terminating impedances at all frequencies where the device has gain, not just in the operating frequency band.
2) Remove unwanted sources of feedback.
3) Do not overdrive the amplifier.

SMALL SIGNAL SWEPT GAIN MEASUREMENT

Before discussing some of the causes of instability it would be instructive to present a method of measurement that will graphically display regenerative tendencies. Small signal swept gain measurements are made by injecting a low level secondary signal into the input of an operating RF power amplifier and measuring the amplifier's interaction with that signal. The measuring system

shown in Figure 1 consists of an operating amplifier stage with directional couplers connected to both its input and output. A spectrum analyzer/tracking generator is used in conjunction with a broadband amplifier to produce a sweep signal. The sweep signal is fed into the coupled port of the input directional coupler (reflected port). The resultant signal is picked off at the coupled port of the output directional coupler (forward port). With the sweep signal turned off, the spectrum analyzer is adjusted for a convenient trace. The amplifier is next removed and replaced with a through connector. The sweep signal level is adjusted so that it will not affect the normal operation of the amplifier (typically 30 dB below that used to drive the amplifier). With the amplifier reconnected the small signal swept gain can now be measured. The two spectrum analyzer traces shown in Figure 2 provide a good indication of the power of this measurement method. The lower trace is what is normally seen on a spectrum analyzer plot of the amplifier output. There is an indication that something is going on by the slight peaks in the noise at about ± 40 MHz from the carrier frequency. When the sweep signal is turned on a dramatic result is obtained. At ± 40 MHz two very large gain peaks are seen. The amplifier used for the measurements in Figure 2 is on the verge of oscillation.

The absolute level of the small signal gain is not of great significance. However, two situations can provide an indication of potential instability. These are:

1) Sharp peaks or dips in amplifier gain - (these may change with variations in supply voltage, drive level, source/load VSWR, temperature, and operating frequency).
2) Abrupt changes in the curve which vary under the same conditions as #1.

EXTERNAL FEEDBACK AROUND THE AMPLIFIER STAGE

Regeneration can result from improper power supply decoupling, allowing energy to be fed back from one stage to another. Problems can sometimes result, even with single stage amplifiers, when the power supply impedance is too high. Regeneration resulting from insufficient power supply decoupling is typically low in frequency, where the RF bypass capacitors become high impedances. Figure 3 shows a typical D.C. feed system without proper low frequency bypassing. Currents flowing through the common impedance, Z_2 can result in unwanted coupling between stages. Impedance Z_2 may be inductive as well as resistive in nature, particularly where the D.C. leads are coiled. The best practice, as shown in Figure 4, is to bypass each stage individually with relatively large capacitor values (typically 1 µF to 100 µF).

Poor grounding practices are another cause of instability. Feedback paths result from the flow of RF currents through common impedances in the ground connections of an amplifier (Figure 5a). Common ground impedances may result in energy being coupled, not only between stages, but between the output and input of an individual amplifier stage. As the frequency of operation is increased the amount of coupling is increased in the common impedances. "Ground loops" may cause discontinuities in the frequency response (gain) and "drive up" characteristics (power snap). The best solution is to connect all ground areas on each amplifier stage directly to a large overall ground plane (Figure 5b). The ground plane will provide a low impedance path between all ground points on the amplifier.

Another form of external feedback is caused by RF energy which is radiated from

an antenna connected to the amplifier. If proper shielding and bypassing techniques are not employed the radiated energy may be coupled into the input of the amplifier. Gain variations and oscillation which change with physical positioning of the amplifier and/or the antenna are an indication of radiated feedback.

INSTABILITIES RESULTING FROM FEEDBACK INSIDE THE TRANSISTOR

Proper impedance termination of the transistor at all frequencies is required in order to insure that the amplifier is stable. As the matching networks' performance degrades out of band, D.C. biasing networks can play an important role in keeping the amplifier stable. Conventional design philosophy dictates that D.C. feed chokes have an impedance level of >10 times that of the circuit into which they are connected. The "cold" end of the input feed choke is often tied directly to ground while the "cold" end of the output feed choke is heavily bypassed (Figure 6). At lower frequencies where the transistor has very high gain the D.C. feed chokes, along with the feedback capacitance of the device, can form a classical oscillator circuit (Figure 7). In order to reduce the probability of this type of regeneration, the D.C. feeds need to be modified to provide a low Q termination to the device at low frequencies.

Figure 8 shows three different feed networks along with a Smith chart plot of their impedances from 1MHz to 1 GHz. A 25 nH inductor (L_1, Figure 8a) is seen to present an impedance that stays at the edge of the Smith chart (ie. high Q).

Adding a ferrite bead, L_2 in series with the feed inductor is a common technique for reducing the Q of the network, even at low frequencies (Figure 8b). This has proven quite successful in improving stability. Unfortunately the bead absorbs power from the amplifier due to RF currents flowing in the bead at the operating frequency.

Another configuration that works even better than the simple addition of a ferrite bead is a technique known as "double decoupling" (Figure 8c). As the name implies, the combination of L_1 and C_1 "decouple" the feed network at the operating frequency while L_2 and R_1 "decouple" at low frequencies. The value of R_1 is selected to approximate the transistor impedance (typically 1 Ω to 100 Ω). The inductor L_2 should have an impedance close to the value of R_1 at frequencies of a few MHz. While a high value wire wound inductor may be used for L_2, a ferrite bead based inductor is recommended. The high permeability of the ferrite causes a substantial impedance level to be reached even with a single turn of wire. This is particularly important when the choke must handle large values of D.C. current. Depending on the value of R_1 needed in parallel with L_2, a ferrite bead alone may be sufficient. The capacitor C_1 should have an impedance of about 1/10 the value of R_1 at the operating frequency. The network used in the collector D.C. feed requires the addition of a large value capacitor C_2 (typically 0.1 µF to 1 µF) in order to provide a low impedance ground path for low frequency currents through R_1. When proper double decoupling is applied to both the collector and base/emitter D.C. feeds, a significant improvement in stability can be obtained with little power loss. The exact values needed in the double decoupling networks may be determined by evaluating the stability of the stage with the swept small signal gain technique.

Figures 9 through 12 provide an indication of the stability of an amplifier using the three D.C. feed networks just described. Figure 9 shows a small signal gain sweep of an amplifier using the same feed choke of Figure 8a. Even when the amplifier is operated into a 50 Ω load severe gain peaks are evident. When a 3:1 VSWR is applied to the output of the amplifier regeneration results (Figure 10, sweep signal turned off). The gain sweep in Figure 11 shows the performance of the choke and bead from Figure 8b with the amplifier operated into a 3:1 VSWR. The large peaks at ±40 MHz are gone but at ±8 MHz there are indications of potential instability. Using the double decoupled network of Figure 8c the stability performance improves even more (Figure 12).

Stability of an amplifier can be improved by the application of negative feedback. A series R-L-C network between the collector and base of the transistor can be used to reduce the device gain at low frequencies (Figure 13). The capacitor in the circuit serves as a D.C. block (typical value .01 µF to 0.1µF). The R-L combination is used to provide an increasing amount of negative feedback as frequency decreases. The inductor is selected to have an impedance of > 10 times the device impedance (input or output) at the operating frequency. The resistor value is in the range of 10 Ω to 100 Ω.

The occurrence of in-band (inside the operating bandwidth of the amplifier) regeneration is possible with the proper combination of input/output impedances and the proper amount and phase of feedback. An amplifier with this type of instability will generally sustain its output when drive is removed. A conjugately matched amplifier which exhibits in-band instabilities into VSWRs of 3:1 or less probably has a transistor problem. When developing a device for production a manufacturer optimizes around a test circuit. During the development phase the device will be subjected to various degrees of mistuning. If the device exhibits an inherent tendency toward in-band regeneration it will most likely be discovered here and, hopefully, corrected. If one does not deviate greatly from the match topology used by the device manufacturer the amplifier should not exhibit in-band instabilities into less than a 3:1 VSWR. Reference #2 presents a method using S parameters for analyzing instabilities of this type.

VARACTOR MODES

In a bipolar transistor the collector-base capacitance varies as a non-linear function of the voltage across the junction. This non-linearity can produce spurious signals in the amplifier. One of the most common spurious products results from varactor division. When the collector-base junction of a bipolar transistor is included in a low impedance signal path at a sub-multiple of the carrier frequency spurious energy may be produced, typically at $f_0/2$ (Figure 14). In order to prevent $f_0/2$ regeneration the input and output of the transistor should be terminated with a resistive load rather than a short circuit. Sometimes $f_0/2$ regeneration is caused primarily by the transistor. If it is not possible to eliminate $f_0/2$ regeneration with changes in circuitry the device may be at fault. Some transistors exhibit a "push-pull" mode of $f_0/2$ regeneration due to an imbalance in currents across the die. In such cases selection of an alternate device should be strongly considered. The greater the voltage swing across the transistor junctions the greater the chance for varactor mode instabilities. As the swings are increased the variation in capacitance will increase, improving the chances that an unstable region will be encountered. Applying excessive drive to an RF power amplifier stage at low supply voltages increases the probability of parametric instabilities.

CONCLUSION

The analysis of instability in RF power amplifiers is not an exact science. However, an understanding of the basic causes of instability and the use of proper circuit topology, coupled with the small signal swept gain measurement technique described in this paper, increases the designers chances of success.

REFERENCES

1) H.O. Granberg, "Good RF Construction Practices and Techniques", RF Design, September/October 1980.

2) R. Jack Frost, "Large Scale S Parameters Help Analyze Stability", Electronic Design, May 24, 1980.

3) E. A. Franke and A. E. Noorani, "Improving The Stability Of Mobile Radio Power Amplifiers", IEEE 33rd Vehicular Technology Conference, 1983

4) J. Johnson, et al, "Solid Circuits", Communications Transistor Corporation, March, 1973

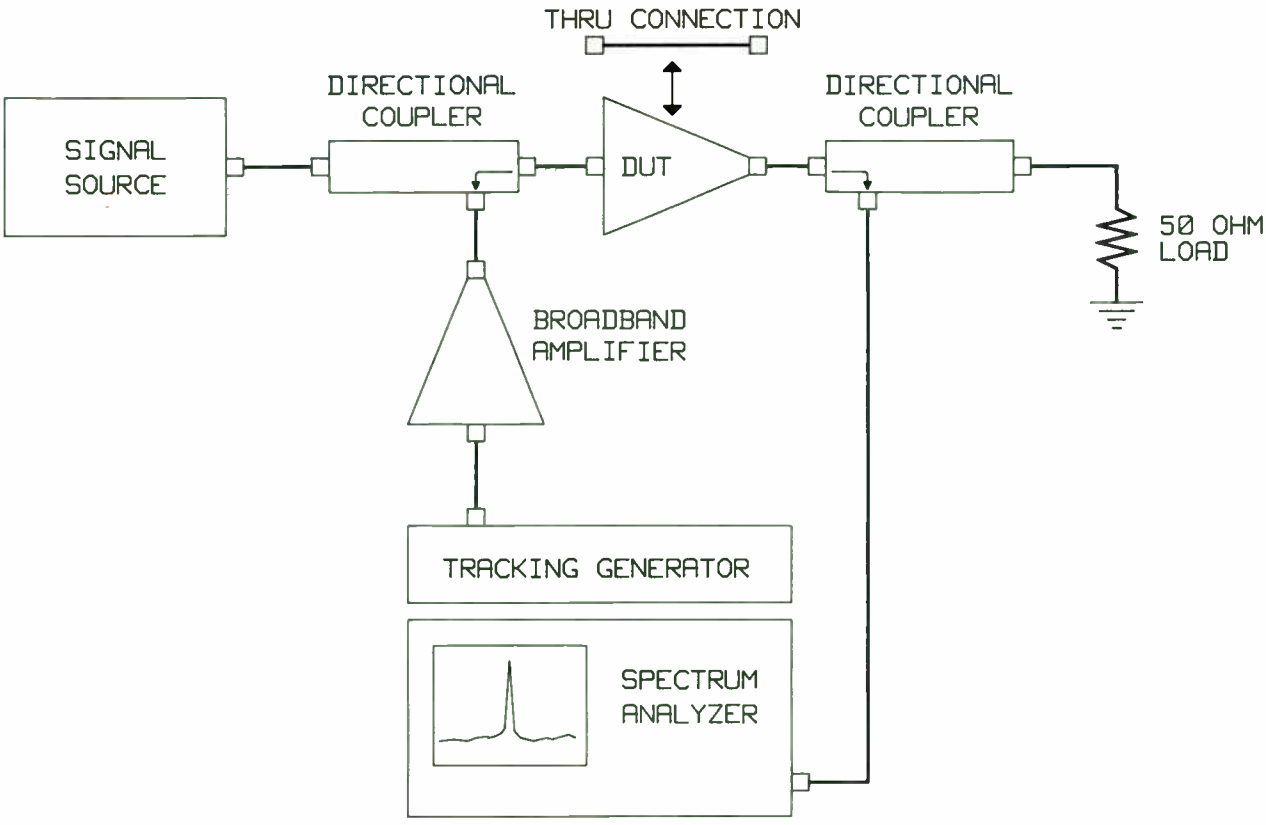


FIGURE 1

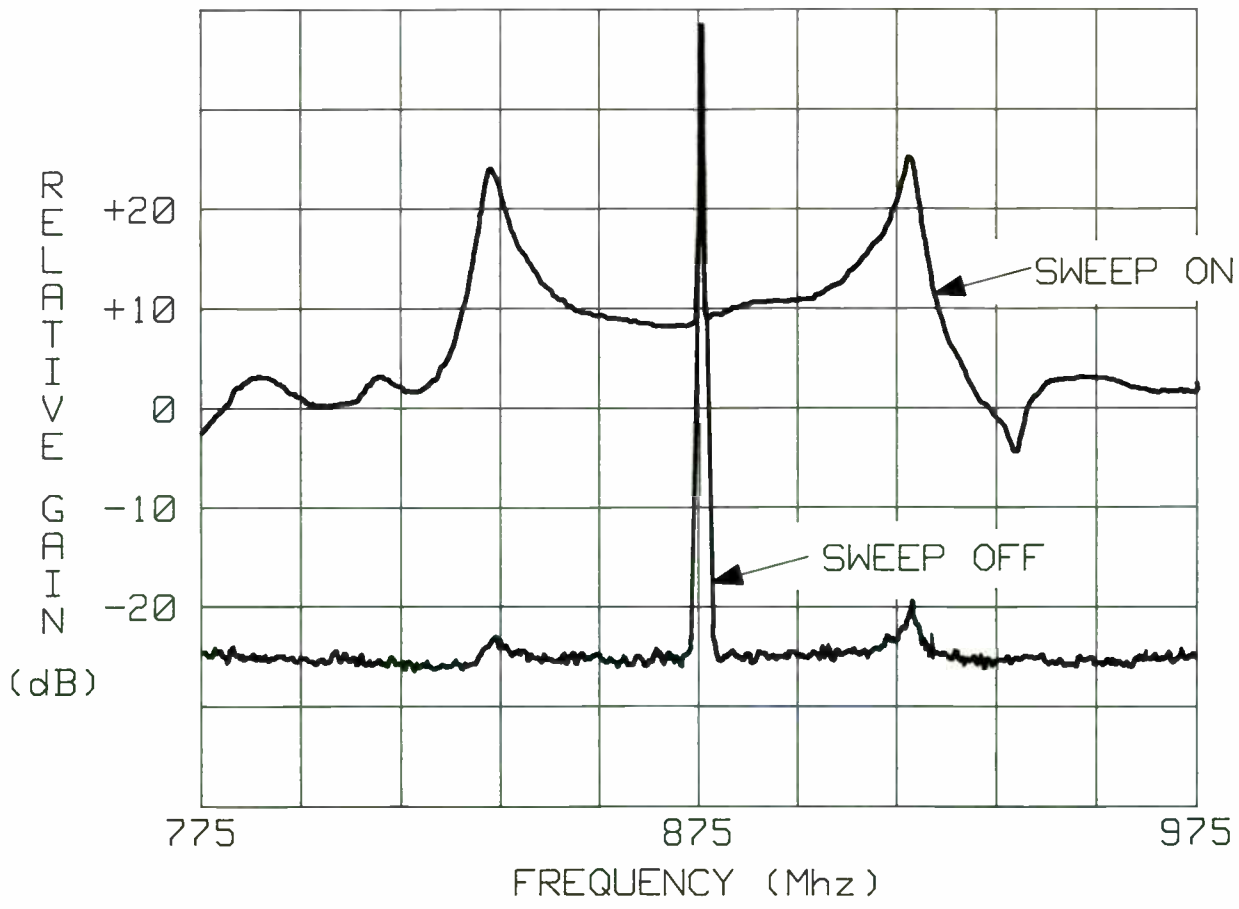


FIGURE 2

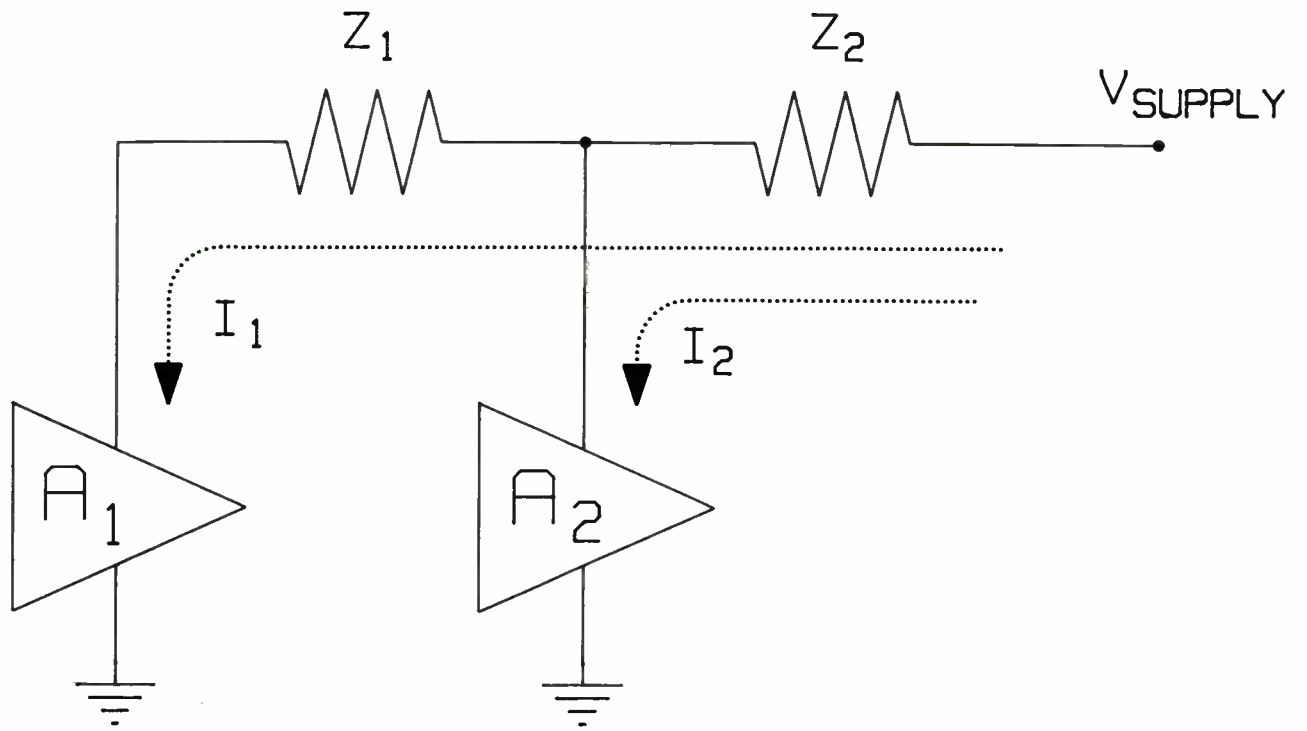


FIGURE 3

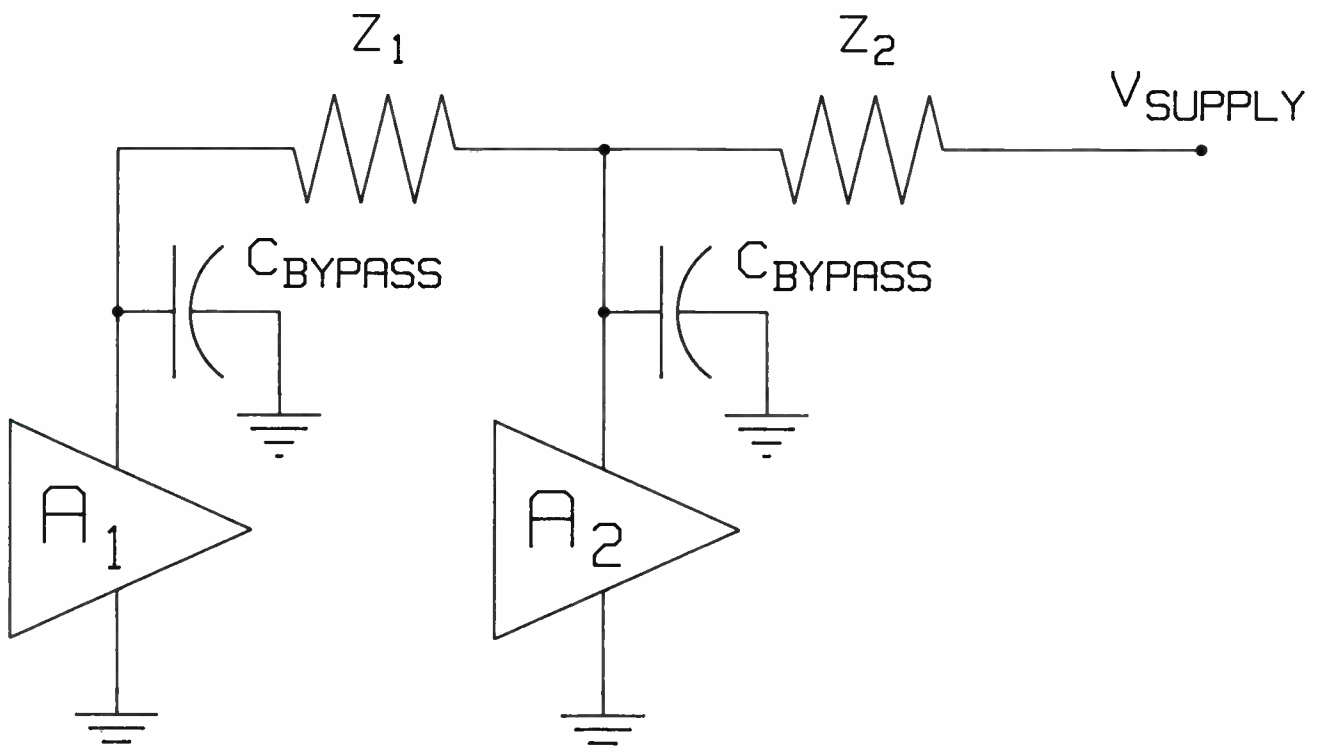


FIGURE 4

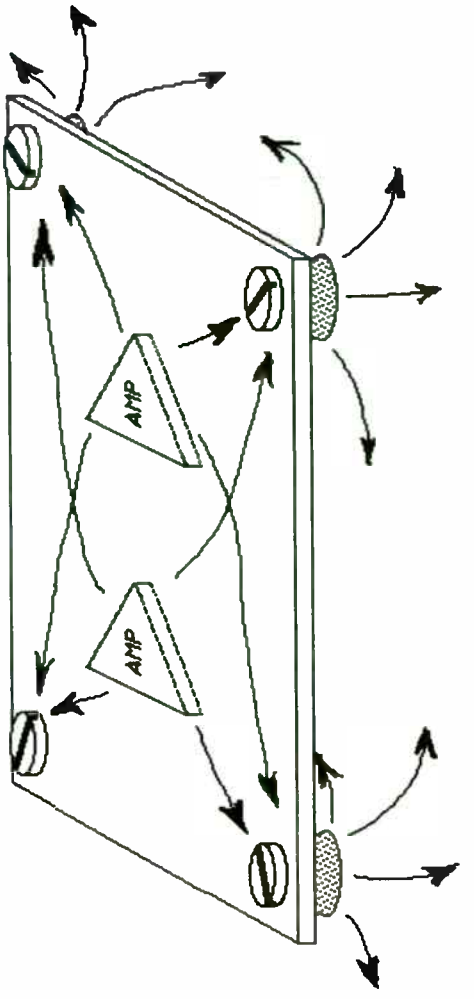


FIGURE 5a

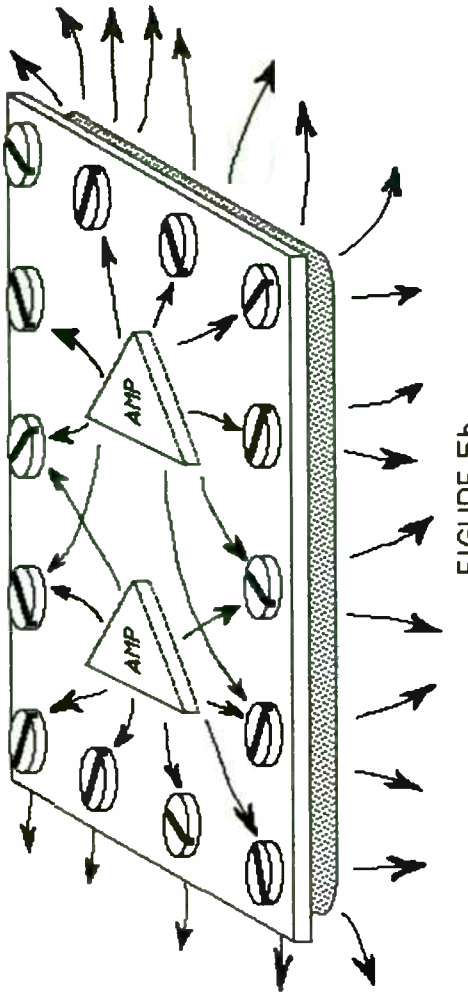


FIGURE 5b

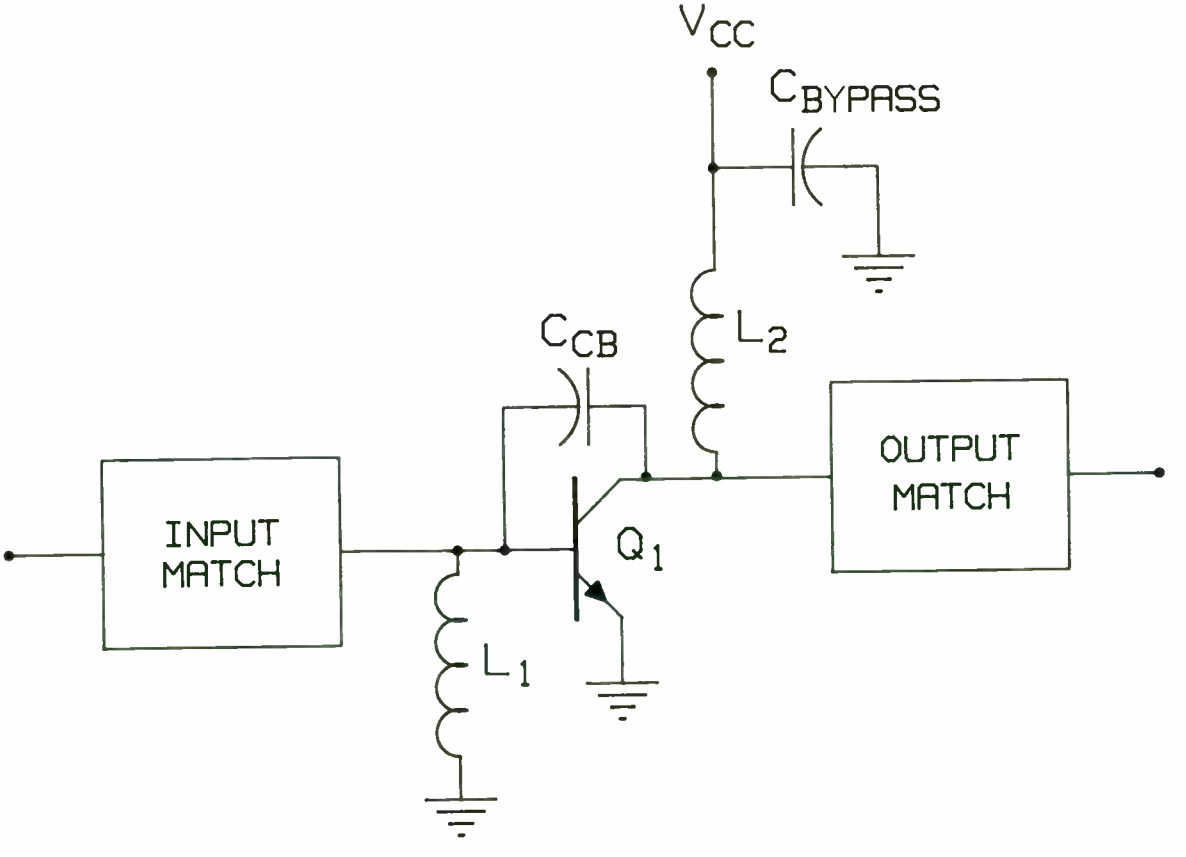


FIGURE 6

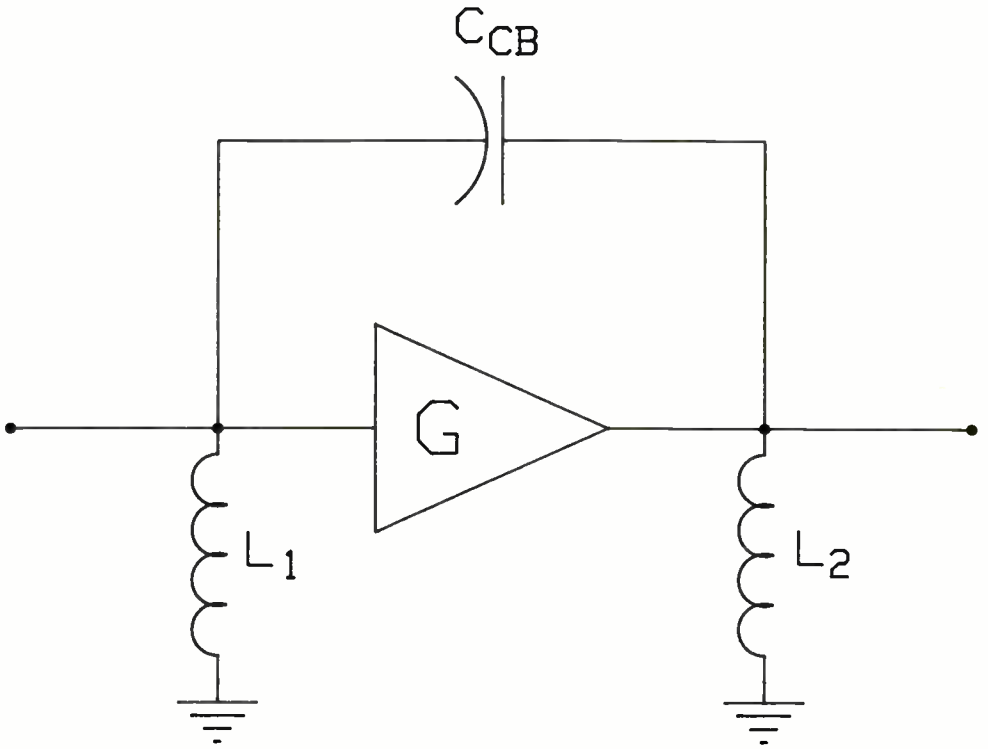


FIGURE 7

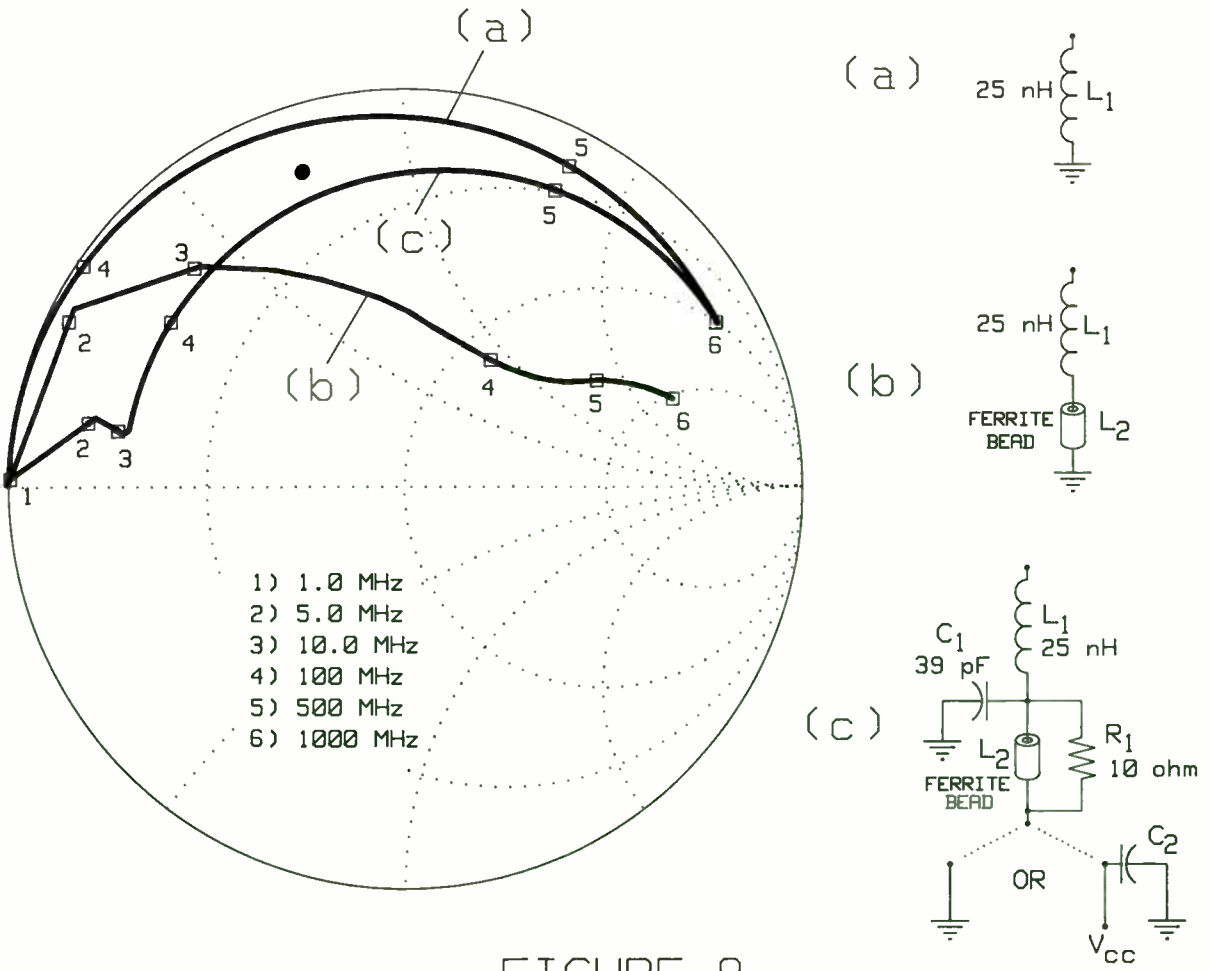
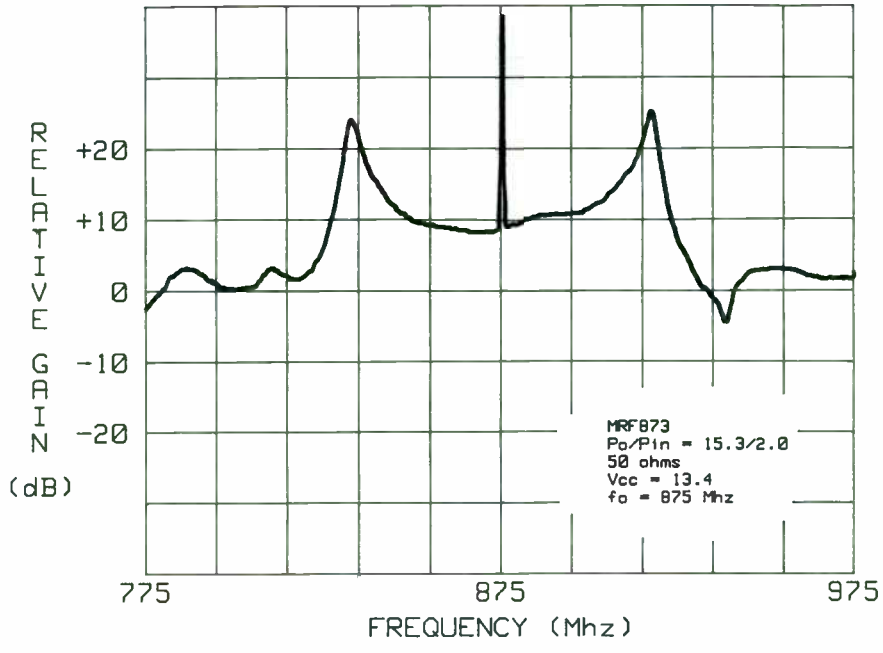


FIGURE 8



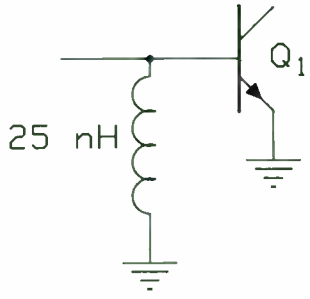
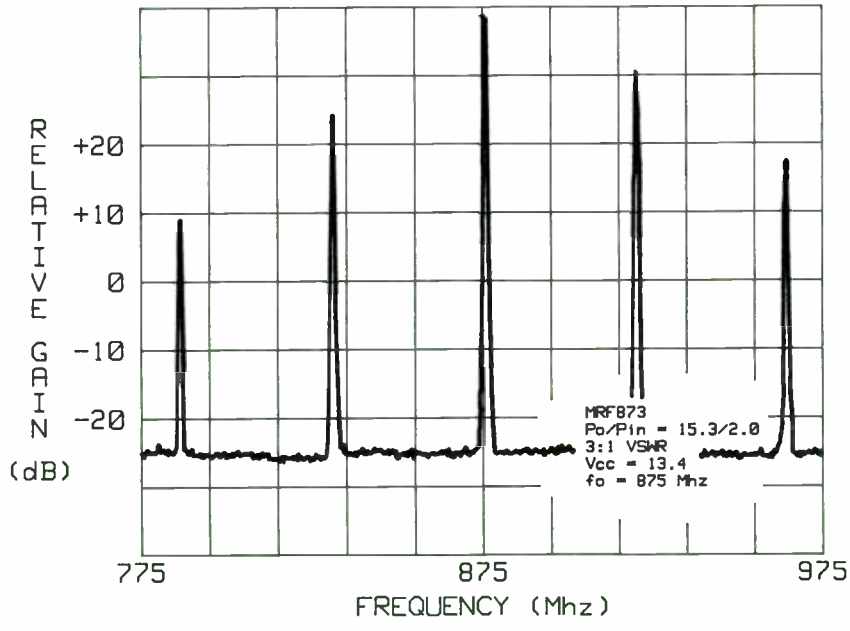


FIGURE 9



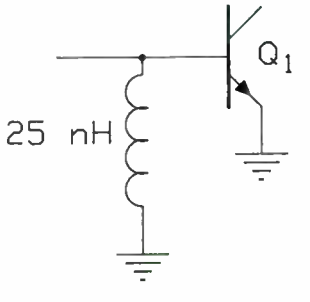


FIGURE 10

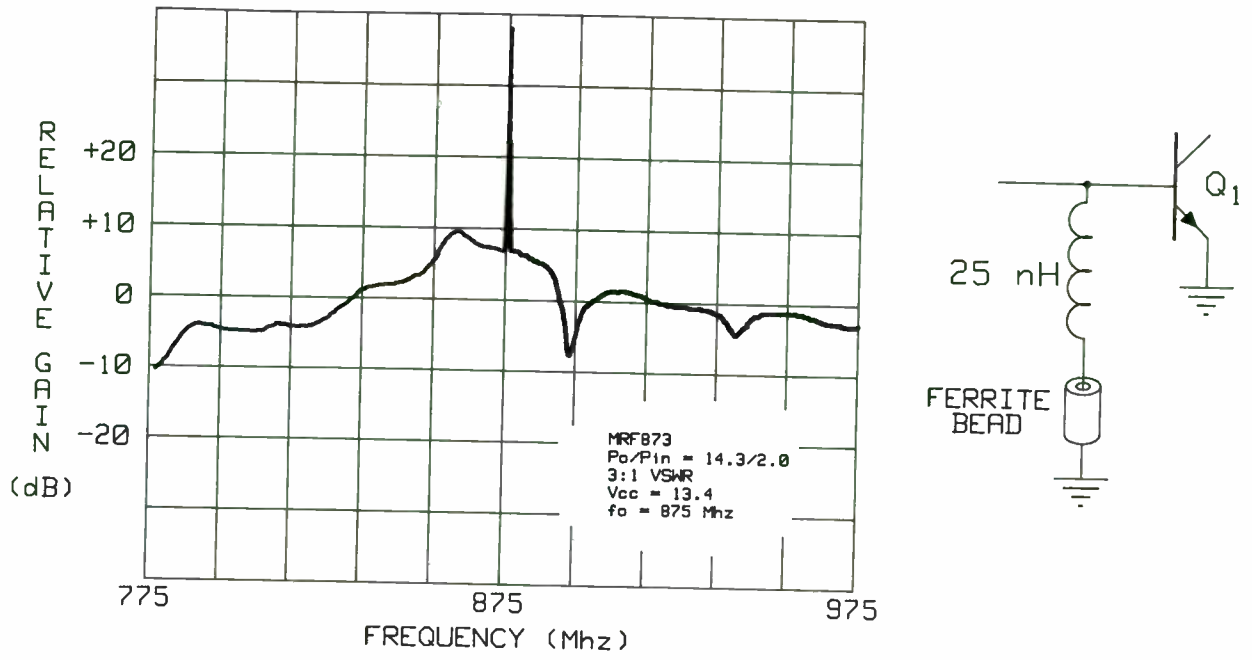


FIGURE 11

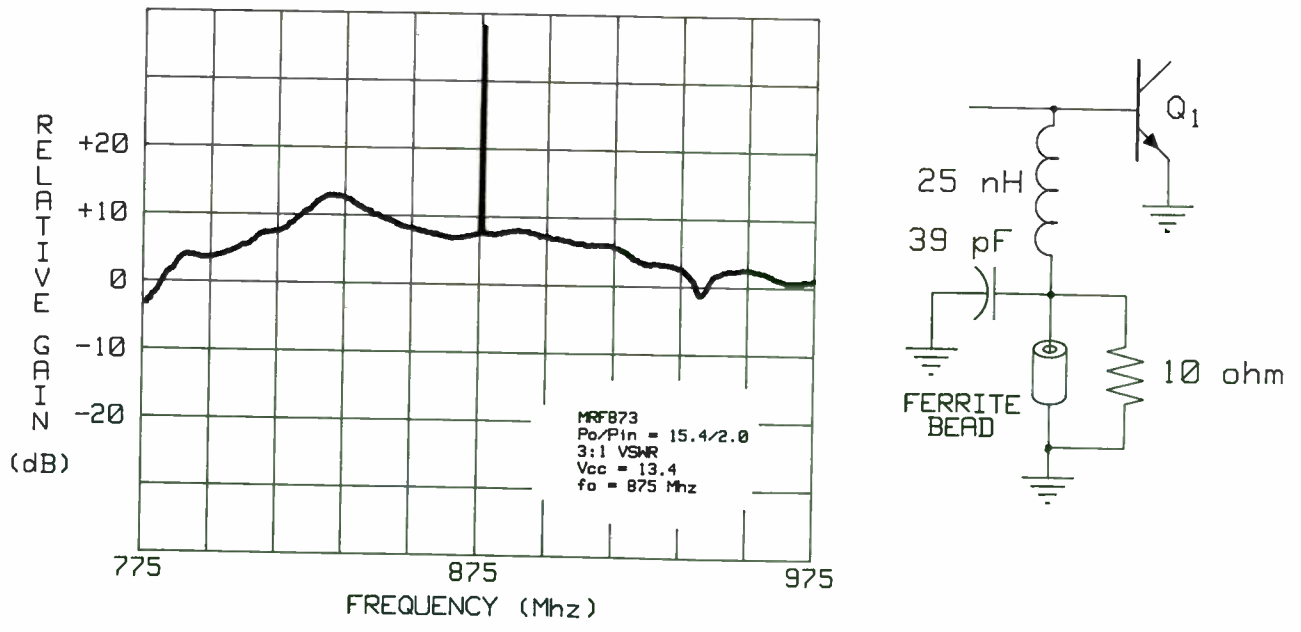


FIGURE 12

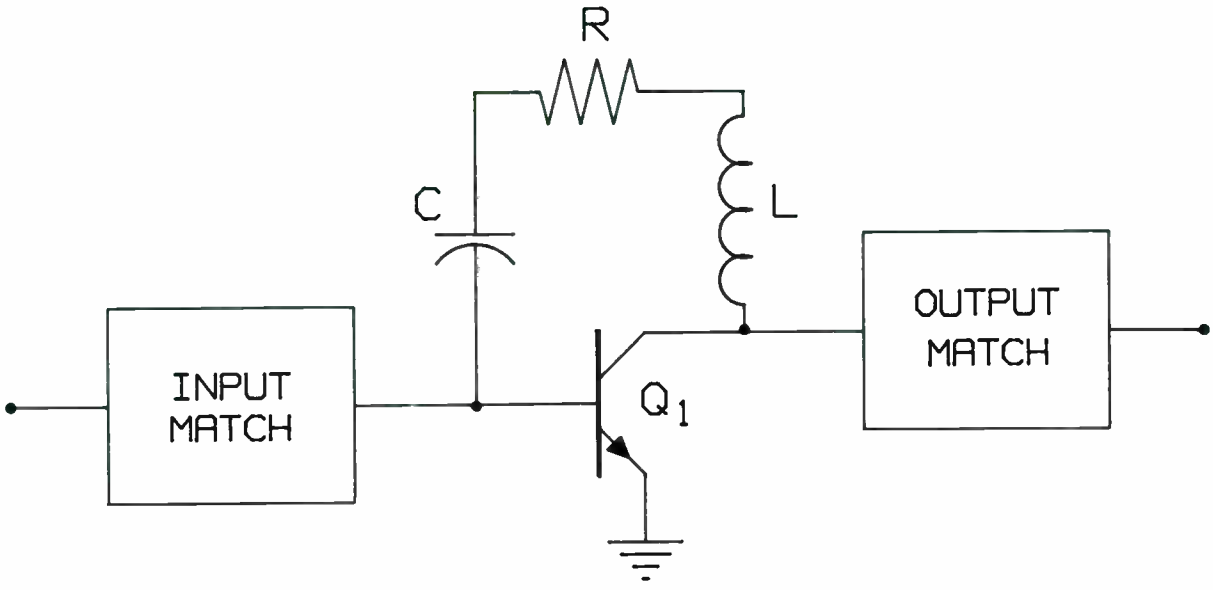


FIGURE 13

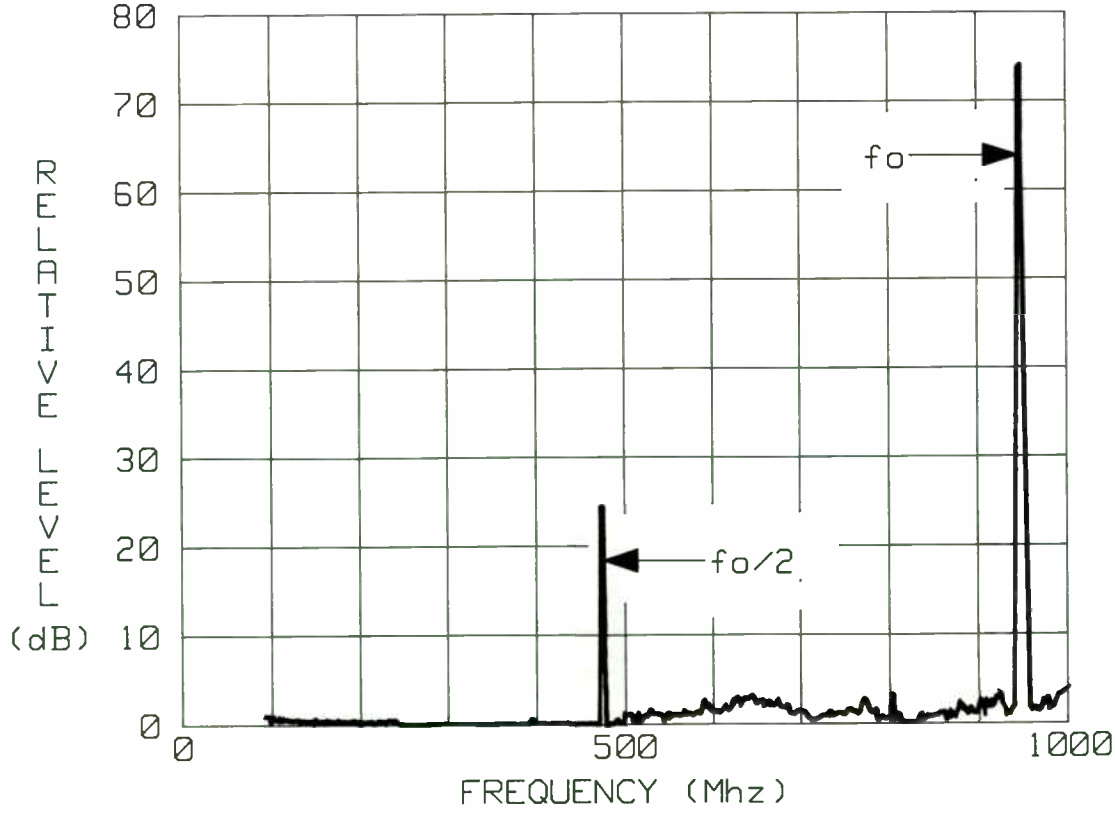


FIGURE 14

Concept and Design of an 800W UHF Power Amplifier

by

Harald W. Wickenhaeuser
Section Leader
R o h d e & S c h w a r z
Radiocommunications Division
Muehldorfstr. 18
8000 Munich 80, West Germany

1. INTRODUCTION

The development target was to expand the R&S power amplifier line by another model. The new addition covers the UHF range from 225 to 400 MHz /1/. The following are the main design objectives:

- The output power in AM mode is 200 W corresponding to 800 W PEP at full modulation (m = 1).
  In FM mode, the amplifier outputs 300 W.
  The values stated are continuous values valid for the whole of the specified temperature range from -20°C to +55°C with a MTBF > 8000 h.

- As the main field of application for the amplifiers will be radiocommunication systems, special attention has to be paid to the amplifier's co-sitability. This is particularly important for the immunity to transmitter backward intermodulation for which a value of -45 dB has been set.
  See section 4 (Applications) for further information regarding this sort of intermodulation. For collocation reasons, too, the amplifier should not degrade the exciter noise floor by more than 3 dB.

- The total conversion efficiency of $\geq$ 15 % which can be reached by amplifiers of this kind must be maintained for high modulation frequencies (data transmission).
  See section 2.3 for further information.

- In contrast to its predecessors, the amplifier must be of the general purpose type, in other words, it should be possible for non-R&S UHF exciters to supply the drive signal (use as booster).
  As an option (e.g. lab applications), the amplifier may be driven from signal generators.
  In line with the general-purpose philosophy, the amplifier system will have a 110V/220V single-phase power supply and a 48V battery supply for mobile, field applications.

- EMC requirements IAW MIL-STD 461 are to be met.

The requirements stated above are the basis for the design concept which is described in the following section.

2. DESIGN CONCEPT

2.1 UHF-Amplifier Unit VD490H1

Please refer to block diagram, fig 1.
The required output power and the RF power transistors that are svailable at the time of development determine the basic design concept for a transmitter power-amplifier. Taking the internal losses of about 2 dB in the harmonic filter, the directional coupler the circulators and the combiners into account, the required output power of 800 W PEP at a VSWR of 2 can be obtained with an installed power of about 1.7 kW. At present, the best UHF-transistors have an output of about 110 W, meaning that 16 transistors are required to give a total power of 1.7 kW. Internally each of these transistors comprise two single-transistor push-pull stages which are sealed in a recently developed twin case which due to its form is called also Gemini housing.
Two transistors of this kind, i.e. a total of 4 transistor systems form an amplifier module. See section 3.1 for further details.
Using a 6 dB Wilkinson power splitter/combiner, four amplifier modules are joined together to form a power amplifier block. On the output side, the 6 dB combiner is followed by a high-power circulator whose purpose is to present a uniform real load to the output stages and to considerably improve the backward intermodulation characteristics of the unit.

The output power from both amplifier blocks is added using a 3 dB Wilkinson combiner. After passing through the harmonic filter and the measuring directional-coupler, the output power is applied to the antenna connector. An T/R relay or a PIN diode switch are optional amplifier components.

Combiner technology, the circulator and the harmonic filter will be dealt with more closely in section 3.

Similarly, when internal losses of about 0.7 dB are taken into account, a driver power of about 280 W is required for the two amplifier blocks. Two amplifier modules, the same as the modules in the output stages, provide this power easily linked with a very good linesrity. Due to this two-stage amplifier configuration, the unit can be driven at its nominal power by any UHF exciter on the market. A preamp for driver powers down to 100 mW and a PIN diode regulator are availsble as options; they can both be accommodated in the same rackmount.

Monitoring of important psrameters are dealt with in the following sections.

2.2 Power Supply IN 490

Please refer to block diagram, fig. 2.
Essentially, the design concept for Power Supply IN 490 is determined by the following technical constraints:

- Collector voltage range: 8 to 28 V
- Current drain: max. 2 x 75 A
- Modulation range: 0 to 10 kHz
- AC supply: 110 V/220 V, single phase

An analysis of these contraints suggests that a secondary switched mode power supply with a high switching frequency is the best solution. On the one hand, a high switching frequency gives a good modulation range, and this is desireable, on the other hand, high switching frequencies give rise to enormous problems (EMC, losses, etc.). A switching frequency of 400 kHz was decided on as a good compromise.

At this point, one should not skip over the unforeseen difficulties that arose when the collector current supply was being developed. Apart from problems with regulation, mastering the high currents and the associated EMC problems was exceptionally difficult; but one should not forget that we are designing a "4.5 kW AM medium wave transmitter" where the transmitter frequency is subsequently rectified and undergoes 80-dB filtering.

This "medium wave transmitter" will comprise identical power modules I and II. The pulse width modulators (PWM), which are accommodated in the power control submodules, are synchronized by means of a common crystal oscillator in the control unit module.

The modulator module supplies the PWM with the AF modulation signal.

The interworking of the modules mentioned above will be described in more detail in section 2.3.

See section 3.5 for a circuit description of the collector current supply.

Apart from the size of the mains transformer, there is nothing special about the other modules, for reasons of economy, they will only be mentioned in passing.

A automatic switch on facility, essentially containing an inrush current limitation circuit has been provided to stop the AC mains circuit breaker blowing when the amplifier unit is switched on. This modular subgroup has been assigned to the transformer module.

2.3 The complete VD 490 plus Exciter

Please refer to block diagram, fig. 3.

First of all, the case where the VD 490 H1 and the IN 490 are connected together using an R&S Transceiver XD 432 as exciter will be considered.

The RF signal flow is straightforward.

Amplitude modulation is carried out by means of comparing the demodulated amplifier RF output voltage with the modulating AF voltage to produce an error voltage. To this purpose, the forward voltage coupled from the measuring directional coupler is demodulated and amplified, this voltage has a DC component proportional to the carrier power and an AF signal proportional to the modulation depth. The modulator in the IN 490 compares this voltage with the modulating AF which gives the nominal value for the modulation depth and the DC component which gives the nominal value for the carrier power.

The error output signal from the differential amplifier controls the instantaneous output power from the exciter so that the condition below is fullfilled:

Demodulated envelope = modulation AF

This modulation technique makes amplitude modulation with a high degree of modulation, low distortion and high efficiency possible. To further increase the units's efficiency, the collector voltage of the power amplifier is matched to the actual instantaneous power requirements by the modulating AF voltage. Thus amplifier unit's total conversion ef-

ficiency ist about 50% greater than in pure linear operation with a fixed collector voltage. This mode of operation has to be chosen in connection with non-R&S exciters.

Each rackmount has a monitoring facility that continuously checks important parameters e.g. supply voltages and-currents, amplifier symmetry, output power and-VSWR, deviations of rated loop values, internal temperatures. If there is a fault, the amplifier performance associated with the parameter in question is degraded to the point where further operation without jeopardizing the equipment is possible. The fault location is internally stored and displayed on the front panel.
A three-step, temperature-controlled blower cools the rackmounts.

EMC problems will be discussed in section 3.6. At this point, it should be pointed out that not any cable will do for connecting up the rackmounts. This is particularly true of cables carrying up to 150 A for the collector current supply. The characteristic impedance of this cable is "clearly noticeable" in the modulated collector voltage mode.

3. CIRCUIT DESIGN DETAILS

3.1 UHF amplifier module

Please refer to circuit diagram, fig. 4.
As mentioned previously, ten of these modules are used in the amplifier rackmount. It therefore forms the "core" of the amplifier and will be described in detail in the following text.

The module comprises two push-pull amplifier stages, connected together using 3 dB-90° hybrids at the input and output to form a balanced amplifier. The balanced amplifier, which has already been described and derived in the literature /2/, has an extremely constant, real input impedance (in this case 50 Ohm), if the hybrid at the module input has identical terminations at the outputs marked "0°" and "90°" (the terminations may also be complex). The two push-pull amplifier stages meet this condition almost perfectly. On the other hand, when considering the output side, both amplifiers always "see" two identical loads.
A piece of 50 Ohm coaxial cable is used for the first step in input-side matching, i.e. the transition from single-ended to push-pull operation. The following 620 pF chip only provides DC decoupling, the 10 Ohm chip which is connected in parallel gives stability at lower frequencies. A four-to-one line transformer, realized using 10 Ohm coaxial cable, down transforms to a 12.5 Ohm level. A combination of stripline and discrete components perform transformation to the low base input-impedance. The input transformation is optimized for the upper operating frequencies.

On the transistor output side, a stripline circuit compensates for the transistor output capacitance at operating frequencies at the centre of the range. One-to-four transformation, DC decoupling and the push-pull - single end transition are similar to the analogous circuits on the input side.

Apart from the collector voltage supply, which contains chokes and by-pass capacitors that come into effect at various frequencies, every amplifier section requires a base bias supply as the amplifier stages are being operated as a class AB linear amplifier taking a quiescent collector current of about 200 mA per transistor. The base bias supply has temperature compensation, and can be set separately for each transistor. It is therefore not necessary to look for transistor pairs with identical characteristics. This is one of the measures to use standard components in the amplifier wherever possible.

The transistors are accommodated as a push-pull pair in the already mentioned Gemini housing. At present, this kind of housing is unrivalled, as far as electrical and thermal characteristics go, for transistors in this frequency and power range.
It is worth mentioning that not one transistor of this type met an untimely end during the development phase.

In conclusion, a summary of this module's electrical data:

Peak output power:		> 200 W
Gain:		> 8.5 dB
Frequency response	225 to 400 MHz:	< 0.5 dB
1 dB compression:		> 180 W
Conversion efficiency:		abt. 50%

A photo of the amplifier module is shown in fig. 5.
The formation of the modules inside the rackmount can be seen in fig. 10.

3.2 Splitter/combiner

As it does not seem likely that single UHF transistors for the kilowatt power level will be available in the foreseeable future, the development engineer is forced to familiarize himself with the splitter/combiner methods whether he wants to or not.

In the final analysis, the choice of the splitter/combiner method will play a large role in determining the electrical quality data and the production costs.
Let us now discuss splitter/combiners, taking the 6 dB Wilkinson model as an example. Four of them are incorporated in the VD 490 H1.

The splitter/combiners in the forerunners of the VD 490 H1 were realized using semi-rigid coaxial cables and own-manufacture compensating thin film resistors on ceramic substrates which were soldered on special milled heat sinks. Although this module satisfied the electrical requirements, it was difficult to produce, required a lot of testing and was the cause of many faults yet. When the project was calculated, it was found that this solution was no longer viable.

As there was no suitable non-R&S component available, we started our own development project which was opposed by the experts. The first tests were very promising and during the later design stages, the Super Compact circuit analysis and optimization program /3/ proved to be a useful tool for experienced RF design engineers.
As shown in fig. 6, printed stripline-triplate technology was used for the final implementation. The compensation resistors were commercially available. To save space SMA technology was used for the connectors, even though from the electrical point of view BNC would do the job easily.

Up till now, we have only received consistently positive batch production reports. The combiner has a very narrow parameter spread and is checked without any adjustments on a GO/NOGO basis. The progress made towards the solution described in the introduction is spectacular.

The 3 dB combiner (see fig. 7) four of which are used in the amplifier is also realized in stripline technology.

3.3 Circulator

To prevent high-frequency pollution in the form of transmitter backward IM products, circulators are to be fitted to all batch-production models as standard. As there was no product on the market which met the requirements for our application when development was started and our resources where insufficient to do our own development work, a high power circulator was developed to the batch-production stage in close cooperation with two leading circulator manufacturers; at the same time development work was being done on the VD 490.

The required wide bandwidth, the insertion loss, the isolation, the temperature characteristics and the immunity to external magnetic fields were the main problems.
Fig. 8 shows the batch-production model from one manufacturer.
The excellent specifications are given below:

Frequency range:	225 to 400 MHz
Insertion loss:	< 0.7 dB
Isolation:	20 dB
Continuous power:	200 W
Peak power:	470 W
Temperature range:	-20°C to +75°C

The circulators are arranged on the driver aide of VD 490 H1 (see Fig. 11).

3.4 Harmonic filter

As the VD 490 is primarily a radiocommunicationa set, harmonica and spurious signal must be suppressed by > 80 dB. To realize a suitable harmonic filter at the amplifier output, the usual UHF/SHF problems, in particular thermal problems, have to be solved. For a filter of this kind, we are talking about removing heat from coils and capacitors at a rate of abt. 40 W.

This problem was solved by using cross capacitors with a beryllium oxide dielectric which is a very good thermal conductor. These capacitors, as fig. 9 shows, are soldered onto a silver-plated copper base. The base is sited in the set to maximize heat extraction. Both the heat produced by the cross capacitors and that from the coils which are wound with thick wire is conducted away very efficiently. It is practically impossible to unsolder the coils with a solder iron without preheating the whole filter. Even in continuous operation at an ambient temperature of 55°C the thermal/electrical stability of the filter is excellent.
The harmonic and spurious signal requirements stated above will be met under all operating conditions.

3.5 Collector supply

The collector supply comprises the following modules:

Transformer module

Control unit

and Power modules I & II

Each power module comprises the following submodules:

High-current unit

Power control

Please refer to fig. 2.

The dominant component in the transformer module, at least from the weight point of view, is a 4.7 kVA toroidal transformer. After the secondary voltage has been rectified an smoothed, a unregulated voltage of about 50 V is applied to the input capacitors. The transformer module also contains the EMC filter (see section 3.6) and the inrush current limiting automatic switch-on facility. Without such a circuitry most attempts to switch on the system would result in triggered AC mains circuit breakers, due to the extremely high phase dependent starting currents of higher powered toroidal transformers.
The transformer module can be clearly seen in fig. 12.

The only component that the control unit contains which is essential for understanding the collector supply is a 6.4 MHz crystal oscillator. This is used to synchronise the clock oscillators in the power control submodule.

Fig. 14 is a simplified block diagram of the power module.
The 400 kHz clock frequency required by the pulse width modulator is obtained by dividing down the 6.4 MHz clock oscillator frequency. If this module is inserted in an IN 490 its oscillator frequency is synchronised by means of the crystal oscillator mentioned previously. This is necessary to prevent the two power module units in the IN 490 from beating, on the other hand, when one of these units is removed for servicing it is still fully operational, albeit without synchronisation.
The pulse width modulator generates interleaved 200 kHz pulses at two separate outputs. The pulse width is proportional to the amplitude of the AF control voltage. By means of a flip-flop, the 400 kHz discharge pulse is derived from two 200 kHz pulses. The further use of these pulses will be discussed in the final section on the high current unit.

It is difficult to dimension overcurrent protection for a power supply of this kind. On the one hand, when a "catastrophic" short occurs the protection circuit must react extremely quickly (< 10µs) to prevent damage that will rapidly lead to complete destruction of the module.

On the other hand, the power supply must be protected from long-lasting overcurrents caused by faulty modules, without, however, being affected by modulation peaks.

This problem was solved by using two current sensors with different current thresholds and response times. The comparator, which has been dimensioned for short-circuit current I_1, reacts within a very short time τ_1. The comparator that has been designed for overcurrent I_2 has a considerably longer reaction time, τ_2.
It is always true that $I_1 > I_2$ and $\tau_2 >> \tau_1$.
Both the overcurrent protection circuit and the temperature monitoring facility, designated , react so efficiently to a fault at the control input of the PWM that the pulse width is immediately reduced to zero.

A few special circuit features in the high current unit are illustrated with the detailed circuit diagram in fig. 15.
Basically we are dealing with a buck converter with a 400 kHz clock frequency. Three power MOSFETs (V1, V2, V3) are connected in parallel as a power switch. The gate-side drive, which uses 400 kHz pulses with different pulsewidths, is provided by the arrangement V15,T1; V16,T2 plus V17,T3, V14.
V15,T1 and V16,T2 are driven on the input-side by the interleaved 200 kHz pulses which have been mentioned previously. On the output-side they are connected in parallel. In each case, the circuit is a 200 kHz forward converter, each of which has a max. duty cycle of 50%. In combination they give a 400 kHz forward converter which, theoretically, can reach a duty cycle of 100%. The well-defined trailing edge of the 400 kHz gate-control signal for V1 to 3 is generated by the discharge pulse using V17,T3, V14.

V1 to V3 are connected in parallel by means of the triple choke L2. This choke has an inductance that is inversely proportional to the load current to prevent tearing off-current operation during low load current periods. When the transistors are off, the load current flows through the free-wheel diodes V4 to V6.

C20 to C24 are smoothing capacitors carrying a already moderately "clean" DC voltage. However, because spurious signals must be 80 dB below the carrier for radiocommunications equipment and because the 200 kHz and 400 kHz components of the collector voltage appear directly as high frequency sidebands in the transmitted signal, this DC voltage has still to be very carefully filtered.
Dimensioning a filter of this kind is a difficult problem from many points of view.

- Both the input and output (= RF section) impedances are highly non-linear and complex.

- The series coils must be designed to handle any load-current peaks (= 75 A) without saturation effects.

- The filter has a very destinct effect on the high current unit's dynamic characteristics.

Each filter modification influenced modulation range and loop stability in a direction difficult to predict.

The Cauer low-pass in the diagram has poles at 200 kHz and 400 kHz. It is the result of exhaustive, computer aided simulations, many lab-tests combined with the experience of "old foxes".

Finally the current-sensing resistor R40 and the output voltage limiting diode V40 should be mentioned. The former is part of the already described overcurrent protection circuit.

The development of the collector supply for the UHF power-amplifier was the most difficult and resource-consuming stage in development. Mastering the high currents, the associated EMC problems and the modulation dependent control of the collector voltage gave rise to many unexpected problems, in particular problems with nonlinear loop characteristics. The gigantic current capability of the transformer module caused electrical mayhem during development by melting down power diodes and even destroying whole modules, luckily only the electronics suffered.

The realized version of the high current unit shows fig. 16.
Two such units are used in IN 490 (see fig. 13) to supply the total collector-current.

3.6 Some specific EMC problems

As mentioned in the introduction, the complete VD 490 amplifier system, comprising the UHF amplifier VO 490 H1 and power supply IN 490, is a piece of radiocommunications equipment and as such must meet the EMC requirements of MIL-STO 461.

In this respect, the RF section VD 490 H1 caused few difficulties because potential EMC problems are well-known and could be anticipated. In this way, all kinds of low-cost solutions were avoided. Working with EMC problems shows the truth of the saying "You can't get something for nothing" time and time again. High-quality feedthrough filters were always used even though mounting these components did not facilitate the production process. It goes without saying that coaxial cables and connectors with a high shielding factor were used.

All this was checked using an "RF sniffer" and any leaks were "plugged". For example, it was possible to increase "RF-tightness" of the 1:4 splitter/combiners, which were mentioned before, beyond the required specifications by arranging a row of contacted-through holes along the outer edge of the PCB. This was an EMC measure with a good cost/benefit ratio.

Oue to the meticulous mechanical design, it was no problem removing the "slot antennas" near the blower and the instrument's top-cover, the only difficulty that was left.

It was much more arduous "taming" the spurious signals caused by the power supply unit IN 490. You will recall the "4.5 kW 400 kHz medium wave transmitter" 200 kHz apectra with unbelievable amplitudes contaminated the collector voltage, the supply line and practical all the internal lines. It goes without saying that the RF spectrum was also unacceptable.

As our development work was far off-target, only a coordinated application of a variety of measures could rectify the situation. The most important steps are outlined in the following:

The high current unit PCB was redesigned in accordance with strict RF design rules. Apart from realisation of a central ground point, it was particularly important to minimize the areas enclosed by printed lines and wires carrying high currents.
The harmonics filter in the high current unit also had to be redimensioned. I have already discussed the problems related to this topic.
Finally, the whole module had to be placed in a screening cage.

The supply transformer was replaced by a version with a more effective copper tape shield for the interfering frequencies between the secondary and the primary side.

There was no question that an efficient supply EMC filter was required. Commercially available filters also turned out to be unsuitable on one or more of the following grounds:

- Attenuation in the range 150 kHz to a few MHz
- Dimensions
- Current loading 35 A_{rms} @ ϑ_a = 75°C

None of the filters on the market were able to comply with the last point. However, as the equipment can be powered from a 110 V AC line, too, it was imperative that this condition was met.

When we started to design our own filter not only had we to keep to the constraints mentioned above but we also had to comply with safety regulations (VDE) which are required by law. As far as the latter was concerned, component selection was critical.

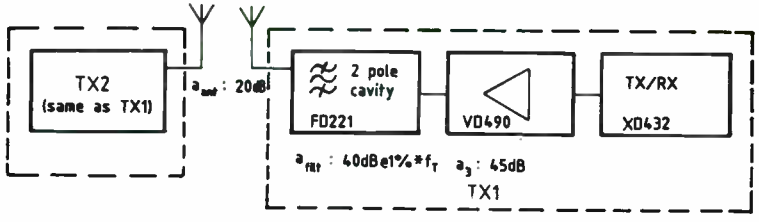
Both the EMC test to MIL-STD 461 and the safety test to VDE were passed succesfully before batch production was started.

4. APPLICATIONS

As mentioned in the introduction, the amplifier (fig. 17) is of the general purpose type. A few realized examples of the many possible applications are described.

4.1 Extreme collocation conditions

In a transmitter site with several 200 W UHF transmitters and a minimal antenna isolation of 20 dB, it was necessary to have a transmitter backward IM rejection of > 140 dB for a transmitter frequency separation of $\geq$ 1%, if reception is not to be impaired.



Example: f_{TX1} 300 MHz

f_{TX2} 297 MHz

IM 303 MHz

294 MHz

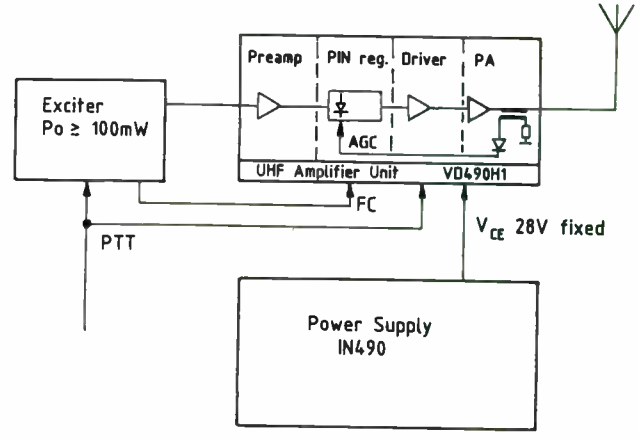
$a_{tot} = a_{ant} + a_{filt} + a_3 + a_{filt}$

$a_{tot} = 20\ dB + 40\ dB + 45dB + 40\ dB = 145\ dB$

It is easy to see that this problem can only be solved using filters and circulators.

Fig. 18 shows the final system /4/.

4.2 Booster for any (low-power) exciter



The preamp (available as an option) is used for exciter output-powers < 7 W.

The PIN regulator, which is also available as an option, compensates for the output power frequency response of the exciter. Level adjustment is only carried out after a frequency change (FC), this taking a few milliseconds when the PTT key is pressed for the first time. The voltage for the AGC is stored digitally until the next frequency change takes place.

4.3 40 kW EIRP UHF high-power transmission system

By combining all the components mentioned in this talk as shown in fig. 19, one obtains a 40 kW EIRP UHF transmitter system, if an antenna with a gain of about 10 dBi is used.
Transmitter systems of this kind are mainly used for military ECM/AJ applications.
A Technical Information sheet which gives an in-depth treatment of the system is available /5/.

5. CLOSING REMARKS

This talk has dealt with the most important design considerations for a 800 W PEP UHF amplifier.

The practical realization of the circuits, and the associated problems that were encountered during the development phase, were discussed using a few selected modules as examples.

Many interesting electrical details, e.g. BITE (Built-In Test Equipment), the modulator circuitry had to be left out, or could only be discussed superficially due to lack of time. The same is true for the large mechanical parts, in particular in IN 490. Optimising the thermal characteristics of the set-infrared thermography was used for the first time - would take up another lecture.

The applications described at the end of the talk were intended to emphasize the system-building and the general-purpose capabilities of the amplifier.

REFERENCES

1. Technical Information. UHF Power Amplifier VO 490
   Rohde & Schwarz, Mühldorfstraße 15, 8000 Munich 80, West Germany

2. Anaren Microwave Components Catalog No. 17, pp. 75-76
   Anaren Microwave, Inc., 6635 Kirkville Road, E. Syracuse, NY 13057, USA

3. Super Compact, User Manual
   Compact Software, Inc., 52 Hillcrest Orive, Upper Saddle River,
   NJ 07458

4. Wickenhäuser, H.: VO 490, a new UHF power amplifier for Radio Equipment
   Family 400.
   News from Rohde & Schwarz (1987) No. 117, p. 46

5. Technical Information. UHF High Power Amplifier System VO 400
   Rohde & Schwarz, Mühldorfstraße 15, 8000 Munich 80, West Germany

LIST OF FIGURES

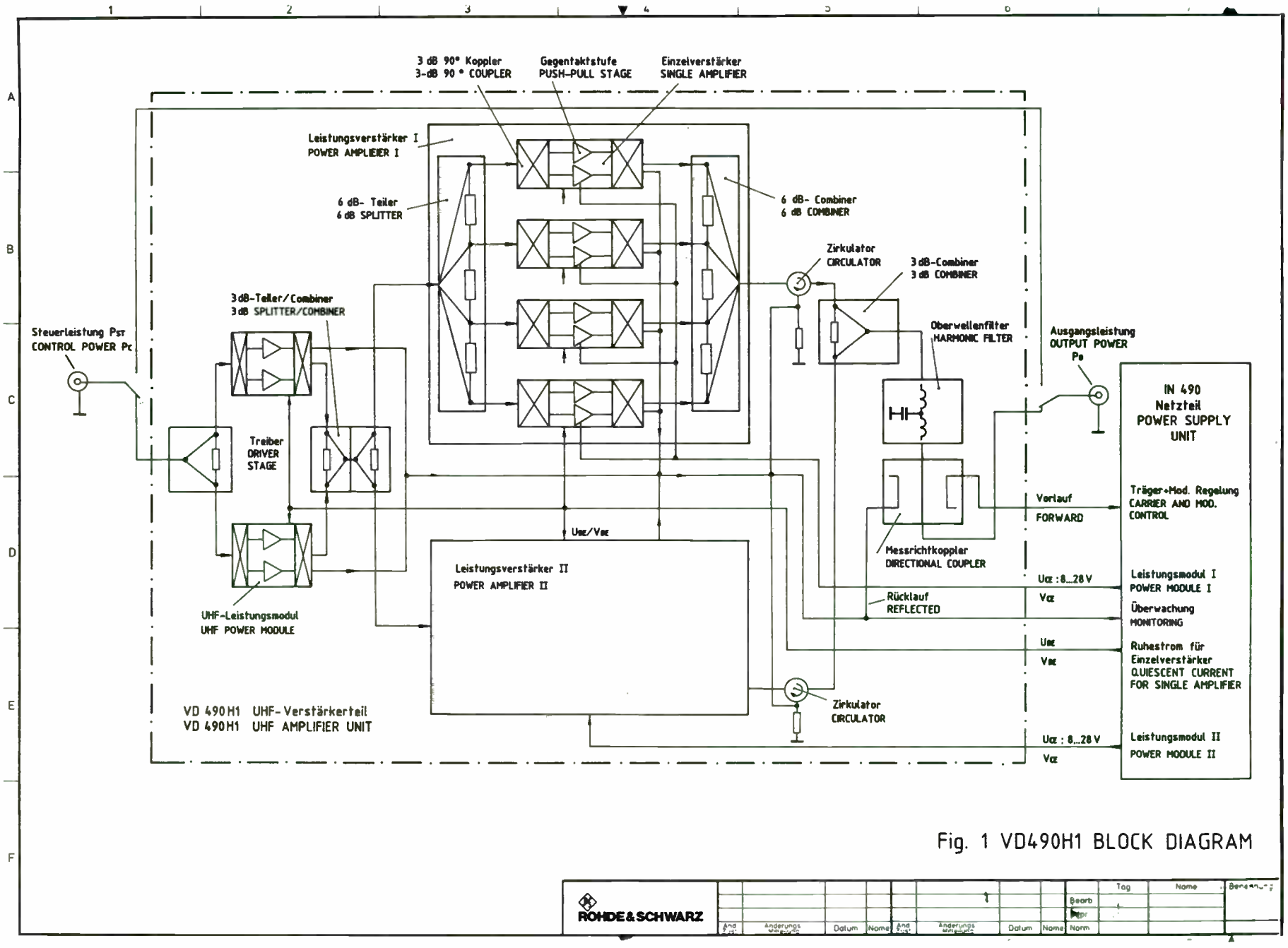


Fig. 1 VD490H1 BLOCK DIAGRAM

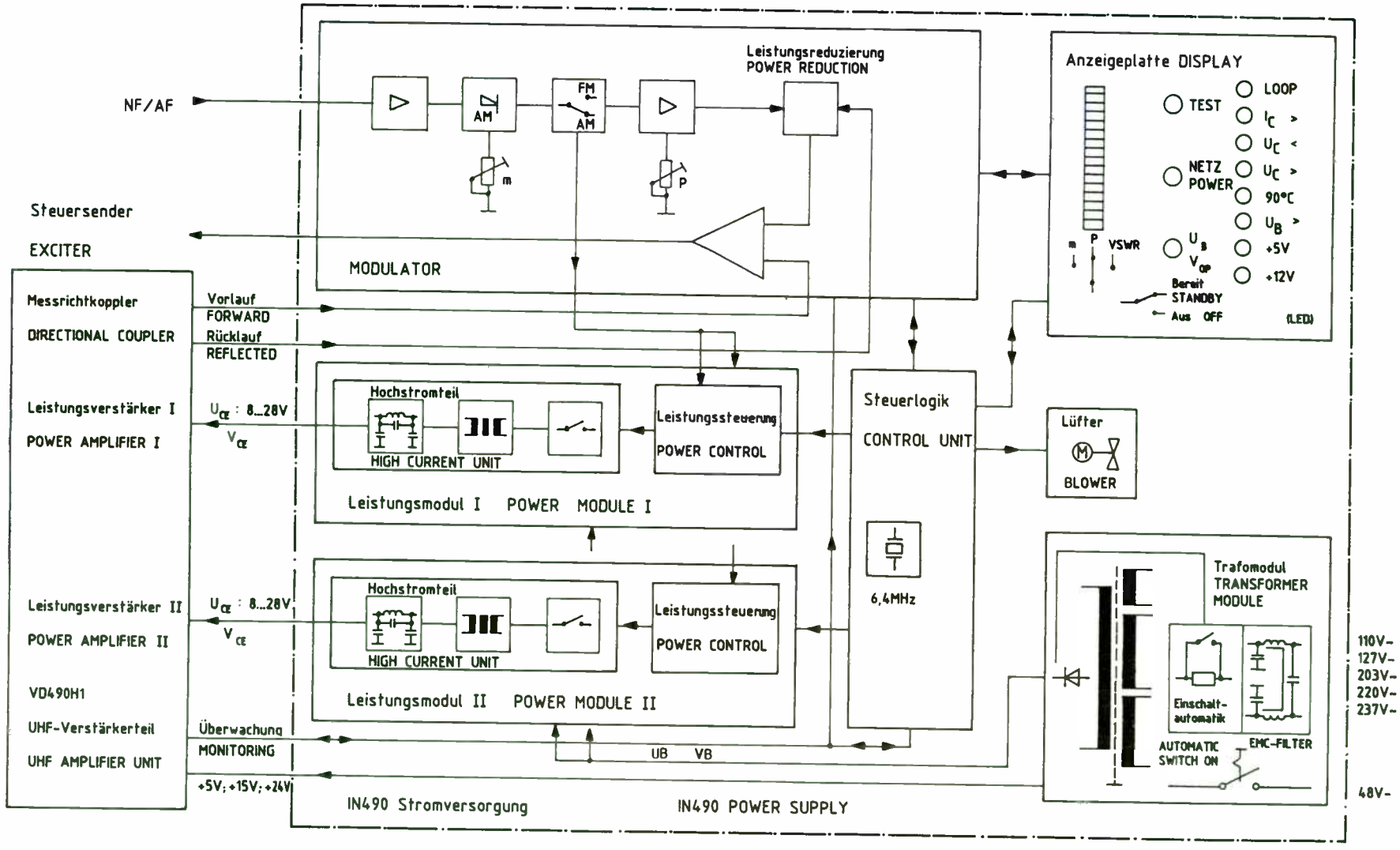


Bild 2 IN490 Stromversorgung Blockschaltbild
Fig. 2 IN490 POWER SUPPLY BLOCK-DIAGRAM

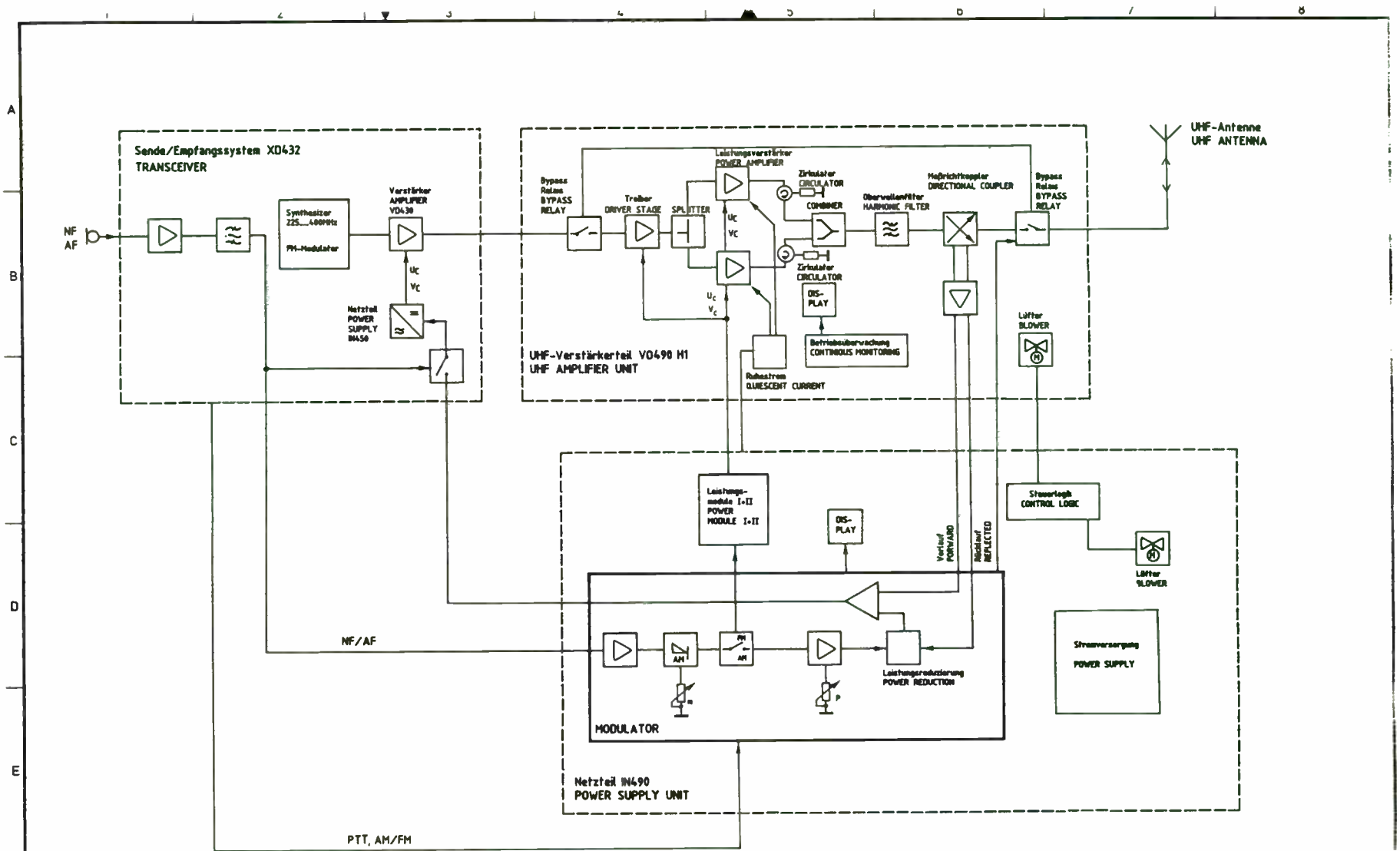
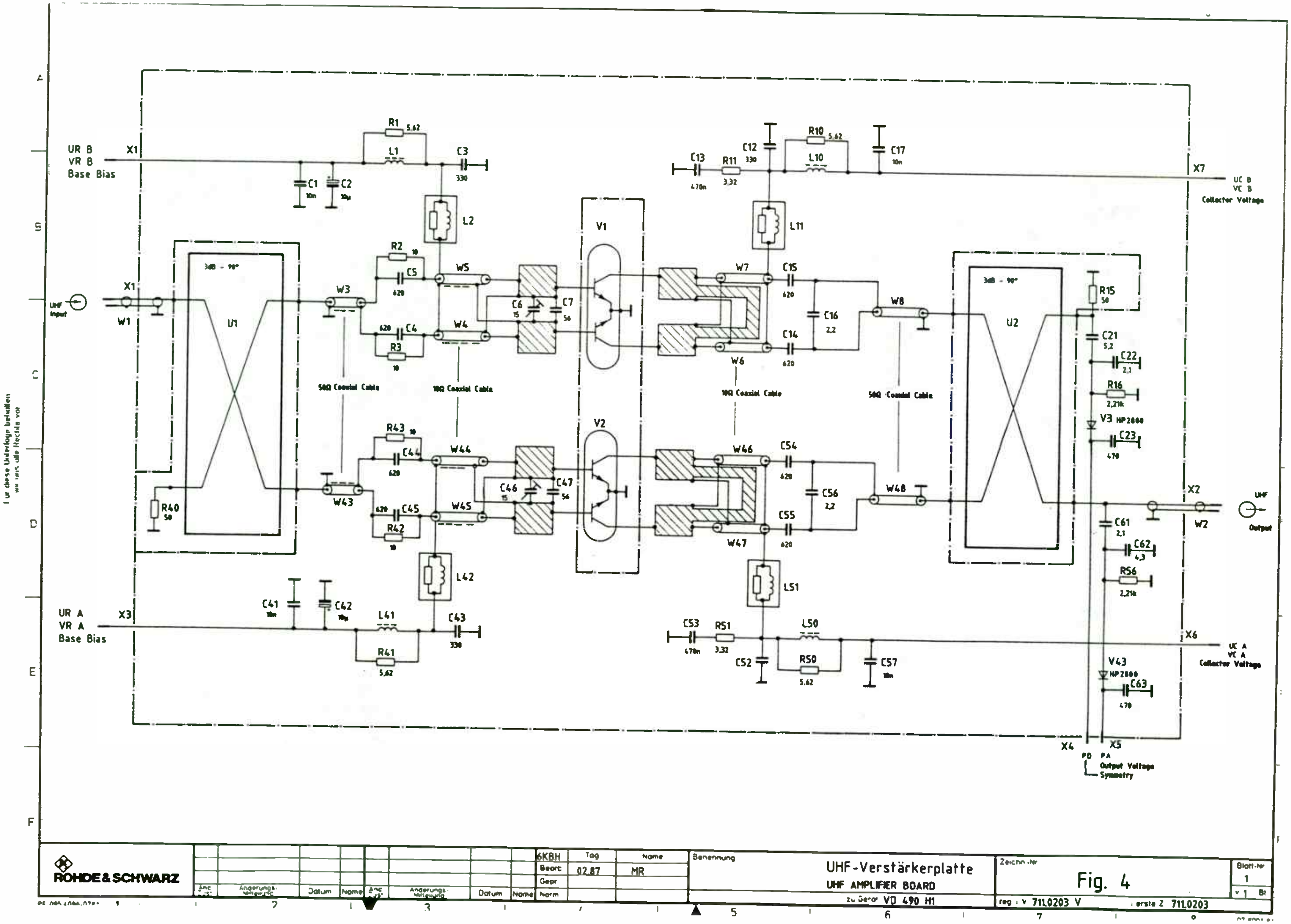


Fig. 3 VD490 WITH XD432 BLOCK DIAGRAM

ROHDE&SCHWARZ												Tag	Name	Benennung		Zeichn.-Nr.		Blatt-Nr
											Bearb			Blockschaltbild VD 490 mit XD 432				v 81
											Gepr.							
And. Zust.	Änderungs-Mitteilung	Datum	Name	And. Zust.	Änderungs-Mitteilung	Datum	Name	Norm					zu Gerät	reg. i. V.		erste Z		



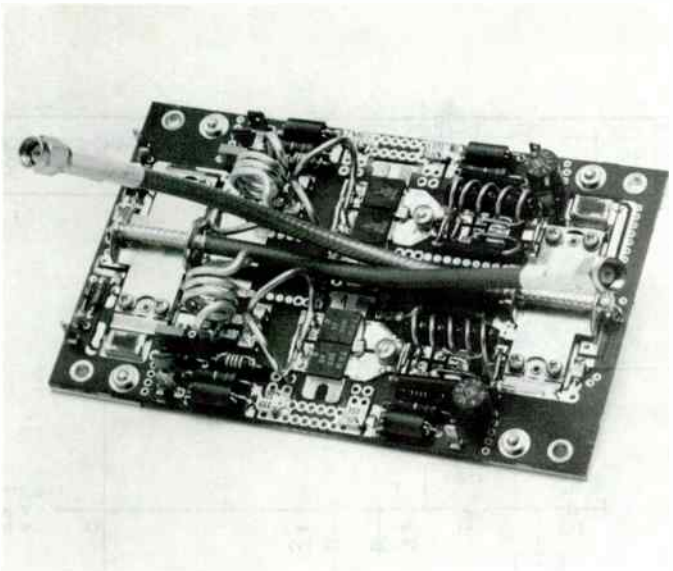


Fig. 5 UHF-Amplifier Module

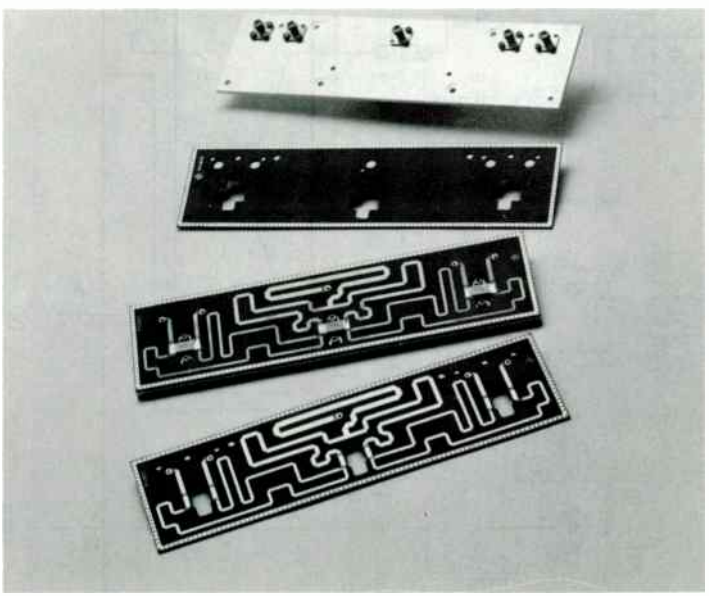


Fig. 6 1:4 Wilkinson
Splitter/Combiner

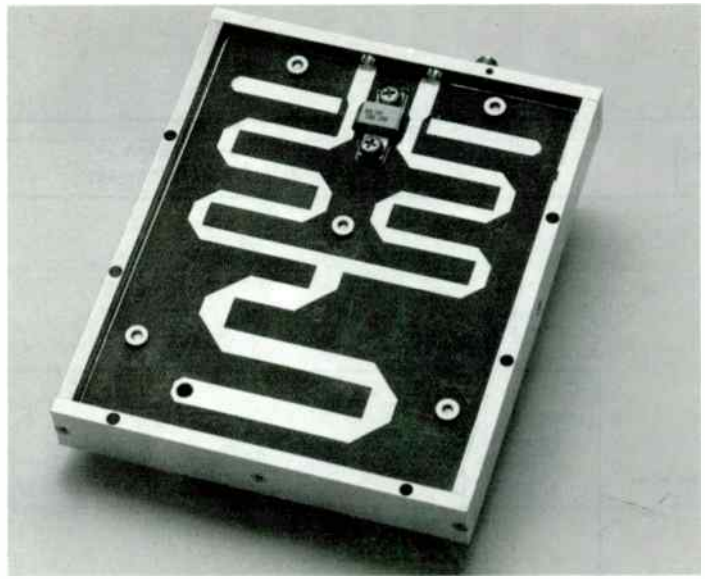


Fig. 7 1:2 Wilkinson
Splitter/Combiner

Fig. 8 200W Circulator

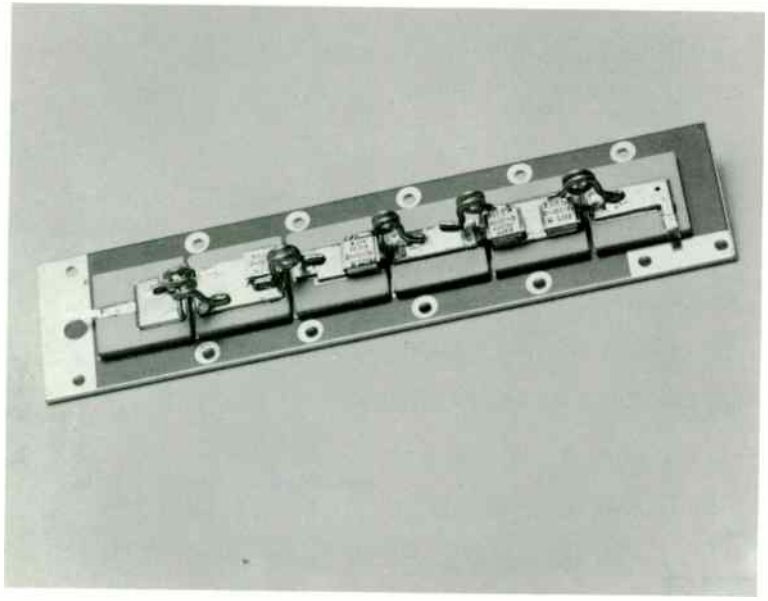


Fig. 9 UHF- Harmonic Filter

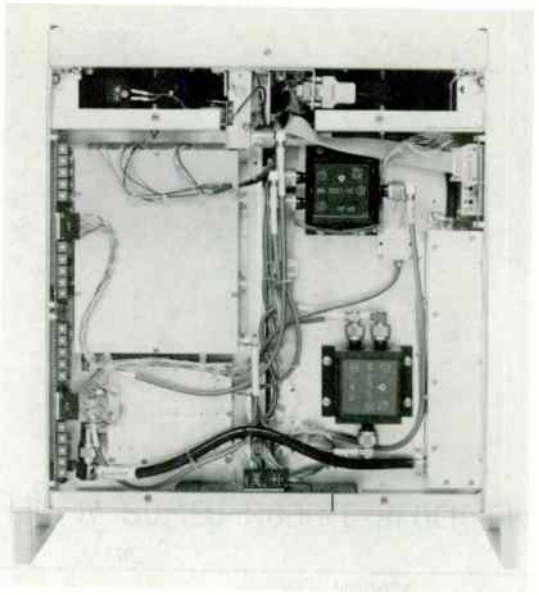


Fig. 11 VD 490 H1, Driver side

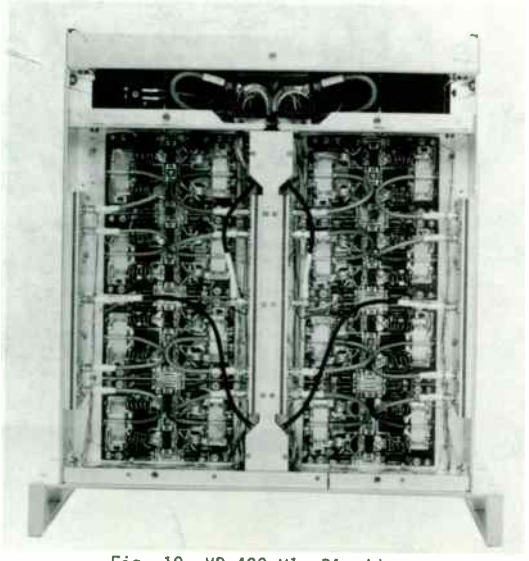


Fig. 10 VD 490 H1, PA-side

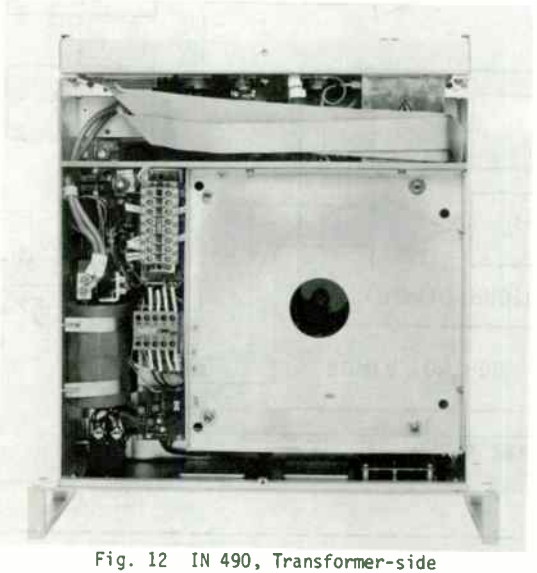


Fig. 12 IN 490, Transformer-side

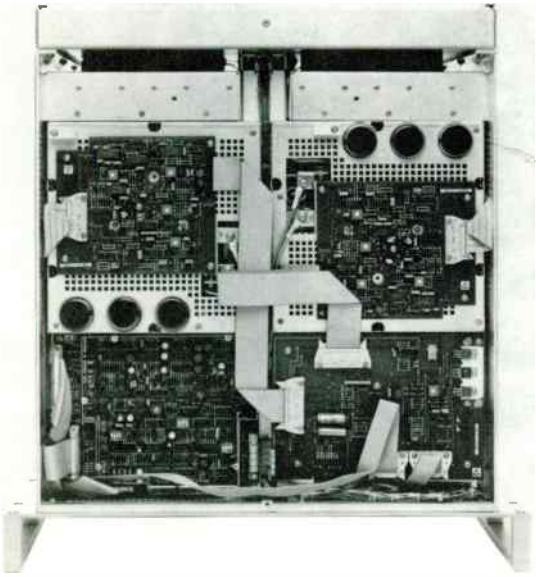


Fig. 13 IN 490 Converter-side

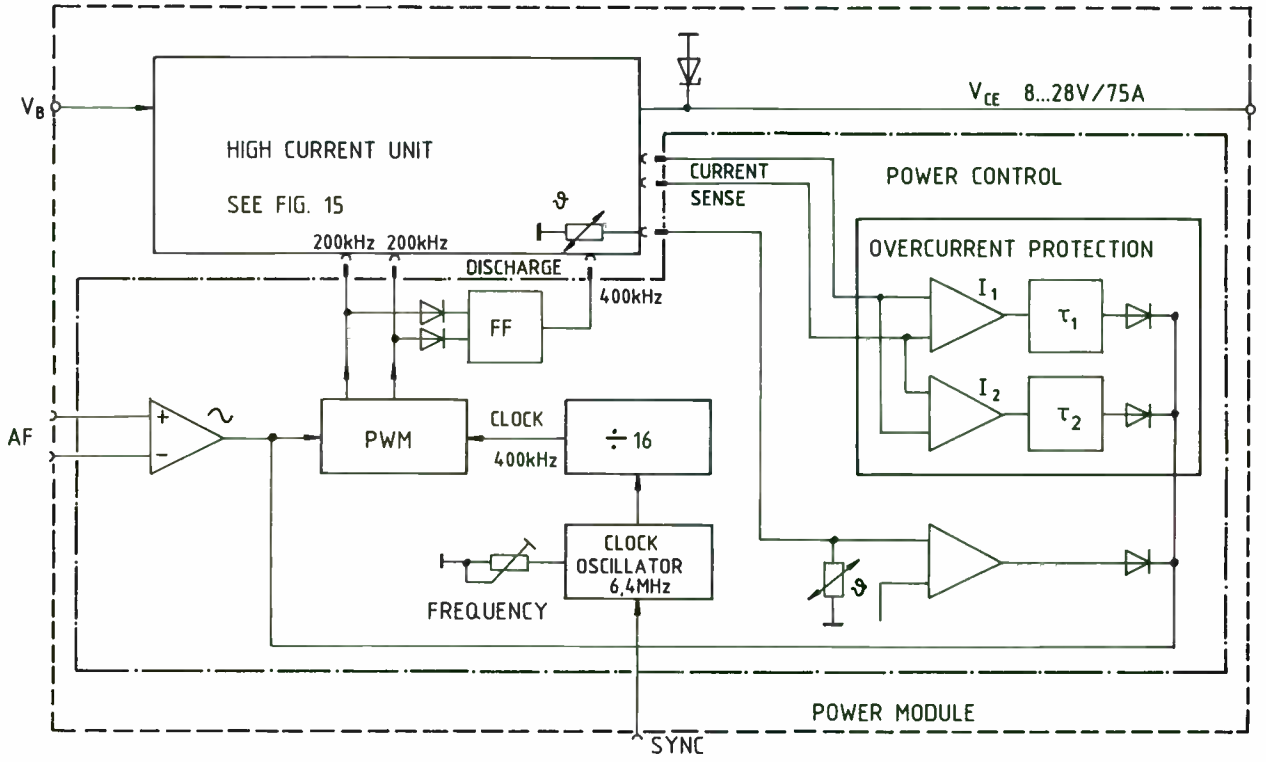
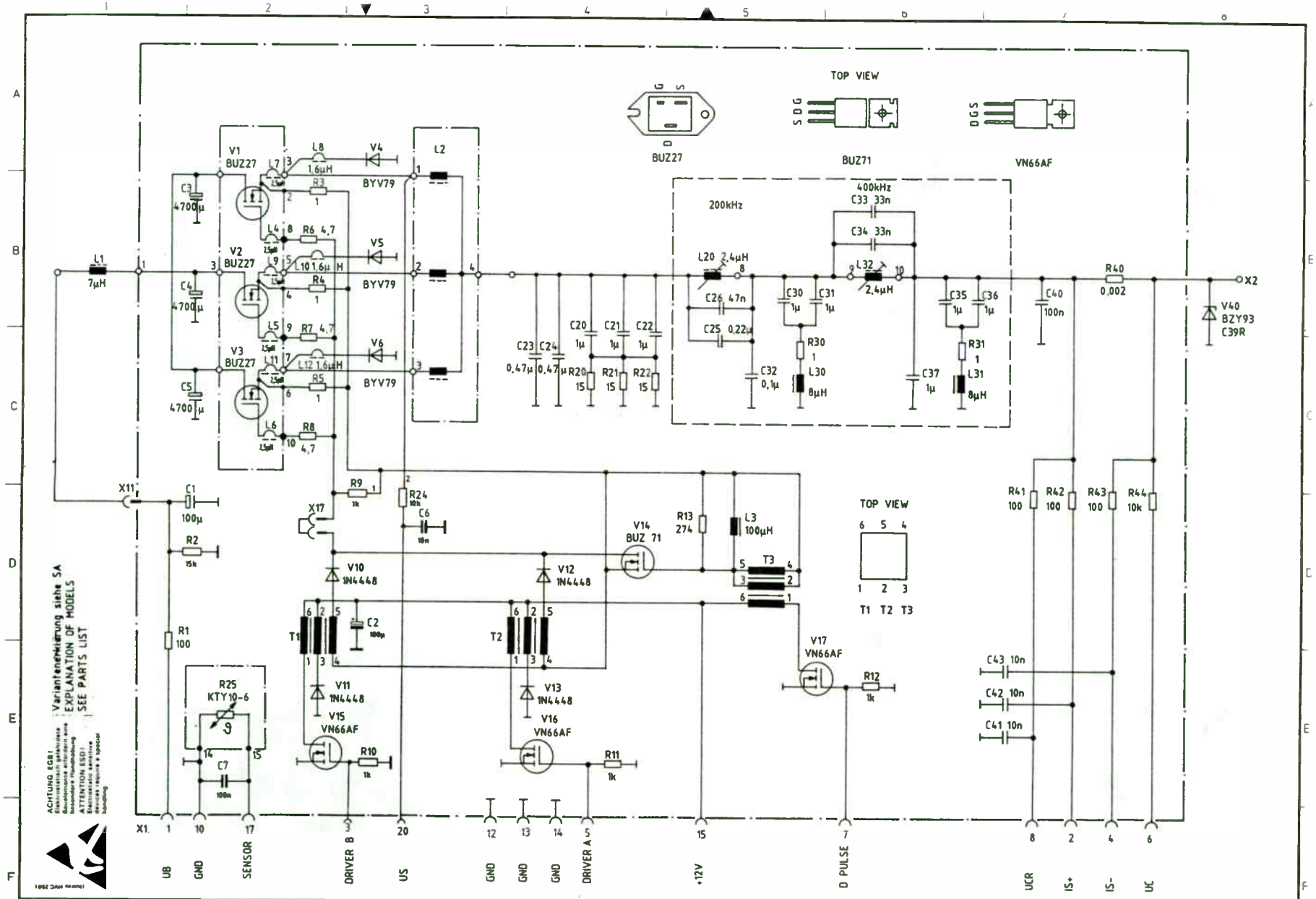


Fig. 14 POWER MODULE BLOCK DIAGRAM



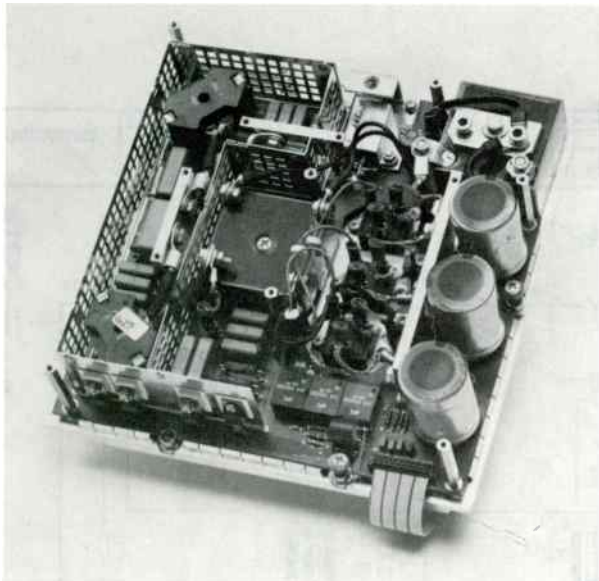


Fig. 16 High Current Unit

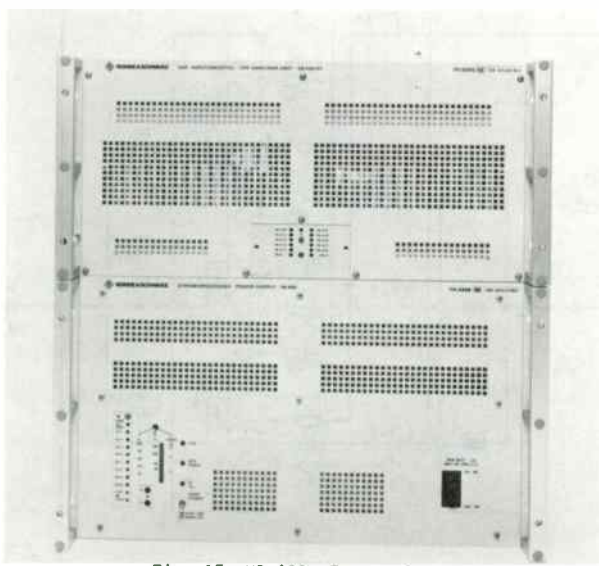
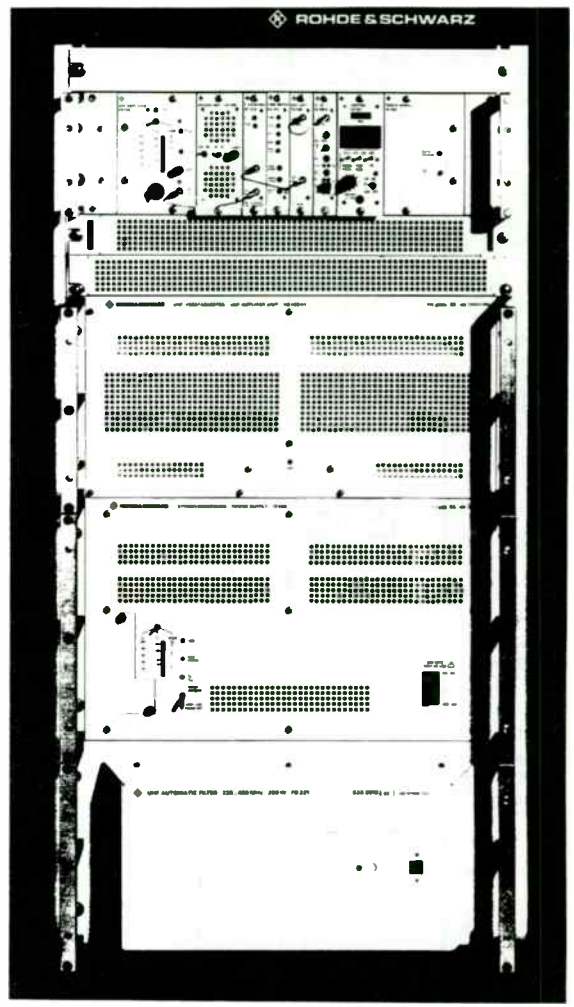


Fig. 17 VD 490, Front view

Figure 18 - 200W UHF T/R System for high collocation demands



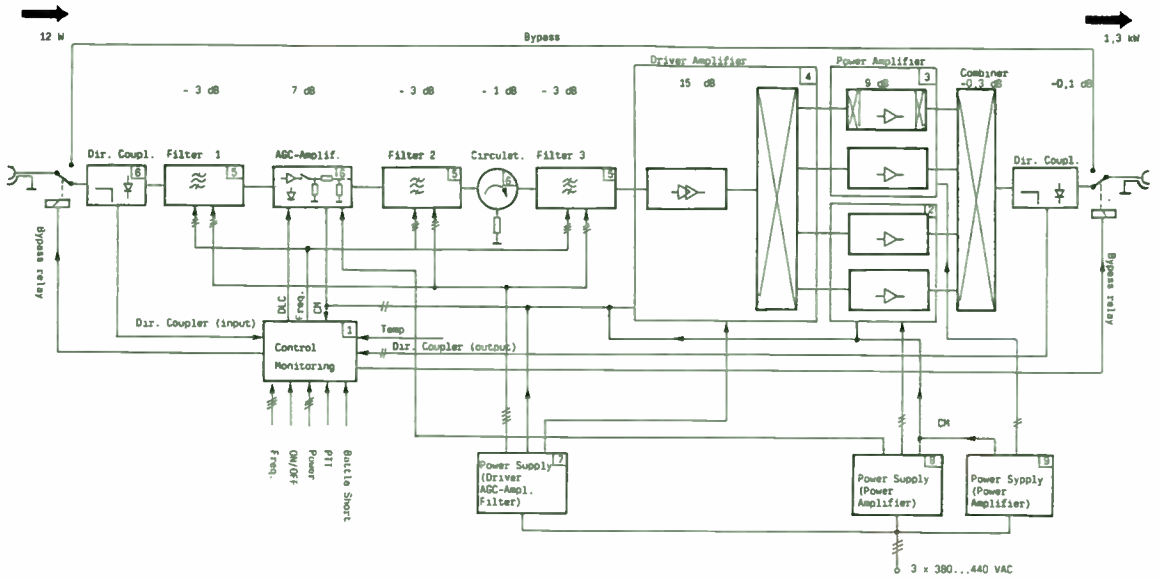


Figure 19 - MPA VD 410, Block Diagram

DESIGNING FREQUENCY SYNTHESIZERS

Combining several frequency synthesis technologies with the well known phase-locked loop (PLL) can result in wide-band, phase-coherent, high-resolution frequency sources. Understanding the possibilities and their technological relationship will produce sound and economical designs.

Prepared for RF TECHNOLOGY EXPO-88
by CORNELL DRENTEA

HONEYWELL INC., HONEYWELL PLAZA
MINNEAPOLIS, MINNESOTA 55408
(612) 870-6658

1.0 INTRODUCTION

It has been noted that most engineers who have not been intensively involved with the design of frequency synthesizers during their careers tend to use simplistic PLL's when faced with frequency synthesis. Although the PLL in its simplest form can be easily understood (it will be assumed that the reader is familiar with its basic concepts), it isn't an exclusive solution to all problems. In addition, complex, high-resolution, fast-switching requirements over wide bands can involve several forms of synthesis which if not properly chosen can minimize results. Oversimplification of the requirements or ill-developed system designs can even result in the wrong technologies being used for the intended purpose. With several forms of PLL's and many other synthesizer forms available, each with their own advantages and disadvantages, the designer is faced with making instant decisions which could prove costly. This writer believes that knowing what technology or combination of technologies need be used in particular applications is as important as circuit design, particularly when low phase-noise, high-resolution/speed-coherent schemes are required.

This paper provides an overview of the available technologies and their relationship, and, more importantly, shows how they can be used together harmoniously. Because of time limitations, the paper will not emphasize circuit details which can be found elsewhere in the literature. Rather, it focuses on the relationship of the different methods used, especially in complex RF signal processing for communications and guidance.

2.0 DEFINING FREQUENCY SYNTHESIS

Frequency-locked and phase-locked loops have been used historically in automatic frequency controls (AFC) and detectors. While, in these applications, they are not commonly referred to as synthesizers (because of the reference frequency being transmitted over the air as a pilot or as modulation) these and other technologies have been used recently in frequency synthesis of local oscillators (LO) in other RF and digital signal processing applications. It is in this context that we refer to frequency synthesis as a means of deriving discrete frequencies from reference oscillators located within, or related to, the user's apparatus for the purpose of providing multiple, stable clocks in digital systems, and also in LO injections to analog RF signal processors involving heterodyning.

Frequency synthesis is defined as the process or processes through which mathematically related frequencies can be created or derived from one or more stable reference frequencies. This definition includes but is not limited to the more simplistic (and inexact) definition of frequency synthesis sometimes found in published material which defines it as "the process providing a finite number of frequencies, all equally spaced".

3.0 SYNTHESIZER FORMS

Although it is impossible to do a total classification of frequency synthesizers, I will discuss proven methods that produce results. The classification will emphasize certain properties of synthesizers which will make us understand when and how to use them collectively.

Frequency synthesizers can be categorized in two major classes, brute-force direct and indirect, and non-brute-force direct and indirect, as shown in Table 1. Further classification is possible as shown. For ease of understanding, the numbers in the table have been keyed with the material in the text.

A brute-force direct synthesizer is usually used in applications where several (mostly up to ten) frequencies are required. It is defined as a process of deriving related frequencies by using consequential frequency generation, such as in the case of inherently-produced nonlinear products in mixers, or by using intentional harmonic generation in frequency doublers, triplers, and the like, and sometimes by simple digital logic division. The brute-force direct synthesizer can be non-coherent, or coherent (fully synthesized) depending on how the various processes use the reference frequencies. The non-coherent synthesizer simply means that the derived frequencies are the result of two or more reference frequencies, each impacting the final stability and resolution independently. A simple example of a non-coherent synthesizer is the mixer type where independent crystal controlled LO's can be combined to produce new frequencies. By contrast, a coherent or fully synthesized process derives all its outputs from a single reference frequency. Thus, the outputs will be coherent with the reference. (Note: A small out of phase relationship between the reference and the output is acceptable within this definition). Full synthesis becomes very important when the designer is trying to achieve total coherence for all the local oscillators in a signal processing system using several kinds of synthesizers. Knowing how

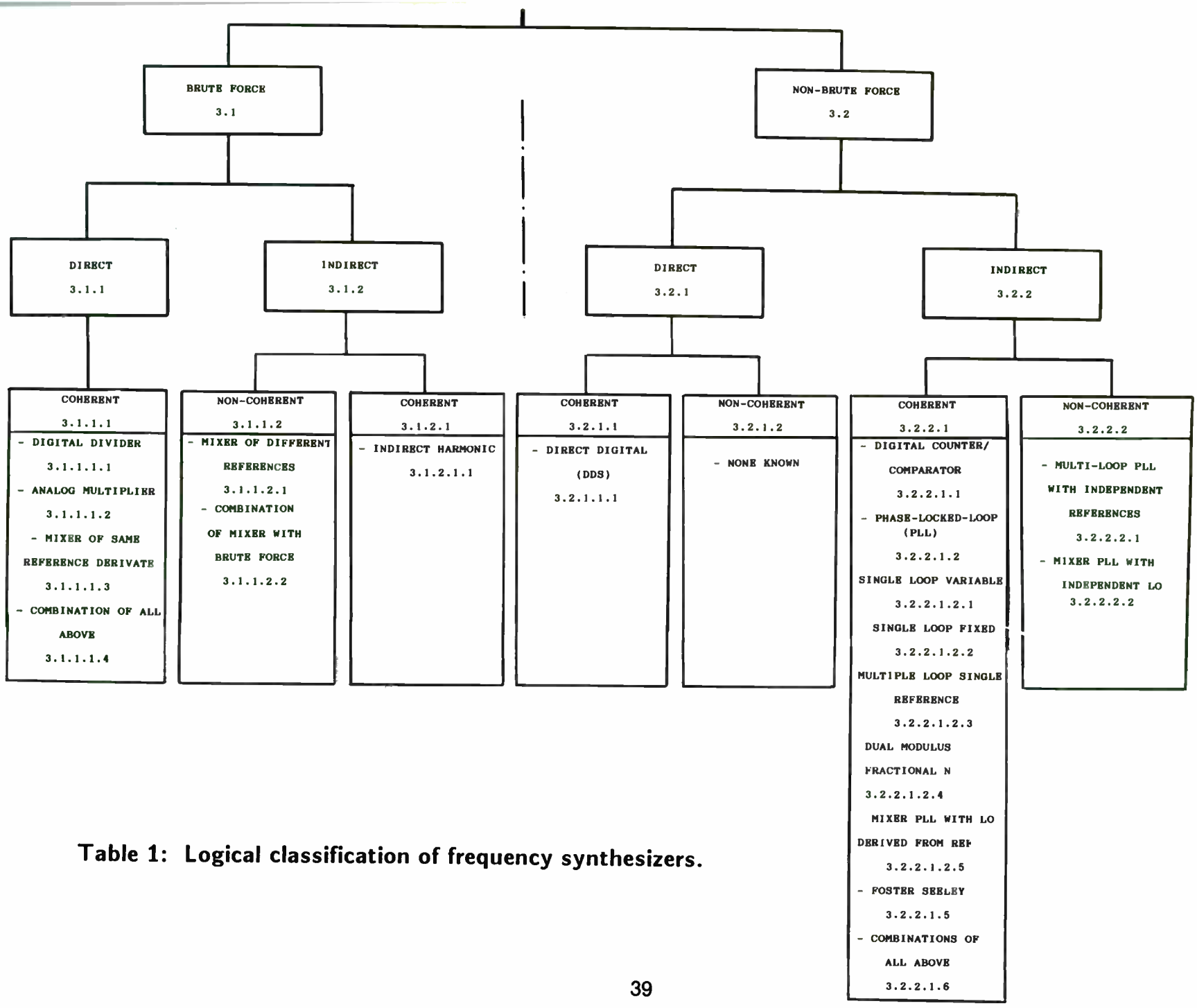


Table 1: Logical classification of frequency synthesizers.

to combine the diverse technologies is of paramount importance in such a complex design case, as will be shown, later in this paper.

As shown in Table 1, most brute-force direct synthesizers with the exception of the mixer type are coherent. The other two general categories of synthesizers from Table 1 are the non-brute-force direct and the non-brute-force indirect. Within the context of the first category, we find the direct digital synthesizer (DDS) as a coherent form (fully synthesized). The Foster Seeley discriminator and the various single PLL's have been identified within the non-brute-force indirect category as coherent types. They are called indirect because they indirectly lock to a reference rather than being used as direct products of a process as is the case with the brute force-direct synthesizers. In addition, various multiple-loop PLL's with independent references, and the mixer type PLL when used with an independent (not reference-derived) LO, have been classified as non-coherent types within the same class. All proven synthesizer technologies have been included in Table 1.

Additional techniques can be added to the table by using the same concept. I will now briefly discuss each type of synthesizer from Table 1.

3.1 BRUTE FORCE

3.1.1 DIRECT

3.1.1.2 NON-COHERENT

3.1.1.2.1 THE MIXER SYNTHESIZER

A good example of a brute-force, non-coherent synthesizer is the mixer type as shown in Figure 1. In this example, two or more crystal oscillators are mixed in a certain mathematical arrangement to produce new frequencies intended as LO injection for channelized receivers, transmitters, or transceivers. This technique is relatively inexpensive and has been used extensively by the CB and maritime transceiver industry. It can also be found in simple broadband local area network (LAN) modems, some microwave communication systems, and radar/navigation equipment. The non-coherent, mixer synthesizer can become coherent when the reference is mixed with its own products and when used with other coherent forms of brute force synthesizers, such as the digital/analog divider type/or the harmonic multiplier. This is shown in Figure 2.

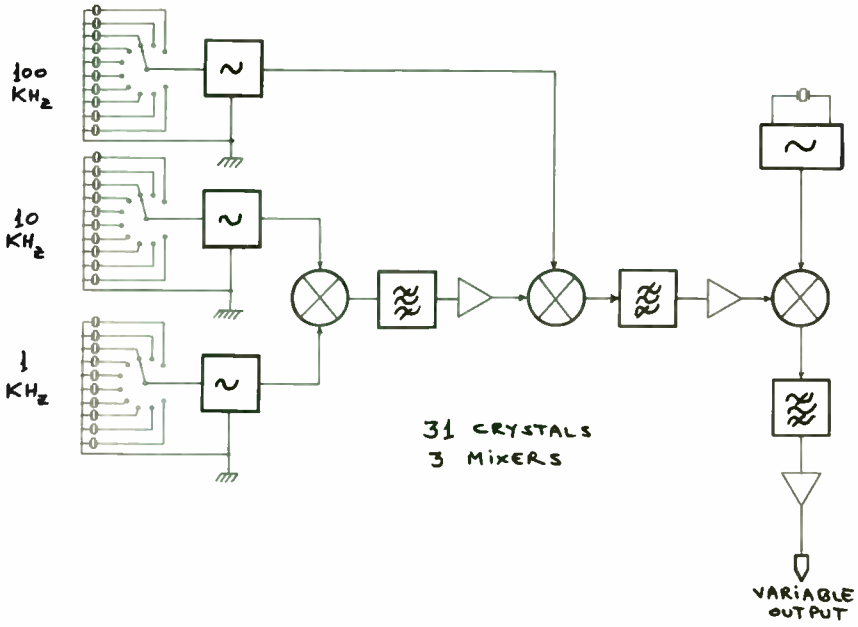


Figure 1: Example of a brute-force direct non-coherent mixer-type synthesizer. Frequency resolution and stability depend on more than one reference.

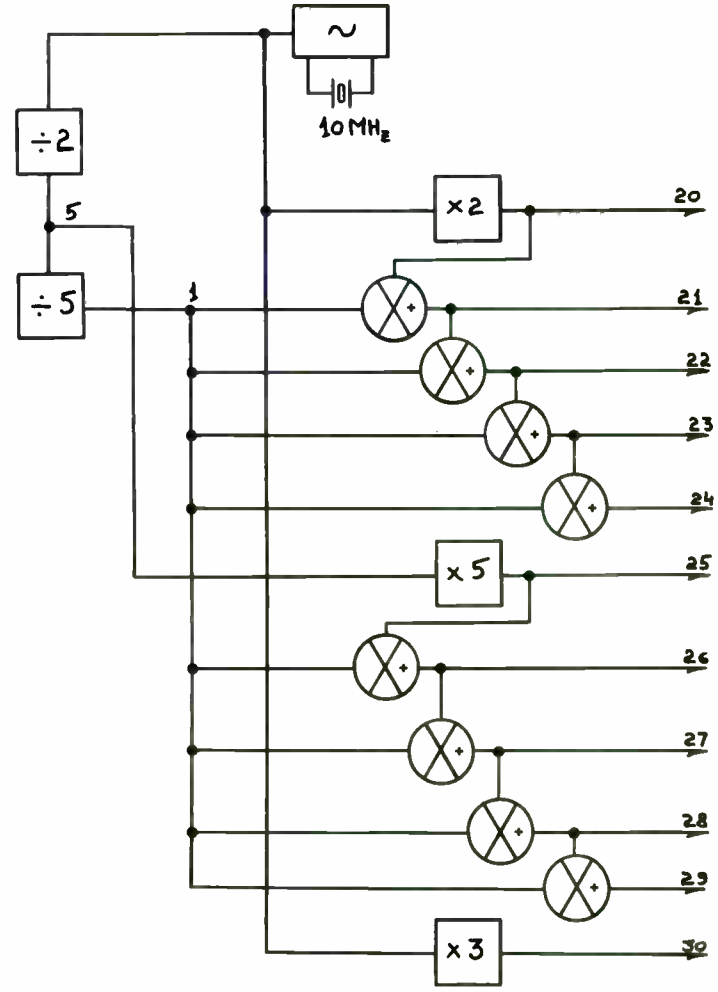


Figure 2: Example of a brute-force direct coherent synthesizer using mixers, dividers, and multipliers. The output resolution and stability are derived from the same reference.

Figures 1 and 2 reveal that although relatively simple at first glance, the mixer synthesizer requires special care in system design. When designing such a synthesizer, it is important to take into consideration the in-band, compounded mixer products (created in the intermediate frequency (IF) chains from combined mixing processes) which carry through the entire system and can be very detrimental to the purity of the final outputs. Design tools for predicting intermodulation distortion in mixers (product charts and computer programs) should be thoroughly used. Only frequency ratios which produce sixth order and higher distortion products should be considered if at all possible, and the process should be carried through the entire mixing chains much like in multi-conversion receiver design. Information on predicting intermodulation distortion in mixers can be found in several of the references at the end of this paper.

Because of the multiple filters required and the shielding involved, the seemingly simple mixer synthesizer can sometimes become a cumbersome piece of equipment. Among its advantages are concurrent parallel outputs (which can translate in fast switching without the lockup characteristic of other forms of synthesis) and straightforward implementation, if the system design is properly developed.

Historically, the mixer synthesizer dates back to the prewar era and is probably the first form of synthesizer ever developed. Its use has diminished with the recent introduction of other forms of synthesizers. However, the mixer synthesizer is finding renewed use when combined with other forms of synthesis in complex hybrid frequency sources.

3.1.1.1 BRUTE FORCE-DIRECT COHERENT

3.1.1.1.1 DIGITAL DIVIDER

Relatively simple to implement, the digital divider can use any of the logic families available to the designer today, depending on the frequencies of interest. However, only the highest frequency logic family required to satisfy the particular application should be used, in order to keep the harmonic-rich square waves from wandering into the sometimes sensitive RF signal processing equipment. A combination of slow rise times, Schmitt triggering, and analog filtering are often used to keep such effects down. One-shots are usually avoided due to their nonsynchronous nature and reliance on temperature dependent RC's. Large-scale integration (LSI) of synchronous counters can be easily implemented, and is highly recommended.

The relative simplicity of digital division makes this method of synthesis very attractive for brute-force synthesizers. On the other hand, it can be noisy and hard to filter and should be carefully implemented.

3.1.1.1.2 ANALOG MULTIPLIER

Straightforward analog multiplication is seldom used alone in frequency synthesis. Exclusive use of multipliers is a sign of poor design and should be avoided unless absolutely required, (exception is the case of multiplied microwave synthesizers). A limited use of multipliers when combined with other forms of synthesis is acceptable

as already shown in Figure 2. The design in this figure shows that when properly combining dividers and multipliers with the mixer synthesizer can result in new, powerful, brute-force coherent synthesizers.

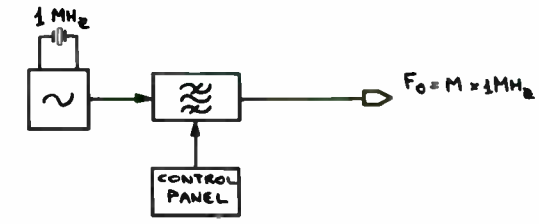
Further combining such synthesizers with direct-digital and/or fixed, phase-locked loops can provide a solid foundation for designing versatile, agile structures as we will discuss later.

3.1.2 BRUTE FORCE-INDIRECT

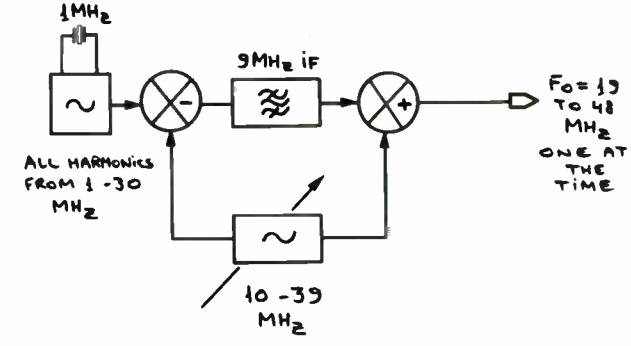
3.1.2.1 COHERENT

3.1.2.1.1 INDIRECT-HARMONIC

A coherent form of brute-force synthesis is the indirect-harmonic. This type of synthesizer locks onto a selected harmonic of a reference frequency, with the help of a variable filter as shown in Figure 3 A, or a double mixing scheme and a fixed IF as shown at B. While many variations of this synthesizer exist, one of its less known applications has been invented by Dr. Barlow Wadley of South Africa.



A)



B)

Figure 3: A) Harmonic synthesizer using a variable, band-pass filter to select multiples of the 1 MHz reference. B) Harmonic synthesizer using a tuned oscillator in a double mixing scheme with a narrow-band IF.

This design, otherwise known as the Barlow Wadley loop, has been commonly used in communications receivers until the recent introduction of the phase-locked loop and can still be found in new applications. The concept is so appealing that engineers have used it repeatedly. The block diagram in Figure 4 shows the implementation of a Barlow Wadley synthesizer in a HF communications receiver. One of the most

interesting features of the Barlow Wadley scheme is the drift cancelling mechanism provided by the double mixing of a free running oscillator (the MHz oscillator) as shown in the figure. A 1 MHz reference oscillator generates LO markers at exactly 1 MHz intervals anywhere between 3 to 32 MHz. The twenty nine harmonics are fed to one side of a loop mixer, as shown. A separate free running variable oscillator named the MHz oscillator generates frequencies between 55.5 to 84.5 MHz, and runs open loop. The output of this oscillator is fed simultaneously to the loop mixer and the first mixer of the receiver where it combines continuously with the incoming RF signal, generating the 1 MHz-wide IF centered at 55 MHz. The signal information which has a bandwidth of 1 MHz (54.5 to 55.5 MHz) is then fed to one side of the second mixer. The other side of this mixer is fed from the loop mixer through a narrow bandpass filter centered at 52.5 MHz. In this approach, the 52.5 MHz signal will only be true at exactly 1 MHz intervals, as a result of the selective mixing process which takes place in the loop mixer. This is shown in Table 2.

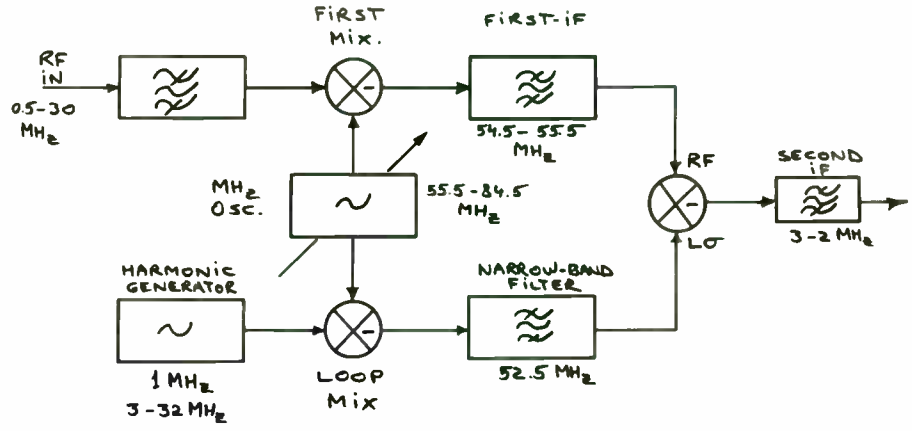


Figure 4: A good example of indirect harmonic synthesis is the Barlow Wadley loop. Stable receiver injection LO is synthesized in 1 MHz steps from a 1 MHz reference. Despite the open-loop nature of the MHz oscillator, any drift is cancelled automatically due to the double mixing process.

The drift of the MHz oscillator is completely eliminated by the subtraction process in the second mixer, providing stable synthesized conversion for the second IF of the receiver.

The advantages of this synthesizer are obvious; simple circuitry can provide many relatively wide and equally spaced frequencies derived from a stable 1 MHz reference. The disadvantages are due to harmonics of the reference oscillator being heard at the beginning and end of each band, a phenomenon which can only be minimized through extensive shielding of the stages and by locating the

harmonic generator as far away from the RF processing circuits as possible.

MHz OSCILLATOR FREQUENCY (MHz)	HARMONIC NUMBER OF FREQUENCY SELECTED (MHz)	LOCAL OSCILLATOR INPUT TO SECOND MIXER (MHz)
55.5	3	52.5
56.5	4	"
57.5	5	"

Table 2: The math behind the indirect harmonic synthesizer process as used in the Barlow Wadley receiver example.

3.2 NON-BRUTE FORCE

3.2.1 DIRECT

3.2.1.1 COHERENT

3.2.1.1.1 DIRECT DIGITAL SYNTHESIZER (DDS)

A relatively new form of direct coherent synthesis, and probably the most promising type of synthesizer implementation today is the DDS (also called the numerically controlled oscillator, NCO). Its concept is simple and effective.

It uses a digital binary adder to increment a preprogrammed constant number representing a phase increment at several arbitrary points in time over a 360 degree cycle. (Note: Some designs sample over several cycles to keep spurious content down at the cost of higher harmonics). Stored digital amplitude values representative of the cycle are sampled every phase increment (the Nyquist criteria calls for a minimum of two samples) from a read-only memory (ROM) and then presented to a digital-to-analog converter which translates them into a sine wave output. Low-pass filtering is used to attenuate harmonics and some of the spurious products caused by the digital switching.

The concept can be viewed as a wheel in a gear train which turns continuously since the cycle repeats every 360 degrees due to digital overflow. When a frequency change is required, a different gear ratio is engaged with the transition keeping the new frequency in phase with the old one. The block diagram of a DDS is shown in Figure 5.

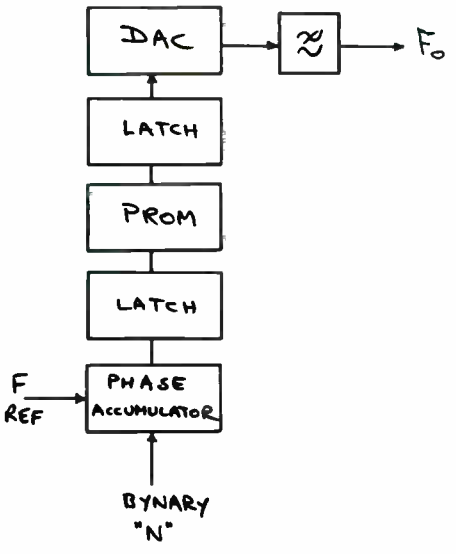


Figure 5: The concept of DDS is based on sampling amplitude values at various points over a full-wave cycle of 0 to 360 degrees (sometimes over several cycles). The number of sampled points can vary but can not be less than the Nyquist rate of two. Once completed, the process rolls over to 0 degrees, and repeats. Implementation is sequential-parallel, digital-to-analog (D/A) conversion.

One of the biggest advantages of this synthesizer type is its ability to achieve high-step resolution because of the sampling technique. In addition, fast, coherent switching of frequencies is possible, which in turn lends to frequency hopping and frequency-shift keying (FSK). Problems usually associated with phase-locked loop settling characteristics and, consequently, phase-noise components are virtually inexistent in this form of synthesizer, making it a design choice expected to replace the PLL in the near future.

Contrary to popular belief, the limiting factor in this type of synthesizer is not the speed of available digital-to-analog products, but the rather poor, spurious performance of this technology when approaching the higher switching speeds which is caused by the so-called "glitch energy". It is not unusual to find DDS implementations which incorporate RC and LC, low-pass filters in the switching lines to reduce the spurious content. Today's state-of-the-art DDS has reached practical frequencies of 130 MHz (300 MHz sampling rate) and higher at resolutions of 1 Hz. The Honeywell HDAC51400 (used by the DDS industry) is a good example of a high performance D/A technology designed to meet the requirements for such high frequency DDS. Recent developments in Gallium Arsenide (GaAs) DAC's are expected to push this technology into the 400 MHz range and higher (1 GHz sampling rate). Honeywell's Solid State Electronics Division (SSED) is looking at new ways of producing higher speed DAC's with reduced switching noise content.

Low-spurious, DDS's can be manufactured to several megahertz. Judicious use of this technology in combination with the phase-locked and mixer approaches can make for cost-effective, multi-octave, high-resolution synthesizers at up to the microwave frequencies. As much as 60 dB of spurious attennation has been realized in such high-resolution hybrid approaches. Figure 6 shows an example of how the DDS can be used effectively to obtain high resolution in a single loop PLL.

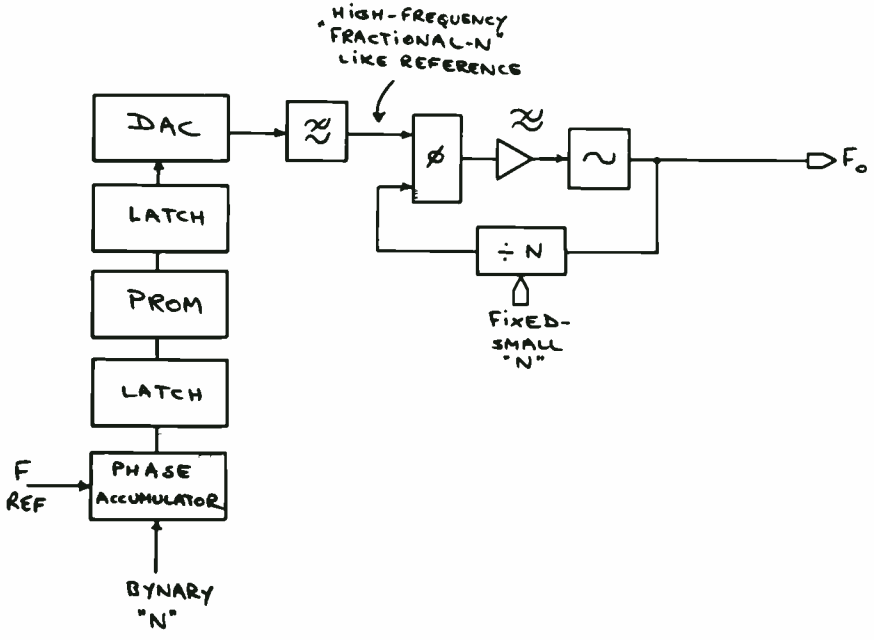


Figure 6: A PLL which uses a DDS in its reference can achieve high resolution and fast switching, with relatively small loop gain. The phase-noise performance of such synthesizer can be superior over other forms.

Today's DDS can provide phase-noise performance which exceeds that of the PLL. A typical example yields -120 dBc at 1 KHz from the carrier. Latch synchronization can be important. As previously stated, the spurious content of a DDS is directly proportional to the DAC's speed, and can be degraded if the system's latches are not synchronized with the reference as noticed in some of the published designs. Prediction of spurious response for a DDS is primarily based on the so-called "image" frequency which is determined by the reference frequency and its harmonics. Due to the Nyquist rate this product is located at least one octave away from the highest frequency output expected and should be greatly attenuated by the synthesizer's low-pass filter. The mixing effects of this product can be attenuated by the half-octave filters usually present at the input or output of the RF signal processing equipment. As much as 80 dB of spurious rejection has been observed in synchronized low frequency DDS's, while as little as 40 dB of spurious rejection has been reported in VHF-DDS systems. Today's DDS uses binary arithmetic inputs which make for fractional reference frequencies which can not be easily derived from even frequency reference oscillators (e.g. 4, 5, 10 MHz). Additional translation which can take the form of a microprocessor is used if binary-coded decimal (BCD) inputs are required. The popularity of DDS synthesizers is expected to increase in the near future.

3.2.2 NON-BRUTE FORCE INDIRECT

3.2.2.1 COHERENT

3.2.2.1.1 DIGITAL COUNTER/COMPARATOR LOOP

One of the less known synthesizer forms, the digital counter/comparator (not to be confused with the Digiphase ® approach) is defined as a coherent type. It is characterized as using the reference of a high resolution digital counter and digital comparators to correct the frequency of a VCO in closed loop. The locking mechanism is achieved by activating an up/down counter connected to a D/A converter through the greater than, less than, or equal outputs of a cascaded comparator chain which constantly compares the VCO's output frequency to a preset number. Consequently, fast-locking transitions

can be achieved with a "sliding effect" caused by the speed of the steering counter. This effect can be useful in certain applications. Because of the relatively large number of cascaded counters required for high resolution this synthesizer type seldom exceeds four decades of resolution. Consequently, the output frequency is open loop between the steps, and simple frequency discriminators can be used in mixing schemes to achieve final stability.

Among its advantages, this type of synthesizer achieves fast switching with relatively low noise content. Its true digital nature makes it a robust candidate for very produceable and reliable signal processors without the problems usually associated with PLL's manufacturing. Large-scale integration (LSI) is feasible and recommended.

3.2.2.1.2 THE PHASE-LOCKED LOOP (PLL)

3.2.2.1.2.1 SINGLE-VARIABLE

The PLL is defined as a coherent type synthesizer under the non-brute-force indirect category. It uses phase/frequency comparison with a reference to correct the frequency of a VCO in closed loop, a simple concept which has many ramifications. The introduction of the PLL has at least temporarily obscured many of the other forms of synthesizers. Because of its apparent simplicity, the PLL has been viewed by many as a solution to all frequency generation problems without a real understanding of the new problems it created and its obvious limitations. Among them, the phase-noise performance of the PLL has been one of major concern. What is phase-noise?

Frequency stability is defined as the degree to which an oscillator produces the same frequency throughout a specified period of time. Every source has a certain amount of frequency instability which can be broken down into two components, long-term and short-term stability. Long-term stability refers to the frequency variations that occur over relatively long periods of time, expressed in parts per million per hour, day, month, or year. By contrast, the short-term stability contains all discrete elements causing frequency changes about the nominal frequency with a duration of less than a few seconds. In a PLL, the long-term stability is assured by the reference through digital loops and phase detectors. In a simple, single-loop variable PLL, digital transients created in the dividing chains combined with the noise sensitive phase comparators, and low Q VCO's, can degrade the phase-noise performance characteristic of such system directly proportional with its loop gain (the wider the loop, for greater frequency coverage, and the smaller the reference, for finer resolution, the greater the problem). The loop filter which is usually intended to keep transients from entering the VCO lengthens the switching time of the synthesizer and a compromise low-pass function results with poor phase-noise performance. With these contradictory requirements, the problem of wide-band, high resolution can only be solved through the application of multi-loop PLL's with improved phase detection, higher references, and switched VCO's for higher Q's. In addition, hybrid designs are often used involving all other forms of synthesizers.

Despite its inherent problems, the single-loop variable-PLL has been considered a solution for many other problems. Some overly enthusiastic articles have publicized unrelated synthesizers using, for instance, frequency-to-voltage detectors in mixing schemes as phase-locked, a mistake which shows the belief that the PLL is viewed

by many as a replacement for all other types of synthesis. Other published material on the subject go so far as to mention the obsolescence of the brute force methods described earlier.

Such is not the case. While the PLL can offer great versatility, it is far from perfect and can not be used by itself except in simplistic applications. On the other hand, unless the DDS is improved from a spurious performance point of view, many forms of synthesizer technologies will still be combined to produce superior results especially when high-resolution, fast-switching parameters are required.

Low-reference, single-variable PLL's are practical in digital clock generation with relatively unpretentious phase-noise performance. An effective implementation of such phase-locked loop is shown in Figure 7.

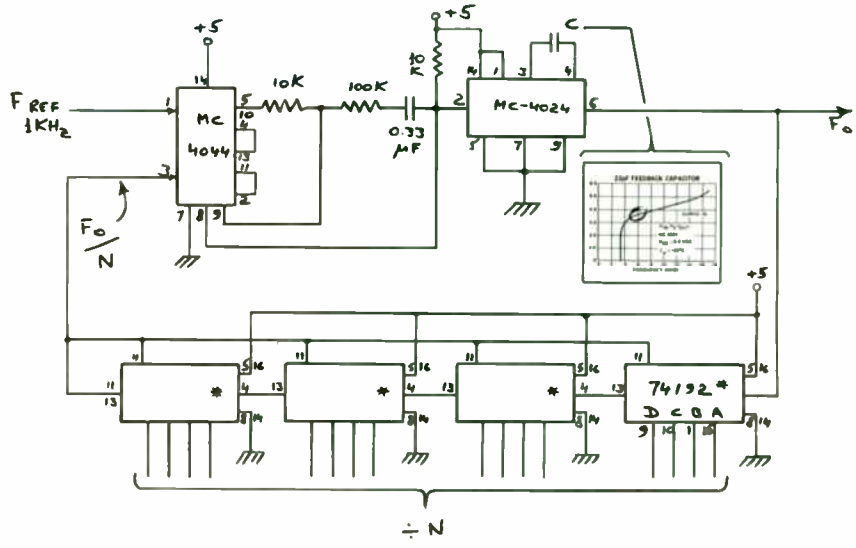


Figure 7: Design of a single loop, digital, variable PLL with a reference/resolution of 1 KHz. Such design can be used in digital systems requiring clock generation.

This simple digital PLL can work in any frequency range within the specifications of the Motorola part MC 4024 (typically up to 24 MHz with the indicated part, or up to 85 MHz with the Texas Instruments 54S124 part). The capacitor C is chosen from the linear region of the Motorola specifications at 77% of the highest required locked frequency (a fact not usually found in the literature). Control over the range is then possible over the remaining range.

In addition to its digital applications, the single-loop variable PLL has analog applications in RF signal processing when the step

resolution is increased to large frequency values, say 10 MHz or 1 MHz. With such high reference frequencies, simple, single-loop variable PLL's have been used extensively in signal processing and communications equipment. The phase-noise performance of these PLL's is outstanding by comparison. However, while the first conversion of such equipment is usually PLL synthesized in 1 MHz steps, the fine tuning and beat-frequency oscillator (BFO) functions usually remain open-loop because of the major design complexity encountered to achieve high resolution-full synthesis.

A typical example of a synthesized/open loop system is shown in Figure 8.

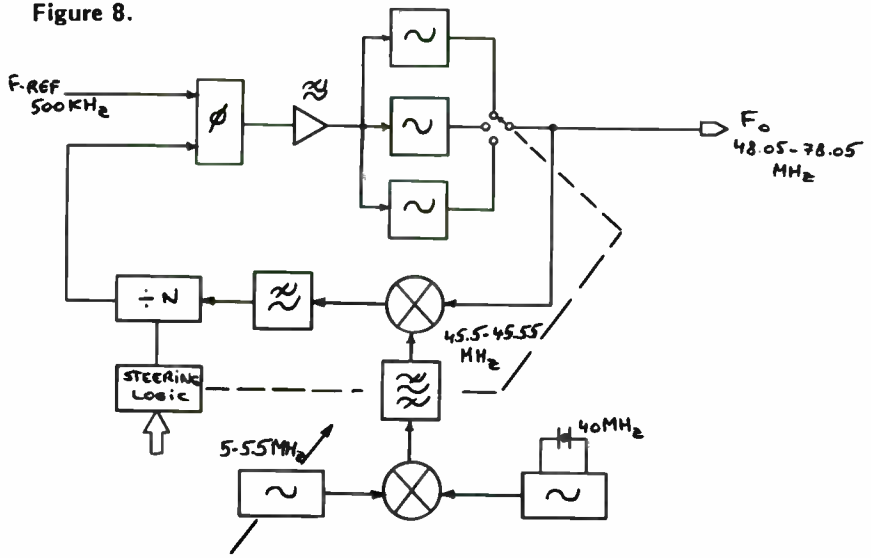


Figure 8: Typical example of not-fully synthesized frequency scheme in a receiver. While a 500 KHz step and high Q VCO's are selected with a PLL, the high resolution is obtained from two open-loop sources through a mixer type synthesizer.

3.2.2.1.2.2 SINGLE-LOOP FIXED

The single-loop fixed PLL can be considered as a variation of the single-loop variable PLL. Its use is somewhat limited to simple digital applications. Of much more importance is the single-loop fixed PLL when used in special applications at VHF frequencies. The example in Figure 9 shows an ideal implementation of such a PLL.

Its output can be used either for frequency multiplication in microwave frequency generation or divided down as a fixed LO injection especially in fully synthesized systems. Our example uses a simple crystal-controlled 88 MHz series resonant Colpitts oscillator to guarantee initial start-up almost on frequency before locking occurs. The output of the oscillator is amplified and converted to an ECL level to be divided, first by eleven and then eight, producing a 1 MHz, fifty-percent duty cycle digital signal which is compared with the 1 MHz reference.

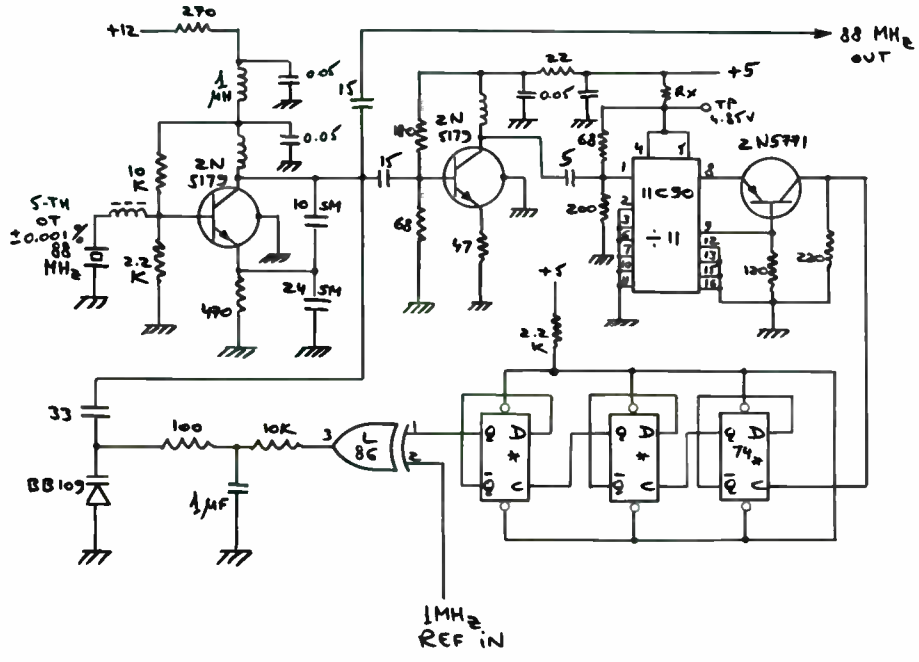


Figure 9: A single-loop fixed PLL at VHF. Such synthesizer is very practical if used in coherent high-resolution synthesizers, multiplier-type synthesizers for microwave frequency generation, divided down for BFO injection, or simply as fixed injection in fully synthesized frequency schemes.

Simple, exclusive-OR phase detectors can be used with this synthesizer as long as the fifty percent duty cycle of the divided digital signal is maintained. Although this paper will not deal with the many types of phase detectors available (full explanations are available in the references), the point to be made is that the designer should use the best phase detector for the particular application. For the situation at hand, it would be a mistake to chose one of the more elaborate phase detectors which would provide locking over several π's. Since the crystal oscillator is guaranteed to start almost on frequency, it will always be within the capture range of the much simpler, exclusive-OR phase detector ($2\overline{\pi}$).

The design shown is typical of fixed phase-locked sources operating at up to 140 MHz. The limitation of such design is the fabrication of overtone crystals at the VHF frequencies. This method of synthesis is particularly useful when fixed loops are required in coherent, multi-loop synthesizers. The phase-noise performance of this synthesizer is excellent and has been measured at better than -70 dBc/Hz at +/- 100 Hz from the carrier. Discrete spurious products are usually 60 dB below the carrier, or better.

3.2.2.1.2.3 MULTIPLE-LOOP, SINGLE REFERENCE

The requirement for fine-resolution PLL's by using high frequency references has been achieved through multiple-loop designs. In addition, deriving references for the various loops from the same crystal oscillator to maintain coherence in systems, has taken advantage of brute-force digital dividers and single-loop fixed PLL's. To obtain the required resolution, other forms of synthesizers, such as the brute-force mixer, and the DDS have been used in conjunction with the PLL. Consequently, complex synthesizers have resulted. A good example of such synthesizers is the famous dual-loop Collins (Rockwell International) approach shown in

Figure 10 for which a patent has been issued. In this approach, high resolution is achieved through the use of two relatively high-frequency references which have been coherently derived from the same TCXO. They have been chosen such that when the two PLL outputs are subtracted from one another in a mixer, their final resolution is determined by the delta frequency between the two references, in our case 100 Hz.

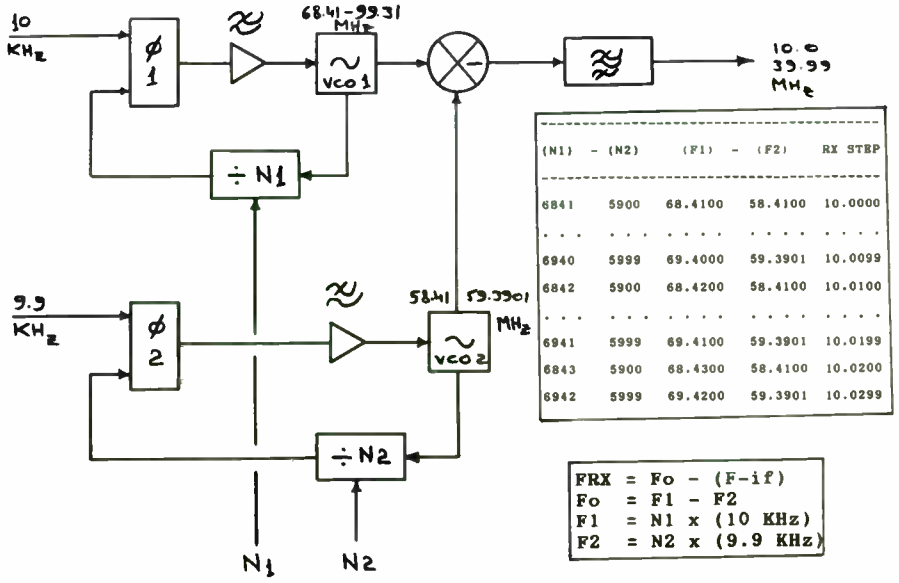


Figure 10: A dual-loop synthesizer provides high resolution and relatively fast switching. The references which are derived from the same crystal oscillator are 10 KHz and 9.9 KHz respectively. The resolution is obtained by subtracting the two VCO outputs and consequently their references in a mixer-type synthesizer.

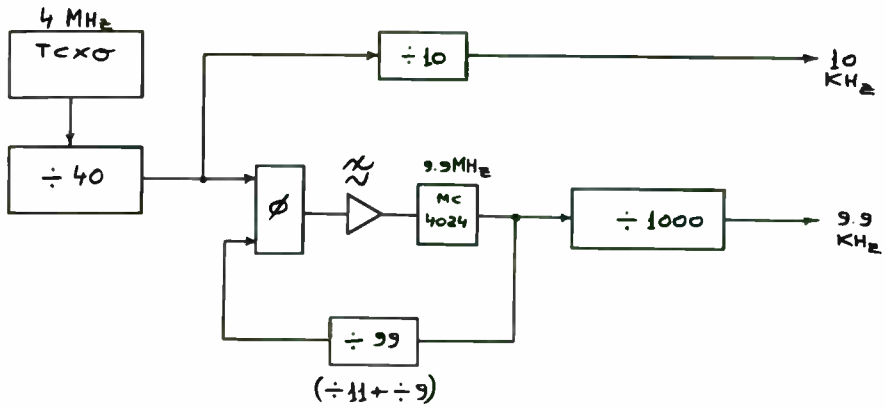


Figure 11: Deriving the two reference frequencies for the dual-loop synthesizer, from a 4 MHz TCXO. A combination of brute-force and fixed PLL forms are used here.

Until the recent introduction of the DDS, the multiple-loop PLL synthesizer has been used extensively in wide-band, high resolution synthesis in receiver and transmitter design. Although some of the loops have been recently replaced with the DDS, the multiple-loop synthesizer remains the choice of designers for high resolution.

3.2.2.1.2.4 DUAL-MODULUS (SWALLOW TECHNIQUE)

When a PLL synthesizer is designed for VHF or UHF frequencies, programmable dividers can not be used directly. Direct programmability over the entire range of a VCO operating at frequencies beyond the capabilities of conventional counters can be achieved through the dual-modulus (P/P+1) prescaling process (also known as the swallow technique). The process is usually implemented in ECL

technology. Dual-modulus counters allow the division ratio to be changed under logic control swallowing of one or two pulses from the total train of pulses provided by the VCO in a loop. To understand this technique, refer to Figure 12 and the following explanation.

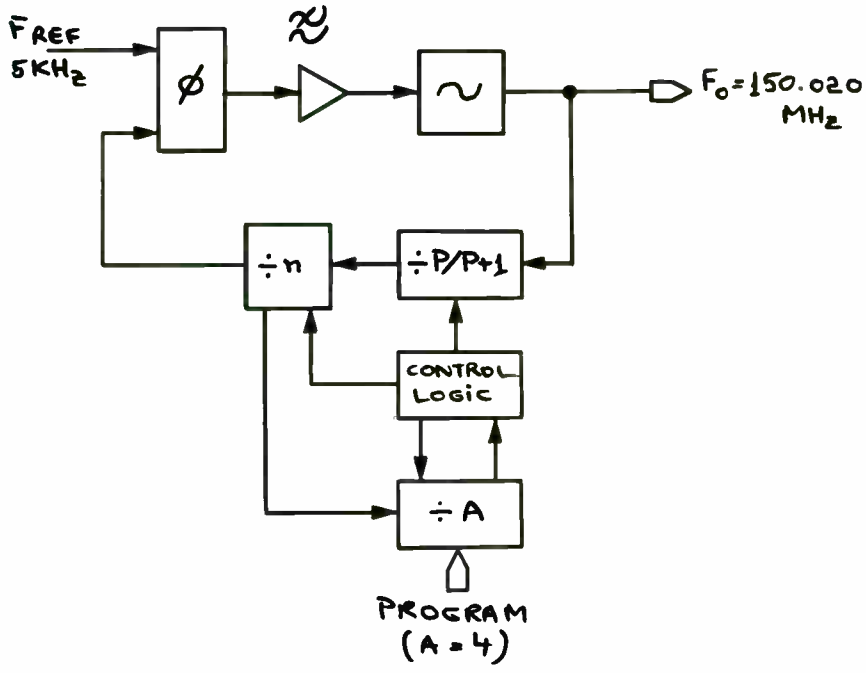


Figure 12: The dual-modulus approach.

Assume that an output frequency (Fo) of 150.020 MHz is expected with a reference of 5 KHz. From the simple phase-locked loop equation Fo = F r x N,

Then:

$$N = Fo/Fr = 150.020 \times 10 \text{ Hz}/5 \times 10 \text{ Hz}$$
$$= 30{,}004$$

Assume the dual-modulus values P and P + 1 are the numbers 20 and 20 + 1 = 21. If N is divided by the P modulus, the resulting quotient is the n program value and the A program value in the remainder.

$$N/P = 30{,}004/20 = 1500 \text{ and remainder of } 4$$

$$n = 1500 \qquad A = 4$$

The key to the operation of this synthesizer is the P + 1 function of the dual-modulus prescaler. If the above value for n and A are programmed into the synthesizer, the loop will perform in the following manner:

The P/P+1 divider is initially set up to divide by 21 and will output a pulse to the divide-by-n counter as well as to divide-by-A counter for every 21 pulses it receives. Since divider A is programmed to divide by 4, this cycle will continue for another three times (each cycle equals 21 VCO pulses at the P/P+1 counter), at which time divider A outputs a command (count of four) to the control logic which instructs divider P/P+1 to change the divisor to P = 20. This change is executed and an inhibit signal is fed back to the divider A, preventing it from any further count until the total cycle is repeated. Meanwhile, divider n which has already counted to four, continues to receive pulses from divider P/P+1 at a ratio of 20:1. After the remaining 1496 pulses have been received (1500 - 4), divider n finally outputs a pulse to the phase comparator as well as to the control logic which resets all counters and the process repeats. It can be seen from this simple explanation how the swallow terminology was born. To check the math involved, use the following equations.

$$N = (P + 1)A + (n - A)P$$

If N = 30,004

Then: 30,004 = (21 x 4) + (20 x 1496)

Various division ratios such as 5/6, 8/9, 10/11, 15/16, 32/33, 40/41, 64/65, 80/81, 100/101 and 128/129 are available in ECL technology operating at frequencies exceeding 1 GHz.

3.2.2.1.2.4 DIGIPHASE® AND FRACTIONAL-N

Alternatives to the multiple-loop synthesizer are the Digiphase® and the Fractional-N approaches. They achieve high resolution with a relatively high-frequency reference. Although this paper will not go into detail about their workings (ample information is available in the references), resolutions of 1 Hz have been achieved with references of 100 KHz by using these techniques. The fractional-N synthesizer is basically similar to the Digiphase® approach, except that it is more like the traditional PLL. The rational behind the fractional-N approach is as follows. If it was possible for N to take on fractional values, the output of a PLL could also be changed in fractional increments. Because fractional division is not possible with standard digital dividers, a different mathematical approach has to be adopted. By dividing the output frequency by N + 1 every M cycle, and dividing by N the rest of the time, an effective division ratio function would become:

$$fo = (N + 1/M)fr$$

This expression shows that fo can vary in fractional increments of the reference frequency by varying M. This is implemented through a fractional phase value which is added to the phase detector's output as shown in Figure 13.

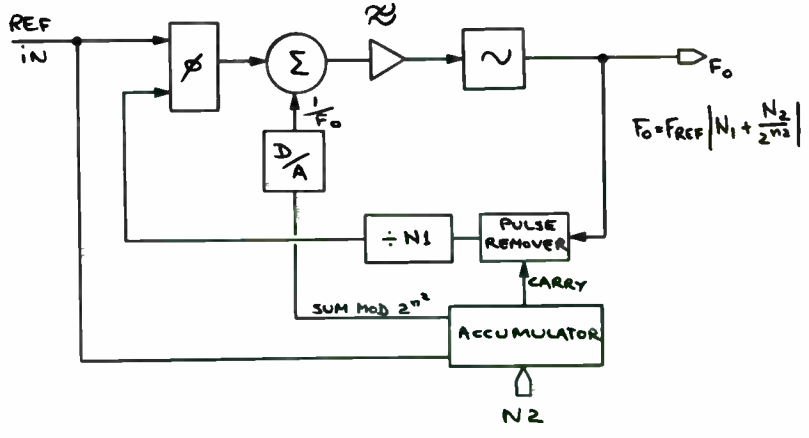


Figure 13: Fractional-N synthesizer approach.

3.2.2.1.2.5 MIXER PHASE-LOCKED LOOP WITH SAME REFERENCE

When the dual-modulus or fractional-N approaches are not economically practical, or the VCO frequency is so high that there is no digital component available to perform the prescaling operation, the mixer-PLL is used. An implementation is shown in Figure 14.

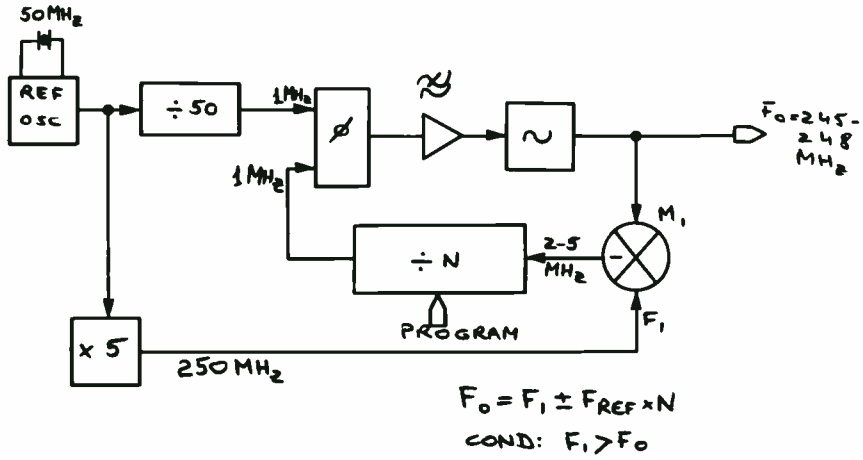


Figure 14: The mixer PLL synthesizer is used to eliminate VHH/UHF prescalers, but limited frequency coverage results.

The mixer PLL approach has the advantage of allowing the use of conventional counters without VHF/UHF prescaling (or other techniques) fact that makes for economical PLL's. An important advantage of this synthesizer is the possible mathematical reversion of "N" due to the mixer subtraction mode which in turn improves the loop dynamics. On the negative side, the second input required by the mixer needs to be derived from the reference frequency, a sometimes difficult task. Although an attractive solution in certain applications, only a limited, loop bandwidth is possible with this approach. The mixer PLL can be used in conjunction with all other forms of synthesizers.

3.2.2.1.5 FOSTER SEELEY DISCRIMINATORS

As previously mentioned, Foster Seeley and ratio detectors have been used in FM demodulators and AFC's. Judicious use of such circuits can help achieve coherent fine tuning in many forms of synthesizers as we have seen in the digital/counter comparator example. Although digital forms of discriminators exist, the operation of analog detectors is usually based on the vector sum of two rectified voltages created across an inductor. One of the voltages is 180 degrees out of phase from the input signal, while the other is only be 90 degrees out of phase. The vector combination of the two, after being filtered, can be used in a feedback circuit of other forms of synthesizers, usually through a mixing process. Quick response, and limited range are the primary characteristics of this form of synthesis.

3.2.2.1.6 COMBINING COHERENT SYNTHESIZER FORMS

Today, practical high-resolution full synthesizers combine all the coherent brute-force and non-brute force types to achieve fast switching and low-phase noise. The experienced designer will mix technologies such as to take advantage of the various properties of the different techniques explained here in order to derive all frequencies from a single reference. Shown in Figure 15A is an example of an earlier approach to a full synthesizer required for a general coverage HF transceiver with a first IF of 75 MHz and a second of 9 MHz. Several of the technologies presented here have been used. For example, the dual-modulus PLL, the mixer PLL, the fixed-loop PLL, and the various brute-force approaches involving the mixer and the digital divider can be recognized in the design.

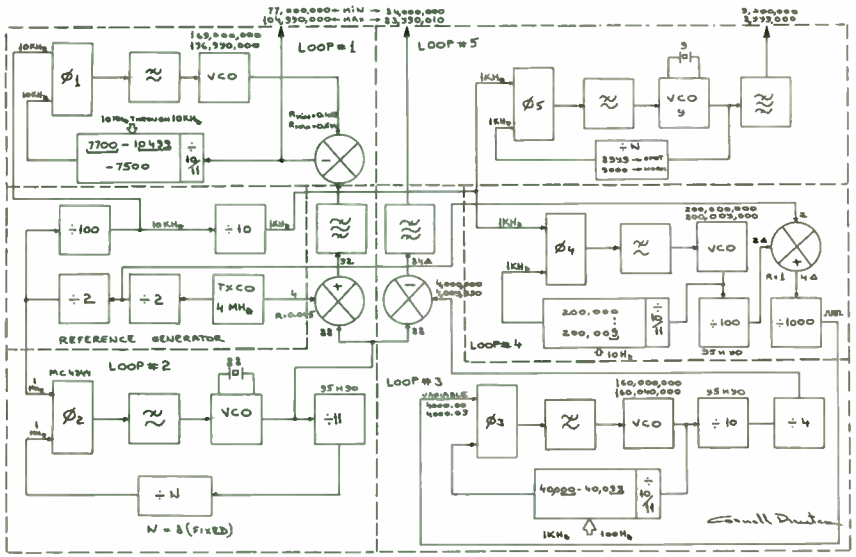


Figure 15A: A coherent high-resolution synthesizer for an HF communications system with first IF at 75 MHz and the second at 9 MHz involves several of the synthesizer forms previously discussed. The large loop gains realized in the PLL's are responsible for relatively high phase-noise content despite the use of VCO presteering.

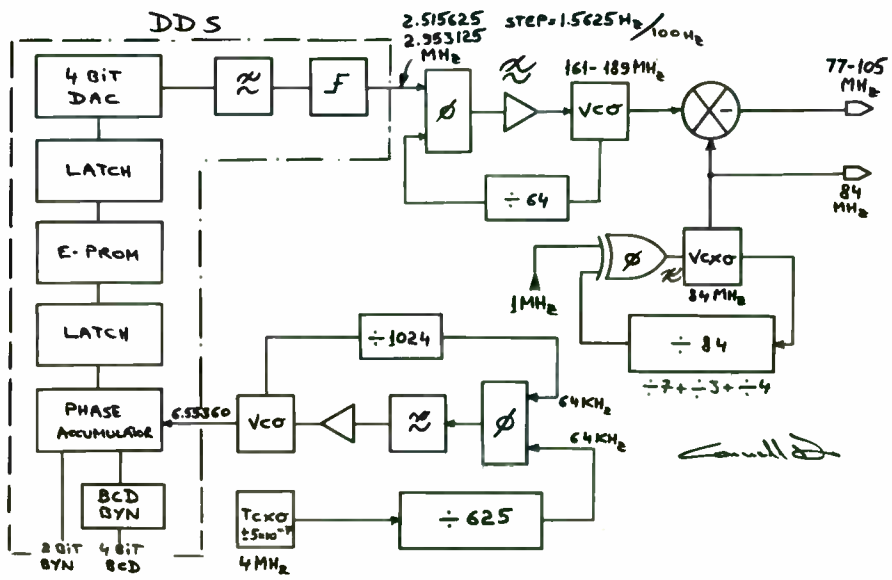


Figure 15B: A better approach to the high resolution synthesizer as developed by the author. Considerably smaller loop gains are realized for improved phase-noise performance with the use of a DDS as a variable reference.

When designing complex synthesizers such as the one shown in Figure 15B, loop gains should be kept small. Although this rule was implemented to some extent in Figure 15A, a better design was realized

by taking advantage of the DDS in the reference frequency of a variable loop PLL as shown at B. In addition, a high reference frequency was used to achieve fractional loop increments which resulted in the final high resolution.

However, a new mathematical problem was introduced into another area of the design, namely the synchronous generation of fractional reference from the 4 MHz standard. This reference was derived through a fixed small-loop PLL and through specially chosen binary divisions. The final design shows clearly how several technologies have been combined.

The above discussion points out that the design of a complex synthesizer is not a simple task when one considers the mathematical and practical impacts of changing various elements in favor of others. Impractical mixing ratios, or uneconomic and sometimes unrealizable prescaling requirements, combined with fractional reference requirements in DDS's, can easily send a good engineer back to the drawing board several times before a good compromise is achieved. While loop behavior can be modeled accurately through computer programs, predicting phase-noise performance through analysis can be inexact and sometimes frustrating.

More often, as in the case of large projects, the designer views the synthesizer as a specialized piece of equipment for which there is no time to design. As various vendors of general purpose synthesizers are approached, a choice is usually made based on knowledge of the requirements at the time. Understanding the technologies presented here can help in choosing the right products and eliminate costly mistakes.

3.2.2.2 INDIRECT-NONCOHERENT

3.2.2.2.1 MULTI-LOOP PLL, INDEPENDENT REFERENCES

3.2.2.2.2 MIXER PLL WITH INDEPENDENT REFERENCE

There are instances when full synthesis in a high-resolution multi-loop synthesizer is not a requirement or can not be achieved. Although the various loops remain coherent in parts of the design, the extra effort required to derive all frequencies from one reference is not viewed as a goal primarily for economic reasons. In such cases, references are generated at will through various oscillators resulting in non-coherent indirect approaches. Such designs show up in various multi-loop and mixer PLL's in which the references or mixer inputs are simple crystal oscillators chosen at random. There is nothing wrong with such approaches as long as the designer follows good RF system design procedures. Special care is needed in the choices of numbers such that frequency sources will not fall in the bandpass of the signal processing scheme. In addition, using multipliers of crystal oscillators to derive fixed injections for mixers should be avoided because of their inherent phase-noise multiplication. Careful choice of technologies will pay off in the long run.

Sometimes it is more economical to design a fully synthesized system than a non-coherent type. Complex analysis of trade-offs is then required since it is much harder to design a fully synthesized frequency scheme than one with random references.

3.3 CONCLUSION

This paper was written for the RF or digital engineer faced with the task of designing or specifying frequency synthesizers. It presents a

new methodology for defining, combining, and implementing various forms of synthesizers. A key part of this methodology is the designer's responsibility of distinguishing between specific coherent and non-coherent applications and applying the technologies accordingly. If properly used, the methods presented here will produce sound results.

REFERENCES

1.0 Drentea Cornell, Radio Communications Receivers, TAB, 1982.

2.0 Manassewitsch Vadim, Frequency Synthesizers Theory and Design, John Wiley & Sons, 1980

3.0 Rohde L. Ulrich, Digital PLL Frequency Synthesizers Theory and Design, Prentice-Hall 1983.

4.0 Egan F. William, Frequency Synthesis by Phase Lock, John Wiley & Sons, 1981.

5.0 Hardy K. James, High Frequency Circuit Design, Reston, 1979.

6.0 Drentea Cornell, High Stability Local Oscillators for Microwave Receivers and other Applications, Ham Radio, November 1985

7.0 Blanchard Alain, Phase-Locked Loops Application to Coherent Receiver Design, John Wiley & Sons, 1976.
Haykin Simon, Communication Systems, John Wiley & Sons, 1978.

FREQUENCY SYNTHESIZER CONTRIBUTIONS TO TRANSMITTER NARROWBAND SPURIOUS EMISSIONS

by
Douglas J. Hughes
Senior Engineer
IIT Research Institute
185 Admiral Cochrane Dr.
Annapolis, MD 21401

ABSTRACT

An investigation was conducted concerning the narrowband spurious emission impact of a small, but representative, number of military communications transmitters and transceivers upon the electromagnetic compatibility (EMC) of colocated receivers. Each transmitter analyzed incorporated an indirect frequency synthesizer using phase locked loop oscillators. The synthesizer designs were detailed, and various sources of spurious emission signals were identified. Comparisons of these designs against applicable measured spurious emission data revealed that six synthesizer subcomponents contributed to spurious emissions: oscillators, mixers, frequency multipliers, frequency dividers/prescalers, intermediate frequency amplifiers, and phase detectors. Synthesizer designs featuring a voltage controlled oscillator (VCO) operating at the transmitter output frequency were found to have less spurious emission impact on colocated receivers, as scored using the IEEE EMC Figure of Merit (FOM), than those featuring heterodyning of the VCO output to the operating frequency.

INTRODUCTION

Experience over the 27-year history of the Department of Defense Electromagnetic Compatibility Analysis Center (DoD ECAC) operations in Annapolis has shown that the task of assigning frequencies for deployments involving many communications-electronics (C-E) equipment must be approached by first considering the colocated equipment. Once the constraints associated with colocated equipments have been identified and appropriate frequencies have been assigned, the assignment process for the remaining equipments is greatly simplified. An automated analysis model (1) has been developed to aid in the assignment process for these colocated C-E systems. It is empirically based and includes assignment constraints based upon the following large-signal interactions: receiver crossmodulation/desensitization, receiver spurious responses, receiver intermodulation, transmitter noise, transmitter intermodulation, and transmitter narrowband spurious emissions. Development of this analysis model has been continuous for the past 20 years.

Efforts to categorize the sources of narrowband spurious emissions from frequency synthesized transmitters and lessons learned from them are the subject of this presentation.

SCOPE

A review of current-generation C-E systems involved with EMC analyses at ECAC revealed that the frequency determining source within all but single-frequency transmitters and receivers is a frequency synthesizer. A few crystal-controlled systems are still being built, but the advent of inexpensive

large-scale-integration (LSI) integrated circuits has made the frequency synthesizer increasingly popular.

The subject of frequency synthesis is broad and beyond the scope of this paper. Excellent references (2-7) are available for the interested researcher. Frequency synthesis may be defined in the most general sense as being a system where circuitry operates on one or a few oscillators to produce many outputs that are different from the original frequencies. While frequency synthesis has many virtues, it can introduce design problems. The greatest of these problems are undesired narrowband outputs and poor noise performance in all but the best designed systems.

FREQUENCY SYNTHESIS (This section is condensed from Reference 4)

Frequency synthesizers are generally divided into two types: direct and indirect. A direct synthesizer is one using frequency mixing and multiplication with appropriate filtering to achieve the desired output. Indirect synthesizers utilize phase-locked loops (PLL) with voltage controlled oscillators. There is overlap between the two types. In fact, most modern C-E systems have synthesizers that incorporate a combination of both designs.

Direct Synthesis

The direct synthesizer generates a number of fixed, stable frequencies by multiplication, division and mixing of a reference signal. Precision reference oscillators are often at a frequency of, or near, 5 MHz because of commercial availability and desirable crystal oscillator stability properties in this frequency range. For example, the reference frequency might be divided to 1 MHz and then multiplied to 3 MHz and 27, ... 36 MHz as required. One of the frequencies from 27 to 36 MHz is selected and mixed with a fixed 3 MHz signal. The resulting sum frequency is divided by 10 and, in the process, the variable digit is shifted from MHz values to 100 kHz values. The output of the divider is approximately 3 MHz, so the scheme can be repeated indefinitely.

Indirect Synthesis

The PLL synthesizer, in effect, multiplies the reference frequency by a variable number. It does so by dividing a voltage-controlled oscillator (VCO) output frequency by that variable number and adjusting the output frequency so that, after division, it is equal to the reference frequency from a stable crystal controlled oscillator. A simple loop is shown in Figure 1a. Under conditions of lock, the two inputs to the phase detector have a constant phase relationship and must, therefore, have the same frequency. The output frequency, F_{out}, must therefore be

$$F_{out} = N \times F_{ref} \qquad (1)$$

If the output frequency increases, the frequency out of the divider ($\div N$) will exceed F_{ref}, the phase difference will increase, and the phase-detector output will decrease. The tuning voltage to the VCO will result in a decrease of the output frequency, thus countering the frequency increase that began the sequence. The loop filter is present to suppress undesired components produced in the phase detector so that they do not cause unacceptable modulation on the VCO. The loop filter has an important effect on

noise suppression, on acquisition of lock, on response speed, and on loop stability.

The mathematical control-system representation of the loop is shown in Figure 1b. The variable at the output is frequency. This is divided by N and subtracted from F_{ref}. The difference frequency is integrated, which is shown in Laplace notation as division by s, to give phase difference. The phase is converted to voltage in the phase detector. K_ϕ expresses the ratio of voltage to phase. The phase-detector output voltage is multiplied by the loop-filter transfer function $G_{LF}(s)$, and the resulting voltage controls the frequency of the VCO. The ratio of frequency to control voltage is given by K_v. There is a one-for-one correspondence between the blocks in Figures 1a and 1b except that the phase detector (PD) in Figure 1a includes the frequency-subtracting function and the two PD blocks shown in Figure 1b. Thus, the voltage from the phase detector is proportional to the phase difference,

$$v_p = K_\phi \, (\phi_{REF} - \phi_s), \tag{2}$$

which is related to the frequency difference:

$$v_p = K_\phi \, (\int F_{ref} \, dt - \int F_s dt \,)$$

$$= K_\phi \int (F_{ref} - F_s) \, dt \tag{3}$$

Figure 1c shows some modifications of the block diagram that are common in practical synthesizers. None of these, in itself, changes the mathematical representation. The same basic loop could operate at a higher output frequency if the output were mixed down to the original frequency, thus leaving the values of N and F_{ref} unchanged. This frequency offset is not shown in the mathematical representation which is concerned only with response to a change from steady state and not with constant offsets which only affect steady state.

Often the reference frequency will be derived from a higher frequency basic reference oscillator by division. This divider (÷M) usually need not be included in the mathematical representation.

A digital-to-analog (D-A) converter may be included, as shown, to coarse tune the VCO to approximately the correct frequency in order to increase speed somewhat or to aid in acquiring lock. Once again, this is an additive signal and may not appear in the mathematical representation. However, if the response to a change in D-A output were being analyzed, this change would be shown as a step in voltage injected after the phase detector.

Synthesizer output spectral purity far from spectral center is basically that of the VCO and, sufficiently close to center, is that of the reference multiplied up to the output frequency. Care must be taken to keep broadband noise in the loop components from corrupting the VCO spectrum. Grounding and shielding are of great importance here, since low frequencies can frequency modulate the VCO if added to the tuning voltage. It is particularly difficult to suppress discrete sidebands separated from the output frequency by the reference frequency. Sidebands due to local low-frequency signals, such as power supply frequencies, are also difficult to suppress. While a few significant sidebands may appear relatively close to spectral center, the output of the PLL synthesizer will be relatively clean far from spectral

center. First, the discrete sidebands are primarily at multiples of the reference frequency and tend to fall off at least 12 dB/octave of separation from the spectral center.[2] Secondly, whereas the direct synthesizer output is basically that of a multiplied crystal oscillator and therefore reaches a low-level plateau of noise at fairly small frequency offsets, the VCO noise of the PLL synthesizer continues to diminish quite far from the spectral center and finally reaches a plateau that is relatively low both because of the higher power in the oscillator and because it is not multiplied. The subject of VCO noise is adequately addressed in the available literature, including references 4 and 8-11.

SPURIOUS EMISSION SOURCES WITHIN FREQUENCY SYNTHESIZED TRANSMITTERS

The designer, or the EMC analyst who is predicting spurious emissions from synthesized transmitters, is advised to consider a multitude of potential sources. Because of the variable frequencies involved in the synthesizer as the output frequency is generated, the resulting spurious emission frequencies will often not experience a one-to-one frequency change relationship to changes in the output frequency. Hypothetical spurious emission sources within the synthesizer are detailed in the subparagraphs to follow. Note also that these same sources can be sources of spurious responses in a synthesized receiver.

Oscillators

The two oscillator types found in frequency synthesizers are the VCO and crystal controlled reference oscillators. The inductor-capacitor (LC) free-running oscillator is almost never found in a synthesizer because of stability problems. A third oscillator type that is sometimes used in microwave and electronic warfare transmitters is the surface-acoustic wave (SAW) oscillator. It may be treated the same as a crystal oscillator for the purpose of spurious emission predictions.

The fixed frequency of a crystal or SAW oscillator can usually be determined by review of a C-E system block diagram. The tuning range of a VCO can likewise be determined from the block diagram. Each oscillator, unless carefully shielded, can potentially couple to transmitter driver and output circuitry to generate spurious emissions.

Oscillator noise may also require consideration. (8-11) The VCO is perhaps the most critical noise source in a PLL synthesizer. If it is spectrally clean, having very low phase noise, it may be controlled with a narrow bandwidth PLL with no compromise in overall noise performance. The major limitation will then be response time. A noisy VCO may be "cleaned up" with a PLL. If controlled by a low division ratio loop, it will be stabilized and the spectral purity of the reference will be transferred directly to the output for spacings within the loop bandwidth.

The best VCO designs are those with the smallest tuning sensitivities. This, of course, leads to a restricted tuning range. The overall range is then expanded by one of a number of methods. One is by switching reactances into the circuit. An additional method to achieve both large tuning range and good oscillator noise performance is to use several switched VCO's to cover the C-E system tuning range.

The designer or EMC analyst is advised to consider VCO tuning range as a function of the range center frequency as a measure of the amount of VCO noise

to be expected in the transmitter output. The general rule is that the smaller the tuning range (higher Q), the lower the oscillator noise level will be.

Mixers

Frequency heterodyning is the primary source of non-harmonic spurious emissions. A mixer is used to combine signals at two different frequencies to arrive at a third frequency that is desired to be at the sum or difference of the two input frequencies. In actuality, the output of a mixer contains all combinations of sums and differences of the two input frequencies and their harmonics due to the nonlinear characteristics of the mixer. Stated mathematically:

$$F_{SE} = p\,F_{LO} + q\,F_{IF} \qquad (4)$$

where

F_{SE} = frequency of the spurious emission output, in MHz

F_{LO} = frequency of one of the mixer inputs, usually a local oscillator (LO), in MHz

F_{IF} = frequency of the other mixer input, often an intermediate frequency (IF), in MHz

p,q = any integer, including zero, denoting the arithmetic sign and harmonic number of the nonlinear mixer response to F_{LO} and F_{IF}, respectively.

The order of the mixer spurious emission is:

$$\text{order} = |p| + |q| \qquad (5)$$

and denotes the coefficient number and exponent value of the power series nonlinear term contributing the applicable mix.

Bandpass filters are used following mixers to pass only the desired mix from Equation (4). The effectiveness of the filter determines which undesired products result. The EMC analyst or designer should assume that any of the Equation (4) mixes could potentially be present at a transmitter output unless measured data are available to prove adequate attenuation due to bandpass filters.

Frequency Multipliers

Frequency multipliers are used in transmitters to generate a signal at an integer multiple of an oscillator or intermediate frequency. They are circuits driven or biased so that significant harmonic distortion occurs. A filter at the output is used to select the desired harmonic output while suppressing the others as well as the driving input signal.

The filtering requirements at the output of a multiplier may be reduced considerably through the use of balance in the circuit. A balanced frequency doubler is essentially a traditional full-wave rectifier circuit with an RF choke at the output to short the DC output component to ground. The suppression of the fundamental driving input is limited by the diode matching and transformer symmetry, with values of 40 dB being realized with careful construction. (4) The usual circuits employ hot-carrier diodes. A virtue of this circuit is the broadband response. It is useful in some applications with

no output filtering if the presence of high-order even harmonics is not objectionable. However, some filtering is often useful. Even a single resonator will produce a surprisingly clean output. The circuit has a typical loss of 7 to 9 dB and requires an input drive power sufficient to overcome the diode turn-on voltages.

Frequency triplers or even quadruplers may be built with transistors. Single devices with suitable output filters may be used. Alternatively, balance may be employed to suppress some of the output components. Multiplication beyond a factor of 4 or 5 per stage is generally not recommended with transistors.

High-order frequency multiplication is most often performed with varactor diodes, or step-recovery diodes. The step-recovery diode is a junction device with large amounts of charge storage. The fundamental driving signal will forward bias the diode over part of the operating cycle. As the input polarity reverses, the charge stored in the diode does not disappear immediately. When the driving polarity reverses, current continues to flow just as if the device were a resistor. However, when the stored charge is finally "swept out" of the junction, current ceases. The sudden decrease leads to a large output voltage of short duration. The pulse is very rich in harmonics. The output voltage pulse results from driving the diode with an inductive source.

The output of a step-recovery diode multiplier may be filtered to achieve a desired frequency. The filter must generally be of relatively narrow bandwidth owing to the high order of multiplication employed. Step-recovery-type multipliers are well covered in Reference 6.

Numerous other methods are available for frequency multiplication. All of them will degrade the phase noise characteristics of the source. The designer and EMC analyst should assume that any or all of the undesired frequency multiplier harmonics could potentially generate spurious emissions.

Frequency Dividers

The advent of inexpensive LSI digital dividers has greatly affected synthesizer designs. It has made frequency division feasible. Programmable dividers have been produced commercially that perform well at input frequencies in the hundreds of MHz. The highest-frequency digital dividers operate at input frequencies in excess of 1,000 MHz but function only as fixed dividers, or prescalers, with no logic inputs. An arbitrary number of dividers may be cascaded to produce very low frequency outputs. They may also be programmed to divide by any integer number over quite a large range.

Fixed dividers, or prescalers, that are used ahead of programmable dividers have been found to be sources of spurious emissions. These dividers usually operate at frequencies in the hundreds of MHz and have outputs in the tens of MHz. Unless carefully shielded, the prescaler outputs can couple to other transmitter circuitry and be transmitted directly or else frequency modulate the VCO. The prescaler output is rich in harmonics; therefore, all harmonics should be considered by the designer or EMC analyst.

Intermediate Frequency (IF) Amplifiers

Intermediate frequency amplifiers find application in frequency

synthesizers whenever the VCO output is heterodyned to a different frequency before being divided and compared with the reference. Figure 1c illustrates such a case. This IF is an additional frequency that is present within the synthesizer and that must be considered as a spurious emission source. It will change frequency with a change in the operational frequency. Review of the system block diagram will allow determination of the IF as a function of transmitter output frequency. An IF amplifier usually follows a mixer; therefore, undesired mixer products will also be present and amplified unless the IF bandpass filtering is adequate. Equation 4 may be used to determine if unwanted mixer products may fall within the IF amplifier passband. If so, these products, in addition to the IF, must be considered as potential sources of spurious emissions.

Phase Detectors

There are numerous circuit types that find application as phase detectors in frequency synthesizers. They include: balanced modulators, balanced mixers, high-speed samplers, exclusive ORs, flip-flops, sample-and-hold devices, and phase-frequency detectors. The most popular phase comparing integrated circuits are of the phase-frequency type. They use two edge-triggered flip-flops and a gate to generate a digital signal that contains a number of pulses proportional to the phase difference between the two input signals. The detector output filter removes the pulsing waveform while retaining the average direct current value. This DC signal is the error voltage used to fine tune the VCO.

All the phase detector types contain outputs that are rich with harmonics. A lowpass filter is required to attenuate these harmonics to prevent them from frequency modulating the VCO. None of the filters are perfect devices and harmonics of the reference frequency will be present on the VCO output at reduced levels. These carrier sidebands must be taken into consideration when doing an EMC analysis unless the reference frequency is low enough to restrict the sidebands to within a few tenths of a percent of the carrier frequency. Receivers are rarely assigned to operating frequencies this close to the operating frequency of a colocated transmitter.

Harmonics

Spurious emissions at frequencies that are an integer multiple of the transmitter output frequency (harmonics) are always present at various levels dependent upon the effectiveness of the transmitter output filtering.

DATA AVAILABILITY

Measured spurious emissions data were obtained for seven transmitter nomenclatures that each feature a PLL synthesizer as part of the design. The data recorded from the 50-ohm, closed-system tests were the frequency of each emission and a corresponding power level, in units of dBm. The selected samples are used to show the significant sources of spurious emissions. The seven transmitters were the: AN/ARC-159, AN/ARC-164, AN/ARC-182, AN/PRC-124, AN/WSC-3, AN/WRT-2, and AN/GRC-144. TABLE 1 lists the manufacturer, frequency coverage, output power capabilities, nomenclature date, quantities of test samples, number of tuned frequencies tested and synthesizer figure number for these seven communication systems.

TABLE 1

COMMUNICATIONS TRANSMITTERS/TRANSCEIVERS EVALUATED

Nomenclature	Frequency Coverage (MHz)	Manufacturer	Nomenclature Date	Output Power (Watts)	Number of Test Samples	Number of Tuned Frequencies	Synthesizer Figure Number
AN/WRT-2	2-32	Westinghouse	1959	1000	1	3	4
AN/ARC-159	225-400	Rockwell/Collins	1972	10	2	3	5
AN/GRC-144	4400-5000	ITT	1973	0.25	2	4	6
AN/WSC-3	225-400	E-Systems	1974	100	1	1	7
AN/ARC-164	225-400	Magnavox	1978	10	3	7	8
AN/ARC-182	30-88 108-174 225-400	Rockwell/Collins	1979	10 10 10	3 3 3	3 6 5	9
AN/PRC-124	30-88	Rockwell/Collins	1984	5	1	2	10

The data were obtained in order to evaluate the relative significance of the seven hypothetical spurious emission sources.

DATA REDUCTION TO CATEGORIZE SOURCES

The measured spurious emissions data were evaluated to determine sources within each transmitter that generate significant spurious emissions. The measured power levels were also evaluated in order to determine if an analytic model could be developed that would be superior to the log-normally distributed statistical power level model presented in Reference 1. Several steps were involved in this categorization process.

Transmitter Block Diagram

Block diagrams were obtained from technical manuals or from ECAC project personnel for the seven transmitters studied. Each oscillator, mixer, frequency multiplier, frequency divider, IF amplifier and phase detector that could potentially generate spurious emissions was identified. Operational frequencies of each potential source were identified and compared to the measured spurious emission frequencies. The measured spurious frequencies were readily correlatable with the identified sources. TABLE 2 summarizes which sources were present within each of the seven transmitter types. Each identified source was found in at least three of the seven.

Multiple Test Samples

As indicated in TABLE 1, data representing more than one serial number were found for four of the seven analyzed transmitter types. The spurious emission sources and relative levels were confirmed by evaluation of data for the multiple test samples when available. The spurious emission sources were generally found to be independent of serial number. Some variations in levels, measured for the same spurious emission, only from a different transmitter sample (same nomenclature), were found.

Multiple Test Frequencies

Six of the seven transmitter types analyzed were tested at multiple operating frequencies. This allowed the data to be examined for spurious emission source level variations with tuned frequency. A particular source was found to have a fairly constant output level in the transmitters containing broadbanded transmitter driver and amplifier filtering. These transmitters were the AN/ARC-164 and AN/ARC-182. The effects of filtering in the remaining transmitters may be examined, with the exception of the AN/WSC-3. Only one AN/WSC-3 operating frequency (272.6 MHz) was tested; therefore, the effectiveness of the AN/WSC-3 voltage-tuned bandpass filters cannot be determined without additional data taken for other tuned frequencies. However, most of the measured spurious emissions for the AN/WSC-3 were found within 10 MHz of the operating frequency. This implies some filter effectiveness.

Large standard deviations (> 8 dB) from mean values (in dBm) were found for most measured harmonic spurious emission rejection levels. This suggests that a log-normal distribution statistical model is inadequate for harmonic level predictions. Additional analysis of harmonic attenuation features of transmitters is planned in order to predict harmonic EMI power levels more accurately.

TABLE 2

SPURIOUS EMISSION SOURCES PRESENT IN THE SEVEN TRANSMITTERS ANALYZED
(X Indicates Presence of the Source)

Spurious Emission Sources	Transmitter Nomenclature						
	AN/ARC-159	AN/ARC-164	AN/ARC-182	AN/WSC-3	AN/WRT-2	AN/GRC-144	AN/PRC-124
Oscillators							
Crystal	X		X		X	X	X
VCO	X		X	X	X	X	X
Mixers	X		X	X	X	X	X
Frequency Multipliers	X				X	X	
Frequency Dividers/ Prescalers	X	X	X			X	X
IF Amplifiers	X		X			X	
Phase Detector Reference Frequency and its Harmonics		X	X	X			X
Harmonics	X	X	X	X	X	X	X

Large standard deviation values were also found for the AN/ARC-159, AN/WRT-2 and AN/GRC-144 transmitter nonharmonic spurious emission levels as a result of multiple tuned frequencies. This suggests a need to add transmitter tuned filter effects to allow prediction accuracies better than the Reference 1 log-normal spurious emission model.

Synthesizer/Driver Design Categorization

The transmitter block diagrams were used to categorize the synthesizer/driver designs. It was discovered that each design could be assigned to one of three broad categories as a function of the PLL synthesizer voltage controlled oscillator (VCO) frequency and its relationship to the transmitter output frequency. These categories were as follows:

Synthesizer Category	Feature
A	The VCO output is at the transmitter output frequency.
B	The VCO output is mixed with a second oscillator to arrive at the transmitter output frequency.
C	The VCO output is multiplied and then mixed with a second oscillator to arrive at the transmitter output frequency.

A fourth category is obvious for future study. It is the case where the VCO output is multiplied to arrive at the transmitter output frequency. No measured data was found for transmitters representative of this category. It is known that the synthesized versions of the AN/GRT-21 and AN/GRT-22 RIVIT SWITCH transmitters fall into this category "D"; however, only crystal controlled versions of these transmitters have been measured to date.

Spurious Emission Figure of Merit (FOM)

The measured data were used to calculate an in-band, an out-of-band, and a composite spurious emission FOM score for each of the seven transmitters. The methods of Reference 12 were used. This method is used to score the EMC FOM of transmitter spurious emissions with scores that range from below zero to above 100. A volunteer panel of EMC engineers derived the scoring method so that the worst transmitter they reviewed received a zero or lower score and the best received a score of 100 or greater. The method was derived so that a score of 50 would represent an average transmitter. The lower the FOM the more a colocated receiver would be impacted by spurious emission EMI and the more assignment channels would be denied. Reference 12 also presents FOM scoring techniques for other receiver/transmitter EMI characteristics, such as: intermodulation, adjacent signal, spurious response, and transmitter noise. Only the spurious emission FOM was applicable to this study.

In order to determine the FOM, all closed-system measured spurious emissions in given frequency ranges that exceed a -70 dBm level are counted and their measured levels, in dBm, are averaged. The out-of-band and in-band frequency ranges are shown in Figure 2. A small excluded band is allowed around the tuned frequency where measured spurious emissions are not counted. The reasoning is that receiver adjacent signal EMI interactions would dominate spurious emission interactions at these small frequency separations.

The FOM calculation is stated mathematically as:

$$S_{SEI} = \frac{170}{(1.08)^{N_{SEI}}} - \bar{P}_{SEI} - 80 \qquad (6)$$

$$S_{SEO} = \frac{170}{(1.08)^{N_{SEO}}} - \bar{P}_{SEO} - 80 \qquad (7)$$

and

$$S_{SE} = (S_{SEI} + S_{SEO})/2 \qquad (8)$$

where

S_{SEI} = in-band FOM score

N_{SEI} = number of in-band measured spurious emissions found to exceed -70 dBm

P_{SEI} = the average spurious emission level (in dBm) for all in-band sources found to exceed -70 dBm

S_{SEO} = out-of-band FOM score

N_{SEO} = number of out-of-band measured spurious emissions found to exceed -70 dBm

P_{SEO} = the average spurious emission level (in dBm) for all out-of-band sources found to exceed -70 dBm

S_{SE} = average FOM score.

The FOM calculation results for each transmitter are summarized in TABLE 3. It is obvious from the in-band FOM scores that category A synthesized transmitters generate fewer significant in-band spurious emissions than do the other two categories. Those transmitters featuring a low-pass harmonic filter at the output have higher out-of-band FOM scores than those that do not. The high AN/GRC-144 FOM scores are due to the excellent transmitter output filtering designed into the unit.

TABLE 3

SUMMARY OF TRANSMITTER SPURIOUS EMISSION FOM SCORES

Transmitter	Synthesizer Category	S_{SEI}	S_{SEO}	S_{SE}
AN/ARC-159	C	-15.6	57.0	20.7
AN/ARC-164	A	45.2	104.4	74.8
AN/ARC-182				
30-88 MHz	A	80.5	45.4	63.0
118-174 MHz	A	92.9	42.4	67.7
225-400	A	24.5	66.9	45.7
AN/WSC-3	B	-25.7	69.1	21.7
AN/WRT-2				
Bands 1-6	B	35.2	10.2	22.7
Bands 7-12	C	0.6	-22.6	-11.0
AN/GRC-144	C	89.8	131.7	110.7
AN/PRC-124	B	69.7	104.6	87.1

Spurious Emission Frequency Model

The output frequencies of all spurious emission sources analyzed in this study could be described by a linear equation as a function of the transmit frequency and of constant oscillator frequencies. The Reference 1 model for spurious frequency determination was found to be adequate for all spurious emission types presented. Each source frequency was identified by a linear equation of the following form.

$$FSE_{i,j} = A_{i,j} F_j + B_{i,j} \text{ , MHz} \qquad (9)$$

where

$FSE_{i,j}$ = Ensemble of i spurious emission frequencies from transmitter j

F_j = Operating frequency of transmitter j, in MHz

$A_{i,j}$ = Unitless constant

$B_{i,j}$ = Constant, in MHz

Due to space limitations, the individual spurious emission frequencies and levels categorized are not presented here.

Spurious Emission Level Model

A refined model resulted that used empirical methods to estimate transmitter driver and final amplifier selectivities rather than circuit analysis. The model was similar to the log-normal level model of Reference 1 with the exception that additional rejection terms would be added to account for selectivity. Mean and associated standard deviation rejection levels would be assumed at a point prior to filters in the transmitter, and the total rejection would be the sum of the spurious emission rejection and any filter rejections. Figure 3 is a simplified presentation of the model. The number and types of inter-driver stage filters are estimated from transmitter block and schematic diagrams. The selectivity functions for each filter are to be represented by simple low-pass, high-pass, or bandpass functions, as applicable. Multiple measured data points are required in order to estimate the various attenuation and rejection functions required by the model. Limited modeling efforts, for the AN/ARC-159 and AN/PRC-124, have shown expected improvement over the Reference 1 mean value model. The standard deviations of predicted values from measured values were approximately half the value for the original (Reference 1) model.

RESULTS AND CONCLUSIONS

1. A majority of the transmitters involved in EMC analyses at ECAC utilize PLL frequency synthesizers for frequency control.

2. Transmitter PLL frequency synthesizers, unless carefully designed, generate a significant number of narrowband spurious emissions at significant power levels resulting in low spurious emission FOM scores.

3. Frequencies of narrowband spurious emissions may be determined based on the transmitter configuration as shown in a system block diagram.

4. The power levels from individual in-band narrowband spurious emission sources are fairly consistent (within 6 dB) between serial numbers and independent of tuned frequency for transmitters with broadbanded driver filtering.

5. Tracking narrowband filters, in the driver stages of transmitters so configured, apply greater attenuation to modify the large frequency separation spurious emission levels, suggesting that the emission sources are located prior to the filtering.

6. All the potential narrowband spurious emission sources hypothesized as being present within synthesized transmitters were found to exist within three or more of the seven transmitter types analyzed. These sources were:

Oscillators

Mixers

Frequency Multipliers

Frequency Dividers/Prescalers

Intermediate Frequency Amplifiers

Phase Detectors

Harmonics

7. Simplified empirical spurious emission frequency and level models were developed. They are given in Figure 3.

8. PLL frequency synthesized transmitters have improved spurious emission characteristics if the synthesizer voltage controlled oscillator (VCO) operates at the transmit frequency. Synthesizers with VCO frequencies that are heterodyned to the transmit frequency have more significant spurious emissions.

9. Statistical variations found for transmitter harmonic levels suggest that an accurate harmonic level model is not feasible without additional study.

ACKNOWLEDGEMENT

The work contained herein was performed under Contract No. F-19628-85-C-0071 for the Department of Defense at the Electromagnetic Compatibility Analysis Center, Annapolis, Maryland.

REFERENCES

1. Lustgarten, M. N., "COSAM (CO-SITE ANALYSIS MODEL)," pp 394-406, 12th G-EMC, IEEE Electromagnetic Compatibility Symposium Record, Anaheim, CA, July 14-16, 1970, IEEE 70C28-EMC.

2. Egan, William, F., Frequency Synthesis by Phase Lock, New York, NY, John Wiley & Sons, 1981.

3. Gorski-Popiel, J. (editor), Frequency Synthesis: Techniques and Applications, New York, NY, IEEE Press, 1975.

4. Hayward, W. H., Introduction to Radio Frequency Design, Englewood Cliffs, NJ, Prentice-Hall, Inc., 1982.

5. Kroupa, Venceslav, Frequency Synthesis Theory Design and Applications, New York, NY, John Wiley & Sons, 1973.

6. Manassewitsch, Vadim, Frequency Synthesizers Theory and Design, New York, NY, John Wiley & Sons, 1976.

7. Viterbi, Andrew, Principles of Coherent Communication, New York, NY, McGraw-Hill, 1966.

8. P.N.A.P. Rao, G. Boopathy, and Jagdish Prasad, "EMC Math Models of Voltage Controlled Oscillators and Frequency Synthesizers Analysis and Measurements," pp 445-454, Paper 17PB4, IEEE 1984 International Syposium on Electromagnetic Compatibility, Tokoyo, Japan, October 16-18, 1984, IEEE 84-CH2097-4.

9. Borris, Jaime A., "Improve Synthesized Transceiver Performance and Reliability by Simple Screening of the VCO Active Device," pp 87-96, Proceedings of RF Technology Expo 86, Anaheim, CA, January 30-February 1, 1986.

10. Foot, Norman J., WA9HUV, "Low-Noise Phase-Locked UHF VCO Part 1: The Noise Problem," pp 33-38, HAM RADIO MAGAZINE, July 1986.

11. Robins, W. P., Phase Noise In Signal Sources (Theory and Applications), London, U.K., Peter Peregrinus Ltd., 1982.

12. "Special Issue: Final Report of Ad Hoc Committee on an Electromagnetic Compatibility Figure of Merit (EMC FOM) for Single-Channel Voice Communications Equipment," IEEE Transactions on Electromagnetic Compatibility, Volume EMC-17, Number 1, February 1975.

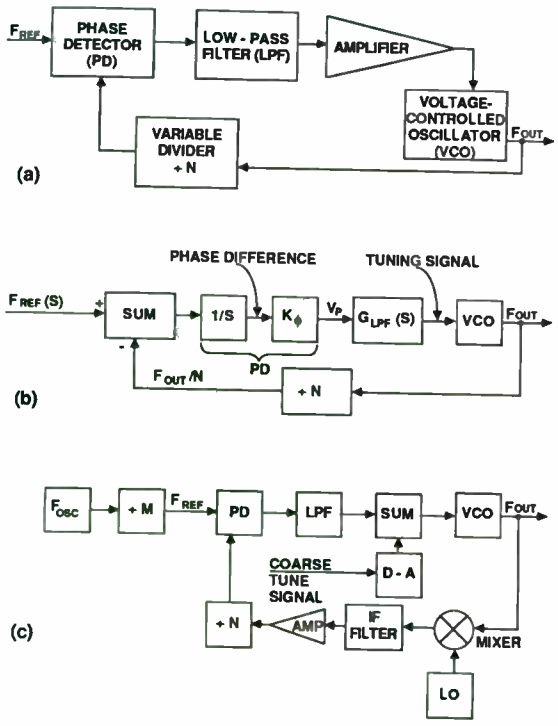
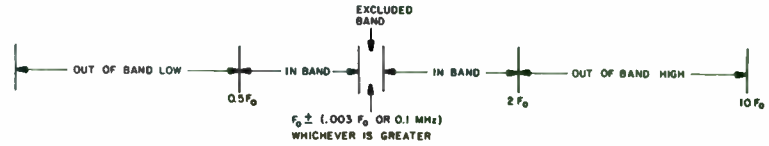


Figure 1. Synthesizer phase locked loops: (a) Basic synthesizer; (b) Mathematical representation; (c) With added components to cover a different output frequency, to allow use of a higher reference oscillator frequency, and to allow more rapid tuning.



Excluded Band:

$F_o \pm (0.003\ F_o$ or 0.1 MHz) (whichever is greater).

Emissions in this range are not considered

In-Band Range:

0.5 F_o to 2 F_o (inclusive, except for excluded band).

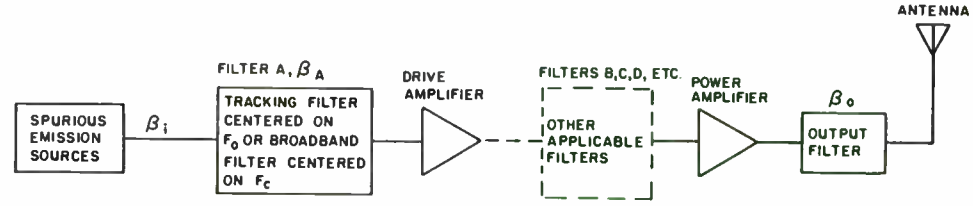
Includes second harmonic and first subharmonic.

Out-of-Band Ranges:

Low Band: 0.1 F_o to < 0.5 F_o.

High Band: > 2 F_o to 10 F_o.

Figure 2. Illustration of band definitions.



$$\beta_{SE(i)} = \beta_i + \beta_A\ (FSE_{i,j}) + \beta_B\ (FSE_{i,j}) + \ldots + \beta_o\ (FSE_{i,j}),\ \text{dBc} \quad (10)$$

where

$FSE_{i,j}$ = spurious emission frequency as defined in Equation (9), MHz

$\beta_k\ (FSE_{i,j})$ = rejection offered by filter k at a frequency of $FSE_{i,j}$, dB

$\beta_o\ (FSE_{i,j})$ = rejection offered by the output filter at a frequency of $FSE_{i,j}$, in dB

β_i = attenuation of the i'th spurious emission source, in dBc

Figure 3. The proposed spurious emission level model.

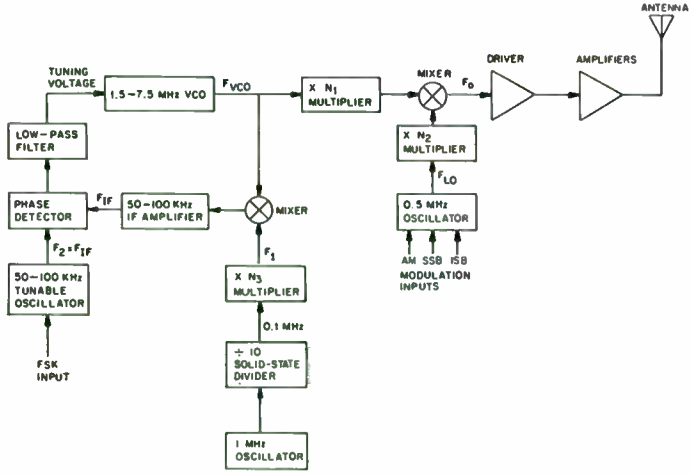


Figure 4. AN/WRT-2 frequency synthesizer block diagram.

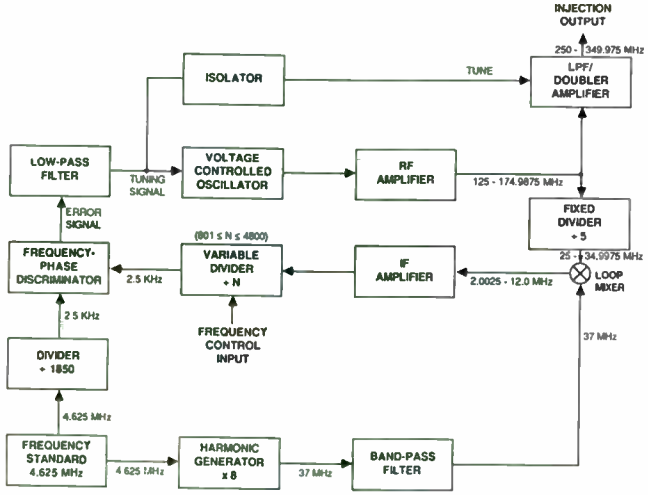


Figure 5. AN/ARC-159 frequency synthesizer block diagram.

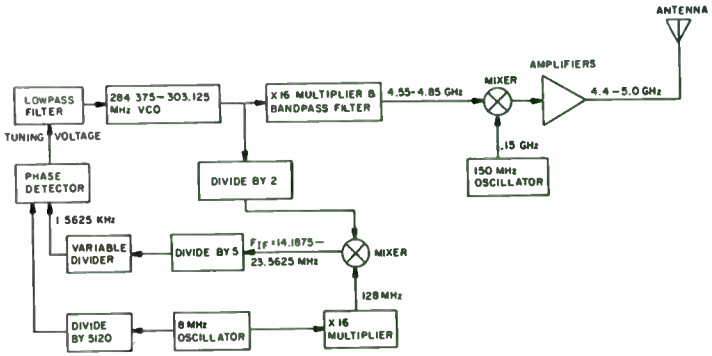


Figure 6. AN/GRC-144 synthesizer functional block diagram.

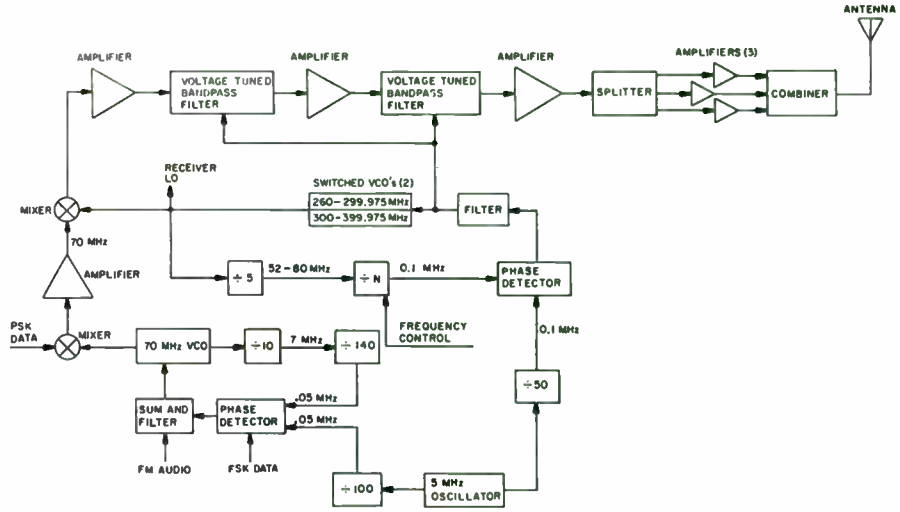
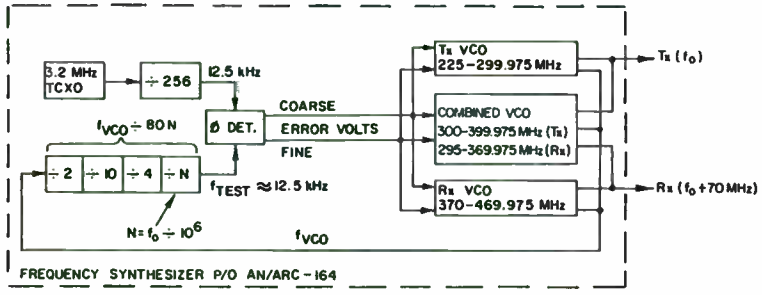


Figure 7. AN/WSC-3 synthesizer functional block diagram.



FREQUENCY EXAMPLE

f_o = 310 MHz

f_{VCO} ÷ 80 = 3.875 MHz

$f_{TEST} = \frac{3.875 \times 10^6}{310} = 12.5$ kHz

Figure 8. AN/ARC-164 frequency synthesizer simplified block diagram.

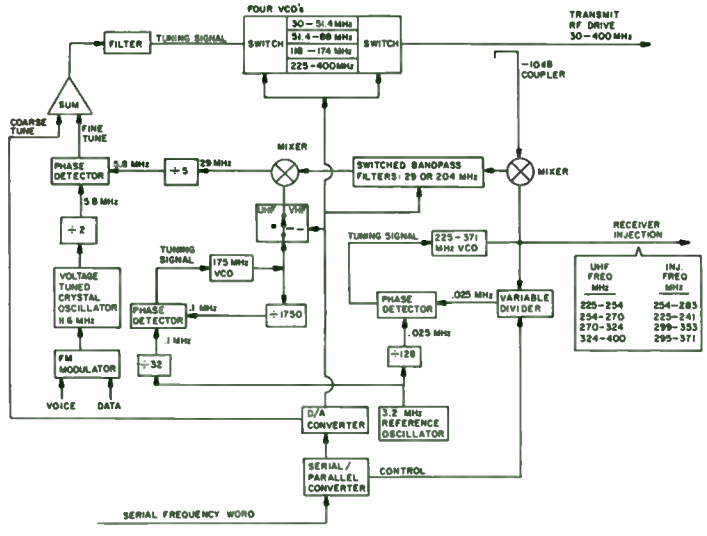


Figure 9. AN/ARC-184(V) frequency synthesizer block diagram.

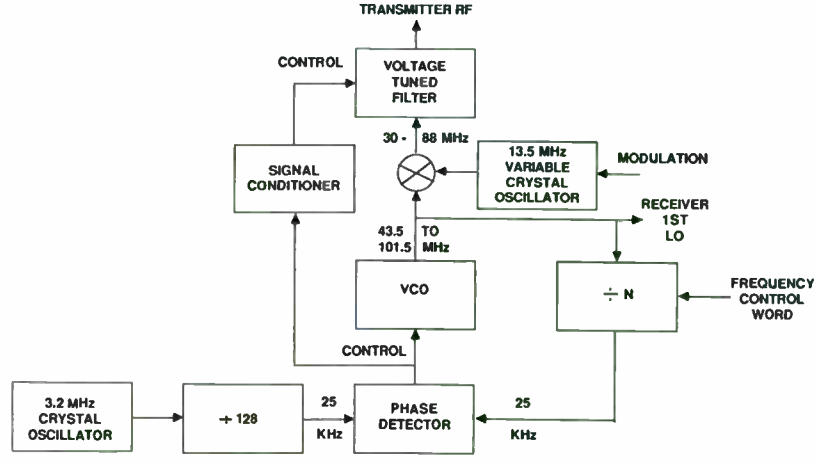


Figure 10. AN/PRC-124 frequency synthesizer block diagram.

A FREQUENCY AGILE BPSK DEMODULATOR

By
R.D. Roberts
Staff Engineer
Harris Corporation
Government System Sector
Communication Systems Division
MS 1-5855
Box 91000
Melbourne, Fl 32902

ABSTRACT

This paper describes a frequency agile BPSK demodulator which results from embedding a frequency synthesizer inside a Costas loop. The resulting receiver does not use an IF stage and is suitable for use in a system where all frequency generating sources can be phased locked to a common reference signal. An example of such a system would be a broadband local area network. The step response of such a configuration is investigated.

INTRODUCTION

Traditionally, demodulation of BPSK is accomplished using either a Costas loop or a squaring loop. Both of these methods generally work best at some IF frequency since the carrier recovery loop usually has a narrow tuning range. If the demodulator itself were tunable over a given frequency band than a cost savings could be realized by the omission of the IF stage. For example, such a demodulator would be useful in a RF modem designed for use in a broadband local area network, where the signal dynamic range is relatively small and signal to noise ratios are generally very good.

The approach shown in Figure 1 achieves frequency agility in a BPSK demodulator by incorporating a frequency synthesizer inside a Costas demodulator loop. The synthesizer forms a minor control loop while the Costas demodulator forms a major control loop. This configuration is feasible as long as all frequency generating sources in the system are phase locked to a common reference. Figure 2 shows how this concept can be utilized in a two cable broadband local area network.

A HEURISTIC EXPLANATION

Referring again to Figure 1, one can see that the frequency synthesizer loop (encompassing loop filter H2) is perturbed by the Costas loop. In principle, the Costas loop is introducing a static phase shift in the frequency loop filter as explained next.

Figure 3 is a simplified block diagram of the synthesizer control loop. The loop is considered locked to a static reference input if the signal to the loop filter (out of the phase detector Kp) is zero. Unless this condition is true, the output of the loop filter will slew in such a manner as to reduce the filter input signal to zero.

If the voltage V (referring to Figure 3) is zero, the loop locks with zero phase difference between the reference signal and the output of the divide-by-N counter since there is no offset current injection. If the voltage V is changed to a non-zero value, the resulting current flow at the summing junction of the loop filter will cause the output of the loop filter to slew in a direction necessary to reduce the summed current flow at the filter input to zero. The steady state phase of the oscillator will be such that the phase error signal at the output of the phase detector cancels the offset current due to the voltage V. Thus, the phase of the oscillator can be shifted by a static amount while still maintaining the same frequency.

We next consider the phase error output of the Costas loop (from the IxQ multiplier shown in Figure 1). This phase error signal can conceptually replace the voltage V of Figure 3. The Costas phase error signal must first pass through an integrator (loop filter H1) so that a means is provided to establish a fixed voltage input to the summer prior to loop filter H2 (Figure 1). The output of loop filter H1 is proportional to the phase error between the incoming signal (scaled by a factor 1/N) and the reference signal, while the output of loop filter H2 is proportional to the frequency error between the VCO (also scaled by a factor 1/N) and the reference signal.

Referring once again to Figure 2, it can be seen that since the frequency synthesizers in both modems A and B use the same reference source, frequency coherency can be guaranteed (provided the synthesizers are tuned to the same frequency) with the phase being randomly distributed between plus and minus 180 degrees (i.e. a function of the propagation path length). Thus, rapid phase acquisition can occur once the synthesizers have settled (i.e. the synthesizer loop bandwidth is significantly greater than the Costas loop bandwidth).

ANALYSIS

Figure 4 can be used to analyze the performance of this demodulator. The outer loop formed by the signal path through both H1 and H2 is a major control loop while the inner loop comprised of just H2 forms a minor control loop. There are two inputs to the demodulator, the incoming BPSK carrier that will be demodulated (θi) and the reference signal to the frequency synthesizers (θr). We can make use of the fact that all the synthesizers on the system are slaved to the same reference frequency and relate the two signals as

$$\theta r(s)=(\theta i(s)-\mathcal{F}(s))/N$$

The use of this relationship will transform the control loops into a form that has feedforward compensation from the input θi to the minor control loop. The net result is a third order control loop with feedforward compensation. Figure 5 illustrates the resulting control loop model.

Using the techniques of major/minor control loop analysis with feedforward compensation (Ref. 5), the state equations can be derived and a simulation model for a phase step input can be constructed as shown in Figure 6. The simulation parameters are given in terms of physical constants (defined in Figure 4) as:

$$A0 = 2KoKp/T1T3$$

$$A1 = \frac{2(T2+T4)KoKp}{T1T3} + \frac{KoKp}{NT3}$$

$$A2 = \frac{2KoKpT2T4}{T1T3} + \frac{KoKpT4}{NT3}$$

$$B0 = 2KoKd/T1T3$$

$$B1 = \frac{2KoKd(T2+T4)}{T1T3} + \frac{KoKp}{NT3}$$

$$B2 = \frac{2KoKdT2T4}{T1T3} + \frac{KoKpT4}{NT3}$$

$$Z1 = KoKp/NT3$$

$$Z2 = KoKpT4/NT3$$

A closed form solution for these state equations has also been generated (Ref. 5) but it is too involved and is beyond the scope of this paper. (The final results from the two techniques are identical). Comparison of predicted and measured results from hardware consistantly corresponded within about 10 percent.

CONCLUSIONS

A frequency agile BPSK demodulator can be built by embedding a frequency synthesizer inside a Costas loop. The resulting configuration forms a third order control loop with feedforward compensation and can be analyzed using major/minor loop analysis. The frequency control loop (the minor loop) provides frequency coherence with the incoming signal while the Costas type loop (the major loop) provides phase coherence. As long as all frequency sources are phase locked to a common reference signal, the system will provide statisfactory demodulation of BPSK data without the use of an IF stage in the receiver. The loop parameters can be adjusted to provide the desired step response.

ACKNOWLEDGMENTS

This work was supported by the Harris Corporation, Government Systems Sector, Melbourne, Fl and was performed in partial fulfillment of the MSEE degree from Florida Institue of Technology, Melbourne, Fl. I want to thank Dr. N. Tepedelenlioglu of FIT for his advice and comments.

REFERENCES

1. D'Azzo, J.J. and C.H. Houpis Linear Control System and Design Conventional and Modern. New York: McGraw Hill Book Co., 1975

2. Eveleigh, V.W. Introduction to Control Systems Design. New York: McGraw Hill Book Co., 1972

3. Gardner, F.M. Phaselock Techniques. New York: John Wiley and Sons, 1979

4. Motorola Semiconductor Products Inc. Phase-Locked Loop Systems Data Book, 1973

5. Roberts, R.D. Step Response of a Third Order Phase Locked Loop with Feedforward Compensation. MS Thesis, Florida Institute of Technology, 1985

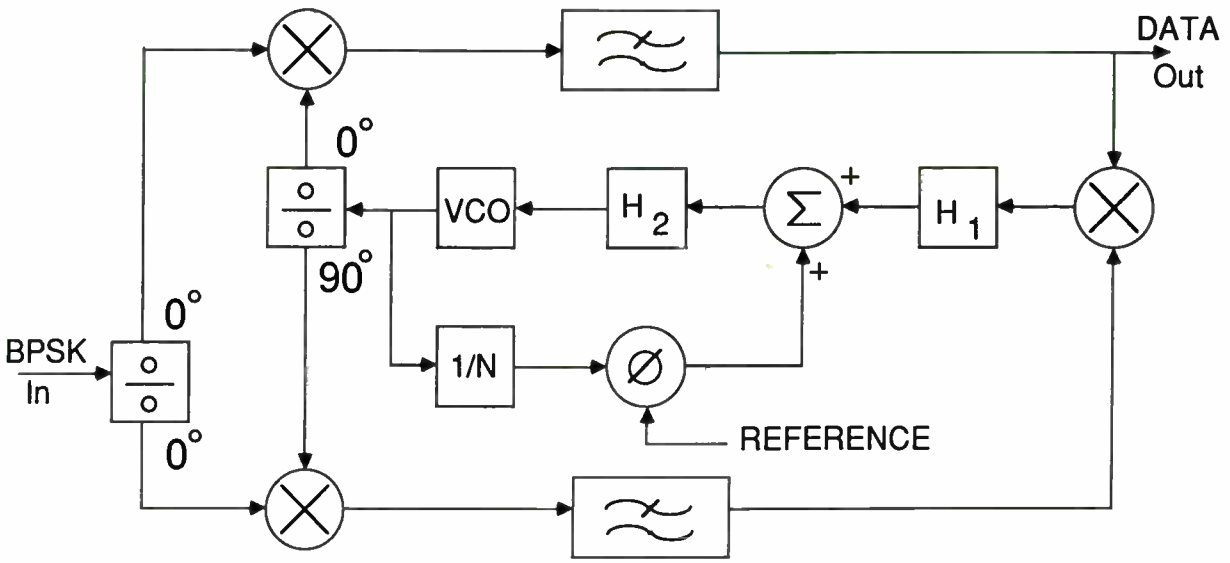


Figure 1
Block Diagram of the Frequency Agile BPSK Demodulator

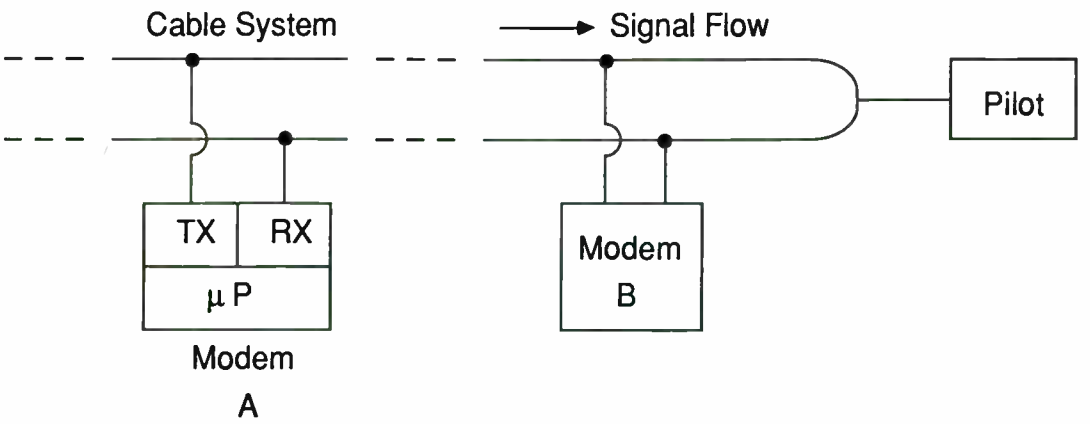


Figure 2
A Two Cable System Broadband Local Area Network with a Common Reference Tone

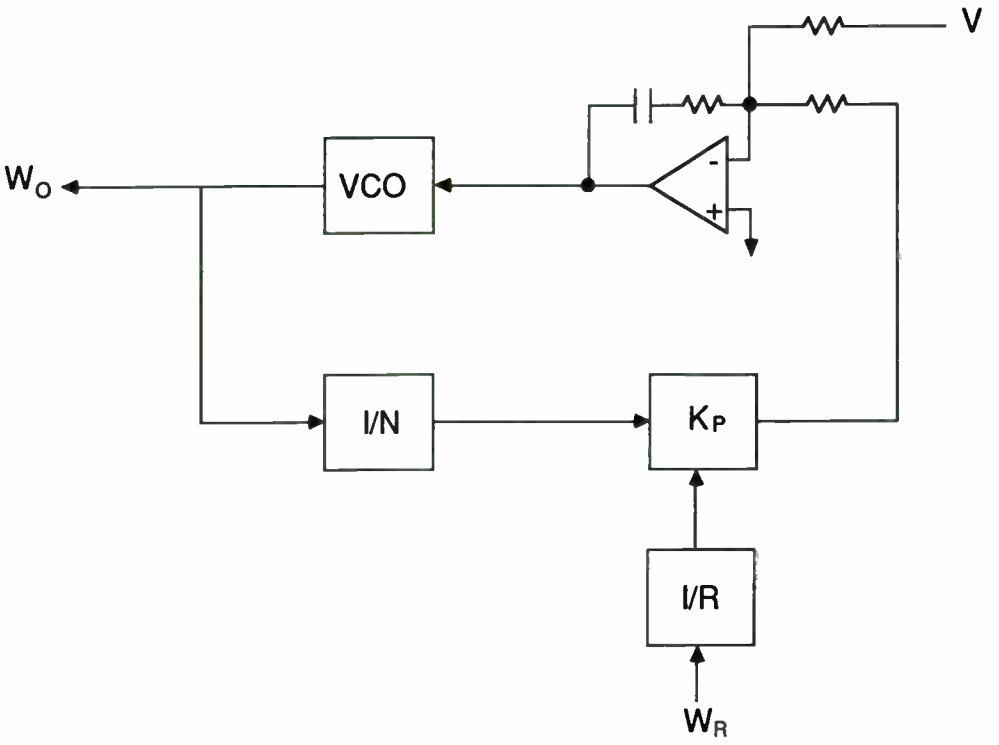


Figure 3
A Synthesizer Control Loop with an Injected Offset Voltage

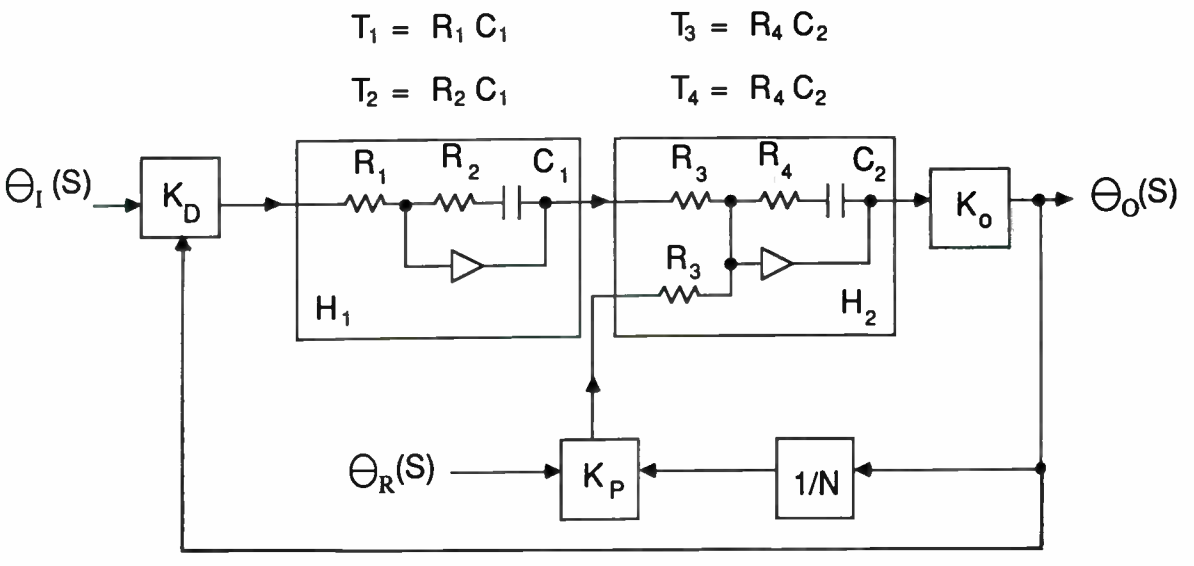


Figure 4
The Major/Minor Control Loop Equivalent

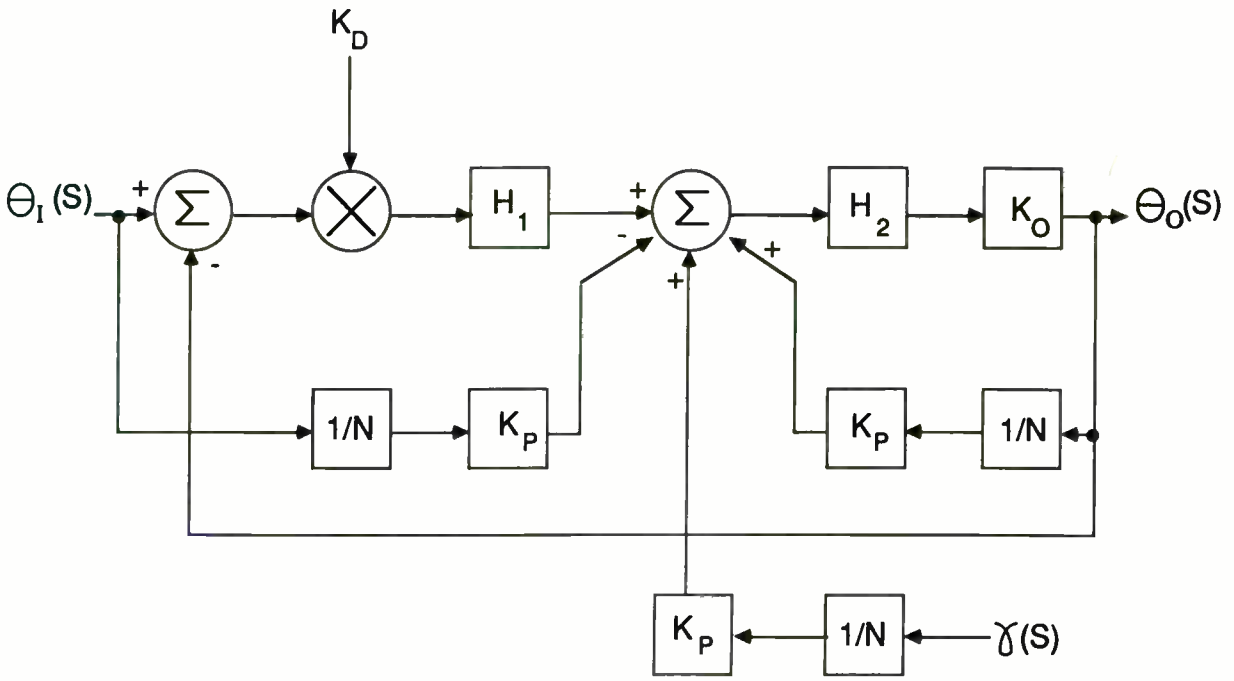


Figure 5
Linear Major/Minor Control Loop Model of the Frequency Agile
BPSK Demodulator with Feedforward Compensation

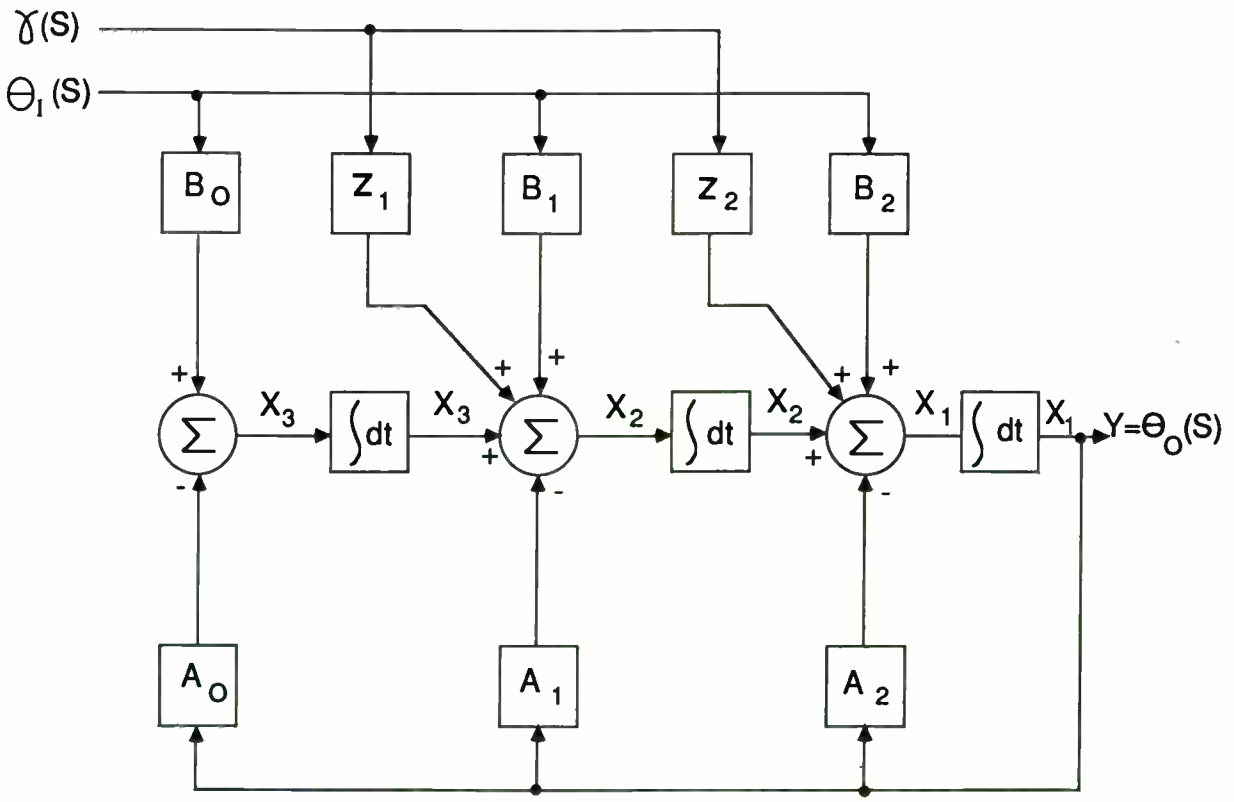


Figure 6
The Simulation Diagram for the Frequency Agile BPSK Demodulator

A GHz RF MODULATOR WITH HIGH ON/OFF RATIO

by

Seymour N. Rubin
President
SR ASSOCIATES
1553 N. Commonwealth Ave.
Los Angeles, CA 90027

INTRODUCTION

Applications for high data rate high frequency RF modulators are become more common. One such application is the driving of acousto-optical modulators and deflectors. These devices require power levels ranging to 1-2 watts and frequencies extending into the 1-2 GHz region.

The system described here used off-the-shelf RF components. This provided flexibility and rapid design. Proprietary acousto-optical modulator matching techniques allowed the use of standard devices without having to compensate for mismatches.

Pulse modulation of RF signals in the 1-2 GHz range with data rates from DC to over 250 MHz was achieved. These signals were clean with excellent power on/off ratios at high PRFs and 1 ns rise and fall times. As described, the output power was 1.6 watts though by using other output amplifiers a wide variety of power levels could be achieved.

Even though the design was conceived for use with an acousto-optical modulator, its excellent performance would allow use wherever fast modulation of GHz signals is required.

BLOCK DIAGRAM

Figure 1 presents a block diagram of the demonstration system showing the major elements and the RF levels present at each node. The configuration was optimized for switching operation, that is, the output signal either full on or full off.

RF was provided by a 1.2 GHz oscillator though any +14 dBm source could have been used. Series mixers were used to provide the required on-off ratios. The modulation input signal was split into 2 channels to drive the mixers. The phase relationship between the modulation inputs to the two mixers and their RF inputs was set by adjusting cable lengths in the two channels. This phase relationship and the RF levels to each mixer were key to the success of the design.

The RF from the oscillator drove the first mixer through an isolation and level setting attenuator. The output of this mixer was sent through an attenuator and an isolation amplifier to set the optimum drive level for the second mixer.

Output from the second mixer fed a driver amplifier through another level setting attenuator. The driver output was split into two channels, each with a high-power, wide-band amplifier. The signals from the high-power amplifiers were summed by a power combiner to provide the RF output. An attenuator in one channel matched the saturation output level of the two amplifiers so that optimum power output was achieved.

MODULATION INPUT LEVELS

Setting the modulation input signal at a +1.4 volt transition from a +0.2 volt baseline resulted in a +0.3 volt peak pulse from a +0.075 volt baseline at the mixer inputs.

These optimized levels were important to the on-off performance of the system. The non-zero baseline at the mixers was adjusted to compensate for mixer offsets and imbalances and provided maximum carrier attenuation for no modulation input.

SYSTEM PERFORMANCE

The system performance was measured using a 250MHz pulse generator with 1 ns rise and fall times at up to 5 volts output into 50 ohms as a source and a sampling oscilloscope with 100 picosecond rise time to observe the response. The duty factor was set to 50% and the pulse train was un-interrupted.

Designing a system of this nature required consideration of several key parameters: rise and fall times, on-off (or extinction) ratio, pulse aberrations, maximum modulation rate and power output. Several of these parameters are shown in Figure 2 which is a composite photograph of the RF envelope.

The rise and fall times of the envelope measured from Figure 2 were less than 1.2 ns. The extinction ratio was more than 25 dB below the peak at 2 ns after turn-off and more than 30 dB below thereafter. This was true for all measured modulation rates.

Pulse top aberrations were less than +/-10% and were somewhat dependent upon the modulation rate. As a result, there was a slight envelope amplitude variation with modulation frequency.

Figure 3 demonstrates that the amplitude change with modulation rate was minimal. It shows the envelope characteritics at various modulation rates. Note that the extinction ratio performance was good to at least 250 MHz. At that rate, the envelope stays more than 20 dB below the peak for 0.9 ns.

Peak power output was measured at 1.6 watts from DC to over 250 MHz. Higher power could be achieved either by using different amplifiers or by adding more of the same type in parallel.

RESPONSE IN AN ACOUSTO-OPTICAL SYSTEM

To predict the response of an Acousto-Optical (AO) modulator to the pulses generated by this system, a sine squared function of the voltage was computed. This was done because AO modulators respond to power and approximately conform to the sine squared function. Data from Figure 2 was used in the calculations

It was assumed that the pulse modulated RF signal would drive the AO modulator to its maximum response point. A plot of linear and sine squared rise times is shown in Figure 4. It is seen that the 10-90% rise time was 1.18 ns for the linear voltage plot and 0.82 ns for the sine squared response. The overshoot was 7% for the linear case.

When the rise time of the modulating signal is taken into account the calculated voltage rise time is 0.63 ns. However, since there were nonlinearities in the mixer, this result may not be totally accurate. It is evident, however, that the rise time would have been less than 1 ns if a sub-nanosecond drive signal were available.

A similar process was used to plot the fall time. This is given in Figure 5. From this data it is seen that the 10-90% fall time was 1.07 ns for the linear waveshape and 0.7 ns for the sine squared function. This plot also shows that the extinction ratio was quite good. At 3.5 ns after the 90% point on the fall time, the envelope had dropped more than 30 dB below peak. At 2.5 ns it was 25 dB below the peak.

SUMMARY

A high speed RF modulation method with high on-off ratios, fast modulation rates, fast rise times and low pulse aberrations has been demonstrated. Though the system was developed for use with acousto-optical modulators its speed and flexibility offer advantages for many other applications such as data communications or diode laser modulation.

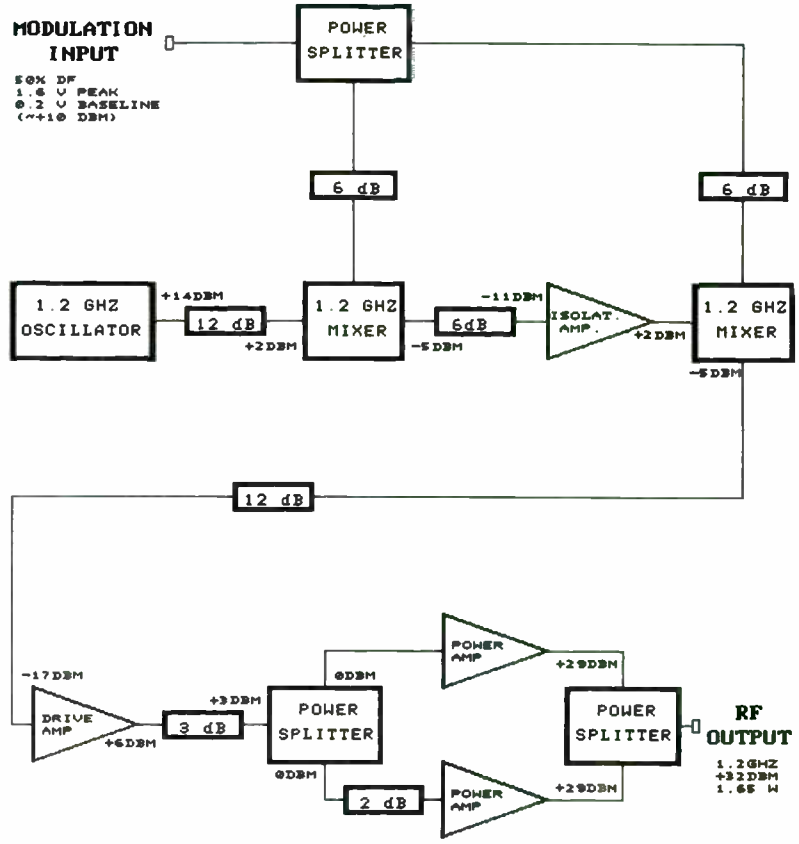


Figure 1: System Block Diagram

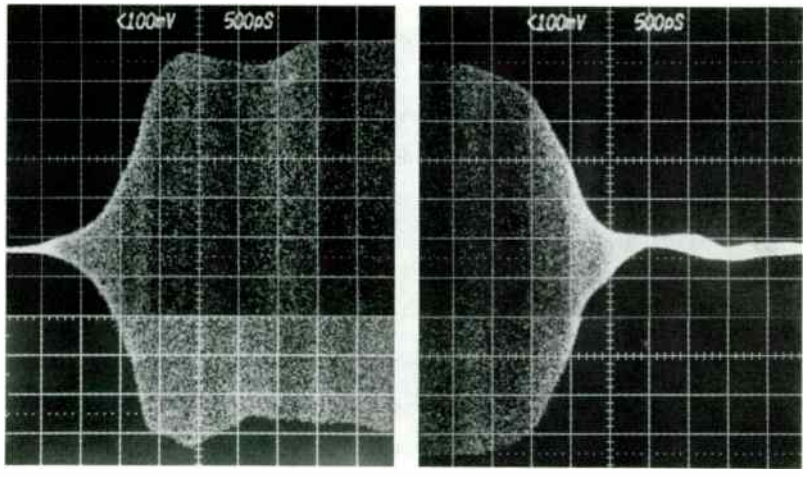
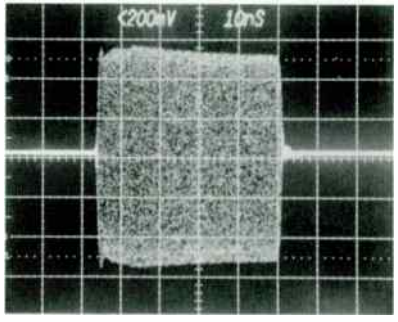
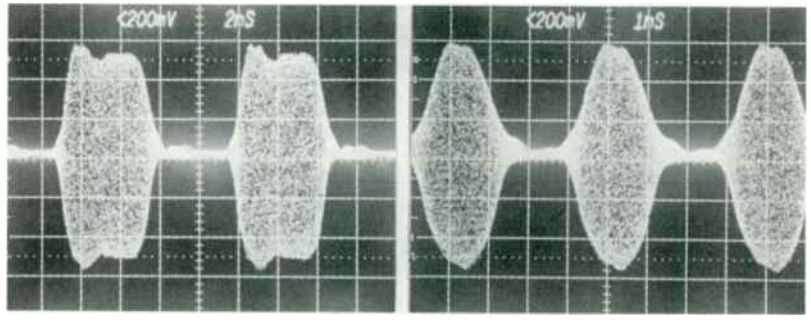


Figure 2: Composite photograph of the RF output envelope.



(a)



(b) (c)

Figure 3: Waveshapes at various modulation rates: (a) 10 MHz. (c) 100 MHz. (d) 250 MHz.

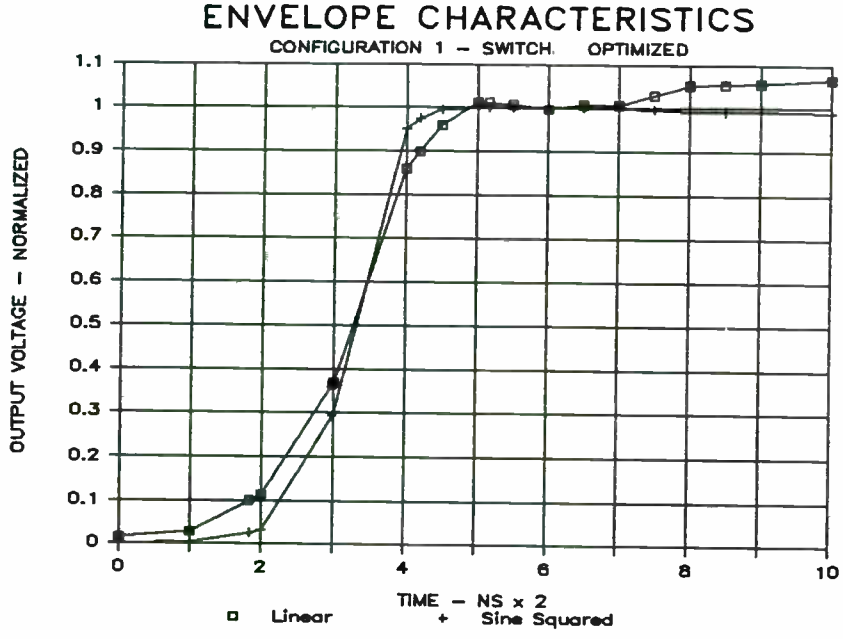


Figure 4: Calculated envelope rise time for the linear and sine squared cases. Data from Figure 2 was used.

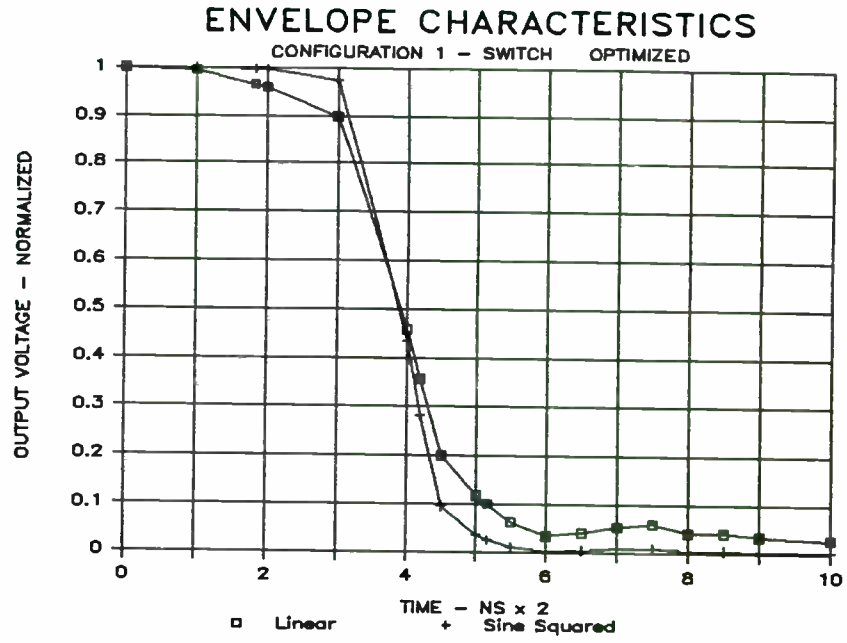


Figure 5: Calculated envelope fall time for the linear and sine squared cases. Data from Figure 2 was used.

FILTER DESIGN TECHNIQUES USING PERSONAL COMPUTERS

Michael K. Ferrand
Senior Microwave Engineer
Microlab/FXR
Ten Microlab Rd.
Livingston NJ 07039

ABSTRACT

Intelligent design and layout along with careful documentation can take the mystery out of microwave filter design. Two program listings are offered to speed design of elliptical and cavity filters. In addition discussion on design techniques is given.

The Microwave Filter

The ubiquitous microwave filter, often the systems designers after thought plays a major role in every part of todays communications equipment. From the HAM radio to a complex ECM pod, filters are used to shape, discriminate or combine signals. The types of microwave filters are as varied as their uses. The huge variety of technologies and techniques makes microwave filters one of the most complex and often misunderstood parts of todays microwave systems. Filters, like the other components in systems have continued to shrink in size over the years. The advances in low-noise semiconductors have allowed engineers to replace large high Q cavity and waveguide filters with relatively low Q devices with lumped element or distributed components. As we look to the future, filters will become integral parts of sub-systems in an effort to reduce parts count and enhance reliability. Incorporation into MMIC's is expected before the end of this decade.

The 1950's saw,the advancement of filter theory and techniques. Point by point measurements using slotted lines and crystal detectors were used. Filters were often aligned with hammers and screw drivers and thus was born the "Guru" who seemingly coould make these parts work as if by magic. In the next two decades we enjoyed the use of swept measurement techniques and tuning could be done in "Real Time". Many of the novel techniques of the 1950's were made common place by advancements in manufacturing techniques. The1980's saw filters being aligned with the help of computers. Instant information on S-Parameters and circuit analysis became available on desk top computers. The 1990's and beyond will no doubt see filters created in the foundry. Their capacitors, inductors and lengths of transmission lines incorporated into MMIC's. The next generation of computer programs must provide not only more accurate element models but account, and compensate for circuit parasitics. Only with the advent of this "intelligent" software give us the ability to produce these parts on such a scale.

As we look back at the evolution of microwave components we see the human element involved. Time and time again the success failure of a project may hinge on this "guru" who could perform their magic with a razor blade or knife. This leaves projects vulnerable to the health, punctuality and temperament of this person. Often they may feel that their talent (and not sharing its secrets) gives them job security and leverage in performance reviews. In reality this attitude benefits no one. By sharing such knowledge with your co-workers you extend your worth to the company. Often steps must be taken to insure the repeatability of a design regardless of the personnel involved. Some of the ways to reduce the problems trying to repeat a design are as follows:

Documentation: Complete manufacturing drawings, procedures and notes from the engineer, technician and assembler.

Electrical Data: Charts, graphs and point to point data will be useful the next time the job is done.

Sample Unit: Although not always practical, its good to have an extra working unit.

Enter, The Computer

With the advent of the personal computer, programming has been taken from the realm of the computer scientist and handed to the engineer. Filter design which was once the domain of the slide rule and graph paper has become a fast and accurate on a PC. For an established engineer being "computer literate" is a benefit, for a beginning engineer it's a must! Most PC's available today come packaged with at least the BASIC language but as one looks to the software market a cornucopia of languages are available. Basically all computer languages contain the logic statements necessary to

perform elementary math functions, conditional testing, looping, formatted input/output. Thus the choice of which language is used depends largely on the users preference for the syntactical differences. The four languages that are in the greatest use in the engineering environment are FORTRAN, BASIC,PASCAL and C. Since much of our programming is going to be math intensive the extent of the mathematical functions supported by the language may be a factor. Only FORTRAN will directly support complex numbers and PASCAL and C do not include X^y. These minor differences are easily compensated for by user defined functions. It is usually assumed that the person using the program has familiarity with it and the type of design they are attempting. A person entering design parameters that are wrong can only receive results that are wrong, recall the axiom "Garbage In, Garbage Out". Engineering programs are notoriously "User Un-Friendly" and usually no thought is given to the range of acceptable values. Given a limited set of input values one can create algorithms that work quicker and more accurately.

The Design Examples

Two design examples are given to show the usefulness of the self written programs to the working engineer. While the designs and programs shown are certainly not "State of the Art" they do represent practical applications of existing design information. The program listings (in BASIC) will be available at the RF Expo Show.

EXAMPLE I Elliptical Highpass Filter

This example shows the usefulness of a unique design approach for a ultra-wideband highpass filter.

Passband: 1-18 GHz.
Stopband: ≥ 60 dB at 0.5 GHz. (must provide D.C. block)
Passband Insertion Loss: ≤ 1.0 dB
Passband VSWR: ≤ 2.0:1

This filter was design by using the program ELLIP.BAS. The design criteria are entered and the program calculates the element values. The filter is constructed with capacitively coupled sections of 50 Ohm transmission line. The inductor/capacitor combination gives the transmission zeros to complete the elliptical circuit.

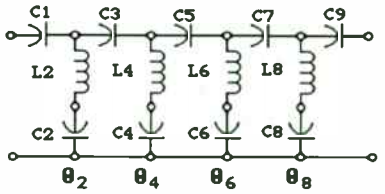


FIGURE 1

The capacitively coupled sections are constructed by using a teflon rod through the centers, each section being threaded on the rod. The capacitor are formed by using teflon spacers to form the dielectric separating the lengths of transmission line. The needed spacing is calculated by the following formula:

$$\text{Spacing (ins.)} = \frac{Cpf * a}{.2246 * er}$$

where: Cpf is the desirad capacitance in picofarads.
a is the total area of the capacitor plates
er is the dielectric of the insulator (2.1 for teflon)

The traps are formed by a series inductor-capacitor combination to ground. It was important to keep inductor lines (high impedance sections) as short as possible. Line lengths must be less then 1/4 wavelength up to highest passband frequency. To provide good performance to 18 GHz. line lengths had to be kept less then 0.164 ins. The shunt capacitors were standard value 25 mil. chips. Chips were soldered into the housing, then after completing assembly the inductor wires were attached.

The actual response for a production unit is shown at the end of the article. This unit shows insertion loss of ≤ 0.6 dB. to about 17.5 GHz. increasing to 0.9 dB. at 18 GHz. Stopband attenuation at 500 MHz. was measured with a spectrum analyzer and found to be 63 dB. The return loss is typically better then 12 dB. until about 17.5 GHz. The passband and return loss degrades rapidly above 18 GHz. as the line lengths approach 1/4 wavelength.

Example II. Quadraplexer Filter

This example shows the use of a common type of microwave filter in a

rather unusual application. Four signal bands are to be filtered and combined to form a single wide band signal for processing. The requirements for the bands are as follows:

Band 1 7.700 - 10.225 GHz.
Band 2 10.450 - 12.875 GHz.
Band 3 13.000 - 15.525 GHz.
Band 4 15.700 - 18.225 GHz.
Cross-over points: 10.3375, 12.9375, 15.6125 GHz.
Insertion Loss over passband: ≤ 4.0 dB.
Stopbands: ≥ 60 dB @ X-Over +/- 15%
≥ 60 dB. from D.C to 7.25 GHz.
≥ 60 dB. form 20.0 to 26.5 GHz.

This design used the program INTCL.BAS to design the 4 separate interdigital filters. Bands 3 & 4 required 19 sections, Band 2, 21 and Band 1 23 sections to achieve the desired response. In addition to this a lowpass filters were attached to the ends of the filters outputs to provide harmonic rejection and signal combination. A partial schematic is shown below:

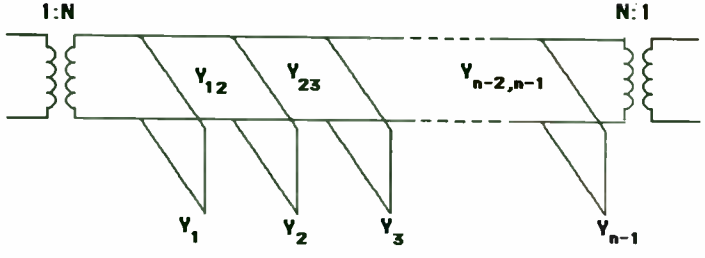


FIGURE 11

The complete design of this filter was inordinately complex so each filter was design separately as was the filter/combining network.The input-output ports of the filters were realized by selecting tap points on the first and last resonators. The exact position of these "tap" points was accomplished by use of a vector network analyzer. The design of the prototype unit allowed the resonators to be shorted at the end. With the first post shorted a zero phase reference was set. Then the first tuning screw was moved away from the short. A bandwidth measurement was made to determine the range over which the phase changed 180 degrees. The loading or "Q" bandwidth is determined by:

$$Q_{in} = \frac{1}{G_1 * \text{Bandwidth}} \qquad Q_{out} = \frac{1}{G_n * \text{Bandwidth}}$$

where: G_1 & G_n are normalized lowpass element values

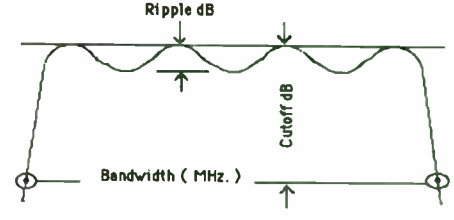
After the correct tapping point is located the width / spacings are confirmed by opening each section sequentially and measuring each additional 180 degree change in phase. These bandwidths (Coupling Bandwidths) are calculated from the following formula.

$$K_{j,j+1} = \frac{1}{\sqrt{G_j * G_{j+1}}} * \text{Bandwidth}$$

where: j=1 to n

G_j & G_{j+1} are normalized lowpass element values

To design with these normalized coupling bandwiths the equil-ripple bandwith of a chebyshev filter is multiplied by a factor "H" to scale to a 3 dB bandwidth.



$$H = \text{Cosh}\left[\text{Cosh}^{-1}\left(\frac{1}{EPS3} * EPSR\right)\right]$$

where $EPS3 = \sqrt{10^{0.3} - 1}$

$EPSR = \sqrt{10^{R} - 1}$ R= Ripple in dB.

This technique is particularly valuable for prototyping interdigital and

combline filters. The production units have a response typical to the curve shown at the end of the article.

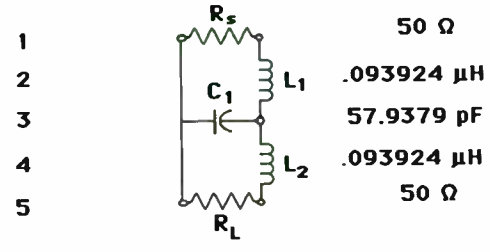
S-Parameter Analysis

For lumped component designs the S-Parameter method is used for analysis of the filter designs. The program begins by defining the circuit in terms of its general circuit parameters:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix}$$

where **A** is a voltage transformer
B is a series impedance
C is a shunt admittance
D is a current transformer

Let's suppose we wish to analyize this lowpass circuit assuming infinite Q.



A series impedance is represented in **[ABCD]** format as:

$$\begin{vmatrix} 1 & Z \\ 0 & 1 \end{vmatrix}$$

Each inductor in our example would be represented this way:

$$\begin{vmatrix} 1 & jwL \\ 0 & 1 \end{vmatrix}$$

A shunt admittance is represented in **[ABCD]** format as:

$$\begin{vmatrix} 1 & 0 \\ Y & 1 \end{vmatrix}$$

The capacitor in our example would be represented this way:

$$\begin{vmatrix} 1 & 0 \\ jwC & 1 \end{vmatrix}$$

To cascade we simply multiply:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} 1 & jwL \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ jwC & 1 \end{vmatrix} \begin{vmatrix} 1 & jwL \\ 0 & 1 \end{vmatrix}$$

Next a transformation is made from **[ABCD]** to S-parameters. parameters are directly related to the measurements made on test equipment. The

Scattering matrix may be defined as follows:

$$\begin{vmatrix} S11 & S12 \\ S21 & S22 \end{vmatrix}$$

where **S11** = Input Reflection Coefficient
S12 = Reverse Voltage Transfer Coefficient
S21 = Forward Voltage Transfer Coefficient
S22 = Output Reflection Coefficient

$$S11=\frac{A*Z2+B-C*Z1*Z2-D*Z1}{A*Z2+B+C*Z1*Z2+D*Z1}$$

$$S12=\frac{2*(A*D-B*C)*\sqrt{Z1*Z2}}{A*Z2+B+C*Z1*Z2+D*Z1}$$

$$S21=\frac{2*\sqrt{Z1*Z2}}{A*Z2+B+C*Z1*Z2+D*Z1}$$

$$S22=\frac{-A*Z2+B-C*Z1*Z2+D*Z1}{A*Z2+B+C*Z1*Z2+D*Z1}$$

with Z1 = source impedance
Z2 = termination impedance

Finally, VSWR, attenuation, phase and delay can be calculated.

$$VSWR = \frac{1+|S_{11}|}{1-|S_{12}|}$$

$$Atten\ (dB.) = 20 * Log_{10}\left|\frac{1}{S_{21}}\right|$$

$$Phase\ (deg.) = -Atn\left[j\frac{A*Z2+B+C*Z1*Z2+D*Z1}{A*Z2+B+C*Z1*Z2+D*Z1}\right]$$

$$Delay = \frac{\Delta\ phase}{2*\pi*\Delta\ freq.}$$

Conclusion

Practical design experience combined with the careful layout of components are the basic steps to a successful prototype design. In addition to this, documentation of techniques and details can make future production runs easier. The application of user written programs, even those based on published information can be a great design tool and time saver.

References:

Program: INTCL.BAS

1. Conformal Transformation Combined with Numerical Techniques, with Application to Coupled-Bar Problems
Ralph Levy
IEEE Microwave Theory and Techniques. April 1980

2. Tapped-Line Coupled Transmission Lines with Applications to Interdigital and Combline Filters
Edward G. Cristal
IEEE Microwave Theory and Techniques. December 1975

3. A Study of the Phase and Filter Properties of Arrays of Parallel Conductors Between Ground Planes
JT Bolljohn
IEEE Microwave Theory and Techniques. March 1962

4. Coupled Circular Cylindrical Rods Between Parallel Ground Planes
Edward G. Cristal
IEEE Microwave Theory and Techniques. July 1964

5. Coupled Rectangular Bars Between Parallel Plates
William Getsinger
IEEE Microwave Theory and Techniques. June 1962

6. Microwave Filters, Impedance Matching Networks and Coupling Structures
Matthaei, Young & Jones
Artech House

7. Improved Single and Multiaperature Waveguide Coupling Theory, Including Explanation of Mutual Interactions
Ralph Levy
IEEE Microwave Theory and Techniques. April 1980

8. Waveguide Handbook

N. Marcuvitz
MIT Radiation Lab Series

Program: ELLIP.BAS

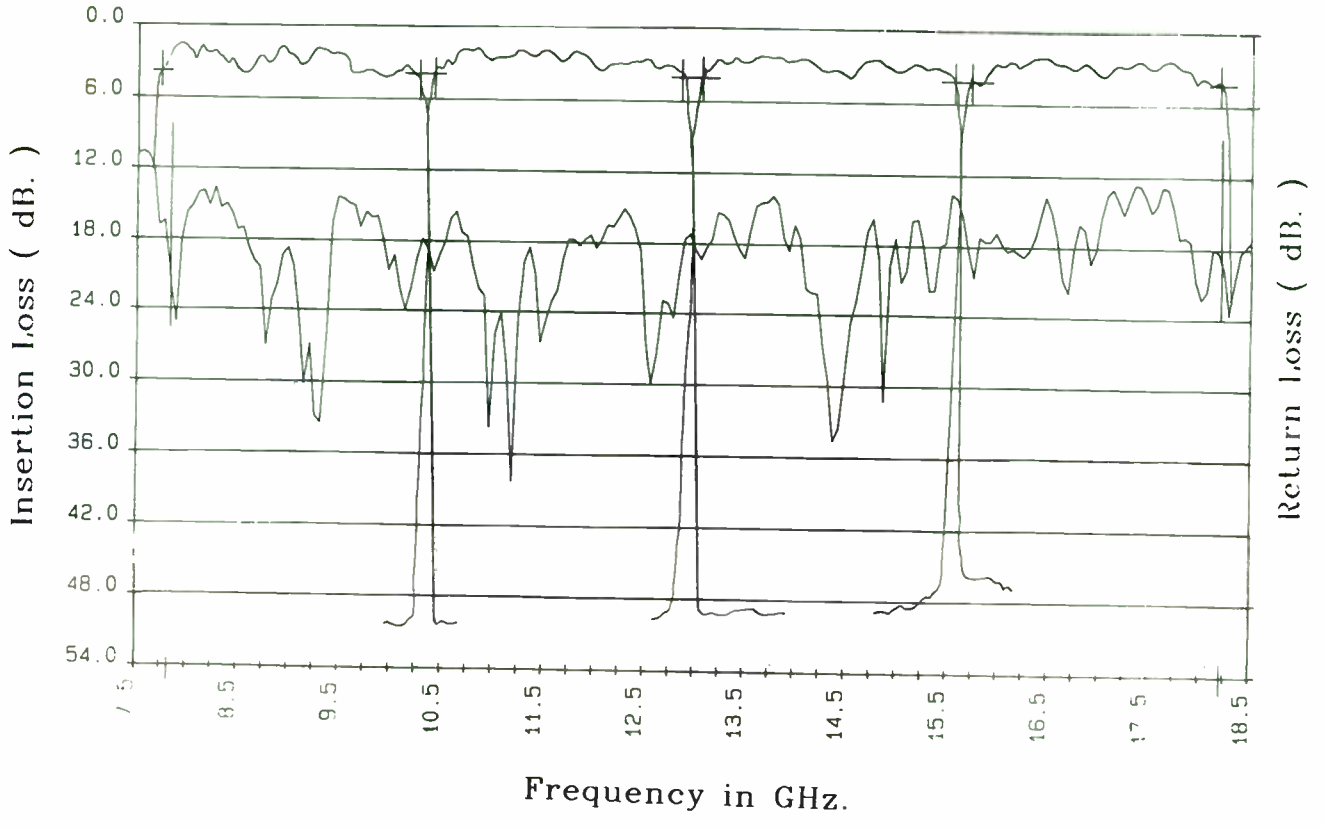
1. Handbook of Filter Synthesis
   Anatol V. Zverev
   Wiley-InterScience

2. Reference Data for Radio Engineers
   ITT

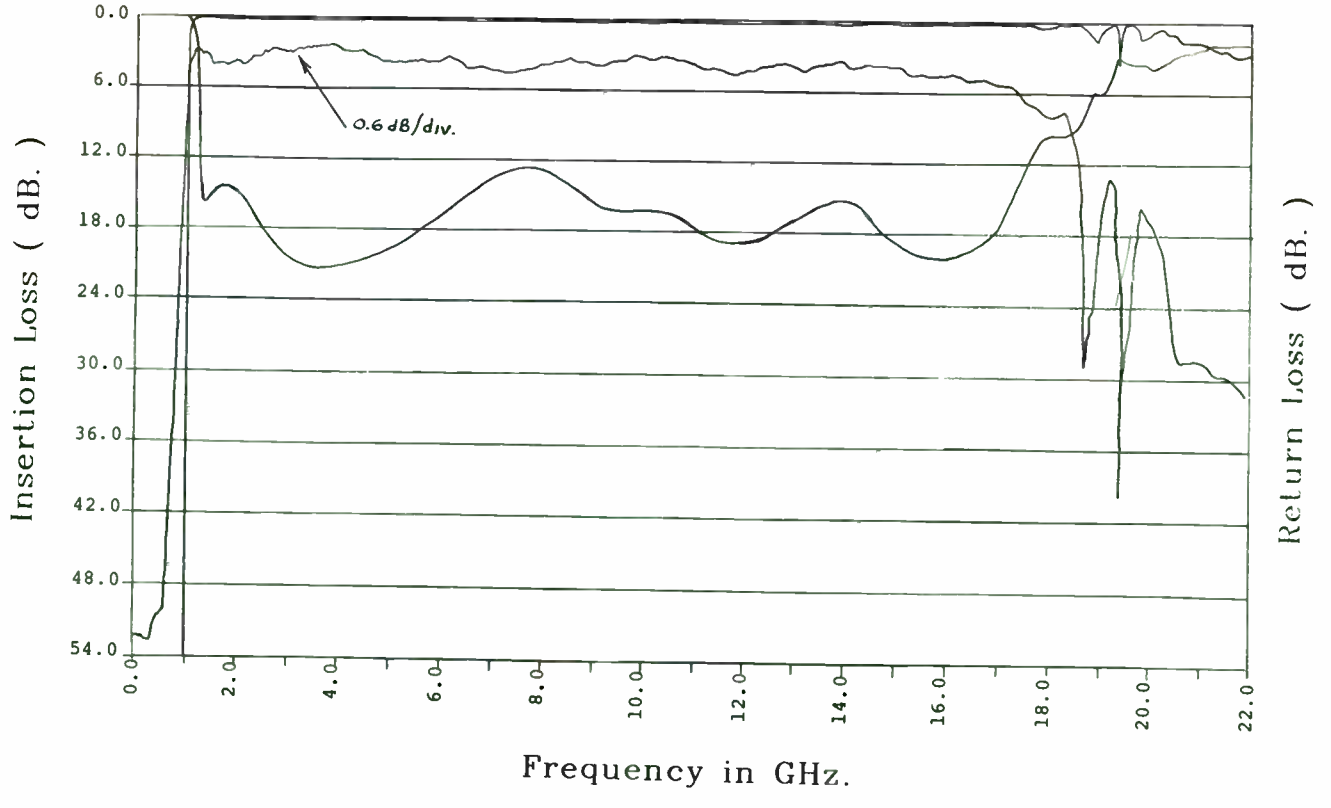
3. Filtering in the Time and Frequency Domains
   Blinchikoff and Zverev
   Wiley-InterScience

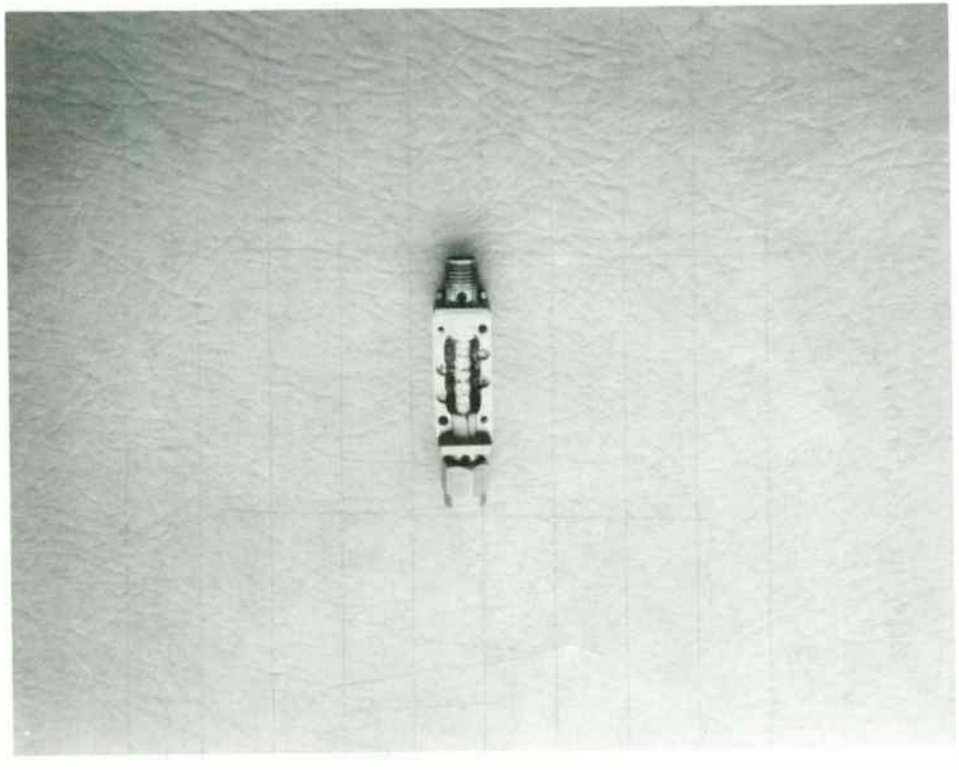
4. Elliptic Approximation and Elliptic Filter Design on Small Computers
   Pierre Amstutz
   IEEE Trans. on Circuits and Systems Dec. 1978

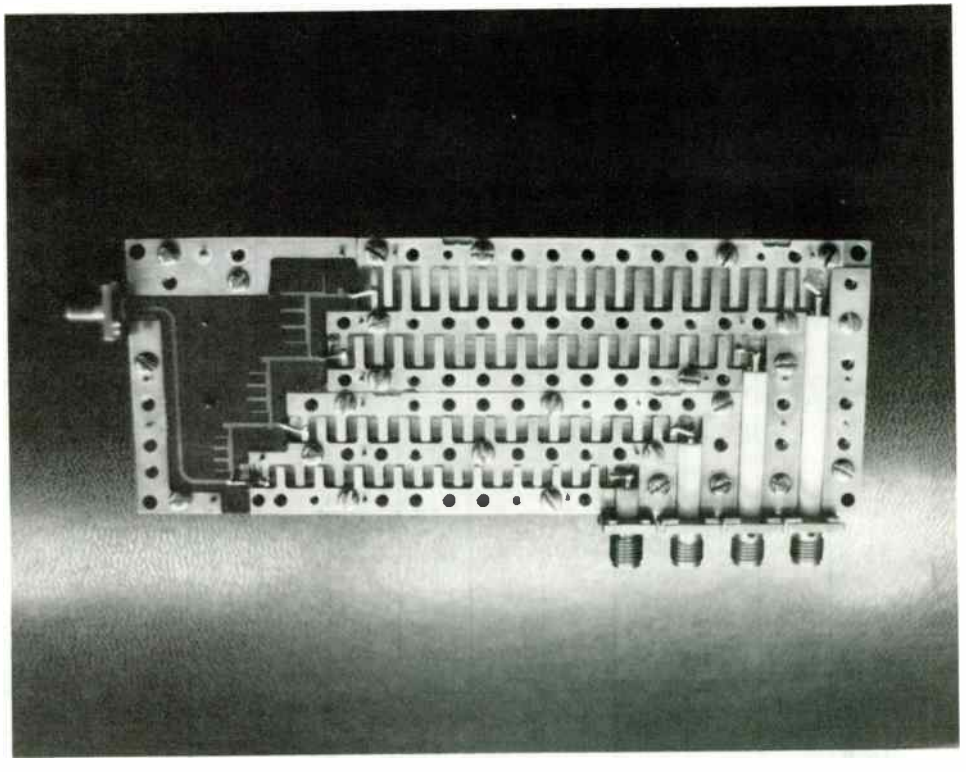
Microlab/FXR Part#: BQ-A77 Serial#: 017
Tester: rms Date: 2/22/87



Microlab/FXR Part#: HA-A75 Serial#: 003
Tester: mkf Date: 3/5/87







Operation of A Synthesizer Settling Time Test Fixture

L. M. Tichauer, J. C. Lunsford, B. J. Tarnowski & J. T. Bowles
Hughes Aircraft Company
P.O. Box 3310
Fullerton, Ca. 93634

BACKGROUND

The performance of a frequency synthesizer's phase settling time is a critical design parameter when it comes to a frequency agile, fast hopped phase modulated communication system. The magnitude or degrees of phase variation allowed before phase settling is declared is therefore very important. Phase settling is defined within the system specification. Any phase noise impressed upon the system evidences itself as phase jitter of the received signal that can obscure the exact point of a zero crossing. In an m-ary phase modulated system this is a critical parameter in determining the system's ability to differentiate between the various allowed states.

It is important that this specification be met by each frequency regardless of what the previous frequency was. This refers to the fact that some systems do not dump the residual charge on the integration capacitor in the feedback loop when changing frequencies. When this residual voltage is present, the initial conditions for a given frequency's settling time may vary. Therefore for a system with N frequencies, there will be N^2 possible frequency combinations. To properly characterize the system each combination must be checked out for each deliverable synthesizer.

It was this need that precipitated the need for an automated test set. To assure measurement accuracy it was important to remove the element of human fatigue, due to the sheer magnitude of the number of combinations to be tested. It was decided that a statistical capability also be made available for future analysis.

In addition to phase settling, spurs, harmonics and noise levels also need to be characterized by this test system and a record kept of any failures discovered during the testing process. If this was to be an automated test set, a system had to be designed that would minimize the need for human interface and make provisions available for automatic recording of the measurements. When an error is encountered, instead of generating an error signal that would have to be cleared by an operator action, a method had to be devised to make a detailed record of the error and allow the system to continue on with the measurements.

THE SYSTEM

Figure 1 shows a block diagram of the system used to carry out the task of completely characterizing the synthesizers. The controller is used to coordinate the interaction of the various pieces of test equipment required to carry out the measurements. It is also the operator interface where the initial conditions for the tests to be carried out are entered. The menu driven program asks for the serial number of the unit under test and also asks for the degree of documentation to be supplied by the test system.

The printer is used to log the test results and to record the time and type of error encountered in the measurement process. For spur and noise measurement errors, the error display is automatically captured by the spectrum analyzer and sent to the controller for storage on a floppy disk. If so desired a plot could automatically be plotted out by the system plotter with a time tag attached along with the serial number of the failed unit.

A switch matrix was also built to allow for the automatic steering of signals to the various pieces of test equipment. This removed the element of error involved with constantly having to reconnect the test set up for the various tests that had to be carried out. In addition, all of the paths were calibrated and error traces were stored within the computer for future downloading to the spectrum analyzer. This feature automatically cancels out any vestigial test set artifacts from the measured data. Also since there are no connectors to be constantly changed, variations due to improper connection are also eliminated. Care had to be taken in designing the switch matrix to minimize the amount of crosstalk between the various paths.

The system interface is needed to convert the IEEE - 488 signals to usable inputs to the system under test. This interface is very user specific in nature and will not be described in detail. Standard interface cards were used wherever possible and custom cards were designed and fabricated as required.

The frequency counter was used for frequency measurement as well as pulse width measurement. The statistical functions which are built into the counter helped immensely in reducing the speed of the measurements since the bus would not have to download each measurement to the controller for further processing. By proper partitioning of the tasks, the measurement process execution time was greatly reduced.

The spectrum analyzer chosen for this task also has various built-in functions which helped greatly in reducing task execution time. Some of the valuable features were direct downloading of display data to a plotter and the ability to subtract one set of data from another internally as well as display the result.

SETTLING TIME TEST FIXTURE

A diagram of the basic test fixture is shown in Figure 2. As can be seen, the unit under test's output signal is mixed with another signal generator tuned to the same frequency. The resultant phase deviation is then characterized. It is important that the synthesizer under test be phase locked to the test's reference generator such that the resolution or sensitivity of the system is optimized. Since the settling time of many commercial generators are measured in hundreds of milliseconds, it is important that the settling time of the test generator not hold up the testing process.

These synthesizers were being tested in an integration area, therefore each unit was previously sold off as being functional before being placed within the system. Consequently, it was

assumed that the long term settling time of the individual synthesizers was within specification. By using a comparable synthesizer as the reference generator, it was further assumed that the short term drift characteristics would track with the unit under test.

The test synthesizer was given approximately twice the maximum settling time to stabilize before the unit under test is strobed. The resultant output of the detector is then divided into two paths. One path is a straight through path while the other is delayed by approximately 50 nanoseconds.

The delayed channel is used as the reference for the floating window comparator and the straight through path as the signal to be compared. The delayed phase detector output is summed with two DC offsets to allow the window to "float". In this way, any fixed phase offset between the two synthesizers can be ignored. The magnitude of the window was derived from the phase specification, since the overall phase variation over time is gradual in nature near the point of settling, for the synthesizers under test.

The strobe signal, generated by the controller triggers the measurement via the system interface and the switch matrix which initiates the frequency load function in the synthesizer under test. The strobe signal also addresses the test fixture, causing a rising edge to be generated at the test fixture output as well as triggering an internal timer. This timer is used to force an end to the measurement if settling is not declared within twice the maximum time specified for settling. This avoids the problem of hanging up the system. The output of the test fixture goes low when settling has been declared. This occurs when the output of the phase detector falls within the dynamic window of the comparator for a predetermined period of time which is then subtracted out by the software.

The output of the fixture is fed to a pulse measurement device which will digitize the pulse width of this signal. This process is repeated by the system inteface until notified otherwise by the controller. The controller is instructed by the counter when the measurement has been completed. The counter is preprogrammed to take 100 samples and then report the results to the controller. The final average value, as well as any requested stastistical data are what is sent across the IEEE - 488 bus to the controller. In this way the measurement process is not bogged down with a lot of bus overhead. Finding a counter capable of performing the pulse width measurement capability along with the statistical functions proved to be extremely valuable in streamlining the time required for this measurement.

After the measurement has been performed for one pair of frequencies, another combination is set up by the controller and the measurement process continues. The controller also controls a switch matrix which steers signals within the system to allow for complete automation of the process. The test results are then sent to the printer with the option being available to print out only failures or each and every value. The maximum settling time as well as the minimum settling time for each synthesizer was also broken out as a separate portion of the printer output.

The rest of the information was stored on disk for future reference.

CONCLUSION

By accomplishing the task of automating the phase settling time measurement process, many calendar months of schedule as well as man-months of effort have been eliminated. The only other alternative would be to test a much smaller subset of frequencies and assume that all of the rest of the frequency combinations will be acceptable.

Even in systems where the integration capacitor is initialized in each instance, this system is still valuable in that accuracy and actual man-month efficiency will greatly increase in production. Since the system is self documenting, complete characterization of the unit under test can be accomplished.

Automatic spur and noise characterization have also been measured within the overall measurement test setup of which the settling time test fixture is a subset.

SETTLING TIME MEASUREMENT SETUP

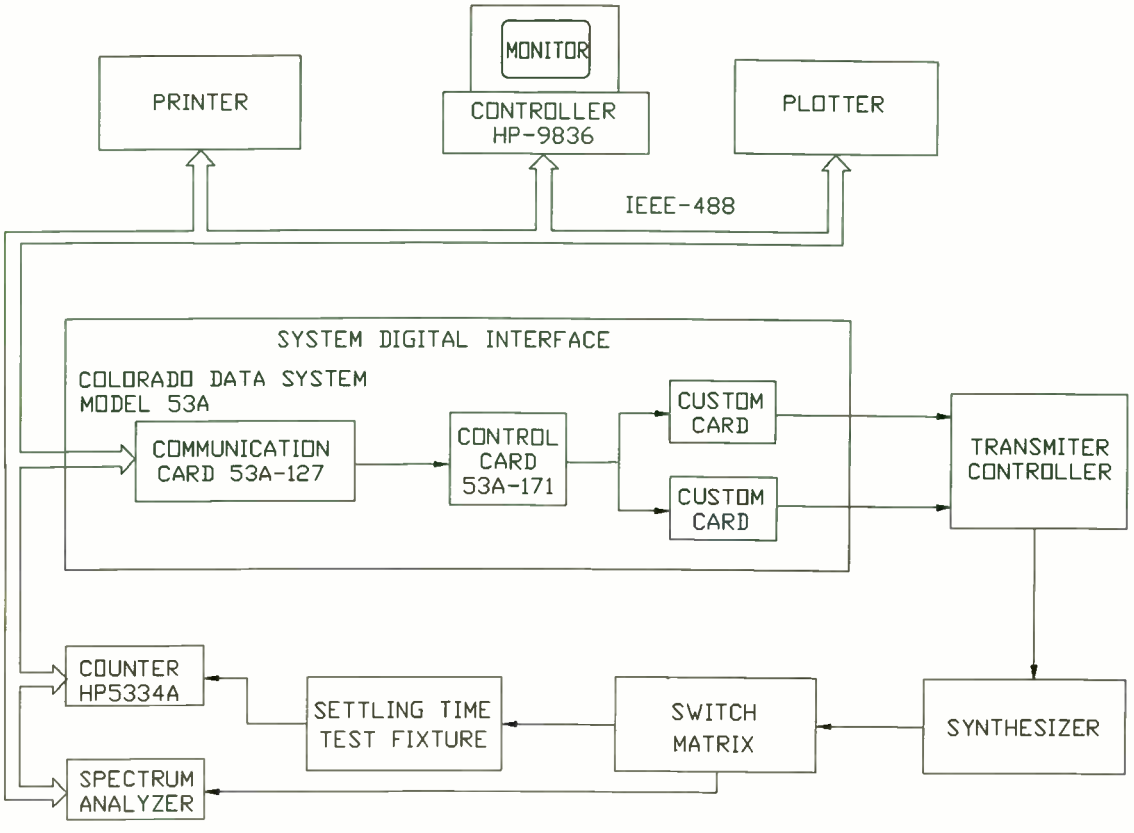


Figure 1

SETTLING TIME TEST SET BLOCK

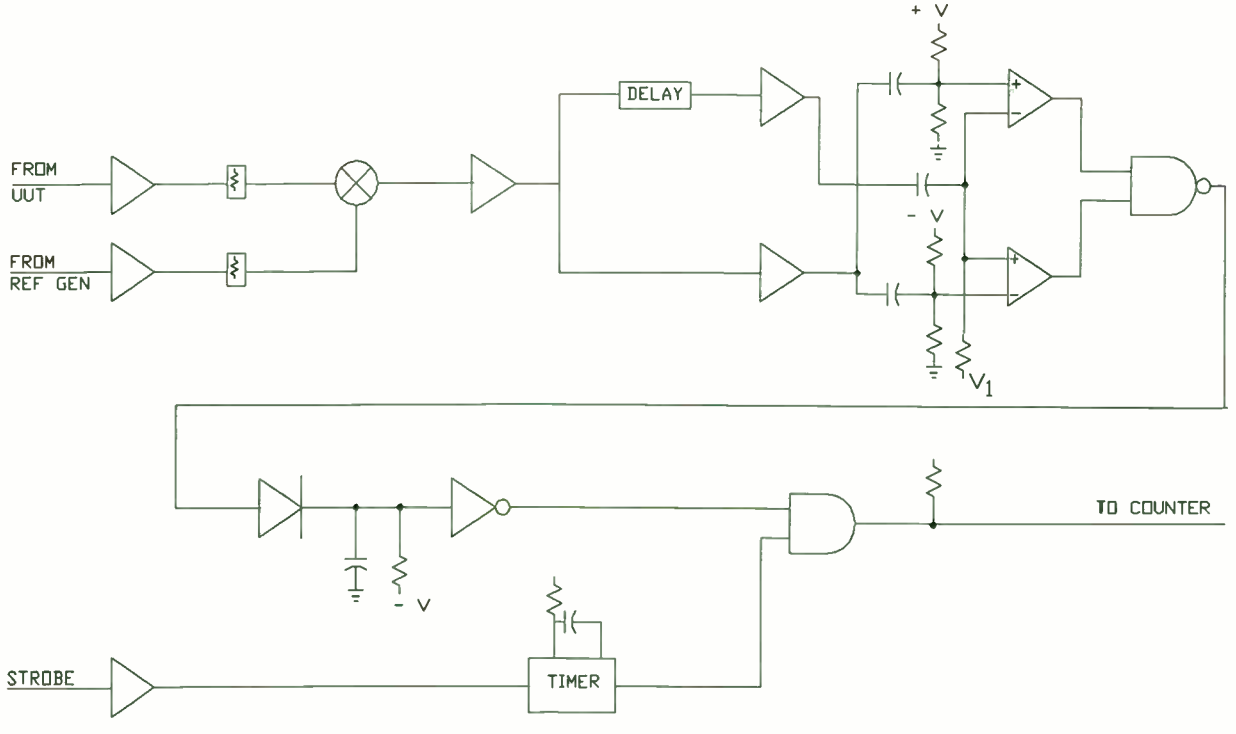


FIGURE 2

Digital RF Synthesis: Theory and Application of a Booming Technology

by

Greg Lowitz

Hewlett-Packard Company

1501 Page Mill Road

Palo Alto, California 94304

Abstract: *Tough demands placed on today's electronic systems require flexible signal generators capable of highly-complex signal simulation. The marriage of emitter-coupled logic (ECL) and GaAs DAC technology gives birth to a new generation of sophisticated off-the-shelf digital RF synthesizers especially suited for applications in radar/EW, communications, disc testing, and others. An introduction to digital-synthesis theory coupled with a survey of real-world applications gives the newcomer an exciting look at current trends and opportunities in high-performance system testing.*

Introduction

The demand is growing for accurate and complex-waveform generators that overcome the shortfalls of conventional function and RF signal generators. For test applications requiring complex *real-world* signals, basic sine and square waves are no longer sufficient to thoroughly calibrate and characterize a system's performance. Furthermore, as designers race to lead the competition, new testing techniques become necessary to fully stress product susceptibility and reaction to imperfect environments. All too often products are developed and tested under ideal conditions, resulting unwittingly in designs that are less robust than was otherwise possible. In extreme cases, a design defect might be found where it hurts most: in the customer's hand.

To circumvent the lack of commercially-available solutions, design labs and manufacturers often settled for inferior test-setups or resorted to expensive, home-brew signal generators with application-specific architectures. A typical production test station, for example, might house several specialized signal sources, each one different from the other. These custom-built "black boxes" imply many hidden costs including design, assembly, documentation, training, and maintenance -- operations that detract from a manufacturer's primary objective: to sell products.

As engineers and managers seek more efficient and cost-effective testing techniques, a new generation of flexible, reconfigurable test equipment is working its way onto designers' benchtops: *Digital Synthesis*. Digital synthesis combines synergistically the best attributes of analog and digital technology into one powerful combination. High-speed multiplexed memories, digital accumulators, fast multipliers and RF DACs make possible this growing revolution in testing technology. Compared to other waveform-synthesis techniques, digital synthesis offers the following benefits:

Benefits of Digital Synthesis

1- Arbitrary waveform creation in both time or frequency domain

2- Stable and repeatable signals with digital precision

3- Software reconfigurability

4- Phase-continuous frequency switching

5- Amplitude and frequency agility

6- Low noise and distortion

7- Wide modulation bandwidth

8- Potential for size reduction through integration

9- Reasonable cost

At last count, roughly half-a-dozen manufacturers were advertising some form of digital synthesizer, spanning a wide range of operating frequencies, modulation capability, harmonic

performance, and user-interface features. Most of these products have appeared only within the last two years, leading one to take stock of the growing interest in such technology. One commercially-available example from Hewlett-Packard features a 12-bit, 125 MHz bipolar DAC, and GaAs differential sampler. Typical two-tone intermodulation distortion at 10 and 45 MHz is better than -65 dBc, demonstrating the high spectral purity achievable even at RF.

Specifying the Right Architecture

While each digital synthesizer has its own characteristic features, most digital synthesizers employ one of two principle architectures:

1- Waveform look-up (memory based)

2- Direct digital synthesis (accumulator based)

Figure 1 displays the simplified architecture of the HP 8770A Arbitrary Waveform Synthesizer -- a memory-based design. 512K of fast, multiplexed dynamic RAM outputs 12-bit waveform data samples to the DAC every eight nano-seconds. A GaAs half-order-hold sampler freezes the DAC output after switching transients have decayed. This essential element significantly reduces the effects of sinc(x) rolloff and helps to ensure a spurious-free output to maintain high signal integrity. A phase-equalized elliptical brickwall-filter transforms the sampled data into a continuous time-domain RF analog signal. Finally, a low distortion output amplifier provides enough current to drive a 50-ohm load.

A household analogy to waveform look-up is the compact disc player whose music (waveforms) is stored in disc form (memory) and replayed through an audio-quality digital-to-analog converter. Although the bandwidth of a compact disc player too low to be useful for RF applications, the basic concept is the same. As a point of interest, several key parallels between the HP 8770A and a compact disc player are highlighted in the table 1 on the following page.

Comparison to the Compact Disc

Feature	HP 8770A	Compact Disc Player
Bandwidth	50 MHz	20 kHz
Sampling Rate	125 MHz	44.1 kHz
Number of Bits	12, linear	16, linear
Dynamic Range	72 dB	90 dB
Two-tone Intermod	< -65 dBc @ 10 MHz	< -90 dBc @ 1 kHz
Anti-aliasing Filter	brickwall elliptical analog	analog and digital filter
Number of Channels	1	2 (L+R)
Waveform Storage	512K words fast memory	> 1 Gigabyte optical disc
Recordable/Erasable	Yes	No
Cost/channel/kHz BW	$0.52	$10.00

Table 1

Conversely, direct digital synthesis uses a digital phase-accumulator with a sine look-up table to generate waveforms that are carrier based. The phase accumulator is a simple digital integrator. A constant value *k* at the accumulator input causes the output to count up with a slope of *k*. For example, to generate a ramp of slope three, simply input a binary three to the accumulator. Because the accumulator output is added to its input during each new clock cycle, the output continues to increment according to the pattern 0,3,6,9,...and so forth.

Feeding this ramp into a sine-wave look-up table (phase-to-amplitude converter) results in an output sinewave whose frequency is directly proportional to the slope of the accumulator output. The absolute output frequency is related by the equation: $Freq = k*Fc/(2^N)$, where *k* is the accumulator input, Fc is the system clock frequency, and N is the number of bits in the

accumulator. Because the accumulator represents a phase index, no signal glitching occurs when the accumulator overflows. Rather, the accumulator continues to ramp up at the same rate assuming the input remains fixed.

With some added hardware, direct digital synthesizers can be designed to provide independent, arbitrary modulation of frequency, phase, and amplitude. This makes generation of modulated carriers straightforward, such as coherent radar pulses, and complex shift-keyed signals like QAM and FSK. This is easy to visualize since the input to the phase accumulator need not remain fixed as previously described. A changing accumulator input modulates the carrier frequency. This simple elegance obviates the need to vary the clock frequency to produce carriers of different frequencies, allowing a fixed output filter for waveform reconstruction.

Similarly, a simple post-accumulator adder is all that's necessary to add phase modulation capability. Amplitude modulation, however, is not as easy since ultra-high-speed multipliers are not readily available. However, with the military push for VHSIC (Very High Speed Integrated Circuits) for use in fast radar/EW signal processors, the state-of-the-art is advancing rapidly. For current applications, a variety of techniques can be used to effectuate multiplication at very high data rates, but discussion of these is beyond the scope of this paper. For comparison to the architecture of figure 1, figure 2 illustrates a simplified version of a direct digital synthesizer with provisions for FM, PM, and AM. The HP 8904A is a low-frequency implementation of direct digital-synthesis.

Ultimately, the application at hand will dictate which architecture provides the greatest utility. In many cases, either architecture will solve a particular problem equally well, but developing waveforms may be easier in one case over another. Table 2 summarizes the relative advantages of the two approaches:

Comparison of Two Digital-Synthesis Techniques

Important Parameters	Waveform Look-up (memory based)	Direct Digital Synthesis (accumulator based)
Frequency Resolution	Moderate	Excellent
Frequency Hopping	Limited	Excellent
Multiple Tones	Excellent	Limited
Modulation Flexibility	Moderate	Excellent
Real-time Control	Limited	Moderate
Bandwidth	High	High
Complexity	Moderate	High

Table 2

As the preceding table illustrates, direct digital synthesis compares favorably in many parametric areas. However, its higher modulation flexibility results in greater circuit complexity and accordingly higher price. Furthermore, for applications that require multiple tones, jamming signals, clutter, and highly-unusual time-domain waveshapes such as commonly found in magnetic recording applications, memory-based systems excel by comparison. Furthermore, for baseband applications where carrier synthesis is unnecessary, memory-based systems provide the best flexibility -- when driving PIN diode modulators or the I and Q inputs of a vector modulator, for example.

Although memory-based synthesizers can simulate carriers with modulation, memory use is inefficient, placing practical limitations on waveform variety. Just which technique is optimum for a given situation depends critically on the desired waveform. In either case, advanced memory sequencing can make efficient use of memory by looping and repeating certain sections of a waveform. For example, consider a pulsed waveform with a 1 us pulse width and a 10 ms off time (figure 3). Rather than storing the entire waveform in memory (which would require over 1.2 Megawords to recreate assuming an 8 ns clock period), a short segment of pulse off-time could be

looped as many times as desired. In total, the entire waveform could be recreated using only a fraction of the available waveform memory, making possible long chains of complex signals.

Technical Considerations of Digital Synthesizers

Regardless of the chosen architecture, both techniques can use the same output stage including a DAC, sampler, anti-aliasing filter, and amplifier. This output stage ultimately determines critical specifications such as distortion, spurious outputs, dynamic range, amplitude flatness, phase linearity, and output power. With clock frequencies in the hundred-megahertz region, retaining 12-bit performance is a non-trivial task. Interactions and reflections between the DAC and sampler are significant. Furthermore, careful attention to grounding and shielding are essential in minimizing ground-current modulation and sampling-clock feedthrough.

Open literature reports new developments in GaAs DAC technology pushing clock rates into the 1 GHz region. The difficulty in supplying digital data at these rates is compounded by the difficulty in designing low-distortion wideband DACs and amplifiers that sport sub-nanosecond rise times. However, as the demand for ultra-wideband modulation grows, industry requirements will drive new developments in test equipment.

In addition to the DAC/sampler combination, the choice of output anti-aliasing filter directly affects the usable bandwidth and spectral purity. As shown in figure 4, a sampled system produces a repeating spectrum spaced in integral multiples of the clock frequency. By filtering the spectrum with a theoretical brickwall cutoff at one-half the clock frequency (Nyquist frequency), it is possible to recover only the essential signal information, while throwing away out-of-band spectra.

As a practical matter, however, the ideal brickwall filter does not exist. In this case, a certain amount of leakage can fold over into the passband region (aliasing), adding error to the signal spectrum. How much leakage occurs depends both on the signal bandwidth and the filter transfer function. For a high-performance 12-bit digital synthesizer, it is desirable to have greater than 60 dB of rejection at the Nyquist frequency. Therefore, to allow for some finite filter rolloff from passband to stopband, available bandwidth can be as high as 0.4 * $Fclock$. As a result, the overall system performance is directly proportional to the quality of the output filter.

User Interface Requirements

Unlike products with well-bounded markets and functionality, digital synthesizers are so general purpose that it becomes difficult to define a set of front panel buttons that don't artificially limit the potential applications. Even within a given market, every designer has unique testing requirements, making it even more difficult to develop a universal interface.

As a result, digital synthesizers beg for a software-reconfigurable user-interface that is both general purpose and customizable. Hewlett-Packard's answer to this dilemma is the HP 11776A Waveform Generation Software -- which, when teamed with an HP series 200/300 computer, enables fine control over all waveform parameters, whether described mathematically, graphically, or in the time or frequency domain.

Using a computer or graphics tablet for waveform data entry, a complete set of math and waveform manipulation commands enable a user to design very complex waveforms with minimal training. For the first time, sophisticated tools for complex waveform development are at the fingertips of every engineer. For example, the single commands $TOFREQ$ and $TOTIME$ are all that's necessary to transform a signal between the time and frequency domain (Figure 5). This makes it easy to add signal imperfections, noise, multipath, and transients custom-tailored to a given application. In addition, the software has provisions that enable users or third-party vendors to develop software shells that run application-specific programs or macros.

In some cases, rather than designing waveforms from scratch, it is desirable to capture a real-life waveform using a waveform digitizer. The user can subsequently replay the waveform through the digital synthesizer or use the waveform generation software to modify the waveform in almost

any way imaginable. Figure 6 depicts the complex spectrum of a real-life radar pulse that has been digitized and replayed through the HP 8770A.

For high-performance system testing, digital synthesis offers unmatched flexibility and precision control of all waveform parameters. With these tools, designers and production engineers can increase engineering productivity while speeding and improving testing quality.

Real Customers, Real Applications

Now that the concept of digital synthesis is more clear, it is stimulating to look at more examples based on real customer requirements. Three examples from radar, EW, and disc drive testing exemplify just a few of the potential uses for digital synthesizers.

Radar/EW

Consider the simplified block diagram of a typical microwave receiver (figure 7). The designer is interested in a variety of characteristics that ultimately define the accuracy and functionality of the receiver. The radar engineer might be interested in simulating a complex chirp or phase-coded pulse to test the signal processor's pulse-compression circuitry. Or, perhaps, the designer wants to simulate the return of one or more test targets with varying amounts of noise to test the CFAR (constant false-alarm rate) -- a test of a receiver's ability to adapt to changing noise levels and jamming signals. Figure 8 illustrates the effects of pulse compression for distinguishing closely-spaced radar target returns. From the first signal it might appear that there are only three targets. However, after compression, six targets with different radar cross-sections clearly emerge -- an important distinction, say, for an air-traffic controller! Finally, by upconverting the signal to uW frequencies, it is possible to test the entire receiver chain from the antenna to the digital signal processor.

On the other hand, the EW designer is interested not only in detecting emitting radars, but in active ECM (electronic countermeasures) such as spot jamming, deceptive jamming or communications jamming. With memory-based digital synthesis it is a simple matter to define waveforms with well over several-thousand *simultaneous* tones with precisely defined frequency and phase relationships. Figure 9 shows a view of 1000 simultaneous tones spaced over a 30 MHz bandwidth. A close-up of this spectrum in figure 10 reveals that certain tones are intentionally missing. This would enable a user to selectively jam certain channels, while leaving others free for transmission.

Disc Testing (Magnetic Recording)

The magnetic recording industry is highly competitive, forcing manufacturers to look for ways to trim operational costs, speed media certification and provide higher quality overall. Figure 11 illustrates a typical read channel functional block found in many of today's disc drives. The HP 8770A faithfully reproduces complex signals necessary to test disc read/write heads and servo loops. Furthermore, waveforms can be manipulated to exercise the margins of operation and susceptibility to missing bits, extra-bits, bit-shifts, and surface contamination.

Figure 12 shows a unique waveform that tests the recovery of a disc drive's AGC (automatic gain control) circuitry. The rapid gain and frequency change at midscale accompanies a changing bit pattern. Similarly, figure 13 shows an actual captured monopulse signal from a thin-film head found in a commercial disc drive.

Low distortion, low spurious outputs, and wide bandwidth are crucial qualifications for validating today's sensitive high-speed drives and dense disc media. Any unwanted signal perturbation could result in erroneous test results, making device failure rate look higher than it actually is.

Other Applications

Table 3 highlights by application a variety of other uses of digital synthesis. This table is not exhaustive, but is intended to stimulate clever ideas for other applications.

Unique Digital-Synthesis Applications

Radar/EW	Simulate IF and video signals in radar/EW receivers. Generate chirps, LFM, NLFM, and Barker codes. Construct a radar main bang with pseudo-random PRI jitter. Test receiver sensitivity to noise, clutter and Doppler shift. Simulate complex target returns with multipath. Generate jamming signals, noise and interference.
Comms	Generate JTIDS, BPSK, QPSK, QAM, PCM, T1, AM and FM signals. Modulate frequency, phase and amplitude simultaneously. Generate reference standard modulation indeces.
Disc Test	Calibrate disc certifiers and sensitive magnetic recording devices. Test for missing bit, extra-bit, and servo response.
Component Test	Test amplifier intermodulation distortion with two-tone signals. Characterize ADC linearity. Check AGC and servo loop dynamics.
Filtering	Define Nyquist filters for digital communications Convolve signals with windows such as Hamming, Hanning and Kaiser.
Noise	Generate Gaussian noise with specified mean and standard deviation. Specify signals with precise signal-to-noise ratios.

Table 3

Conclusion

Digital synthesis technology is booming and is changing the way engineers test products. High-speed digital logic carefully merged with wideband DAC and RF technology can provide innovative, flexible and extremely precise digital synthesizers at a reasonable cost. Complex test signal requirements are opening the door for a new generation of off-the-shelf test equipment capable of providing real-world waveforms to speed production testing and improve product quality. A comparison of two digital synthesis techniques coupled with several actual customer applications gives the newcomer to digital synthesis an exciting glimpse into current and future trends in testing technology.

Useful References

1- R. Hassun, A. Kovalick and W. Sagun, "The Theory of RF/Analog Waveform Synthesis by Digital Means," RF & Microwave Symposium, April 1987.

2- G. Lowitz, "HP 8770S Signal Simulator System," Hewlett-Packard data sheet #5954-8890, July 1987.

3- R. Armitano and G. Lowitz, "Use Predistortion to Improve Amplitude and Phase Accuracy of Digital Synthesizers," Hewlett-Packard white paper, November 1987.

4- A. Kovalick, "HP 8770A Applications in Magnetic Recording," Hewlett-Packard white paper, November 1987.

5- R. Hassun, D. Kreitter and J. Minck, "Benchtop Signal Simulation Becomes a Reality," Defense Electronics, November 1987 PP. 129-143.

6- P. Thysell, "Effective Use of the HP 8770S Signal Simulator System," Hewlett-Packard product note (PN 8770S-2), 1987.

7- K. Kafadar, "Digital Signal-Synthesis Tools Model Real-World Environments," EDN, November 12, 1987 PP. 239-248.

8- G. Lowitz, "Generate Precise Modulation and Complex Spectra with the HP 8770A," Hewlett-Packard white paper, August 1987.

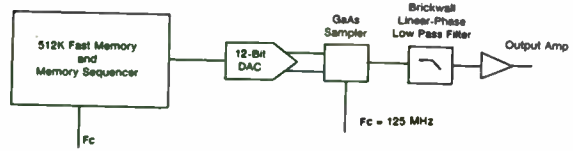


Figure 1 Waveform look-up digital synthesizer (memory based).

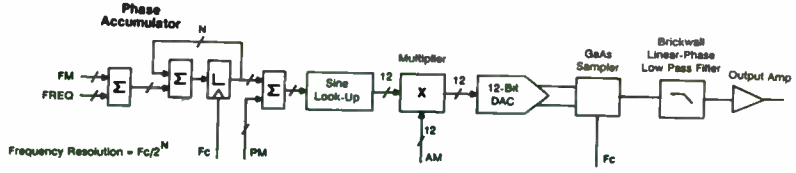


Figure 2 Direct digital synthesizer (accumulator based).

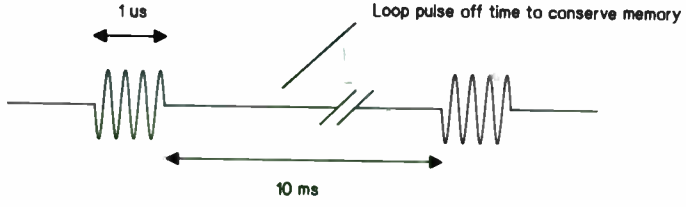
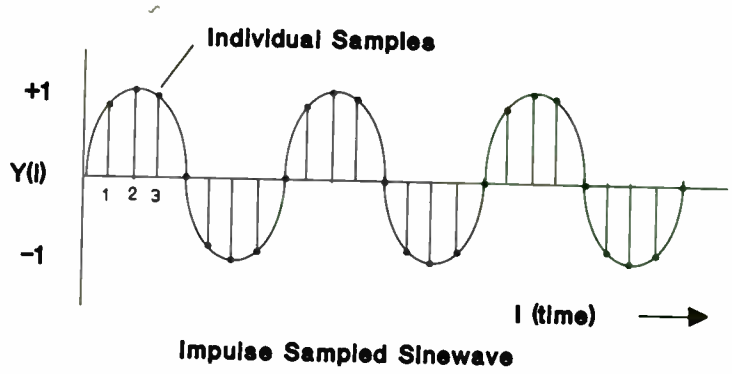


Figure 3 Sequencing ensures efficient use of memory.



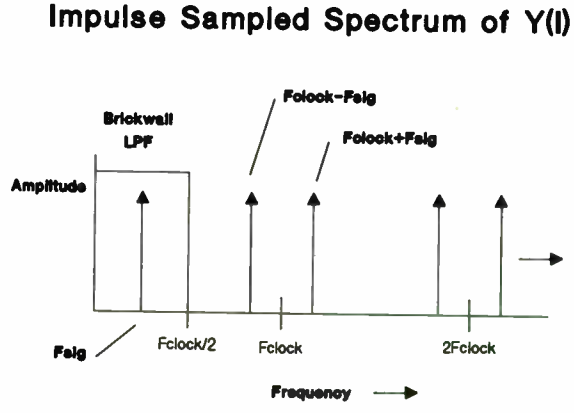
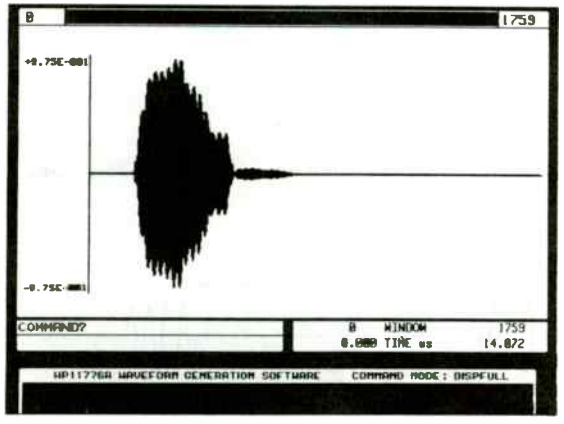


Figure 4 Basic theory of sampled systems.



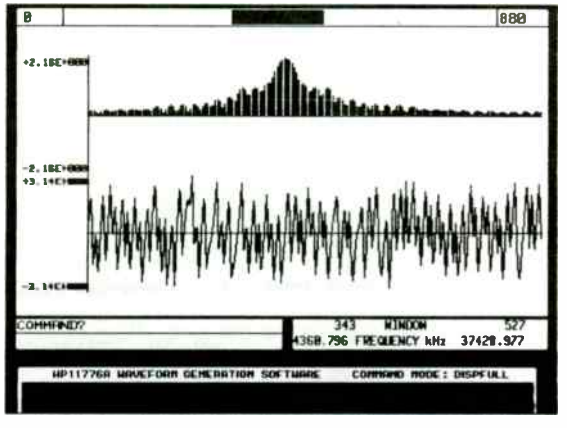
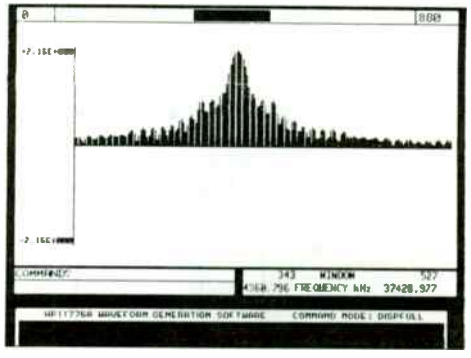


Figure 5 Advanced development tools permit waveform creation in the time or frequency domain. Here a complex pulse is analyzed for its magnitude and phase content using the HP11776A Waveform Generation Software.



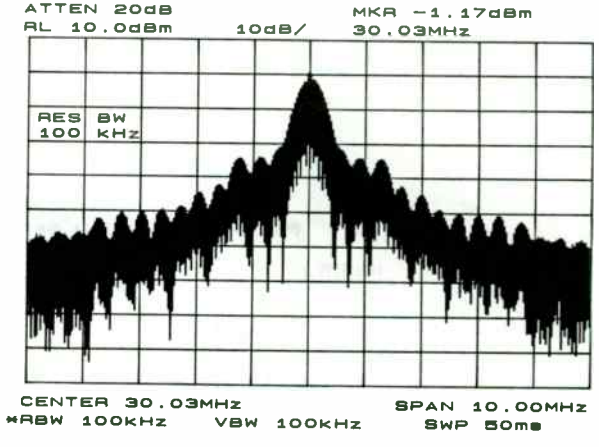


Figure 6 Spectrum analyzer display of an actual radar pulse that has been digitized, modified (figure 5), and downloaded to the HP 8770A Arbitrary Waveform Synthesizer.

Advanced Receiver Test for Radar/EW/Communications

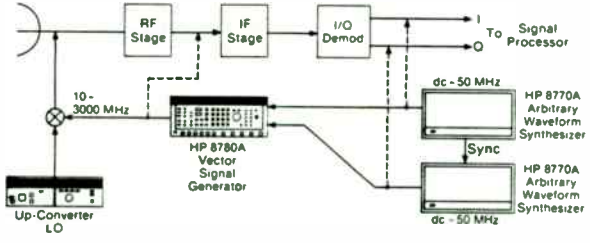
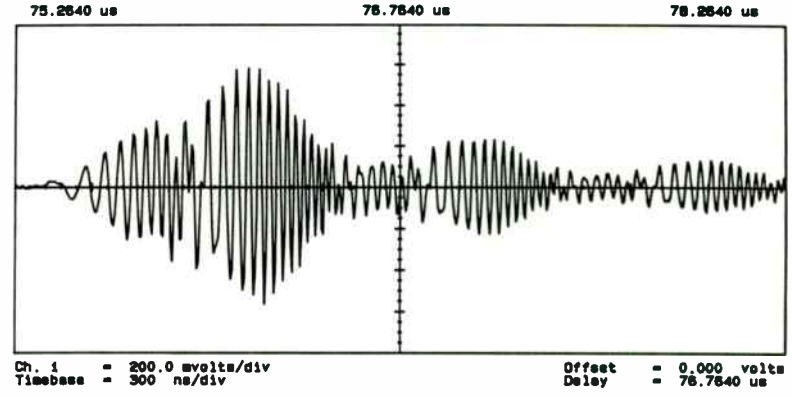


Figure 7 Improve receiver testing with realistic digital-synthesis simulation techniques. Here two HP 8770A synthesizers are used to modulate the I and Q inputs of the HP 8780A Vector Signal Generator, enabling very wideband, complex signal generation from 10 MHz to 3000 MHz. Adding an external local oscillator extends complex signal capability to higher microwave or millimeter-wave bands.



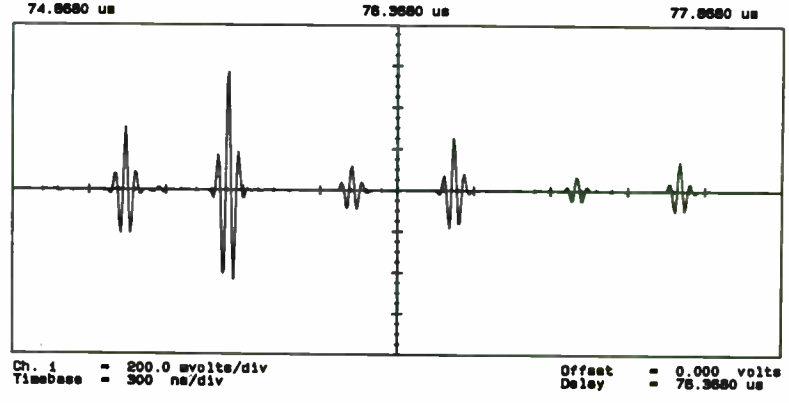


Figure 8 The first figure is a complex target return from a radar with a 122 kHz PRI, 50 MHz linear chirp and a 1 us pulse width. The second figure is the same return after compression. Several closely spaced targets are now clearly visible. Both waveforms were simulated using digital synthesis techniques.

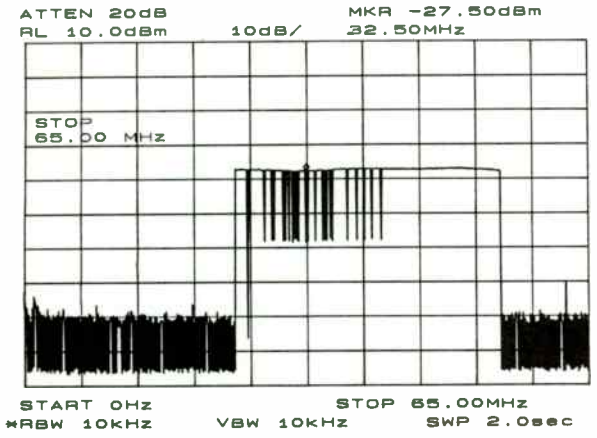


Figure 9 A 1000-tone frequency comb useful for jamming applications.

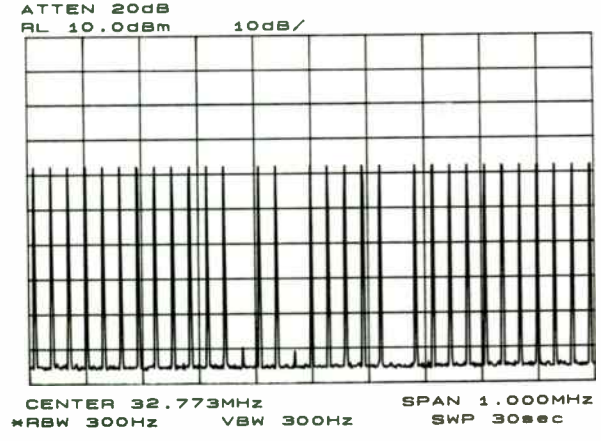


Figure 10 Close-up of the frequency comb of figure 9 has some channels intentionally void of interference, permitting free communication on selected channels.

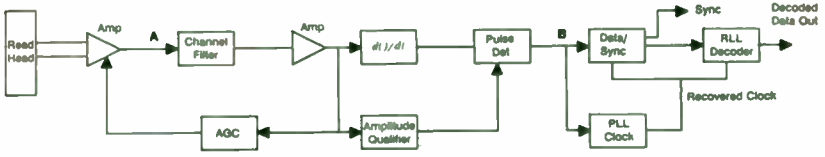


Figure 11 Block diagram of a typical disc-drive read channel.

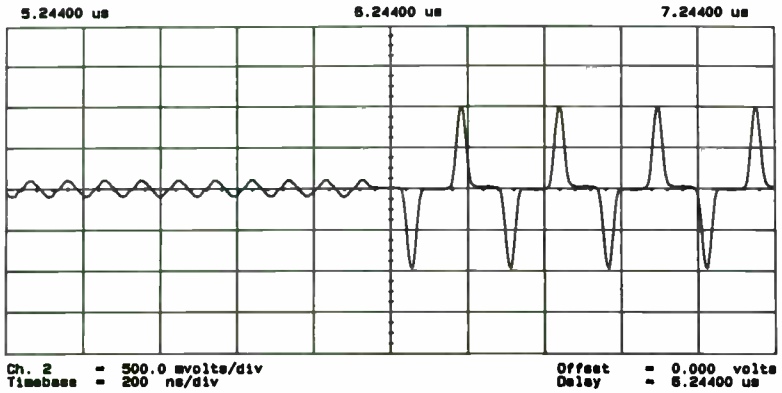


Figure 12 Complex waveform tests disc drive AGC (Automatic Gain Control) recovery time.

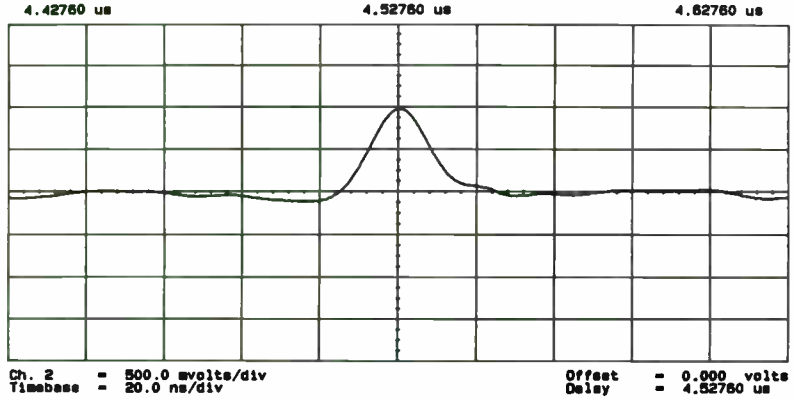


Figure 13 Regenerated monopulse from a thin-film disc-drive read head.

A SHORT DESCRIPTION OF AN ATE SOFTWARE SYSTEM FOR RF MODULE TESTING

BY

J.C. Lunsford, J.T. Bowles, B.J. Tarnowski and L.M. Tichauer
Hughes Aircraft Company
P.O. Box 3310
Fullerton, Ca. 92634

OVERVIEW:

The primary intent of this system design was to reduce the technician/engineer time overhead associated with rf performance testing of a spread-spectrum communications system. Given the use of four high-speed synthesizers switching across 51 frequencies, the job of running a complete test on one system was projected to require two technician man weeks and one engineer man week. Obviously, the occurrence of faults requiring retesting would increase this time dramatically. Not so obviously, test reliability would begin to suffer fairly quickly due to test operator fatigue and data transcription errors.

The answer to this type of problem is test automation. An Automated Test Equipment (ATE) system was designed to perform the repetitive test tasks and flag out of tolerance conditions for operator action. Additionally, such a system can automatically log to paper or magnetic media (or both) all of its readings for performance documentation and fault histories. As it evolved, it became apparent that the system could come fairly close to operating in a completely hands-off mode. With some operator interface in the test fixture calibration phase of the test, the actual performance testing could run unattended. In its finished version, the system allowed the testing of a set of rf modules in approximately 36 continuous hours of operation only one of which required attendance by a relatively unskilled operator.

BASIC SYSTEM OPERATION:

The operation of this ATE system is simple and straight forward. All system functions are menu driven and, where operator action and/or data entry are required, directly prompted. These prompts are displayed on the computer monitor complete with connector designations and useful allowable data range information. An important aspect of this system is that all operator inputs are qualified by the system as to their basic reasonability before they are acted upon.

Each test and evaluation function is available as a separate soft key activated command or as part of a linked test sequence performed as a total system performance test. This allows direct testing of a known problem to be done with the same software that, during general performance testing, found the problem in the first place.

INITIAL OPERATOR ACTION:

At the start of a general system test, the operator is prompted for a categorical qualification of the rf module group to be tested (in this case any one of three different systems). He is further prompted for the serial number of the unit under test (UUT), and whether a printed test report is desired or not. If a report is selected, all test data are printed in the report. If a report is not selected only test

failure data is printed.

At this point the operator has the option of calibrating the test system for the rf losses in the cabling and switch matrix. As long as there have not been any physical changes in the cable connections or in the switch matrix itself, there should be no reason to check the calibration. If calibration is desired, the operator is prompted through the process of sending a known swept frequency signal through each possible rf signal path. The losses in that path are read as a frequency response curve from the spectrum analyzer and stored on one of the computer's disk drives for offset reference during actual testing. This test fixture compensation feature is completely transparent to the operator.

FULLY AUTOMATED TESTING:

From this point on, the test system runs without operator intervention. First the output of a built in test function is monitored to test for detected error conditions. Next a minimum discernible signal test is run by attenuating a stream of test data until the bit error rate of that data exceeds test specifications.

FREQUENCY AND POWER TESTS:

Thus far there have not been any dramatic savings of either test or operator time. Next, however, each frequency of each synthesizer (four synthesizers, fifty-one frequencies each) is tested against its specifications for both frequency accuracy and power output. This entire process including a full report listing each synthesizer/frequency combination, its actual frequency and power output, and whether or not both of those values are within specifications takes approximately 20 minutes.

This is possible because the synthesizers are digitally controlled and the frequency counter and spectrum analyzer used to measure the signals can be accessed over the IEEE-488 buss by the computer to report their findings. The entire test process is handled and reported by the test system without any need of connection changes, instrument reading, or data transcription by the operator. This is where the ATE concept starts paying for itself.

SETTLING TIME TESTS:

The next test is critical to the spread spectrum concept of data communications. Each synthesizer must be capable of settling on its assigned frequency in microseconds. The rf system design requires the measurement of synthesizer settling time for 10,000 different synthesizer/frequency combinations! This is slow tedious work under the best of circumstances and is generally consigned to spot checks to avoid operator fatigue and the resultant meaningless test results. With the ATE system these measurements are made and reported, as in the frequency and power testing, without operator intervention. Total test time is less than thirty hours.

SPUR AND NOISE TESTS:

Finally, each synthesizer/frequency combination must be

tested against acceptable spur and noise profiles. Up until now, the major function of the ATE system was to set synthesizer frequencies and directly read test instruments. In this instance, the nature of the measurement to be made is much more complex and requires another order of magnitude of data analysis to be performed. Here is where the power of microprocessor controlled instrumentation truly comes into play.

The spectrum analyzer is capable of performing data analysis by virtue of having its own control computer and digital trace memories. The ATE system control computer has a set of control programs for the spectrum analyzer stored on its own disk drive. When the command for spur and/or noise testing is encountered, the system controller selects the appropriate spectrum analyzer control program and down loads it to the analyzer. Additionally, mask traces which describe the performance envelope that the synthesizer output must meet are down loaded at the same time.

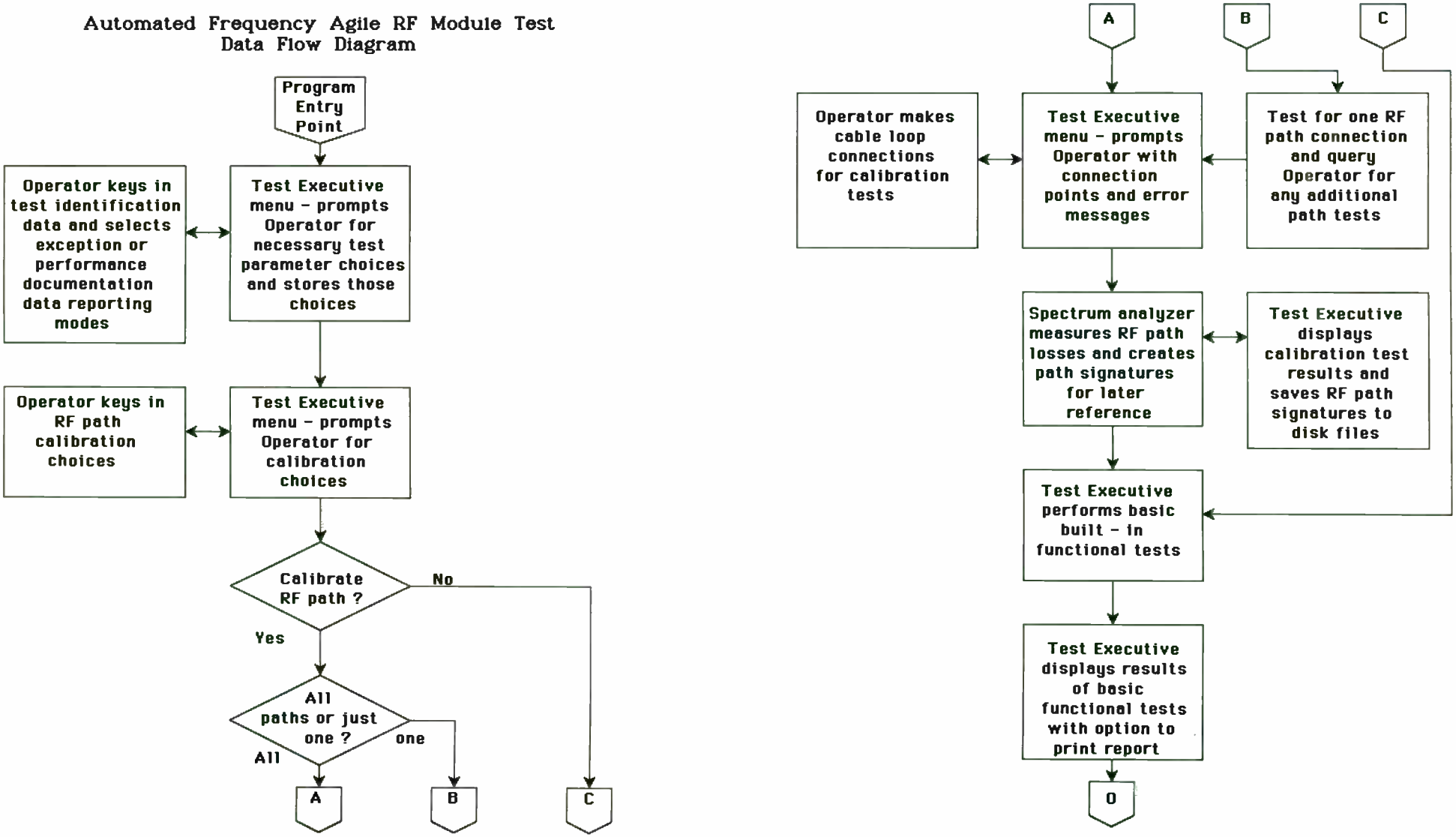
The spectrum analyzer now examines the output spectrum of each synthesizer/frequency combination that is presented to it by the ATE system. It directly tests each signal for excessive spurs and noise and sends the trace display with a date and time stamp back to the ATE system control computer for each test failure. The control computer stores and indexes these error traces on its disk drive for later review by the test operator. A complete test of all synthesizer/frequency combinations takes about five hours, assuming that no faults are found.

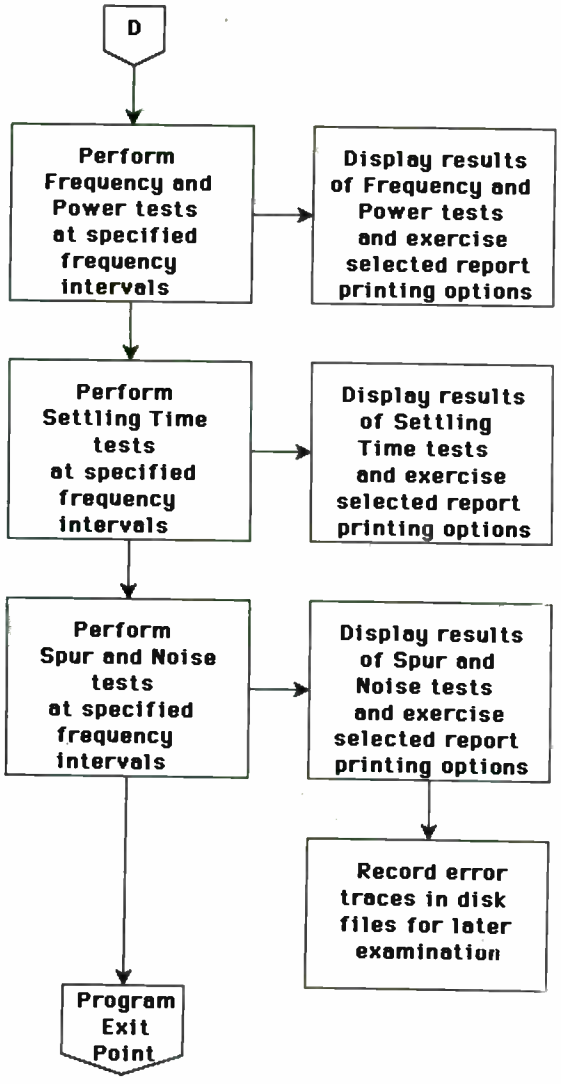
These saved spectrum analyzer traces are now available for off line analysis by the test operator (or the design engineer) in order to isolate and correct UUT faults. The traces, both the original signal and the mask against which it was measured, may be displayed on the computer monitor, printed on a plotter, or restored to the spectrum analyzer for detailed evaluation and permanent records of performance testing.

ATE ADVANTAGES:

The primary advantages of ATE systems in this type of application are:

- No erroneous test results occur from operator fatigue and transcription mistakes.
- Major savings in man hours and calendar test time are realized.
- All possible operating parameter variations can be tested to assure greater reliability.
- Creation and maintenance of complete performance test records becomes a realizable goal.





INDUSTRY MEASUREMENT OF NOISE PARAMETERS

by

William Mueller
Sr. Applications Engineer
Avantek Inc, Microwave Semiconductor Division
3175 Bowers Ave Santa Clara Ca 95054-3292

This paper discusses a practical method of generating noise parameters from tuned device measurements. Background theory and terminology are presented first. This is followed by a discussion of influences the measurement setup has on the measured noise performance. The methodology of noise parameter generation is given next. The paper closes with a brief discussion of technique limitations, and some alternate methods of noise parameter generation.

Theory and Terminology

Figure 1 depicts a generalized noise matching scheme and the associated terminology. The device being characterized is shown as a two port, with a 50Ω set of s parameters s_{11}, s_{21}, s_{12}, and s_{22}. A *reflection coefficient* Γ_G is presented to the device by the input matching circuit, corresponding to a *generator impedance* $Z_G = R_G + j X_G$, or to a *generator admittance* $Y_G = G_G + j B_G$. Similarly, the reflection coefficient presented to the output of the device is Γ_L, representing a *load impedance* Z_L or a *load admittance* Y_L. When the device being characterized is embedded in these matching networks, the reflection coefficient seen looking into the device input port becomes

$$s_{11}' = s_{11} + \frac{s_{12} s_{21} \Gamma_L}{1 - s_{22}\Gamma_L}, \qquad (1)$$

and the reflection coefficient seen looking into the output port becomes

$$s_{22}' = s_{22} + \frac{s_{12}s_{21}\Gamma_G}{1 - s_{11}\Gamma_G}. \qquad (2)$$

The *device noise figure* is given by the ratio of the signal-to-noise ratio at the input of the device to the signal-to-noise ratio at the output of the device. It is designated by the symbol F, and can be expressed as a linear quantity or in dB, where

$$F_{dB} = 10 \log_{10} (F). \qquad (3)$$

Noise performance can also be expressed as an equivalent noise temperature in Kelvin. The noise figure exhibited by a device is established by the input match presented to the device; the lowest noise figure a device can achieve is designated F_{min} (also frequently designated F_{opt}). The reflection coefficient presented to the device to realize this best noise performance is designated Γ_{opt}.

The gain exhibited by a device with input matched to Γ_G and output matched to $s_{22}'^*$ (i.e. conjugately matched) is the *available power gain*, G_A. It is a function of device s parameters and Γ_G, and

is given by

$$G_A = \frac{|s_{21}|^2 (1-|\Gamma_G|^2)}{|1-s_{11}\Gamma_G|^2 (1-|s_{22}'|^2)} \qquad (4)$$

When $\Gamma_G = \Gamma_{opt}$, this gain is also referred to as the *associated gain* of the device.[1]

It is possible to define a parameter called the *equivalent noise resistance*, designated R_N, that indicates the rate at which the noise performance degrades as the input match moves away from the optimum input reflection coefficient Γ_{opt}. This parameter has the units of ohms, though it is often normalized (divided by the system characteristic impedance, usually 50 Ω) and expressed as a ratio. The noise performance of a device can be calculated from

$$F = F_{min} + \frac{R_N}{G_G} |Y_G - Y_{opt}|^2. \qquad (5)$$

The parameters F_{min}, Γ_{opt}, and R_N are called the *noise parameters* of a device, and provide a complete small signal noise description of the device in the same manner as s parameters provide a complete small signal gain description.[2]

Measurement Setup

The measurement setup can be divided into three sections: the noise source, the device under test, and the noise receiver.

The *noise source* provides a known amount of noise at two distinct temperatures, a requirement of modern noise figure meters. In a solid state noise source, the two noise temperatures are established by alternately applying bias to and removing bias from a noisy semiconductor, usually a diode biased into avalanche. The amount of noise generated by a noise source is called the *excess noise ratio* or ENR; ENR is measured in dB and will, of course, vary with frequency.

Since noise performance is a function of input match, it is important that the impedance changes of the noise diode between on and off states not effect the Γ_G presented to the device being measured. Thus most noise measurement systems will isolate the active device from noise source impedance changes by placing either a 50 Ω attenuator or a ferrite isolator between the active device and the noise source. The new generation of 5 dB ENR noise sources are 15 dB ENR noise sources with an additional 10 dB of attenuation incorporated internally to provide this kind of isolation. Any isolating component added after the noise diode will decrease the amount of noise reaching the receiver; the ENR of the isolated noise source must therefore be adjusted by the loss of the isolating element when calibrating the noise receiver.

The "*device under test*" consists of the fixtured semiconductor,

tuners for establishing the noise and gain matches, bias "Tees" for applying dc bias to the device under test, and any adapters necessary to interconnect these items. The noise receiver will measure the noise figure of the entire system of bias Tees + tuners + fixture + semiconductor. Any losses occurring between the semiconductor and the noise source add directly to the noise figure of the semiconductor as measured by the receiver. They also contribute to errors in associated gain measurements. Losses after the semiconductor only contribute to errors in the determination of semiconductor associated gain. Using low-loss tuners, bias Tees, and connectors is an important step towards minimizing the error in noise figure measurement. Although these losses will be accounted for through measurement and calculations, the corrections will necessarily be imperfect, and corrections for small errors are therefore to be preferred to corrections for large errors.

The *noise receiver* is the final portion of the noise measurement setup. This unit will have a limited frequency range over which it can measure noise performance directly. If the characterization being made lies within this frequency range, no additional components are necessary in this portion of the setup. If the device is to be characterized outside this frequency range, a frequency shift must be effected in the signal being measured. This is accomplished through the use of an external mixer, with an LO (*local oscillator*) frequency selected to place the IF (*intermediate frequency*, or output signal from the mixer) in the frequency band the noise receiver can directly measure. The noise receiver will now be measuring the noise performance of the "device under test" portion of the setup cascaded with the frequency conversion section. To minimize the contribution of the frequency conversion section, a *low noise amplifier* (LNA) should be placed between the relatively noisy mixer and the "device under test". As with losses in front of the device under test, the noise figure of the frequency conversion circuitry (and consequently, of the LNA) should be kept as low as possible to minimize errors incurred while correcting for its contribution to the measured noise performance of the device under test.

Note that when the output tuner of the "device under test" is adjusted, it will alter the generator impedance presented to the LNA. This can change the noise performance of the LNA, and consequently change its contribution to the noise figure of the device under test as measured by the noise receiver. Thus if the noise performance of the LNA is sensitive to drive impedance, an isolator should be placed between the LNA and the output tuner. An attenuator in this position is usually not an adequate solution, because it would raise the noise figure of the frequency converter too much, defeating the purpose of using an LNA in the first place.

A narrow bandpass filter is needed after the LNA to insure that no spurious signals get converted into the frequency range in which the noise receiver is measuring. The isolation of this filter should be on the order of 40 dB. The LO frequency should be selected to insure that the image frequency is not detected by the receiver; usually this translates into insuring that there is at least 30 MHz

spacing between the RF signal and the LO.

Test Measurements and Noise Parameter Calculations

The measurement setup used is shown in Figure 2. The portions of the setup are defined as follows. The "noise source" consists of the noise diode plus any isolator or attenuator following it, with the ENR of the "noise source" adjusted accordingly. The "*input tuner*", or "*tuner A*" consists of everything (tuner, bias Tee, connectors) between the noise source and the fixture holding the semiconductor being tested. The "*output tuner*" or "*tuner B*" consists of everything between the output of the fixture holding the semiconductor and the "second stage." The "*second stage*" consists of the frequency converting section of the setup along with its LNA and any isolator used to protect the LNA from impedance effects of the output tuner. The "noise receiver" is the piece of test equipment that the output of the second stage is attached to.

F_{min}

To determine the minimum noise figure of the device under test, the noise receiver is used to measure the noise figure F_{meas} of the system consisting of "input tuner" + fixtured device + "output tuner" + second stage. This measurement is corrected for second stage contribution and "input tuner" loss, yielding F_{min}. Note that this approach ignores any losses in the test fixture. Calculations indicate conductor losses are less than .005 dB for a standard coaxial test fixture, so this error should be small.

The second stage contribution to F_{meas} can be calculated from the cascade noise equation. If two noisy two ports are cascaded, the noise performance of the cascade is given by:

$$F = F_1 + \frac{F_2 - 1}{G_1} \quad , \tag{6}$$

where Fn (n = 1 or 2) is the noise figure of the nth stage and Gn is the available gain of the nth stage; all terms in this equation are expressed as numerics, not dB. To apply this equation, we must know F, F_2, and G_1 to calculate F_1.

Once F_1 is known, it must be corrected for the loss of "tuner A" to get to the device noise figure. The loss of the tuner can be calculated from its available gain, which is a function of its s parameters and its terminating impedances. In the general case, this gain can be calculated from equation (4) above. The assumption that $\Gamma_G = 0$ (i.e. the generator impedance is exactly 50 Ω) can be made in most cases. Measurements of the the s_{11} of a typical 5 dB ENR noise source yield reflection coefficients on the order of .01, small enough to add no significant error to the associated gain calculation. With this simplifying assumption, the associated gain (in dB) is given by:

$$G_{A\,dB} = 10 \log_{10} \left[\frac{|s_{21}|^2}{1 - |s_{22}|^2} \right] \tag{7}$$

Measurements of the s parameters of the "input tuner" will allow this calculation to be made.[3]

Measurements proceed as follows.

Second stage noise figure F_2 is measured by attaching the noise source directly to the second stage and reading the noise figure from the receiver. The gain displayed by the receiver at this point is adjusted to 0 dB.

The configuration of Figure 2 is re-established, and the device is placed in the test fixture. DC bias is applied. The output tuner is adjusted for maximum gain. The input tuner is adjusted for minimum noise. Tuning is iterated as necessary until optimal noise performance is achieved. The noise and gain indicated on the noise receiver are recorded as $F_{meas\ dB}$ and $G_{meas\ dB}$. These numbers are used (in numeric form) as F and G_1 in equation (6) to allow us to calculate F_1 in numeric. Equation (3) allows this number to be translated into dB ($F_{1\ dB}$).

"Tuner A" is taken to a vector network analyzer, and its s parameters are measured. The noise source s_{11} is measured to determine whether or not Γ_G can be ignored in the calculation of the tuner available gain. If so, Equation (7) can be used to calculate the available gain (loss) in dB of tuner A; if not equation (4) must be used, and the result converted to dB. Define the correction for the loss in dB of "tuner A" as $C_{A\ dB}$; this will be a negative number.

The corrected noise figure of the device being characterized is then given by

$$F_{min\ dB} = F_{1\ dB} + C_{A\ dB}. \qquad (8)$$

Γ_{opt}

Γ_{opt} can be calculated from the cascade of the noise source s parameters, the s parameters of "tuner A," and a description of the input side of the test fixture. The noise source and tuner s parameters are available from the measurements done above to generate F_{min}. A description of the test fixture will complete the necessary information.

At low frequencies (below 6 GHz) it is usually adequate to describe the test fixture as a length of perfect 50 Ω line. The length of the line can be determined by measuring the electrical delay of the input trace of the test fixture on a network analyzer. One way to do this is to measure the input rpe (reference plane extension) of the fixture. To make this measurement, the test fixture is connected to the network analyzer with no device installed. The phase of s_{11} is displayed on the network analyzer, and the electrical delay is adjusted until the phase plot shows zero degrees phase shift versus frequency. The rpe can then be read from the network analyzer display. (This method ignores capacitive fringing effects at the open socket of the fixture. Comparisons with de-embedded measurements show that the fixture used by the author

actually exhibits about 1 degree of phase shift per GHz. Adjusting the electrical delay for a phase versus frequency plot with a downwards slope of 1 degree per GHz would correct for this effect. Note that this correction is only 6 degrees at 6 GHz.)

At higher frequencies, it may be necessary to measure the s parameters of the input circuit directly. If the fixture description is generated from measurements of the fixture with the device socket open, remember to correct for the difference in fringing capacitance between an open circuited fixture and one with a device installed. Also be sure that any fixture lid is closed or cover is installed when making these measurements.

Equation (2) can be used to calculate the generator reflection coefficient presented to the fixture. S_{11} of the noise source becomes Γ_g in this equation, and the s parameters are the s parameters of "tuner A" (s_{22}' of "tuner A" is Γ_G for the test fixture). If the rpe description of the test fixture is used, the correction to this reflection coefficient to get Γ_{opt} is only a phase correction.

Therefore $$|\Gamma_{opt}| = |s_{22}'_{\text{tuner A}}|, \qquad (9)$$

and $\angle \Gamma_{opt}$ can be found from

$$\frac{\Delta ang}{360^\circ} = \frac{rpe}{\lambda} \qquad (10)$$

where $$\lambda = \frac{c}{f} = \frac{3 \times 10^{10} \text{ cm/sec}}{\text{frequency in Hz}} \text{ cm} \qquad (11)$$

Thus

$$\Delta ang = (rpe \times 12 \times f_{GHz})^\circ \qquad (12)$$

and

$$\angle \Gamma_{opt} = \angle \Gamma_{G\ \text{tuner A}} - (rpe \times 12 \times f_{GHz})^\circ \qquad (13)$$

where f_{GHz} is the frequency in GHz at which rpe is measured.

If the test fixture is described by a set of s parameters instead of by a reference plane extension, we can once again apply equation (2). Now Γ_g in equation (2) is replaced by the value of Γ_g for the test fixture calculated above, and the s parameters in equation (2) become the s parameters of the test fixture. The resulting value of s_{22}' for the input half of the test fixture will be Γ_{opt}.

R_n

If the noise performance of the device being characterized is measured at some Γ_G not equal to Γ_{opt}, equation (5) can be used to calculate R_n. A convenient choice for such a Γ_g is 0 / 0, corresponding to a generator impedance of 50 Ω. All the terms in equation (5) are now known except for R_n, and the equation can be re-arranged to give

$$R_n = 12.5\Omega \times (10^{F50dB/10} - 10^{Fopt dB/10}) \times (1 + |\Gamma_{opt}|^{-2} + (2\cos[\angle\Gamma_{opt}])/|\Gamma_{opt}|) \quad (14)$$

where F_{50dB} is the noise performance in dB of the device being characterized with $\Gamma_s = 0 \angle 0$, and $F_{opt dB}$ is F_{min} in dB.[4] Note that F50dB must be corrected for the loss that "tuner A" has when tuned to 50Ω.

Technique Limitations and Alternative Measurement Methods

A significant limitation of this technique is its heavy reliance on accurate s parameter characterization of various portions of the test setup. There are very definite limits on the accuracy with which the s parameters of any network can be measured at high frequencies; these accuracies degrade rapidly due to sensitivity limitations in the measuring equipment as the reflection coefficients of the networks being measured become large.

The mathematics involved when calculating the loss of "tuner A" from its s parameters can lead to errors if the s parameters are not known to a sufficient number of significant figures. If $|s_{22}|$ is near 1 and $|s_{21}|$ is near zero, the calculation of $|s_{21}|^2/(1-|s_{22}|^2)$ involves the ratio of two small numbers, and can be difficult to determine accurately.

Since "tuner A" must be disconnected from the test setup and carried to a network analyzer for measurement, there is an implicit assumption that the connections to the network analyzer are equivalent to the connections to the noise setup. The constant disconnections and reconnections are very time consuming, and can cause significant wear to the connectors involved. If the tuner settings are recorded while making a series of noise measurements, then there is an assumption of setting repeatability when the tuners are measured later. Many tuners are mechanical devices exhibiting significant hysteresis in their tuning, and the technique by which the settings are reproduced becomes significant in this situation.

When Γ_{opt} is determined from a description of the input half of the test fixture, there is an assumption that the test fixture has infinite isolation. Cascading the three measurements of s parameters (of noise source, tuner, and fixture input) has a larger error bound than does a single direct s parameter measurement of the entire input network.

The calculation of R_n from 50Ω measurements cannot be solved if $\Gamma_{opt} = 0$, and runs into difficulties as Γ_{opt} approaches this value. Under these circumstances measurements should be taken with a generator impedance other than 50Ω and the appropriate values used in equation (5) to solve for R_n. Equation (14) will no longer apply.

There is no correction for temperature in this technique. Temperature corrections to device noise figure can be on the order of tenths of a dB. Most noise receivers use room temperature as the "cold" reference temperature, and assume a value of 290 K. Some receivers will allow the operator to enter a measured "room temperature" (off temperature) of the noise diode to help correct for these errors. If no corrections are made, measurements made in cold rooms

(e.g. early in the morning in an air conditioned room that has not been occupied overnight) will yield lower values for F_{min} than will "normal" measurements.

A useful way to verify that the noise parameters generated by this technique are valid is to compare the measured G_A of the device being characterized with the value calculated from device s parameters and Γ_{opt}. Remember to correct the G_A of the device for the loss of both the input tuner and the output tuner; as in the F_{min} determination, equation (7) for available gain should be used. Note that for Tuner B the loss is given by $|s_{12}|^2/(1-|s_{11}|^2)$ if port 1 is to the left and port 2 is to the right.[5]

An alternative technique for determining noise parameters is based on equation (5). Noise performance is measured at a minimum of four known generator impedances, and equation (5) is used to establish a set of four equations with the four unknowns of F_{min}, $|\Gamma_{opt}|$, $\angle \Gamma_{opt}$, and R_N. Simultaneous solution of these equations yields the noise parameters. This technique seems to work well, with the proviso that Γ_{opt} is within the region on the Smith chart surrounded by the four generator impedances selected. Solutions requiring mathematical extrapolation to determine Γ_{opt} seem to be somewhat unreliable. If redundant generator impedances are used, the statistical method by which the measurement set is reduced to four equations can have a significant impact on the values generated for the noise parameters.

An interesting variant on this technique involves the use of a programmable input tuner. First four known "standard" generator impedances are presented to the device under test, and the set of equations is solved to yield an initial guess at Γ_{opt}. The tuner is then instructed to tune to four impedances surrounding this prediction of Γ_{opt}, and a "refined" value for the noise parameters results. The tuner is finally set to this second determination of Γ_{opt}, and the measured results of F_{min} are compared to the predicted. [6]

These multiple known generator techniques have the advantage that they are not as dependant on direct s parameter measurements, and do not involve constant dis-assembly and re-assembly of the test setup. They of course assume that the tuners used (or impedances presented) are perfectly characterized.

The best verification of noise parameters is to use them to design a low noise amplifier, then compare amplifier measurements to predicted performance. The above described method was used to characterize the Avantek gallium arsenide field effect transistor ATF-13135. The resulting 12 GHz noise parameters were:

$$F_{min} = 1.2 \text{ dB}$$
$$\Gamma_{opt} = .37 \angle -65$$
$$R_N = 40\ \Omega$$

The amplifier design used a short length of 50Ω transmission line, then a quarter wave transformer to match the input of the FET up to Γ_{opt}. The measured amplifier performance at turn on was F = 1.5 dB at 12 GHz with 8.4 dB associated gain. This compared well

with the predicted value of F=1.455 dB, G_A = 8.875 dB for the computer simulation of the amplifier. Further, emperical tuning on the finished amplifier could not improve significantly on this performance, i.e. for the topology and materials selected, the amplifier was near optimum "as designed."

REFERENCES

1. George Vendelin, *Design of Amplifiers and Oscillators by the S Parameter Method*, J Wiley, 1982.
2. *Avantek High-Frequency Transistor Primer Part II: Noise and S Parameter Characterization*, Avantek, Inc.
3. Strid, "Measurement of Losses in Noise-Matching Networks", *IEEE Transactions on MTT*, Vol MTT-29, March 1981 pp 247-252.
4. Appendix 1
5. George Vendelin and William Mueller, "Noise Parameters of Microwave Transistors", *Microwave Journal*, November 1987, pp 177-186.
6. Lane, "A 0.5-18 GHz Semi-Automatic Noise Parameter Measurement Technique", *ARFTG Proceedings*, June 1982.

Industry Measurement of Noise Parameters: APPENDIX I

Derivation of equation (14)

$$R_N = 12.5\Omega \times (10^{F_{50dB}/10} - 10^{F_{optdB}/10}) \times \left(1 + |\Gamma_{opt}|^{-2} + 2\cos(\angle\Gamma_{opt})/|\Gamma_{opt}| \right) \quad (14)$$

Let $\Gamma_{opt} = \Gamma \angle \theta$

then $$\Gamma_{opt} = \frac{Z_{opt} - Z_o}{Z_{opt} + Z_o}$$

so $$Z_{opt} = \frac{1 + \Gamma_{opt}}{1 - \Gamma_{opt}}$$

and $$Y_{opt} = \frac{1}{Z_o} \frac{1 - \Gamma_{opt}}{1 + \Gamma_{opt}}$$

$$= \frac{1}{Z_o} \frac{(1-[\Gamma\cos\theta + j\Gamma\sin\theta])}{(1+[\Gamma\cos\theta + j\Gamma\sin\theta])} \frac{(1+\Gamma\cos\theta - j\Gamma\sin\theta)}{(1+\Gamma\cos\theta - j\Gamma\sin\theta)}$$

$$= \frac{1}{Z_o} \left[\frac{1-\Gamma^2(\cos^2\theta+\sin^2\theta)}{(1+2\Gamma\cos\theta+\Gamma^2)} + j \frac{-2\Gamma\sin\theta}{(1+2\Gamma\cos\theta+\Gamma^2)} \right]$$

$$= G_o + j\, B_o$$

In general,

$$F = F_{opt} + \frac{R_N}{G_o} |Y_o - Y_{opt}|^2$$

so $$F = F_{opt} + \frac{R_N}{G_o} \left| (G_o + jB_o) - \frac{1}{Z_o} \left[\frac{1-\Gamma^2}{1+2\Gamma\cos\theta+\Gamma^2} + j \frac{-2\sin\theta}{1+2\Gamma\cos\theta+\Gamma^2} \right] \right|^2$$

$$F = F_{opt} + \frac{R_N}{G_o} \left[(G_o - \frac{1}{Z_o} \frac{1-\Gamma^2}{1+2\Gamma\cos\theta+\Gamma^2})^2 + (B_o + \frac{1}{Z_o} \frac{-2\sin\theta}{1+2\Gamma\cos\theta+\Gamma^2})^2 \right]$$

solving for R_N,

$$R_N = \frac{(F-F_{opt})G_G}{\left[G_G - \frac{1-\Gamma^2}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2 + \left[B_G + \frac{2\Gamma\sin\theta}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2}$$

Let $Y_G = 1/Z_o + j\,0$ (corresponds to $Z_G = Z_o$). Then

$$R_N = \frac{(F-F_{opt})(1/Z_o)}{\left[\frac{1}{Z_o} - \frac{1-\Gamma^2}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2 + \left[\frac{2\Gamma\sin\theta}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2}$$

$$= \frac{(F-F_{opt})(1/Z_o)}{\left[\frac{1+2\Gamma\cos\theta+\Gamma^2-1+\Gamma^2}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2 + \left[\frac{2\Gamma\sin\theta}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2}$$

$$= \frac{(F-F_{opt})(1/Z_o)}{\left[\frac{2\Gamma(\cos\theta+\Gamma)}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2 + \left[\frac{2\Gamma\sin\theta}{Z_o(1+2\Gamma\cos\theta+\Gamma^2)}\right]^2}$$

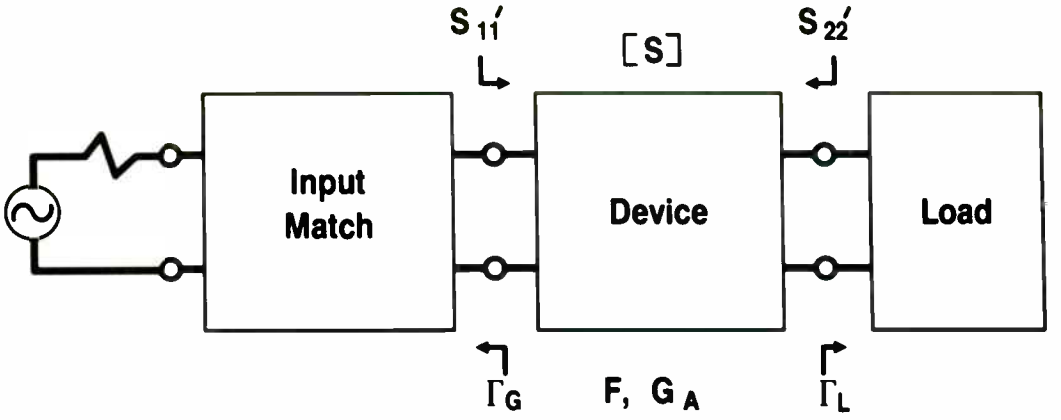
$$= \frac{(F-F_{opt})(1/Z_o)}{\frac{4\Gamma^2(\cos^2\theta+2\Gamma\cos\theta+\Gamma^2+\sin^2\theta)}{Z_o^2(1+2\Gamma\cos\theta+\Gamma^2)^2}}$$

$$= \frac{(F-F_{opt})(Z_o)(1+2\Gamma\cos\theta+\Gamma^2)}{4\Gamma^2}$$

$$= (Z_o/4)\,(F - F_{opt})\,\left(1 + 1/\Gamma^2 + 2\cos\theta/\Gamma\right)$$

If $Z_o = 50\Omega$, and expressing R_N in terms of noise figures measured in dB,

$$R_N = 12.5\Omega \times (10^{F_{50dB}/10} - 10^{F_{optdB}/10}) \times \left(1 + |\Gamma_{opt}|^2 + 2\cos(\angle\Gamma_{opt})/|\Gamma_{opt}|\right)$$



$\Gamma_G = |\Gamma_G| \angle \Gamma_G \qquad Z_G = R_G + jX_G \qquad Y_G = G_G + jB_G$

Figure 1 Noise Match Terminology

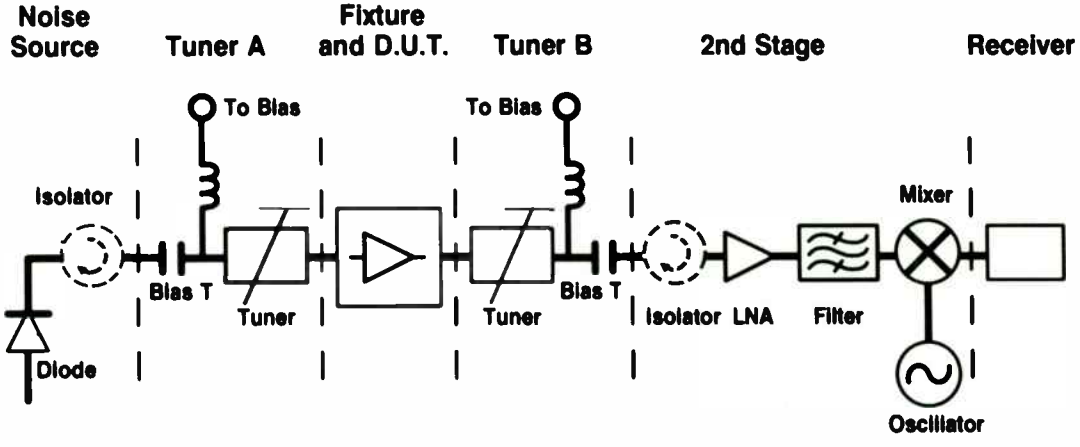


Figure 2 Noise Parameter Measurement Setup

A K Band Dielectric Resonator Oscillator

by

Ching Ho and Pramode C. Kandpal
Rockwell International Corporation
Advanced Technology, Telecommunications
P.O. Box 10462
Dallas, Texas 75207

ABSTRACT

This paper presents a computer aided design and analysis of a K band feedback dielectric resonator oscillator (DRO) used as a local oscillator in an 18 GHz digital radio system. Excellent correlations between calculated and measured data were observed. This design approach is quite general and can be used for any type of oscillator application.

INTRODUCTION

The dielectric resonator oscillator is becoming very useful as a high Q, miniature circuit component in various microwave applications. This paper presents the design and analysis of a K band feedback dielectric resonator oscillator. The DRO produces 10 dBm output power at a frequency of 17.54 GHz. A 99 MHz tuning range was obtained by mechanically tuning with a metal plate. A 16 MHz electronic tuning was obtained by changing gate voltage of the FET. The frequency stability is 7.25 ppm/^{o}C over the temperature range from -30 ^{o}C to 80 ^{o}C, which can be easily compensated by electronic tuning. A -115 dBc/Hz phase noise is observed at 100 kHz from the carrier frequency at 17.54 GHz. The frequency can be easily changed by replacing the dielectric resonator in the frequency range 17.0 - 18.5 GHz.

DIELECTRIC RESONATOR MODEL

The dielectric resonator characteristics are measured using HP8510 Automatic Network Analyzer (ANA) system in a fixture with 50 ohm microstrip line and a quartz spacer (20 mils) above the AT/Duroid 6010.5 substrate. The measurement setup simulates a transmission resonance method. The model parameters of the dielectric resonator are calculated according to Walworth [1]. The equivalent circuit and the calculated parameters of the dielectric resonator are shown in Figure 1. The calculated and measured S-parameters (S11 and S21) are shown in Figures 2(a) and 2(b). The model fits well and is valid for a narrow band around the resonance frequency (17.53 to 17.56 GHz) of the dielectric resonator.

FET MODEL

The S-parameters of the NE045 FET are measured using an HP85041A transistor fixture and HP8510 ANA system. The equivalent circuit parameters are calculated using Liechti's MESFET model [2] with some modifications for package parasitics. The equivalent circuit is shown in Figure 3. The calculated intrinsic and extrinsic parameters are shown in Table 1. The calculated and measured S-parameters show a good agreement as shown in Figures 4(a)-4(d).

DIELECTRIC RESONATOR OSCILLATOR MODEL

The parallel feedback DRO is realized using feedback through a dielectric resonator [3], [4] as shown in Figure 5. For the DRO to oscillate, these two conditions must be met:

(1) Closed loop gain $|\beta A| > 1$

(2) Phase of the closed loop gain $\theta_A+\theta_\beta = 2N\pi$

The Supercompact software program for the DRO using an FET as the active device and feedback through dielectric resonator is shown in Table 2. The reflection coefficient at output port is shown in Figure 6. Oscillation at 17.5430 GHz is obtained and verified by measurement at 17.5435 GHz, showing a good agreement between computed and measured data.

RESULTS OF MEASUREMENT

A feedback DRO was fabricated on a duroid substrate with an NE045 FET as the active device and a dielectric resonator on a 20 mils quartz as a feedback element as shown in Figure 7. The DRO produces 10 dBm output power at 17.54 GHz. A metal tuner was used for mechanical tuning. A 99 MHz tuning range with less than 1 dBm power change was observed as shown in Figure 8.

The electronic tuning was implemented by changing the gate voltage of the FET. A 16 MHz tuning range was observed with less than 1 dBm power change as shown in Figure 9.

For frequency stability test, the frequency drifted 14 MHz over the temperature range from -30 ^{o}C to 80^{o}C as shown in Figure 10. The frequency stability is 7.25 ppm/^{o}C.

The phase noise measured was -115 dBc at 100 kHz from the carrier frequency 17.54 GHz as shown in Figure 11. A summary of the performance of the 18 GHz DRO is shown in Table 3.

CONCLUSION

The design approach for an 18 GHz feedback type dielectric resonator oscillator has been described. The oscillations were predicted quite accurately, however, the power can not be predicted by using this small signal analysis model. An attempt to predict the power is being undertaken by using the large signal analysis model. The measured performance of the feedback DRO was acceptable for local oscillator applications in the 18 GHz digital radio system. The design approach is quite general and can be applied to any type of oscillator application.

REFERENCES

[1] J.H. Walworth, "Theory of Operation of the DRO", RF Design, pp 626-631, Jan. 1985.

[2] C.A. Liechti, "Microwave Field-Effect Transistors", IEEE Trans. Microwave Theory Tech., vol. MTT-24, pp 279-300, June 1976.

[3] K.K. Agarwal and Ching Ho, "Predicting Long Term Frequency Drift in FET Oscillators Using Device Modeling", 1987 IEEE MTT-S, Las Vegas, Nevada, pp 959-962, June 1987.

[4] K.K. Agarwal and Ching Ho, "Analysis of Long Term Frequency Drift in FET Oscillators", IEEE Trans. Microwave Theory Tech., vol. MTT-35, Dec. 1987.

TABLE 1

Equivalent Circuit Parameters of an NE04583 FET

INTRINSIC ELEMENTS:

g_m = 92.2 mmho
t_o = 5.3 ps
F_{3B} = 69.9 GHz
C_{gs} = 0.57 pF
C_{gd} = 0.048 pF
C_{dc} = 0.006 pF
R_I = 3.85 ohm
R_{ds} = 178.5 ohm
R_{gs} = 1 Mohm

EXTRINSIC ELEMENTS:

C_{ds} = 0.28 pF
R_g = 1.4 ohm
R_d = 0.17 ohm
R_s = 1.6 ohm
L_g = 0.32 nH
L_d = 0.27 nH
L_s = 0.016 nH
C_p = 0.015 pF
R_p = 1.4 ohm

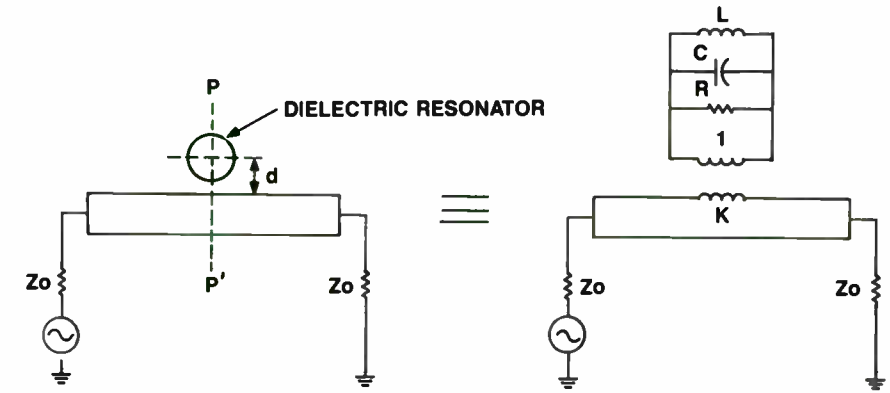
TABLE 2

```
* MODEL OF 18GHZ DRO
* USE FET MODEL AND DIELECTRIC RESONATOR MODEL
BLK
SRC 1 0 R=1.4342 C=.15001PF
SRL 1 2 R=1.4362 L=.32109NH
CAP 2 3 C=.57059PF
RES 3 4 R=3.8505
SRL 4 R=1.5927 L=.01608NH
CAP 2 5 C=.04814PF
SRL 5 6 R=.16661 L=.2735NH
CAP 5 3 C=5.6223E-3PF
CAP 5 4 C=.28455PF
VCG 2 5 3 4 G=.09218 R1=1E6 R2=178.54 F=69.864E9 T=5.2654E-12
TRL 13 1 Z=50 E=67 F=18GHZ
TRL 6 15 Z=50 E=67 F=18GHZ
RES 12 14 R=791.25
IND 12 14 L=1.0255PH
CAP 12 14 C=80.26PF
TRF 11 12 13 14 N=1
TRF 15 12 16 14 N=1
RES 11 0 R=50
A: 1POR 16
END
FREQ
 STEP 17.53GHZ 17.56GHZ 5MHZ
 STEP 17540MHZ 17545MHZ 100KHZ
END
OUT
PRI A S
END
```

TABLE 3. A Summary of The Performance of The 18 GHz DRO

Frequency	17.54 GHz (17.4-18.5 GHz)
Output Power	10 dbm
Frequency Stability	
-30 °C to 80 °C	7.25 ppm/°C (14MHz)
Mechanical Tuning	
Within 1 dbm Power Change	99MHz
Electronic Tuning	
Within 1 dbm Power Change	16MHz
Phase Noise	
at 10 KHz	-90 dbc/Hz
at 100 KHz	-115 dbc/Hz

FIGURE 1. MODEL OF DIELECTRIC RESONATOR



PARAMETERS OF DIELECTRIC RESONATOR

L = 1.026 pH
R = 791.2 ohm
C = 80.3 pF
K = 1

FIGURE 2(A). S11 OF THE DIELECTRIC RESONATOR

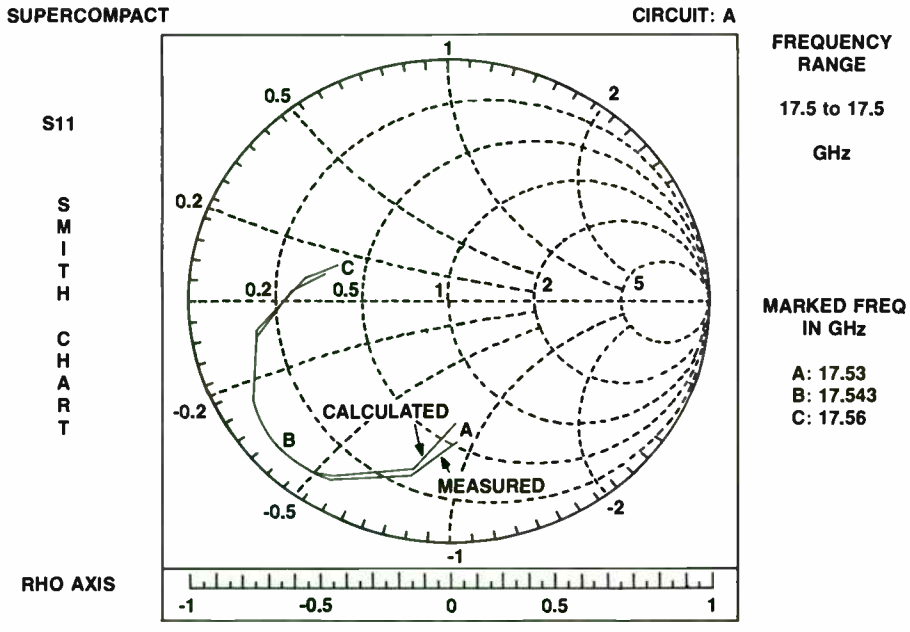


FIGURE 2(B). S21 OF THE DIELECTRIC RESONATOR

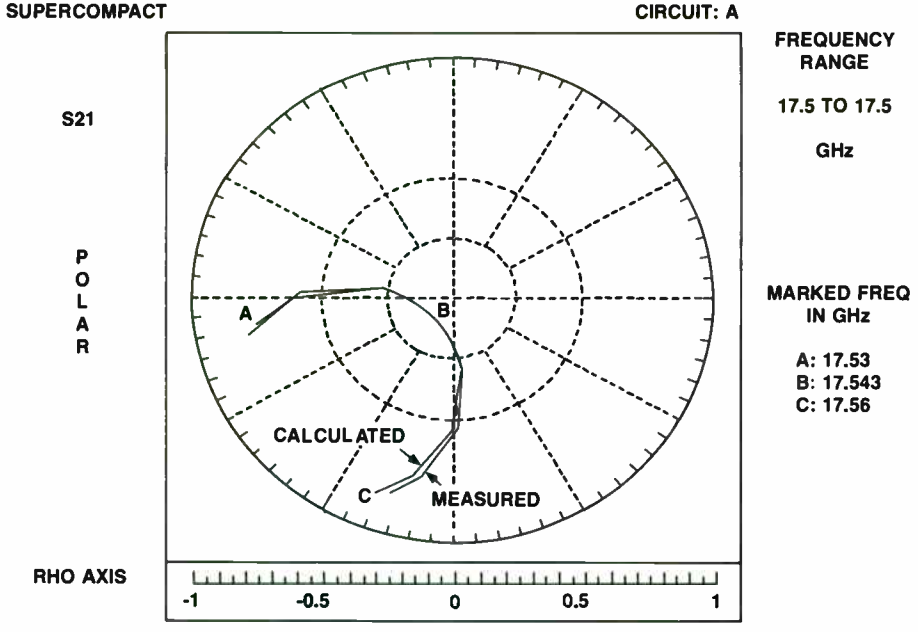


FIGURE 3. MODEL OF A GaAs MESFET

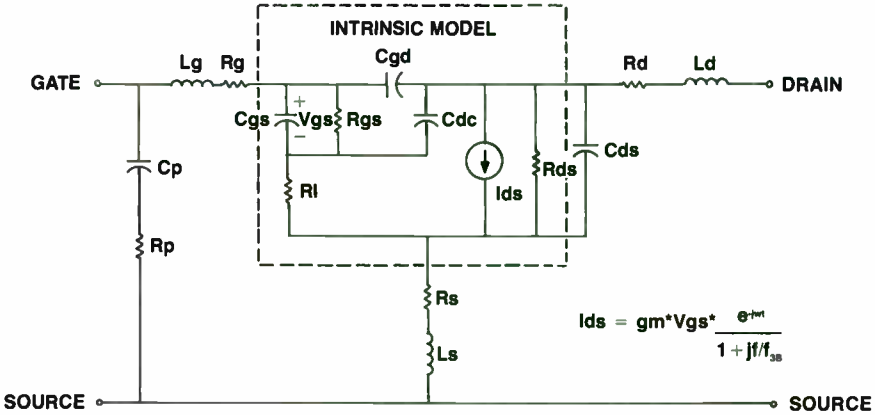


FIGURE 4(A). S11 OF NE04583 FET

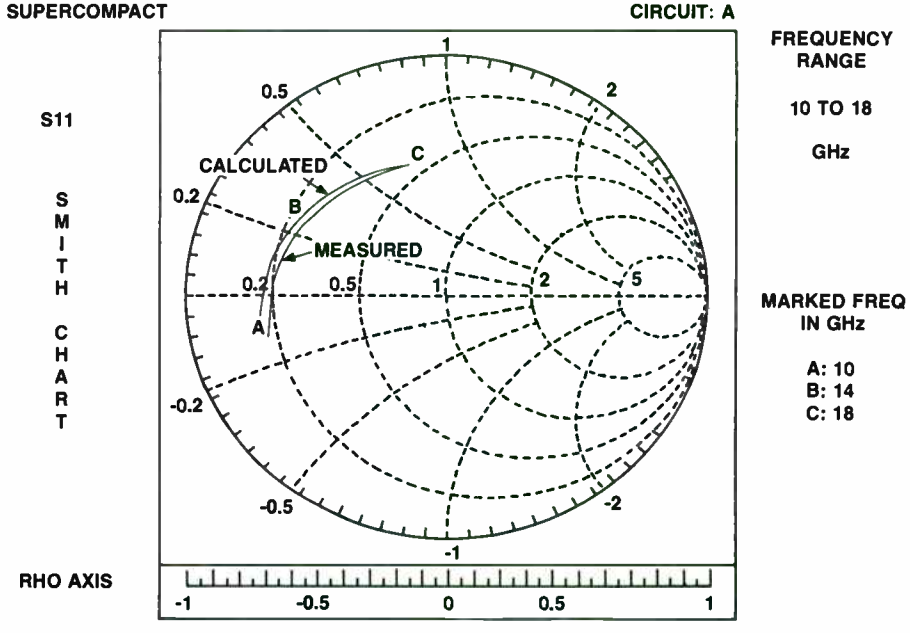


FIGURE 4(B). S21 OF NE04583 FET

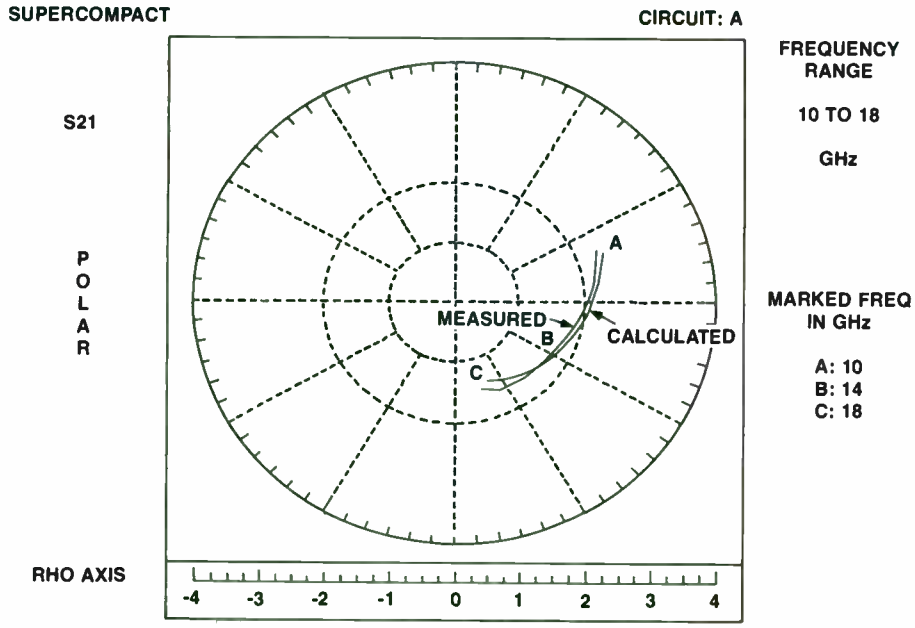


FIGURE 4(C). S12 OF NE04583 FET

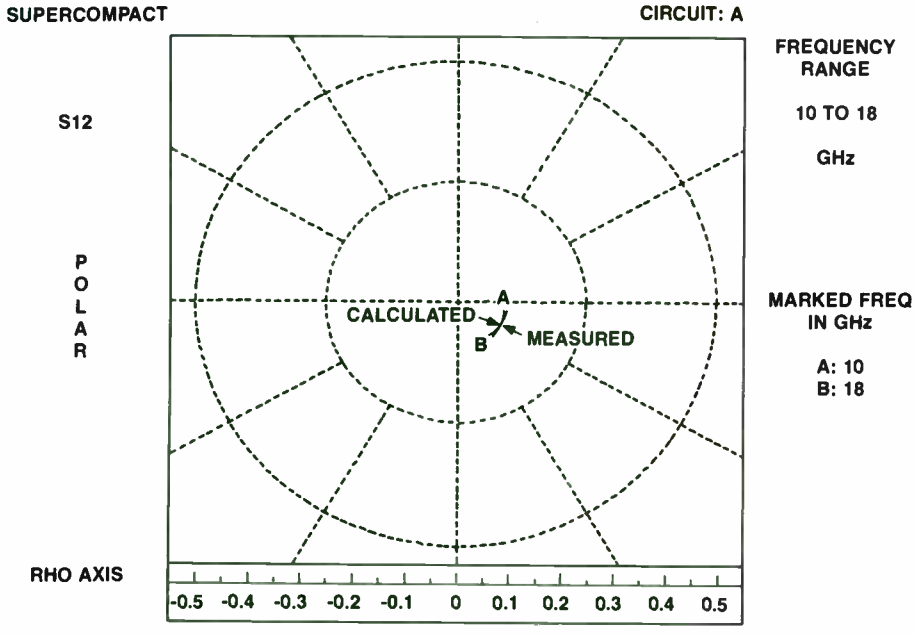


FIGURE 4(D). S22 OF NE04583 FET

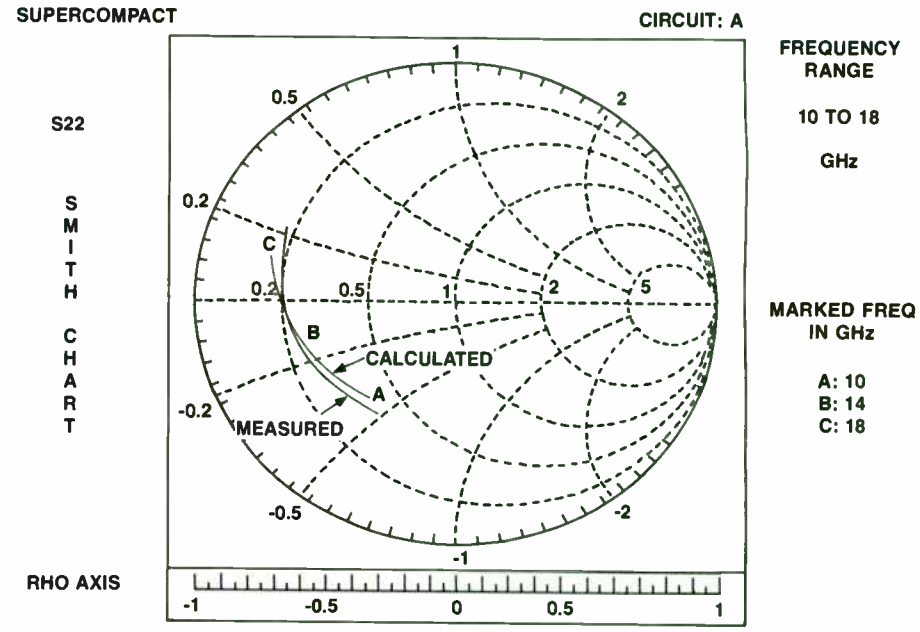


FIGURE 5. FEEDBACK CIRCUIT FOR DRO USING FET

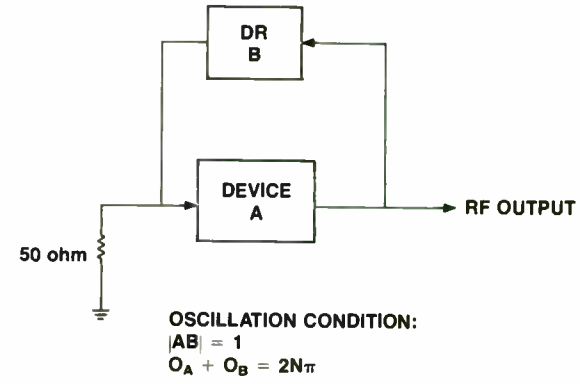


FIGURE 6. REFLECTION COEFFICIENT OF DRO

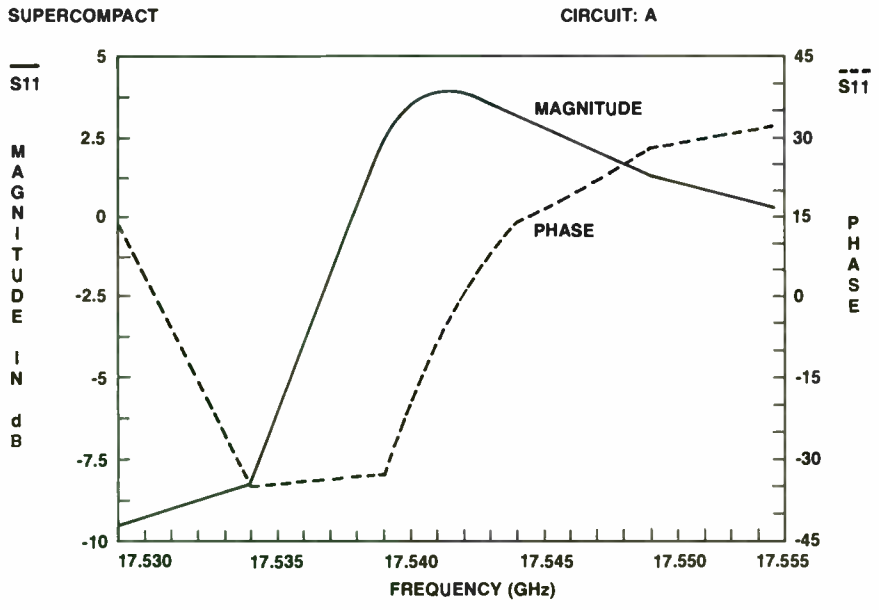


FIGURE 7. PICTURE OF A FEEDBACK DRO

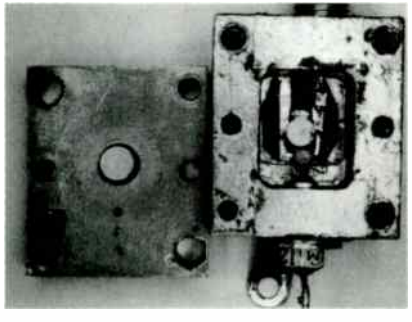


FIGURE 8. 18-GHz DRO MECHANICAL TUNING WITH METAL PLATE

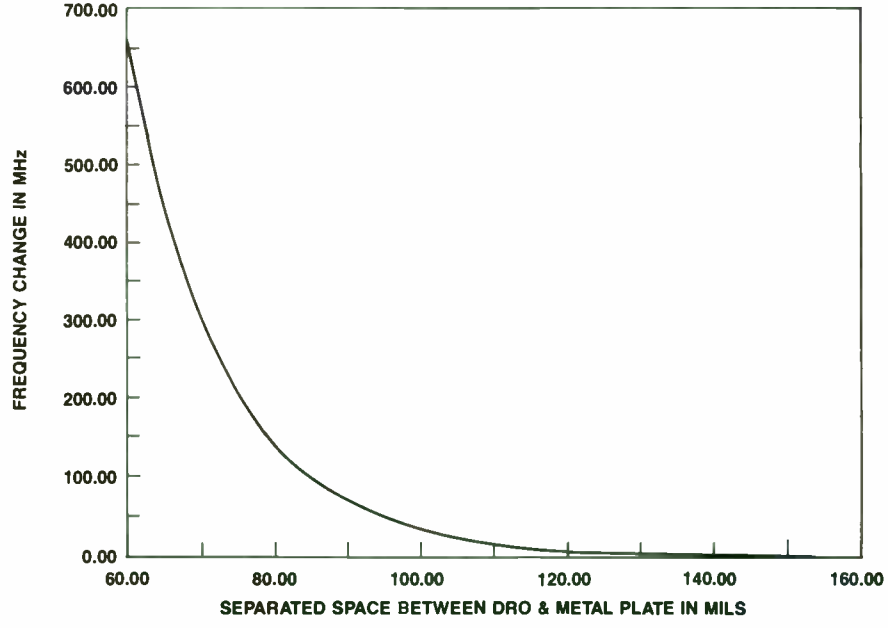


FIGURE 9. 18-GHz DRO ELECTRONIC TUNING WITH GATE VOLTAGE

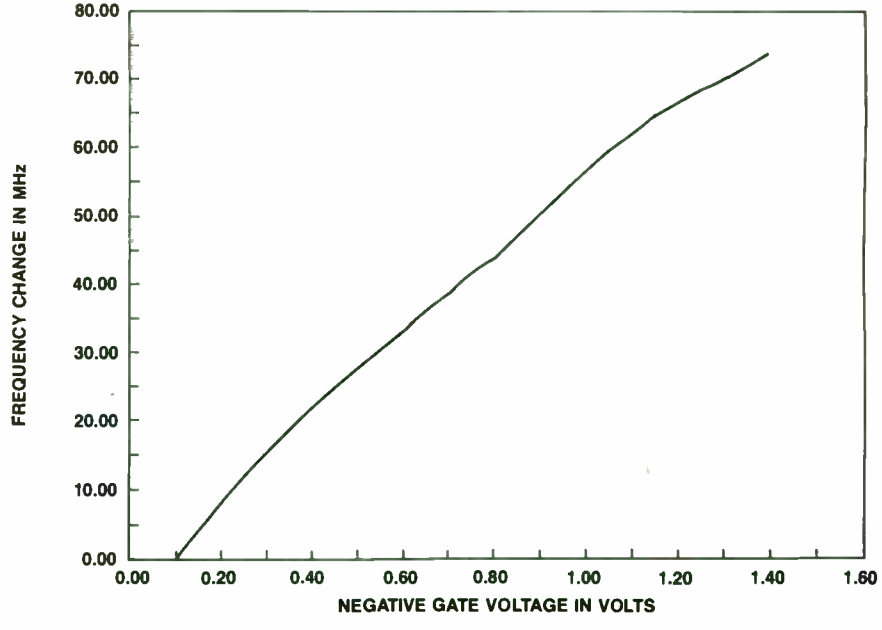


FIGURE 10. 18-GHz DRO FREQUENCY STABILITY

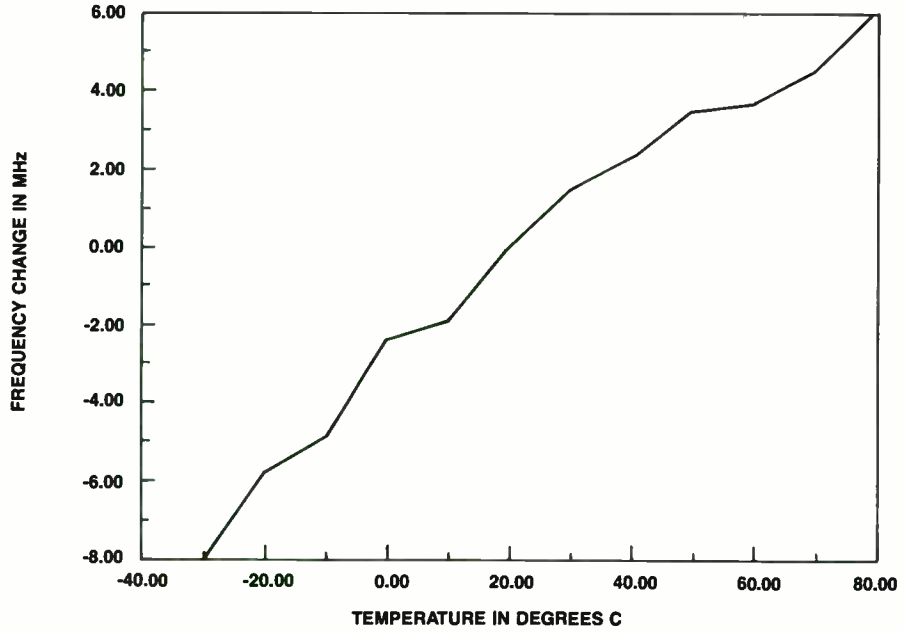
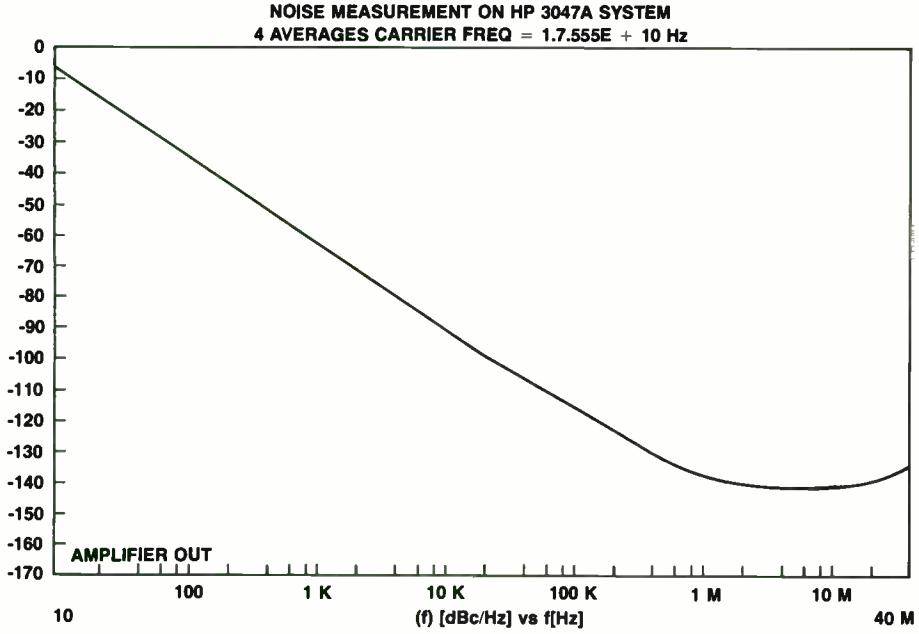


FIGURE 11. PHASE NOISE OF THE DRO



Statistical Analysis of Allan Variance, Aging, Phase Noise, and Gravitational Sensitivity of Quartz Crystal Frequency Standards

Robert R. Zeigler, Jr.
Operations Manager
Piezo Crystal Company
100 K Street
P. O. Box 619
Carlisle, PA 17013

ABSTRACT

Manufacturers of frequency standards have traditionally made claim to meeting various parameters by stating maximum performance limits, i.e. less than 5E-10 per day aging rate. This paper proposes a non-traditional but improved method of specifying important parameters such as phase noise, Allan Variance, g sensitivity, and aging.

A statistical analysis will be presented on actual accumulated data on precision quartz crystal oscillators in the frequency range of 4 MHz to 110 MHz.The investigation examines Allan Variance, long term aging, phase noise and gravitational (g) sensitivity measurements.

Data accumulated on several hundred units will be presented. This paper will show plots of the distribution of each measurement with calculations for $\bar{x}$ (mean value) and sigma (standard deviation), 2 sigma, and 3 sigma limits.

If the user systems engineer and the manufacturer of the product are aware of these statistics, each is better equipped to evaluate the performance of the standards. The data presented will show the dangers in specifying and manufacturing only to maximum limits without taking into consideration the statistical chances of meeting that limit.

Statistical process control was introduced by Dr. Deming and utilized with great success by the Japanese. Manufacturers and users in the United States are now becoming more aware of the importance of process control and the value of maintaining each process step to within measurable statistical limits.

This author is proposing that a 3 sigma value be the minimum manufacturing limit on all specified parameters.

INTRODUCTION

Precision frequency control has become increasingly important in current high accuracy systems. Communications, radar and time synchronized systems are just a few examples where precision control is crucial to performance.

Four major parameters will be investigated to define frequency instability in high accuracy quartz crystal controlled frequency standards.

- Time domain short term frequency stability per second
- Time domain long term frequency stability per day
- Frequency domain single sideband phase noise 100 Hz from the carrier
- Frequency domain single sideband noise rise due to vibration induced into the crystal and oscillator circuitry

This paper addresses each of the above topics by the testing of frequency standards operating in the frequency range of 4 MHz to 110 MHz. For this study only 3rd overtone SC cut precision Piezo Crystal Company quartz crystals were used. The completed frequency standards described are usually ovenized and constructed within a volume ranging from 3 cubic inches to 16 cubic inches.

Each area of instability will be illustrated by the following:

- Block diagram describing the test
- Formulas from which the data is derived
- Typical data from the test
- Statistical analysis of the data for various frequencies
- Statistical analysis charted for all frequencies from 4 MHz to 110 MHz

TIME DOMAIN SHORT TERM STABILITY

Allan Variance is a measurement of the frequency jitter in short periods of time, normally from 1 second to 1000 seconds. A one second time period was chosen because it represents the most often specified sampling of time. It is also in the area of best stability (typically 0.1 second to 10 seconds) demonstrated by quartz crystal oscillators being currently manufactured. For time periods of less than 0.1 seconds and greater than 10 seconds the instability is worse. Stability specifications for time periods greater than 1000 seconds are usually considered long term stability measurements.

Figure 1

Block diagram of Allan Variance test.

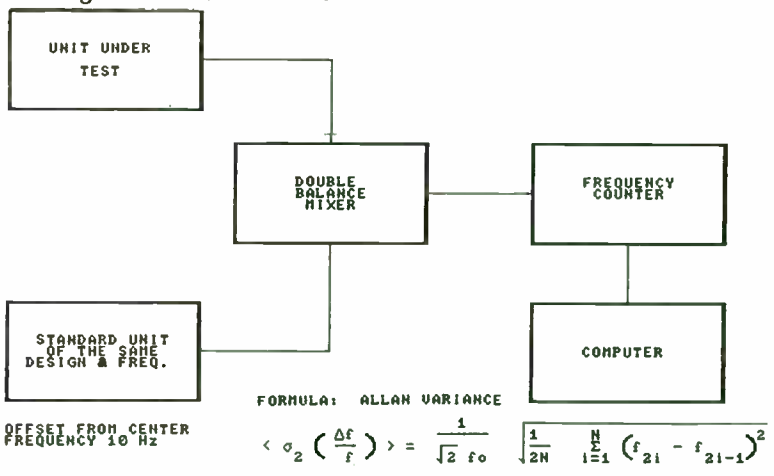


Figure 2

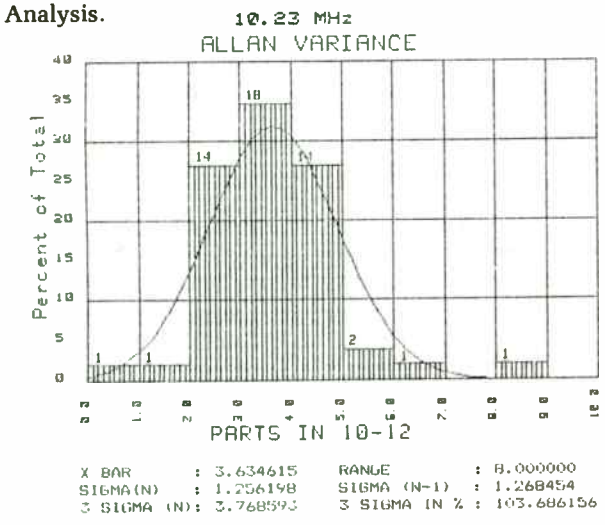
Typical Data: Frequency stability for gate times of 1 second.

```
PROGRAM NAME AVAR3
VERSION      2.01
SERIAL #     M235
TEST NUMBER= 5
BASE FREQ.= 10238310 HZ
20 SAMPLES AT 1 SEC EACH
WARM-UP TIME= 1 MIN.
AXIS OF STUDY: Y TAP
ALLAN VARIANCE VARIABLES

  3.28125760537E-12
 ---------------------
```

Figure 3

Statistical Analysis.

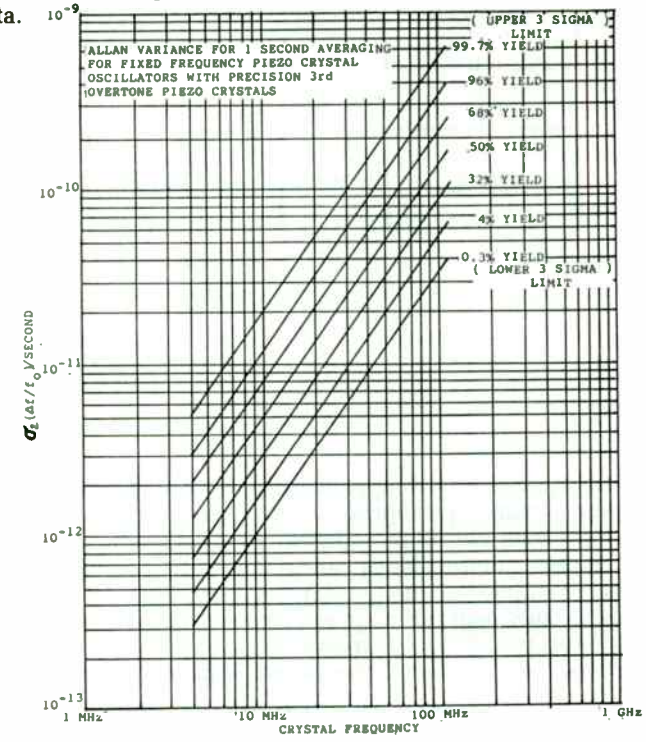


The spread one can expect with units of the same design in the same lot is normally 10:1. For this reason, a statistical analysis of the data shown in Figure 3 is most helpful. This analysis clearly demonstrates the chances of meeting the specified limits (i.e. 10 x 10^{-12}) to both the manufacturer and the user.

The above data and data from hundreds of other units of various designs were analyzed. These produced the statistical analysis shown on Figure 4 at all frequencies between 4 MHz and 110 MHz.

Figure 4

The various curves shown represent the 1 sigma, 2 sigma and 3 sigma limits of the data.



TIME DOMAIN LONG TERM FREQUENCY STABILITY

Figure 5

Block diagram of test.

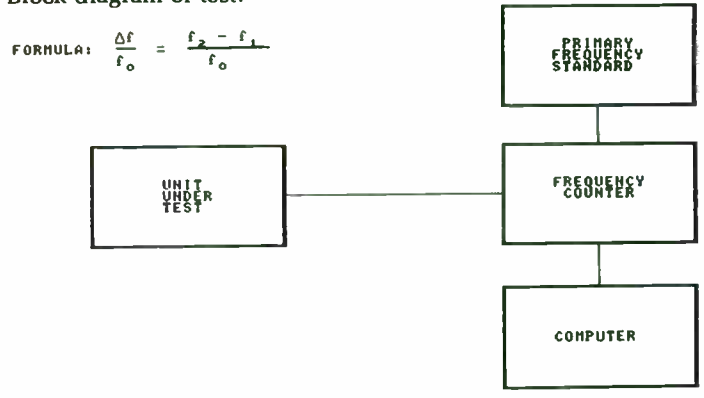
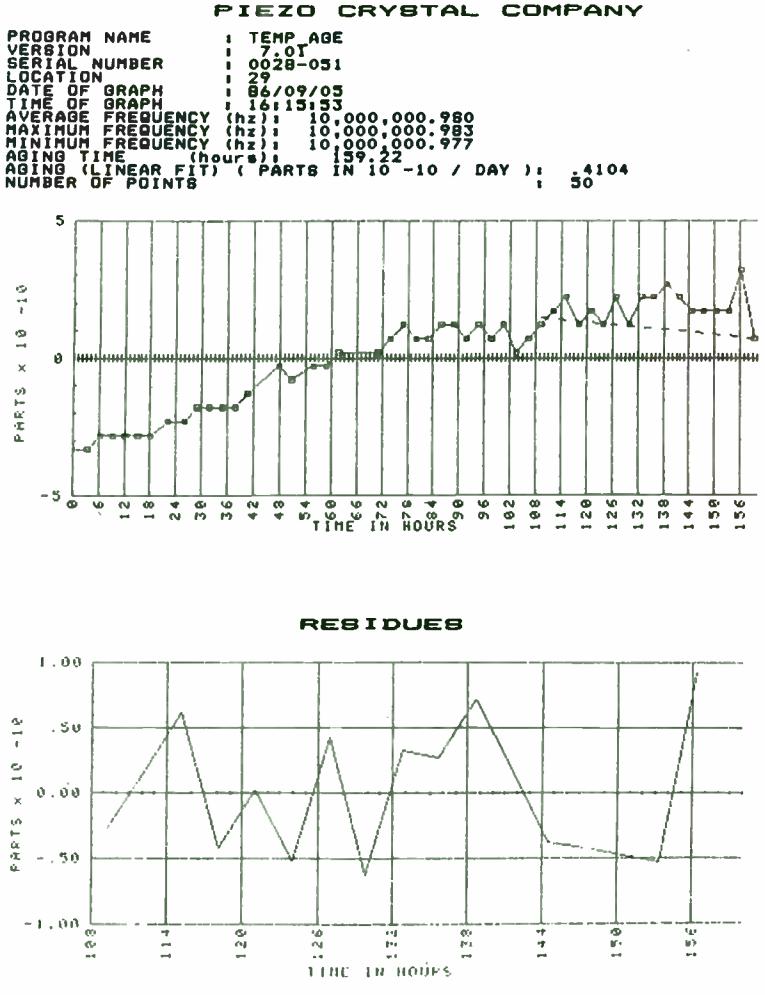


Figure 6

Typical Data: Drift rates per day at the time of shipment.

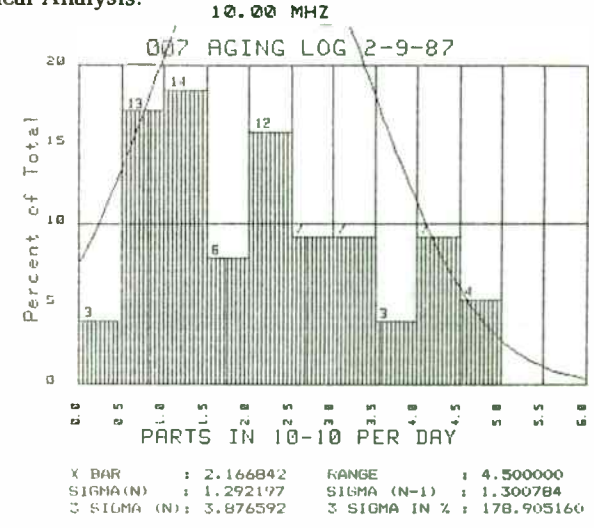


The long term frequency drift is measured by sampling the frequency periodically over a period of days or weeks. The counter sample time is usually in the range of 10 seconds to 1000 seconds, depending upon the number of digits the counter can display and the accuracy desired for the test. It is desirable to sample the frequency as often as practical throughout the time period of the test.

Frequency drift can take many forms. It can drift positive or negative, or the combination of both. For example, it may drift positive for the first 90 days, then drift negative for the balance of the life of the unit. When the drift reaches its final direction, the curve shape can be either linear, exponential, or in some cases accelerating (not normal). This can, and does, happen on low drift rate standards, i.e. better than 1×10^{-10}/day, of the same design in the same lot.

Figure 7

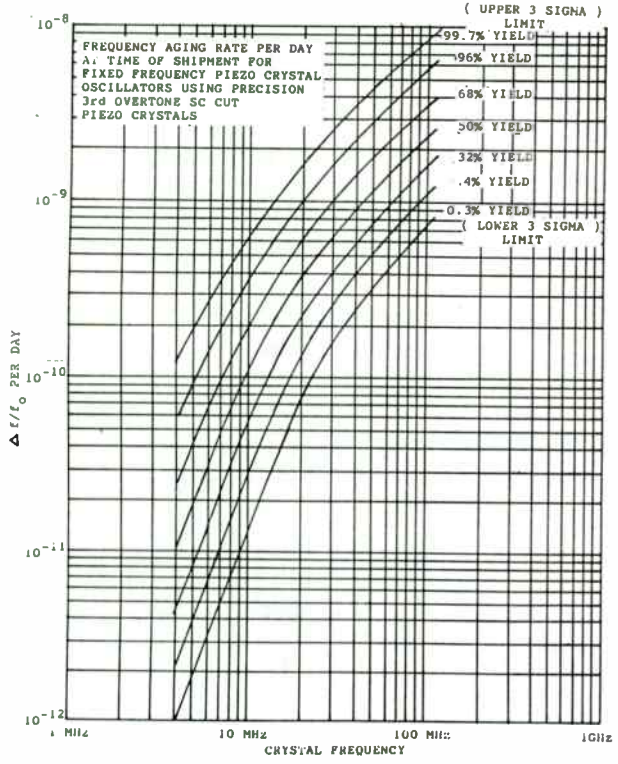
Statistical Analysis.



The curves shown in Figure 8 represent the 1 sigma, 2 sigma and 3 sigma limits of the data. It should be noted the upper 3 sigma (99.7%) limits are the preferred specification limits. Any other specified limits will result in lower yields and increased costs. Limits with 32% yield or worse must be avoided until better processing or technology is discovered.

Figure 8

Statistical Analysis of units in the frequency range of 4 MHz to 110 MHz.



FREQUENCY DOMAIN SINGLE SIDEBAND PHASE NOISE

Phase noise sidebands in frequency sources are caused by phase fluctuation due to the unintentional modulation of the oscillator frequency. The low frequency noise is mainly a measure of the loaded Q. The high frequency noise is mainly due to the signal to noise ratio of the individual active stages.

Figure 9

Block diagram of phase noise test.

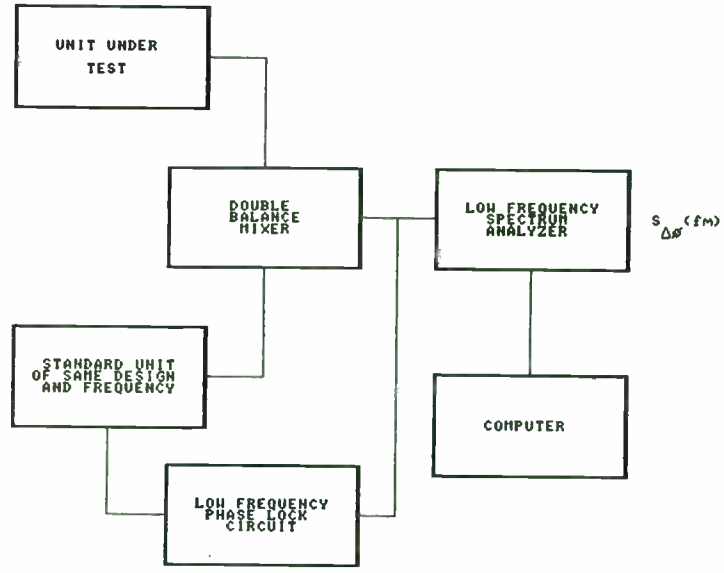


Figure 10

Typical single sideband phase noise.

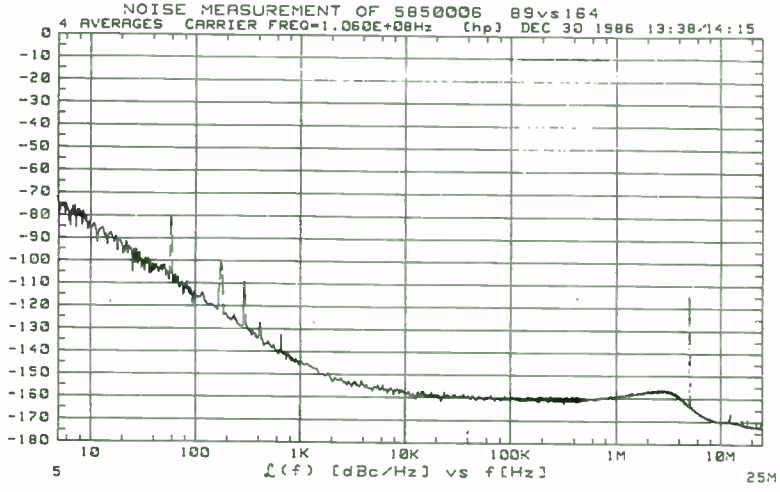


Figure 11

Statistical Analysis.

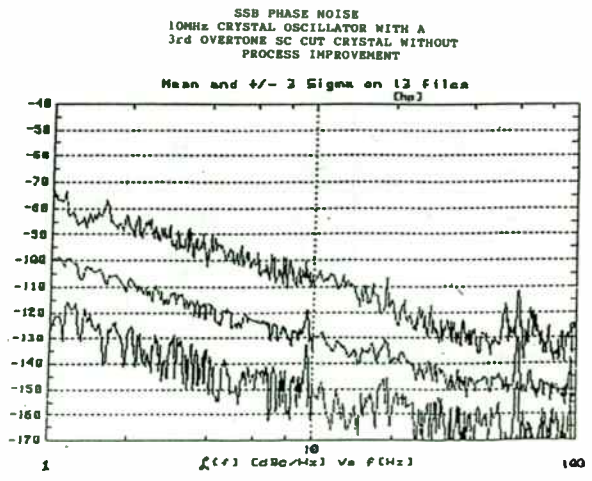


Figure 12

Statistical Analysis.

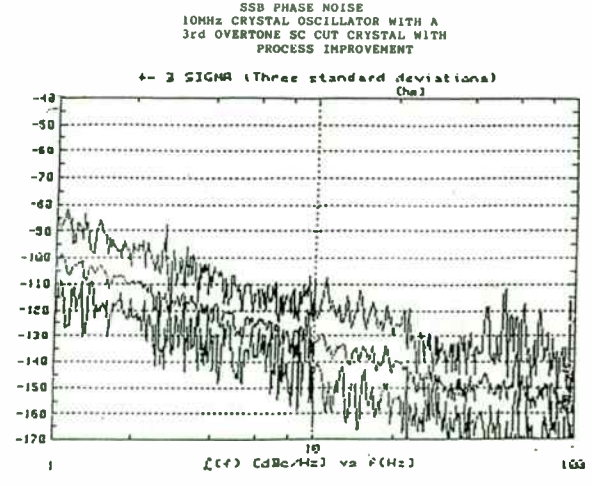
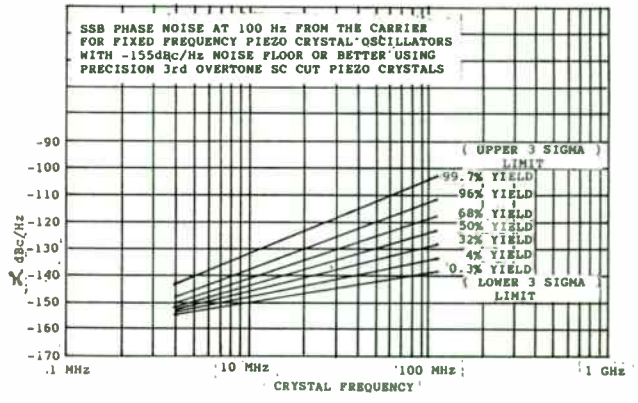


Figure 11 and 12 are of particular interest because they represent the before and after data of an experiment designed to improve the 3 sigma limits with improved crystal processing.

Figure 13

Statistical Analysis: units in the frequency range of 4 MHz to 110 MHz.

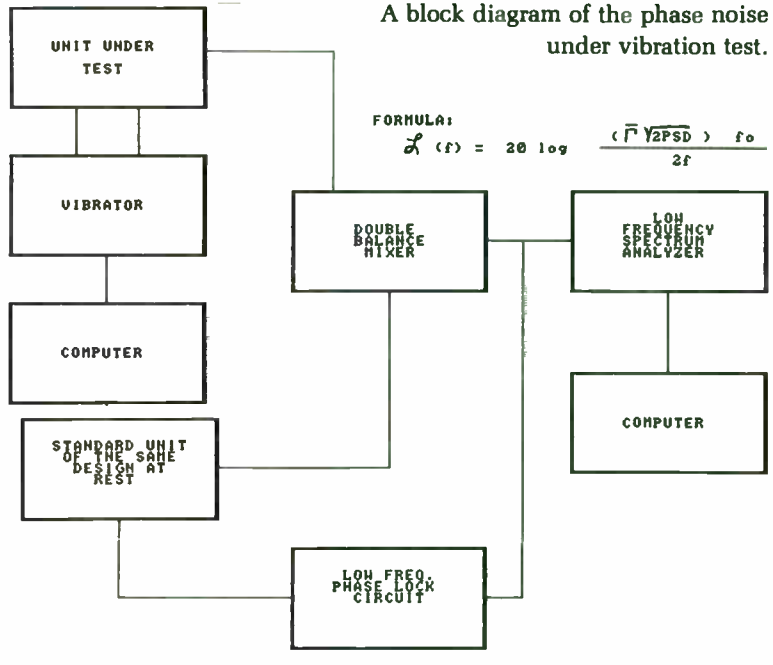


The chart in Figure 13 represents the optimum phase noise limits a designer should specify at 100 Hz removed from the carrier frequency.

FREQUENCY DOMAIN SINGLE SIDEBAND PHASE NOISE UNDER THE INFLUENCE OF VIBRATION

Figure 14

A block diagram of the phase noise under vibration test.



When quartz crystal oscillators are under the influence of random vibration they exhibit a rise in the sideband noise spectrum. The amount of rise is mainly caused by the gravitational (g) sensitivity of the quartz crystal due to the piezoelectric effect. The change in pressure on the quartz plate causes a frequency shift. The resultant sideband is at the offset frequency at which the crystal is being externally excited (normally 5 Hz to 2000 Hz). Printed circuit board and structural resonant rises also contribute to the sideband increase. Figures 15 through 19 show the typical effects of vibration on phase noise of a 10MHz unit under various conditions.

Figure 15

Phase noise of the unit under test at rest: no vibration.

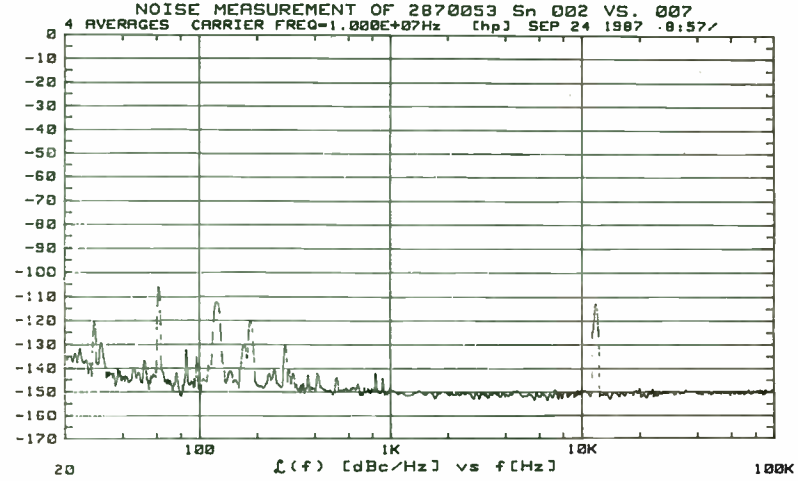


Figure 16

Random vibration spectrum to which the unit is subjected.

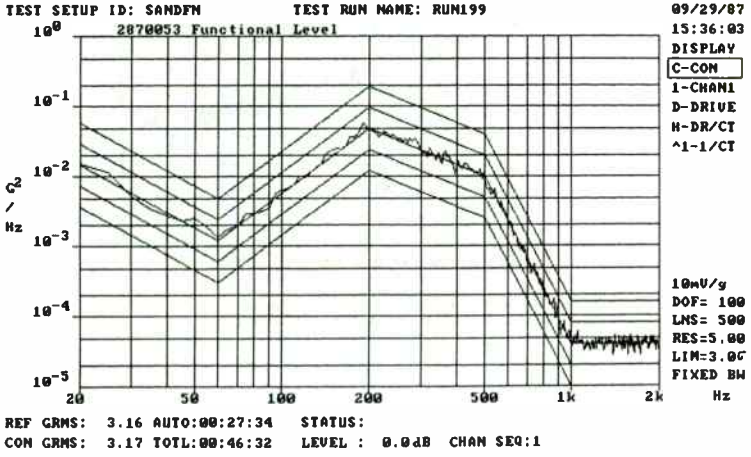


Figure 17

Sideband rise in the X axis when subjected to the vibration spectrum of Figure 16.

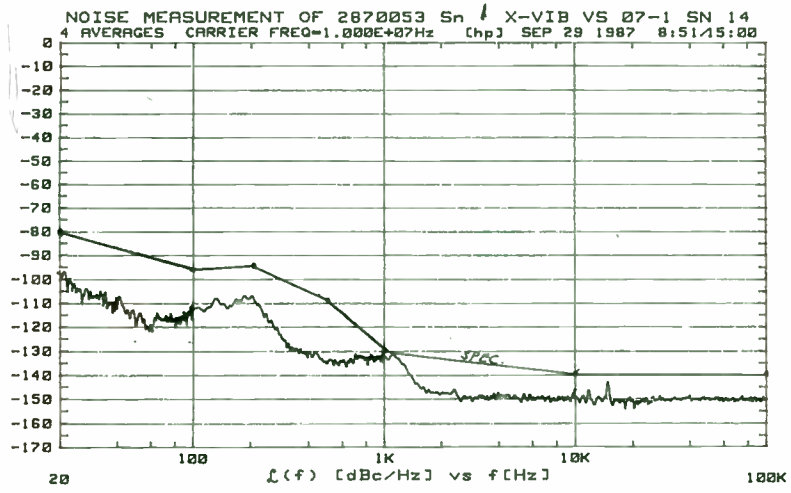


Figure 18

Sideband rise in the Y axis when subjected to the vibration spectrum of Figure 16.

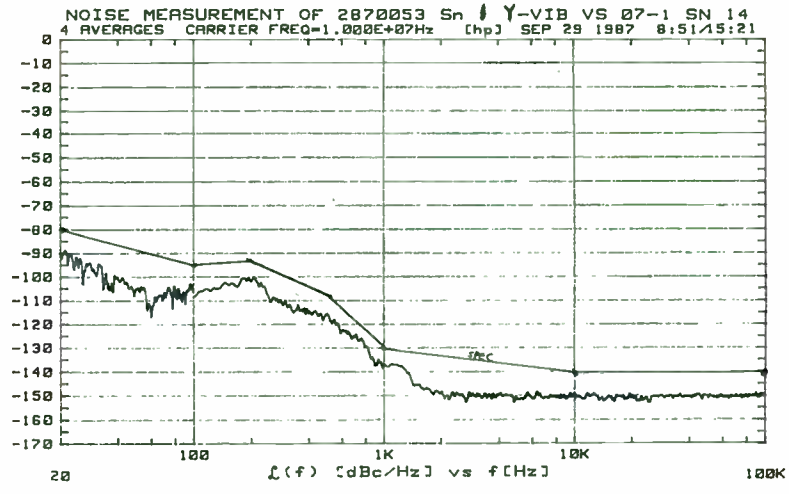


Figure 19

Sideband rise in the Z axis when subjected to the vibration spectrum of Figure 16.

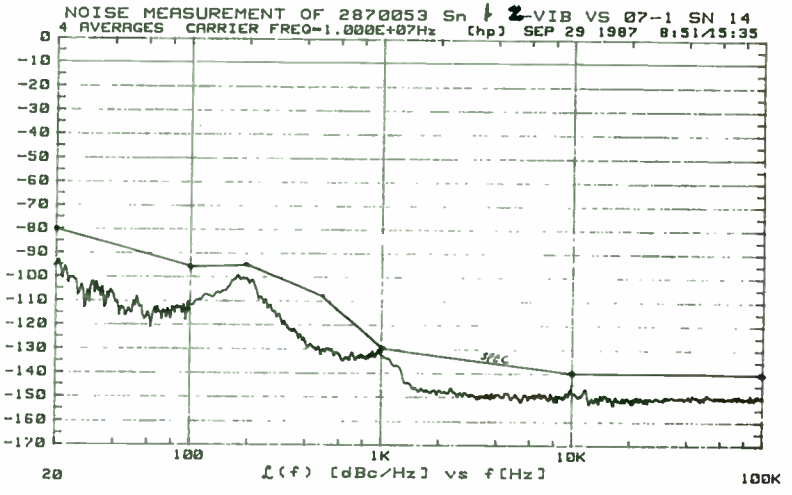


Figure 20

Statistical Analysis: g sensitivity, (Γ) is calculated from the amplitude of the sideband due to random vibration. The result can be statistically displayed.

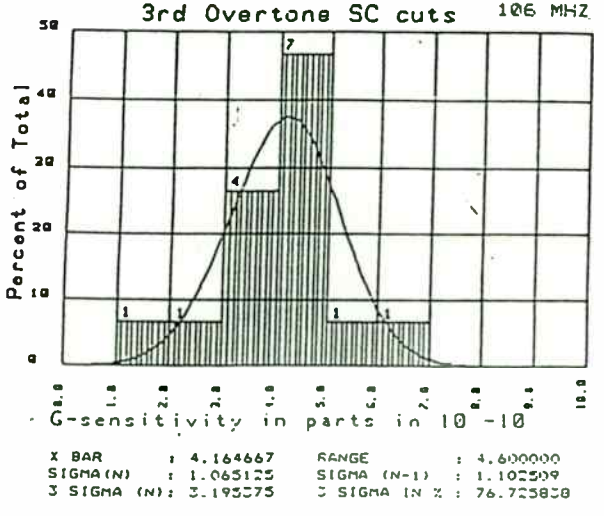
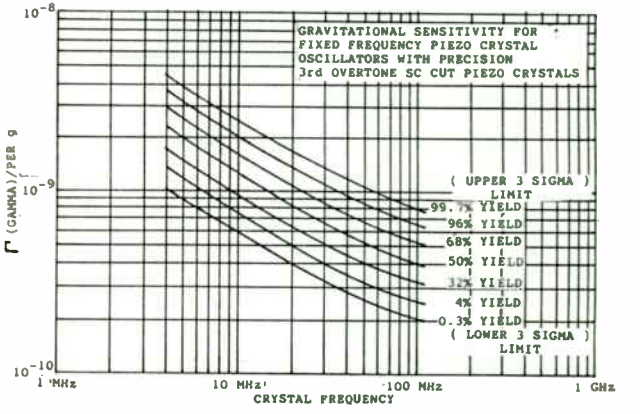


Figure 21

Statistical Analysis: g sensitivity for units in the frequency range of 4 MHz to 110 MHz.



CONCLUSION

Statistical analysis provides a more complete and in-depth picture of the data presented. The statistical chances of meeting a specified limit become immediately apparent.

This author proposes that all specifications be considered as the upper 3 sigma manufacturing limit. Systems designers evaluating manufacturers should insist on data that proves a 3 sigma capability. Use of the 3 sigma capabilities will ensure that the original specifications have been clearly stated and understood by both parties. More importantly, it ensures that the manufacturer is in control of all manufacturing processes and will ship a product that meets specified limits, with consistent quality, and a consistent rate of delivery.

Improved Accuracy for PLL Transient Analysis

by

David M. Badger
Senior Staff Engineer
Wavetek RF Products, Inc.
5808 Churchman By-pass
Indianapolis, Indiana, 46203

ABSTRACT

This paper discusses the limitations of existing PLL transient analysis. A new phase detector model is developed along with a more general form for the loop filter. The resulting model is used to simulate the frequency step response for two different PLL designs. The results are compared to the actual response for these designs. Additional model refinement is suggested.

INTRODUCTION

Since UNIVAC computer aided engineering (CAE) has been useful to engineering practices. When personnal computers came of age in the early part of this decade CAE became affordable to a large part of the engineering community. Now with shorter and shorter development cycles necessary to stay competitive in the marketplace CAE is essential. CAE saves time by allowing solutions to be tested before they are implemented thereby averting trial and error methods. Computer generated tolerance studies can help prevent surprises when a design is built in large quantities. The effectiveness of less experienced engineers can be increased since CAE allows them to perform complicated analysis without extensive mathematical skills.

Of course in order to be this useful a computer model has to be accurate. Even a small offset or gain error means the engineer has to "calibrate" the program results to the real world. For sophisticated models the errors may not even be linearly related, inotherwords a 10% error in one simulation does not imply a 10% error in another. This is the case for many PLL transient analysis programs. This paper will describe two techniques that enhance the accuracy of PLL transient models.

PLL ANALYSIS

PLL analysis can be divided into two types: steady-state and transient. Steady-state analysis is usually performed in the laplace (frequency) domain and tells the designer if the loop is stable, how it will affect sinusoidal angle modulation and VCO phase noise, and how much reference signal will be on the tuning line. Transient analysis on the other hand tells the designer how the loop responds to large external changes such as sudden changes in reference frequency or divider ratio. Steady-state analysis is the more frequently used as every PLL design should be analyzed for stability. Also closed loop bandwidth calculations give an

estimate for lock time provided the phase detector does not exceed its linear operating range. There exist several PLL analysis programs that perform accurate steady-state analysis. If the designer needs to know exactly how long and how a PLL will respond to a step change in frequency, however, he is less fortunate.

Many transient analysis programs assume a mathematical form of the response waveform based on simplifying assumptions. The actual response is determined by inserting the design parameters into the equation. Due to the complexity of the math these response solutions do not exist for systems of greater than third order. Even the solutions for low order systems assume linear characteristics for the phase detector that result in significant error when simulating large transients.

Another technique available to the designer is to model the loop as a circuit network and utilize a time domain circuit analysis program such as SPICE. This technique has no limit on the complexity of the loop transfer function but non-linear components must still be simplified. Also the designer must now translate the PLL to a circuit equivalent which is something of a inconvenience. It would seem that the PLL designer would deserve a program tailored to his needs; one that contains accurate non-linear models for expected components such as a phase detector.

NUMERICAL ANALYSIS

In reference 1 Mr. Zubiel provides the base for such a program. Why solve a complicated set of differential equations that describe a PLL when a computer can find the solution numerically? This is precisely what Mr. Zubiel's program does for a second order type II loop. Most real PLL designs however contain extra filtering for reference and noise suppression. With some improvement in the phase detector model and expansion in the loop filter complexity this idea results in very accurate simulations.

In his article Mr. Zubiel describes a way to simulate the transfer function of the loop filter using accumulator registers. Briefly stated, the loop filter integrator in figure 1 can be described by the equation:

$$Vo = -R2/R1 - (1/CR1) \int_0^t Vin\ dt$$

This integral can be calculated simply by dividing time into small increments equal to the reference period and adding the input voltage to the accumulated sum at each time increment. This numerical technique can be used to allow more general loop filter structures.

Consider for example a circuit "cell" that contains a pole and a zero as described by the laplace transfer function:

$$Y/X = K\ (\ s+z\ )\ /\ (\ s+p\ )$$

where X and Y are the input and output functions of s. This can be expanded into a differential equation by:

$$Y (s+p) = K (s+z) X$$

$$sY + pY = KsX + KzX$$

$$dy/dt + py = K\, dx/dt + Kzx$$

where x and y are the input and output functions of time. Finally this differential equation can be translated into a difference equation for analysis by numerical techniques:

$$[y(n) - y(n-1)] / dt + py(n) = K [x(n)-x(n-1)] / dt + Kzx(n)$$

$$[1/dt + p]\, y(n) = K [x(n)-x(n-1)] / dt + Kzx(n) + y(n-1) / dt$$

$$y(n) = \frac{K [x(n)-x(n-1)] / dt + Kzx(n) + y(n-1) / dt}{[\, 1/dt + p \,]}$$

where y(n) is the output for this increment of time

x(n) is the input for this increment of time

y(n-1) is the output for the last increment of time

x(n-1) is the input for the last increment of time

dt is the time increment

This equation for y(n) is used to calculate the response of the circuit cell to an input sequence x(n) obtained from another part of the program. If a more complicated loop filter is desired many of these cells can be cascaded together; the output of one being the input of the one following. This then would allow the analysis of an arbitrarily large order system.

PHASE DETECTOR MODEL

Being able to model a system with many poles and zeroes will not improve accuracy unless the non-linear elements of the system are modeled correctly. In a PLL the phase detector is the one element that is intended to behave in a non-linear fashion. The most widely used and most difficult to describe type of detector is the phase/frequency comparator such as Motorola's 4344/4044 or 12040 . The following discussion is confined to this type.

For phase errors less than one cycle the phase/frequency comparator provides an output pulse the duty cycle of which is proportional to the phase error. The DC component of the output then is the amplitude of the pulses times their duty cycle. The phase/frequency comparator has two outputs, one active for positive errors and the other active for negative. This leads to the linear phase detector characteristic for errors between -1 and +1 cycle as described by figure 2.

For phase errors greater than 1 cycle consider the case

of a positive frequency step. At t=0 the VCO frequency is too low and the phase error starts to increase. If the phase error is positive the U-not output is active as seen in the timing diagram of figure 3. As the phase error approaches 1 cycle the duty cycle of the U-not output approaches 0%. When the error exceeds 1 cycle the duty cycle returns to 100%. Note that the U-not ouput is still active. The D-not output cannot become active until two negative transitions occur on the variable input within one reference period (figure 4). Clearly this cannot occur until the variable frequency input is greater than the reference. For a reference and variable inputs of slightly different frequencies the phase/frequency detector then has the characteristic of figure 2.

This is accurate as long as the phase error is monotonic with time i.e. forever increasing. Eventually, however, the loop will increase the VCO frequency until it is equal to the reference at which point the phase error will stop increasing. Since the phase detector output would have to be positive according to figure 2 the VCO frequency will continue to increase beyond the reference. This allows two conditions to occur. First the phase error is now decreasing, and second, two negative transitions of the variable input can occur within one reference period. The other output (D-not) will now become active once the phase error decreases below a cycle boundary.

The characteristic in figure 2 must be modified in this situation or it would have the phase detector output jump positive once the phase error decreased below a cycle boundary and thus push the VCO higher in frequency and away from lock. The modification is to subtract the number of cycles skipped from the phase error once the reference frequency has been exceeded. This returns the phase detector model to its linear region between -1 and +1 cycle. Now when the phase error decreases to less than 0 the output becomes negative reducing the VCO frequency and the loop eventually settles to 0 phase error. This static phase/frequency detector model with its dynamic modification are easily implemented in software.

Although considerably more accurate than previous models this technique is still inexact in the more general case. In the model just described the duty cycle will vary from 100% to 0% for large frequency errors to give an average of 50%. The duration of the output pulse for the actual device, however, cannot be greater than the period of the reference. This is illustated in figure 5. For greatly different input frequencies the average duty cycle is then less than 50%. A more precise relation can be derived by assuming the average duty cycle will be one-half the maximum or:

maximum length of pulse = 1 / Tr

period of output pulse train = 1 / Tv

average duty cycle = (1 / 2) (1 / Tr) / (1 / Tv)

= Fv / 2 Fr

The precise manner in which the duty cycle of the output changes for large frequency errors is described by Dr. Best in reference 3. The results of his analysis are shown in figure 6. Note that for frequency errors less than 20% the average duty cycle is very close to 50% and the model developed here is adequate.

COMPARISON TO ACTUAL DATA

The accuracy of these techniques is demonstrated by comparing the program output to actual results. In the first example consider the loop design of figure 7. The tuning voltage transient for a frequency change from 492 MHz to 508 MHz is shown in the photo. The cycle skipping is apparent from the jagged rise in voltage. The computer generated normalized response shown in figure 8 is the instantaneous voltage divided by the final tuning voltage. The cycle skipping is again observed. Also the accuracy of the model is evident when the timing of events such as last cycle skipped, first overshoot, second overshoot, etc. are compared to the actual response.

Another example is shown in figure 9. In this design two extra poles are added to filter the reference from the extremely sensitive VCO resulting in a fourth order system. The computer generated response is shown in figure 10 again with good agreement. Figure 11 shows the step response for this system without the two additional poles. Even though these two poles are a factor of 20 above the loop zero they have a significant affect on the transient.

FUTURE WORK

These examples are for frequency steps of less than 2:1. In the future the phase/frequency detector model will be refined to exactly represent its physical operation. Also planned is the inclusion of complex poles and zeroes to the loop filter model. Many loops use Butterworth, Chebychev, and notch filters in order to reduce reference, fractional division, or aliasing spurs. These transfer functions would require additional levels of differentiation in the model. Finally, there are occasions when other non-linear components are added to a loop structure to realize gains in speed. These could also be modeled.

CONCLUSION

A modified PLL transient analysis program was presented that utilizes an improved phase detector model and more general loop filter in order to provide accurate step response data. The program output was compared to actual data with favorable results. Several further refinements were suggested.

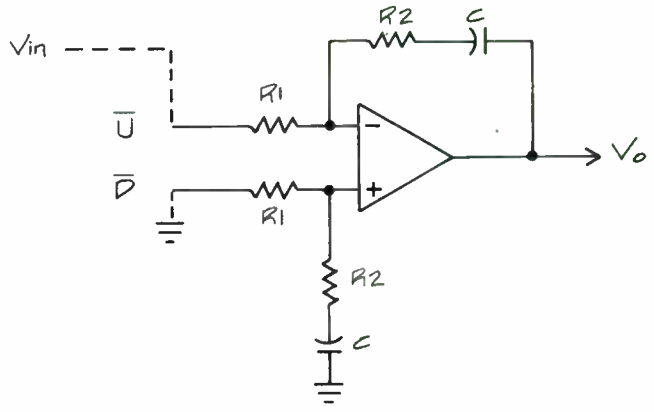
REFERENCES

1. Zubiel, Stan; " Predict PLL Transient Response with Computer Simulation"; R.F. Design, May, 1985.

2. Rohde, Ulrich L.; "Digital PLL Frequency Synthesizers: Theory and Design"; Prentice-Hall Inc., Englewood Cliffs, N.J., 1983.

3. Best, Dr. Roland E.; "Phase-Locked Loops: Theory, Design, and Applications"; McGraw-Hill Book Co., New York, 1984.

4. Gardner, Floyd M.; "Phase Lock Techniques"; Wiley, 2nd edition, New York, 1980.



$$V_{in} = V_{DD} \times \text{DUTY CYCLE}$$

$$V_o = \frac{-R_2}{R_1} - \frac{1}{CR_1} \int_0^t V_{in}\, dt$$

FIGURE 1.
THE STANDARD PLL LOOP FILTER FOR A TYPE II SECOND ORDER LOOP.

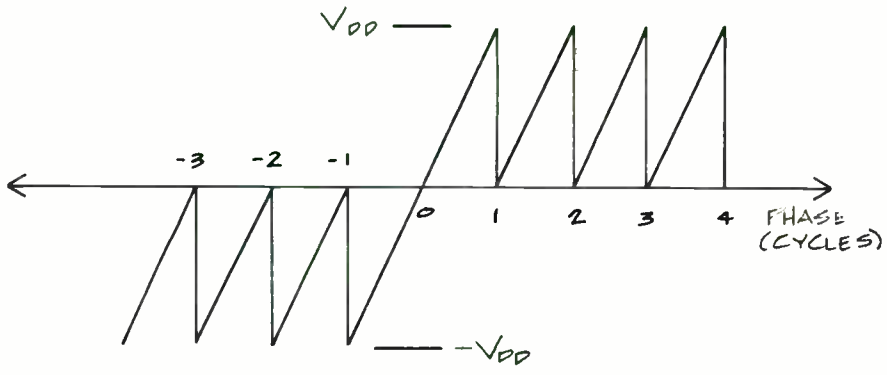


FIGURE 2.
PHASE/FREQUENCY COMPARATOR CHARACTERISTIC

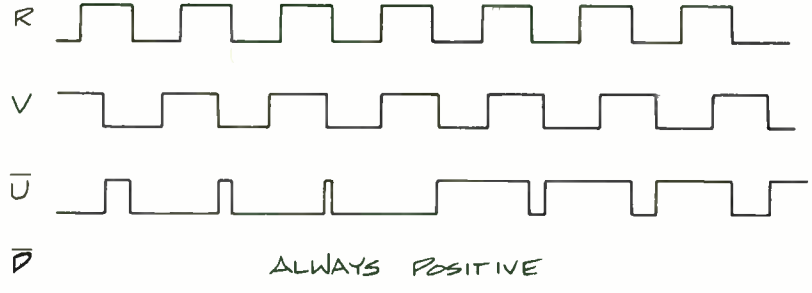


FIGURE 3.
PHASE/FREQUENCY DETECTOR OUTPUT FOR $f_r = 1.125 f_v$

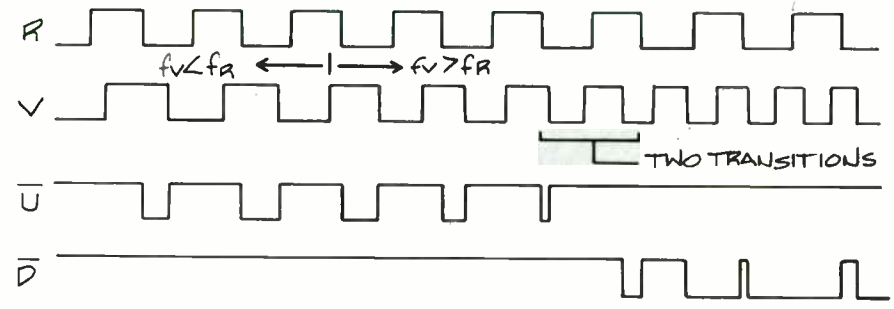
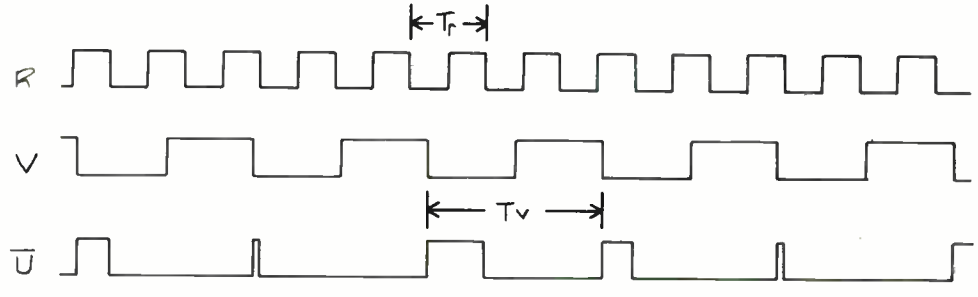


FIGURE 4.
PHASE/FREQUENCY DETECTOR
OUTPUT FOR INCREASING f_v



MAXIMUM LENGTH OF OUTPUT PULSE = $1/T_r$
PERIOD OF OUTPUT WAVEFORM = $1/T_v$

MAXIMUM DUTY CYCLE = $\frac{1/T_r}{1/T_v} = \frac{f_v}{f_r}$

FIGURE 5.
PHASE/FREQUENCY DETECTOR FOR $f_r = 2.333 f_v$

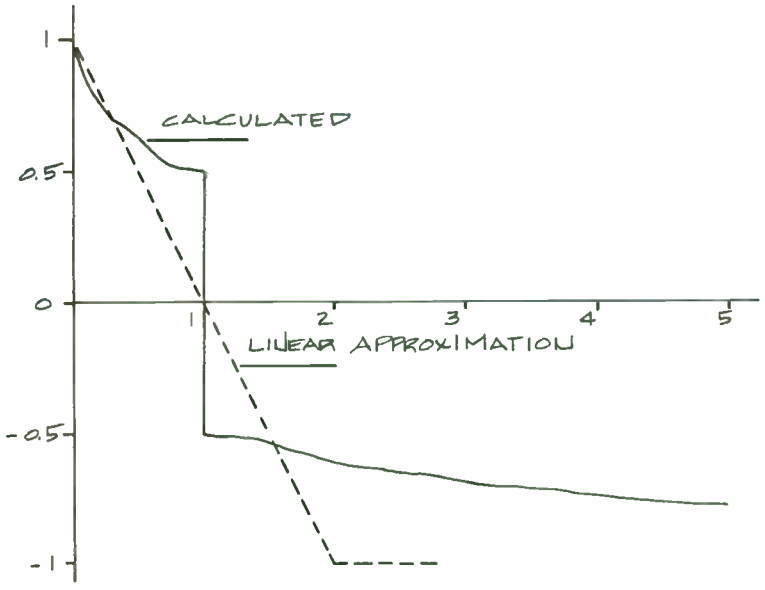
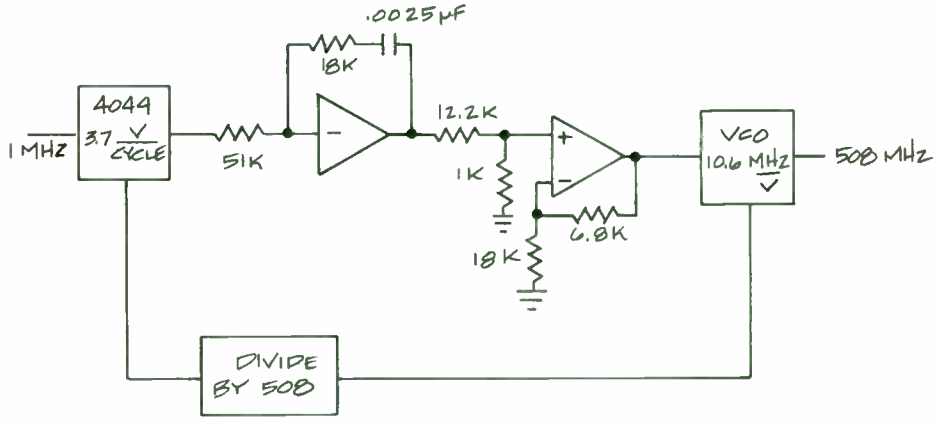


FIGURE 6.

(FROM "PHASE-LOCKED LOOPS: THEORY, DESIGN, AND APPLICATIONS", DR. ROLAND E. BEST; MCGRAW-HILL BOOK CO., 1984. COURTESY OF MCGRAW-HILL)



$$F(s) = \left(\frac{18}{51}\right)\left(\frac{s+22200}{s}\right)\left(\frac{1}{13.2}\right)\left(1+\frac{6.8}{18}\right) = .037\left(\frac{s+22200}{s}\right)$$

$$\omega_n = \sqrt{\frac{(.037)(10.6\times10^6)(3.7)(22200)}{500}} = 8200$$

$$\delta = \frac{8200}{2(22200)} = .18$$

RESPONSE FOR 16 MHZ STEP

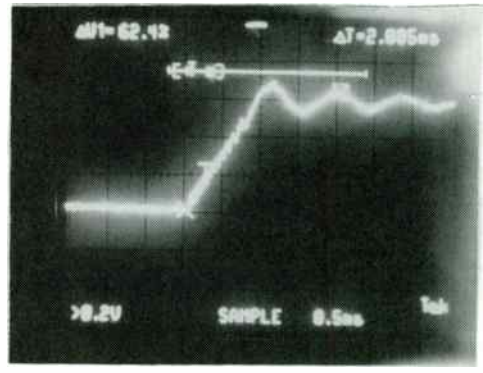


FIGURE 7
EXAMPLE 1

Figure 9

$$F(s) = \left(\frac{667}{s+762}\right)\left(\frac{6.0}{10.5}\right)\left(\frac{s+34.2}{s}\right)\left(\frac{667}{s+667}\right)$$

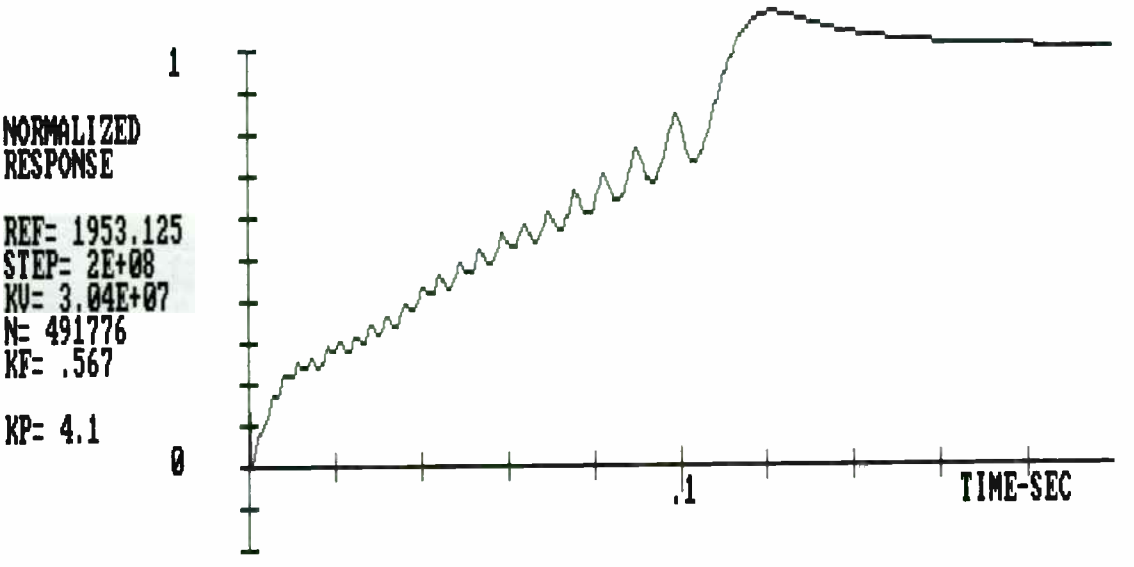
ASSUMING SECOND ORDER RESPONSE

$$\omega_n = \sqrt{\frac{(.567)(30.4 \times 10^6)(4.1)(34.2)}{491776}} = 70.1$$

$$\delta = \frac{70.1}{2(34.2)} = 1.02$$

RESPONSE FOR 200 MHZ STEP

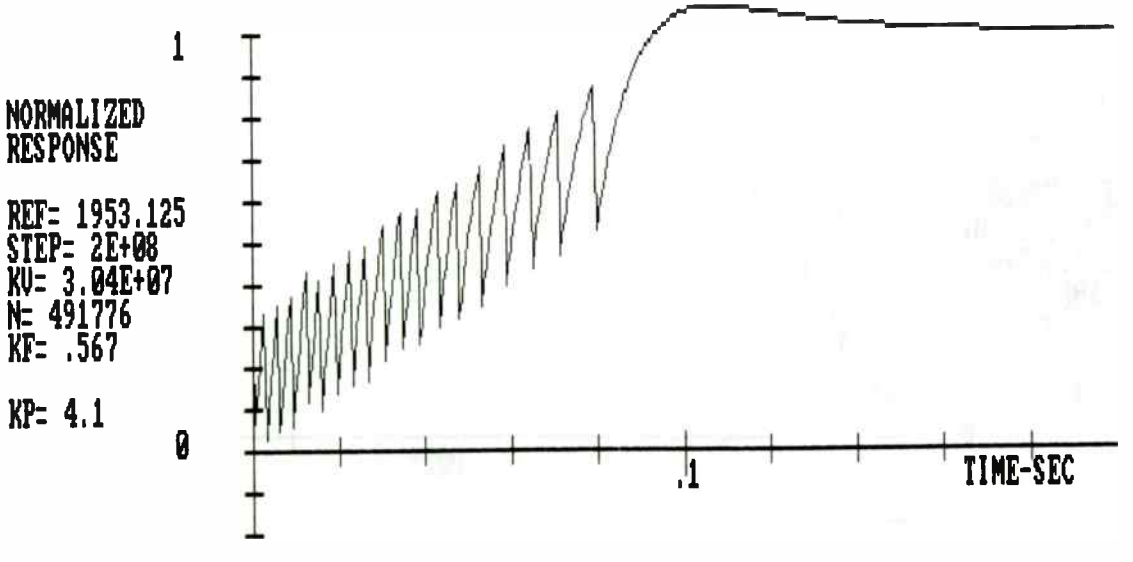
Figure 8



POLE1= 0 | POLE2= 762 | POLE3= 667 | POLE4= | POLE5=

ZERO1= 34.2 | ZERO2= | ZERO3= | ZERO4= | ZERO5=

Figure 10



POLE1= 0 | POLE2= | POLE3= | POLE4= | POLE5=

ZERO1= 34.2 | ZERO2= | ZERO3= | ZERO4= | ZERO5=

Figure 11

Pole-Zero Cancellation Reduces Noise in Phase-Locked Loop Systems

by

David M. Badger
Senior Staff Engineer
Wavetek RF Products, Inc.
5808 Churchman By-pass
Indianapolis, Indiana, 46203

Abstract

A technique is described which can reduce the contamination of noisy components in the feedforward path of a phase-locked loop (PLL). The effects of the loop's transient response are discussed and a design example is given.

Noise in PLL Systems

Just as he or she must pay taxes, the electrical engineer must deal with noise. The trick, of course, is to minimize that noise. Of interest to the PLL designer is the output signal's phase or frequency jitter. This noise can be measured as phase noise, residual FM, Allen variance, spectral density of phase fluctuations, or spectral density of frequency fluctuations, but however measured it is desired to be low.

Any active component can cause noise in a PLL. Usually, however, there is one of two dominant sources. For offset frequencies greater than the loop bandwidth the intrinsic VCO phase noise is dominant. For offset frequencies less than the loop bandwidth the phase noise of the reference determines the output noise. The latter case is true provided the loop bandwidth is large enough to attenuate the close-in phase noise of the VCO to below the multiplied reference (see figure 1). In either case these ideal noise limitations can be disturbed by other noisy elements. Noise sources after the VCO output and before the loop filter such as the divider and phase detector can degrade phase noise within the loop bandwidth. Fortunately these noise sources will only affect the lowest noise designs. Noise sources after the phase detector and before the VCO such as coarse steering DACs and op amps can degrade phase noise outside the loop bandwidth. This paper descibes a way of dealing with excessive noise outside the loop bandwidth as this is the more common problem in narrowband communication synthesizers.

An example would best illustrate our concern. Consider a PLL designed to lock a VCO that covers a frequency range of 250 MHz (see figure 2). A DAC is used to compensate the open loop gain for changes in divide ratio and VCO sensitivity. A phase noise of -100 dBc/Hz @ 10 kHz offset is desired. The DAC, however, has an output noise voltage of 30 nV/ $\sqrt{Hz}$ @ 10 kHz. Using phase modulation theory to calculate the phase noise due to the DAC:

frequency fluctuations due to DAC =

$$(30 \text{ MHz} / \text{V}) (30 \text{ nV} / \sqrt{\text{Hz}}) = 0.9 \text{ Hz rms} / \sqrt{\text{Hz}}$$

phase noise @ 10 kHz due to DAC =

$$20 \log \{ (1.414)(.9)/(2)(10000) \} = -84 \text{ dBc/Hz}$$

So, no matter how clean the VCO, the output phase noise would be limited to -84 dBc/Hz.

In this example the DAC was used to compensate the open loop gain variations of the loop, but a DAC might also be needed to aid acquisition. A system that has a filter in its feedback path will have a pull-in range no greater than the bandwidth of the filter. If this is inadequate, a DAC can be used to coarse steer the VCO to within the filter's passband. A coarse steering DAC might also be used simply to minimize the PLL error signal and thus maximize the loop's hold-in range.

DAC's are not the only potential source of noise after the phase detector. A PLL may be used to lock a remote VCO. A long, high impedance tuning line would be very susceptable to environmental noise. VCO's with large tuning sensitivities can have their phase noise limited by AM-to-FM conversion. The varactors, acting as AM detectors, modulate the tuning line with detected AM noise. The resulting carrier to noise ratio of the FM noise can be considerably higher than that of the AM noise. Placing a shunt capacitor at the tuning line will provide a low impedance and reduce the converted noise, but will also form an extra pole in the loop filter.

Reducing Noise in the Feed-Forward Path

In fact all these sources of noise could be reduced by placing a lowpass filter before the VCO. The cut-off frequency of the lowpass must be well above the loop bandwidth or stability will be adversly affected. Also, the noise will not be substantially attenuated except at frequencies well above the cut-off. The noise then is unimpeded in a wide range of frequencies from well below the loop bandwidth where the loop will correct the noise to well above the cut-off of the lowpass where the noise is attenuated. Since we are primarily concerned with noise outside the loop bandwidth anyway, it may seem reasonable to further reduce the loop bandwidth and the lowpass cut-off in order to attenuate the noise. This, of course, would further reduce the loop's ability to track the reference and may result in unacceptable lock times. This is unnecessary however if an additional zero is added to the loop filter which compensates for the added pole. Now the lowpass filter can have a low cut-off frequency without degrading loop stability or reducing loop bandwidth.

This idea is used effectively in other areas of electronics. The signal-to-noise ratio of a cassette recording is degraded by tape "hiss". The Dolby system accentuates the high frequency content in the recording process and then attenuates high frequencies during playback. This way the inherent tape noise is attenuated while the

fidelity of the recorded material is preserved. In another example, the white noise present at a FM receiver's input is converted to a parabolic spectral density function by the FM detection process. The demodulated noise is comprised mostly of higher frequencies as a result. Some radios "pre-emphasize" the high frequencies before the modulator of the transmitter and then "de-emphasize" the demodulated signal. Again the noise is attenuated while the signal is unchanged.

The usefulness of pole-zero cancellation is evident in the following example. Again consider the PLL design in figure 2 but modified as in figure 3. The cut-off frequency of the lowpass network after the DAC is set to 100 Hz. This provides 40 dB of attenuation on the DAC noise at 10 kHz. The DAC contribution to phase noise is now below the goal of -100 dBc/Hz. The highpass network before the DAC is designed to completely cancel the effect of the lowpass as is evident when the laplace transfer functions of the cascaded networks are multiplied.

Effect on Transient Response

Though very effective in reducing DAC noise, this technique may cause problems during large transients. Since the closed loop transfer function remains unchanged, the transient response should also be unaffected. For this to be the case the transient before the lowpass filter must have a large voltage "spike" to speed up the charging of the lowpass capacitor. This voltage spike may saturate the op amp driving the lowpass causing a large increase in settling time.

The level of this spike can be predicted using the initial value thereom of the laplace transform. It is well known that the closed loop transfer function for a second order type II loop is:

$$H(s) = \frac{(2dws + w^2)}{s^2 + 2dws + w^2}$$

where d is the damping factor and w is the natural frequency. Given this closed loop transfer function the laplace transform of the loop's step response is :

$$Y(s) = \frac{H(s)}{s} = \frac{(2dws + w^2)}{s(s^2 + 2dws + w^2)}$$

The lowpass filter has a transfer function of :

$$P(s) = \frac{p}{(s + p)}$$

If the input to the lowpass is designated by the function x(t) with a laplace transform of X(s) then :

$$Y(s) = X(s)\,P(s)$$

$$X(s) = \frac{s + p}{p} \; ; \; Y(s) = \frac{(s + p)(2 d w s + w^2)}{p s (s^2 + 2 d w s + w^2)}$$

The initial value thereom states that:

$$x(t=0) = \lim_{s \longrightarrow \infty} s X(s) = \frac{2 d w}{p}$$

This expression provides a good estimate of the peak voltage present at the input of the lowpass filter for over-damped systems ($d \geq 1$). For under-damped systems the initial value may not be the peak value. Though offered without proof, it seems intuitive that the peak voltage at the input to the lowpass could be estimated by multiplying its initial value by the normally expected overshoot.

If we use the above expression on our continueing example we find that the peak voltage at the input to the lowpass is 5 times the steady-state value. Thus if 10 volts is required to tune the VCO across its range, the op amp would have to supply 50 volts when tuning from one band edge to the other. Clearly this is impractical in most designs, so either a method to disable the extra pole-zero during transients must be designed or a rather long settling time must be tolerated.

It should be noted that saturation can also occur during large levels of phase modulation (or FM at low rates). If the modulation is driving the VCO directly, this problem can be corrected by summing the phase modulation with the error signal at the input to the integrator.

The discussion above assumes the added pole is canceled exactly by the zero. This is not necessary though it does simplify the analysis considerably. The problem of saturation could be reduced if the added zero was above the added pole but below the loop zero. The open loop response would still cross the 0 dB point at a 20 dB per decade rate providing stability, but the added zero would not accentuate the transients quite as severely. In this situation the added pole acts to reduce the open loop gain and thus the closed loop bandwidth unless compensated by a gain increase in the loop filter. If this approach is used, the open loop Bode plot should be examined closely to avoid stability problems.

Conclusion

A pole-zero cancellation technique similiar to Dolby cassette recordings or pre-emphasis/de-emphasis two-way communication systems can reduce the noise effects of components between the phase detector and VCO. The technique creates considerably larger transients within the loop and so is most useful in reducing noise in narrow bandwidth loops that do not require fast switching, but do require some level of refernce tracking.

References

Gabel, Robert A. and Roberts, Richard A. : "Signals and Linear Systems"; John Wiley and Sons, 1973.

Gardner, Floyd M.: "Phase Lock Techniques"; Wiley, 2nd edition, New York, 1980.

Rohde, Ulrich L.; "Digital PLL Frequency Synthesizers: Theory and Design"; Prentice-Hall Inc., Englewood Cliffs, N.J., 1983.

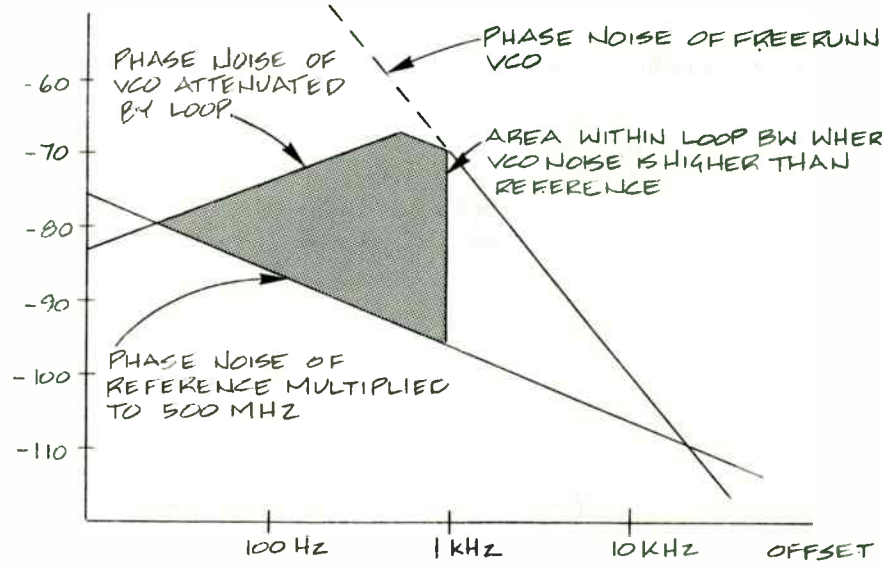
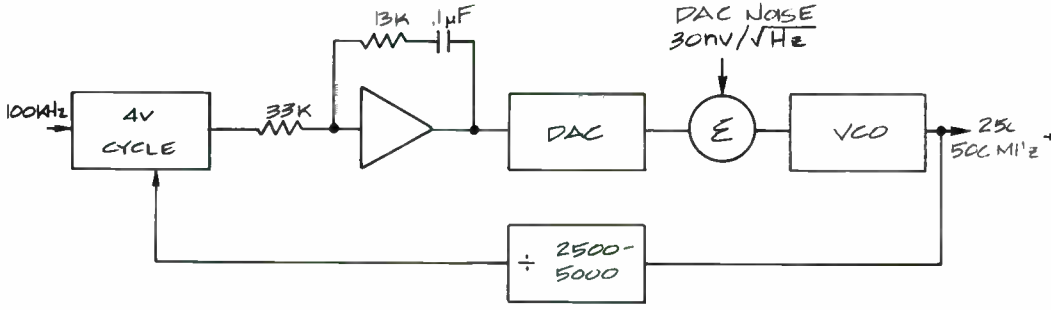


FIGURE 1

PLL DESIGN IN WHICH THE VCO PHASE NOISE IS NOT ATTENUATED TO BELOW THE REFERENCE NOISE. (d = 1, W = 3140)



$K_V = 30$ MHz/VOLT @ 250 MHz
$\quad\quad 10$ MHz/VOLT @ 500 MHz

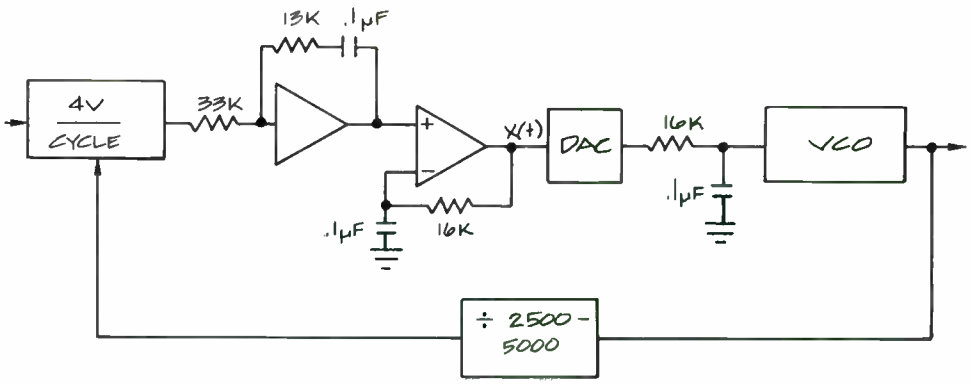
DAC AT FULL SCALE AT 500 MHz

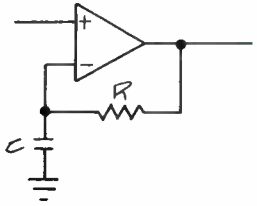
$$F(s) = k_f\left(\frac{s+z}{s}\right) = .39\left(\frac{s+769}{s}\right)$$

$$W = \sqrt{\frac{K_f K_p K_V z}{N}} = \sqrt{\frac{(.39)(4)(10^7)(769)}{5000}} = 1550$$

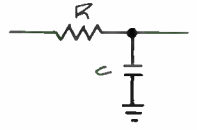
$$d = \frac{W}{2z} = \frac{1550}{2.769} = 1.0$$

FIGURE 2
A TYPICAL NARROWBAND UHF LOOP





$$F_1(s) = RC\left(s + {}^1/_{RC}\right)$$



$$F_2(s) = \frac{{}^1/_{RC}}{s + {}^1/_{RC}}$$

$$F_1(s)\,F_2(s) = 1$$

- OVERALL TRANSFER FUNCTION REMAINS THE SAME
- DAC NOISE REDUCED BY:

$$20\,\text{LOG}\left[\frac{625}{2\pi\,10^4}\right] = 40\text{ dB @ 1ckHz}$$

- INITIAL VALUE OF $x(t)$

$$x(0) = \frac{2dW}{P} = \frac{2(1550)}{625} = 4.96$$

FIGURE 3.
LOOP MODIFIED WITH POLE-ZERO CANCELLATION CIRCUITS

A Phase Lock Loop That Works - Almost: Part 3

by

Michael F. Black

Advanced Microelectronics Division

Defense Systems & Electronics Group

Texas Instruments, Inc.

8505 Forest Lane

P.O. Box 660246 - MS 3149

Dallas, Texas 75266

Part 1 introduced the idea of a phase lock loop which, despite the best efforts and intentions of its designer, fails to work at the expected level of performance. Three basic rules for design evaluation will often eliminate this "almost" phase lock loop.

Rule one states that all the individual component blocks of the loop the VCO, phase detector, etc. must work perfectly by themselves as intended.

Rule two states that these components, when connected together open loop, must function as designed or the closed loop has little chance of working correctly.

Rule three states that the closed loop PLL must work correctly with a CW input before any pulse or modulation performance can be expected.

In Part 1 component evaluation was discussed; in Part 2, the details of an open loop checkout were discussed; in Part 3, the tests of a complete closed loop will be discussed.

Figure 1 illustrates a complete block diagram of the necessary components for a closed loop phase lock loop. From Part 1 the design and gain constants of the individual blocks have been verified; in Part 2 the open loop operation and block interfaces have been verified: The closed loop operation can now be determined. With the loop closed, the integrity of the blocks and interfaces should have left few problems requiring resolution.

Before the loop will attempt to lock the VCO must be within the capture range of the PLL. Capture will occur when enough error signal from the phase detector is amplified by the loop filter to cause a frequency shift by the VCO in the correct direction: this leads to a yet larger phase detector output, and consequently more correction voltage. As the VCO and reference input frequencies move closer to each other, the correction voltage goes through a transient which should drive the frequency difference to zero and the phase difference to a steady state value.

The frequency range over which the loop is capable of capture has been mathematically described by equations for "lock up bandwidth", "drift in bandwidth" and many other bandwidths. In practice, peculiarities of phase lock loop components will, in many cases, overshadow the fine shades of differences between these bandwidths. Linear phase detectors will have a different capture range than cosine wave detectors; type two loops will capture differently from equivalent bandwidth type one loops. Offset voltages in the phase detector output and at the loop filter op amp input can cause substantial unit to unit variations.

To avoid problems in attempting reconciliation of these capture descriptions and measured parameters, force the VCO and reference frequencies together. With a system to drive the PLL into lock the acquisition time will consist of the time required by the loop filter output voltage to ramp to the required level plus the transient time for the loop filter constants to settle to a steady state phase. The transient waveform can easily by read from graphic step response tables of filters with varying damping ratios. Figure 2 illustrates the typical ramp and transient waveform expected from a "drive in" system. Without a positive "drive in", there is nothing to bring the PLL towards lock. The VCO may wander aimlessly until it eventually runs into a limit voltage from the loop filter or perhaps drifts close enough for a lock.

Several different schemes for PLL "drive in" operation are available. Some are more suitable for one application than another. Figure 3 illustrates a simple form of voltage aiding. By knowing the output frequency expected from the VCO and the exact tuning characteristics of the VCO, a ROM and D/A converter can supply the majority of the VCO tuning voltage with the loop filter supplying only a vernier voltage to finish the lock. This system is limited to fixed frequency synthesizer type applications.

A second "drive in" scheme injects sweep voltage into the loop. The sweep voltage can be sine wave, a triangle wave or a sawtooth. The VCO follows this sweep voltage until it comes within the PLL bandwidth. Then, with the correct design constants, the PLL should lock. For some applications it may be acceptable to leave the sweep input permanently connected. This would result in a correct VCO frequency but a phase shift that continuously varies following the injected sweep. Most applications require the sweep to shut down when lock is achieved. A lock detector is required to shut down the sweep; this can be a simple level detector and timing circuit or a much more sophisticated circuit. In either case the VCO would be driven into lock and the sweep turned off. Figure 4 illustrates a sweep aided search circuit.

Very careful limits must be observed on the sweep amplitude. The amplitude must never be great enough to drive the VCO through its lock up frequency and continue on without capture. Figure 5 illustrates this point with a typical circuit. A current source injects a constant current I into the loop filter feedback loop. The constant current and capacitor Cf cause a ramping voltage at the loop filter op amp output. This sweeps the VCO at a rate determined by $V_{sweep} = \frac{I * Time}{C_f}$

The larger the current I, the faster the sweep; however, at lock the VCO control voltage must stop sweeping. A lock detector can't react fast enough to do this. The loop must manage for itself.

The current I is still being injected when the VCO stops sweeping. Something must allow the VCO control voltage to remain constant. An equal and opposite current is developed by the phase detector. As long as I is present and the VCO is held at a constant frequency this offset current must be present. The phase detector can produce V volts peak to peak output. The maximum offset current from the phase detector is then $I_{offset} = \frac{.5 * V}{Rin}$

If I is greater than I offset, then the VCO will sweep past the correct frequency and never lock. In practice I should be kept less than 80% of I offset. This limits the maximum frequency sweep rate of the loop. Careful choice of component values may be required to achieve the desired rate. Figure 6 shows the VCO control voltage and the phase detector output voltage as the loop sweeps, locks and then has the search current removed.

For initial loop checkout, it is convenient to keep the search and sweep circuit as simple as possible. A large value resistor that can be temporarily connected to either a positive or negative supply will do. Figure 7 illustrates this circuit. Connection of R to either V+ or V- will inject the required search current. Lock with the injected current can be verified and then the current removed to reflect normal operation.

Figure 8 shows a preferred "drive in" lock scheme. A frequency discriminator instantly detects any frequency difference between the reference input and VCO. An output voltage forces the VCO in the correct direction for a zero frequency difference. The phase lock loop then holds a constant phase difference. A TTL frequency discriminator/phase detector such as a 4344 IC automatically provides this action and acts as a frequency discriminator or phase detector as required. It can however be confused by non CW inputs. At higher frequencies other types of discriminators are available, but most require some type of switching to turn them on when lock is lost.

Any search scheme can drive the loop towards lock; but, did it lock? This can sometimes be difficult to answer. A DVM will show that the VCO control voltage has stopped at the correct level; a frequency counter will confirm that the long term frequency is right; an oscilloscope will show that the phase detector output is at a steady state; but, the most complete, accurate check will come from connecting the VCO output to a spectrum analyzer.

With a high resolution spectrum analyzer many questions about the loop operation can quickly be answered. The correct operating frequency can be verified, the spectrum can be closely examined for any spurious oscillations, and the output noise spectrum may give more information.

The phase lock loop will attempt to make the output spectrum look like the spectrum of the reference input frequency within the PLL bandwidth. Outside the loop bandwidth the spectrum is essentially the free running spectrum of the VCO. All oscillators have a degree of noise in their output: this noise is a mixture of amplitude, frequency and phase noise. With a modern quality oscillator the AM noise is an insignificant part of the output spectrum; FM and PM noise constitute the bulk with PM noise usually the larger of the two. The spectrum analyzer displays a combination of all these noise effects.

If the reference is relatively clean and the VCO noisy, the PLL will probably not be able to completely clean up the output spectrum. Figure 9 shows the expected display expected for this case. This example uses a relatively wideband PLL for maximum noise suppression from a noisy VCO. Figure 10 displays the same VCO output but the loop bandwidth is far too narrow. The center spectral line is jittery and indistinct. The noise floor rises rapidly and meets the center spectral line only a few dB below the peak. In Figure 11 the bandwidth is too wide and exceeds the modulation capabilities of the components. The peaks on either side indicate severe peaking in the loop response with a very low phase margin. A small change could cause the peaks to rise and the loop to break into oscillation at the frequency of the peaks. All of these examples would show the same frequency on a counter with no hint of any problems.

In examining the output spectrum, the PLL should be varied over its required operating extremes. The spectrum should remain essentially the same with changes in center frequency. With a low spectrum analyzer IF bandwidth, the center frequency line should naturally be at the correct frequency and be jitter free and distinct. The noise floor, if any, should be a smooth envelope free of peaking or extraneous spectral lines. The envelope of noise should intersect the central line no closer than 30dB below the central line. Figure 9 illustrates these conditions. Special applications may require exceptions to these guidelines. Note that, while it may be close, the output noise spectrum does not completely describe the PLL loop bandwidth. The output noise is the result of the influence of the loop bandwidth on the VCO noise spectrum. The spectrum analyzer display is a mixture of both factors.

The phase detector output should be checked to further verify the loop operation. Again, with the spectrum analyzer, the output should be examined for nonharmonically related spectral lines. Even though the spectrum at this point is the result of double sided noise folded down to single sided, only multiples of the carrier frequency should be present. An oscilloscope observation can also be effective. A certain amount of noise

will be present even if a TTL phase detector is being used. This noise should be centered about an average value equal to the offset set by the loop filter. Usually this is zero. Op amp offsets or leakages may be a problem if this is not true. The peak to peak value of the noise should be far less than the peak to peak output available from the phase detector. If the noise level is approaching the maximum output the loop bandwidth is too narrow to correct all the oscillator noise. Noise spikes or slight offset drifts can easily knock the loop out of lock or cause it to slip cycles. The loop parameters need to be examined to see if a better VCO is needed or possibly a wider loop bandwidth.

The output of the loop filter can be checked with a DVM. The measured voltage should match the tuning voltage measured at this frequency when the VCO was characterized. A 10K resistor connected to the actual op amp output provides a good test point and prevents switching transients from the DVM from interfering with loop operation. An oscilloscope and low frequency spectrum analyzer can again be used to look for any previously missed oscillations.

Up to this point the measurements have been basically health checks to see if the loop has locked at the correct frequency, is not unreasonably noisy and is free of spurious oscillations. If the loop can pass these tests, serious design verification procedures can begin. The main parameter is the loop bandwidth. Even if the loop was designed as a type two second order loop extraneous breakpoints have probably pushed the order higher.

In some cases poles may be so far out as to have negligible effects on the loop performance, but factors such as the loop filter open loop gain, op amp break points and the VCO modulation function can easily distort the design goal loop response. A complete PLL frequency domain design program should allow for the op amp gain, two op amp break points, a VCO modulation breakpoint, and a phase detector breakpoint in addition to the required values of VCO gain and phase detector gain. Substitution of actual component values for theoretically derived values will further increase the match with measured results. Simply calculating response using a natural frequency and a damping ratio is a good place to start but the actual measurements may not match what was intended.

From a design analysis program a loop filter response can be predicted. Experience has shown that loops with a design goal damping ratio of two or greater will often tend to be more stable and have less problems with component tolerences and variations. In a type two loop this implies the loop filter capacitor will be relatively large and the filter zero frequency will be many times less than the closed loop 3dB frequency.

Predicting a closed loop filter response is not enough. A good set of measured data must show the intended design was relized. One method is to phase modulate the reference input frequency then demodulate the VCO output. A transfer function for the loop could be derived from the measurement. This is not the best way, because the equipment is very complicated and the analysis of the output response is difficult. A much better method gives a direct real time presentation of loop transfer

function with no more trouble than measuring any low pass filter. This method is insensitive to the loop operating frequency whether it is kiloHertz or gigaHertz. Figure 12 illustrates a basic PLL block diagram with a forward gain path composed of the phase detector gain (Ko), the loop filter transfer function F(s) and the VCO gain Kv(s). Feedback gain for this example is unity. The transfer function for this closed loop is $\frac{V_o}{V_i} = \frac{G}{1+GH}$ or

$$G(s) = \frac{K_o * F(s) * K_v(s)}{1 + K * F(s) * K_v(s)}$$

To demonstrate the measurement technique the block diagram is redrawn in Figure 13 with gain of 1 isolation blocks and a summing junction is provided for a test signal, VA. For an output from the phase detector, the transfer function $\frac{V_B}{V_A} = \frac{G}{1+GH}$

The forward gain path from VA is again

$$G(s) = F(s) * K_v(s) * K_o$$

The feedback path is a unity gain block. So

$$\frac{V_B}{V_A} = \frac{K_o * F(s) * K_v(s)}{1 + K * F(s) * K_v(s)}$$

Thus the transfer function for a signal injected at A with an output from the phase detector at B is the same as the transfer function from the reference frequency input to the VCO output. A video signal connected at A can be used to generate a response a point B. The vector of the ratio VB/VA will produce the data for the phase and magnitude response of the PLL loop bandwidth.

In practice this test is implemented as shown in Figure 14. An additional input resistor is connected to the loop filter input. The input to A is swept over the desired range. An oscilloscope, a gain phase meter, a spectrum analyzer or, preferably a network anaylzer will provide information on the transfer function. After the loop has been locked, the level of the input at A can be adjusted. The level should be as low as possible. 30dBc sidebands at the VCO output are a good goal. The eventual choice of signal level will be determined by the capabilities of the test equipment.

Frequency sweep range should cover at least two decades and extend well into the roll off area, at least 20dB down. The actual magnitude of the loop gain is normally not important. It is easiest to compare plots if the initial gain reading at the lowest measurement frequency is normalized to 0dB. Figures 15, 16 and 17 are the bandpass plots of the three noise spectra of Figures 9, 10 and 11. Note the overdamped peaking in Figure 10 and the very narrow bandwidth of Figure 11. The measured data can now easily be compared against the predicted data. Figure 18 illustrates the degree of match that can be achieved. This is a very important step. All the calculations and predictions will mean little if the measurements don't agree. If the measurements do agree then an accurate model can be generated that can be used for transient testing or further analysis and predictions. If the measurements don't agree something may be missing in the loop model. Aside from the pitfalls discussed in Part 1 and 2, unknown frequency poles and op amp limitations are usually the biggest barriers. Injection lock problems produce erratic peaking and slopes not normally expected. The PLL closed loop frequency response is one of the best overall indications of design verification. The measured bandpass should match the classical graphs for a given natural frequency and damping

ratio modified by the extra higher order poles.

A last point of interest is a further examination of the VCO output noise spectrum. With the spectrum analyzer a spectrum on only one side of the carrier beginning a set frequency removed from the carrier is recorded. This spectrum is displayed with a linear frequency sweep. If the data is reformatted for a log frequency sweep, the noise spectrum can be easily compared against the loop bandpass plots. The same spacing should also be used on the amplitude scale. Figures 19, 20 and 21 plot this data for the three examples previously discussed. Note by comparison that the shape of the output noise spectrum follows the shape of the bandpass plot but is not an exact replica. This implies that the noise spectrum can be useful as a general health check but should not be relied on for meaningful data on loop stability.

Through the three parts of this series on phase lock loops the emphasis has been on careful measurement of the performance of each block to use in building a better more accurate overall model. The three design verification rules offer a systematic method to complete this measurement. The problems with the individual pieces of the loops must be corrected before moving on to the next level. With the problems corrected and accurate data available, the loop can be closed with confidence that the lock will be right, the transient performance will be right and the bandpass will be right.

If the basic type two loop can be made to work as predicted, and if more exotic structures are required, then they can be incorporated. Most PLL problems can be solved with basic hardware detective work. Seldom does the problem lie with loop theory or exotic equations. Perhaps with these thoughts in mind there will be fewer "almost" phase lock loops.

BASIC CLOSED LOOP PLL

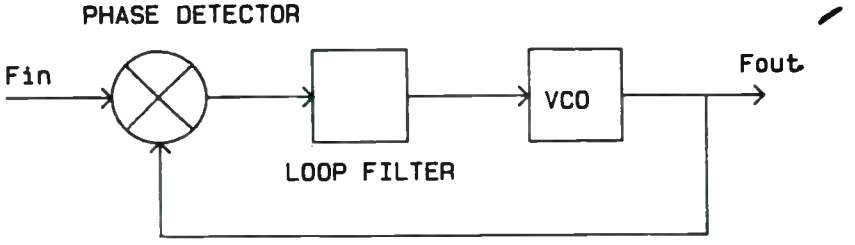


FIGURE 1

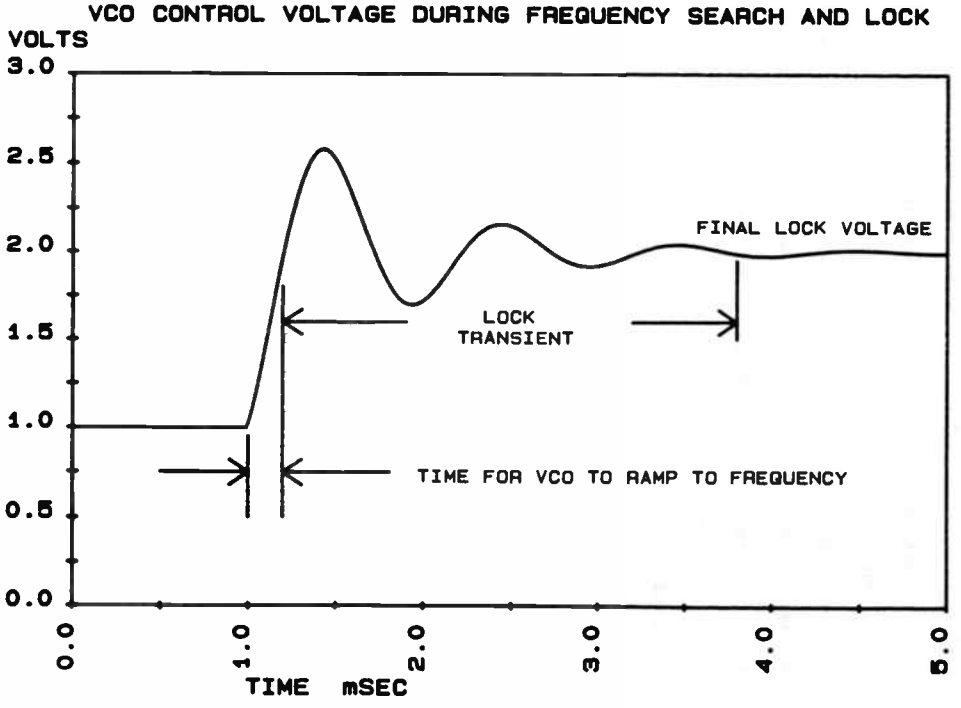


FIGURE 2

PLL WITH VOLTAGE AIDING FOR SEARCH

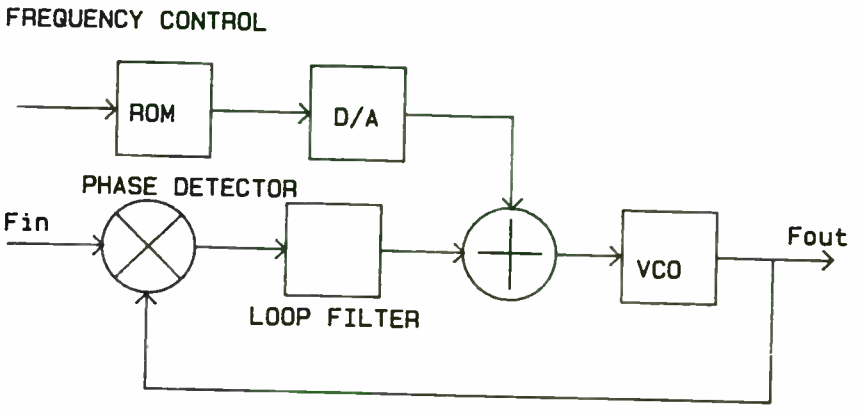


FIGURE 3

171

PLL WITH SWEEP SEARCH FOR LOCK

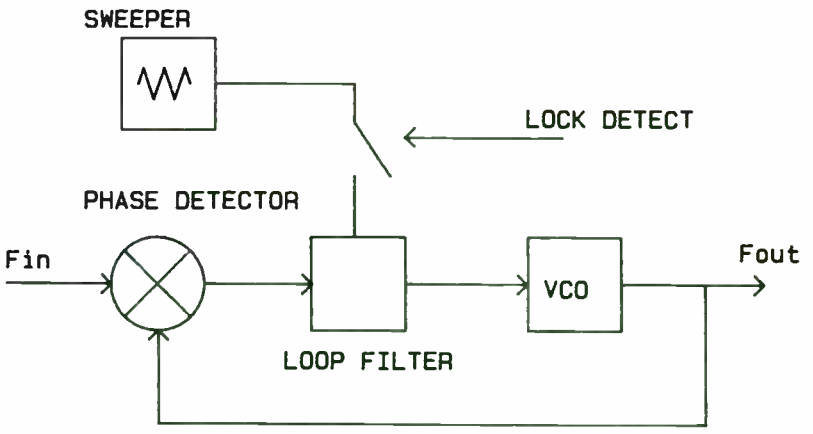


FIGURE 4

PLL WITH CURRENT INJECTION FOR SEARCH

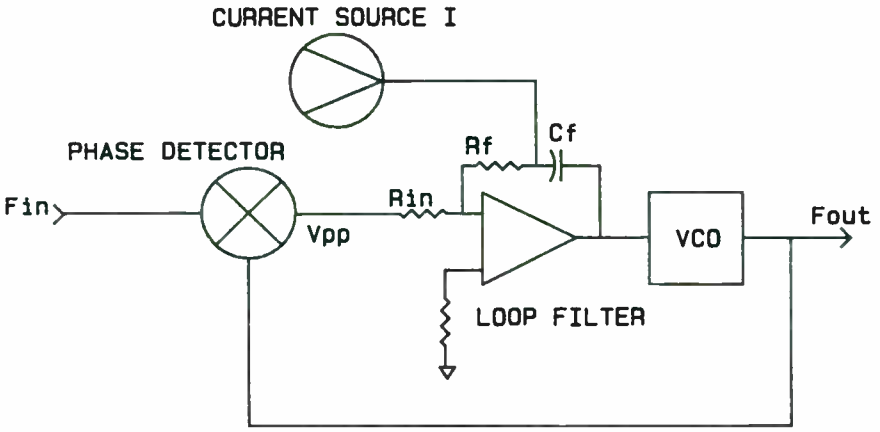
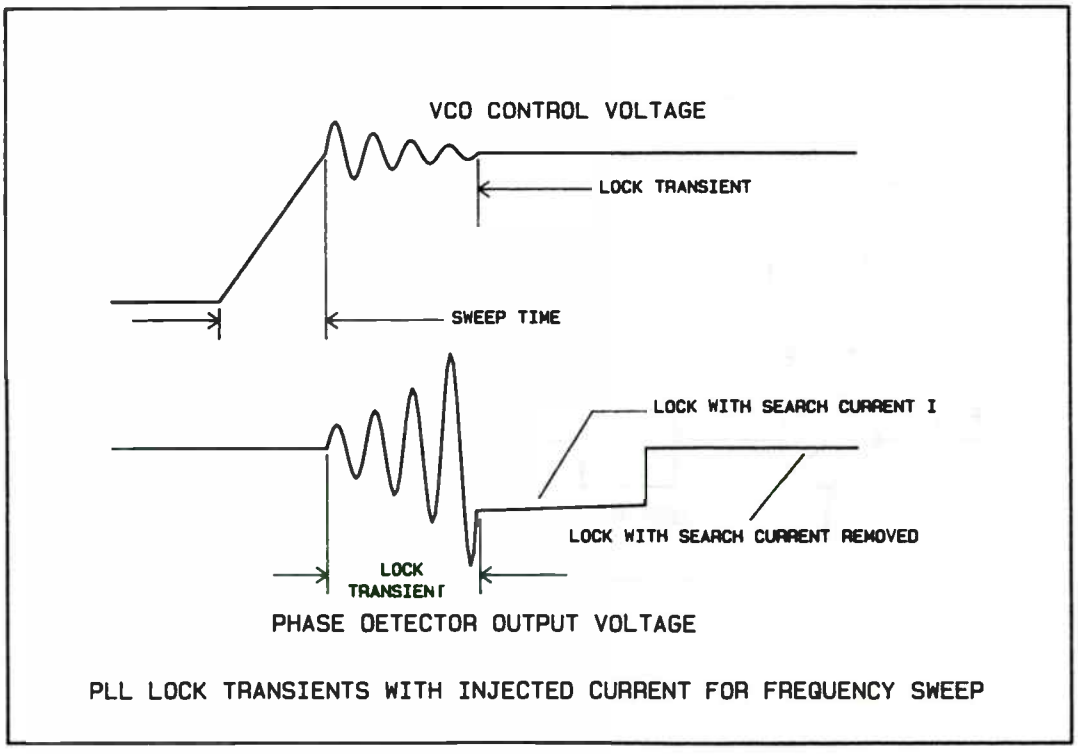


FIGURE 5

172



PLL LOCK TRANSIENTS WITH INJECTED CURRENT FOR FREQUENCY SWEEP

FIGURE 6

PLL WITH MANUAL SWEEP FOR TEST

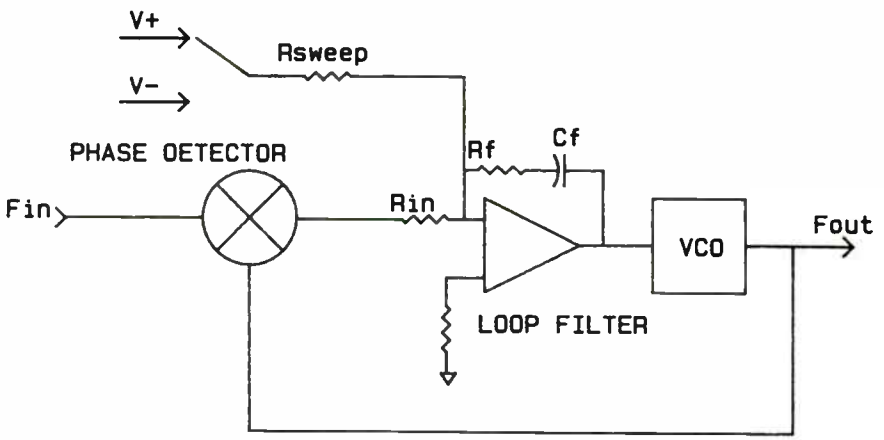


FIGURE 7

PLL WITH FREQUENCY AIDING FOR SEARCH

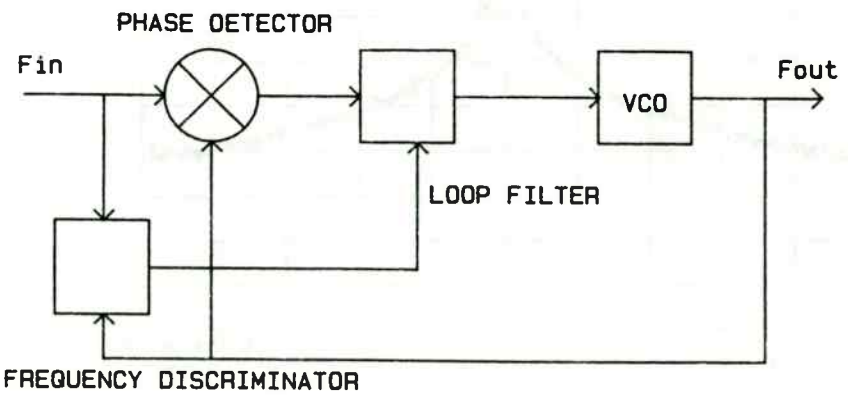


FIGURE 8

VCO OUTPUT SPECTRUM FROM PLL WITH NOMINAL BANDWIDTH

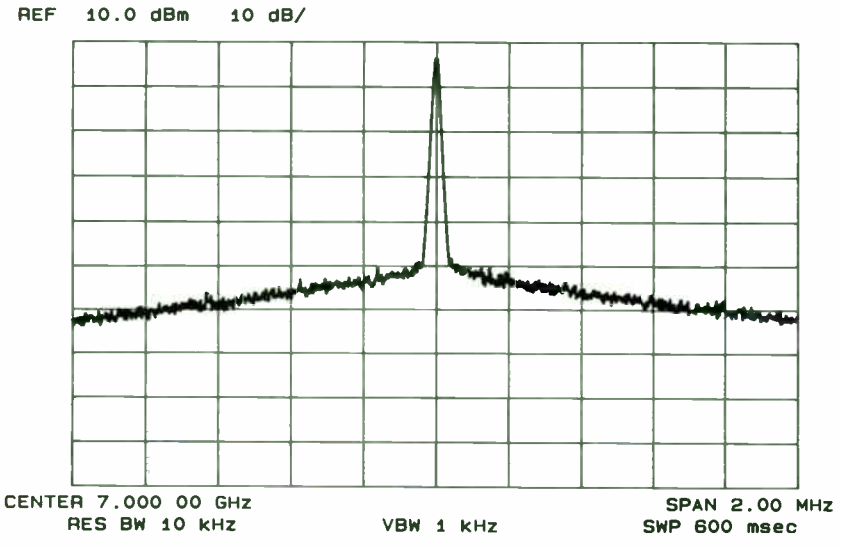


FIGURE 9

VCO OUTPUT SPECTRUM FROM PLL WITH VERY NARROW BANDWIDTH

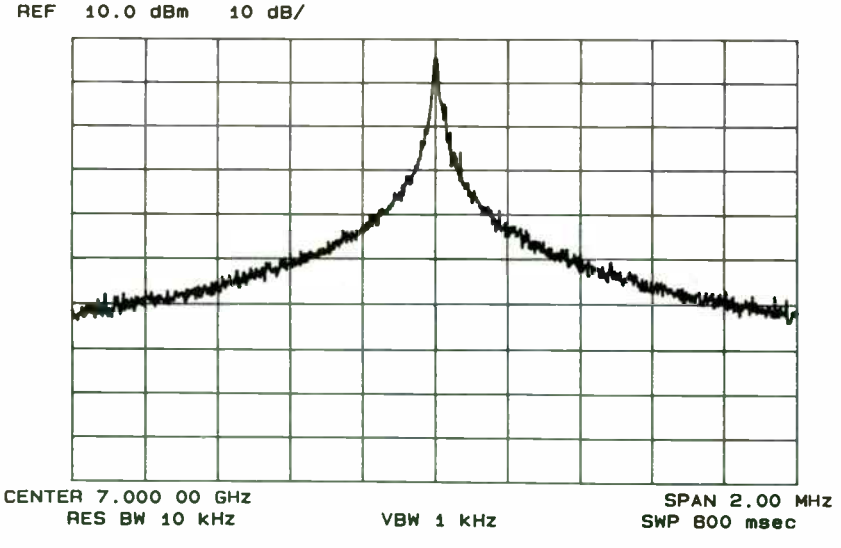


FIGURE 10

VCO OUTPUT SPECTRUM FROM PLL WITH BANDWIDTH SET TOO WIDE

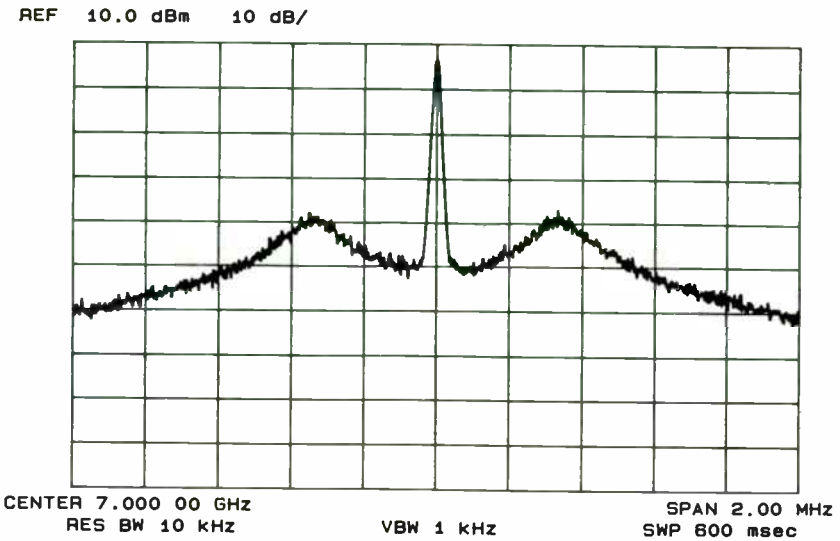


FIGURE 11

PLL TRANSFER FUNCTION BLOCKS

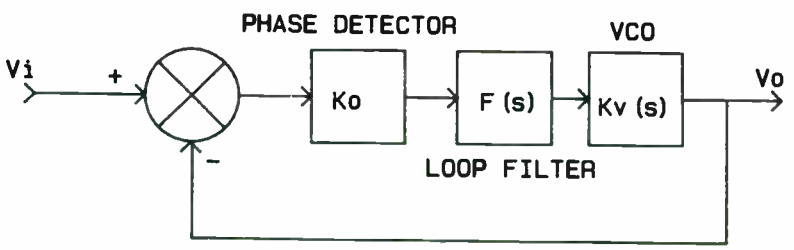


FIGURE 12

PLL TRANSFER FUNCTION BLOCKS FOR BANDWIDTH MEASUREMENT

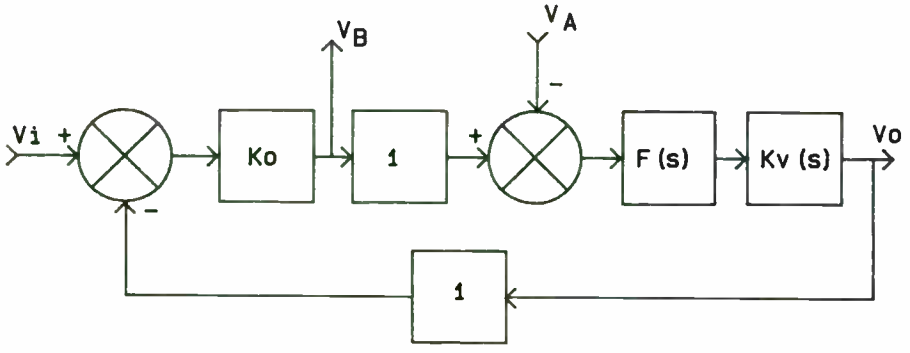


FIGURE 13

TYPE TWO PLL WITH BANDPASS MEASUREMENT CONNECTIONS

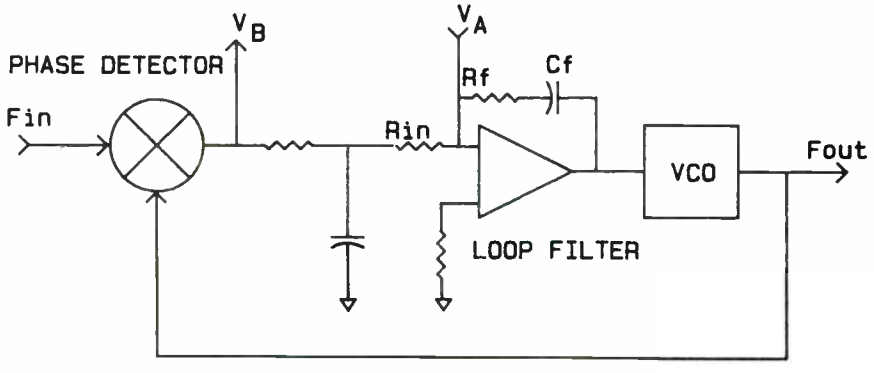


FIGURE 14

PLL CLOSED LOOP NOMINAL BANDPASS

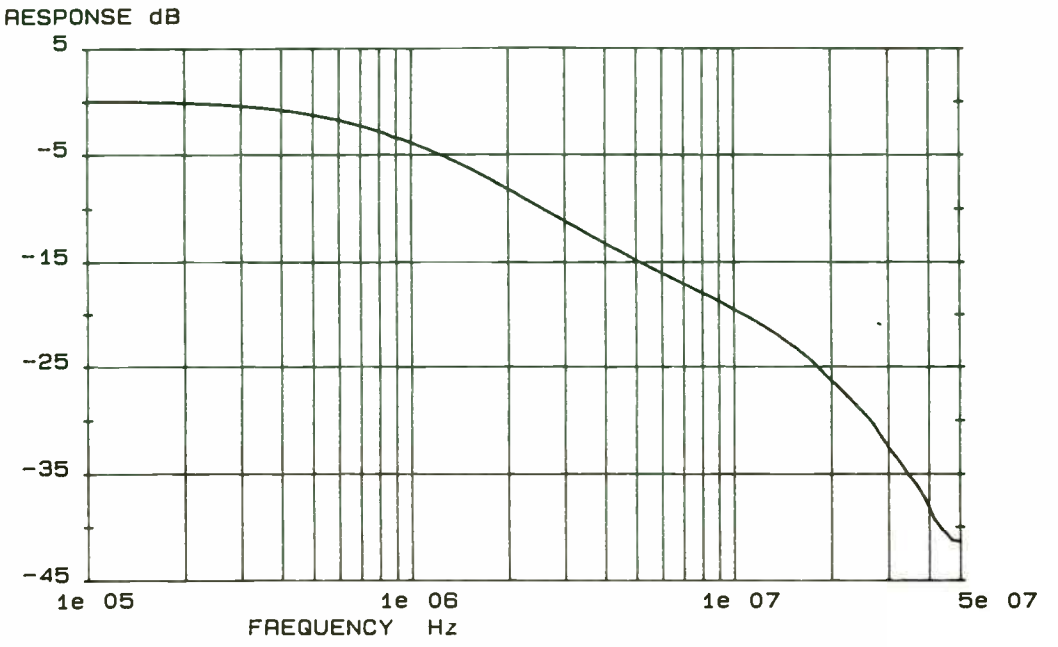


FIGURE 15

PLL CLOSED LOOP BANDPASS FOR VERY NARROW LOOP

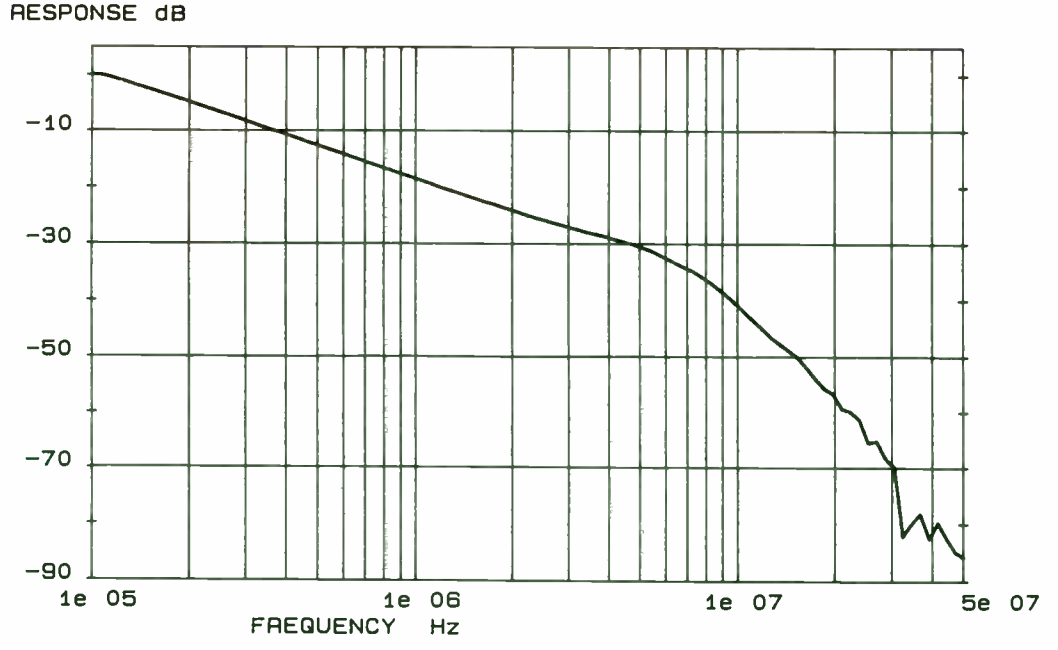


FIGURE 16

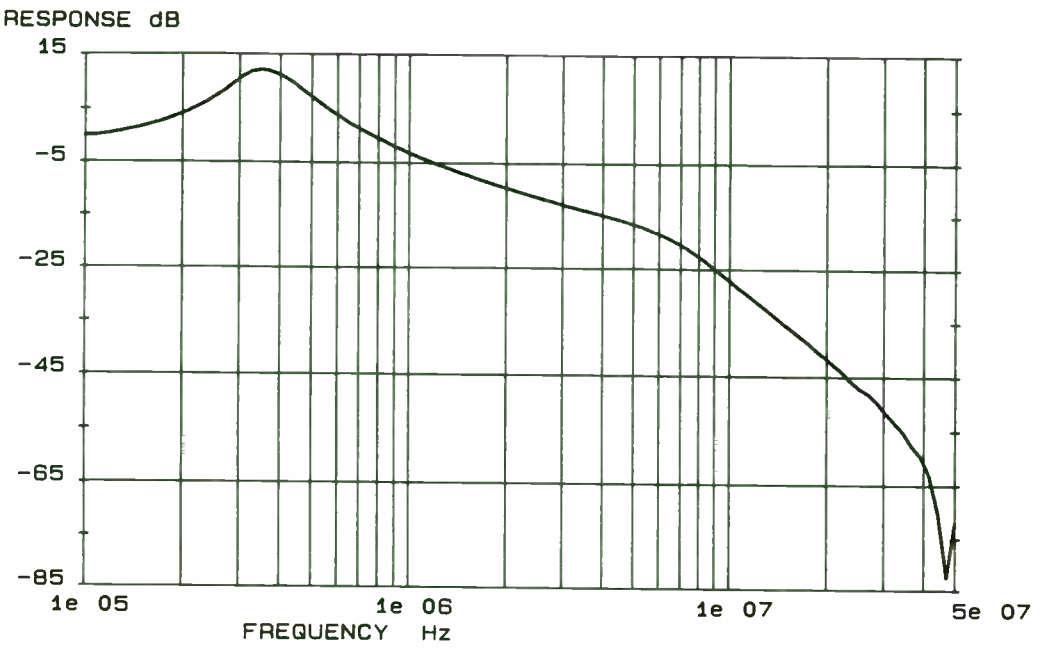


FIGURE 17

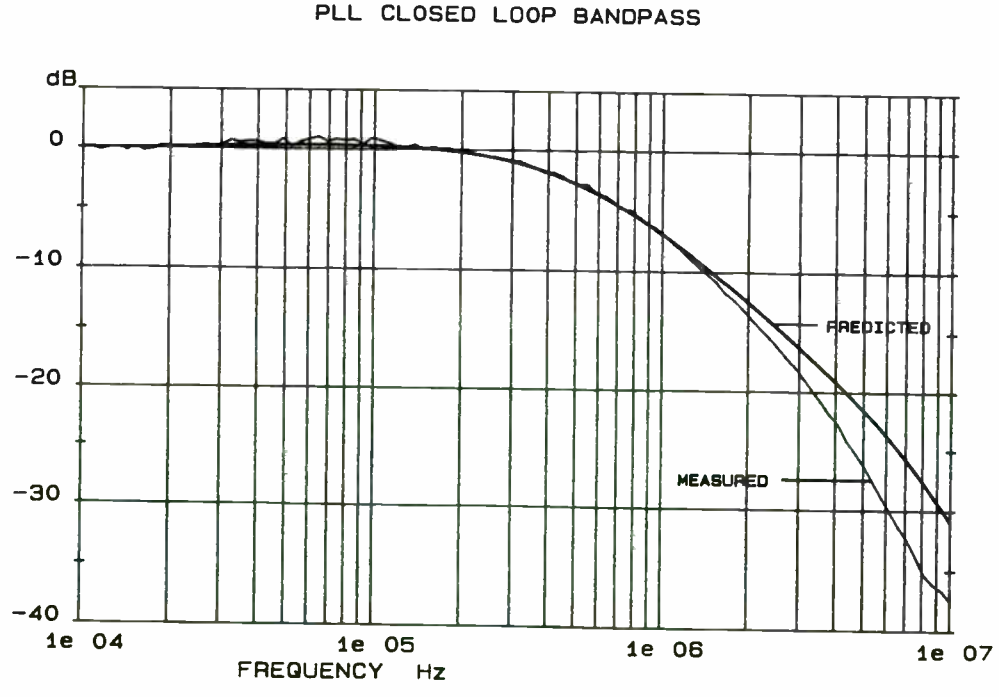


FIGURE 18

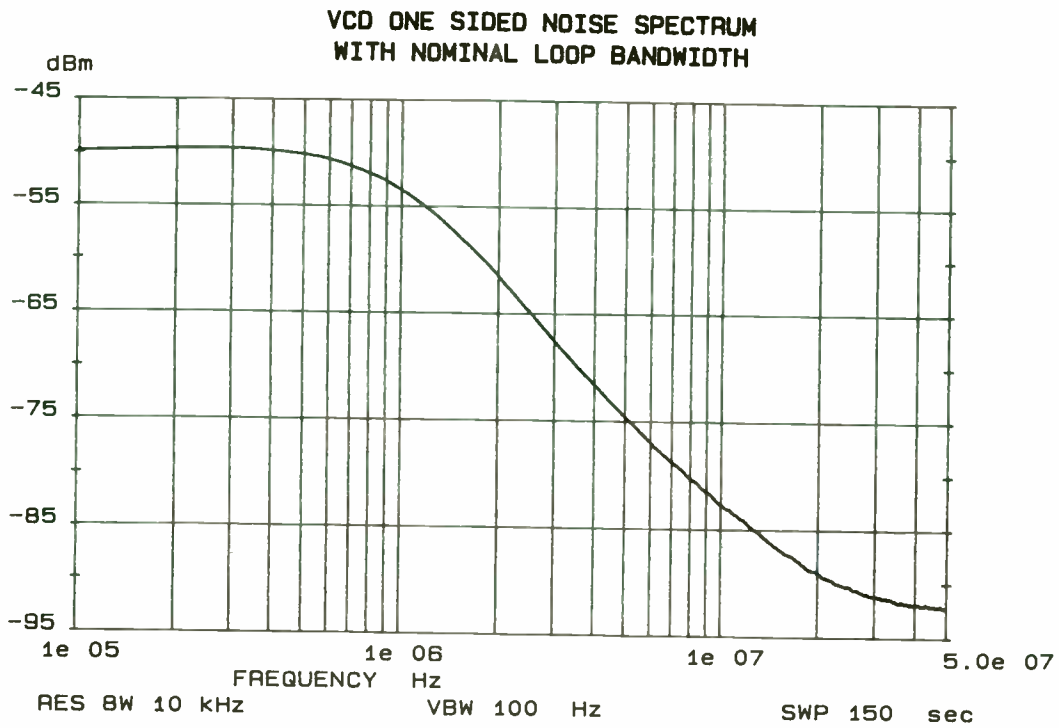


FIGURE 19

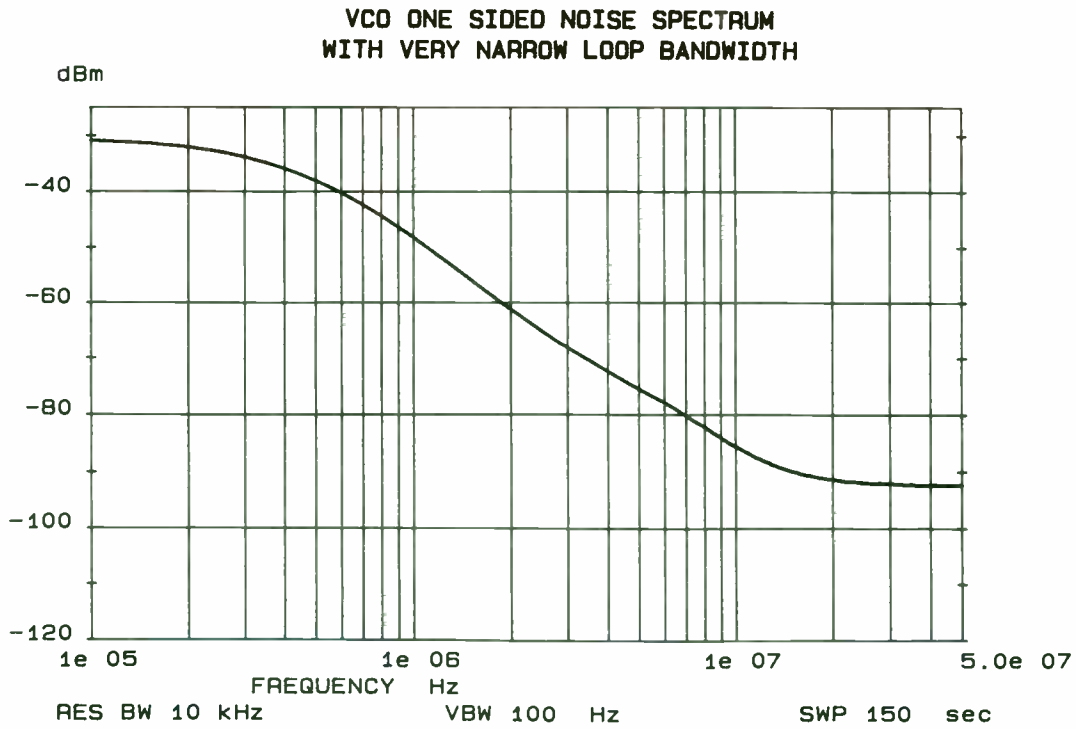


FIGURE 20

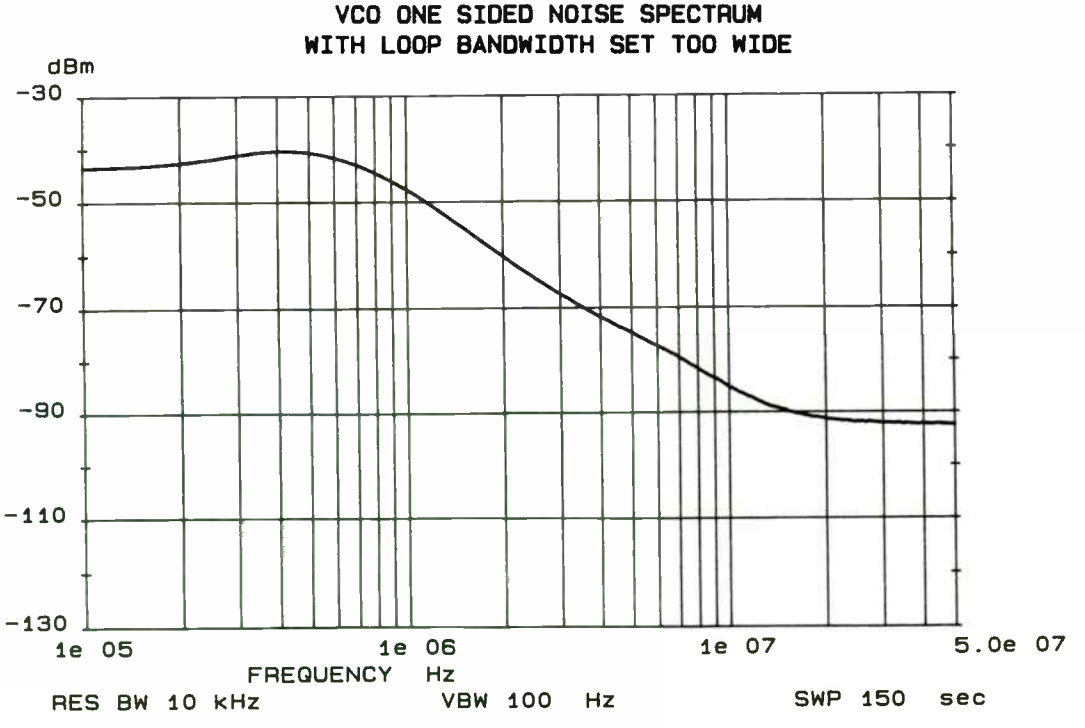


FIGURE 21

DESIGN OF PRACTICAL WIDEBAND TYPE-2, SECOND-ORDER PHASE-LOCKED LOOP FREQUENCY SYNTHESIZERS

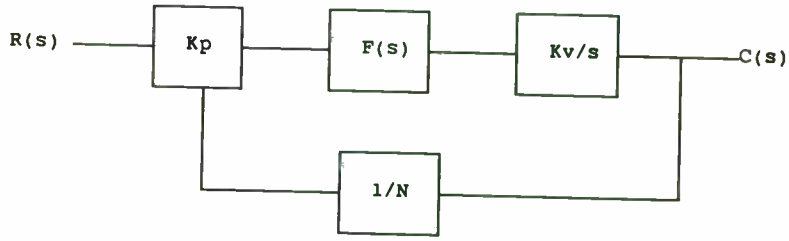
by Jeff Blake
Fairchild Weston Systems Inc.
300 Robbins Lane
Syosset, New York 11791

Although classical, continuous-time analysis is commonly used for phase-locked loop (PLL) frequency synthesizers, it is only applicable for loop bandwidths that are no more than just a few percent of the reference frequency. But for today's high performance, frequency agile synthesizers, requiring much wider bandwidths, this technique is far from adequate. This is due to the discrete-time, or sampling nature of all PLL's that incorporate digital circuitry, such as digital dividers and phase detector. Much has been written to describe what the effect is on loop performance and stability. Some have even attempted to show how to include into the design some sort of "fudge factor". The subject of this paper, however, is not how to design around the PLL's sampling nature, but how to design with it, through exact analysis.

It has been published that "the most important factor when designing a hopping synthesizer is the loop bandwidth." This specific statement will be shown to be false.

Also, time honored concepts such as the damping coefficient, ζ, and the natural frequency, ω_n, will be abandoned for more meaningful, physical design constants such as K, the loop gain, and τ, the loop filter time constant. It is apparent from most literature that K is the popular constant to study. But it will be shown that τ must not be ignored, and in fact, in hopping synthesizers, where the controller must be carefully designed, τ must be deliberately chosen.

To begin, it may be helpful to quickly review classical, continuous-time PLL theory. In Figure 1, we see a common PLL. For this paper, the phase detector used will be a typical "phase/frequency detector", with charge pump, and shunt R-C loop filter. This gives rise to the phase detector and loop filter transfer functions shown in Figure 1.



typically, for a Type-2, second order loop,

$$F(s)=\frac{s\tau+1}{s\tau}$$

Figure 1

Letting the loop gain, K, be defined as

$$K = \frac{Kp \cdot Kv}{N} \quad , \qquad (1)$$

the open loop transfer function is

$$G(s)H(s) = \frac{K}{s} \cdot \frac{s\tau+1}{s\tau} \quad , \qquad (2)$$

and the closed loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{KN(s+1/\tau)}{s^2+Ks+K/\tau} \quad . \qquad (3)$$

The denominator of the closed loop transfer function is known as the characteristic equation. It is the roots of this equation that will determine the stability and performance of the loop. A root locus will enable us to visualize this. It is obvious that the roots are dependant on two parameters, K, the loop gain, and τ, the loop filter time constant. In Figure 2a, the root locus is shown for K being held constant, and τ varying from 0 to ∞. Figure 2b shows the root locus for τ constant, and K is being varied from 0 to ∞. Since the roots are always in the left-hand plane for all K and τ, the system described by these equations is unconditionally stable for all loop bandwidths, and all degrees of damping.

The roots of the characteristic equation are given by:

$$\frac{-K}{2} \pm \sqrt{\left[\frac{K}{2}\right]^2 - \frac{K}{\tau}} \quad , \qquad (4)$$

and, in general, are complex. They become real when

$$\tau = 4/K \quad . \qquad (5)$$

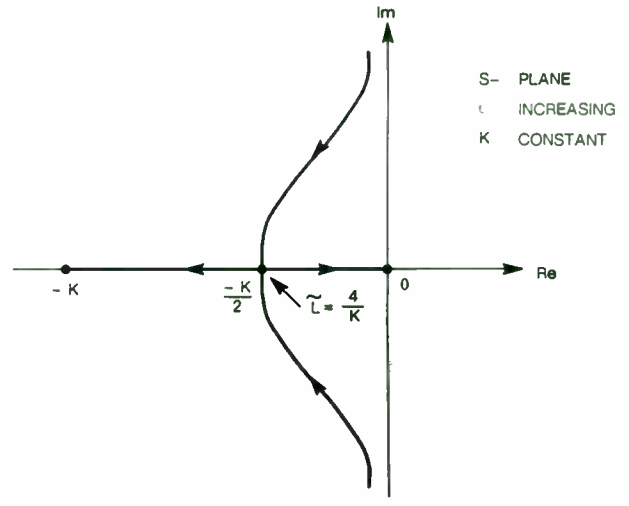


FIGURE 2A

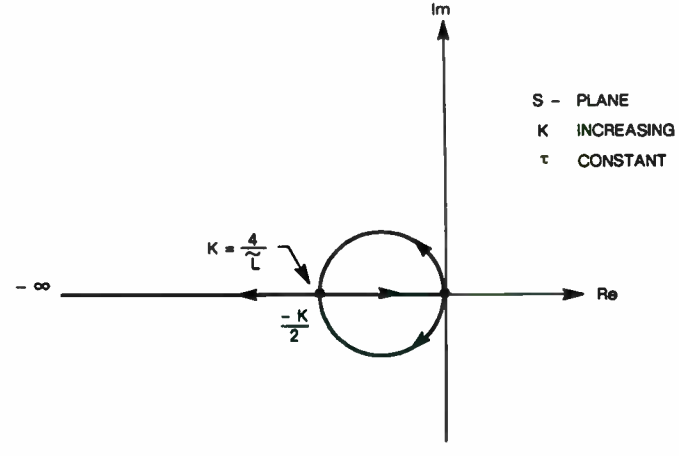


FIGURE 2B

This is the boundary between the underdamped and overdamped step responses, with the boundary itself yielding a critically damped response. In the root locus plots, this boundary is the point of convergence. When designing a hopping synthesizer, a critically damped response is chosen since is yields the fastest response time.

Not "Divider Delay"...

When constructing a PLL, a continuous system is usually never built. Digital dividers are commonly incorporated within the feedback path, interfacing with a digital phase detector. Since information exists only at the logic transitions, our PLL is clearly no longer a continuous system, but a discrete-time system.

This discontinuity has been addressed for some time (References 1, 2, 3). Many attempts have been made to describe this discrete-time behavior by introducing a pure time delay term equal to one reference period (References 4, 5, 6, 7). Commonly it has been introduced into the feedback path, and called "Divider Delay", or "Transportation Lag". This model, as shown in Figure 3, is not the best one to consider.

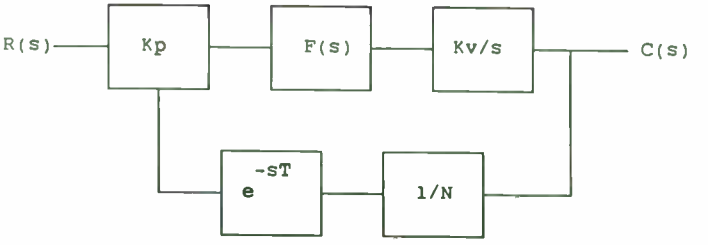


FIGURE 3

The error arises by modeling the discontinuity as an actual time delay: the same delay that would be present in a delay line. Modeling the discontinuity as such was the result of the fact that the logic transitions are spaced by one reference period, and thereby stating that there is a time delay equal to that amount. If one were to persue this line of thinking more, it might be apparent that this would certainly represent worst case, and perhaps the delay term should be less severe. Appendix A develops an equivalent delay term from the exact analysis in this paper, by using a narrowband approximation, demonstrating a factor of one half. This result is also independantly developed in References 8, 11, and 12. The more precise method is then to accept this discontinuity as a phenomenon due to the sampling nature of the loop, and apply an appropriate model.

...But Sampling

To arrive at the appropriate model, it will be helpful to examine the controller of the system, and the mechanism by which the VCO is controlled by it. The two inputs to the digital phase detector, namely the output of the digital divider in the feedback path, and the reference frequency, control the DC voltage applied to the VCO. More specifically, a capacitor is charged to the necessary voltage, and held there. Translation of logic signals into an analog voltage such as this is the behavior of a zero-order hold (ZOH). Figure 4 shows the model of the system to be studied. This model is definitely not new (References 2,3), neither is this model restricted to explicit sample-and-hold phase detectors.

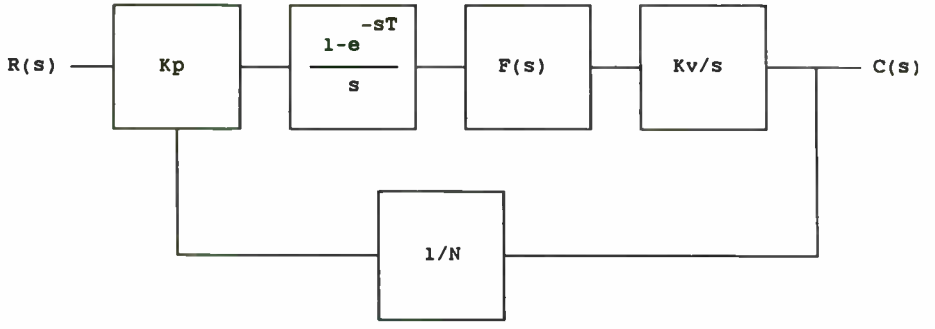


FIGURE 4

For this model, the open loop transfer function is:

$$G(s)\,H(s) = \frac{KpKv}{N}\ \frac{s\tau + 1}{s^2\tau}\ \frac{1 - e^{-sT}}{s} \qquad (6)$$

Since this new exponential term certainly can complicate the analysis, z transforms will be used to vastly simplify things. (This form certainly can be used for computer programs, if one is careful of the units! See Appendix A for a discussion, and Appendices C and D for applications.)

Rearranging, the open-loop transfer function may be rewritten as:

$$G(s)\,H(s) = \frac{K}{\tau}\left[1-e^{-sT}\right]\left[\frac{1}{s^3} + \frac{\tau}{s^2}\right] \qquad (7)$$

Substituting the following transformations,

$$1 - e^{-sT} \leftrightarrow \frac{z - 1}{z}, \qquad \frac{1}{s^3} \leftrightarrow \frac{T^2}{2}\ \frac{z(z + 1)}{(z - 1)^3}, \quad \text{and}$$

$$\frac{1}{s^2} \leftrightarrow T\ \frac{z}{(z - 1)^2},$$

the new open-loop transfer function is therefore:

$$G(z)\,H(z) = \frac{KT}{\tau}\ \frac{\frac{T}{2}(z + 1) + \tau(z - 1)}{(z-1)^2} \qquad (8)$$

(which is in agreement with eq. 5 in Ref. 2 and eq. 30 in Ref.10)

Defining the characteristic equation as

$$1 + G(z)\,H(z)$$

leads to

$$1 + G(z)\,H(z) = \frac{z^2 + \left[\frac{KT}{\tau}\left[\frac{T}{2} + \tau\right] - 2\right] z + \left[K\frac{T}{\tau}\left[\frac{T}{2} - \tau\right] + 1\right]}{(z-1)^2} \qquad (9)$$

Stability (Reference 2)

Using both the open-loop transfer function, and the characteristic equation, a study of the system's stability can be made. In the continuous-time case, for stability, roots of the characteristic equation had to lie in the left-hand complex plane. Now, due to the z-transformation, the left-half plane has been mapped to the interior of a unit circle. Therefore, a stable system, described by these equations in z, must have its roots within a unit circle.

Figure 5a depicts a typical root locus for a constant K with τ the variable, and Figure 5b is a typical root locus for τ being held constant, and K the variable. In Figure 5b, the trajectory of the poles from K = 0 to the convergence point is circular, and whose center is at the open loop zero. Clearly, if the roots along the trajectory are to remain within the unit circle, the center must remain > 0. The center is determined by setting the numerator of the open loop transfer function to zero.

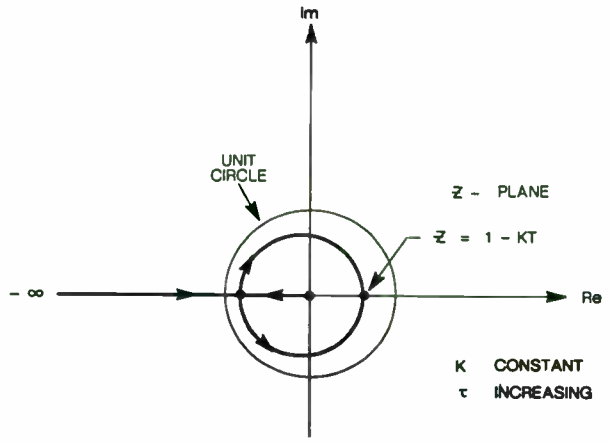


FIGURE 5A

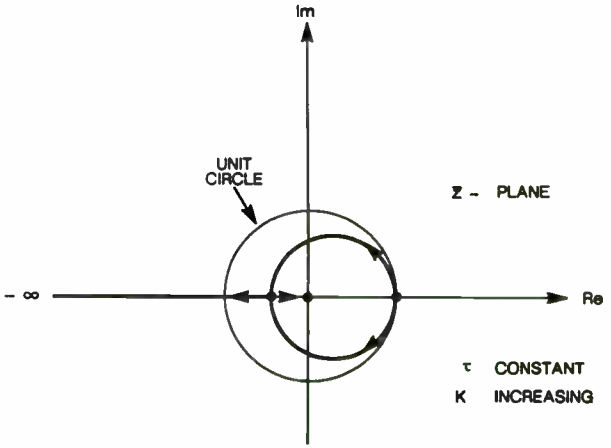


FIGURE 5B

namely,

$$\frac{T}{2}(z + 1) + \tau(z - 1) = 0 \qquad (10)$$

solving for z, we find that

$$z = \frac{2\tau - T}{2\tau + T} \quad . \qquad (11)$$

Clearly, for z to be > 0, it is necessary that

$$\frac{\tau}{T} > \frac{1}{2} \quad . \qquad (12)$$

This determines the bounds for the "damping" of the loop to maintain stability.

To determine the bounds for the loop gain, K, observe the behavior of the root locus in Figure 5b. As K increases, a root will eventually leave the unit circle at z = -1. Substituting z = -1 into the characteristic equation, and solving for K, it can be shown that for stability, it is necessary that

$$KT < 2 \quad . \qquad (13)$$

To summarize, the continuous-time equations demonstrated unconditional stability with respect to K and τ, whereas the more precise discrete-time approach yields a conditionally stable system. The bandwidth of a practical loop cannot be increased without bound, nor can the damping be arbitrarily chosen.

Step Response

Since we are interested in frequency-agile synthesizers, we will now turn our attention to the response of the system to a frequency step at the input. The output frequency of a loop is changed by either stepping the reference frequency, or by changing the divide-by-N number. Egan (Reference 12) has shown that both of these stimuli are analogous.

Choosing to look at the Error transfer function,

$$\frac{E(z)}{R(z)} = \frac{1}{1 + G(z)\,H(z)} \quad , \qquad (14)$$

we get

$$\frac{E(z)}{R(z)} = \frac{(z - 1)^2}{z^2 + \left[K\frac{T}{\tau}\left[\frac{T}{2} + \tau\right] - 2\right]z + \left[K\frac{T}{\tau}\left[\frac{T}{2} - \tau\right] + 1\right]} \quad . \qquad (15)$$

Substituting the z-transform of a frequency step,

$$R(z) = \frac{\Delta\omega\, T}{N}\,\frac{z}{(z - 1)^2} \quad , \qquad (16)$$

into the above equation, yields:

$$E(z) = \frac{\Delta\omega\, T}{N}\,\frac{z}{z^2 + \left[K\frac{T}{\tau}\left[\frac{T}{2} + \tau\right] - 2\right]z + \left[K\frac{T}{\tau}\left[\frac{T}{2} - \tau\right] + 1\right]} \quad . \qquad (17)$$

Taking the inverse z-transform to get the time response, we find that, in general,

$$e(nT) = \frac{\Delta\omega\ T}{N}\ \frac{a^n - b^n}{a - b}\ , \qquad (18)$$

where a and b are the roots of the characteristic equation, namely

$$a = A + jB,\ b = C + jD. \qquad (19)$$

Just as in the continuous-time analysis, the response may be classified into three different categories depending upon the roots of the characteristic equation.

Complex Conjugate Roots: Underdamped Response

Choosing τ such that the roots of the characteristic equation are complex, the response due to a frequency step is underdamped, being a dampened sinusoid in nature. Since, for complex conjugate roots, the real parts are equal and the imaginary parts are opposite in sign, the roots may be rewritten as follows:

$$\begin{aligned} a &= A + jB \\ b &= A - jB \quad . \end{aligned} \qquad (20)$$

Substituting back into (18), we find that

$$e(nT) = \frac{\Delta\omega\ T}{N}\ \frac{R^n}{B}\ \sin(n\theta)\ , \qquad (21)$$

where $$R = \sqrt{A^2 + B^2}\ ,\quad \theta = \arctan\left(\frac{B}{A}\right)\ . \qquad (22)$$

A typical response is shown in Figure 6a.

Real, Equal Roots: Critical Damping

Since critical damping yields the fastest settling time for a given loop bandwidth, this is the most important case. The roots have equal real parts, and zero imaginary parts, namely

$$a = b = A\ . \qquad (23)$$

Substituting back into (18) yields

$$e(nT) = \frac{\Delta\omega\ T}{N}\ nA^{n-1}\ . \qquad (24)$$

Figure 6b depicts a typical response.

FIGURE 6 TYPICAL STEP RESPONSES

Real, Unequal Roots: Overdamped Response

Increasing the damping further, the response becomes more sluggish. The real parts of the roots are now unequal, with zero imaginary parts. The roots may be rewritten as

$$a = A \qquad (25)$$
$$b = C,$$

which yields

$$e(nT) = \frac{\Delta\omega\, T}{N} \; \frac{A^n - C^n}{A - C} \; . \qquad (26)$$

Figure 6c shows a typical response.

Optimum Design

Since achieving critical damping is necessary to minimize settling time, an equation yielding the necessary τ for a given K will now be generated. Critical damping occurs when the roots of the characteristic equation just become real (point of convergence in the root locus of Figure 5b). Solving for the roots of the characteristic equation, we find that

$$a,b = \frac{-\left[K\frac{T}{\tau}\left[\frac{T}{2}+\tau\right]-2\right]}{2} \pm \sqrt{\left[\frac{K\frac{T}{\tau}\left[\frac{T}{2}+\tau\right]-2}{2}\right]^2 - \left[K\frac{T}{\tau}\left[\frac{T}{2}+\tau\right]+1\right]} \qquad (27)$$

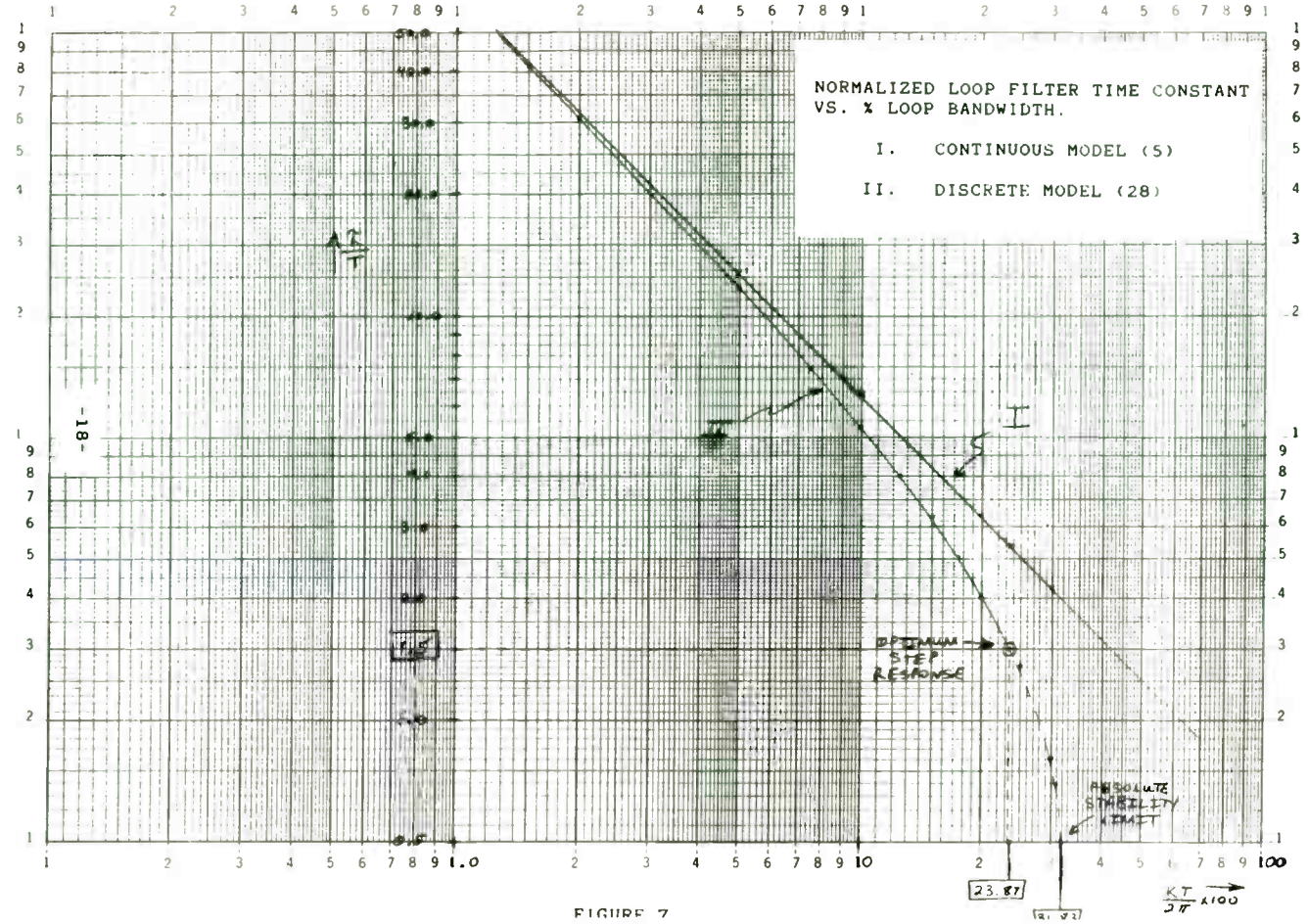


FIGURE 7

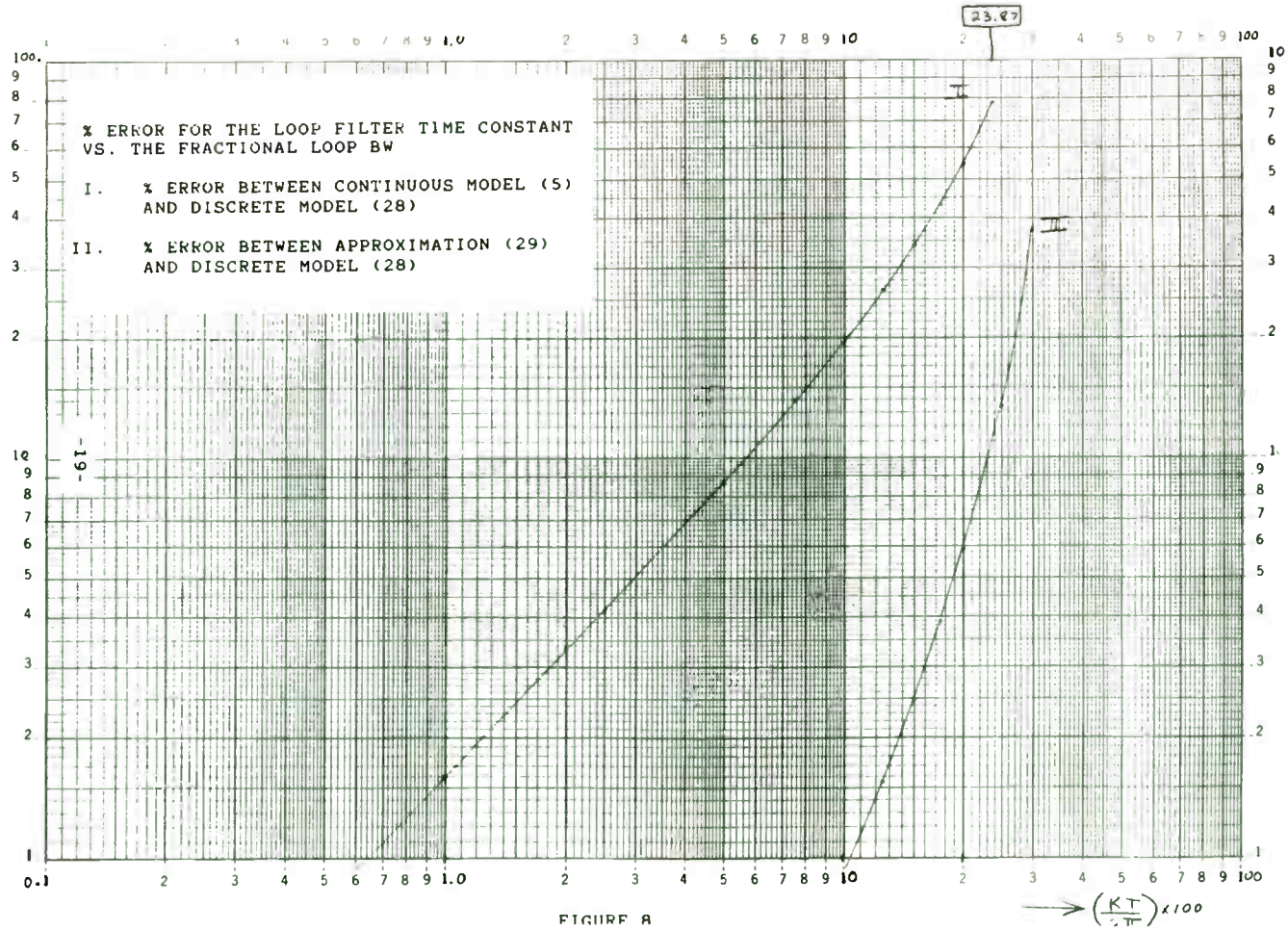


FIGURE 8

Clearly the roots become real and equal when the radicand (commonly known as the discriminant) is equal to 0.

Setting the radicand equal to 0, and solving for τ, we find, therefore, for critical damping,

$$\tau = \left[\frac{2}{K} - \frac{T}{2}\right] + \sqrt{\left[\frac{2}{K} - \frac{T}{2}\right]^2 - \left[\frac{T}{2}\right]^2} \quad . \tag{28}$$

Appendix B shows that for small bandwidth loops (relative to the reference frequency), the above equation can be approximated by

$$\tau = \frac{4}{K} - T \quad . \tag{29}$$

This is quite similar to the continuous time equation except for the reference period term.

Figure 7 is a graph of the normalized loop filter time constant (τ/T) necessary for critical damping, versus the normalized loop gain ($KT/2\pi$), for both the continuous time system (5), and for the discrete-time system (28). Figure 8 shows the percent error expected if τ is chosen using the continuous-time equations, and if τ is chosen using the approximation (29). As can be seen, for a 10% loop bandwidth, the continuous-time equation is about 20% in error from the correct equation for a critical damping time constant. Whereas the approximation (29) is still excellent, yielding less than 1% error.

If the loop bandwidth were to be increased, and the loop filter time constant set to give critical damping in accordance with (28), an absolute minimum in the settling time would eventually be reached. This results when the error response of (15) reverts to a mere delta function (shifted by one reference period). This occurs when the two non-unity coefficients in (15) are zero, namely

$$\frac{KT}{\tau}\left[\frac{T}{2} + \tau\right] - 2 = 0 \quad \text{and} \quad \frac{KT}{\tau}\left[\frac{T}{2} + \tau\right] + 1 = 0 \quad . \tag{30}$$

Solving simultaneously results in

$$\frac{\tau}{T} = 1.5, \quad \text{and} \quad KT = 1.5 \quad . \tag{31}$$

This then gives rise to

$$E(z) = \frac{\Delta\omega\, T}{N}\,\frac{1}{z} \quad , \tag{32}$$

whose inverse z-transform is:

$$e(nT) = \frac{\Delta\omega\, T}{N}\,\delta(t - T) \quad . \tag{33}$$

This demonstrates that a type-2, second order system theoretically may achieve steady state phase settling in 2 reference periods.

Appendix A: Narrowband models for the zero-order-hold

Two narrowband ($sT << 1$) models for the zero-or-hold term,

$$\frac{1 - e^{-sT}}{s} , \tag{A1}$$

will be developed. Before utilizing this term directly (as opposed to using z-transforms), two considerations must be addressed. First, the term must be dimensionless, and secondly, it must have a magnitude of one, at an angle of zero degrees, at DC. Analyzing the zero-order-hold term determines that it has dimensions of seconds, and a magnitude of T, at an angle of zero degrees, at DC. Clearly, dividing by T will render the term dimensionless, with a magnitude of one, at an angle of zero degrees, at DC.

Therefore the term to study is:

$$\frac{1 - e^{-sT}}{sT} . \tag{A2}$$

Substituting the first three terms of the well known expansion,

$$e^{-x} \approx 1 - x + \frac{x^2}{2} - \ldots \tag{A3}$$

into (A2) results in

$$\frac{1 - e^{-sT}}{sT} \approx 1 - \frac{sT}{2} . \tag{A4}$$

The first model results from using the expansion for e^{-x} again, namely

$$1 - \frac{x}{2} \approx e^{\frac{-x}{2}} . \tag{A5}$$

This gives rise to the delay-like model for the zero-order-hold:

$$\frac{1 - e^{-sT}}{sT} \approx e^{\frac{-sT}{2}} . \tag{A6}$$

The second model is generated by using the approximation

$$1 - x \approx \frac{1}{1 + x} , \text{ for } x << 1 . \tag{A7}$$

Applying this approximation to (A4) results in the single-pole model for the zero-order-hold:

$$\frac{1 - e^{-sT}}{sT} \approx \frac{1}{1 + \frac{sT}{2}} . \tag{A8}$$

These two models, (A6) and (A8), may be used as desired to approximate the effect of the discrete-time nature of the loop when doing narrowband continuous-time analysis.

Appendix B: Narrowband approximation for the critical damping time constant equation

If the equation to determine the loop filter time constant that yields critical damping for a given loop gain, (28), is too intimidating, or if its accuracy is not necessary, a simple approximation can be made to significantly reduce it. Rewriting (28) here:

$$\tau = \left[\frac{2}{K} - \frac{T}{2}\right] + \sqrt{\left[\frac{2}{K} - \frac{T}{2}\right]^2 - \left[\frac{T}{2}\right]^2} \qquad \text{(B1)}$$

Rearranging inside the radical yields,

$$\tau = \left[\frac{2}{K} - \frac{T}{2}\right] + \frac{2}{K}\sqrt{1 - \left[K\cdot\frac{T}{2}\right]} \qquad \text{(B2)}$$

Now for small loop bandwidths, relative to the reference frequency, namely KT << 1, the approximation,

$$\sqrt{1 - x} \approx 1 - \frac{x}{2}\ , \text{ for } x \ll 1 \qquad \text{(B3)}$$

is used to achieve the simple equation:

$$\tau = \frac{4}{K} - T\ . \qquad \text{(B4)}$$

Appendix C: Demonstration that K is approximately equal to the loop bandwidth

It will be shown that for a given loop gain, K, and with τ chosen for critical damping in accordance with (28), the loop bandwidth is approximatley equal to K. The loop bandwidth is defined as the frequency at which the magnitude of the open loop transfer function (6) goes to 0 dB. Rewriting (6) here:

$$G(s)\,H(s) = K\ \frac{s\tau + 1}{s\tau}\ \frac{1 - e^{-sT}}{sT}\ . \qquad \text{(C1)}$$

(The T in the denominator resulted from the discussion in Appendix A.)

Taking the magnitude of (C1) and setting it equal to 1 (0 dB) results in:

$$\frac{K}{\omega^3 T\tau}\ \sqrt{2\left[1 + (\omega\tau)^2\right](1 - \cos(\omega\tau))} = 1$$

Solving for ω for a given K, with τ chosen in accordance with (28), the following table results:

predicted % loop BW	actual % loop BW	
100 (KT)	100 (ωT)	% error
0.25	0.257	2.93
0.50	0.515	2.95
1.00	1.03	2.98
2.00	2.06	3.03
2.50	2.58	3.03
5.00	5.15	2.97
10.00	10.24	2.39
12.50	12.74	1.92
15.00	15.20	1.37
17.50	17.64	0.78
20.00	20.04	0.22

This demonstrates that the 0 dB crossover frequency may be approximated, to within a few percent, by K.

Appendix D: Maximize Phase Margin

(proceed with caution)

When settling time is not a primary consideration, τ may be chosen such that the phase margin of the system is maximized for a given K. This is accomplished by taking the angle of the open loop transfer function, differientating, setting it equal to zero, and solving for τ.

Rewriting the open loop transfer function (6), and incorporating the factor of T due to the discussion of Appendix A, we get

$$G(s)H(s) = K \frac{s\tau + 1}{s\tau} \frac{1 - e^{-sT}}{sT} \qquad (D1)$$

Substituting Euler's formula,

$$e^{jx} = \cos(x) + j \sin(x),$$

into (D1), and taking the angle yields:

$$\theta = \arctan(\omega\tau) + \arctan\left[\frac{\sin(\omega T)}{1 - \cos(\omega T)}\right] - 270^\circ \qquad (D2)$$

Differentiating with respect to ω gives rise to:

$$\frac{d\theta}{d\omega} = \frac{\tau}{1 + (\omega\tau)^2} - \frac{T}{2} \; . \qquad (D3)$$

Setting (D3) equal to zero, and solving for τ, we find that for a given K, maximum phase margin will occur when

$$\frac{\tau}{T} = \frac{1}{(\omega T)^2}\left[1 + \sqrt{1 - (\omega T)^2}\right] \; . \qquad (D4)$$

The admonishment to use caution when designing for maximum phase margin, is because the resulting loop time constant may be significantly removed from the critical damping value. Table D1 demonstrates this by comparing the time constants for critical damping (28), and for maximum phase margin (D4). As can be seen, at a percentage loop bandwidth of about 8.5% (KT=.53), the two values coincide.

		τ/T		
		Critical Damping		max pm
% loop bw	KT	eq.(5)	eq.(28)	eq.(D4)
1.0	0.06	63.66	62.66	506.11
1.5	0.09	42.44	41.44	224.66
2.0	0.13	31.83	30.82	126.15
2.5	0.16	25.46	24.45	80.55
3.0	0.19	21.22	20.21	55.79
3.5	0.22	18.19	17.17	40.85
4.0	0.25	15.92	14.90	31.15
4.5	0.28	14.15	13.13	24.51
5.0	0.31	12.73	11.71	19.75
5.5	0.35	11.57	10.55	16.23
6.0	0.38	10.61	9.58	13.55
6.5	0.41	9.79	8.77	11.47
7.0	0.44	9.09	8.06	9.81
7.5	0.47	8.49	7.45	8.47
8.0	0.50	7.96	6.92	7.38
8.5	0.53	7.49	6.45	6.47
9.0	0.57	7.07	6.03	5.71
9.5	0.60	6.70	5.66	5.06
10.0	0.63	6.37	5.32	4.50
10.5	0.66	6.06	5.01	4.02
11.0	0.69	5.79	4.73	3.61
11.5	0.72	5.54	4.48	3.24
12.0	0.75	5.31	4.25	2.91
12.5	0.79	5.09	4.03	2.62
13.0	0.82	4.90	3.83	2.36
13.5	0.85	4.72	3.65	2.13
14.0	0.88	4.55	3.48	1.91
14.5	0.91	4.39	3.32	1.70
15.0	0.94	4.24	3.17	1.50
15.5	0.97	4.11	3.02	1.29
15.9	1.00	4.00	2.91	1.00

Table D1

References

1. Gardner, Floyd M., "Charge-Pump Phase-Lock Loops", IEEE Transactions on Communications, Vol. COM-28, No. 11, November 1980
2. Barab, Stuart and McBride, Alan L., "Uniform Sampling Analysis of a Hybrid Phase-Locked Loop...", IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-11, No. 2, March 1975, pages 210-216
3. Eisenberg, Barry R., "Gated Phase-Locked Loop Study", IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-7, No.3, May 1971, pages 469-477
4. Wetenkamp, Scott F., et. al., "Transportation Lag in PLL's", Watkins Johnson Tech-notes, Vol.5, No.3, May/June 1978
5. Przedpelski, Andrzej B., "PLL Primer, Part III", RF Design, July/Aug 1983
6. Przedpelski, Andrzej B., Letters to the Editor, RF Design, Sept/Oct 1983, Page 6
7. Goldman, Stan, "Divider Delay: The Missing PLL Analysis Ingredient", RF Design, Mar/Apr 1984
8. Egan, W. F., Letters to the Editor, RF Design, Mar/Apr 1984, pages 9A,10A
9. Crawford, James A., "Understanding the Specifics of Sampling in Synthesis", Microwaves + RF, Aug 1984
10. Crawford, James A., "Extending Sampling to Type II Phase-Locked Loops", Microwaves + RF, Sept 1984
11. Crawford, James A., "The Phase/equency Detector", RF Design, Feb 1985
12. Egan, W.F., Frequency Synthesis by Phase Lock, section 3.5, John Wiley & Sons, 1981

Supercomputer-aided Designs for Viable Near- future RF System

by

DR. PERAMBUR S. NEELAKANTA
ASSOCIATE PROFESSOR
DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF SOUTH ALABAMA
MOBILE, AL. 36688, USA
TEL: 205/460-6117

Large-scale computational electromagnetics and the associated algorithms as required by the near-future, complex RF systems employing CAD strategies, indicate the need for supercomputer utilization with appropriate codes / softwares. This paper elaborates the specific areas in which Supercomputer-aided Design (S/CAD) approach is imminent in RF system development. Example(s) of RF design compatible with S/CAD algorithms are presented.

INTRODUCTION

Current trend in RF technology is to use high-level computational electromagnetics to match the state-of-the-art RF systems. Such design strategies broadly enclave large-structured, conventional transmission-lines, radiation components, and the passive/active elements. However, extension of RF design for near-future applications should address miniaturized structures such as high-density package interconnect lines, monolithic integrated circuits and millimeter/submillimeter wave systems. Forecasts on the extendability of such RF designs further suggest the possible inclusion of guided-wave structures of ULSI and/or WSI implementation as well as the superfast networks of gigabit logic systems. CAD application to these anticipated systems will requre algorithms of ultracomplexity with computational need of large-memory and fast-processing capabilities -- only to be met by a supercomputer.

Some basic RF structures which may need S/CAD compatibility are:

- Millimeter & submillimeter wave integrated circuits (MMIC).
- Monolithic phased arrays, signal-processing arrays & conformal antennas.
- High-speed interconnections in high-density packages.
- Multilayered microstrip structures.
- Electromagnetically active exotic surfaces.
- On-chip microwave metrology etc.

The important aspects of CAD-based RF design are modeling, analysis and optimization. Existing computer-aided RF designs are based essentially on quasistatic analysis. The extended technology as indicated above will however, require more rigorous analytical techniques such as FFT, spectral-domain approach, method of moments, double- variation techniques and finite-element/finite-difference time-domain solutions. Compatible algorithms will therefore, be of large-scale and require computational efficiency matching the vector processing approach available in supercomputers. This is because, repeated similarly-structured computational steps are involved in the relevant analyses and optimization algorithms.

Supercomputer- implementation of computational steps in parallel (in pipeline sense) by vector processors would make design optimization significantly more efficient and is likely to reduce the CAD costs. To understand the significance of S/CAD strategies, the fundamentals of supercomputation can be reviewed and the salient features are explained below:

SUPERCOMPUTATION FUNDAMENTALS

The von Neumann, or serial architecture employed in most computers represents a fundamental design bottleneck for achieving significantly greater speeds. This problem is however, overcome in the supercomputers by employing parallel architectures, yielding execution time several orders of magnitude greater than the conventional computers [1-4].

The architecture of the first generation computers are described as von Neumann organization which featured the stored program computer. It is defined by, 1) an input/output device; 2) a single memory for strorage data and instructions; 3) a single control unit for the interpretation of instructions; and 4) a single , simple arithmetic/logic unit for the processing of data. Parts (3) and (4) are known as the central processing unit or CPU. The important feature of this von Neumann computer is that each operation is performed sequentially, be it memory fetch/store, an arithmetic/logic operation or an input/output operation. In order to improve computer performance, three types of parallelism is being introduced into computer architecture. The first is pipelining where assembly line techniques are applied to the control unit and the arithmetic/logic unit. The second is to have several functional units which are independent entitities performing different functions and operating simultaneously on different data, for example, logic and arithmetic operations. The third, and in many ways the most powerful parallelism concept, is to have an array of identical processing units under a common control unit performing the identical operations simultaneously on different data stored in their private/dedicated memories. There is also the practice of mutiprocessing in supercomputation where several processors obey their own instructions and communicate via a common memory, thereby maximizing the thro'put.

An example of multiplying two floating point numbers using the serial, pipeline and array processors [1] is illustrated in Fig. 1. In the addition process of Fig. 1, one can immediately observe the power of the array procedure, its speedup going linearly with the number of processors. Of course, this clearly depends on the optimum balance between the vector length and number of processors, and this occurs when they are equal. Otherwise, the number of processors is too small, and the speedup is diminished. Or, if the vector length is too small, computer power is wasted; this is the issue of load-balancing. For the pipeline procedure, its speedup is linear with the number of suboperations. In the example cited in Fig. 1, it is equal to four. In Fig. 2, computer performance on a vector is given for the simple computers of Fig. 1, in terms of the maximum rate of computation, r(in magaflops) and the vector length to achieve half the maximum rate of computation $n_{1/2}$. The quantity L is the number of suboperations, n is the length of the vector, τ is the clock period (intrinsic speed) of the computer, and S is the fixed startup time required to setup the pipeline for the vector being processed. It is of interest to note the limits imposed by the technology and by the architecture [1] as indicated in Fig. 3.

S/CAD APPROACH VERSUS RF SYSTEM DESIGN

The major problem that plagues the RF system industries is the design-cycle caused due to the delay-factor in design turnover time. This will multiply to a very significant extent in the near-future vis-a-vis the new RF systems envisaged. With the advent of supercomputer-aided approach, the design-cost and the delay will however be reduced to a reasonable level.

The pragmatic aspects of S/CAD implementation for need-based RF systems can be understood from the following considerations pertaining to two particular systems:

A. High-Speed Interconnecting Networks & Packaging : A characteristic of high-speed digital ICs which dictates the methods by which they can be analyzed is their complexity. The maze of wiring and the inherent three-dimensional nature of the circuit would be difficult to be characterized as a simple electromagnetic boundary value problem. Partly for this reason and partly because there appear to be some inherent limitations to the speed-range of these circuits (due to skin effects), a full wave analysis is not often necessary. Rather, analysis using lump coupled elements and/or distributed parameter calculations with subsequent transmission-line analysis is satisfatory.

The most general design requirements of superfast digital systems are to realize reliable circuits with better density and higher speed. These requirements can be translated into specific electrical design protocols as indicated hereunder and detailed in Fig. 4:

- All signals on the system must meet timing requirement. A corollary is that the path lengths should be minimized and the critical path identified.
- The signal waveshapes must be within a given tolerance, since for example, a negative transition may lead to additional switching delay.
- Unwanted signal coupling between wires must be less than an upper bound so that coupled signals (cross-talk) do not cause improper switching of the logic circuits.
- Voltage transients on voltage distribution wires (and ground) induced by switching circuits must be limited to small fractions of the dc levels (Fig. 4).
- External electromagnetic disturbances should not cause false-switching of the digital circuits.

In the last decade an important shift has taken place in the design of high-speed digital hardwares with the advent of smaller and denser ICs and packages. Previously, the hardware components consisted of both physically and electrically large discrete parts. Stray elements and coupling among the components were small in most cases and the interconnections between the components were electrically insignificant. The corresponding electrical network models were highly decoupled and the network analysis matrices sparse. This led to relatively simple analysis models and techniques for the electrical performance of these systems. In contrast, today's high-level of integration can lead to very large and complex systems with extremely small physical dimensions. Therefore, an electrical analysis that excludes coupling among the closely-spaced components is rather invalid. Further the interconnections that once led to insignificant stray elements are now the main parts in the equivalent circuits. Thus, the circuit models for IC systems are extremely complex, with highly coupled hardwares. An electrical analysis of these models without computer-aided design techniques is not possible, especially for high performance systems.

The type of hardware designs which lend themselves to miniaturization are microwave, digital- and analog- type systems. Usually, the overall dimensions of the relevant parts of a subsystem for which a signal or coupling analysis is of interest are small, less than a few centimeters. Often, the highest frequency component contained in the signals which are propagating in the system corresponds to a wavelength (which exceeds the physical and electrical dimensions in many cases) makes the analysis with lumpe- circuit models valid.

The type of analysis required for a particular system depends on its performance and purpose. The electrical analysis may become a very simple one for low-speed or low-frequency circuits since the reactance of the capacitances is high and the inductances are almost short-circuits. Then a simple analysis may suffice which involves a few LCR elements. In contrast, complex models are required to represent high-speed/high-performance systems. The signal transitions in very low-speed digital systems may be in the micro- or even millisecond range. At the other end of the system, one is concerned with the analysis of a Josephson or MESFET technology where the signal transitions are in the picosecond range.

A fundamental quantity which characterizes a particular interconnection technology is what is known as the general impedance level. It is simply the lossless characteristic impedance (Z_0) of the "average" connection in the system. Typical values of Z_0 range from 5 to 200 ohms. In lower-performance FET logic hardware, the devices are typically of a higher impedance than Z_0, and thus the capacitance is the dominant circuit element .

Bipolar transistor logic hardware may exhibit impedances of the order of Z_0, so that both capacitance and the inductance are important. Josephson junctions exhibit a very low internal impedance and the inductance is regarded as the dominant stray element.

The supercomputer-aided electrical analysis of interconnections approach suggested here is based on computational electromagnetics and RF transmission-line theory. The steps involved in the modeling are:

- Preliminary/Approximate Model: The unknowns are expressed in terms of voltage and current using the circuit theory concept.
- Lumped Element Model: RLC representation of the transmission line and computation of the circuit elements.
- Analysis of the modeled circuits to evaluate their electrical performance.

Integrated circuits chips are placed on a chip carrier (Fig. 5) which is sometimes called a space transformer, since it transforms the closely spaced IC chip connections to larger connection points. The connections among the chip carriers are established in the mutiplane board. A logic signal may start with an LSI circuit located on one chip and may be received by a circuit located in the other chip. Thus, the signal may be delayed by both the integrated circuits and the package. The major contribution to the average delay is due to the circuits for lower-performance systems, while the package delay dominates for the average delay for high performance hardware.

Thus, the purpose of computer-aided electrical analysis is to ensure that the hardware at hand meets the electrical design criteria. The analysis of digital system interconnections encompasses a wide spectrum of frequencies. This is quite in contrast to microwave systems, which usually operate at a few band-limited regimes. As a consequence of this, a mixture of static, quasistatic and dynamic models is employed in high-speed systems. Quasistatic models such as lumped equivalent circuits play an important role in the representation of the complex physical geometries. Further, an analysis with an inexhaustive model, though approximate, may lead to valuable information without solving the complete dynamic problem. An example of this, is the analysis of a low impedance voltage supply system with an inductance-resistance model. In contrast to this, a complex capacitance-resistance model may suffice to represent most parts of low current, high impedance FET package.

In the above-mentioned modeling approach(es), three specific computational algorithms are possible: The first category is analytical methods; the second method refers to numerical techniques and the third approach is based on analog simulations.

Analytical methods (including empirical formulations) are the fastest means to obtain the solutions, but are limited currently to simple geometries. By proper hybridization with numerical and/or empirical approaches, 'quasianalytical' methods can be evolved. Depending on the size of the problems (as decided by the number of unknowns/discrete elements), supercomputation would lead to solutions with acceptable error limits.

Pure numerical solutions are quite popular in the existing research, design and development efforts. Integral equation methods [5] to calculate two-dimensional circuit parameters of transmission-lines have been successfully developed.. However, the only bottleneck in the the method is that the state-of the-art computers severely limit the caculations to about 100 elements, lest undue errors are noticed. Futher, computational process time increases enormously when the size of the matrix formulation increases to accomadate more unknowns. Though, improved computational algorithms based on conjugate gradient technique, method of moment etc. have been evolved to handle large-sized problems [6], still computational limits do persist warranting supercomputation and S/CAD methods. The trade-offs between accuracy of analysis and the size of the problem are depicted in Fig. 6.

Another problem associated with packaging/interconnections is the propagation of pulses on transmission-lines. Example propagation problems for chip-to-chip communication and muti-chip interconnection lines and cross-talks are illustrated in Fig. 5. Several problems can be identified which are fundamental to further progress in this area of high-speed networks. One is the development of CAD tools for analyzing propagation on lossy coupled nonuniform transmission lines with nonlinear sources and loads. Another is the development of CAD tools for analyzing cross-talk between parallel and nonparallel transmission lines and multiple reflection phenomena on interconnected transmission lines. In all these cases, the modeling techniques are limited by the state-of-the-art computers (in terms of memory and execution time). With the adoptation of S/CAD, better and accurate solutions can, however be realized.

To summarize, the following are the imminent CAD-needs of high-speed packaging/interconnection systems which can be conveniently and more effectively handled with S/CAD implementation:

- Development of automated design tools which incorporate lumped and distributed parameters-based algorithms into transmission-line/circuit analysis leading to CAD programs.
- Development of methods for analyzing steady-state and transient propagation on lossy coupled nonuniform transmission-lines with nonlinear loads and sources.
- Development of methods to analyze coupling between parallel and nonparallel transmission-lines and multiple reflection phenomena on interconnected transmission-lines.
- Development of methods for accurate estimation of the mutual capacitance matrix corresponding to 100 or more metallic elements.
- Development of stable methods to synthesize equivalent circuit models of complex package structures using measured frequency domain data. Relevant results will be used to analyze packages in the time domain.
- Studies on the limits of applicability of various design-methods for lumped and distributed parameters and the extendability of S/CAD strategies.
- Stochastical characterization of pulse propagation on transmission-lines which are effectively nonuniform with a 'random' propagation constant because they pass through a complex changing environment: S/CAD implementation.

B. **Stripline Problems:** Stripline structures with multiconductor transmission-lines embedded in mutidielectric media (Fig.7) [7] have practical significance in millimeter-wave applications. Existing algorithms are based on conformal mapping techniques, Green's function formulation, variational method, Fourier transform approach, Fourier integral solution and generalized spectral domain analysis.

Compatible to futuristic frequency of operation (towards 100 GHz), the relevant analysis should consider the lossy behavior of the aforesaid structure. Thus the algorithm will refer to a dynamic problem involving Green's function solution (Fig.8) with a scattering matrix solvable by the method of moments [8]. This numerical design approach will require the conversion of an operator equation to a matrix equation for necessary solution. However, the matrix involved will depict such a large systems of equations that use of the state-of-the-art CADs will not suffice. Therefore, a new algorithmic approach (such as the conjugate gradient method [8]) to solve the nonself -adjoint operator equation iteratively with computational requirements of systolic array architecture/vector and/or parallel processing is necessary to obtain an optimal solution. This will necessiate the use of a supercomputer. Relavant to this area, specific problems which can be addressed via S/CAD strategies are as follows:

1. There are a very few numerical methods available for adequate characterization of junctions and bends in transmission-lines carrying RF energy above 30 GHz. What is needed is a full wave solution for analyzing microstrip-lines and efficient ways to compute (supercomputation ?) radiation from microstrip structures.

2. As circuits become clustered together, methods should be developed for accurate characterization of proximity effects, parasitic coupling, effect of discontinuities and junctions. Again, emphasis is to be given towards full wave solution at frequencies higher than 30 GHz, since radiation is an integral part of the model. Use of finite element and/or integral equation approaches with the state-of-the-art computers to handle the size of the problem in question, is not pragmatic considering the memory requirements and the computation time required. Hence, S/CAD is the only viable approach.

3. There is a need to find the physical limits of a high-Q circuit. At most, a quality factor (Q) of 100 is attainable for open resonator structures in planar configurations. The question that is yet to be answered is: What is the ultimate value of Q that is theoretically attainable? Also, are alternate enclosed cavity structures viable in planar applications to offer very high Q-factors? To what extent S/CAD will supplement the relevant studies?

4. What is the typical characteristic impedance one should utilize in the design of microstrip-lines? For example, with a low characteristic impedance line, the circuit is tightly coupled and the losses are high. A high characteristic impedance, on the other hand, implies that fields are very loosely coupled and hence the structure is highly prone to external EMI. The question is what should be proper characteristic impedance and how should this be based on an optimum external interference criteria and losses in the circuit. Considering the complexity of the problem, the optimum solution can be reached only via S/CAD.

5. Research efforts are lacking in the analysis of distributed inductors and transformers in MMIC design and, again when the analyses take into account of radiation effect, the size of the problem significantly increases warranting supercomputational needs.

6. There is also a consideration as regards MMIC chip resonance arising from dielectric resonator phenomena. Though this problem is solvable, for accurate results however, supermachines are needed for computations involved.

7. Currently, there is no nonlinear large-signal device model available. Moreover, when a model does exist it is inaccurate as it is based on static/quasistatic data and therefore does not include any high-frequency or dynamic representations. Also, most numerical models tend to ignore the effect of the dielectric substrate for printed circuits and radiation effects. Hence, for accurate modeling of microstrip circuits at high-frequencies, it is necessary to solve Maxwell's equations with the proper boundary conditions in a more exact fashion. The dynamic nature of the corresponding algorithm pushes the computations involved into the supercomputer domain.

8. For the development of large-scale interactive CAD graphics for electromagnetic analysis, it is highly desirable to have supercomputational feasibilities.

9. Finally, it would be useful to have exhaustive validation and complete S/CAD-based documentation of capabilities and limitations of various electromagnetic models of devices. This will include development of error estimates for various computational algorithms. Also, there ought to be flags within the S/CAD output to indicate the error estimate of the model and information about the reliability of the solution.

CONCLUSIONS

Indicated above are only a few S/CAD versus RF system design considerations. In fact, viewing the whole gamut of fast-growing RF systems, S/CAD will be the "slide-rule" of RF designers in the near-future, without which, possibly no accurate and cost-effective designs can be implemented in the RF domain within a resonable time-frame of design turnovers.

REFERENCES

[1] Hockney, R.W., and Jesshope, C.R., Parallel Computers (1983, Bristol: Adam-Hilger)

[2] Browne, J.C., Physics Today, 37(5), pp.28, 1984

[3] Fox, G.C., and Otto, S.W., Physics Today, 37(5), pp.53, 1984

[4] Seitz, C., and Matsoo, J., Physics Today, 37(5), pp.38, 1984

[5] Jawson, M.A., and Symm, G.T., Integral Equation Methods in Potential Theory and Electrostatics (1977, New York: Academic Press)

[6] Wei, C., et al., IEEE Trans. on Microwave Theory & Tech., MTT-32, pp.439,1984

[7] Weeks, W.T., IEEE Trans. on Microwave Theory & Tech., MTT-18, pp.35, 1980

[8] Sarkar, T.K., et al., Survey of Numerical Methods for Solutions of Large Systems of Linear Equations for EM Field Problems, Report: RADC-TR-81-103 dated June 1981

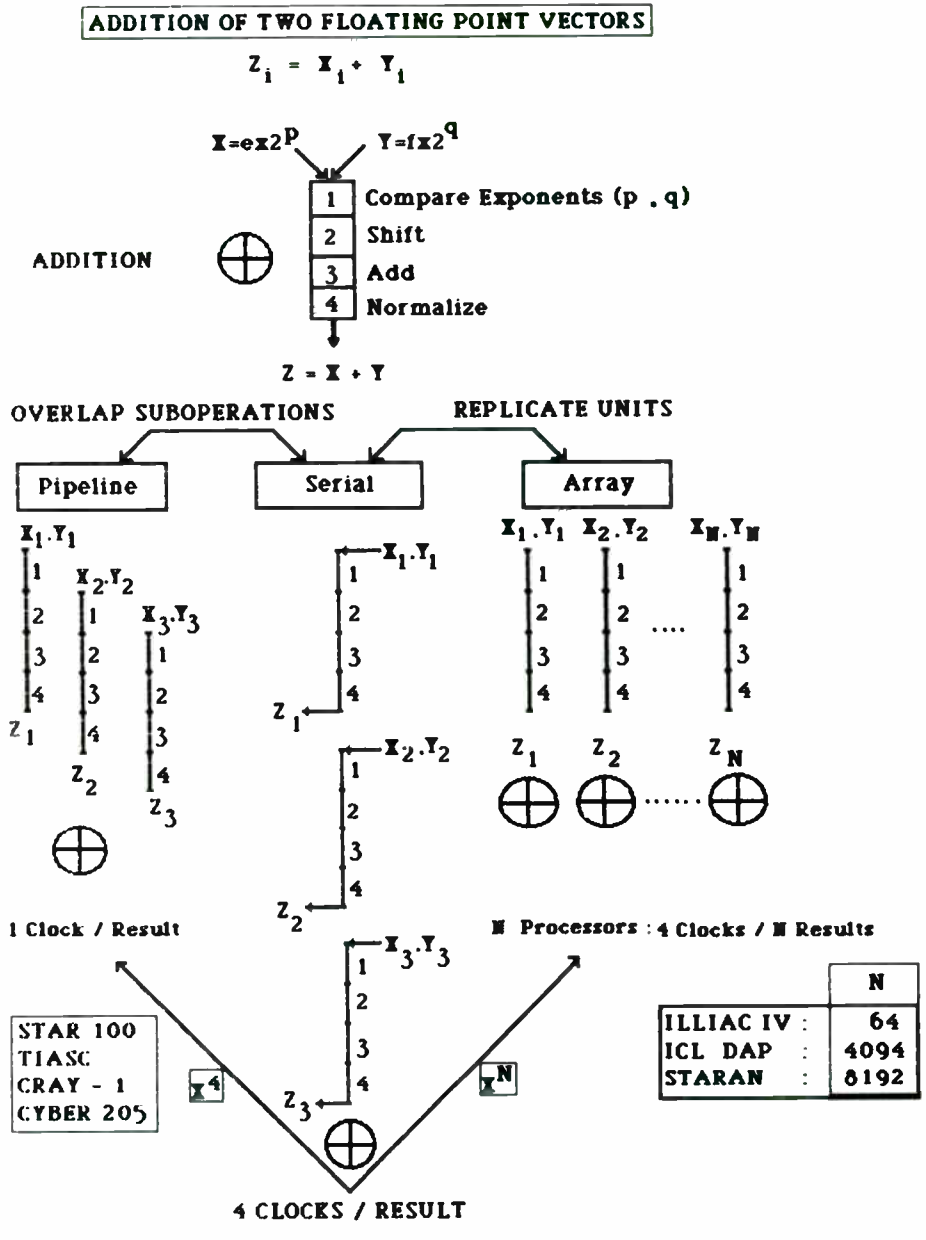


FIG.1: ADDITION OF TWO FLOATING POINTS

r_∞ & $n_{1/2}$ FOR VARIOUS ARCHITECTURES			PARAMETERS
SERIAL	$n_{1/2} = 0$	$r_\infty = 1/L\tau$	r_∞ = vector-length for semimax computn. rate of $n_{1/2}$
PIPELINE	$n_{1/2} = S+L-1$	$r_\infty = 1/\tau$	L = # of suboperations
ARRAY	$n_{1/2} = N/2$ $r = N/\tau$ if $n>N$ $n_{1/2} = n/2$ $r = n/\tau$ if $n<N$		n = vector length
FACTS ● $n_{1/2}$ architecture-limited ● r_∞ architecture & technology limited ● τ strictly technology limited ● Replicating a processor N times, multi-plies r_∞ and $n_{1/2}$ by N			τ = clock period S = startup time for pipelining

FIG.2: COMPUTER PERFORMANCE ON A VECTOR FOR SIMPLE COMPUTERS OF FIG.1

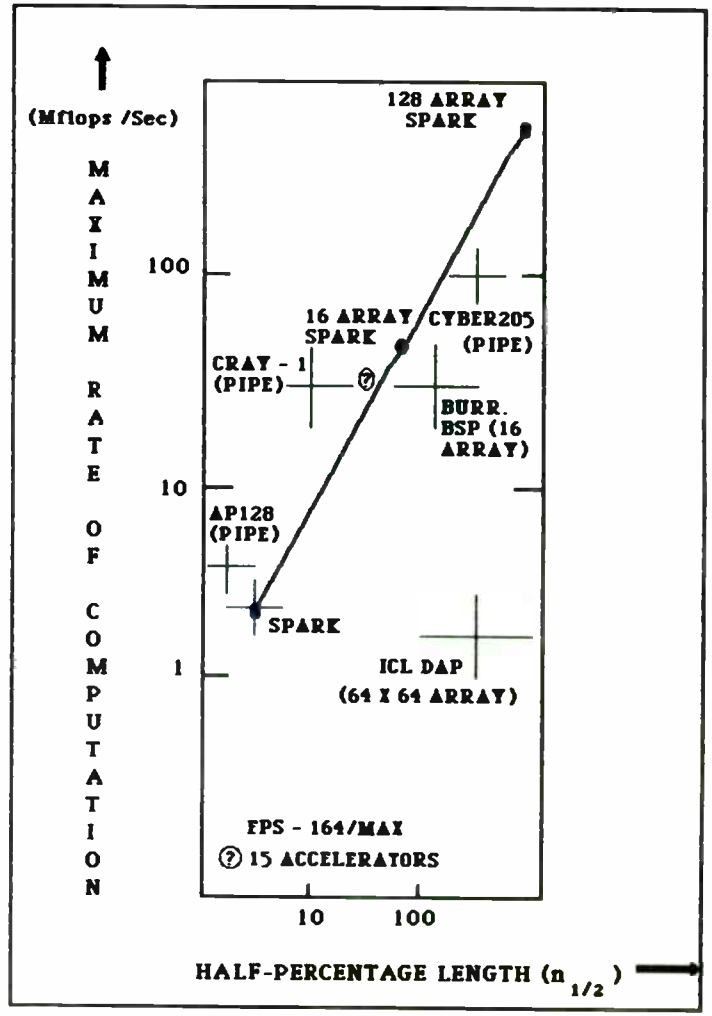


FIG.3: AVERAGE PERFORMANCE OF COMMERCIAL MACHINES

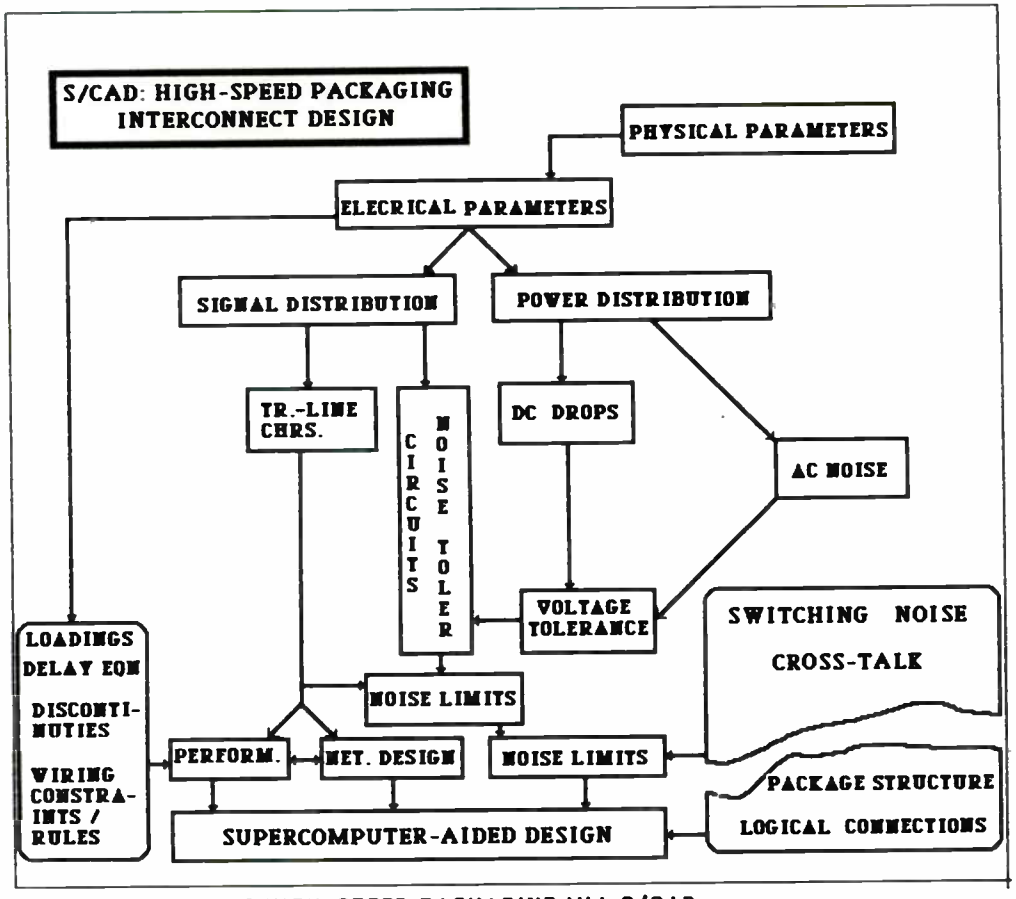


FIG.4: DESIGN OF HIGH-SPEED PACKAGING VIA S/CAD

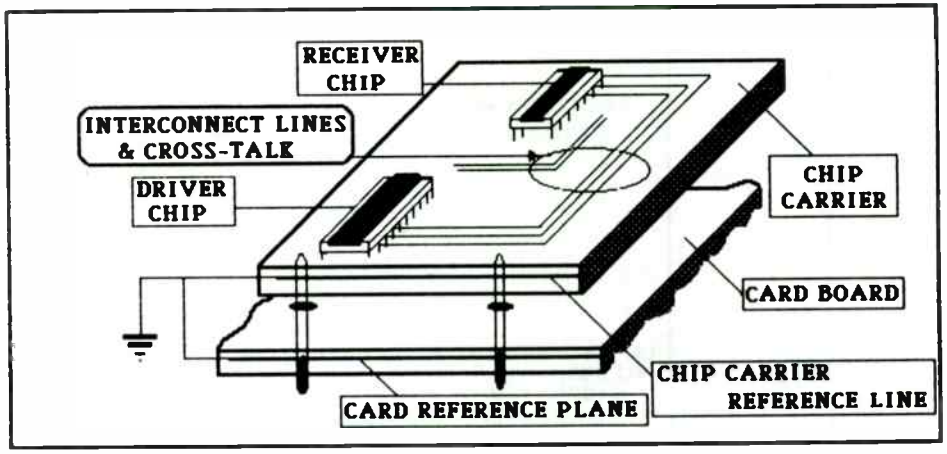


FIG.5: PACKAGE-ENVIRONMENT FOR CHIP-TO-CHIP CMMUNICATION

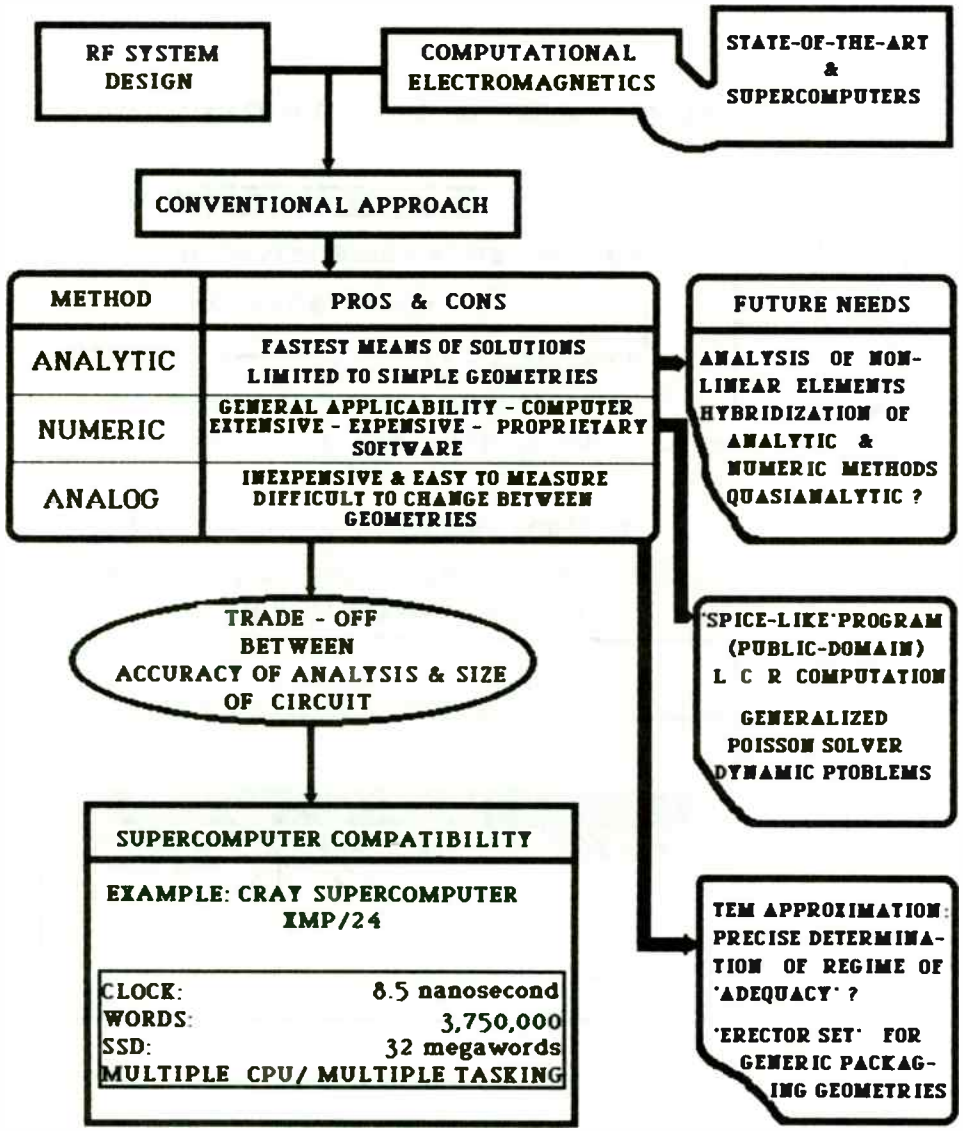


FIG. 6. EVALUATION OF S/CAD STRATEGY VIS-A-VIS RF DESIGNS

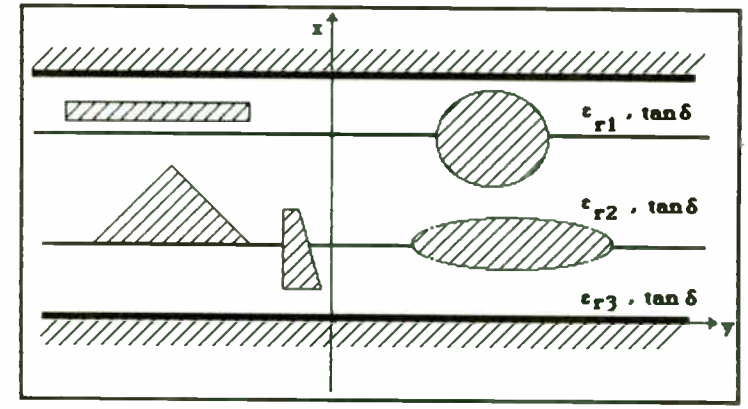
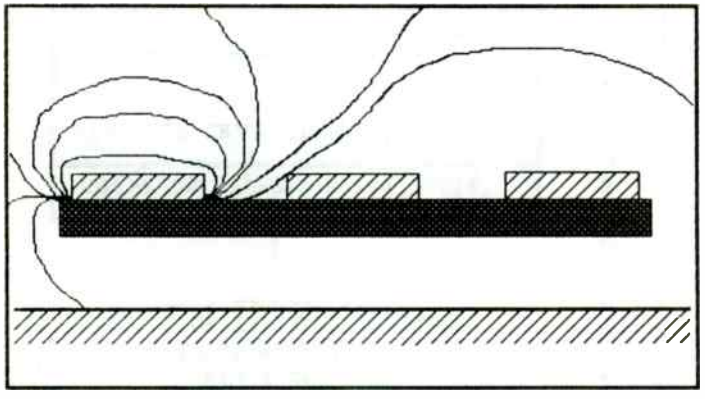


FIG 7: ARBITRARILY-SHAPED MULTIPLE CONDUCTORS IN MULTILAYERED LOSSY DIELECTRICS: A GENERALIZED ASYMMETRIC MICROSTRIP-LINE MODEL



BOUNDARY ELEMENT METHOD

GREEN'S FUNCTION SOLUTION OF FIELD PROBLEMS

$G(r/r') = -\ln|r - r'|$ **(GREEN'S FUNCTION)**

$$\Phi(r) = \frac{1}{2\pi\varepsilon}\int_S \sigma(r')G(r/r')ds' + C$$

POTENTIAL ⟷ CHARGE DISTRIBUTION

(INTEGRAL EQUATION)

GALERICAN PROCEDURE

$\underline{S}\,\sigma = \underline{b}$ **(ALGEBRAIC MATRIX EQUATION)**

⟹ **NUMERICAL SOLUTION**

FIG.8: SIMPLE MICROSTRIP STRUCTURE / ANALYSIS

A GaAs PULSE MODULATOR DESIGN FOR RF SIGNAL GENERATORS

by
Ted J. Dudziak, P.E.
Project Engineering Manager
Wavetek RF Products
5808 Churchman Bypass
Indianapolis, IN. 46203-6109

INTRODUCTION AND BACKGROUND

Amplitude control of RF and Microwave signals has traditionally been accomplished using the PIN diode. There are two control modes : linear and on/off. The linear mode is that used to implement AM or level control while the on/off mode is used in transceiver applications.

This paper discusses the on/off control mode or that mode required for implementing a pulse modulator.

The use of the PIN diode as a current controlled resistor is a result of the forward bias current vs RF resistance characteristic. This characteristic is sufficiently linear over a wide enough range to use it in amplitude modulators. Its main limitation is the lower frequency limit due to the minority carrier lifetime. This characteristic also affects the switching speed of the diode. A compromise is always made between the switching speed and the lower operating frequency limit. Optimizing one parameter always degrades the other parameter.

Another concern is that the control parameter is the bias current. For an amplitude control configuration with several PIN devices, the required control current can be substantial such that there could be a concern about either the control circuitry or the power supply design.

The GaAs FET overcomes these concerns in RF amplitude control applications. However, it has only been recently that it has been as economic a solution as the PIN diode.

The design of general purpose test equipment has its own set of design criteria. It is a situation in which the world is your customer. One would like to satisfy all the customer requirements with one product. The device or technology which allows the test equipment manufacturer to provide as broad a solution as possible will get the most consideration.

Currently pulse modulators are being implemented using PIN diodes in RF products both as stand-alone devices as well as functions within RF signal generators. The resultant performance of pulse modulators in signal generators appears to be adequate for most applications. However, more demanding requirements mean that a stand-alone product be used. These products are designed to deliver the best performance over the widest frequency range. Table 1 contains a typical set

of specifications for a stand-alone product. The drawback of the current PIN stand-alone product offerings is the cost, overall switching speed, and shear bulk from the packaging.

TABLE 1
PULSE MODULATOR SPECIFICATIONS

	STAND ALONE	INTEGRATED
ON/OFF RATIO	> 80 dB	> 60 dB
FREQUENCY RANGE	100 TO 2000 MHz	1 TO 1000 MHz
INSERTION LOSS	< 3 dB	< 2 dB
RISE AND FALL TIMES	< 30 nSec	< 500 nSec

The performance of pulse modulators implemented within RF signal generators is compromised considerably. Table 1 illustrates typical performance of an integrated pulse modulator. In particular is the ON/OFF ratio. Most of these implementations are done with the AM circuitry.

The economic availability of GaAs devices has allowed a high performance pulse modulator to be designed both as a stand-alone product and integrated into an RF signal generator. Since it will be considered as a stand-alone product its specifications should be at least as good as those products currently available. Thus the overall performance within an RF signal generator will surpass anything currently available.

GaAs devices offer several advantages over PIN devices which contribute to a better pulse modulator design. First the switching speed is better. The switching speed is not dependent upon minority carriers as is the PIN. The dominant contributor is the input gate structure. As a result switching speeds approaching one nanosecond are possible. Second the control parameter for the GaAs device is a gate voltage instead of a forward bias current. Large amounts of current are not required for a multiple element pulse modulator design. Third, GaAs devices offer better overall isolation characteristics which lead to a simpler design over a wider frequency range. It takes less devices to achieve the equivalent performance.

Among the design concerns is the insertion loss which is still significant for a design with series devices. For an integrated design, the insertion loss will affect the output level specification since the pulse modulator will be inline. Sufficient output margin should be available from the output amplifier stage to overcome the insertion loss from the pulse modulator. The insertion loss characteristic will vary as the output frequency. Thus the operation of the signal generator should be able to compensate for the frequency variations.

ELECTRICAL DESIGN

Figure 1 is a schematic of the GaAs pulse modulator. The key element for the RF performance of the pulse modulator is the MACOM MA4GM201-500 SPDT switch. Two series devices (SW2 and SW4) and one shunt device (SW3) comprise the basic RF switch. The switches are arranged so that the series elements of SW2

and SW4 are ON and the shunt elements of SW2, SW3 and SW4 are OFF when the pulse modulator is ON and the opposite configuration when the pulse modulator is OFF.

Between SW2 and SW4 are two quarter wave lines at 1800 MHz. SW2 and SW4 also contain shunt elements which provide more isolation when the pulse modulator is OFF. The shunt device, SW3, provides additional isolation when the pulse modulator is OFF. Some peaking of the overall response is also provided by the two quarter wave lines. The GaAs devices look primarily resistive with some parallel capacitive reactance when the series RF path is enabled or the pulse modulator is ON. The impedance transformation of the quarter wave lines cancels some of the reactance thus flattening the overall response. SW1 switches in a 50 ohm resistor which is used to provide a constant impedance to the driving amplifier when the pulse modulator is OFF.

A monolithic FET driver was selected, the Siliconix D169. Although it does not exploit all the speed capability of the GaAs switches its usage did simplify the overall design while providing very good switching speed. A simple JFET device provides the enable/disable function for the pulse modulator.

The polarity of the drive signal can be changed by two simple circuit modifications. First the series and shunt drive lines are reversed at the FET driver. Second, the source is connected to the appropriate reference. If an input of +5 volts is to represent "RF ON" then the source should be connected to ground. The source should be connected to +5 volts if 0 volt input is to represent "RF OFF".

Note that there is no terminating resistor on the pulse input drive line to the D169. The FET driver is a fully TTL compatible device. Any terminating resistor would have ruined any plans to drive the pulse modulator with a TTL driver. TTL totem-pole outputs will not drive a 50 ohm load but an appropriate TTL line driver will drive a 50 ohm load. TTL compatibility is more than voltage levels.

However, it is important to maintain good transmission line terminating practices when operating the pulse modulator as the drive signals should be consistent with the expected rise and fall time characteristics of the gated RF output. The rise and fall time characteristics will be in the order of 30 nSec. The use of 50 ohm back-matching at the pulse generator source will provide a very adequate drive signal to the pulse modulator. Thus there will be minimal distortion to the gated RF pulse due to reflections on the pulse drive line.

Figure 2 illustrates the back-matching technique. The pulse generator source impedance,Zout, and the cable characteristic impedance, Zo, must be the same, preferably 50 ohms. A pulse leading edge propagates down transmission path to the input

of the pulse modulator drive pin. The high input impedance of the drive pin reflects the leading edge of the drive pulse back down the transmission line where it is absorbed by the generator source impedance. A similar result is noted for the trailing edge of the drive pulse.

Although there is distortion of the pulse at the generator end of the transmission path, the pulse at the pulse modulator end of the transmission path is a true representation of the intended pulse. If there is a concern about pulse distortion anywhere along the transmission path a termination can be placed at the pulse modulator drive input. The distortion at the pulse source shows itself as a widening of the pulse width by four times the propagation delay of the transmission line.

THE HYBRID IMPLEMENTATION

The pulse modulator was fabricated using hybrid technology. The circuit traces are screened and fired gold on a 10 mil alumina substrate. The substrate is attached to the hybrid package with a very low resistivity conductive epoxy to ensure that the ground plane of the substrate is uniformly attached to the package. Chip capacitors and thin-film nichrome chip resistors are attached to the circuit trace pads with conductive epoxy. The FET driver and JFET devices are also attached to the substrate with conductive epoxy. All bond connections to the substrate and the active devices are made with 1 mil bond wire. Double bonds were made at all connections to insure a reliable connection and to reduce the overall bond wire inductance. Figure 3A, B, and C illustrate the die attach and bond connections of each type of device. Note that only one bond connection is made to the FET driver since the bond pad will accommodate a singe bond connection.

All active devices and the nichrome resistors are passivated to avoid incidental handling and environmental problems.

The passivated GaAs switches are attached with conductive epoxy to the hybrid package through a hole in the substrate. There is one hole for each device. The hole in the substrate was used to reduce the length of the bond wires and thus their parasitic inductance. By setting the devices into the substrate the length of the bonding wires was reduced. All ground connections to the GaAs devices are made with 3x5 mil ribbon wire. The ribbon wire was used to further reduce the parasitic effects of the bond wire inductance. In addition, multiple bonds are made to the GaAs switch especially the SW2, SW3, and SW4 source connections made to ground. Figure 4 is a picture of the bond connections made to the GaAs switches. Note the large number of bond connections at the left edge of the switch. The inductance of the bond wire was modeled using the Touchstone(tm) RF Design Analysis Program. The purpose of the modeling was to establish the bond length tolerances for the hybrid fabrication process.

There was good correlation between the results predicted by Touchstone and the measured performance.

GaAs devices are extremely sensitive to electrostatic discharge. Handling of these chip devices should only be done with ionized air and teflon tools. Teflon pickup tools are available if an epoxy dispense and die placement station is used.

Further, the bonding method is also critical for GaAs chip devices. A commonly used technique for bonding semiconductor devices is thermal-sonic. This is a combination of temperature and ultrasonics to achieve a bond connection to the semiconductor device. This method is not recommended by the manufacturer of the GaAs switch. Thermal compression bonding is used because the GaAs substrate is very prone to fracturing when sonics are applied. Since there is some variation in the operation of all bonders a consultation with the GaAs chip device manufacturer is recommended to determine the actual bonding considerations. There are usually some guidelines given on the chip data sheet, but caveat emptor.

THE PACKAGING

The package design had to take into account several concerns. First the pulse modulator had to be able to achieve the isolation specification. RF isolation experiments were performed on several package sizes. The purpose of the isolation experiments was to determine the package size required to achieve the ON/Off ratio specification.

A 2.1x1.1 inch hybrid enclosure was designed to accommodate the 1.5x0.525 inch alumina substrate. The initial design was nickel plated brass. Subsequent package designs will done using Kovar material. The use of a Kovar package is necessary to better balance the coefficients of thermal expansion of the hybrid package, the epoxy used for attaching the devices and the alumina substrate.

The second concern is the need for flexibility of the design in different applications. A double package design was decided upon. Figure 5 illustrates the double package design. Note that there is an inner enclosure and an outer enclosure. The inner enclosure can be utilized within a design as a functional module. This can be done on a pc-board with the hybrid connections made directly onto the pc-board. The outer enclosure will be used if the pulse modulator will be used in a standalone configuration such as the implementation on the 2520 signal generator. The outer enclosure has two SMA connectors for the RF connections and one four-pin in-line connector for the power and control signals.

THE IMPLEMENTATION

The intended use of the pulse modulator is in the Model 2520 signal generator. Figure 6 is a block diagram of the

implementation. Note the pulse modulator is located between the output amplifier and the step attenuator assemblies. The pulse modulator is not an integral part of the amplitude control so it is operated at a fixed output level. Any variations in frequency response are compensated for by the AutoCal(tm) routines for the level control. The GaAs performs as well at DC as it does at RF frequencies. However, it is possible to destroy the GaAs devices if greater than 0.5 volts is applied to either RF port. The DC block and RF circuit breaker in the step attenuator provide the necessary protection.

SO HOW DOES IT PERFORM ?

The resultant pulse modulator specifications are provided in Table 2.

TABLE 2
PULSE MODULATOR SPECIFICATIONS

FREQUENCY RANGE	DC TO 2200 MHz
INSERTION LOSS	< 4.5 dB
ON/OFF RATIO	> 80 dB
RISE AND FALL TIMES	< 50 nSec
RETURN LOSS	> 25 dB
VIDEO FEEDTHROUGH	< 5 mV
PULSE INPUT	FULLY TTL COMPATIBLE

Figure 7 and 8 show the isolation and insertion loss performance data for several production units. The isolation performance is typically 90 dB at 2200 MHz and 108 dB at 100 MHz. The insertion loss specification is 4.0 dB at 2200 MHz, 2.5 dB at 1500 MHz and less than 1.6 dB below 500 MHz. This performance correlates well with CAD analysis using Touchstone(tm). A similar analysis of return loss indicates performance better than 25 dB.

ACKNOWLEDGMENTS

The pulse modulator design team was comprised of Jack Coffey, electrical designer, Tim Stucker, project coordinator, Daryl Metzger, Dale Hedgespeth and Donna Watson, mechanical designers, Bill Kennedy, testing and evaluation, Jim McClellan, hybrid engineering, Dave Noble, hybrid technician, and Midge Traylor, hybrid assembler. A lot of effort went into the project which was a learning experience for all. Ray Sokola, engineering manager, is appreciated for his patience.

FIGURE 1 PULSE MODULATOR

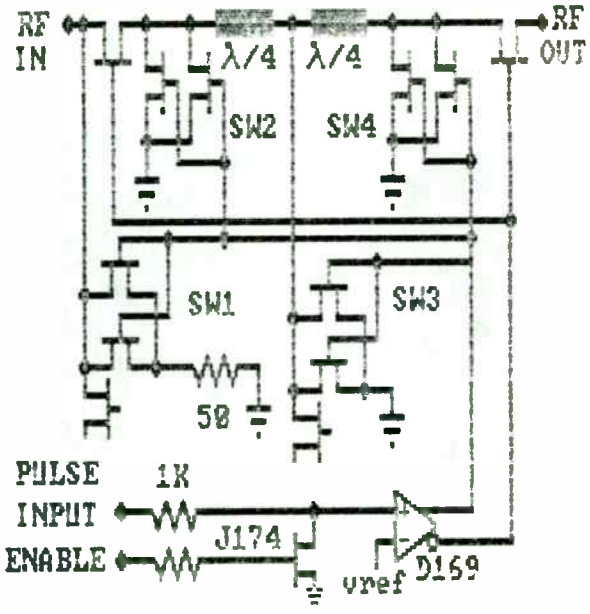
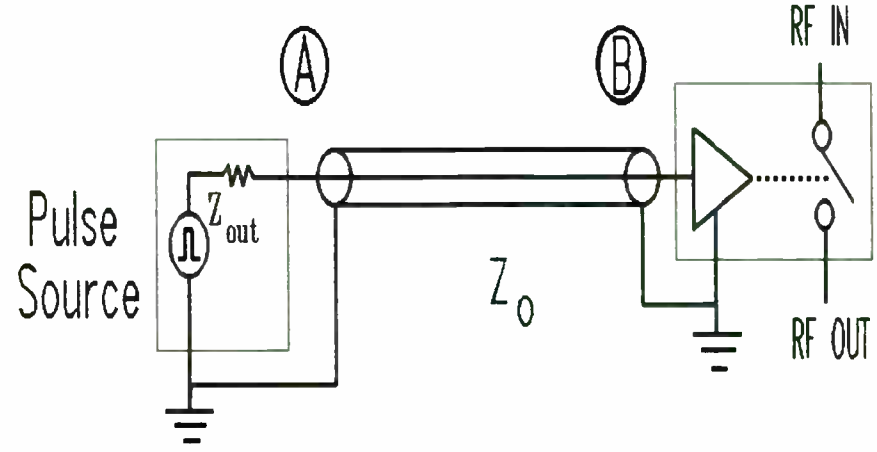
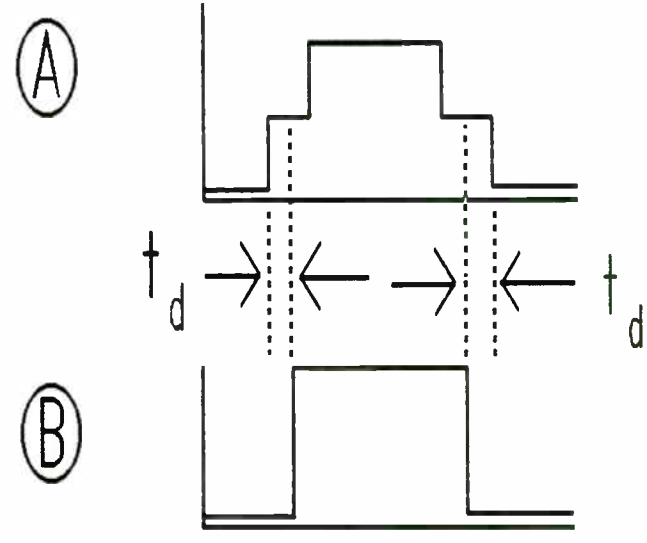


Figure 2 - Back-Matching





t_d = Propagation Delay Time

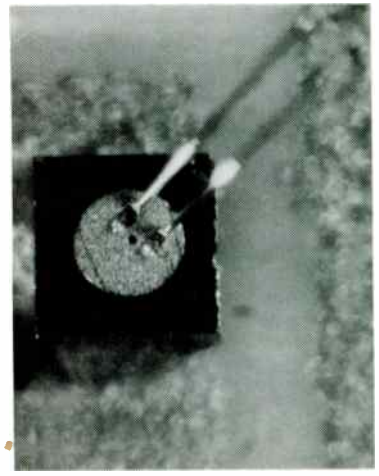


FIGURE 3A.

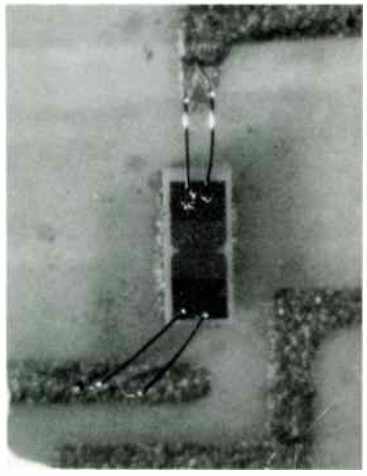


FIGURE 3B.

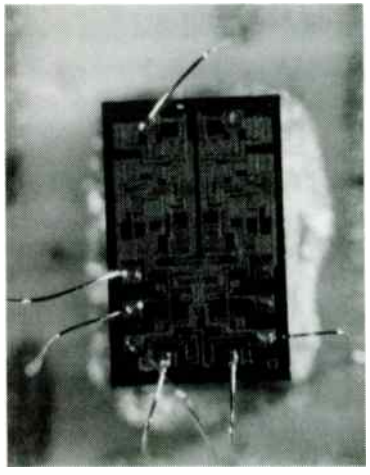


FIGURE 3C.

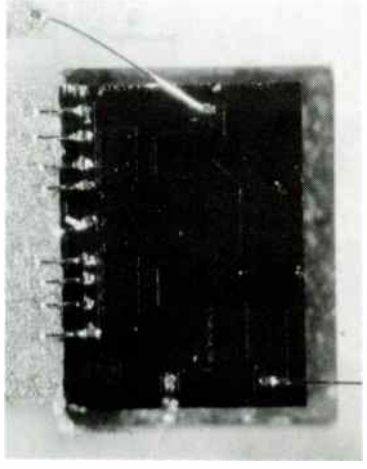
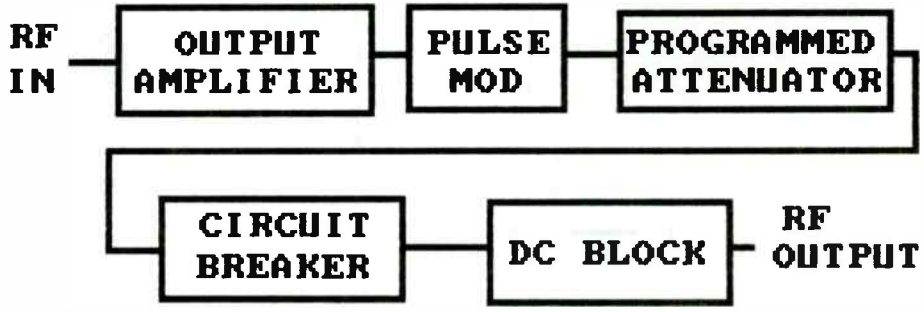


FIGURE 4.

FIGURE 6 BLOCK DIAGRAM OF PULSE MODULATOR IMPLEMENTATION



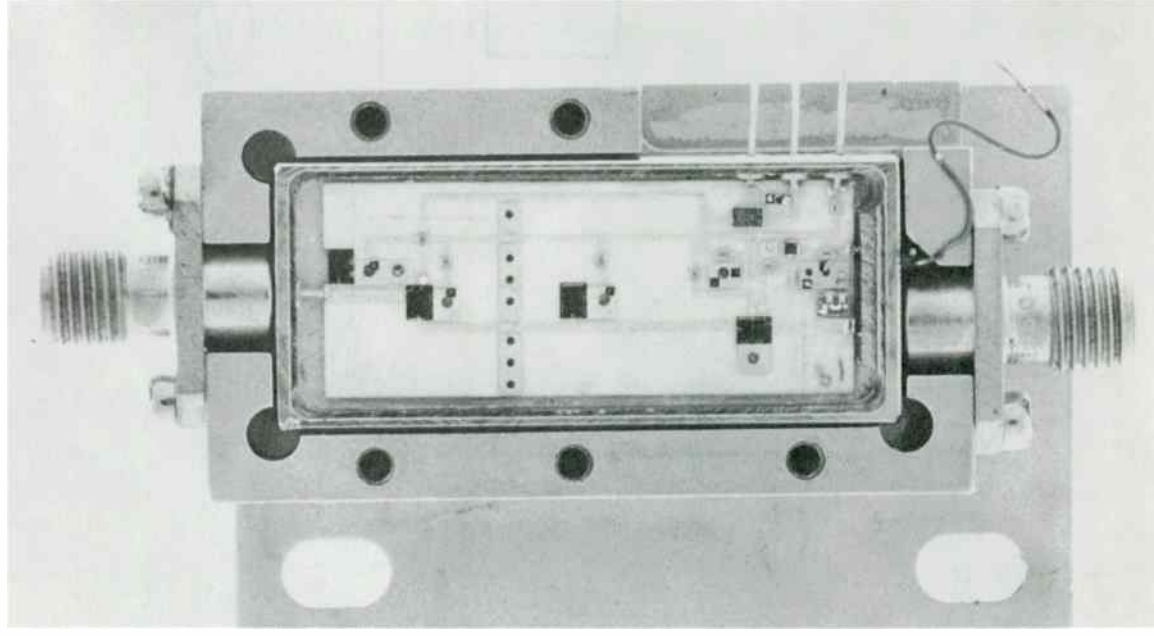


FIGURE 5 PULSE MODULATOR PACKAGE

FIGURE 7 ISOLATION vs FREQUENCY

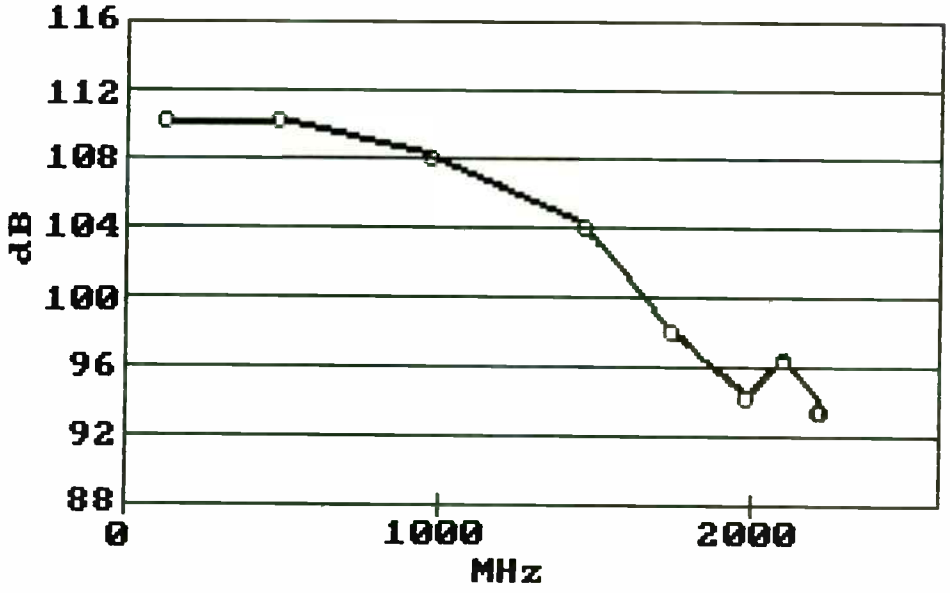
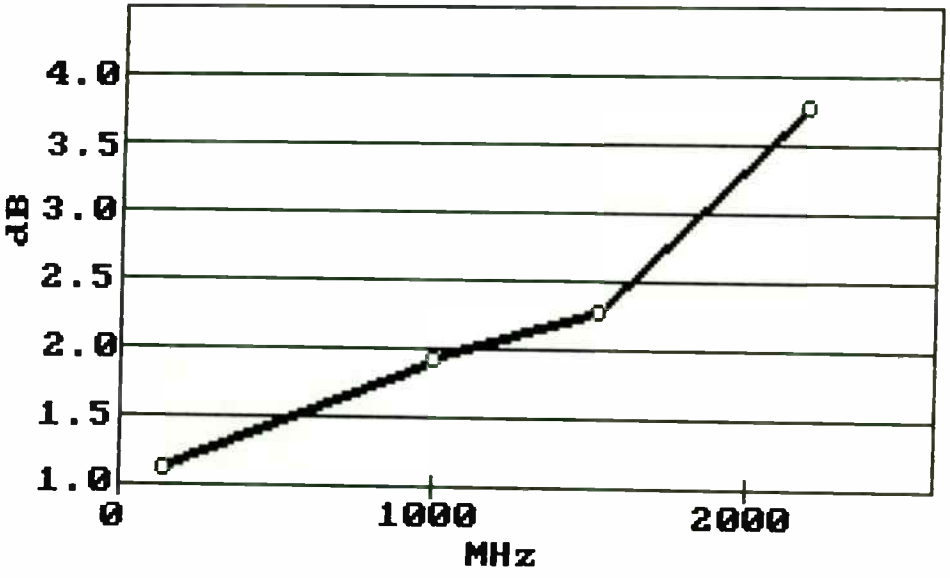


FIGURE 8 INSERTION LOSS vs FREQUENCY



RADIO-WAVE PROPAGATION - A TUTORIAL

PART 1 - BASIC PRINCIPLES AND CHARACTERIZATION

D. R. Dorsey, Jr
Principal Systems Analyst

Network Development Division
Control Data Corporation
Sunnyvale, CA 94089

1.0 INTRODUCTION

What happens to a transmitted radio wave after it leaves the antenna ?

This tutorial is intended as a basic introduction to the processes involved in the propagation of radio waves through the atmosphere of our earth, as well as some of the exotic forms of wave travel that occur. Part 1 will investigate the overall structure of the atmosphere and that part of it that plays the greatest role in supporting radio wave travel, the ionosphere. Fundamental processes and theoretical models are presented, and we will also study the influence of the sun, which makes it all possible.

Part 2 will study propagation at high frequencies, primarily up to 30 MHz, where the influence of the ionosphere is predominant. Part 3 extends the discussion to frequencies at VHF and above, where propagation modes are defined more by line-of-sight paths than by ionospheric refractions.

1.1 STRUCTURE OF THE ATMOSPHERE

The atmosphere of our earth, in the general meteorological sense, is generally characterized by the temperature, pressure and density variations of the air at a given altitude. The pressure and density behavior is governed by the well-known "barometric equation" from fluid mechanics [1,3]

$$dp = -g\rho dh = - Nmg\ dh \qquad (1)$$

where

g = gravitational acceleration
h = height (altitude)
N = number density of molecules
m = mean molecular mass

and the pressure/temperature relation is generally assumed to obey the perfect-gas law

$$p = NkT \qquad (2)$$

where

k = Boltzmann's Constant = 1.372×10^{-16} erg/deg
T = absolute temperature (deg Kelvin)

Combining (1) and (2) we have

$$dp / p = - (mg / kT)\ dh = - dh / H \qquad (3)$$

where $H = kT/mg$ is defined as the "scale height" of the atmosphere.

For an isothermal atmosphere, (3) can be integrated to give the pressure and density relations

$$p = p_0 \exp (- (h - h_0) / H)$$
$$\rho = \rho_0 \exp (- (h - h_0) / H) \qquad (4)$$

such that p is the pressure at an altitude h relative to a reference pressure p_0 at height h_0.

The temperature profile of the atmosphere is shown in Figure 1. In the mesosphere, the increase in T with increasing altitude is due mainly to ionization and disassociation of the gases in the atmosphere by solar ultraviolet radiation. Obviously, this profile will vary diurnally and with changes in solar activity.

Below 100 km, however, the specific heat of the atmosphere is so large that the temperature stays essentially constant, justifying the assumption of an isothermal atmosphere made above.

The chemical composition of the earth's atmosphere up to approximately 90 km is essentially constant, due mainly to turbulent mixing. The primary constituents (by volume and mass percentages) are

	Volume	Mass
	---	---
Nitrogen	78.09	28.02
Oxygen	20.95	32.01
Argon	0.93	39.96
Carbon dioxide	0.33	44.02

Trace amounts of water vapor, ozone, nitric oxide and hydrogen also exist.

1.2 STRUCTURE OF THE IONOSPHERE

How does an ionized layer get formed ?

Primarily by radiation from the sun. The major ionizing agents are ultraviolet radiation and x-rays. During periods of major storm activity on the sun, particle radiation also becomes a consideration. To a lesser degree, cosmic radiation also contributes.

As this radiation passes through the atmosphere, coming from space, it ionizes the rarified gases in its path. As it comes closer to the earth's surface, the gas density increases and consequently the degree of ionization increases. The interaction of the radiation with the matter in the atmosphere does, however, involve attenuation of the incident radiation due to the energy expended in the ionization process. At some point, all the energy in the incident radiation has been used up, and ionization ceases. Consequently, we wind up with an area of maximum ionization with a tapering off of ionization level above and below.

We must also keep in mind that the ultraviolet radiation is not monochromatic and may (spectrally) cover a broad range of frequencies. Also, the atmosphere will absorb this radiation at certain well-defined frequencies. and wavelengths, depending on the chemical composition of the atmosphere at a given altitude. Thus, we can expect several such levels to occur.

As the ionization process continues, free electrons and positively charged ions are produced. The free electrons will ultimately recombine with positive ions, or attach themselves to a neutral molecule, forming a negatively charged ion. These processes change continually, moment to moment, constantly re-defining the ionic nature of the atmosphere. The degree of ionization at any point is never constant, and depends on many factors, including time of day, season of the year, the distance between earth and sun and the current level of activity on the sun.

Therefore, we find the atmosphere rather neatly divided into regions of relatively dense ionization, whose boundaries are fairly well defined. It is these regions that we call the ionospheric "layers". These so-called layers, however, are not completely isolated from one another, and there is some degree of overlap between them. But they are distinct enough that we can look at them individually and study their properties.

The ionospheric region of most importance to the propagation of radio waves extends from about 30 to 400 km. Ions exist in this region in sufficiently high numbers to profoundly influence a radio wave passing through it. In the rarified environment of the ionosphere, free electrons and ions can travel back and forth for long periods of time before recombination with another ion occurs. Electron density provides an excellent characterization of the state of the ionosphere. The average electron density profile is also plotted in Figure 1. Compare this against the average temperature profile in the same figure.

1.3 IONOSPHERIC LAYERS

The primary layers of interest to the RF world are called the D, E, F1 and F2 layers, after the naming convention initially set forth by Sir Edward Appleton in 1924. All of these layers display variations in characteristics both diurnally and with change in seasons. The general layout of the layers is presented in Figure 2. Let's now look at each layer individually.

1.3.1 D-LAYER

The D-layer is the lowest layer, extending from about 55 to 92 kilometers. This part of the atmosphere is very dense relative to the upper regions, and can produce significant attenuation in a signal moving skyward, due mostly to energy loss by heat generation from the electronic collision processes. Since it is also the lowest layer, the ionization efficiency is also very low; the ions produced here tend to recombine very quickly.

Thus, the D-layer is primarily a "daytime" layer. Peak ionization occurs around noon, when the sun is directly overhead and the solar radiation flux density is at a maximum. When the sun sets, this layer virtually disappears.

The D-layer primarily affects very low frequencies, usually < 300 KHz. Above 300 KHz, a signal usually passes right on through the layer and continues on, since the D-layer is not very good at bending or reflecting an HF wave back to earth. Thus, this layer is useful only for communications over short distances. It can act negatively, particularly in times of heavy ionospheric storm activity, causing signal absorption to be so high as to render signals non-existent above the D-layer.

1.3.2 E-LAYER

The E-layer is the next highest region, spanning from about 90 to 125 kilometers, and slightly overlapping the upper boundary of the D-layer. The air density is lower here, and the level of ionization is higher. The level of ionization is strongly dependent on the sun's position relative to the horizon. Significant ionization occurs just after sunrise, increases rapidly toward its maximum at noontime, and drops off abruptly shortly after sunset. Some residual activity remains after sunset, and local minimums appear around midnight.

The E-layer is also primarily a "daytime" layer. It is useful for somewhat higher frequencies than the D-layer, and absorption is not quite as high. Bending and reflection of a radio wave will occur to a slightly higher degree than in the D-layer, allowing communications over a slightly longer distance than is possible with the D-layer.

The E-layer also exhibits a "*Sporadic-E*" effect, which will be discussed in a later section.

1.3.3 F-LAYER

By far, the F-layers, comprised of the F1 and F2 layers, are the ones that exert the most influence on long-distance communications, particularly in the shortwave portion of the spectrum. World-wide communications are possible via this region.

The F-layer is composed of two very well-defined areas, called F1 and F2. The F1-layer sits at about 150 to 250 kilometers altitude during the day. Although the ionization level is higher, The F1-layer exhibits much the same sort of behavior as the E-layer, i.e., maximum ionization at midday and almost complete disappearance of the layer after sunset.

The F2-layer generally is situated above the F1-layer, and has several interesting features. The upper boundary of F2 varies seasonally, with a height of 350 kilometers in winter, and up to 500 kilometers during the summer months, during the daylight hours. At night time, the F2-layer tends to "come down" to about 250 to 420 kilometers. This obviously will affect skip distance and communications over a selected path depending on the hour of the day.

In contrast to the other layers, the F2-layer exists 24 hours a day. Due to its height and air density, ion-electron recombination occurs at a relatively leisurely pace, in contrast to the abrupt changes seen in the other layers. Bending and reflection of radio waves is very efficient in this layer, and its existence around the clock makes it very attractive for long-distance communications.

1.4 IONOSPHERIC VARIATIONS

We've seen that the behavior of the ionospheric layers exhibit variations diurnally and seasonally. We will briefly mention the reasons for these effects and for geographical variations. There are also cyclic variations that are due to the sunspot cycle; these will be discussed in a separate section.

Diurnal variations

The time-of-day variations obviously affect the amount of ionizing radiation present in the ionosphere at a specified moment, and hence the level and efficiency of ionization. A typical 24-hour variation of critical frequency in the F2 layer is shown in Figure 3.

Seasonal variations

The variation in ionospheric properties from season to season are due mostly to the position of a point on the earth relative to the sun, and to the length of the days. Those long winter nights provide more time for the F2-layer, say, to forget its electrical properties; hence, we will see lower critical frequencies during winter nights. During the winter day, ionization levels are high, and we see higher critical frequencies.

During the summer months, however, the days are long and the sun comes up over the horizon more slowly than in winter. Consequently, the ionization level builds up more slowly, so that during the summer days we will expect critical frequencies to be lower than the corresponding winter day values. Since the ionization efficiency is lower, ion-electron recombination occurs at a slower rate, causing the F2-layer to "last longer", even into the night, and with higher critical frequencies than in winter.

Geographical variations

It is well known that the intensity of radiation striking the earth varies directly with the elevation angle of the sun. The maximum occurs in the equatorial regions, and decreases in proportion to the latitude north or south of the equator. In general, we expect higher ionization activity in the equatorial regions, and consequently higher critical frequencies.

1.5 IONOSPHERIC LAYER MODELS

Now that we have discussed the basic structure of the ionospheric layers, let us briefly take note of some of the theoretical models that are generally used to describe them analytically.

The ionosphere is a medium whose refractive and reflective properties can be difficult to characterize. Much work in this area began in the 1920's, the most significant being that of Sir Edward Appleton and D.R. Hartree. They assumed an electromagnetic wave to be simple harmonic, progressive and plane, with a well-defined polarization. The ionosphere was considered to be electrically neutral, of uniform charge distribution, polarizable and electrically anisotropic. Maxwell's Equations are solved first, and then the properties of the magneto-ionic medium are imposed. [1]

The Appleton-Hartree formula for the (complex) refractive index is

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1-X-iZ)} \pm \sqrt{\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2}} \quad (5)$$

where $X = \frac{Ne^2}{\epsilon_0 m \omega^2}$

$Y_L = \frac{eB_L}{m\omega}$

$Y_T = \frac{eB_T}{m\omega}$

$Z = \nu/\omega$

with the subscripts T and L referring to the transverse and longitudinal components of the imposed magnetic field. This equation has been studied exhaustively for most possible cases (see Davies [1]). The main point, however, is that the ionosphere is a complex dispersive medium whose properties will be studied in more detail in Part 2, as it relates to HF propagation.

The most simplistic model of an ionized layer amenable to analytical methods is called the Chapman layer. This model makes the assumptions of

1) Atmosphere composed of only one kind of gas
2) Plane stratification
3) Parallel beam of monochromatic ionizing radiation from the sun
4) Isothermal atmosphere

The basic geometry is indicated in Figure 4. We assume radiation of intensity S_∞ incident upon the outer boundary of the atmosphere, incident at a zenith angle χ . The zenith angle is easily defined in

terms of the geographic latitude ϕ , solar declination δ , and local solar hour angle h .

As the wave penetrates into the ionized layer, absorption occurs due to ionic collisions, and the wave is attenuated. Assuming a number density N and an absorption cross section σ, the energy absorbed in a cylindrical volume of unit cross section in the direction of the incident radiation is

$$dS = S\sigma N dh \sec\chi \tag{6}$$

Integrating,

$$\int \frac{dS}{S} = \sec\chi \int_{\infty}^{h} N\sigma\, dh = -\tau \sec\chi \tag{7}$$

where

$$\tau = -\int_{\infty}^{h} N\sigma\, dh \tag{8}$$

is the optical depth of the ionosphere to the altitude h.

Assuming an ionization efficiency η, the number of ion pairs produced per unit volume per second is

$$q(\chi, h) = N\sigma S_{\infty}\eta \exp(-\tau \sec\chi) \tag{9}$$

After some mathematical manipulation, it can be deduced that

$$q(z_0)(z - h \sec\chi) = \sec\chi\, q(\chi, z) \tag{10}$$

which provides us with scaling information. $q(\chi, z)$ has the same shape as $q(0, z)$, but displaced up by $h \sec\chi$ and decreased as $\cos\chi$. Using a logarithmic scale, all the q curves have the same shape, as in Figure 5.

Peak production occurs at height

$$z_m = h \sec\chi$$

or

$$h_m = h_0 + H \ln \sec\chi \tag{11}$$

and the peak production rate is

$$q_m = q_0 \cos\chi \tag{12}$$

Note that height of peak production is not affected by changes in the incident flux. The complete derivation is available in Davies. [1]

Once the ion production rate is known, one then assumes a process of electron disappearance (recombination or attachment) in order to derive the corresponding electron density distribution. [1]

For a recombinative process

$$\frac{dN}{dt} = q - \alpha N^2 \tag{13}$$

and for an attachment type process,

$$\frac{dN}{dt} = q - \beta N \qquad (14)$$

Several models are often used as approximations to the Chapman layer; the most notable among them are:

- Parabolic layer
- Linear layer
- Exponential layer

In each case, an electron density distribution of the general form

$$N = N_o \exp \frac{1}{2}\left[1 - z - \sec\chi \exp(-z)\right] \qquad (15)$$

is assumed. Further refinements to this distribution are invoked to describe a particular portion of the ionospheric layer, eg, the edge of the layer, the center, etc.

1.6 EFFECTS OF THE SUNSPOT CYCLE

In addition to the daily, seasonal and geographical variations in radio wave propagation that we have seen earlier, there is also a yearly variation. This behavior appears to be related to the number of sunspots that exist on the surface of the sun at any given time. [2]

What exactly IS a sunspot ? Observationally, it looks like a dark (usually greyish-black to black) blotch, quite different from the remainder of the solar surface. They rarely occur singly; they usually are found in groups or small clusters. To a visible observation, they also appear to move across the solar disk. This apparent motion is due to the solar rotation, which has a period of about 27.5 days, and is generally viewed as an east to west travel. The size of a sunspot can range from several hundred to 80,000 miles in diameter.

A sunspot is essentially a mass of gaseous plasma in a highly eruptive and stormy state, being moved around in a swirling fashion. There are very strong magnetic fields associated with these regions (4500 gauss). As the plasma moves about, it is cooled to a temperature (4400 degrees K) less than the surrounding area (6100 degrees K) and hence a darker area appears. The confining nature of the strong magnetic fields effectively inhibits convective energy transfer from the surrounding regions, keeping the temperature lower. The underlying physical mechanisms of the solar "dynamo" [14] still provide fertile areas of research for solar scientists.

Sunspot activity seems to peak and ebb in an approximate 11-year cycle. At the peak of a cycle, there are large numbers of spots, and the ionosphere displays strong electrical activity. At a cycle minimum, only a few spots exist and the ionosphere tends to lose it electrical "spunk", and consequently its ability to effectively refract and reflect radio waves. Just ask any radio amateur operator who chases DX to describe the difference in communications path reliability and signal quality between the top and bottom of a sunspot cycle !!

Recorded observations of sunspot activity began with Galileo in 1611. Keeping records of this activity on a regular basis is credited to the German pharmacist Hendrik S. Schwabe in the mid 1700's, and he is considered to have discovered the sunspot cycle. Around 1749, Rudolf Wolf, then the director of the Zurich Solar Observatory, defined a standard method for viewing sunspot activity, taking into account differences in the size and type of telescope used and local viewing conditions.

The Wolf relative sunspot number is based on daily observations, and is defined by [2]

$$R = k\,(10\,g + f) \qquad (16)$$

where

g = number of observed sunspot groups

f = total number of observed sunspots (individual or in groups)

k = observer difference factor

Since solar activity can fluctuate wildly from day to day, trends can be hidden. To better observe monthly and yearly trends, we also define a "smoothed" sunspot number. One such indicator is the monthly mean relative sunspot number, R_m, which is the monthly average of the daily numbers. [2]

Further smoothing occurs by averaging the monthly numbers into a 12-month running average R_s : [2]

$$R_s = \frac{\frac{1}{2}R_{m1} + R_{m2} + R_{m3} + \cdots + R_{m12} + \frac{1}{2}R_{m13}}{12} \qquad (17)$$

which is centered on R_{m7} .

Plotting this number over many years brings out the cyclic nature of sunspots, and it correlates quite well with observed ionospheric conditions. This plot is presented in Figure 6.

Cycle 1 was defined by the Zurich group as having begun in 1755. We are now at the end of Cycle 21. Current wisdom and observation indicates that this cycle bottomed out early in 1986, and that we are

now embarking on Cycle 22, moving upward toward the high activity peak part of the cycle. Jacobs and Cohen [2] give an excellent breakdown of Cycles 19 and 20.

1.7 PARAMETERS USED TO CHARACTERIZE SOLAR ACTIVITY

Although sunspot-watching can be an interesting activity in itself, there are other more accurate and meaningful indicators of solar activity that we can use to try to predict what radio propagation conditions will be like at a given time. These indices are both more objective and more readily available.

The SOLAR FLUX is an excellent index, being a measure of the amount and intensity of the radiation reaching the earth from the sun. It was found in WWII that the noise amplitude at 2695 MHz, due to solar flux, closely correlated with the ionization density found in the F-layers. Since 1947, this flux has been measured by a Canadian observatory in Ottawa. The relationship between solar flux and the daily sunspot number at 2695 MHz (10.7 cm) is almost linear: [2]

$$SF = 63.7 + 0.73\,R + 0.0009\,R^2 \qquad (18)$$

as derived by Stewart and Leftin. [4]

The solar flux will contain particle radiation, such as electrons and protons, which can induce a current, and hence geomagnetic fields in the ionosphere. A measure of the geomagnetic activity is contained in the equivalent range index, or A-index for short, which is a value

determined daily at a mid-latitude site, such as Fredericksburg, VA or Anchorage, Alaska.

Another index, the K-index, is the value of geomagnetic activity taken at 3-hour intervals, and shows us a wide range of activity in a single digit. The A-index is a value obtained at one site only. There is actually a network of data-collecting stations worldwide, and their values are averaged to give the planetary 3-hour index (a_p) and the daily planetary index A_p.

The A- and K-indices are inter-related as [5]

A	K
0	0
3	1
7	2
15	3
27	4
48	5
80	6
140	7
240	8
400	9

The generally accepted standards for propagation prediction purposes are defined as follows:

QUIET	A <= 7	K <= 2
UNSETTLED	7 < A < 15	K <= 3
ACTIVE	15 < A < 30	K <= 4
MINOR STORM	30 < A < 50	K = 4 OR 5
MAJOR STORM	A >= 50	K > 6
SEVERE STORM	A >= 100	K > 7

The current values for solar flux and the A- and K-indices are reported hourly by WWV. For the do-it-yourself propagation predictor, there are also several prediction charts (QST, CQ) and publications available, such as the weekly *PRELIMINARY REPORT AND FORECAST OF SOLAR GEOPHYSICAL DATA* issued by SESC/NOAA.[6] Other services exist worldwide. [2]

1.8 SOLAR FLARES, DISTURBANCES AND RADIO STORMS

The ambient background noise level from solar emissions may stay relatively constant, until interrupted by some sort of disturbance on the sun, which may occur suddenly, producing a sudden ionospheric disturbance (SID), or which might develop in a definite gradual pattern. Such solar disturbances can have a direct effect on ionospheric conditions, ranging from momentary fading to complete radio blackout. The major effect is to decrease the signal to noise ratio, sometimes so much so that a signal can disappear altogether. Or it may be noticed as sporadic fading. Disturbances may last several hours to several days.

It is believed the major part of this activity is attributable to solar flare activity. A flare is a somewhat explosive outburst that occurs in the solar chromosphere in the vicinity of a sunspot, hurling tremendous amounts of plasma outward from the solar surface, along with prodigious amounts of ultraviolet, X-ray and cosmic radiation.

The radiative components generally take only 8 or 9 minutes to reach the earth after the inception of a flare, leading to an SID. If the point on the solar surface where the flare occurs is pretty much in line with the earth, then the radiative components penetrate the higher ionospheric layers, causing a dramatic increase in the ionization level of the D-layer, with significantly increased signal absorption. In this case, the lower radio frequencies are affected before the higher ones. At this particular moment, the dark side of the earth would not be affected at all.

The heavier charged particle contingent, however, may take 18 to 36 hours to arrive. This onslaught is usually referred to as an ionospheric or radio "storm". Particles approaching the earth are deflected by the geomagnetic field toward the magnetic poles. The resultant particle drift is more rapid and pronounced in the upper layers of the atmosphere, causing the F-layers, and consequently the higher HF frequencies, to be affected first. The F-layers may lose the ability to reflect most signal frequencies that are normally possible; the frequency range may be reduced up to half the normal range. Ionization levels in the F-layers may appear to changing rapidly. The intensity of a storm could be so severe as to weaken the ionosphere to the point of total radio blackout, especially in the polar regions, where charged particle density is at a maximum. The ionospheric storm does not respect the dark side of the earth boundary.

Cosmic radiation, at high intensity levels, may cause severe polar cap absorption (PCA). [8] Cosmic rays are also deflected by the geomagnetic field toward the polar regions. If the intensity is high enough, the upper ionospheric layers will be penetrated, and most of the ionization and absorption activity will take place in the lower layers. This may lead to polar blackouts with durations of several minutes to several days. PCA will affect only those signals passing through the polar regions.

Solar flares are classified according to size, brightness, X-ray class and type of activity (Figure 7),and can persist up to an hour in length. Statistically, it has been found that the number of flares per solar rotation is related to the mean sunspot number by [2]

$$N_F = a (R - 10) \qquad (19)$$

Noise storms are classified on the basis of frequency and wavelength of emitted radiation and the time after flare inception, as shown in Figure 8, which is the NOAA/SESC standard classification.

1.9 MAXWELL'S EQUATIONS AND THE NATURE OF RADIO WAVES

The basic equations governing the behavior of radio waves are Maxwell's equations of electromagnetism:

$$\begin{aligned} \vec{\nabla} \cdot \vec{D} &= \rho \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} &= 0 \\ \vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} &= \vec{J} \end{aligned} \qquad (20)$$

which are treated exhaustively in most any textbook on electromagnetic theory. [7]

The electric (E) component of an EM wave is the most important one as far as radio waves are concerned. As a radio wave goes past a point, the E-field at that point varies with the frequency of the wave. The term polarization is used to describe how E varies in both magnitude and direction. A plane electromagnetic wave has E & H vectors that are mutually perpendicular, with vector E X H describing the direction of propagation. Such a wave is said to be plane-polarized. If the E-field magnitude stays constant but its direction changes, we have circular polarization. When both magnitude and direction change, we have the general case of elliptical polarization.

With respect to antennas, we can speak of them as having vertical or horizontal polarization, depending on the orientation of the E-field

with respect to the earth's surface. Most vertical type antennas have E perpendicular to the ground, and hence have vertical polarization. Most dipole antennas, Yagi-Uda arrays, Log-periodic arrays, etc, have E parallel to the ground, and thus show horizontal polarization. The choice of antenna can often be an important factor in obtaining the desired propagation path characteristics.

In the electromagnetic spectrum, radio waves are classified by frequency as:

VLF	- Very low frequency	- 10 - 100 kHz
LF	- Low frequency	- 100 - 300 kHz
MF	- Medium frequency	- 300 kHz - 3000 kHz
HF	- High frequency	- 3.0 - 30 MHz
VHF	- Very high frequency	- 30 - 300 MHz
UHF	- Ultra high frequency	- 300 - 3000 MHz
SHF	- Super high frequency	- 3 - 30 GHz
EHF	- Extremely high frequency	- 30 - 300 GHz

1.10 BASIC RADIOWAVE PROPAGATION MODES

The *SPACE WAVE* is usually considered to be composed of two waves [7] -- a direct wave and a ground reflected wave (Figure 9a). This mode of travel is useful only when transmitter and receiver are in a direct line-of-sight relative to each other, and have no obstructions between them. This is the primary mode of propagation at VHF and above.

The *SURFACE* or *GROUND WAVE* [8,9] moves very closely to the surface of the earth and indeed the ground may act as part of the transmission circuit. This type of wave tends to follow the curvature and contours of the earth (Figure 9b), and can be used between two stations over the horizon from each other. This wave disrupts the magnetic field at the surface, inducing a current in the earth's surface. This usually leads to a power loss in the signal; after enough distance has been covered, the signal would eventually be completely diminished by this power loss.

A surface wave is greatly influenced by the nature of the surface over which it travels. Over a good conductor like seawater, surface absorption is quite minimal, and a wave can travel great distances before becoming significantly attenuated. The surface wave is useful for frequencies below about 500 kHz. This mode of propagation is very popular for maritime communications and navigation, which generally operate in the range 10-100 kHz.

The predominant mode of travel in the range 3-30 MHz is the *SKY WAVE* or *IONOSPHERIC WAVE*. The mechanism for radiowave travel is by the bending (refraction) and reflection of the wave from the ionosphere, bringing the wave back to earth. Further details of this mechanism will be discussed in part 2 of this paper. (Figure 9c)

The *SCATTER WAVE* involves forward (or even backward) scattering of a wave from either the troposphere or from an object such as a meteor trail. This mode is not all that common, and will be looked at again under the heading of exotic propagation modes.(Figure 9d)

1.11 REFERENCES

1) Davies, Kenneth, *IONOSPHERIC RADIO PROPAGATION*, Dover, 1966.

2) Jacobs, G. and Cohen, T., *THE SHORTWAVE PROPAGATION HANDBOOK*, 2nd Edition, CQ Publishing, Inc., 1982

3) Tverskoi, P.N., *PHYSICS OF THE ATMOSPHERE - A COURSE IN METEOROLOGY* , US Dept. of Commerce, 1965

4) Stewart, G. and Leftin, M., *RELATIONSHIP BETWEEN OTTAWA 10.7 cm SOLAR RADIO NOISE FLUX AND ZURICH SUNSPOT NUMBER*, Telecommunications Journal, Vol. 39, No. 3, 1972

5) US Dept. Of Commerce, *NOAA, DESCRIPTIVE TEXT - CONTENTS OF PRELIMINARY REPORT AND FORECAST OF SOLAR-GEOPHYSICAL ACTIVITY*, 1 MAY 1984.

6) US Dept. Of Commerce, NOAA, *CONTENTS OF PRELIMINARY REPORT AND FORECAST OF SOLAR-GEOPHYSICAL ACTIVITY*, (weekly)

7) Corson,D. and Lorrain, P., *INTRODUCTION TO ELECTROMAGNETIC FIELDS AND WAVES*, W. H. Freeman and Company, 1962.

8) Wiesner, L., *TELEGRAPH AND DATA TRANSMISSION OVER SHORTWAVE RADIO LINKS - FUNDAMENTAL PRINCIPLES AND NETWORKS*, 3rd Edition, Siemens AG/John Wiley and Sons, 1984

9) Collin, R.E., *ANTENNAS AND RADIOWAVE PROPAGATION*, McGraw-Hill, 1985.

10) American Radio Relay League, *THE ARRL ANTENNA BOOK*, Fourteenth Edition, 1982.

11) McLean, D.J. and Labrum, N.R., *SOLAR RADIOPHYSICS*, Cambridge University Press, 1985

12) Budden, K.G., *THE PROPAGATION OF RADIO WAVES*, Cambridge University Press, 1985

13) Shibuya, S., *A BASIC ATLAS OF RADIOWAVE PROPAGATION*, John Wiley and Sons, 1987

14) Newkirk, G. and Frazier, K., "The Solar Cycle", *PHYSICS TODAY*, April 1982

15) Gibilisco, S., "Propagation of Radio Waves", *HAM RADIO*, August, 1982

16) Jansz, E., "Radio Wave Propagation", *HANDS-ON ELECTRONICS*, June, 1985

17) Bixby, C.L. and Morris, J., "The Art and Science of DXing", *QST*, January, 1979.

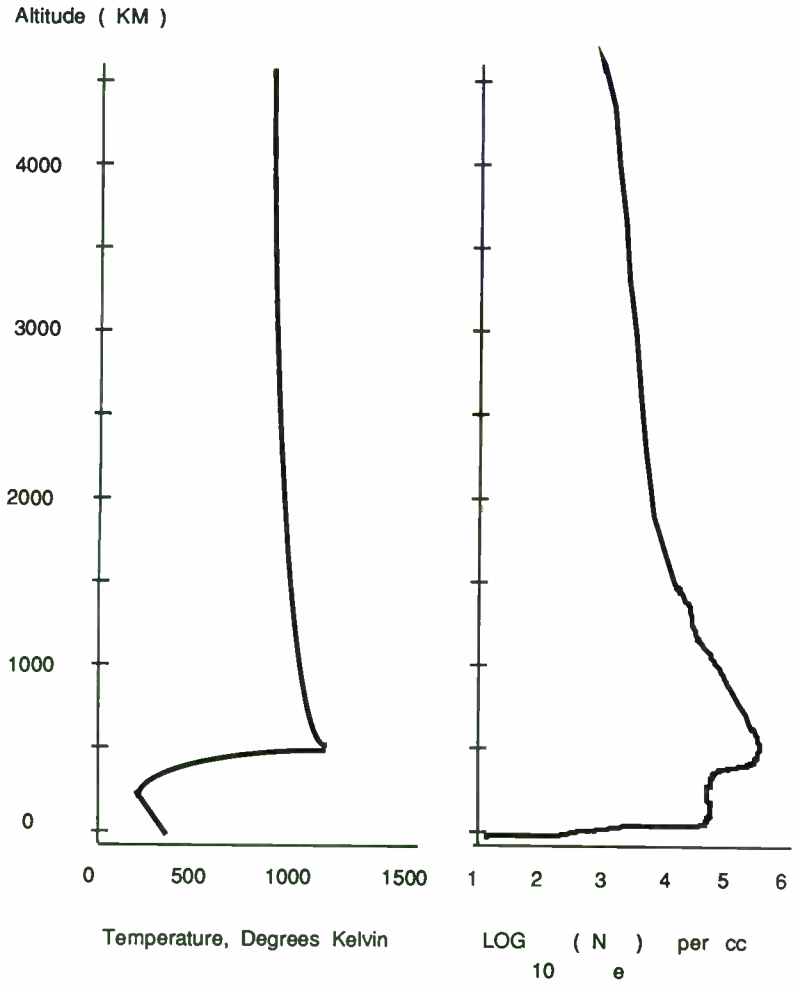


Figure 1. Temperature and Electron Density Profile of the Earth's Atmosphere

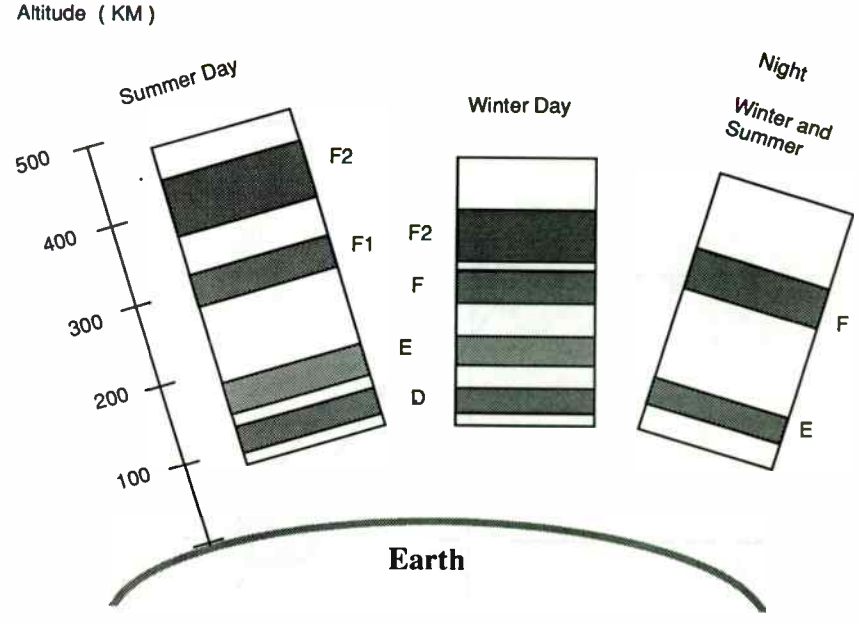


Figure 2. Ionospheric Layer Structure

(Daily and Seasonal Variations)

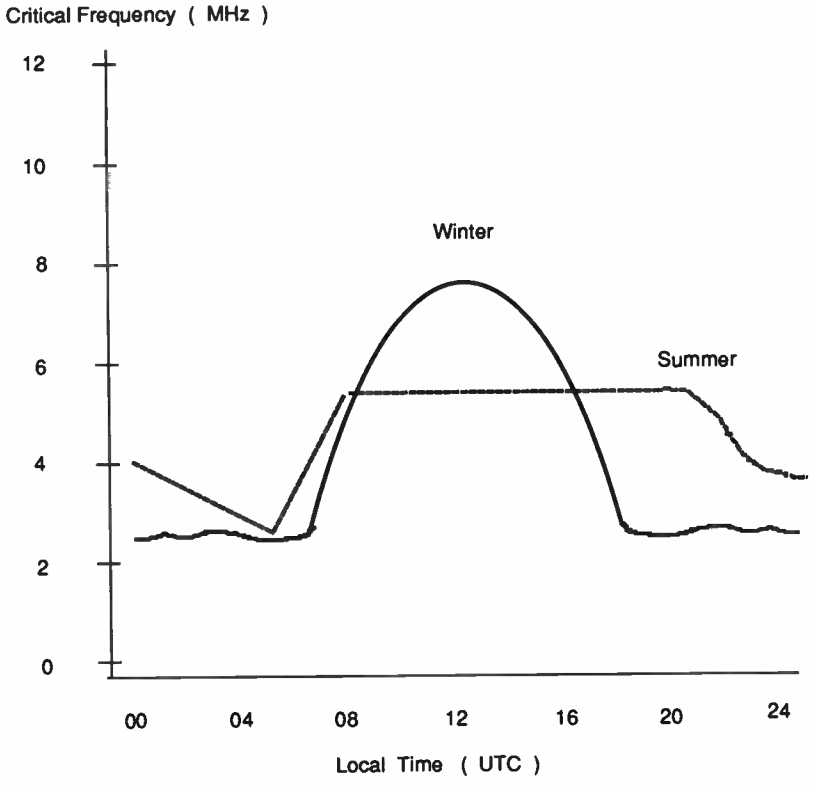


Figure 3. Typical Variation of Critical Frequency in the F2 - Layer

(Daily and Seasonal Variations)

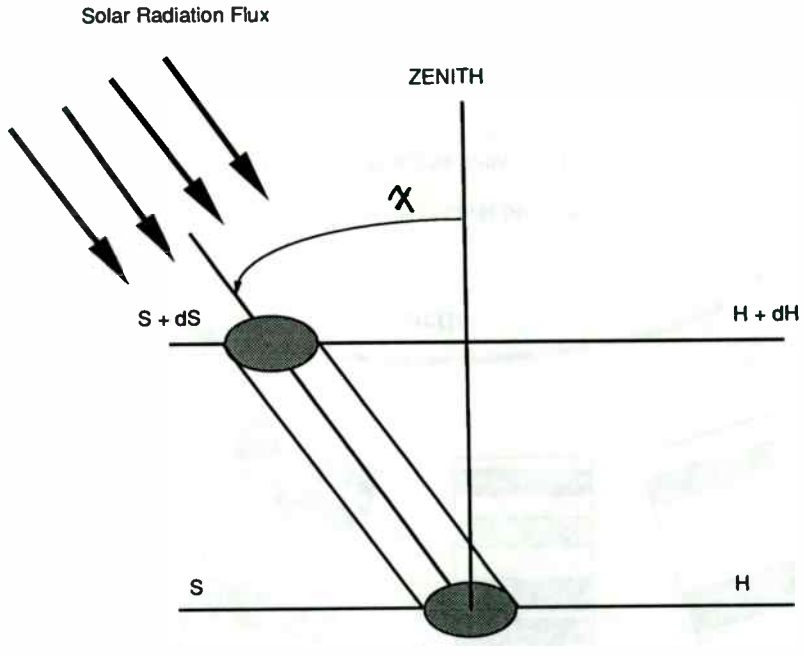


Figure 4. Geometry for Formation of Chapman Layer

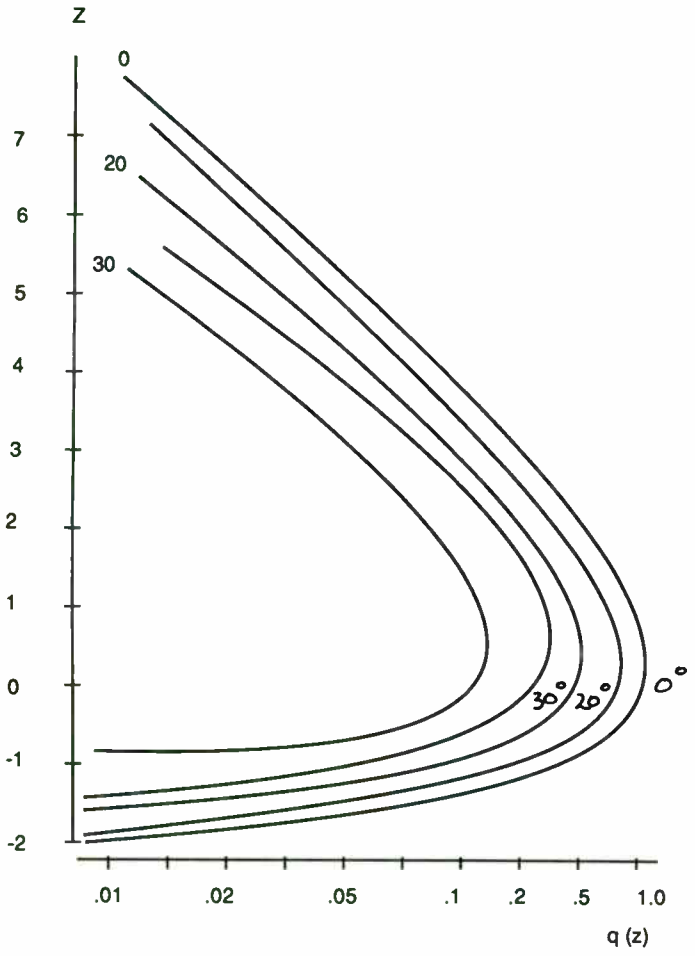


Figure 5. Normalized Photoionization Rate
(Chapman Theory)

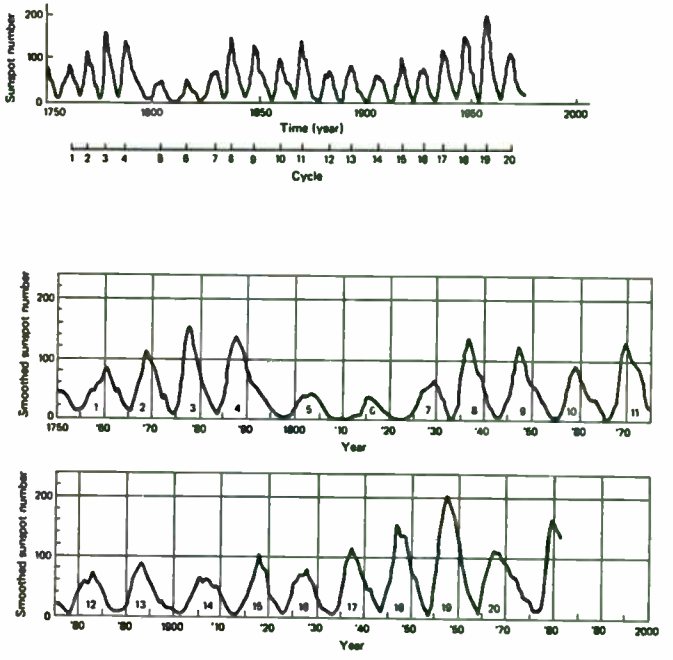


Figure 6. Twelve-Month Running Smoothed Sunspot Numbers since the Year 1750

(From Reference 2 , Jacobs and Cohen,
"The Shortwave Propagation Handbook")

SESC X-RAY CLASSIFICATION SCHEME

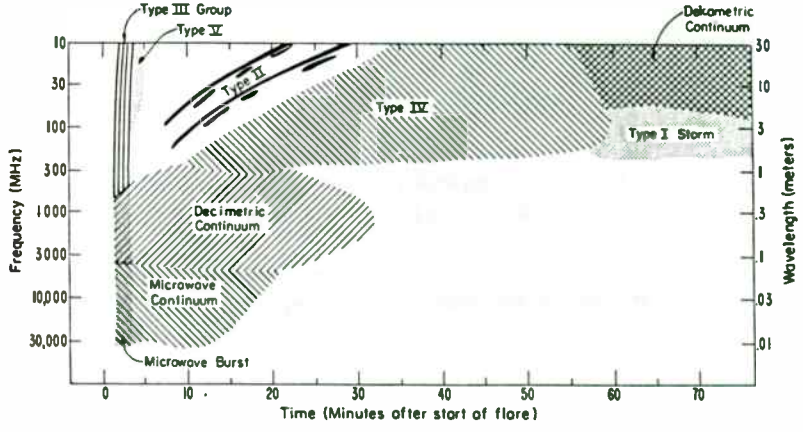
Class	Peak, 1-8Å (Φ W m^{-2})	Peak, 1-8Å (Φ erg $cm^{-2}s^{-1}$)
A	$\Phi < 10^{-7}$	$\Phi < 10^{-4}$
B	$10^{-7} < 10^{-6}$	$10^{-4} < 10^{-3}$
C	$10^{-6} \leq \Phi < 10^{-5}$	$10^{-3} \leq \Phi < 10^{-2}$
M	$10^{-5} \leq \Phi < 10^{-4}$	$10^{-2} \leq \Phi < 10^{-1}$
X	$\Phi \geq 10^{-4}$	$\Phi \geq 10^{-1}$

TYPE: Type of activity.
- FLA - Solar flare
- EPL - Eruptive prominence on the solar limb
- BSL - Bright surge on the limb
- SPY - Spray (including ejecta)
- LPS - Loop Prominence System
- DSF - Disappearing solar filament
- INC - Incomplete; no optical observation but the source region is known
- UNK - Unknown; no optical observation and source region unknown

IMP: Importance.
- S - Subflare (area ≤ 2.0 square degrees)
- 1 - Importance 1 (2.1 ≤ area ≤ 5.1 square degrees)
- 2 - Importance 2 (5.2 ≤ area ≤ 12.4 square degrees)
- 3 - Importance 3 (12.5 ≤ area ≤ 24.7 square degrees)
- 4 - Importance 4 (24.8 ≤ area in square degrees)

A brightness qualifier F, N, or B is usually appended to the importance character to indicate Faint, Normal, or Brilliant.

Figure 7. NOAA / SESC SOLAR FLARE CLASSIFICATION SCHEME



Type II: Slow drift burst
Type III: Fast drift burst
Type IV: Broadband smooth continuum burst
Type V: Brief continuum burst generally associated with Type III bursts
Type CTM: Long-lived dekametric continuum associated with storms of Type IV bursts
Numeral: Relative importance on a scale of 1 to 3. An importance-1 sweep event would be considered small or minor. An importance-3 event is a major or significant event.

Figure 8. NOAA / SESC NOISE STORM CLASSIFICATIONS

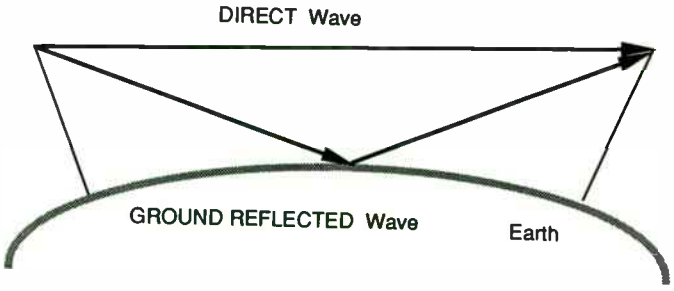


Figure 9a. Space Wave Propagation

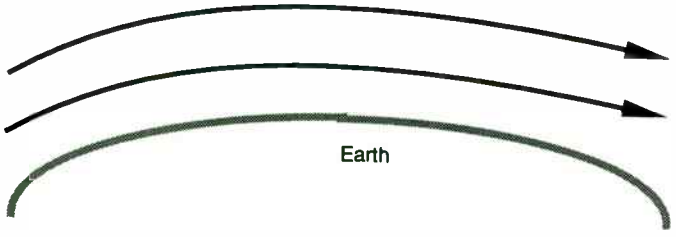


Figure 9b. SURFACE Wave Propagation

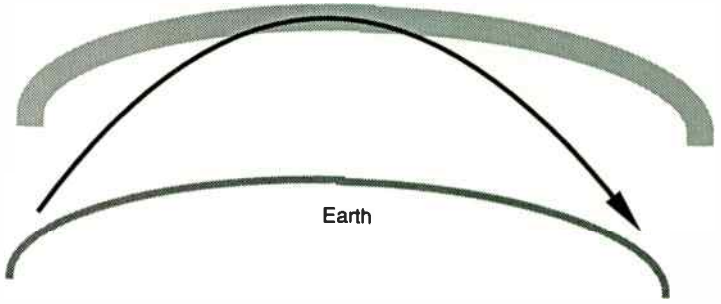


Figure 9c. Ionospheric (SKY) Wave

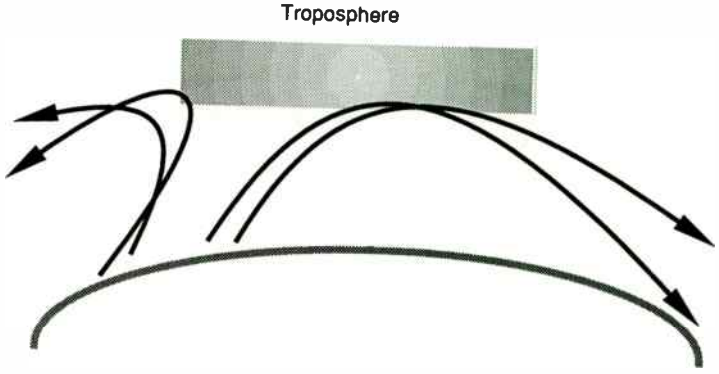


Figure 9d. Scatter Wave Propagation

RADIO-WAVE PROPAGATION - A TUTORIAL

PART 2 - PROPAGATION AT HIGH FREQUENCIES

D. R. Dorsey, Jr
Principal Systems Analyst

Network Development Division
Control Data Corporation
Sunnyvale, CA 94089

2.0 INTRODUCTION

In part 1 of this tutorial, we studied the structure and characteristics of the earth's atmosphere and ionosphere. We now move on to consider radio wave propagation at the "high" frequencies (3 - 30 MHz), where the influence of the ionosphere is most pronounced.

First, the propagation of an electromagnetic wave in an ionized gas is considered. Then we look at the case of oblique propagation of radio waves, the dominant mechanism of propagation at HF. This case is extended to include the effects of earth and ionospheric curvature, and its effects on the prediction of propagation conditions and signal quality. Last but not least, several "exotic" modes of propagation are introduced.

2.1 PROPAGATION OF A PLANE ELECTROMAGNETIC WAVE IN AN IONIZED GAS.

In an ionized gas, currents are generated by the presence of free electrons and ions in the path of an incident electromagnetic wave. A free ion in motion is subjected to both Coulomb and Lorentz forces. Assuming a low pressure gas in which there are few ionic collisions, ionic motion and the conductivity of the ionized gas are due almost entirely to the E-field. Consequently, the Appleton-Hartree formula with no magnetic field reduces to [1]

$$n^2 = (\mu - i\chi)^2 = 1 - \frac{X}{1-iZ} = 1 - \frac{X}{1+Z^2} - \frac{iXZ}{1+Z^2} \qquad (1)$$

where

$$X = \frac{Ne^2}{\epsilon_0 m \omega^2}$$

$$Y_L = \frac{eB_L}{m\omega}$$

$$Y_T = \frac{eB_T}{m\omega}$$

$$Z = \nu/\omega$$

with the subscripts T and L referring to the transverse and longitudinal components of the imposed magnetic field. At this point, the refractive index is still complex.

From basic electromagnetic theory, we recall that the phase velocity of an electromagnetic wave is

$$v = f\lambda = 1/\sqrt{\mu\epsilon} \qquad (2)$$

and the corresponding free space speed of light is

$$c = 1/\sqrt{\mu_0\epsilon_0} \qquad (3)$$

where refractive index $= n = c/v$

f = Frequency (Hz)
ϵ, ϵ_0 = permittivity of medium (free space)
μ, μ_0 = magnetic permeability of medium (free space)
λ = free space wavelength
ω = angular frequency = $2\pi f$
k = wave number = $2\pi/\lambda$

A wave travelling in the x_1-direction, say, is of the form

$$E = E_0 \cos(\omega t - kx_1)$$
$$E = E_0 \exp i(\omega t - kx_1) \qquad (4)$$

Using (1), (2) and (3), this can be re-written

$$E = E_0 \exp i\left(\omega t - \frac{\omega}{c} n x_1\right)$$

or

$$E = E_0 \exp\left(-\chi \frac{\omega}{c} x_1\right) \exp i\left(\omega t - \mu\frac{\omega}{c}\right)x_1 \qquad (5)$$

For $\chi \neq 0$, this is a wave having an exponential decrease of amplitude with distance, ie, absorption with an absorption coefficient $K = \omega\chi/c$.

Then from (1),

$$K = \frac{\omega}{c}\frac{1}{2\mu}\frac{XZ}{1+Z^2}$$

or

$$K = \frac{e^2}{2\epsilon_0 mc}\frac{1}{\mu}\frac{N\nu}{\omega^2+\nu^2} \qquad (6)$$

showing clearly that absorption is frequency dependent.

In the upper part of the atmosphere, the collision frequency is very small at frequencies > 1 MHz. Therefore in Equation 5 of Part 1, we can assume Z = 0, so that

$$\mu^2 = 1 - \frac{2X(1-X)}{2(1-X) - Y_T^2 \pm \sqrt{Y_T^4 + 4(1-X)^2 Y_L^2}} \qquad (7)$$

For a given value of X, a large collision frequency tends to diminish the refractive properties of the medium. Also, when collisions are present, the real part of (7) never vanishes entirely; at X = 2, it has a minimum value

$$\mu_m^2 = 1 - \frac{1}{(1+Z^2)} \qquad (8)$$

Since Z, and hence the refractive index μ, vary with frequency, we are then led to the phenomenon of *DISPERSION* in the ionosphere, similar to the dispersion displayed in optics. Refractive effects are always present in the ionosphere, and will vary from point to point. If the change in refractive index over a given distance is sharp enough, it may appear that we are witnessing a "total reflection" instead of a refractive phenomenon.

Under the influence of the wave's electric field, an ion then begins to oscillate, in simple harmonic motion, at an angular frequency [2]

$$\omega_p = \sqrt{\frac{N_e Q_e^2}{\epsilon_o m_e}} \qquad (9)$$

also termed the plasma angular frequency, corresponding to a frequency of

$$f_p = 8.98\sqrt{N_e} \qquad Hz/sec \qquad (10)$$

which depends only on the properties of the gas considered. Also, a resonance condition exists at this frequency that causes the total current, containing convective and displacement components, to become zero. The ionic oscillation produces a new electric field out of phase with the exciting wave's field. A phase displacement is created, and the phase velocity is [2]

$$v = \frac{c}{\sqrt{1-(\omega_p/\omega)^2}} \qquad (11)$$

implying distinctly a refraction of the wave.

2.2 OBLIQUE PROPAGATION

If the ionized gas had a definite boundary and a uniform electron density throughout its volume, then refraction and reflection at its interface would be very easy to describe. We could simply consider the ionized gas as a dielectric medium with an index of refraction less than unity, and apply the usual laws of reflection, refraction

(Snell's Law) and the Fresnel equations directly. [2] But since the ionosphere is inhomogeneous, the electron density, and hence conductivity, varies from point to point, forcing us to solve a nonlinear wave equation in the general case.

However, we can make some simplifying assumptions to ease the analysis. For the moment we will neglect earth curvature effects, and assume that the index of refraction, as well as the electron density varies slowly with altitude.

Consider first the case of a plane wave at normal incidence to the layer, as in Figure 1. As the altitude and electron density increase, so does the absorption. At some point, $K = 0$ and wave propagation stops, being reflected back to earth. [3] When $K = 0$, the wave frequency just equals the plasma frequency. The altitude at which this occurs is the idealized "*VIRTUAL HEIGHT*".

Now referring to Figure 2, consider a wave entering the ionized layer from below at incidence angle Ψ_i .[3] Consider the medium as stratified into infinitely thin layers in which the refractive index varies continuously. As we saw in Part 1, the electron density and refractive index increase, reach a peak value and then decrease again with increase in altitude.

Snell's law is written

$$n_1 \sin\Psi_1 = n_2 \sin\Psi_2 \quad \text{or} \quad \sin\Psi_i = \sqrt{K(z_n)}\,\sin\Psi_n \qquad (12)$$

so the quantity $n \sin\Psi$ must be conserved across each idealized layer interface. As refractive index increases, the wave is bent more away from the normal, until at the highest altitude $\Psi(z) = \pi/2$ at which point the wave is horizontal. For a given angle Ψ_i , a higher electron density is required, at higher frequencies, to refract the wave further, back earthward. Also, for a given electron density maximum, there is a maximum value of Ψ_i which will return the wave to earth. Hence, there exists an upper limit on the frequency that can be returned to earth. The critical value of electron density [3] is

$$N_c = \left(f^2 \cos^2\Psi_i\right) / 81 \qquad (13)$$

We saw that for normal incidence the wave was reflected back to earth at a value of electron density that caused $K = 0$. Equivalently, for the oblique incidence case, there is a value N_{max} that yields $K=0$. This corresponds to the critical frequency f_c :

$$f_c = 9\sqrt{N_{max}} \qquad (14)$$

Therefore, from (13) we have

$$f = 9\sqrt{N_{max}}\ \sec\varphi_i = f_c \sec\varphi_i \qquad (15)$$

which is the well-known " *SECANT LAW* ". Since this is the frequency at which a wave will just return to earth, it is termed the *MAXIMUM USABLE FREQUENCY*, or *MUF*, when φ_i is at its maximum value. Most propagation predictions are geared toward an accurate determination of the MUF.

The influence of incidence angle is illustrated in Figure 3. [5] Waves 1 and 2, of grazing incidence, suffer no reflections at all and continue on to outer space. Wave 3 enters the ionosphere at its lower edge, where the refractive index is slight, but finite; the refraction is gradual, but sufficient to eventually return the wave to earth. Wave 4 enters at an angle just slightly higher than Wave 3, and exits sooner due to the higher refractive index. Wave 5 enters at an angle such that it is refracted to the horizontal near the center of the layer, where the index of refraction remains relatively constant. Thus, it may stay within the layer for quite a long distance before it is finally refracted out of the layer; this phenomenon is often referred to as *DUCTING*. Wave incidence angle can often be controlled at the transmitter site, using horizontally polarized antennas (such as the Yagi-Uda array), which are easily designed for specific radiation patterns, beamwidth and beam take-off angle.

Figure 4 illustrates the frequency dependence. Refractive effects diminish with increase in frequency, and above the critical frequency, a radio wave simply continues on into space.

The distance between the transmitter site and the point where the wave finally returns to earth is called the *SKIP DISTANCE*. In between is a dead zone, where no signal is received. Radio waves obey an inverse square intensity law as they propagate; in addition, ionospheric and other propagation loss factors contribute to a degradation in signal quality over a given communications path.

Naturally, we would prefer our signal to arrive at its intended destination in the minimum possible distance traversed. The case just discussed is known as a *SINGLE-HOP* path. When the wave returns to earth, it may be re-reflected from the earth back towards to repeat the process over again, as shown in Figure 5, perhaps even several times. Of course, signal quality will degrade further on each hop, and could disappear altogether before reaching its intended destination. Such a path is referred to as a *MULTI-HOP* path. Skip distances can change moment to moment with ionospheric variations, leading to sporadic fading and other maladies that impede reliable communications.

Amateur radio operators are very aware of such conditions, especially in DX work. A contact between two points would ideally be conducted over a path affording single-hop propagation. Using a directionally rotatable antenna array, we would point the antenna

along a beam heading that corresponds to a great circle path between our location and the desired station. This is called using the *SHORT PATH*; if conditions are favorable, a single hop may suffice, but it is by no means a sure thing. Under certain conditions, using the *LONG PATH* heading may provide a more favorable skip situation. This author once attempted to work McMurdo Sound, Antarctica from Virginia on the 14 MHz amateur band. The short path would normally have my antenna pointed almost due south. On the short path, however, the Antarctica station was very weak, about 0.1 dB above minimum discernible signal strength and readability. Having decided to try the long path, the Yagi was swung 180 degrees around effectively going over the North pole and back down again to the South Pole. It worked, and voila, there was McMurdo Sound with a tremendous signal, armchair copy at about 20 dB over S9 ! Obviously, this is not a standard situation, but it does prove the existence of alternate, and possibly even better paths based on the current situation in the sky. Maximum skip distance is achievable if one knows the current skip distance and virtual height, and can vary the antenna beam take-off angle.

2.3 EFFECTS OF NON-PLANAR EARTH AND IONOSPHERE

Thus far, we have considered both the earth and ionosphere to be flat planar structures. In reality, the curvature of both must be taken into account. In a curved ionosphere, Snell's Law would be of the form [1]

$$\mu r \sin \varphi_i = \mu_0 r_0 \sin \varphi_0 \qquad (16)$$

where r is the length of the radius vector from the center of the earth to the point where the refractive index is μ , φ is the angle between r and the direction of wave propagation and a is the radius of the earth, as shown in Figure 6. The subscript '0' refers to reference values; we usually select ground level as the reference so that $\mu_0 = 1$, $r_0 = a$, and $\varphi_0 = \pi/2 - \Delta$, where Δ = elevation angle. According to Davies [1],we then have

$$\mu r \sin \varphi_i = a \cos \Delta \qquad (17)$$

so that

$$\left(\frac{f_c}{f}\right)^2 = 1 - \left(\frac{a \cos \Delta}{a + h_r}\right)^2 \qquad (18)$$

This additional correction shows clearly the dependence of critical frequency on the height of reflection h_r. We would also need to correct the secant law to read $f = k f_c \sec \varphi_o$, where k is a correction factor taking values between 1.0 and 1.2, dependent on distance and actual reflection altitude.

Collin [3] approaches the problem a bit differently, more from the standpoint of reflection and transmission coefficients and the interference effects associated with signal field strength. Plotting relative field strength as a function of distance from the transmitter leads to a *COVERAGE DIAGRAM*, similar to an antenna field-strength radiation pattern. A sample diagram is shown in Figure 7, and earth curvature has been included in the analysis.

2.4 MUF / LUF / FOT CALCULATIONS

An accurate calculation of MUF is an involved process; in practical situations, it is best done empirically using up-to-date ionospheric data available from several sources [15,16,17,18,20]. MUF forecasts are based on the standards set by the Comite Consultatif International des Radiocommunications (CCIR). Done manually, the computations are lengthy and tedious. Several computer programs, in varying scales of grandeur and accuracy, are available. Exhaustive descriptions of the methods used in such forecasts are contained in Saveskie [19], Budden [7], Davies [1] and especially in the OT/TRER 13 report [15] . These methods employ F2 layer data primarily, generally for a specified path distance; calculations to other distances can be extrapolated. Relatively painless do-it-yourself prediction methods are available to the amateur radio service for communications between 3 - 30 MHz [16,17,18].

The MUF imposes the upper limit on frequency usage via ionospheric propagation. The lower limit is defined by the *LOWEST USABLE FREQUENCY (LUF)*, which is effectively set by the daytime D-layer attenuation [1] The optimum frequency of transmission (*FOT*) is considered that frequency which should provide the most reliable and consistent communications; it lies between MUF and LUF and is generally taken as FOT = 0.85 MUF.

2.5 SIGNAL STRENGTH AND PATH LOSSES

Signal strength may refer to either the relative field strength (amplitude) or the power density received at the antenna [1] . In any case, it is usually expressed in dB as the ratio of signal level to the ambient noise level, ie, the signal to noise (S/N) ratio. Signal strength can be diminished by a variety of factors, some of which we will discuss here. The use of a high-gain antenna system may help to offset some of these effects.

PATH LOSS

Path losses may consist of ionospheric absorption (which we have already discussed in detail), dispersive loss and power focusing and defocusing.

We have already alluded to the inverse square propagation loss, which simply means that the power flux falls off inversely as the square of distance from the source of radiation. This is referred to as a dispersive loss. In decibels, this loss is expressed as [1]

$$L_d = 20 \log s, \qquad s = \text{distance from source} \qquad (19)$$

Ionospheric focusing (or defocusing) [1,13] involves the convergence (or divergence) of wave rays that were originally adjacent and parallel to each other. This phenomenon is usually observed in the immediate skip distance area. The situation is similar to the reflection of parallel light rays from a spherical mirror, and is very dependent on take-off angle from the antenna, as shown in Figure 8. As we saw earlier, a wave ray entering at a certain incidence angle will go horizontal near the center of the layer and stay there for a while before being refracted earthward. For a set of rays close to each other, refraction will be such as to effectively re-focus the beam on its way down. Assuming the phase relationships are correct, we may even see constructive interference, so that at the receiver site, we are actually getting signal enhancement ! Now, if the take-off angle is sufficiently high, rays may be refracted near the edge of the layer, in which case the refractive index will not be enough to sufficiently re-focus the set of rays. The rays will eventually return to earth, but have been noticeably defocused and divergent by the time they impinge on the earth. This sort of behavior can degrade signal strength markedly, even up to 9 dB. As with other ionospheric factors, this also may change from moment to moment.

The more or less "continuous" stratification of the ionosphere that we have assumed so far is not always valid. Near sunrise and sunset, the ionization levels in the layer may be in a pronounced state of flux, enough to produce distortions in the ionosphere, as far as wave propagation is concerned. This situation also occurs in the equatorial regions of the earth. These distortions produce ionospheric "*tilts*"

which may cause marked assymetries in ray paths traversing that area of the layer, leading to pronounced focusing or defocusing.

POLARIZATION MISMATCH

A radio-wave propagating through the ionosphere may experience changes in polarization, particularly in the far-field region far from the antenna. Wave dispersion may cause an (initially) linearly polarized wave to become elliptically polarized while inside the layer, and to be circularly polarized when it leaves the layer. The reverse case may occur for a wave that is circularly polarized initially. The polarization sense at the receiving antenna, then, may not be the same as the wave it is receiving, leading to a polarization mismatch. These types of losses usually amount to about 6 dB. [1]

FADING

The ever-changing state of the ionosphere and its effects on propagation will cause field strength changes at the receiving antenna [1,5]. These changes are manifested in the S/N ratio at the receiver input terminals, causing input voltage fluctuations which are generally perceived as a fading of the received signal. The causes are many and the periods for which they last varied. It is a major factor in signal intelligibility and in the case of digital data transmission can produce significant bit error rates. We will look at a few of the major fading factors here.

ABSORPTION FADING is due to variations of ionization levels in the ionosphere. It may occur randomly, and usually happens suddenly, markedly changing the reflecting capability of an ionospheric layer. It is suspected this type of fading is closely linked with large solar flares and ionospheric storm activity.

INTERFERENCE (SELECTIVE) FADING arises from the propagation of several wave rays over different paths during its travels. The reflection of a beam of rays does not occur at one single point in the ionosphere, but rather is spread out over a finite area therein. Minute irregularities in ionospheric constitution may also exist in that area, causing individual rays to be reflected differently. Consequently, each ray travels a slightly different path from its neighbor. When this set of rays arrives at the receiver, their relative phases may differ enough to introduce interference effects.

Real antennas do not radiate all their energy at a single beam take-off angle, either. An actual signal will consist of a conglomeration of high and low angle rays; the incidence angle of each ray will determine exactly how it is reflected from the ionosphere, leading to further interference effects. The relative phases determine the degree of interference, ranging from enhancement to complete wave cancellation. Ionospheric focusing and defocusing is a prime example of this.

POLARIZATION FADING is introduced by polarization mismatch, since the receiving antenna is usually selective for a particular polarization type, and does not respond well to a different type.

Fading is characterized by three parameters: its DEPTH (signal strength depression in dB), its RATE and its DURATION. There are several diversity methods in common use intended to get around fading problems, each of which have their relative merits and disadvantages.

2.6 EXOTIC PROPAGATION MODES

There are several modes of propagation that occur that obey the laws we have studied here, but that occur on a sporadic basis and are considered "non-standard" modes of propagation. They are interesting in themselves, and can be used to advantage if one understands the underlying concepts and knows how to use them.

SPORADIC-E PROPAGATION

Sporadic-E (Es) propagation [4,6] is due mainly to isolated patches of ionization whose intensity is much higher than normal. They tend to form just below the E-layer, and may range 50-100 miles in diameter. They crop up randomly and may last only a few hours. Due to the high ionization level, these small "layers" are capable of propagating waves whose frequencies are much higher than the frequencies normally reflected by the E- and lower F-layers. This phenomenon does occur at HF frequencies, but is usually more noticeable at VHF frequencies. (Figure 9)

Since these patches form beneath the E-layer, the altitude effectively restricts the skip distance to about 1500 miles, so we associate sporadic-E with fairly short-skip communications. The physical extent of the patch usually keeps wave travel at a single-hop path; If the patch is sufficiently large, however, it is possible for multi-hop paths to occur, extending the range of communication considerably. Sporadic-E clouds seem to drift in a more or less

westerly direction at 150-250 miles per hour, probably due to the winds existing at that altitude. Wind shear has been suggested as a possible mechanism for this motion. This movement can lead to rapid changes in skip zones, with a resultant fast fade pattern.

Signal strength is hardly diminished in a sporadic-E reflections, so that strong signals may occur, even with low transmitter power levels. During the summer, sporadic-E is primarily a daytime thing, whereas in the winter, it may be useful far into the night-time hours. The effect is most noticeable in the polar and equatorial regions.

AURORAL PROPAGATION

As we saw in Part 1, the solar particle radiation tends to concentrate itself in the polar regions, due to the effects of the geomagnetic field [4,6]. The subsequent ionization caused by this flurry of particles is responsible for the familiar auroral light displays of varying colors and intensity. Such activity is more pronounced in periods of ionospheric storms and high solar flare activity.

The extremely high ionization levels that exist also produce very high signal absorption, almost to the point of being a barrier. There appears to be a low-frequency "hum" superimposed on signals going over the poles that causes a strange fading pattern that is generally called "*arctic flutter*". This flutter may be severe enough to render a signal almost unintelligible.

At HF, auroral displays usually hinder propagation. But, like sporadic-E, this situation may actually improve signal quality in the VHF 50-150 MHz range.

GRAYLINE PROPAGATION

Near the bottom of the sunspot cycle, propagation conditions are depressed enough to significantly make frequencies above about 15 MHz generally unreliable for long-distance communications via normal ionospheric reflections. During such periods, we look to the "lower" frequencies for effective propagation. The lower frequencies are more useful for short-skip work as a rule of thumb. There is, however, a propagation mode that produces enhanced signal qualities. Although it is a daily phenomenon, it may be available for only a few hours a day, right around sunrise and sunset. This mode is called *grayline propagation* [4,12].

The line separating the day and night sides of the earth is called the ***terminator.*** It is not distinct, but rather is a "fuzzy" area about 1000 miles wide, extending all the way around the circumference of the earth. We can think of it as the "twilight zone".

Since the position of the earth with respect to the sun changes seasonally, the position of the terminator changes as well, swinging an arc of about 47 degrees (relative to the poles) in one year. Thus, different points of the earth may show up on the terminator from day to day.

In Part 1 we discussed the dependence of the absorptive and refractive properties of the ionosphere on the angle at which solar radiation reaches the earth. The higher the angle, the higher the radiation intensity, as a general rule. At grazing angles, however, the refraction effects may not keep pace with the absorption effects. Thus, we have rapidly changing conditions in the ionosphere at the day-night interface, allowing certain frequencies to be propagated very efficiently, mostly in the 2 -10 MHz range.

The ionospheric situation in the vicinity of the terminator is shown in Figure 10. A frequency of 7 MHz will be assumed. On the daylight side, the D- and F-layers are close to their maximum ionization levels. In this case, the D-layer is strongly absorptive. Not too much gets through to the F-layers, which at this point are in good reflecting shape. On the darkness side, there is no D-layer due to the lack of solar radiation. The F-layer is still there, but at night the MUF will generally have fallen below 7 MHz. Therefore, a 7MHz signal will not be reflected and simply heads out into space.

Now let's look at the terminator region. Waves radiated toward either the day or night regions suffer the effects mentioned above. But ***WITHIN*** the terminator itself, waves that stay within the confines of the terminator are privileged indeed, being able to traverse a long distance with minimal path losses. Going "across" the terminator does little good, while going "along" it reaps rewards. This is essentially the mechanism that supports grayline propagation.

Grayline propagation is possible *ONLY* between two points on the earth's surface that are simultaneously with the grayline region. It is relatively easy to determine sunrise and sunset times at any point in the world on a given day, and to determine which of those points (worldwide) will simultaneously be within the grayline zone, and whether each is at local sunrise or local sunset. The position of the terminator, in general, will not be at the same place at local sunset as it was at local sunrise, on any given day.

Computer programs are easily written to predict grayline locations for a specified day. Armed with that knowledge, it is exciting to actually go and experience it. Although grayline is a restricted mode of propagation, it is of considerable interest.

2.7 REFERENCES

1) Davies, Kenneth, *IONOSPHERIC RADIO PROPAGATION*, Dover, 1966.

2) Corson,D. and Lorrain, P., *INTRODUCTION TO ELECTRO-MAGNETIC FIELDS AND WAVES*, W. H. Freeman and Company, 1962.

3) Collin, R.E., *ANTENNAS AND RADIOWAVE PROPAGATION*, McGraw-Hill, 1985.

4) Jacobs, G. and Cohen, T., *THE SHORTWAVE PROPAGATION HANDBOOK*, 2nd Edition, CQ Publishing, Inc., 1982

5) Wiesner, L., *TELEGRAPH AND DATA TRANSMISSION OVER SHORTWAVE RADIO LINKS - FUNDAMENTAL PRINCIPLES AND NETWORKS*, 3rd Edition, Siemens AG/John Wiley and Sons, 1984

6) American Radio Relay League, *THE ARRL ANTENNA BOOK*, Fourteenth Edition, 1982.

7) Budden, K.G., *THE PROPAGATION OF RADIO WAVES*, Cambridge University Press, 1985

8) Shibuya, S., *A BASIC ATLAS OF RADIOWAVE PROPAGATION*, John Wiley and Sons, 1987

9) Gibilisco, S., "Propagation of Radio Waves", *HAM RADIO*, August, 1982

10) Jansz, E., "Radio Wave Propagation", *HANDS-ON ELECTRONICS*, June, 1985

11) Barkstrom, B., "What Time Does the Sun Rise and Set?", *BYTE*, July, 1981

12) Wells, B., "Fundamentals of Grayline Propagation", *HAM RADIO*, August, 1984

13) Graf, C.R., "Ionospheric Focusing", *HAM RADIO HORIZONS*, December, 1977

14) Hall, J., "High-Frequency Propagation Estimations for the Radio Amateur", *QST*, March, 1972

15) Leftin, M., *IONOSPHERIC PREDICTIONS*, VOL 1, Office of Telecommunications, Telecommunications Research and Engineering Report 13, US Dept. of Commerce, Sept. 1971

16) Stonehocker, G., "DX Forecaster", *HAM RADIO* (monthly)

17) Jacobs, "Propagation", *CQ* (monthly)

18) White, "How's DX?", *QST* (monthly)

19) Saveskie, P.N., *RADIO PROPAGATION HANDBOOK*, TAB Books, 1980

20) NOAA/SESC reports.

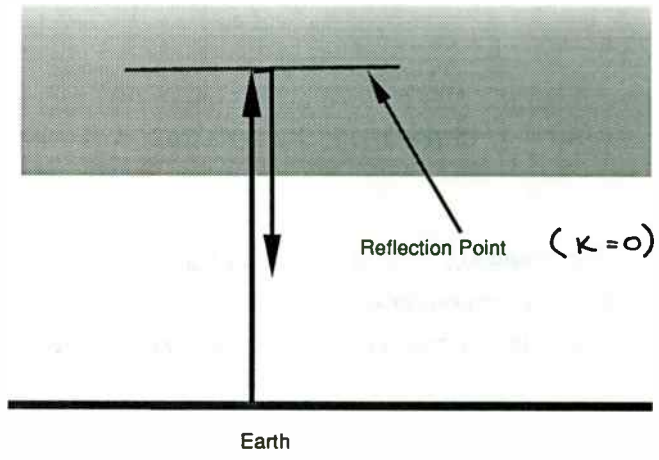


Figure 1. Ionospheric Reflection at Normal Incidence

(Ionosonde Method)

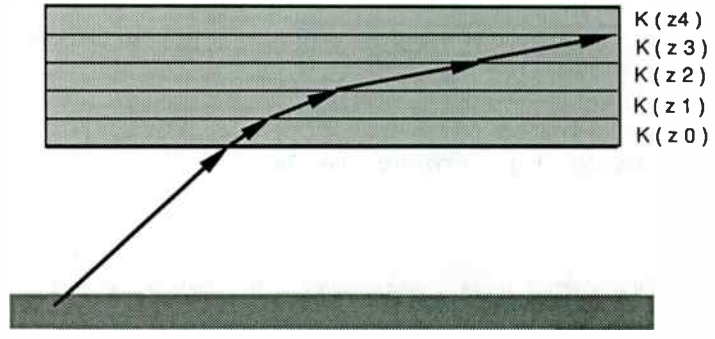


Figure 2. Ionospheric Propagation at Oblique Incidence

(Continuously Stratified Layer Model)

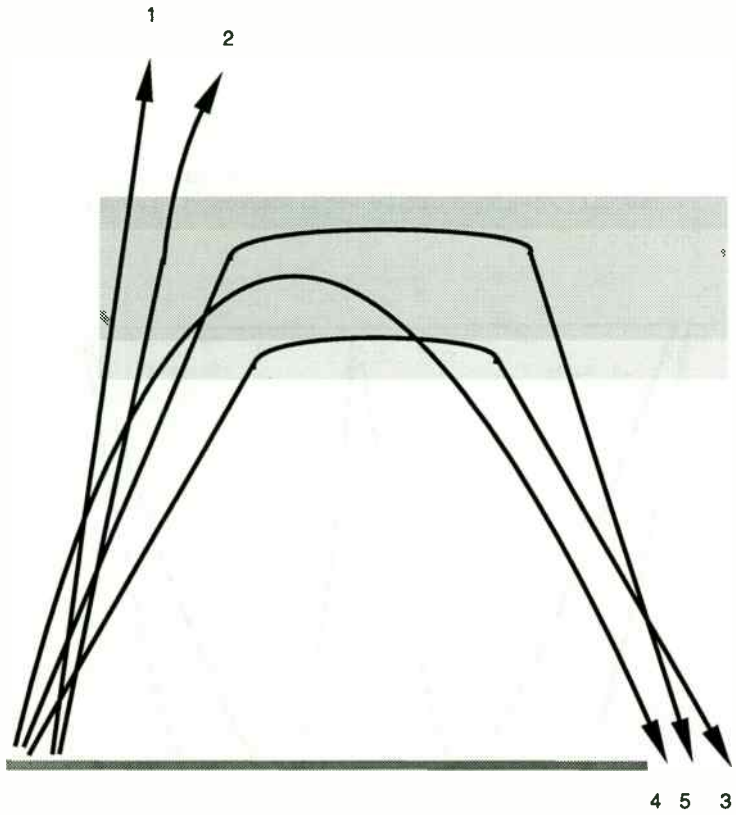


Figure 3. Propagation at Oblique Incidence

Influence of Incidence Angle

F4 > F3 > F2 > F1

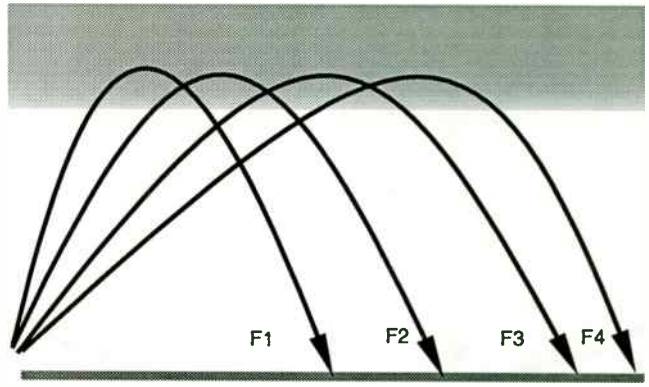


Figure 4. Frequency Dependence of Radiowave Propagation

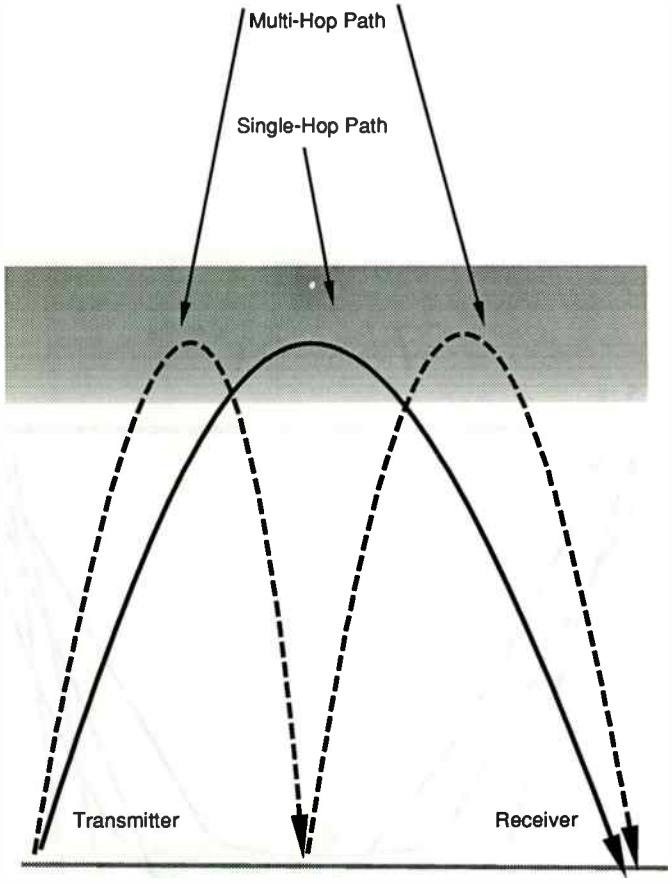


Figure 5. Comparison of Single-Hop and Multi-Hop Propagation Paths

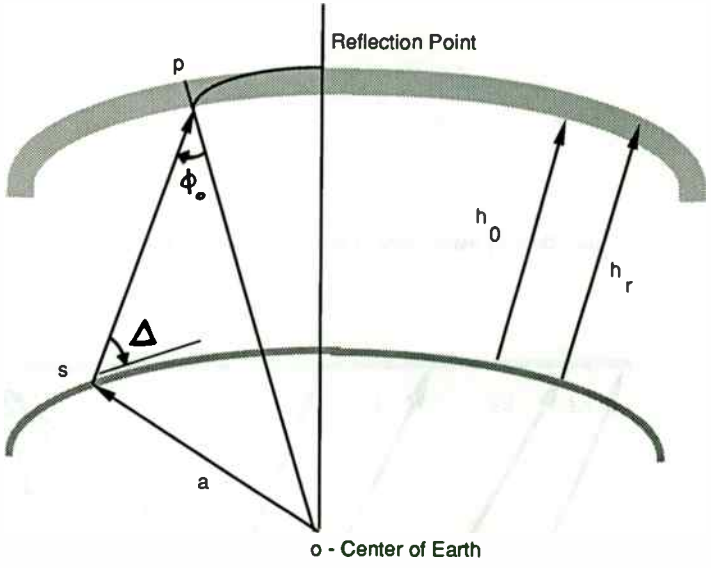


Figure 6. Propagation with Curved Ionosphere and Curved Earth

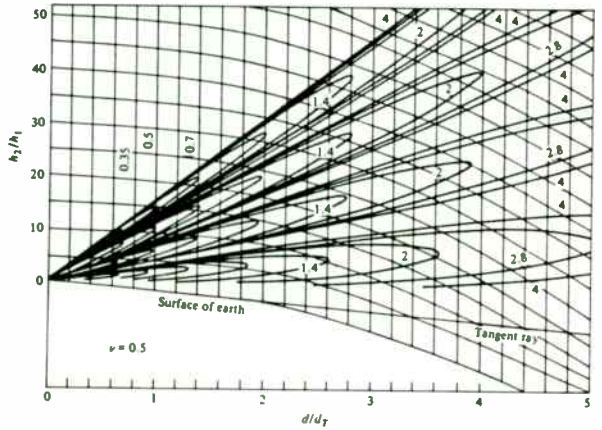


Figure 6.17 Coverage diagram for $\nu = 0.5$. [*From C. R. Burrows and S. S. Attwood (eds.), Radio Wave Propagation, Academic Press, New York, 1949.*]

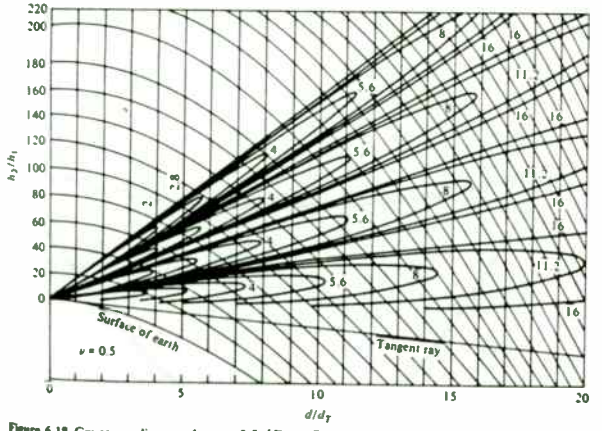


Figure 6.18 Coverage diagram for $\nu = 0.5$. [*From C. R. Burrows and S. S. Attwood (eds.), Radio Wave Propagation, Academic Press, New York, 1949.*]

Figure 7. Typical Coverage Diagram for Radiowave
with Earth Curvature effects included

(From Reference 3, Collin, " Antennas and Radiowave Propagation ")

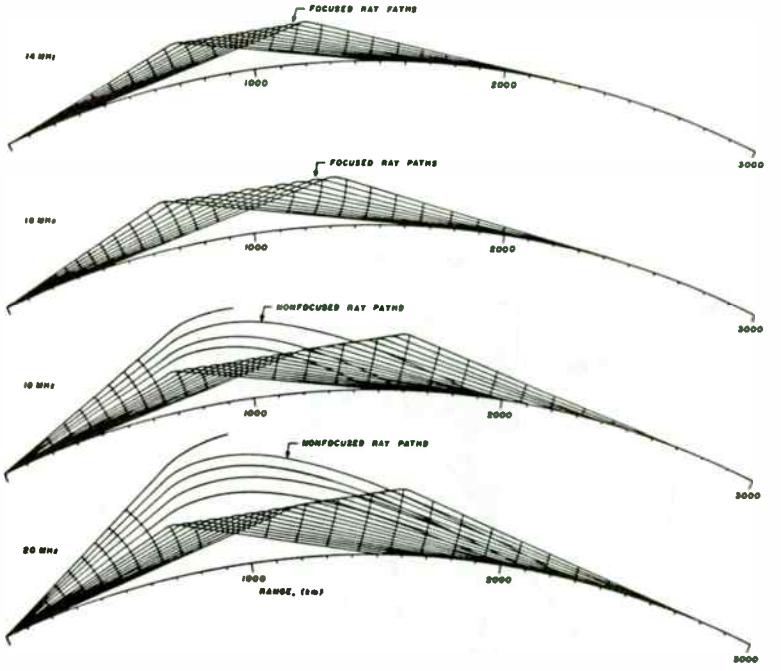


Figure 8. Ionospheric Focusing and Defocusing

(From Reference 13, Graf, " Ionospheric Focusing ",
HAM RADIO HORIZONS, December 1977)

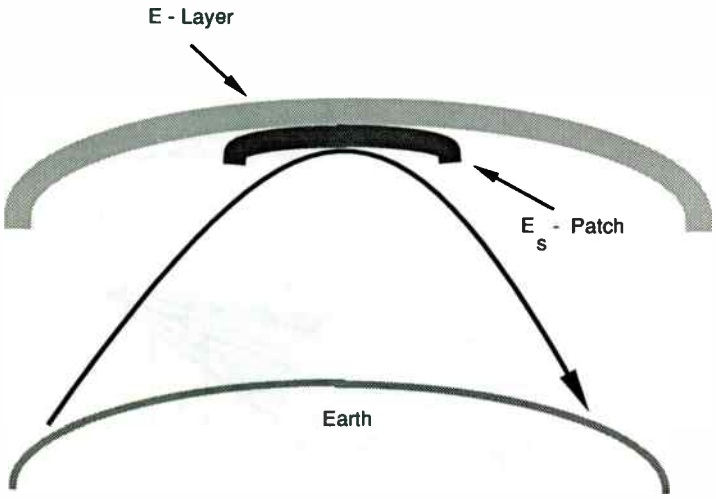


Figure 9. Sporadic - E (E_s) Propagation

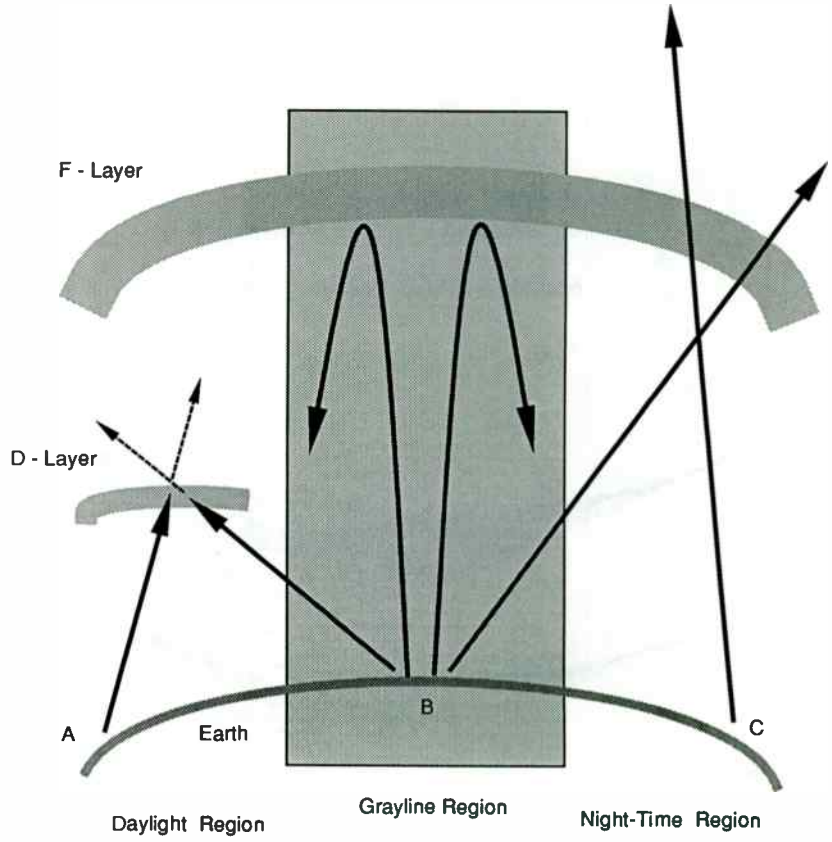


Figure 10. Geometry of GRAYLINE Propagation

RADIO-WAVE PROPAGATION - A TUTORIAL

PART 3 - PROPAGATION AT VHF AND ABOVE

D. R. Dorsey, Jr
Principal Systems Analyst

Network Development Division
Control Data Corporation
Sunnyvale, CA 94089

3.0 INTRODUCTION

Parts 1 and 2 of this tutorial are concerned primarily with radiowave propagation via the ionosphere. At frequencies in the VHF range and above (> 50 MHz), the ionosphere is essentially transparent, and not much of a factor. Propagation becomes predominantly line-of-sight, and we are now more concerned with the atmospheric characteristics that affect that mode of propagation. The effects of rain, snow, ice, fog and atmospheric gases may be profound on such very high frequencies. Special problems associated with satellite communications are also considered.

3.1 PROPAGATION CHARACTERISTICS AT VHF AND ABOVE

Above about 50 MHz, the wave frequency is much greater than the plasma angular frequency that the ionosphere becomes essentially transparent to radiowaves of those frequencies. Compared to HF, propagation is primarily direct-wave, line-of-sight travel. Interference effects from ground reflected waves occur, but not nearly as strongly at HF; they are generally not a major factor.

At these very to extremely high frequencies, wavelengths begin to take on dimensions comparable to the particulate constituents of the atmosphere. As frequency increases, the more severe these effects become. We now become more concerned with attenuation and depolarization effects than anything else. Attenuation due to rain, snow, ice, clouds, fog, water vapor and atmospheric gases must all be taken into account and to some degree, scattering from these entities.

For most purposes, attenuation is defined [2] as the difference, in dB, between the power under ideal (clear weather) conditions and the power actually received at the antenna after traversing an attenuative medium.

3.2 DEPOLARIZATION

In Part 1 we mentioned that most antennas possess a specific sense of polarization, and that they tend to reject signals not having the same polarization sense. For example, a horizontally polarized antenna does not see a vertically polarized signal very well. This is true of most orthogonally polarized systems. There is an infinite number of such polarized pairs; horizontal and vertical are but a subset that is very easy to discuss. Since horizontal and vertical polarization mutually exclude each other, it is then possible to radiate one signal containing both types simultaneously. Such a scheme is called *CO-POLARIZATION* and allows two channels in one signal, with almost negligible mutual coupling between channels. Depolarization arises when a co-polarized wave enters an anisotropic medium and suffers an alteration in its polarization characteristics such that coupling between polarization senses is introduced. This cross-polarization produces crosstalk and other undesired interference effects across channels. The situation is illustrated in Figure (1) [2,3] for a wave initially copolarized with horizontal and vertical components. Referring to this figure, the depolarizing medium generates an unwanted component E_{ba} into the desired E_{aa} and an undesired component E_{ab} into the E_{bb} state; E_{ab} and E_{ba} are called the *CROSS-POLARIZED* states.

CROSS-POLARIZATION ISOLATION is the ratio of co-polarized to cross-polarized power in the same channel. It is usually expressed as a decibel ratio:

$$XPI_V = 20 \log_{10} | E_{aa} / E_{ba} | \text{ dB} \qquad (1)$$

$$XPI_H = 20 \log_{10} | E_{bb} / E_{ab} | \text{ dB}$$

These values we would like to maximize.

Similarly, *CROSS-POLARIZATION DISCRIMINATION* is defined

$$XPD_V = 20 \log_{10} | E_{aa} / E_{ab} | \text{ dB} \qquad (2)$$

$$XPD_H = 20 \log_{10} | E_{bb} / E_{ba} | \text{ dB}$$

Both of these parameters give us a measure of the undesired coupling that may exist between polarization states.

3.3 ATTENUATION BY RAIN

Rain attenuation arises predominantly from absorptive power losses in the lossy dielectric medium of water. To a far lesser degree, scattering of a wave by a rain droplet also represents a loss mechanism. Radiowave interactions with rain are influenced by the rate of rainfall (raindrop density) and on raindrop size and shape distributions.

Raindrop size distributions have been experimentally determined to be exponential of the form

$$N(r) = N_0 e^{-\Lambda r} = N_0 e^{-[cR^{-d}]r} \qquad (3)$$

where

r = median drop radius

R = rate of rainfall (mm/hr)

N_0 , c, d, Λ = empirically determined constants

Of course, actual raindrops are not of uniform size or shape, so the total attenuation coefficient would involve integrating over all drop sizes. Drop size and shape are closely related; in general, the larger the drop, the more it departs from a spherical shape toward an oblate spheroidal shape. It is analytically more convenient to assume a spherical drop shape, but it is not the most realistic model. In addition, storm conditions will generate wind gradients that will also influence drop size and shape distributions.

Most treatments of rain attenuation assume the intensity (power density) of the incident wave to diminish exponentially

$$P = P_0 e^{-KL} \qquad (4)$$

over a distance L in an attenuating medium with attenuation coefficient K , which is often expressed

$$K = \rho Q_e \qquad (5)$$

where ρ is the drop density and Q_e is the attenuation cross section.

Contributions from individual drops are considered to be additive and mutually independent.

In general, the attenuation due to a volume of rain of extent L in the direction of wave propagation is

$$A = \int_0^L \alpha \, dx \quad (\text{dB}) \qquad (6)$$

where α is the specific attenuation, almost universally expressed in the form

$$\alpha = aR^{b} \quad (\text{dB / km}) \qquad (7)$$

where R is the rainfall rate, and a and b are coefficients that are strongly frequency and temperature dependent. Typical values of these coefficients are presented in Figure 2. The specific attenuation has been studied extensively; the treatments given by Ippolito [3] and by Pratt and Bostian [2] are particularly excellent.

Since rainfall varies spatially and temporally, many attenuation models extend the analysis into a statistical model based on cumulative probability distributions of rainfall, and the so-called *EXCEEDANCE* curve P [3]. In this case, we consider an effective path length L_{eff} and an attenuation law of the form

$$A(P) = a\,[\,R(P)\,]^{b} L_{eff} \quad (\text{dB}) \qquad (8)$$

In most cases and especially in satellite communications work, the direction of wave propagation and the direction of rainfall are not orthogonal, as shown in Figure (3). Attenuation calculations must then take into account the effective path length over which the wave will travel through a finite volume of rain. This path is called the *SLANT PATH*. Figure (3) is also the basis for the SAM model [4] which considers the height of both the storm and the earth station, and the elevation angle between earth station and satellite. In the SAM model, the effective path length through rain is

$$L = [\,H_e - H_0\,] \,/\, \sin\theta \quad (\text{km}) \qquad (9)$$

Several statistical models for the prediction of rain attenuation are available, including the Rice-Holmberg model [5], the Dutton-Dougherty model [6], the Lin model [7], the Crane global rain model [8], and the CCIR model [9]. Most all of them take the rain rate as measured on the earth's surface as the statistical variable, and assume an attenuation law as defined by Equation (8).

3.4 DEPOLARIZATION BY RAIN

Rain depolarization is a direct result of the shape of large raindrops in a storm center. Large drops take on an oblate spheroid shape (just like M&M's candy) due to the surface tension and aerodynamic forces acting on them. The orientation of an incident radiowave with respect to the major and minor axes of an oblate spheroid (and the orientation and alignment of the drops relative to a reference direction) determines exactly how much water it will traverse [Figure 4]. Hence, the differently polarized components of a wave may be attenuated and depolarized in different amounts --- a process known as *DIFFERENTIAL ATTENUATION*.

The cross-polarization discrimination has been empirically related to the amount of attenuation as [2,3]

$$XPD = U - V \log_{10} A \qquad (dB) \qquad (10)$$

where U and V are empirically determined coefficients that depend on frequency, polarization angle, elevation angle and drop tilt angle. The values for U and V, as specified by the CCIR, are

$$U = 30 \log_{10} f - 40 \log_{10} [\cos \theta] - 20 \log_{10} [\sin 2\tau] \qquad (11)$$

$$V = 20 , \quad 8 < f < 15 \text{ GHz}$$

$$= 23 , \quad 15 < f < 35 \text{ GHz}$$

where

f = Frequency

θ = Path Elevation Angle

τ = Angle between the received electric field vector and the local horizontal

3.5 ATTENUATION BY CLOUDS AND FOG

Microwave and millimeter-wave attenuation by clouds and fog are defined by the same relations as rain attenuation. The basic difference is in the size of the water droplets. Rain drops range in size from 0.1 mm to 10 mm in diameter. Fog is a suspended mist of minute water droplets with a mean diameter of 0.02 mm to 0.1 mm. Clouds are also composed of small water droplets, but within a cloud the relative humidity is almost 100 percent.

The water content of fog may range from 0.01 to just under 1.0 gm/cubic meter, depending on how "thick" the fog is. In clouds, the water content may vary from 0.05 to more than 2.0 gm/cubic meter. Even at frequencies as high as 300 GHz, the attenuation produced by fog or clouds rarely exceeds 1.0 dB/km. Typical values are in the 0.01 to 1.0 dB/km range.

3.6 SNOW AND ICE EFFECTS

Solidified water, such as snow and ice crystals, has a dielectric constant much different than liquid water. In general, microwave attenuation in dry snow or ice is probably an order of magnitude less than in rain, while in wet snow it is approximately comparable to attenuation rates in rain. As we all know, it would be sheer folly to attempt the definition of a "standard" snow crystal. Consequently, we have almost nothing in the way of an analytical model to define attenuation by, and instead must be satisfied with experimental determinations.

Although snow and ice crystals may form in random patterns, it appears that that such formation may occur along certain preferred directions. It has been observed that snow and ice display a very strong depolarization that is associated with a very small co-polarized attenuation.

3.7 ATTENUATION BY ATMOSPHERIC GASES

The primary mechanism responsible for radiowave attenuation by the gaseous constituents of the atmosphere is molecular absorption. This absorption is a result of a change in the quantum rotational energy of the molecule. Oxygen and water vapor have several resonant absorption lines in the centimeter and millimeter wave regions. At these particular resonance frequencies, attenuation is very high, while on either side there will be window areas where absorption is significantly less.

3.8 SCINTILLATION EFFECTS AND SCATTER PHENOMENA

Scintillation is a phenomenon seen as rapid variations in the amplitude, phase and/or polarization of a signal. It may occur in the upper parts of the ionosphere and in the troposphere, and is usually due to irregularities in electron density or refractive index.

Ionospheric scintillations are due to electron density fluctuations in the F-layers, and are seen mainly in the equatorial and polar regions, especially in periods of high sunspot activity. This is primarily a low frequency effect, but has been observed on satellite links in the equatorial regions for frequencies as high as 6 GHz, and with fluctuations up to 5 dB.

On the other hand, tropospheric scintillations tend to occur more from refractive index variations caused by gradients of high humidity and thermal inversion layers. They occur randomly, depending on local weather and climate conditions. Turbulent layers in the troposohere may also play a part in the fluctuations.

In the troposphere, the index of refraction is considered [3] as horizontally stratified. For radiowave frequencies, the refractive index depends on pressure, temperature and water vapor content, and has a value near unity. It is conveniently described in terms of *REFRACTIVITY* [3],

$$N = (\mu - 1) * 10^{-6} = (77.6 / T)\left[p + 4810\, e / T\right] \qquad (12)$$

where

p = pressure (millibars)

e = water vapor preesure (millibars)

T = temperature (degrees Kelvin)

The variation of N with altitude h, as suggested by the CCIR, is exponential of the form

$$N = 315\, e^{-(h / 7.36)} \qquad (13)$$

At VHF and above, scattering may also (occasionally) take place due to electron density fluctuations induced by turbulence and wind shear in the D-layer region.

The existence of VHF signal scattering is also observed by reflections from an ionized meteor trail at altitudes of 80-120 km. These trails are produced almost continuously, but they last for only a few seconds. Such reflections are very dependent on signal wavelength and the electron density level in the meteor trail.

3.9 TROPOSPHERIC RADIOWAVE PROPAGATION

There are several modes of propagation that occur in the troposphere, that portion of the atmosphere extending from roughly 50 to 100 km. This region is low enough that weather and climate conditions can have some effect on radiowave transmission. These modes of propagation may provide communications up to 2500 miles on VHF and UHF frequencies, although they must not be considered as reliable modes of transmission. Most are related to the changes in refractive index discussed above.

Under normal circumstances, temperature and the water vapor content of the atmosphere decrease with altitude. In a temperature inversion, however, there is a sudden rise in temperature, and a dramatic decrease in moisture content. Figure 5 shows these variations for a normal atmosphere and for two types of inversions.

A *RADIATION INVERSION* [12] is formed over land as the earth cools during night-time, while the air higher up stays relatively warm. The depth of this inversion rarely exceeds 500 meters above the surface of the earth. The change in refractive index over this distance is so slight that wave bending is almost negligible.

A *SUBSIDENCE INVERSION* [12] is associated with a large high pressure weather system. A high pressure system is a rather large air mass whose pressure is higher than the surrounding air. There are steady winds moving in a clockwise fashion associated with this type of system, which moves the air within the high downward and out toward the boundaries of the system. This air mass is generally stable and the air is therefore gradually sinking. This *subsidence* of air produces compression and heating as the air sinks, leading to very strong inversion, whose depth is about 200 meters with a base between 4000-5000 feet. During the day, solar heating of the earth's surface generates rising air masses which tend to counteract the sinking behavior in the high pressure system. Once the sun has set, however, the ground cools again and subsidence prevails. Therefore, subsidence inversions tend to be much more stable and stronger and reside closer to the ground during the night-time hours. Refractive index can change abruptly in this type of inversion.

Tropospheric phenomena occur when the change in refractive index is high enough to cause significant radiowave bending. The simplest case is ordinary tropospheric bending (Figure 6), which takes place near a frontal system and tends to cause signal beam spreading. Closely related to this is *TROPOSPHERIC SCATTER*, in which the waves are scattered from a region irregular refractive index variation (Figure 7). These fluctuations are very weak, but a signal with sufficient radiated power may be able to scatter from the layer and provide greatly enhanced communications over an over-the-horizon path for a short period of time. The rapid fluctuations in atmospheric properties will generally introduce enough phase variations in a signal to produce significant multipath fading.

In a strong inversion, the change in refractive index will suffice to reflect a wave back earthward by the process of super-refraction. If the boundaries of neighboring air masses are very well-defined (Figure 8), then prolonged reflections may occur at the boundaries over a large distance. In this case, we have the phenomenon of *TROPOSPHERIC DUCTING*. Ducts usually appear at some altitude up to 5000 feet (elevated duct), although it is possible for it to be very close to the ground (surface duct), with the ground being the lower edge of the duct. Ducts may vary in depth from a few to several hundred feet. Duct formation springs mainly from the strong effect of water vapor content on the index of refraction, so most ducts will form over large bodies of water.

A duct is sort of a natural microwave waveguide in which signals may propagate for very long distances with negligible attenuation or degradation of signal strength. Of course, to be useful, both the transmitting and receiving stations must be within the confines of the duct, or least capable of propagating into it. Most ducting takes place at UHF and microwave frequencies, although the phenomenon is occasionally observed even down to HF frequencies.

3.11 SATELLITE COMMUNICATIONS

As we have seen in Parts 1 and 2 of this paper, ionospheric reflection is predominant below about 30 MHz. The ionosphere is essentially transparent to radiowaves above 30 MHz. For this and many other reasons, most communications between earth and a spaceborne satellite occur at frequencies well above this lower limit, usually in the GHz range. Of course, we now have to worry about the propagation factors at those frequencies that we have discussed so far here in Part 3.

In addition, radio noise due to atmospheric gases, clouds, rain, surface emissions and noise sources of extraterrestrial origin must also be considered in the overall performance of the satellite communications system. Since noise is also frequency dependent, some satellite link problems can be alleviated by the use of split frequency channels between uplink and downlink.

Inasmuch as most satellite systems today are moving toward the use of wide bandwidths at very high carrier frequencies, we must also consider the effects of the atmosphere on the bandwidths that can be handled reliably [3]. This is especially true for satellite systems using digital methods that typically require a large channel bandwidth. Amplitude and phase dispersions may produce severe signal degradation through the entire bandwidth. If these dispersions are sufficiently large, bandwidth coherence may be

profoundly affected, with a resultant increase in signal distortion and possibly even total loss of signal. These factors can impose an upper limit on the channel capacity or information bandwidth that can be used in a satellite system. Frequency dispersion due to ionospheric effects is less of a problem; it has been observed experimentally [3] that rain attenuation in excess of 100 dB would be seen before frequency dispersion became significantly noticeable For most carrier frequencies above 10 GHz, a bandwidth greater than about 3.6 GHz can be supported before bandwidth incoherence becomes significantly bothersome.

In the case of orbiting satellites, elevation angle can also be a problem. For elevation angles less than 5 degrees or so, the propagation path can be very long, particularly through the troposphere. Such a long path may give atmospheric turbulence ample opportunity to produce scintillation effects. The long propagation path may also give rise to multipath signals being intercepted by the receiving antenna, leading to a multipath fading problem. Site diversity methods, using several receiving antennas feeding to a common receiver, may help to alleviate the multipath problems to some degree [2,3]. A low elevation angle will also aggravate the problems of rain, cloud, fog and atmospheric gas attenuation due to the greatly extended propagation path. Therefore, low elevation angles are usually to be avoided as much as possible, particularly for an orbiting satellite system in which signal acquisition with the satellite may be possible for only limited periods of time.

Another area of concern exists for linearly polarized radio waves used in satellite systems. This is the phenomenon of *FARADAY ROTATION* in the ionosphere. The ionosphere is characterized as an anisotropic and inhomogeneous plasma; consequently, the study of wave propagation through it becomes a nonlinear phenomenon which is very difficult to approach analytically. The overall effect, however, is that in a linearly polarized wave, the direction of the electric field vector will rotate slowly about the direction of propagation. Hence, the polarizations on each end of the satellite link will differ, appearing to have changed by some finite angle. The magnitude of the angular change in polarization sense is directly proportional to the path length, electron number density, geomagnetic flux density and the orientation of the direction of propagation with that of the geomagnetic field. Faraday rotation also varies inversely as the square of the operating frequency [2], and is noticed mostly in the 29 - 146 KHz frequency range [14].

3.12 REFERENCES

1) Collin, R.E., *ANTENNAS AND RADIOWAVE PROPAGATION*, McGraw-Hill, 1985.

2) Pratt, T. and Bostian, C.W., *SATELLITE COMMUNICATIONS*, John Wiley and Sons, 1986

3) Ippolito, L.J., *RADIOWAVE PROPAGATION IN SATELLITE COMMUNICATIONS*, Van Nostrand Reinhold Company, 1986

4) Stutzman, W.L. and Dishman, W.K., " A Simple Model for the Estimation of Rain-induced Attenuation", *IEEE Transactions on Antennas and Propagation*, AP-26, 318-329 (March 1978)

5) Rice, P.L. and Holmberg, N.R., "Cumulative Time Statistics of Surface Point-Rainfall Rates",] *IEEE Tranactions on Communications*, Vol. COM-21, No. 10, October, 1973

6) Dutton, E.J. and Dougherty, H.T., "Modeling the Effects of Cloud and Rain Upon Satellite-to-Ground System Performance", *OT Report 73-5*, Office of Telecommunications, Boulder, CO, March, 1973

7) Lin, S.H., "Empirical Rain Attenuation Model for Earth-Satellite Paths", *IEEE Tranactions on Communications*, Vol. COM-27, No. 5, May, 1979

8) Crane, R.K., "A Global Model for Rain Attenuation Prediction", EASCON '78 Record, *IEEE Pub. 78CH 1354-4 AES*, Arlington, VA, Sept. 1978

9) CCIR, Report 564-2, "Propagation Data Required for Space Telecommunications Systems", in Volume 5, *PROPAGATION IN NON-IONIZED MEDIA*, Recommendations and Reports of the CCIR-1982, International Telecommunications Union, Geneva,.pp 331-373, 1982

10) American Radio Relay League, *THE ARRL ANTENNA BOOK*, Fourteenth Edition, 1982.

11) Shibuya, S., *A BASIC ATLAS OF RADIOWAVE PROPAGATION*, John Wiley and Sons,

12) Pocock, E. , "The Weather That Brings VHF DX", *QST*, May 1983.

13) Gibilisco, S., "Propagation of Radiowaves", *HAM RADIO*, August, 1982.

14) American Radio Relay League, *THE SATELLITE EXPERIMENTERS HANDBOOK*, 1984.

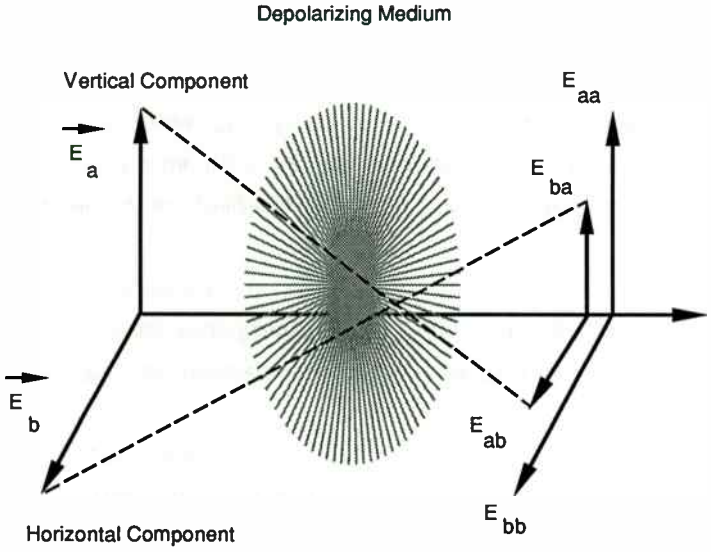


Figure 1. DEPOLARIZATION -- Geometry / Vector Components

Frequency	Coefficients		Specific Attenuation		
(GHz)	a	b	(dB / km)		
			R = 10	R = 50	R = 100
2	0.000345	0.891	0.003	0.1011	0.021
4	0.00147	1.016	0.015	0.078	0.158
6	0.00371	1.124	0.049	0.30	0.657
12	0.0215	1.136	0.29	1.83	4.02
15	0.0368	1.118	0.48	2.92	6.34
20	0.0719	1.097	0.90	5.25	11.24
30	0.186	1.043	2.05	11.0	22.7
40	0.362	0.972	3.39	16.2	31.8
94	1.402	0.744	7.78	25.8	43.1

Figure 2. Attenuation Coefficients and Specific Attenuation for Rain

(From Ref. 3, Radiowave Propagation for Satellite Transmission L. J. Ippolito, Table 4 - 2, page 44)

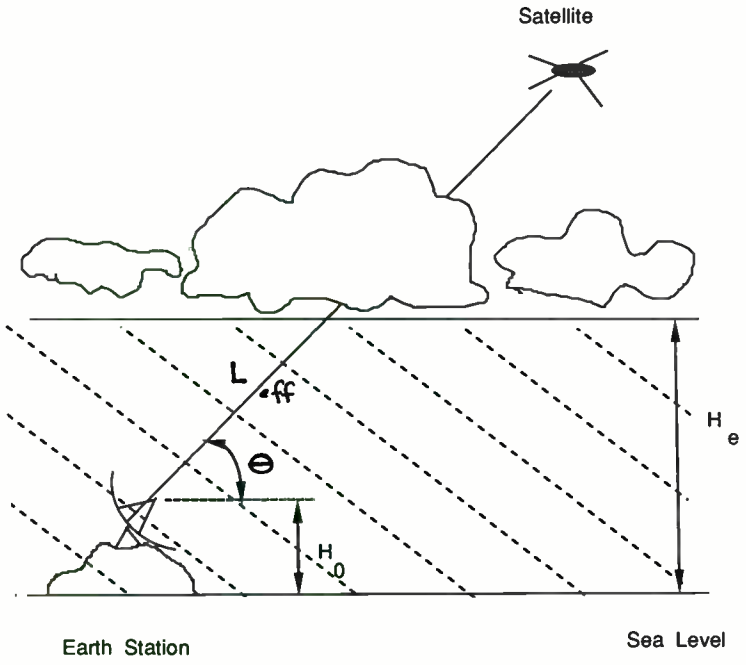


Figure 3. Slant Path and Elevation Angle Dependence in Rain Attenuation Model

(Reference 2 - Pratt and Bostian - Figure 8.17)

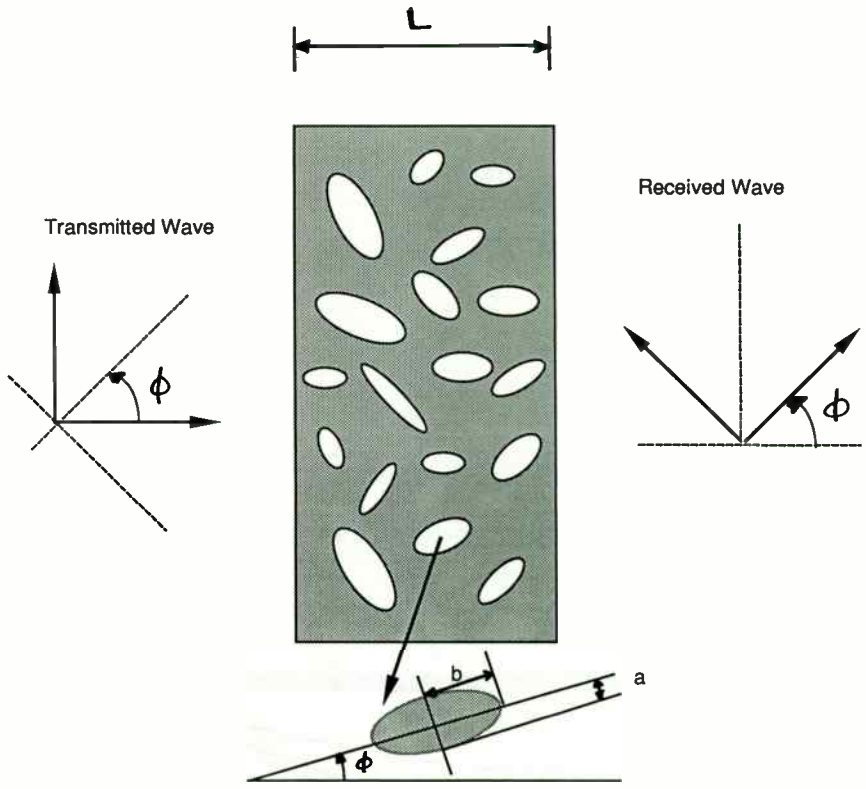


Figure 4. Rain Depolarization Model - Oblate Spheroids

(Reference 3 - Ippolito - Fig. 6-2, p. 96)

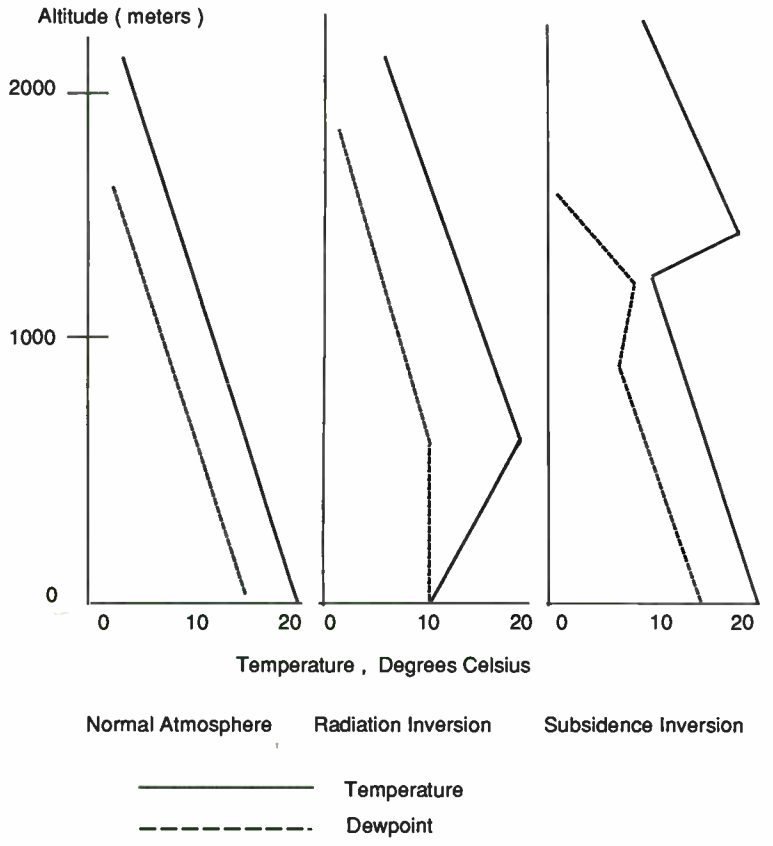


Figure 5. Temperature and Moisture Content Profiles
Normal Atmosphere and Inversions

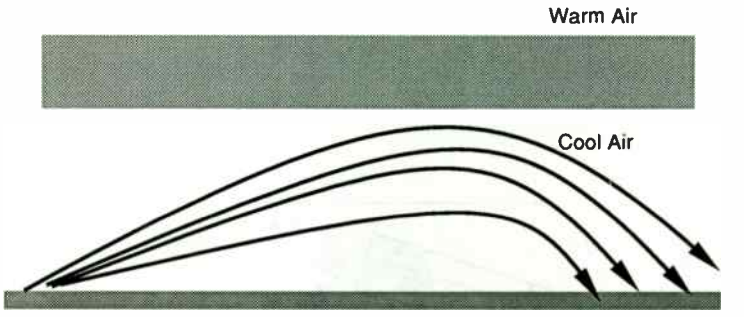


Figure 6. Tropospheric Bending

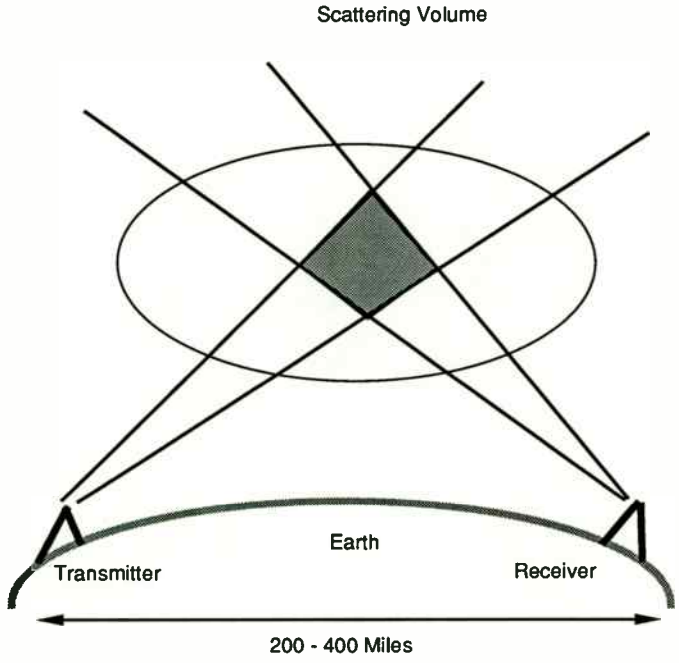


Figure 7. Geometry of TROPOSPHERIC SCATTERING

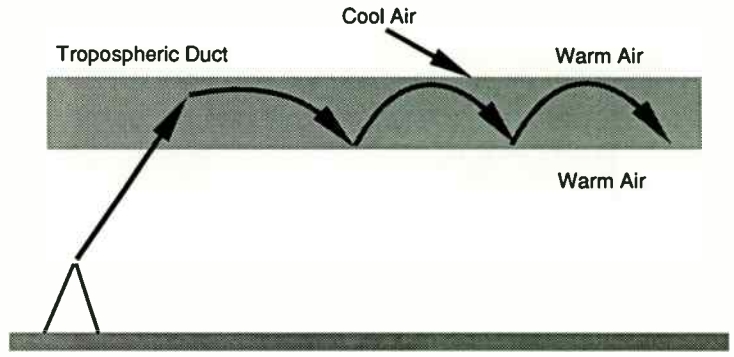


Figure 8. TROPOSPHERIC DUCTING

Accuracy Considerations in RF Network Measurements

Lorenzo Freschet
Product Manager
Hewlett-Packard
Network Measurements Division

When using a reflectometer setup with a simple frequency response calibration, a measurement can have peaks and ripples in the results. How accurate is this measurement? Does the device under test (DUT) exhibit these rapid variations with frequency? Unfortunately, these variations are caused by limitations in the measurement system.

FIGURE 1, FIGURE 2

Measurement limitations (or errors) prevent measured data from being a true representation of the unknown test device. In all applications, measurement errors can influence the application goals. Since measurement accuracy is determined by the errors in the system, this paper will discuss the process of identifying, characterizing and mathematically removing as many errors as we can. There are many classes of errors. Systematic errors are those which are stable and repeatable. Random errors are those which are random in nature and cannot be characterized and removed. Drift (or environmental) errors are those associated with temperature, humidity, pressure, or other factors related to time. This paper will focus on systematic errors.

The ripple in our initial measurement is largely caused by an error called directivity. At the peaks, the device's reflection signal adds in phase with the directivity signal producing a data maximum signal. At the low valleys, the directivity signal and the device data are approximately the same magnitude but out of phase with each other, producing a data minimum. Between the peaks and valleys, directivity and device data combine somewhere between the out-of-phase and in-phase conditions.

This effect is often caused by using an adapter or cable before the Device Under Test. As an example, consider a DUT which has an SMA male connector. Suppose the measurement system has a directivity specification of 40 dB with a 7 mm output connector. But in order to measure the DUT, we will have to adapt from the 7 mm connector to SMA. The resultant overall directivity is the sum of the measurement system directivity plus the reflection coefficient of the adapter. Therefore, a precision 7mm to SMA adapter with a SWR of 1.06 reduces the overall directivity from 40 dB to 27 dB. Other combinations of adapters can reduce overall directivity even further. Overall directivity then is the limit to which a DUT's RL can be measured, but even then, the error is 100%.

Reflection frequency response is also contributing an error. It's actually a composite of the frequency response of the coupler (coupler tracking), the test cables and the input mixers or detectors. And finally, looking back into a measuring system, including all of the adapters and cables, there probably is not

an ideal match. When a signal reflecting off the test device comes back into the test system, an error loop is formed with signals bouncing back and forth. Since this error is related to the product of these terms, it is especially visable when the test device has a high mismatch. For the case of semi-conductors or out of band filter measurements, it can be an error approaching 10%.

Here is the model we use to evaluate each of these errors and remove their effects.

FIGURE 3, FIGURE 4

From this error model and it's equation we can look at the factors which contribute to the difference between the measured value and the actual value. From the equation, it is easy to see that for devices which have very small reflection coefficients, directivity is the major error. For devices which have large reflection coefficients, source match is the most significant error.

The measurement uncertainty equation establishes the worse case uncertainty when the only information we have is worse case magnitude data (as with scalar analyzers). That is, we do not know the exact phase or magnitude of the error at each frequency, but we do know the worst case magnitude value of the errors over the frequency range of interest.

To improve measurement accuracy, we could buy better hardware which would reduce these errors, but that could be very expensive or impossible. We do know, however, that these errors are vector signals and we could mathematically remove them if we could characterize them independently. Since we have three unknown signals, a technique can be used in which three different standard devices are measured to characterize these errors. Theoretically, any three known devices will suffice. This procedure can be used with network analyzers to perform a technique called vector accuracy enhancement.

This calibration procedure entails measuring a "perfect" load. At lower frequencies (< 2-3 GHz) this might be a fixed termination whose reflection is known to be extremely low or at higher frequencies it might be a sliding load whose reflection is known but whose phase can be varied. At microwave frequencies, it is difficult to obtain a load which has excellent return loss over a broad frequency range. One alternative to characterizing directivity over this broad frequency range is to use a sliding load. The directivity vector at a given frequency is determined by sliding the load on the airline to create a "circle" of data points. The center of this circle is the directivity vector at that frequency. To determine the other errors, reflection tracking and the source match terms, two additional devices are used: a short circuit and an open circuit. To extract s_{11a} from our measured value, we use the measured (or extracted) values of directivity, frequency response and source match which we have stored in the analyzer.

This is what corrected data looks like. Notice the data has lost its ripple component and is much more realistic.

FIGURE 5

The effective improvement (for each of the errors we have discussed) is dependent upon the quality of standards we use in the calibration process. To determine the measurement accuracy, it is helpful to work a quick example. The one-port DUT we will use has an s_{11a} of 12 dB of Return Loss. We will use typical overall worse case error values for our system. The resulting value is the uncertainty: +/- 3 dB before accuracy enhancement and +/- .22 dB afterwards. Phase uncertainty is a function of both s_{11} (the result after calibration) and its uncertainty. The worse case phase error occurs when the uncertainty of s_{11} is perpendicular to the value of s_{11a}.

Finally, consider measuring two-port devices. In many cases, we are interested in both the reflection characteristics and the transmission characteristics of our unknown device. This presents some interesting interactions which we need to account for.

In making transmission measurements of our two-port device, we have all the errors discussed so far plus additional transmission mismatches which add with frequency response and a leakage signal (crosstalk). This leakage error becomes significant when the transmitted signal level is reduced substantially by the test device (i.e. high loss devices).

FIGURE 6

In summary, there are many levels of measurement calibration available for accurate measurements on a network analyzer. Which level chosen depends upon the device under test and the level of accuracy needed. A response calibration compensates for the difference in the magnitude and phase tracking. It does not, however, compensate for any port match, crosstalk, or directivity effects. A one-port calibration provides for complete measurement calibration for one-port devices. A full two-port calibration compensates for these reflection errors and additionally compensates for transmission errors, thus providing the maximum level of accuracy enhancement for two-port devices.

FIGURE 1

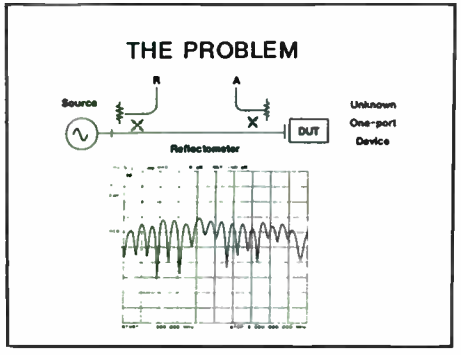


FIGURE 2

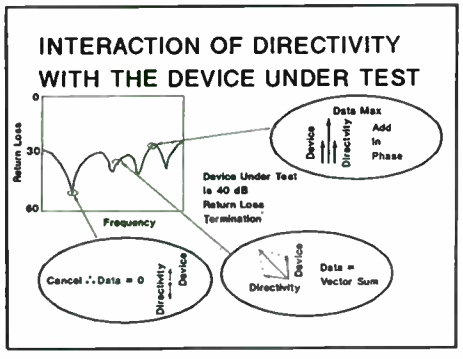


FIGURE 3

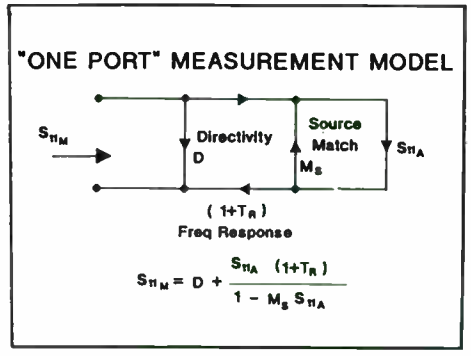


FIGURE 4

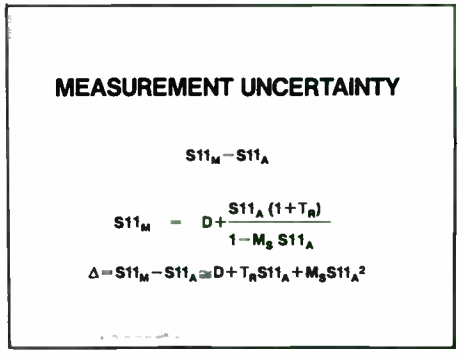


FIGURE 5

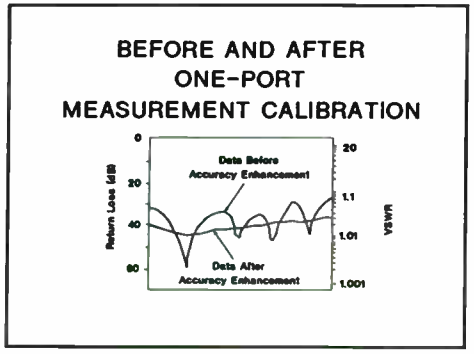
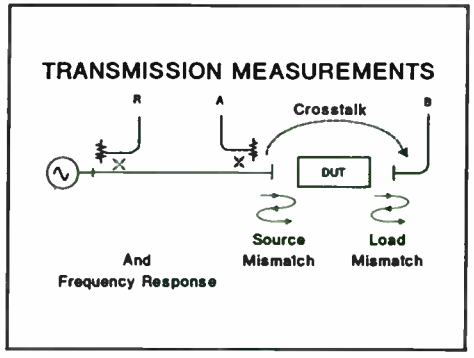


FIGURE 6



HIDDEN ELECTRONICS DETECTION

Michael K. Ferrand / John Bramick

Microlab/FXR
Ten Microlab Rd.
Livingston NJ 07039

Abstract

The recent bugging of the U.S. embassy in Moscow has caused widespread attention to be focused on eavesdropping as a modern security threat. A brief history of electronic eavesdropping is given, and techniques used by persons wishing to gain confidential information are described. A device capable of transmitting an extremely pure fundamental signal and "listening" to harmonic signal reflected by semiconductor devices will be explained and demonstrated.

History of Electronic Eavesdropping

Electric eavesdropping was born with the first use of electricity for communication. The telegraph was the first device to be spied on. During the Civil War telegraph wires were tapped to obtain battle strategies. No doubt that soon after this discovery coded messages and countermeasures were in place. By the 1880's relatively sophisticated methods of protecting both telephone and telegraph conversations were in use. In the 1890's electric eavesdropping became a standard tool of both law enforcement agencies and by persons whom today we would characterize as dishonest, but at the time did not violate any specific statutes. Throughout this century, vast improvements have been accomplished in electronic communication techniques. With the development of the transistor in the 1950's the high voltage power supplies needed to power vacuum tube were no longer needed. The development in the 1970's of the low-drain transistor, the long life battery, the micro-cassette, and the commercially available miniaturized voice actuator have given the eavesdropper another major step forward.

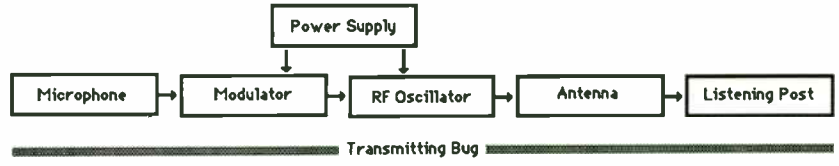
The Electronic Eavesdropping Detector

The problem of detecting electronic bugs has been of paramount concern to government agencies and law enforcement officials for many years. The device described, which we will call the "B1" uses the well known properties of semiconductors to locate listening devices whether or not they are in operation. It had been speculated for years that when a pure signal illuminated a semiconductor, harmonics or distortions of this pure signal would result. At that time, no one was able to confirm this speculation because all the available signal sources were so full of their own harmonics that the tiny harmonics that might result from an illuminated semiconductor were lost. Microlab/FXR investigated this problem in the late 1960's and then embarked on a research program that continues to this day. Using the technology we had developed in the microwave filter area, filters were developed to reduce the harmonic level of an output oscillator to 10^{-17} times the fundamental. Novel circuitry was developed to prevent harmonics from being developed inside the circuitry itself and to interpret the signals that might be returned from an illuminated area.

Bugging, Where, When and How

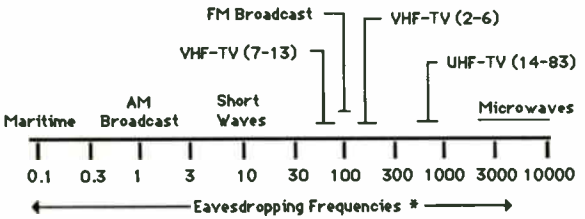
The RF transmitting bug is believed to be the most prevalent eavesdropping device by a wide margin. These are tiny devices that are reported as being placed inside the martini olive, inside fountain pens,walls,furniture, virtually anywhere the mind can imagine. The RF transmitter sends its signal to the eavesdropper's listening post where the information may be heard or recorded on tape.
In their most simple form, these bugs consist of the following:

1) a microphone
2) an oscillator to generate an RF signal that will serve as a carrier
3) a means to modulate the output of the microphone onto the RF signal.
4) an antenna to radiate the modulated RF signal to a distant listening post.
5) a power supply to drive the oscillator and the modulator.



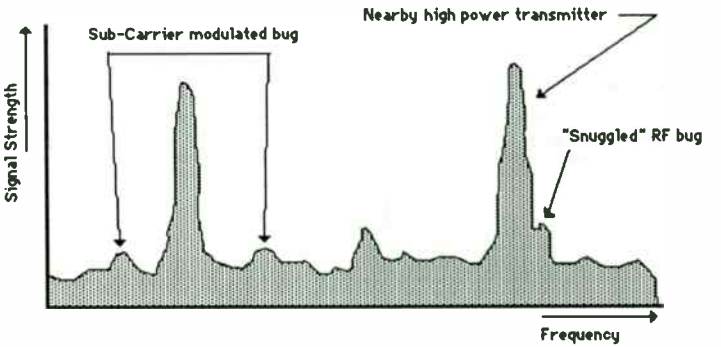
At first glance the detection of an RF bug with a radio receiver appears simple. Indeed a number of companies advertise receivers that "light up" whenever they are in the same room as a bug. After all, even a tiny bug must generate a detectable signal when you're right on top of it.

As a practical matter, RF transmitters are very difficult to detect with RF receivers. First, the eavesdropper can select his operating frequency anywhere from below the AM broadcast band up through the short wave, TV,FM, VHF, UHF, and on to the microwave region. Many thousands of signals can be detected over these frequencies at any time. Each signal must be separately examined to assure that it is not a threat.



* These frequencies represent 500,000 different eavesdropping channels, each having telephone grade bandwidth

The eavesdropper can further conceal his presence by selecting his operating frequency so as to "snuggle". In this case the bug frequency is placed as close as possible to that of a nearby high-power commercial transmitter. The associated listening post uses a highly selective narrow band receiver to separate the bug's signal from that of the transmitter. Countermeasures receiver must cover wide frequency ranges and generally cannot be made so selective.



"Snuggling" and Sub-Carrier Modulation

The eavesdropper can also minimize detection by using one or more techniques of RF switching to activate and deactivate the bug. These techniques keep the bug from generating unnecessary RF radiation which could lead to its detection. They also extend the life of the batteries in the bug. One such method is called voice actuation. This method automatically turns the bug on when there is a nearby conversation. When there is no conversation, the bug simply does not transmit and thus cannot be detected by a conventional "listening" device.

Remote switching is another method of RF switching. These bugs may remain absolutely silent for months or years, until the eavesdropper determines that he wants to listen in. Then the bug is remotely or electronically switched to it's "on" position and begins to transmit. The eavesdropper later turns the bug off when he has accomplished his mission. Remote switching makes it much more difficult to detect bugs by conventional receiver means. Storage and burst represents a third type of RF switching. An entire conversation is stored onto a tape recorder using a voice actuator to conserve tape. At a safe time the eavesdropper turns on the RF bug which transmits a playback of the tape to the listening post. This playback can even be accomplished at high speed so as to minimize the broadcast time required.

The eavesdropper can alternately employ timers to limit his RF transmission to preselected hours. A transmitter can also be triggered to transmit only when the lights are on and the room is occupied. Some of the foregoing methods of RF switching are totally beyond the control of a sweep team using a countermeasures receiver, such as storage, burst, timers and remote switching. Others, such as voice actuation and light switching, may be somewhat controlled by a sweep team but only at the expense of alerting

the eavesdropper.

Numerous methods of modulation are available to the eavesdropper, ranging from very simple to rather complex. The variety of these modulation methods assists the eavesdropper in his attempts to avoid detection by a countermeasures receiver. Some of these modulation methods are noted:

Amplitude modulation (Conventional or Single Sideband)
Frequency Modulation
Angle Modulation
Phase Modulation
Subcarrier modulation (including pulse-position, pulse-width, pulse-frequency, pulse-amplitude, and various pulse-code modulations)
Translated modulation

No countermeasures receiver is capable of verifying bugs using all of the available methods of modulation.

The microphone greatly influences the performance of any bug. More sensitive microphones with wider frequency response pick up more distant conversations. Directional microphones exclude sounds from unwanted directions. Audio filters exclude specific unwanted sounds. Microphone miniaturization makes visual detection more difficult.

All electronic eavesdropping devices require some form of power supply. The miniaturized long-life battery is by far the most widely used power supply because it is so small, readily obtained, and easily deployed. But the life of a battery is limited so other power supplies are also used.

Ordinary house current is convenient for the eavesdropper to power his bug. Thus bugs are often located in wall receptacles, light fixtures and the like because the eavesdropper there finds a reliable and unlimited source of power. He may also place a bug inside a wall and draw his voltage from any nearby power cable. Telephone wires also provide a reliable source of voltage for the bug. Thus RF and other types of bugs are often located inside telephone instruments, telephone wall boxes, telephone switching circuits, or along telephone wires. These bugs may or may not pick up the conversations on the telephone lines. They are primarily intended to use the drive power provided by the telephone. Other power supplies include solar batteries, acoustic or barometric batteries, and other sophisticated sources of energy. These supplies are not believed to be in common use at the present time. Every transmitting bug requires some form of antenna. Long antenna wires of several inches or longer are commonly used and often lead to the discovery of the bug. Smaller ferrite loaded antennas are also used, as are clips that attach to and use existing wires in a building.

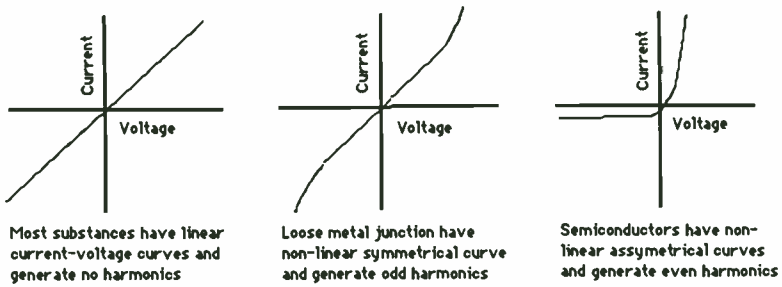
Several methods of packaging the bug are available to the eavesdropper. Generally the electronic components are mounted on a small circuit board which may be unprotected, wrapped in tape, contained within a metal, wooden, or plastic enclosure, or sealed in epoxy. Metal boxes are usually avoided because the eavesdropper wants to escape detection by a metal detector. Epoxy encapsulation is attractive because the bug may not be identified even if it is discovered.

Simple RF bugging transmitters are commercially available and sold under such names as "wireless microphones" and "baby sitters". The most sophisticated transmitting bugs are readily constructed by people with minimal electronic background. The seemingly high technology required and the illegality of the RF bug does not prevent it's wide deployment. The technology that is used in the B1 unit does not rely in any way on the RF transmission coming from the bug.

Theory of Operation

The B1 unit uses an electronic detection system that illuminates a suspected area with a low power UHF signal. The reflected signal is examined for harmonic content. All transistors and other modern solid state devices produce a positive response from the device. Normally only these devices produce this positive response. The magnitude of this positive response increases as the electronics detector is moved closer to the solid-state device even when the device is completely hidden from vision and whether or not such a device is operating.

Corroded metal parts in loose contact with each other produce a negative response from the B1 unit. This non-linear phenomenon has been termed the "rusty bolt" effect and is a source of electromagnetic interference. Only such corroded parts can produce this negative response, the amplitude of which generally varies rapidly with time. All other objects such as wall, ceilings, furniture, fixtures and personnel produce no returned harmonic signal.



Most substances have linear current-voltage curves and generate no harmonics

Loose metal junction have non-linear symmetrical curve and generate odd harmonics

Semiconductors have non-linear assymetrical curves and generate even harmonics

Current-Voltage curves for various materials

The B1 evaluates the returned signal from the returned area and determines if any solid-state devices or corroded metal parts are in the suspected area. The appropriate readout is employed to display the desired information. Typically, a device containing transistors can be detected at a range of 5 to 10 feet and corroded metal parts detected at a range of 3 to 5 feet. The B1 electronic detection system can be broken down into the following sub-systems:

A. A solid-state microwave transmitter
B. Antenna assembly
C. A microwave receiver
D. IF Amplifier circuits
E. Processing circuits
F. An AC-DC power supply

The Microwave Transmitter

The fundamental transmitter power is generated by a transistor oscillator operating at a nominal frequency of 915 MHz. It produces a CW power output of 500 mW.

The output of the oscillator is fed through a 15 dB. directional coupler to a ferrite isolator, which prevents frequency pulling caused by large impedance mismatches at the face of the antenna. The side arm of the coupler provides a reference signal for a phase locked loop (PLL) from which the receiver local oscillator power is derived. The output of the isolator in the main line of the transmitter is fed to an electrically variable attenuator. This circuit consists of PIN diodes in a balanced configuration so that a reasonable RF impedance match results at all attenuation levels. When the bias on the attenuator is set for low power output most of the unused power is absorbed by an internal termination. The bias for the attenuator is obtained from a driver circuit in the indicator box. The output of the variable attenuator is fed through harmonic band reject filters and a lowpass filter to a TNC connector on the front panel of the housing. These filters provide over a 120 dB. of attenuation to signals at the harmonic frequencies while allowing the fundamental signal to pass.

All the transmitter components, except for the transistor oscillator are packaged in a single transmitter stripline assembly. The complete transmitter, including the regulator for the oscillator is contained in a completely shielded module of the B1 unit.

Antenna Assembly

The output of the transmitter module is fed to the transmit port of the boom antenna assembly. This assembly contains a resonant slot antenna for transmitting and a separate slot for each of the second and third harmonic receive frequencies. The transmit signal is fed through a pair of band-reject filters to a balun where the unbalanced impedance of the input coaxial line is transformed to a balanced impedance at the slot. The band-reject filters stop any second or third harmonic signals which may be generated in the cables or connectors between the electronics unit and the antenna. Second and/or third harmonic signals which are generated by a target are captured by the receive slots which are diplexed together onto a common transmission line. Connected in this line is a band-reject filter to provide greater rejection to the transmitter signal than is provided by the decoupling of the antenna elements. This rejection prevents sufficient transmitter power from reaching the antenna receiver port to generate harmonics in the connectors or cable between the antenna and the receiver port of the B1 unit.

Microwave Receiver

The receiver input to the B1 unit is connected to a diplexer which splits the input into two bands. One is at the second harmonic and the other is at the third harmonic of the transmitted signal. Bandpass filters in the diplexer are tuned to each harmonic and provide further rejection of the transmitted signal. The output of each filter is fed to the signal input port of a balanced mixer circuit which provides over 15 dB. rejection of local oscillator signals at the receiver input port and also very high attenuation at

the IF port to AM signals which may be present at the local oscillator port.

The coupled arm of the transmitter directional coupler is connected to a lowpass filter, frequency multiplier and bandpass filter circuits. The lowpass filter allows only fundamental signals to pass to the multiplier and also prevents harmonics which are generated from flowing back to the source. The bandpass filter is tuned to the fourth harmonic and allows only this signal to pass to a fixed attenuator. The level of the fourth harmonic output signal is as low is as practical to minimize feed through or leakage of this signal into the receiver.

The fourth harmonic output from the transmitter module is the input signal to a balanced mixer in the PLL module, operating at a frequency of 3660 MHz. The local oscillator (VCO) which operates at a frequency of 914.5 MHz. (nominal). The output of the VCO is fed through a 3 dB. coupler, a times-four multiplier circuit, similar to one in the input line, and an attenuator which sets the proper level. The difference frequency output of the PLL mixer is used as the input signal in the PLL comparison circuits. This signal is amplified and limited and is applied to a 2 MHz. filter. The output of the filter is fed to a phase detector where it is compared to a 2 MHz. reference crystal oscillator. Any phase offset between the two signals causes voltage output which is amplified in an offset amplifier and applied to a summing amplifier.

The other input to the summing amplifier is a sawtooth sweep voltage from an astable multivibrator. The output of the limiter above is also fed to a narrow band filter tuned a 2 MHz. The filter output is integrated and then drives a switch which disables the multivibrator sweep circuit when the difference signal out the PLL mixer approximates 2 MHz.

The output of the summing amplifier is fed to a loop lowpass filter which provides a PLL lock-in range of 100 KHz. The filter output is applied to the frequency control input of the VCO.

When the equipment is turned on, the frequencies of the VCO and transmitter oscillator are typically several MHz. apart. The output of the PLL mixer will not be 2 MHz. and no signal is available out of the 2 MHz. filters for the phase detector ot the sweep channel switch. Therefore, the sweep voltage operates, causing the VCO to be swept through the transmitter reference frequency. As the VCO frequency approaches the reference frequency a signal builds at the output of the sweep channel filter and integrater and begins to turn off the sweep circuit. As the difference frequency between the two oscillators approaches 0.5 MHz. (2 MHz. at the PLL mixer output) the phase detector loop is activated and the output of the offset amplifier is such as to hold the required frequency difference. At the same time the sweep voltage is completely disabled. The VCO is now phase locked at a frequency which is 0.5 MHz. away from the transmitter frequency.

The other output of the 3 dB. coupler at the output of the VCO is fed in the receiver module through a lowpass filter to a frequency multiplier. Two multiplier outputs are fed through bandpass filters tuned to the second and third harmonics, and fixed attenuators. These output signals serve as the local oscillator signals for the receiver balanced mixers. The microwave components of the receiver are packaged in a single receiver stripline assembly. The PLL microwave components are contained in a single PLL stripline assembly and the electric components are contained on a single printed circuit board. Both circuits plus the VCO are packaged on a completely shielded module of the B1.

The transmitter input signal and the VCO output are frequency multiplied by four in the PLL module so that the difference frequency (2 MHz. in this case) is higher then either the IF frequency in the receiver. In this way harmonics of the difference frequency do not fall into the IF bands and cause spurious responses.

IF Amplifier Circuits

The output of the second harmonic mixer is a signal at 1 MHz. (2 times the transmitter-VCO offset frequency). This signal is fed through a mixer bias circuit and an IF pre-amplifier. The level at the output of this preamp is sufficiently high so that further noise contributions are negligible and the receiver noise figure is established.

The IF signal is further amplified by 25 dB. and fed to a 1.0 MHz. crystal filter through an impedance matching driver stage. The filter bandwidth is approximately 150 Hz. which is the noise bandwidth of the receiver. The filtered output signal is further amplified and applied to an amplitude detector.

The output of the third harmonic mixer is a signal at 1.5 MHz. This signal is processed similarly to the second harmonic channel with exception that the crystal filter is tuned to 1.5 MHz. Also an electronic switch is included at the output of this channel so that the third harmonic output can be disabled.

Processing Circuits

The outputs of the amplitude detectors of the two channels are combined in a difference comparator circuit. If the amplitude of the second

harmonic channel is larger the the third harmonic channel a positive output results from the comparator. If the third harmonic output is larger, the comparator output is negative. This output signal is fed to three indicator circuits.

One output feeds a 1 Hz. lowpass filter which provides post detection integration. This integration improves system sensitivity by restricting the noise bandwidth and smooths out amplitude variations due to rapidly changing input signal levels. The output of the filter is fed to a threshold comparator where its amplitude is compared to a pre-set reference level. When the signal is more negative (the second harmonic exceeds the third) then the reference, an indicator light driver is triggered which illuminates an LED. The illumination of this light indicates that a semiconductor target has been detected. The output of the light is also available as a external output.

A second output of the difference comparator is fed to a log driver which reverses the phase and provides a logarithmic signal variation on a zero center scale of the signal strength meter. The meter reads to the right (positive) when the second harmonic exceeds the third. The larger the reading the greater the difference between the second and third harmonic. Conversely a reading to the left (negative) signifies a larger third harmonic return the target. A pot is provided in series with the meter to make fine adjustments of the meter scale factor.

When the COMPARE switch on the front of the B1 unit is off (disabling the third harmonic channel) the signal strength meter only reads in the positive direction.

The third output of the difference comparator is fed through an electronic switch and serves as a control voltage to a amplitude detector and bandpass amplifier stage. The input to this stage is obtained from the broadband output of the 25 dB. amplifier in the second harmonic channel. The 1 MHz. carrier is fed through a bandpass filter to remove spurious signals and is detected to extract any amplitude modulation which is present. The modulation components are fed through a headphone driver stage which has a passband from 500 to 2000 Hz. The output audio signals are fed through a VOLUME control and a HEADSET phone jack on the B1 unit.

The output of the threshold comparator in the detection indicator light circuit serves as an enable voltage for the electronic switch in the control voltage line. When the threshold comparator output triggers the light driver , the switch is also closed. This applies the difference comparator output as a control bias to the amplitude detector of the audio channel. An increasing negative bias voltage (second harmonic exceeds the third) causes the detector operating point to increase which raises the level of the modulation output. In this manner the level of the modulation does not change until the light driver is triggered and then it increases rather rapidly with increasing positive signal strength indication on the meter until saturation occurs.

The bias for the variable attenuator is derived from a voltage at the wiper of a potentiometer which is connected between plus and minus 9 volts through a series resistor string. One resistor of the string has a switch connected across it. This switch is activated by the potentiometer shaft and is open for the full CW rotation of the shaft. As the shaft is turned CCW the switch closes which changes the bias of the attenuator resulting in 4 dB. output power reduction. Further CCW rotation continuously reduces the power output until a minimum of 20 dB. attenuation is provided at the full CCW stop. The voltage at the potentiometer wiper is fed to the non-inverting input of an op. amp. driver. The feedback around this stage is chosen to linearize the attenuation verses shaft rotation. The output of the driver is fed to the bias terminal of the transmitter stripline assembly. The receiver stripline assembly and the IF amplifier and processor printed circuit boards are packaged in a completely shielded module of the B1 unit.

AC-DC Power Supply

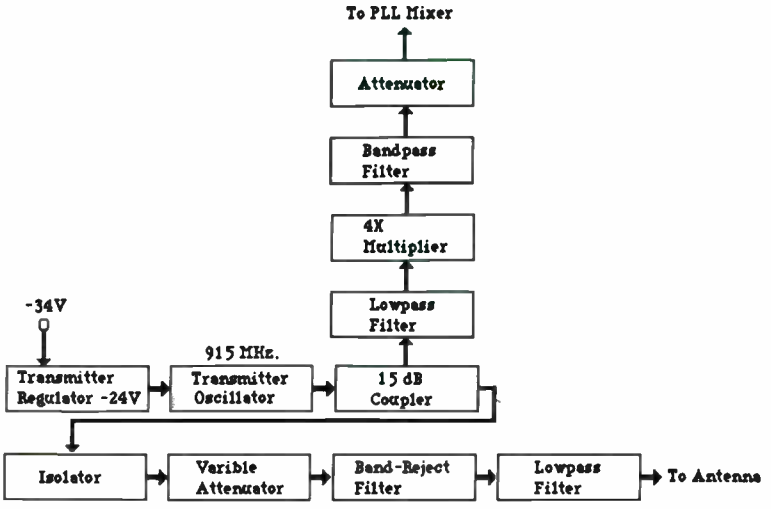
A conventional AC-DC power supply provides -34V and +/- 17V unregulated to the B1. The -34V line is fed to a -24V regulator in the transmitter module which provides the operating voltage for the transmitter oscillator. It is also fed to a second -24V regulator in the PLL module which provides the collector voltage for the VCO. The +/- 17V lines are fed to separate +/- 9V balanced regulators. One regulator provides the operating voltages to the receive and indicator processing circuits and the other regulator provides the operating voltages to the PLL circuits.

The AC supply also contains an output voltage for charging the battery pack. The charge current is 450 mA during most of the charge time. As the batteries approach full charge the temperature rises. There is circuitry contained within the battery pack which senses temperature. This circuit applies a voltage to the AC power supply which is proportional to temperature. When this voltage exceeds a preset reference, a driver stage is turned off which opens a relay. The open contacts of the relay place additional series resistance in the charge circuit which reduces the charge current to a trickle rate.

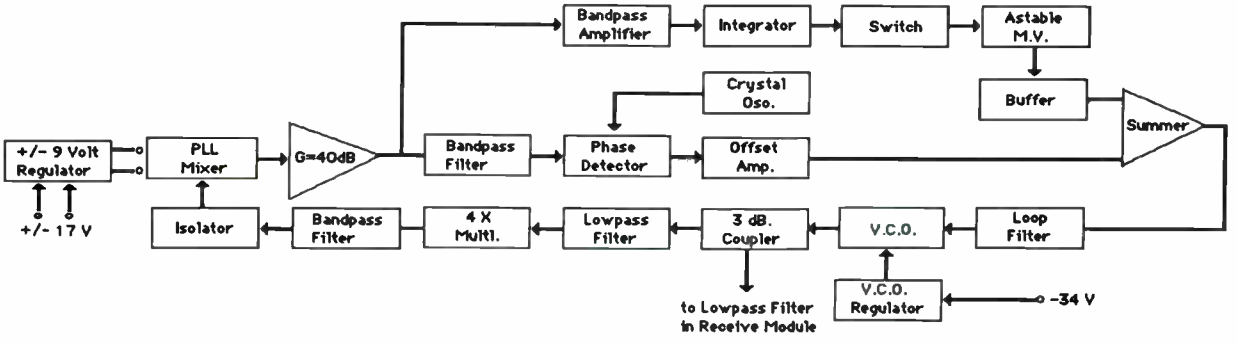
All of the components of the power supply are packaged in a removable module which plugs into the B1 unit.

Conclusion

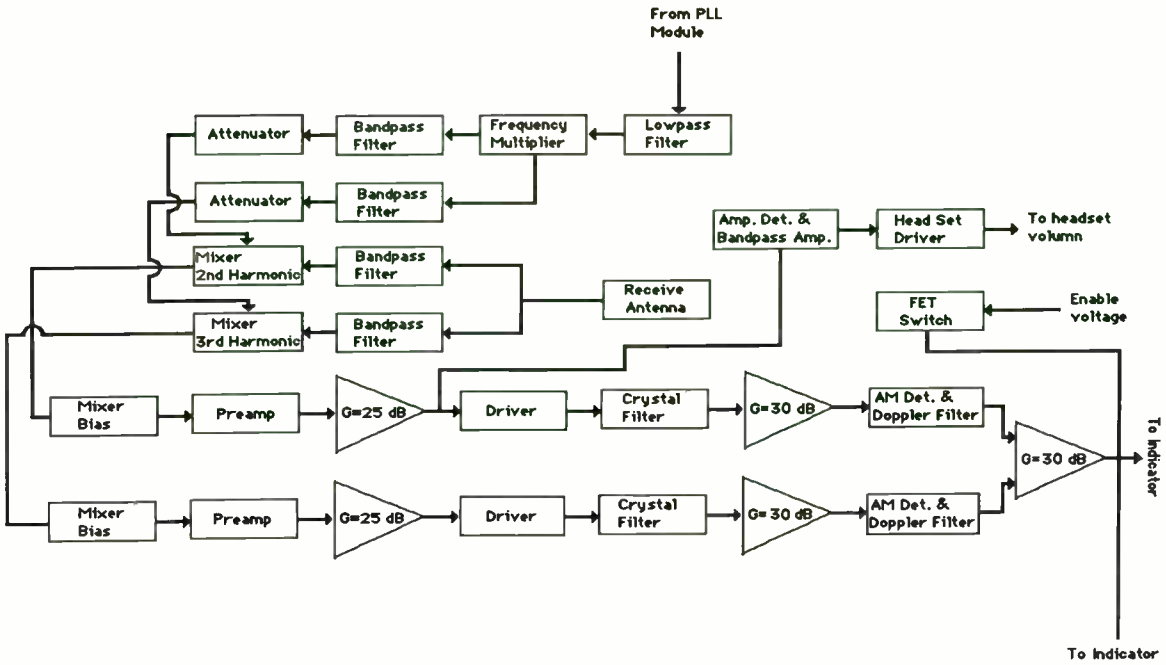
A novel approach for detecting electronic eavesdropping devices has been described. The increasing sophistication of the devices used for bugging has spurned new methods for protecting confidential information. Based on a relatively simple phenomenon exhibited by all semiconductor devices, this device has overcome some basic problems associated with traditional methods of electronic eavesdropping detection.



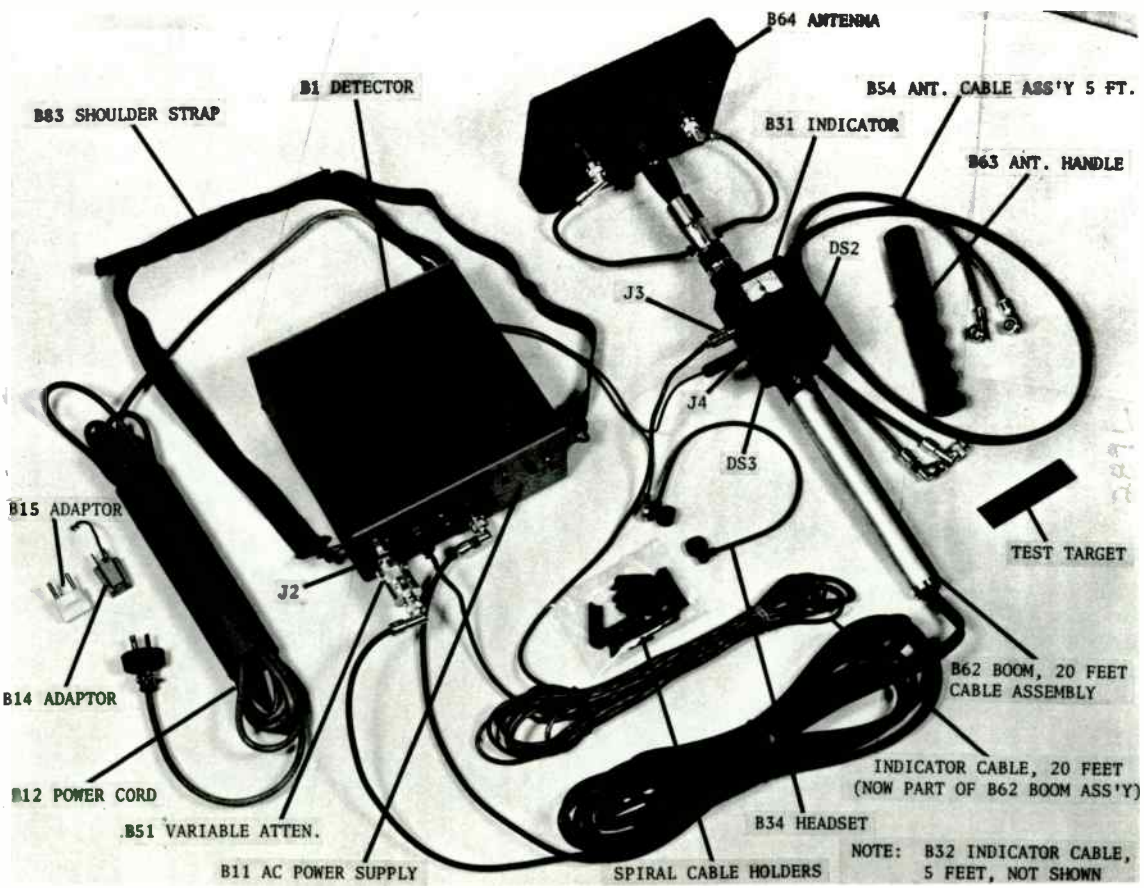
Transmitt Module



PLL Module



Receive Module



A Glass Packaged Varactor As A Hi-Reliability Device

by

John C. Howe
Product Engineer, Motorola Inc.
5005 East McDowell Road
Phoenix, Arizona 85008

INTRODUCTION

The DO204AA (also known as DO7) glass package is an ideal vehicle for the encapsulation of a varactor diode die to be used in an environment requiring hi-reliability. The package provides the die with a hermetically sealed inert atmosphere. The die is metallurgically bonded on both sides, eliminating the inherently unreliable pressure contacts.

PIECE PARTS

The DO204AA package consists of three (3) piece parts. These are the glass envelope assembly (Fig. 1), solder coated preform (Fig. 2) and the glass beaded and tapered lead (Fig. 3). The glass envelope assembly consists of a borated G12 glass tube, open at one end, which is sealed at the other end to a .040" dia. Dumet (per ASTM F29) stud. This stud is in turn welded to a .020" dia. Dumet lead. All metal surfaces are electroless nickel plated to a nominal thickness of 100 micro inches . The second part of the package is a .0040" thick iron-nickel cobalt alloy (per ASTM F15) preform which is coated on both sides with 650 micro inches of 5-95 tin/lead solder. This preform is used to provide a thermal match between the silicon die and the Dumet stud. The third part of the package is a .020" dia. glass beaded and tapered Dumet lead. The lead is electroless nickel plated to a nominal thickness of 100 micro inches.

ASSEMBLY

Fig. 4 is a simplified assembly flow diagram for the DO204AA package. The first step is to load the solder coated preform and die into the glass envelope assembly. These assemblies are then run through a belt furnace using appropriate fixtures at a temperature sufficient to melt the solder. The second step is to coat the tapered end of the glass beaded lead with 10-90 tin/lead solder. The glass beaded and tapered lead is then inserted into the glass envelope assembly. These assemblies are then placed in a programmable sealer using graphite fixtures. The sealer makes the final glass to glass seal and reflows the 10-90 tin/lead solder from the tapered end of the glass beaded lead on to the die. Fig. 5 shows the internal construction of a fully assembled DO204AA, and Fig. 6 is a dimensional outline.

PRODUCTION TEST AND SCREENING REQUIREMENTS

The initial electrical test of the assembled package can be accomplished using a handler designed for an axial lead package which can be interfaced to automatic test equipment designed to test discrete components. Since capacitance and capacitance ratio measured between two applied voltages are important parameters of a varactor diode, it is critical that the handler be capable of being interfaced to a fast response capacitance meter. Following the initial electrical test, all product is screened according to the requirements of Table I. The product then undergoes a final electrical test, after which it is body coated, marked and serialized in preparation for transfer to Quality Assurance.

QUALITY ASSURANCE PROCESSING

Table II shows the 100% processing accomplished by Quality Assurance. Table III

describes the electrical tests to be performed before and after HTRB and the limits against which the parts are measured. The lot which this paper describes had an actual percent defective of approximately one order of magnitude less than the PDA of 10% to limit 2 of Table III after HTRB. The lot is then subjected to a 100% electrical test to Table IV, Subgroups A3 and A4. The actual percent defectives for this inspection was zero. A defective is defined as a unit which fails an electrical limit required by Table III and/or IV.

GROUP A, B, AND C INSPECTION

Group A inspection was performed per Table IV to a combined LTPD of 3 with a maximum accept number of 1. There were no rejects from the Group A inspection. All parts tested as part of Group A inspection had their electrical parameters read and recorded. These readings provided initial electrical data for Group B and C post inspection testing. Table V describes Group B inspection. The parts subjected to Subgroups 3 and 4 were tested to Table III, limit 1 upon completion of the inspections in their subgroups. There were zero rejects in all subgroups of the Group B inspection. Table VI describes Group C inspection. All subgroups were tested to Table III, limit 1 upon completion of their inspections. There was one mechanical reject in Subgroup 1 due to the terminal strength screen. (Refer to Table VI - Group C Inspection: Subgroup 1, Mil-Std. 750 Method 2036.) There were no electrical rejects. Subgroup 3 was a continuation of Group B Subgroup 4 for a total of 1000 hours HTRB.

SUMMARY

The results of this testing show that this glass axial lead package is well-suited to the requirments of hi-reliability applications. Because of its relatively small size, low weight, and hermetic seal, varactors assembled in this package find application in satellites, missile guidance systems, and all types of communications equipment requiring high reliability.

DEFINITIONS OF ABBREVIATIONS

HTRB: High Temperature Reverse Bias

PDA: Percent Defective Allowed

LTPD: Lot Tolerance Percent Defective

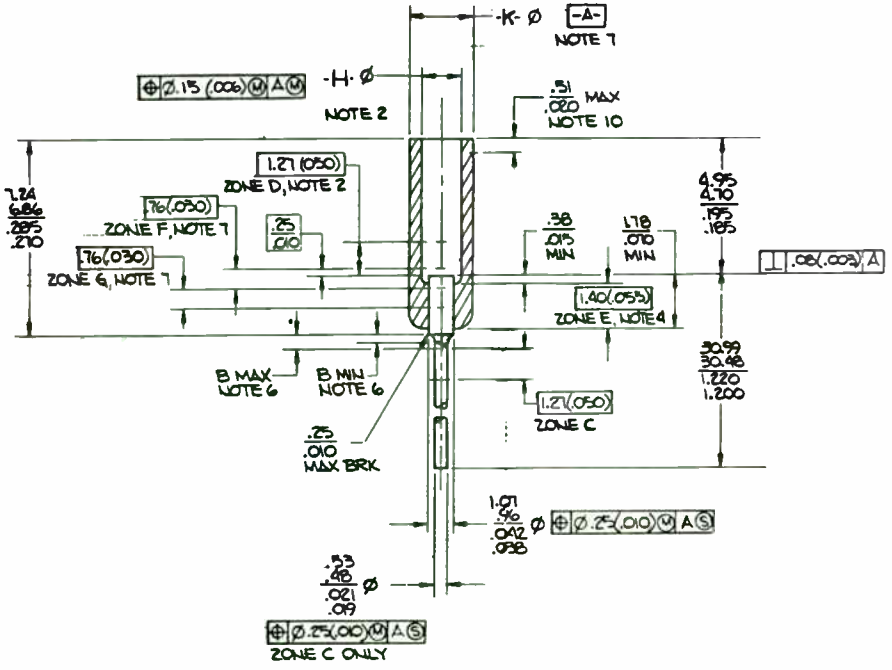


FIGURE 1: GLASS ENVELOPE ASSY

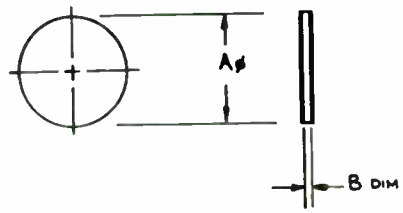


FIGURE 2: SOLDER COATED PREFORM

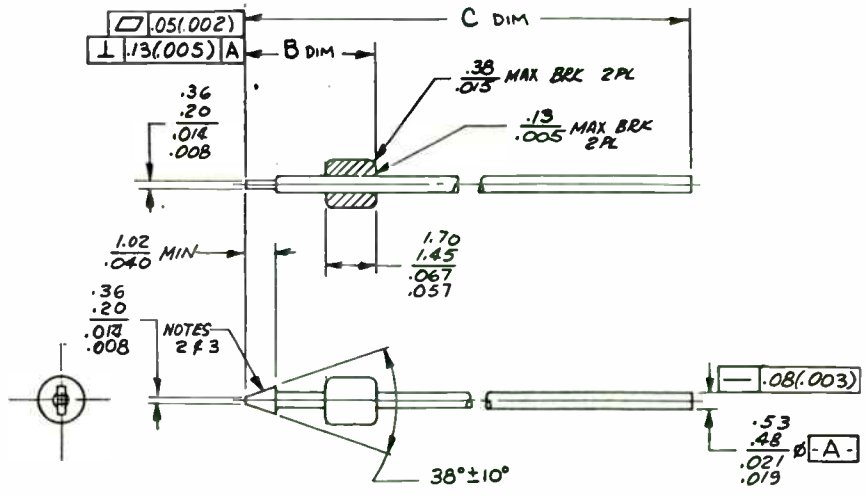


FIGURE 3: BEADED/TAPERED LEAD

SUBJECT/ACTION	FLOW SYMBOL
DISC, SOLDER	
INCOMING QA	
GLASS SEAL	
INCOMING QA	
DIE	
SOLDER ASSEMBLY	
BEADED LEAD	
INCOMING QA	
SOLDER DIP	
QA INSPECT	
SEAL	
IN PROCESS Q.A.	
COUNT	
QA VISUAL & GATE	
PLATING (LEAD FINISH)	
TRANSFER TO FINAL TEST.	

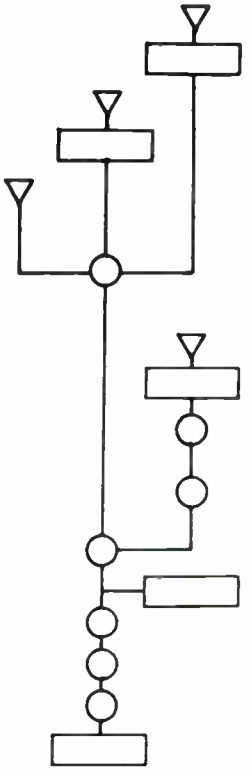


FIGURE 4: ASSEMBLY FLOW DIAGRAM

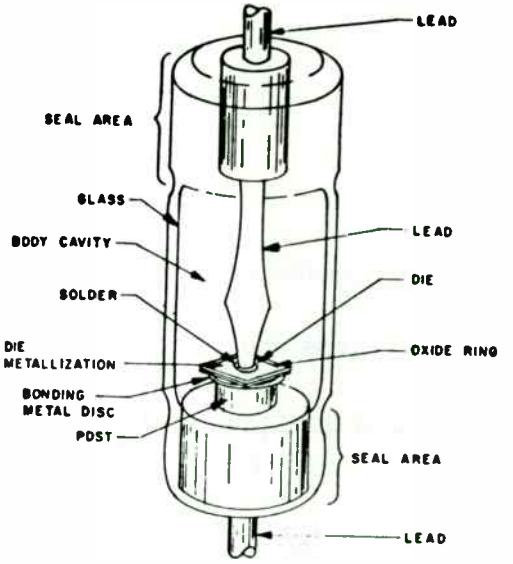
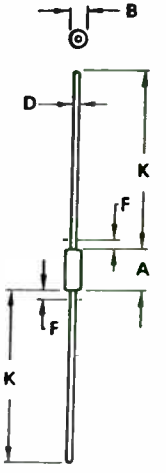


FIGURE 5: INTERNAL CONSTRUCTION



DIM	MILLIMETERS		INCHES	
	MIN	MAX	MIN	MAX
A	5.84	7.62	0.230	0.300
B	2.16	2.72	0.085	0.107
D	0.46	0.56	0.018	0.022
F	–	1.27	–	0.050
K	25.40	38.10	1.000	1.500

FIGURE 6: DIMENSIONAL OUTLINE

TABLE I

PRODUCTION SCREENING

SCREEN	MIL-STD 750 METHOD	CONDITION
Internal Visual Inspection	2074	
High Temp Storage	1032	t = 24 Hours T_A = 150 °C
Thermal Shock (Temp Cycling)	1051	Cond. C 20 Cycles t (Extremes) = 15 min.
Constant Acceleration	2006	20 K G Y1 Direction No one minute hold time
Hermetic Seal Fine Leak	1071	G Leak Rate $\leq$ 5 X 10^{-8} ATM cc/sec
Gross Leak		D

TABLE II

QUALITY ASSURANCE SCREENING

SCREEN	MIL-STD 750 METHOD	CONDITION
Elec. Test		Table III Limit 1
HTRB	1038-A	T_A = 150°C t = 240 Hrs. VR = 10V
Elec. Test		Table III Limit 2 PDA = 10%
Elec. Test		Table IV PDA = 10%
External Visual	2071	

ELECTRICAL TABLE III*

SYMBOL METHOD MIL-STD 750	CONDITIONS T_A 25 °C UNLESS OTHERWISE SPECIFIED	INITIAL 1 LIMITS MIN	INITIAL 1 LIMITS MAX	END POINTS 2 LIMITS MIN	END POINTS 2 LIMITS MAX	UNITS	
BVR1 4021	IR = 10uA	12	--	12	--	V	
ΔBVR1	--	--	--	--	±15%	change from initial	
IR1 4021	VR = 10V	--	0.10	--	0.10	uA	
ΔIR1	-- (Whichever is greater)	--	--	--	±100% or ±50	change from initial nA	
CT 4001	VR = 2.0V, f = 1.0MHz	200	300	200	300	pF	

* END POINT TABLE, 100% HTRB, GROUP B, GROUP C

GROUP A INSPECTION - TABLE IV

SYMBOL METHOD MIL-STD 750	CONDITIONS TA=25°C UNLESS OTHERWISE SPECIFIED	LIMITS MIN	LIMITS MAX	UNITS	ACC/LTPD	TEST SAMPLE SIZE	ACC. ON
	SUBGROUP A-1		LTPD / MAX. ACC.		15 / 0		
2071	VISUAL MECHANICAL INSPECTION	--	--	-			
	SUBGROUP A-2		* LTPD / MAX. ACC.		3 / 1		
BVR1 4021	IR = 10uA	12	--	V			
IR1 4016	VR = 10V	--	0.10	uA			
	SUBGROUP A-3		* LTPD / MAX. ACC.		3 / 1		
BVR2 4021	IR = 10uA, TA = 100°C	12	--	V			
IR2 4016	VR = 10V, TA = 100°C	--	20.0	uA			
BVR3 4021	IR = 10uA, TA = 0°C	12	--	V			
IR3 4016	VR = 10V, TA = 0°C	--	0.10	uA			
	SUBGROUP A-4		* LTPD / MAX. ACC.		3 / 1		
CT 4001	VR = 2.0V, f = 1.0MHz	200	300	pF			
Q 4036	VR = 2.0V, f = 1.0MHz 1/	200	--	-			
TR	f = 1.0MHz 2/	10	--	-			

1) $Q = 2\pi f C/G$ 2) TR = CT(2V)/CT(10V)

TABLE V

GROUP B INSPECTION

SCREEN	MIL-STD 750 METHOD	CONDITION	LTPD
Subgroup 1	2066		10/1
Physical Dimensions			
Subgroup 2			15/1
Solderability	2026		
Resistance to Solvents	1022		
Subgroup 3			10/1
Temp Cycling	1051	F 10 Cycles t (Extremes) $\geq$ 15 min.	
Hermetic Seal Fine Leak	1071	G Leak Rate $\leq$ 5×10^{-8} ATM cc/sec	
Gross Leak		D	
Elec. Test		Table III Limit 1	
Subgroup 4			10/1
HTRB	1038-A	T_A = 150°C t = 340 Hrs. VR = 10V	
Elec. Test		Table III Limit 1	

TABLE VI

GROUP C INSPECTION

SCREEN	MIL-STD 750 METHOD	CONDITION	LTPD
Subgroup 1			10/1
Thermal Shock (Temp Cycling)	1051	F 100 Cycles t (Extremes) $\geq$10 min.	
Terminal Strength Tension Lead Fatigue	2036	 A 8 oz., 15 sec. E	
Hermetic Seal Fine Leak	1071	 G Leak Rate 5 X 10^{-8} ATM cc/sec	
Gross Leak		D	
Moisture Resistance	1021	Omit Init. Cond.	
External Visual	2071		
Elec. Test		Table III Limit 1	
Subgroup 2			10/1
Shock	2016	1500G 0.5 mS 5 Blows Y1 Axis	
Vibration, Var. Freq.	2056		
Constant Acceleration	2006	20KG One Minute Min. in Y1 Axis	
Elec. Test		Table III Limit 1	
Subgroup 3			10/1
HTRB	1038-A	T_A = 150°C t = 660 Hrs. VR = 10V	
Elec. Test		Table III Limit 1	

Pin-Fet Primer
Regarding
Analog/RF Considerations

Jack Koscinski
General Optronics Corp.
4 Olsen Avenue
Edison, New Jersey 08820

I. Scope

In the design of an optical fiber communications system, one of the key elements is the receiver. The front-end of an optical receiver consists of a photo-detector and an associated preamplifier. The function of the photo-detector is to detect the incident light and convert it into an electric current. The preamplifier then multiplies the current into a usable signal while introducing a minimum amount of noise.

The focus of this paper will be to review a very popular (long wavelength) optical receiver preamplifier. This is the Pin-Fet. Both "circuit theory of operation" and "application considerations" will be reviewed with the main objective of informing the user/designer on the performance tradeoffs with Pin-Fets.

II. Introduction

Although there are several types of lasers and fiber available, this discussion refers exclusively to 1300 nm single mode lasers and single mode fibers. These selections are becoming de facto standards for fiber optic analog/RF systems since they provide low loss links which are immune from multi-mode noise phenomena (modal noise).

Before starting a detailed discussion of the Pin-Fet preamplifier, it is useful to briefly review the generic block diagram of a fiber optic analog/RF link. This will highlight the major system components and identify where the Pin-Fet is used in the fiber optic system.

A block diagram for an FDM (frequency division multiplexed) analog/RF fiber optic link is illustrated in Fig. 1. The major system components are: 1) the optical transmitter, 2) fiber optic cable, and 3) the optical receiver.

1. Optical Transmitter

An optical transmitter accepts the individual RF FDM analog/RF inputs and provides the signal conditioning necessary to drive the semiconductor laser diode. Fig. 2 shows the major optical transmitter functional components.

First, the RF combiner sums the multiple analog inputs which are to be transmitted. The RF levels for each carrier should be equalized prior to the optical transmitter. Otherwise, optimum noise and distortion performance for each carrier will not be achieved. However, a slope compensation stage may be provided after the combiner stage to adjust for normal cable slope effects.

Broadband amplifiers, AGC controlled, provide the necessary signal level to drive the laser diode. The composite RF drive level to the laser must be precisely controlled to optimize system noise and distortion performance.

A DC bias is applied to the laser to provide a linear operating point. This DC bias current will determine the average optical output power out of the laser diode which is typically 0.5 mw for single mode 1300 nm lasers.

Laser power is sensitive to changes in temperature and laser aging. To preserve a constant average optical output power, two control circuits are commonly provided in the laser transmitter; a laser temperature controller and an automatic optical power controller.

The optical transmitter may have a significant impact on system noise and distortion performance. In an analog/RF system, the time varying composite RF signal s(t) directly modulates the optical source about a bias current point I(B) as shown in the laser LI (light versus current) curve (Fig. 3). With no input signal, the optical output power is Pave. When the signal s(t) is applied, the optical output power P(t) is:

$$P(t) = Pave\ [1 + m\ s(t)] \quad (1)$$

m = modulation index

$$m = \Delta I / I(B') \quad (2)$$

$$I(B') = I(B) - I(T) \quad (3)$$

I(B) = Laser Bias Current

I(T) = Laser Threshhold current

The current ΔI is the RF current modulation about the laser bias point.

To prevent distortions in the output signal, the modulation must be confined to the linear region of the curve. Furthermore, if I is greater than I(B'), the lower portion of the signal extends below threshhold and severe distortion results. Typical (m) values for analog applications range from (.25) to (.50).

However, using a higher modulation index (m) will provide a higher RF carrier/noise ratio since the received RF carrier is proportional to (m). Thus, a direct tradeoff exists between system distortion and noise performance as contributed by the optical transmitter.

Inherent laser source noise defines the maximum achievable carrier to noise obtainable in a fiber optic system. Minute fluctuations in optical emission are exhibited when biased above threshhold. This noise phenomena is referred to a relative intensity noise (RIN). The intensity noise is neither thermal nor strictly shot noise in nature. It is the response of the laser to modulation by the intrinsic shot noise which results from the granular nature of light and electricity. Typical values of laser RIN are (-120) to (-140) dB/Hz. The resultant carrier/noise ratio of a laser will be 3 dB less than the RIN component (m=1).

2.Fiber Cable

Single mode, 1300 nm fiber cable is recommended for analog RF systems. Single mode fibers have optical losses of less than 1 dB/Km at 1300 nm. Connectorization and splicing practices for single mode fiber has proven to be operationally effective with field installation equipment readily available.

3.Optical Receiver

An analog/RF fiber optic receiver is shown in Fig. 4. The Pin-Fet is used as the front-end of this receiver. A Pin-Fet is simply a unity-gain pin-diode coupled to a preamplifier using a low-noise, field-effect transistor (FET).

A post amplifier, with AGC, follows the pre-amp to provide sufficient gain to establish a unity "system gain" for the FO link. AGC is utilized to maintain constant output power independent of optical loss variations which may occur due to fiber resplicing, adding an optical tap etc.

III.Pin-Fets

1.Types

There are two popular types of Pin-Fet receivers: high-impedance (HZ) and transimpedance (TZ). They differ primarily in their use of the circuit resistance employed in the amplifying circuits.

In the high-impedance design (Fig.5) the photocurrent from the pin-diode is passed through a very high resistance. The signal voltage generated across the resistance is than amplified by a voltage amplifier. The main advantage of this design is that,

because of the high circuit impedances, this design generates very low levels of circuit noise. High-impedance Pin-Fets have the lowest receiver noise levels achievable.

However, the high-impedance creates several problems: bandwidth and dynamic range. Wide bandwidth is inconsistant with the long RC time constant (pole) which develops at the amplifier input due to the unavoidable input capacitances of the amplifier and pin-diode. High-impedance designs must be followed by an equalizing amplifier to realize flat responses at bandwidths up to several hundred Megahertz.

Dynamic range is also a limitation of the high-impedance design. Because of the long RC time constant, the signal may be integrated at the front end. Thus, when strong low frequences components are present, the front end is prone to overload (especially in digital applications where long "1" or "0" bit strings are present).

Transimpedance Pin-Fets provide an elegant solution which avoids these bandwidth and dynamic range limitations through utilization of negative feedback. In this design (Fig. 6), the detected photocurrent is fed into a current-to-voltage amplifier circuit which is a high gain, high impedance amplifier utilizing shunt feedback. Although the transimpedance amplifier is less sensitive than the high-impedance amplifier, the difference is usually only (2-3) dB for most wideband designs.

Since transimpedance Pin-fet's provide low noise (high sensitivities) along with wide bandwidths and high dynamic range, they are generally more popular than high-impedance Pin-Fets. The remaining sections of this paper will deal exclusively with the design and application considerations of transimpedance Pin-Fets.

2.Circuit/Component Considerations

A schematic of a transimpedance Pin-Fet circuit is shown in Figure 7. The two basic sections consist of the "optical photodetector" (pin-diode) and the "preamplifier circuit".

A.Pin-diodes

The photosensitive element used in the Pin-Fet circuit is a low-leakage, high efficiency, highly reliable InGaAs PIN photodiode. Pin-diodes are used for the detector due to their simplicity, stability, and bandwidth. Highly conductive (p) and (n) zones on either side of a low conducting intrinsic zone characterize the pin-diode. Figure 8 illustrates the structure of a planar InGaAs pin-diode.

The key circuit parameters of the pin-diode are: 1) responsivity, 2) capacitance, 3) dark current, and 4) noise equivalent power (NEP). The definitions and key considerations associated with these parameters are enumerated below.

*Responsivity

Responsivity is defined as (detected photocurrent/incident optical power). Responsivity will vary as a function of optical wavelength.

$$Re = \frac{I(pd)}{P(opt)} \ ; \ A/W \qquad (4)$$

Typical "Re" value : 0.8 (A/W) at 1300 nm

*Capacitance

Low capacitance is essential to minimize the noise of the Pin-Fet. Detector capacitance varies directly with the size of the detector active areas.

Typical "Cd" value : 0.5 pf @ 75 um (dia); 5V

*Dark Current

In additional to the desired signal current, a pin-diode also generates a temperature dependent "dark" (no light present) current. This current typically doubles in magnitude for every 10 C increase in temperature.

Typical "Id" value : 5na @ 5V, 25 C, 75 um (dia)

*Noise Equivalent Power

The dominant noise source in a pin-diode is shot noise which is generated by the statistical process of electrons passing through a pn-junction. In a photodetector, this noise is closely related to the statistical arrival of photons at the detector. The (mean square) noise current is directly related to the diode's average current:

$$\overline{i(n)^2} = 2e \cdot I \cdot Bw \qquad (5)$$

e = electron charge

Bw = noise bandwidth

I = average diode current

The current I consists of current related to the signal (Is), and the dark current (Id):

$$\overline{i(n)^2} = 2\ e \cdot Bw \cdot [Is+Id] \qquad (6)$$

The dark current thus defines the smallest optical power P (min) which can be detected. Using equations (4 & 6):

$$P(min) = \sqrt{(2e \cdot Bw \cdot Id)} \ / Re \qquad (7)$$

When the bandwidth is normalized, P(min) is referred to as the noise equivalent power (NEP; W/Rt-Hz) of the photodiode.

$$NEP = \sqrt{(2e\ Id)}\ /Re \qquad (8)$$

This is the fictitious optical power (per root-Hertz) which would generate a photocurrent equal to the dark current value in the photodetector.

Substituting the typical pin-diode (InGaAs) values establishes a typical NEP:

Typical NEP value : $5x10^{-14}$ (W/Rt-Hz)

B. Preamplifier Circuit

As shown in Fig. 7 the Pin-Fet transimpedance preamplifier circuit consists of a cascaded three stage transistor circuit with shunt (resistance) feedback applied between (Q3) and (Q1). Since the input signal is a current, the preamplifier acts as a current to voltage converter. The dynamic passband transfer characteristic for the preamplifier circuit (not including the pin-diode) is equal to the magnitude of the feedback resistance.

$$\Delta Vo/\Delta Iin = (- Rf) \qquad (9)$$

The term transimpedance refers to the conversion of the input current to an output voltage by means of the feedback resistance (Rf).

The three stage cascaded amplifier configuration is implemented as a cascode (common source Mes-Fet driving a common base bipolar transistor) followed by an emitter follower "buffer" output stage. The cascode, with its low Miller capacitance, is utilized because of the requirement to minimize the circuit input capacitance which is extremely important for noise considerations.

Similarly, proper selection of the active device for the first stage (Q1) is of prime importance because this component is a major noise contributor. GaAs (Mes-Fets) are utilized because they offer better noise performance than their silicon counterparts due to these devices having higher transconductances and lower capacitances.

The second and third stage transistors (Q2,Q3) are microwave bipolar transistors. The main requirement is for each device to have a high Ft (several GHz). Low output capacitance is also very important since a pole is present at the output of (Q2) due to the bias resistor (R4) and the output capacitances of Q2 and

Q3. This pole actually becomes one of the main limitations on the maximum available Pin-Fet bandwidth.

C.Output Interface Circuit

Although the emitter follower (Q3) provides a low output impedance, the pre-amplifier circuit does not perform optimally if directly loaded by 50 ohms. The recommended load impedance is 1000 ohms. A recommended interface circuit is illustrated in Fig. 9.

D.Package Considerations

The optimum performance of a Pin-Fet circuit is realized when the circuit parasitic capacitances are minimized. Thus, high performance Pin-Fet circuits are available as hybrid integrated circuits which utilize conventional dual-in-line (DIL) packages.

The circuit is implemented as a thick film integrated circuit (Fig. 10) where the active circuit components are mounted to a ceramic substrate. This substrate also contains the circuit conductive and resistance traces. Large value capacitors and variable resistances are directly attached to the substrate using chip components.

An optical fiber pigtail is utilized to interface the signal "light" to the photodetector. A 50/125 um multi-mode fiber pigtail (1 meter length) is inserted through a tube on the DIL package, aligned to the pin-diode, then secured.

The DIL package is hermetically sealed which provides long term integrity and reliability. Being made of Kovar, the DIL package provides the Pin-Fet shielding immunity against RFI and other electromagnetic interference.

3.Key Pin-Fet Analog/RF Performance Parameters

The key analog/RF performance parameters are: 1) transfer gain, 2) bandwidth, and 3) noise performance. Distortion and dynamic range are also important parameters. However, distortion performance is generally limited by the "optical source" and the Pin-Fet dynamic range is relatively wide (typically greater than 20 dB optical). Thus, these Pin-Fet parameters normally have a negligible impact on the analog/RF system performance.

A.Pin-Fet Transfer Gain

Transfer gain (Tg) defines the "dynamic" transfer characteristic of the Pin-Fet and has the units of (volts/watt).

$$Tg = \Delta Vo/\Delta Pin \text{ (volts/watt)} \quad (10)$$

Vo = Change in voltage out of Pin-Fet

Pin = Change in optical power into Pin-Fet

Using equations (4 and 9) :

$$Vo = (-k \cdot Re \cdot Rf)\ Pin \quad (11)$$

The factor (k) is the optical coupling coefficient. This coefficient accounts for the efficiency of coupling light from the fiber to the detector active area. Since the fiber core is 50 um and the detector active area diameter is 75 um, the resulting efficiency is quite high. Typical values are greater than 90%.

Therefore:

$$Tg = -k \cdot Re \cdot Rf \quad (12)$$

An application where one uses (Tg) is in determining the AC output voltage of the Pin-Fet for a given time varying (modulated) opticalinput signal. The equation for a modulated optical input signal is:

$$Pin(t) = (1 + m\ s(t))\ Pave \quad (13)$$

where ;

m = modulation index

s(t) = composite RF modulating signal (sinusoidal)

Pave = (average) DC optical power

Thus, the AC output voltage of the Pin-Fet may be written:

$$Vo = (m/\sqrt{2}) \cdot Pave \cdot Tg \text{ (rms volts)} \quad (14)$$

B.Pin-Fet Bandwidth

Bandwidth of the transimpedance amplifier is mainly determined by three parameters: 1) total input capacitance, 2) feedback resistance, and 3) the preamplifier circuit open loop gain. Shown in Fig. 11 is a simplified circuit model used to illustrate these bandwidth limiting components. The equation defining bandwidth for a transimpedance amplifier is:

$$Vo(f)/Iin(f) = -Rf/(1+w \cdot Rf\ [Cf+(Cin/A)]) \quad (15)$$

where ;

Cin = Ca+Cd (amplifer input + detector capacitance)

Cf = Parasitic capacitance of feedback resistor

Rf = feedback resistance

A = (amplifier open loop voltage gain)

w = $2\pi f$

Generally Cf is compensated in the feedback loop. Therefore:

$$w(3dB) = A/(Rf \cdot Cin) \quad (16)$$

Thus, the bandwidth of the transimpedance circuit is equal to (A) times the value of the (Rf·Cin) time constant. To achieve high bandwidths it is necessary to minimize the input capacitance and

provide high open loop gain (A) at the required operating frequencies. It is not desirable to arbitrarily reduce the feedback resistance in order to extend bandwidth since this will increase the "resistance (Rf)" noise power. Pin-Fets are available with operating bandwidths up to 400 MHz.

C. Pin-Fet Noise Performance

A simplified noise model for determining Pin-Fet noise performance is shown in Fig. 12. This model contains the primary noise sources which limit Pin-Fet noise performance. One should realize these noise sources consist of both signal dependent and signal independent components. Quantum noise (diode shot noise related) is dependent on the level of received optical power. Quantum noise limits the ultimate sensitivity of an optical receiver. Dark current, resistance (Rf), and circuit noise components are familiar noise components which are independent of the optical signal power.

The equation for determining the carrier/noise ratio at the output of the Pin-Fet is:

$$\frac{C}{N} = \frac{\overline{i(s)^2}\ \text{input signal power}}{\overline{i(nt)^2}\ \text{total input noise power}}$$

$$\frac{\overline{i(s)^2}}{\overline{i(nt)^2}} = \frac{(m \cdot Re \cdot Pave)^2/2 \quad (A^2)}{\overline{iq^2} + \overline{id^2} + \overline{iRf^2} + \overline{ic^2} \quad (A^2/Hz)} \tag{17}$$

where ;

$\overline{iq^2}$ = $2e \cdot Re\ Pave$; quantum noise power

$\overline{id^2}$ = $2e \cdot Id$; dark current noise power

$\overline{iRf^2} = \frac{4K \cdot T}{Rf}$; resistance noise power (20)

$\overline{i_C^2}$ = $[(2e \cdot Ig) + (\frac{4K \cdot T \cdot Tau \cdot w^2 \cdot Cin^2}{gm})]$;

circuit noise power

also ;

e = electron charge

Re = effective responsivity of photodiode (A/W)

Pave = (average) DC optical power (W)

Id = pin-diode dark current (A)

K = Boltzmanns constant

T = temperature (K)

Rf = feedback resistance (ohms)

Ig = gate leakage current (A)

Tau = Fet channel noise related (=1.1)

w = 2 πBw (1/s)

Cin = total input capacitance: (Cd+Ca+Cstray); (F)

gm = tranconductance of Fet (1/ohms)

The Pin-Fet's primary noise components are 1) resistance (Rf) noise and 2) circuit noise. Quantum noise is primarily a function of average optical power and thus mainly independent of the Pin-fet. Dark current noise is generally negligible when utilizing InGaAs pin-diodes.The relative magnitudes of the signal independent noise components are illustrated in Fig. 13 for a typical 120 MHz Pin-Fet (25 C). Signal dependent (quantum noise) contributions will be illustrated in the next section on system performance.

V.Fiber Optic System Illustration

Having identified the key Pin-Fet performance parameters, a fiber optic system application will be discussed to illustrate how this information may be applied.

1.System Requirement

An optical link is required to transmit (4) VSB/AM television channels over a distance of (10-20) kilometers. Channel frequencies are (channels 2, 3, 4, &5) which occupy a bandwidth of (54-82) MHz. What are the key performance considerations? Assume that a Pin-Fet optical pre-amp is available with performance parameters identical to those given in this paper.

2.System Performance Analysis

The key fiber optic analog/Rf performance parameters are noise, bandwidth, and distortion. For this system requirement, the Pin-Fet's noise performance is the main potential contributor to system performance.

System distortion performance is generally limited by the optical source. As discussed in the section on optical transmitters, the modulation index for linear transmission characteristics is generally set at (0.50). Third order intermodulation levels of -55 dB for (4 channels) are achievable from high performance analog laser transmitters and Pin-Fets. Second order distortion products will not be a problem since they will fall out of the passband.

Since the bandwidth requirement is (54-82) MHz, a 120 MHz (3 dB bandwidth) Pin-Fet is selected. This will provide a flat passband over the operating frequency range of (54-82) MHz.

Once the Pin-Fet bandwidth is determined, the values for (Rf) and (Tg) may be determined. Shown in Fig. 14 is a representative curve which shows the relationship between Pin-Fet bandwidth and transimpedance resistance (Log Rf). This curve indicates a transimpedance (Rf) for a 120 MHz Pin-Fet is around 30 kilo-ohms.

Having determined the transimpedance value one may now determine the AC output voltage of the Pin-Fet for a given optical input power.

$$Vo = Tg \cdot Pin \quad \text{(rms volts)} \tag{18}$$

$$Vo = (-k \cdot Re \cdot Rf \cdot m \cdot Pave/\sqrt{2}) \tag{19}$$

Using typical specifications values:

$$Re = 0.8$$

$$k = 0.9$$

$$m = 0.5$$

$$Rf = 30k$$

$$Vo = (7.7 \times 10^{3}) \; Pave \; \text{(rms volts)} \tag{20}$$

Pave is determined from the optical loss budget for this system. A loss budget for this example is shown below using standard values for the optical transmitter power levels, fiber loss, and other system parameters.

Optical Loss Budget	10Km	20Km
*Launched Opt. Power (Tx)	-2.0 dBm	-2.0 dBm
*Fiber Loss (.7 dB/Km)	7.0 dB	14.0 dB
*Splice Losses	1.0 dB	2.0 dB
*Connector Losses	2.0 dB	2.0 dB
*System Margin	1.0 dB	3.0 dB
*Total Optical Loss	11.0 dB	21.0 dB
*Received Optical Power		
-dBm	-13.0 dBm	-23 dBm
-watts	50.0 uw	5.0 uw

By substituting these values of optical power into equation (24), the Pin-Fet AC output voltage levels (rms) are determined.

$$Vo \; (max) = 386 \; (mv)$$

$$Vo \; (min) = 38.6 \; (mv)$$

Note that a 10 dB change in optical power produces a 20 dB change in electrical power. This is due to the relationship (Ipd= Re Popt) in which photodiode current is directly proportional to optical power. Thus, a (20) dB minimum electrical AGC range is required to accommodate the (10) dB optical input power range.

The carrier/noise performance at the output of the optical link

is determined by using the Pin-Fet noise model. These individual signal and noise components have been calculated at various received optical power levels from (5-500) uw. These results are summarized in Table A. Dark current, (being a negligible contribution) , has been ignored in this analysis.

The resultant carrier/noise ratios for this link are shown in Fig. 15. Four curves have been plotted. They illustrate: 1) the carrier/noise ratio due to quantum noise alone C/N(q); 2)The carrier/noise ratio due to all the receiver noise components; 3) the carrier/noise ratio due to the optical source alone (RIN noise is equal to -133 dB/Hz); and 4) the cumulative carrier/noise ratio of the entire system C/N (*). These carrier/noise ratios are referred to a bandwidth of 4 MHz which is the standard signal bandwidth of an NTSC video signal.

Using this graph, the carrier/noise ratios (per channel) for the received optical powers of 5 uw and 50 uw are 41 dB and 53 dB respectively. It should be noted that the modulation index (typically 0.5) was divided by (4) in this analysis to account for the (4) channel loading.

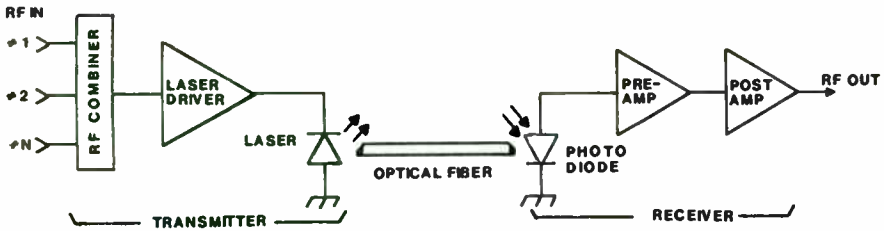
VI.Summary

Pin-Fet transimpedance preamplifiers are high performance optical receiver front-ends. They provide wide bandwidths, high dynamic range, and high sensitivities (carrier/noise ratios). Pin-Fets are available as hybrid integrated circuits, hermetically sealed, and mounted in standard 14 pin (DIL) packages.

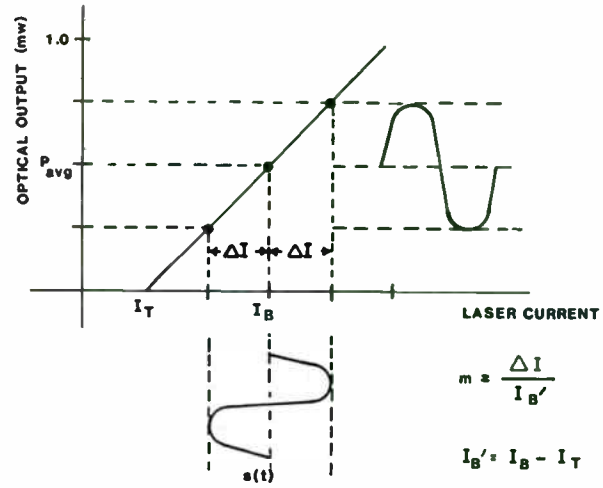
Pin-Fets are easily integrated into RF subsystems once the users understand the key performance considerations. Hopefully this primer has helped clarify some of these design considerations for prospective Pin-Fet users.

ACKNOWLEDGEMENTS:

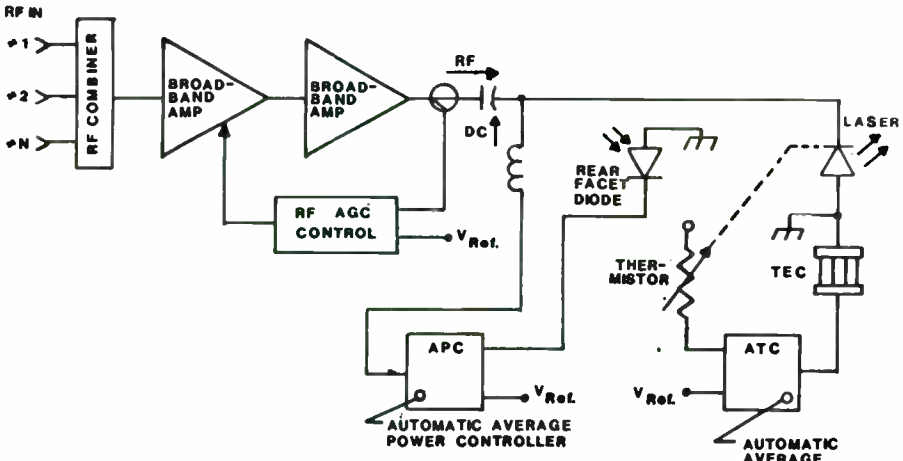
Special thanks to N. Penland, P. Fung, and S. Stankewicz for their contributions in preparing the text, illustrations, and presentation materials for this paper.



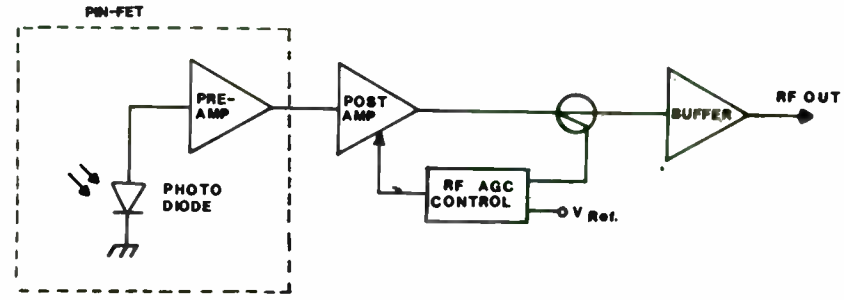
FDM ANALOG/RF TRANSMISSION
FIBER OPTIC LINK
FIG. 1



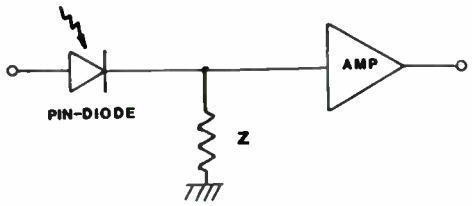
LASER LI CURVE
FIG. 3



ANALOG/RF OPTICAL TRANSMITTER
FIG. 2

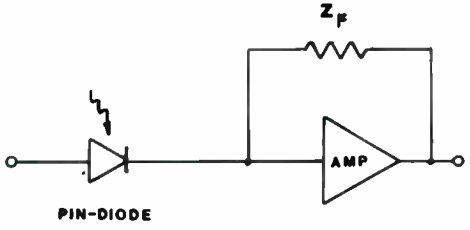


RF ANALOG
OPTICAL RECEIVER
FIG. 4



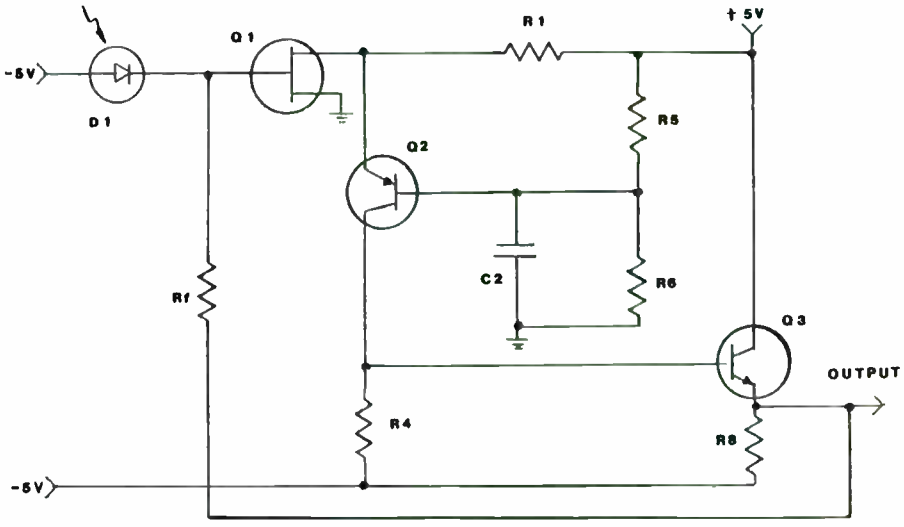
HIGH-IMPEDANCE (HZ) PINFET

FIG. 5



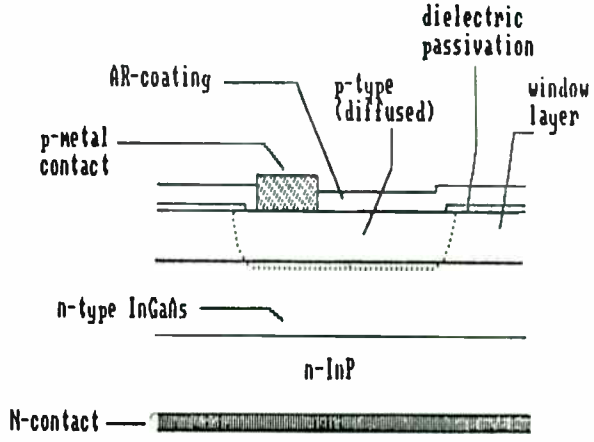
TRANSIMPEDANCE (TZ) PIN-FET

FIG. 6

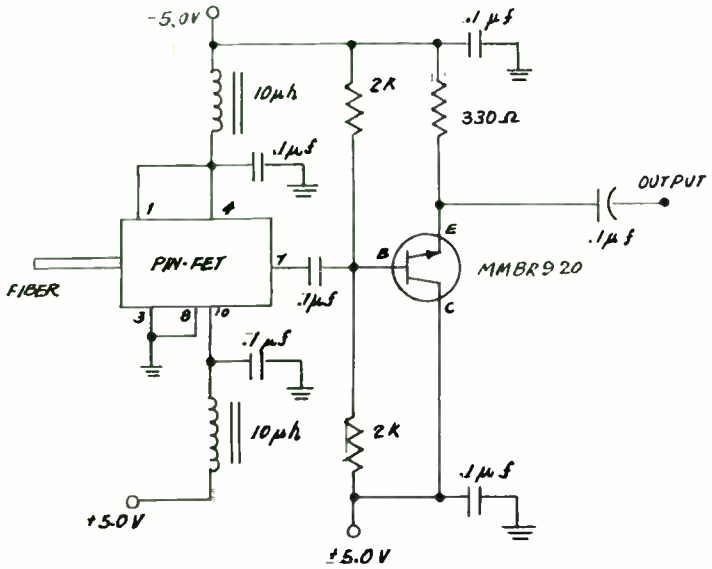


PIN-FET SCHEMATIC

FIG. 7

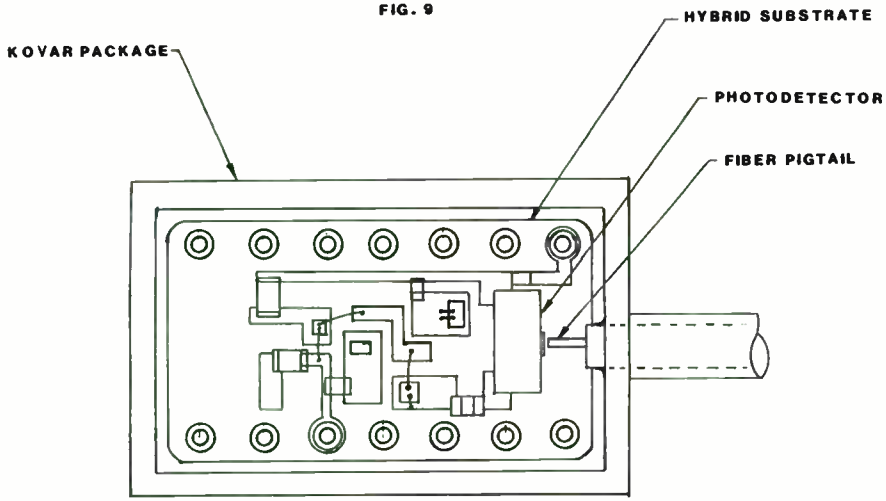


PIN DIODE
PLANAR STRUCTURE
FIG.8



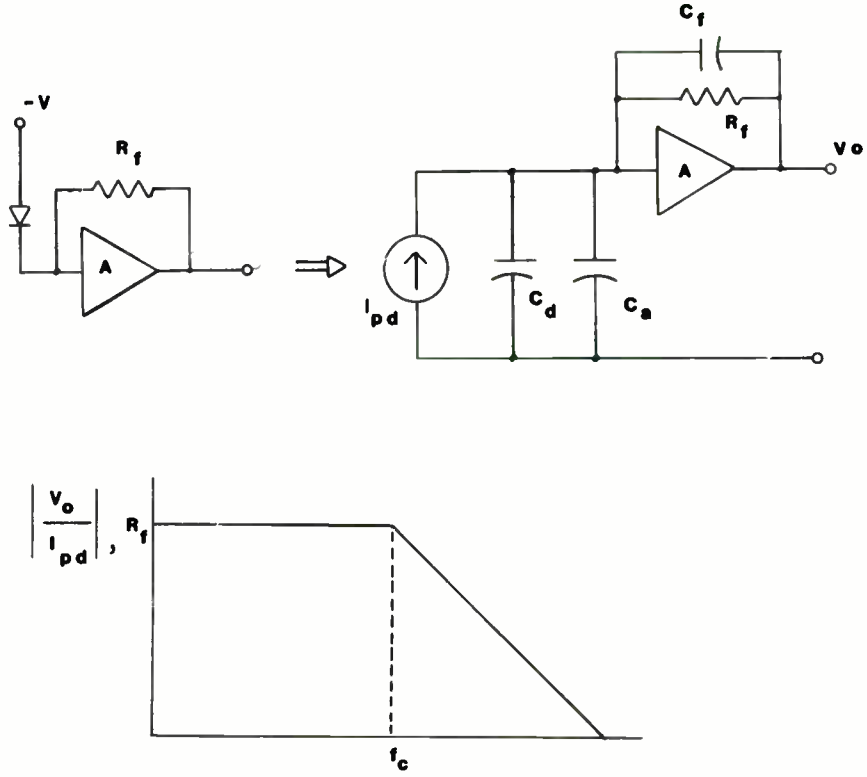
PIN-FET OUTPUT INTERFACE CIRCUIT

FIG. 9



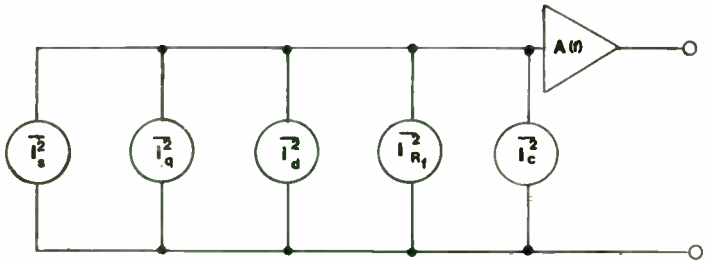
PINFET PACKAGE

FIG. 10



TRANSIMPEDANCE CIRCUIT
BANDWIDTH MODEL

FIG. 11



$\overline{i_s^2}$ = Signal power (A^2)

$\overline{i_q^2}$ = Mean sqaure quantum noise (A^2/Hz)

$\overline{i_d^2}$ = Mean square dark current noise (A^2/Hz)

$\overline{i_{R_f}^2}$ = Mean square resistance noise (A^2/Hz)

$\overline{i_c^2}$ = Mean square circuit noise (A^2/Hz)

$\overline{i_{N(total)}^2} = \overline{i_q^2} + \overline{i_d^2} + \overline{i_{R_f}^2} + \overline{i_c^2}$ Total mean square noise power (A^2/Hz)

PIN-FET NOISE MODEL, FIG. 12

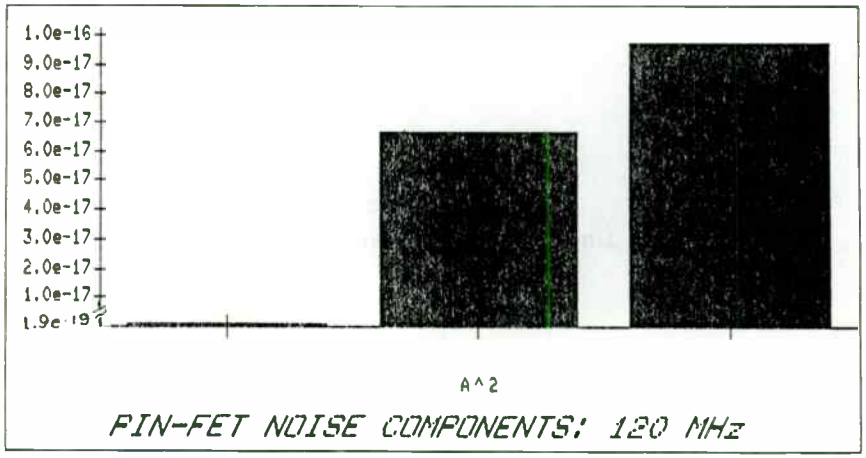


FIG. 13

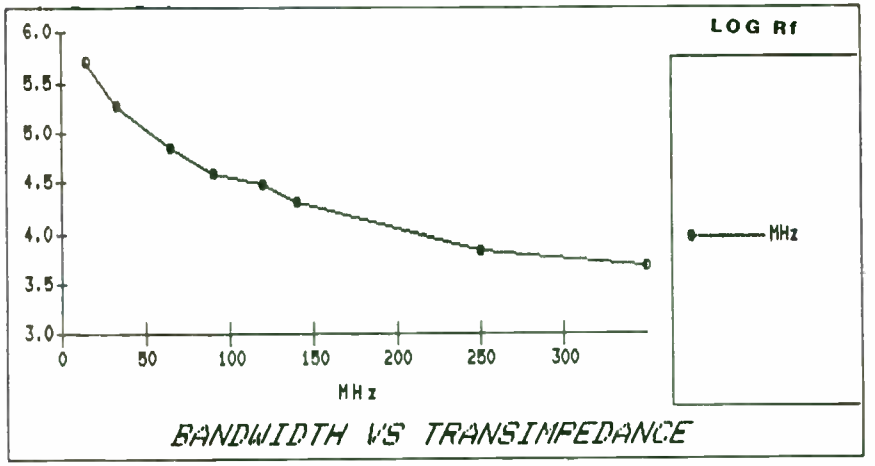


FIG. 14

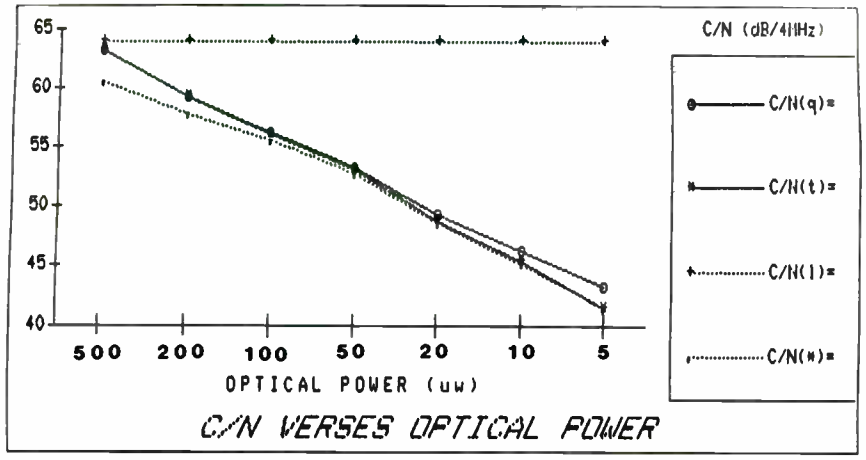


FIG. 15

C/N

m=	0.125	i(s)^2=((m*M*Re*Pave)^2)/2
M=	1	
Re=	0.7	i(q)^2=2q(Re*Pave)(M^2)*F(M)*BW
Pave=	var	
q=	1.6e-19	i(Rf)^2=(4*K*T*BW/Rf)
I(d)=	5.0e-09	
F(M)=	1.0	i(c)^2=(4KT*Tau*w^2*Cin^2*BW/gm)+(2e*Ig*BW)
K=	1.4e-23	
T=	300	
Ft=	4.0	
Rf=	3.0e+04	
BW=	4.0e+06	
x=	0.95	
Tau=	1.1	
Cin=	1.5e-12	
gm=	3.0e-02	
Ig=	1.e-07	
RIN=	133	

500 (5,2,1)

	5.0e-04	2.0e-04	1.0e-04	5.0e-05	2.0e-05	1.0e-05	5.0e-06
i(s)^2=	9.6e-10	1.5e-10	3.8e-11	9.6e-12	1.5e-12	3.8e-13	9.6e-14
i(q)^2=	4.5e-16	1.8e-16	9.0e-17	4.5e-17	1.8e-17	9.0e-18	4.5e-18
i(Rf)^2=	2.2e-18	2.2e-18	2.2e-18	2.2e-18	2.2e-18	2.2e-18	2.2e-18
i(c)^2=	1.3e-19	1.3e-19	1.3e-19	1.3e-19	1.3e-19	1.3e-19	1.3e-19
i(t)^2=	4.5e-16	1.8e-16	9.2e-17	4.7e-17	2.0e-17	1.1e-17	6.8e-18
C/N(q)=	63	59	56	53	49	46	43
C/N(t)=	63	59	56	53	49	45	41
C/N(I)=	64	64	64	64	64	64	64
C/N(*)=	61	58	56	53	49	45	41

TABLE A

THE CONVOLUTED LOOP

by
Robert T. Hart
Senior Principal Systems Engineer
Haris Corporation
PO BOX 334
Melbourne, Florida 32902

INTRODUCTION

A small loop antenna can provide excellent performance for both transmitting and receiving.\1/ The Convoluted Loop (Figure 1) is a single conductor configured to produce two orthogonal loops, resulting in an antenna with high performance, very small size, and an almost ideal radiation pattern. This antenna is designed for mounting at ground level over a small counterpoise and has a height of less than 0.04 wavelengths. The features, and theory, of this unique antenna are detailed in the following paragraphs.

THEORY

The equations specifically developed to define the Convoluted Loop are listed in Table 1, and are the basis of the computer program attached as Appendix 1. A detailed explanation of the equations will lead the reader to an understanding of the antenna. This will, in turn, allow good engineering judgment to be used when selecting design parameters for a particular application.

The efficiency of any antenna is defined as the ratio of the radiated power to input power, conveniently expressed as the Radiation Resistance divided by the sum of the Radiation Resistance plus Loss Resistance. Since small antennas are characterized by low radiation resistance, efficiency is a major concern. On the other hand, large antennas have a high radiation resistance, compared to the loss in the antenna conductor.

The Radiation Resistance for a small loop antenna is dependent on the area enclosed by the conductor, and the operating frequency. \2/\3/\4/ The antenna will become self resonant if the conductor length is greater than 1/3 wavelength due to distributed capacity. This sets the maximum length of the conductor. Further, the equations reflect the fact that a single conductor forms two loops for this antenna. When a reflecting screen is placed under the loop, the effective area of the loop doubles due to the image concept. In the equation for area, the multiplier of 4 covers both the dual loops and their images due to reflection. For a square loop design, each side of one loop is the total conductor length divided by 8. The maximum area is achieved for a given conductor length when the conductor is formed to make a circle. The area is reduced to 87% for an octagon shape, and to 78% for a square, compared to a circle. For mechanical simplicity, a square loop with a reflecting screen is used for the example in this article.

The small area of the loop results in low values of Radiation Resistance. The primary component of Loss Resistance is due to the loss in the conductor, with a small component due to ground loss, which will be discussed later. Therefore, a low loss conductor is required. To minimize loss, copper pipe should be used. The equation for Loss Resistance includes skin effect loss

for a copper pipe, which varies as a function of frequency.

Although efficiency is a major design parameter for any antenna, for small antennas with high efficiency the Q of the antenna must be considered. Large diameter conductors allow the Q to be sufficiently high to affect the instantaneous bandwidth such that it may be too narrow for the type of modulation desired. High Q is also an indication of very high voltages across the tuning capacitor. Small conductors, having high loss resistance, will result in low Q and low efficiency. It is therefore necessary to perform trade offs in the design of the Convoluted Loop for a particular application. For most applications in the HF frequency range, 3/4 inch copper pipe (0.9 inch outside diameter) is a reasonable compromise. Figure 2 presents efficiency versus frequency for various size conductors, indicating only small improvements for larger, more expensive, copper pipe.

The equations for Inductance and distributed capacity are based on data derived from various size convoluted loops at various frequencies, primarily between 1.8 and 10 MHz. The tests were performed on the latest version of MININEC3 \5/, an antenna analysis program. As of this time, they have been derived only for 3/4 inch copper pipe and square loops. Once the inductance and distributed capacity have been calculated, the Convoluted Loop is considered a simple resonant circuit, allowing the inductive reactance and the tuning capacitor value to be calculated. The equation for inductance is multiplied by a value of 1.13 to cover the effect of the matching network. The matching network is an autoformer type of match, having both series and mutual inductance.

The Q of the antenna can be calculated once the inductance and resistance are known. The standard equation for Q is divided by 2 to include the effect of the loading of the transmitter/receiver, since it is the system Q that is important, not the Q of the antenna as a stand alone component. The calculated bandwidth of the antenna is the +/- 3 dB bandwidth, assuming a perfect match (VSWR=1.0/1) at resonance. At the 3 dB frequencies the calculated VSWR is 5.1/1,and the resistance and reactance values are equal, resulting in a 45 degree phase shift of the equivalent resonant circuit.

The voltage across the tuning capacitor is a function of the transmitter power and the impedance of the antenna. Although the voltage can be very high, it is not excessive for available tuning capacitors. Although vacuum variables are preferred, the spacing for an air variable is calculated based on 30,000 volts per inch spacing.

APPLICATION

The following comments are based on experimental results of the Convoluted Loop, or from data calculated by MININEC3. This antenna is readily constructed by cutting sections of copper pipe and joining them with copper elbows. Pieces of clear plastic in the center at the top and bottom provide good insulation and mechanical rigidity. The length of the longer pieces of pipe are

equal to the total conductor length divided by 8. At the top the shorter pieces are cut to length to allow 1/2 inch between the 90 degree bends, which are made of two 45 degree elbows joined with a short section of pipe. At the bottom a short section of pipe is removed for installation of the tuning capacitor.

Each joint in the antenna must be soldered or brazed to minimize joint resistance. The tuning capacitor must be low loss. A vacuum variable has this feature. If the loop ends are connected to the two stators of an air variable split stator or butterfly capacitor , placing them in series with RF currents flowing through the rotor, there is no loss due to wiper contacts. However, all pieces of the capacitor must be welded or brazed. In other words, the typical variable capacitor has plates installed and held in place only by mechanical pressure, resulting in high resistance at each mechanical joint. The high loss will make the loop inoperable. One way to correct this deficiency is to remove all aluminum plates and replace them with copper or double sided print circuit board material. Straps can then be used to bond the plates together and provide connection to the loop conductor.

Although the antenna may be tuned over a wide frequency range, due to it's narrow instantaneous bandwidth, it must be tuned to the operating frequency. This can be easily accomplished using a stepper motor to rotate the shaft of the variable capacitor. The prefered unit has a 7.5 degree step angle and includes a small 300:1 gear head reducer and is manufactured by Hurst Manufacturing loacted in Princeton, Indiana. The part number is 3004-001. They also produce an integrated circuit (part # 22001) that makes the ideal driver simply by adding a few components and a power supply. For a vacuum variable the gear ratio should be reduced, since numerous revolutions of the shaft are required to vary the value of capacity from minimum to maximum.

The Convoluted Loop must be fed as a balanced antenna. For 50 ohm coax feed, a 1:1 balun transformer is required. The feeders are made of 1/4 inch copper tubing wrapped with black plastic electrical tape for insulation. Each feeder is the length of one of the long sides of the loop, and is formed to the inside of the loop antenna orthogonal to the loop that includes the tuning capacitor. The spacing between the feeder and loop conductor and the length of the feeders will determine the feed point impedance. Once the ends of the feeders are soldered to the loop, bending the feeders to vary the spacing will easily allow a very low VSWR to be achieved. This is only one of many ways to feed this antenna, but the author finds it the most convenient \6/.

The loop develops a very high magnetic field. If placed close to ferrous metal, such as reinforcing material in concrete or rain gutters or antenna towers as examples, RF energy will be coupled into the ferrous material and be transformed into heat. This reflects a low impedance into the loop, increasing it's loss resistance and it's efficiency.

Due to it's magnetic properties, the Convoluted Loop is not sensitive to electrostatic fields, the major cause of reception of man made noise. Therefore, a significant improvement in signal to noise reception will be noticed in noisy areas. In theory, the value is 26 dB. Also, due to the high Q, the antenna serves as a preselect filter prior to the receiver. This improves reception in the presence of impulse noise especially from lightning during thunderstorm activity.

It was previously suggested that the antenna should be used in conjunction with a counterpoise. Figures 3, 4, and 5 present radiation patterns derived from MININEC3 for a 10 foot tall Convoluted Loop operating at 4 MHz, having a counterpoise made of 120 radials each having a length equal to twice the height of the antenna (20 feet). Since it is only the reflected energy, not conducted, the radials do not need to be connected to the loop. However, the antenna should be grounded for safety reasons. All patterns presented in this article are over a ground with a dielectric constant of 10 and a conductivity of 0.002 mhos, which represents sandy soil in Florida. Further, the patterns are based on antennas with perfect conductors. Actual patterns can be determined by reducing the pattern gain by the efficiency calculated for a particular size antenna. Figures 6, 7, and 8 are presented for comparison based on a dipole over the same ground. Figure 9 is a Convoluted loop at 7 MHz over 120 1/4 wavelength radials, and figure 10 is a 1/4 wavelength vertical with 120 1/4 wavelength radials, presented for comparison.

It is important to note (see figure 9) that hemispherical coverage is achieved, allowing the antenna to be used for both local and long range communications. Due to it's magnetic properties, a loop antenna produces radiated energy at low elevation angles. This also confirms that there is very little ground loss for this antenna.

On the air tests in comparison with other antennas has confirmed the performance of the Convoluted loop and verified the performance indicated by the patterns derived from MININEC3. In addition, measurements on the antenna confirm the accuracy of the equations. An example run of the computer program is presented in figure 11, to assist those who desire to develop a Convoluted Loop antenna for a particular application. Figure 12 presents a set of data run from the program.

The Convoluted Loop is a result of efforts to design a high performance antenna that requires very little space. The author has hope that others will modify this design, perhaps by putting a "twist" to the conductor, and achieve an antenna design that is truly non-directional.

Robert T. Hart W5QJR
Senior Principal Engineer
Harris Corporation
PO Box 334
Melbourne, Florida 32902

\1/ Hart, R.T., "Small High Efficiency Loop Antenna," June 86 QST

\2/ Barrick,D.E., "Miniloop Antenna Operation and Equivalent Circuit," IEEE Transactions on Antennas and Propogation, vol AP-

34, No 1, January 1986

\3/ King, R.W.P.,"The Loop Antenna for Transmission and Reception," Antenna Theory, Part 1, McGraw-Hill, 1969, ch-11

\4/ Dunlavy,J.H.,"Wide range tunable transmitting antenna," U.S. Patent 3 588 905, June 28, 1971

\5/ Campbell, D.V.,"Personal computer applications of MININEC," IEEE Antennas Propogation. Soc Newsletter, vol 26, pp 5-9,1984

\6/ Hart, R.T., "Small High Efficiency Antennas alias The Loop," Eagle Press 1986

\7/ Schelkunoff & Fries, "Antenna Theory and Practice," Wiley 1952

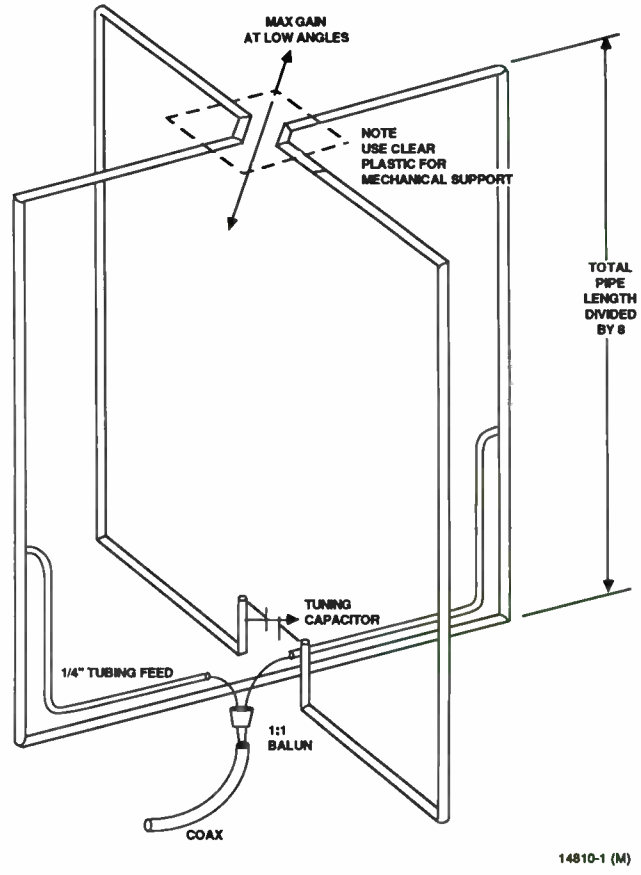


Figure 1. The Convoluted Loop Antenna

$\eta = R_R/(R_R + R_L)$	EFFICIENCY OF AN ANTENNA
$R_R = 3.38*10^{-6}\ (F^2A)^2$	RADIATION RESISTANCE - OHMS
$A = 4(S/8)^2$	TOTAL AREA - SQUARE FEET
$R_L = 9.96*10^{-4}\ \sqrt{F}\ S/D$	LOSS RESISTANCE - OHMS
$Q = X_L/(2(R_R + R_L))$	Q
BW = F/Q	BANDWIDTH - MHz
$L = 1.13(5.6624*10^{-7}\ S^{6984})$	INDUCTANCE - hys
$X = 2\pi FL$	INDUCTIVE REACTANCE - OHMS
$CS = 1.1374*10^{-11} + 4.684*10^{-13}S$	DISTRIBUTED CAPACITY - pfd
$CT = 1/(2\pi F)^2L)$-CS	TUNING CAPACITOR - pfd
$VC = \sqrt{PQX}$	VOLTAGE ACROSS TUNING CAPACITOR
SC = VC/75000	SPACING FOR CT - INCHES
	F - OPERATING FREQUENCY - MHz
	D - CONDUCTOR DIAMETER - INCHES
	S - CONDUCTOR LENGTH - FEET

14810-4 (M)

Table 1

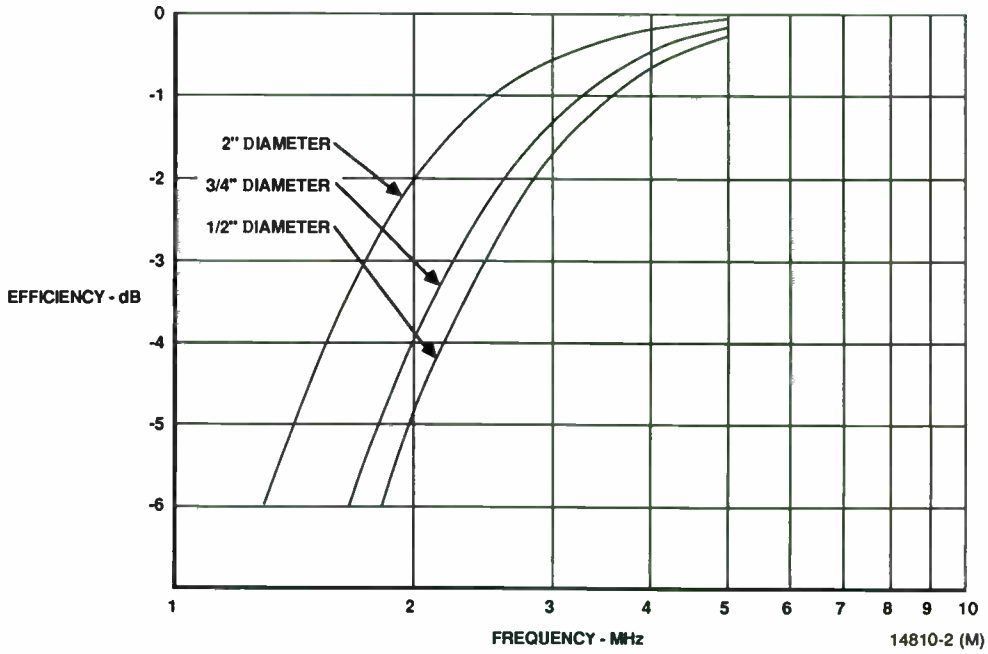


Figure 2. Effeciency vs. Conductor Diameter

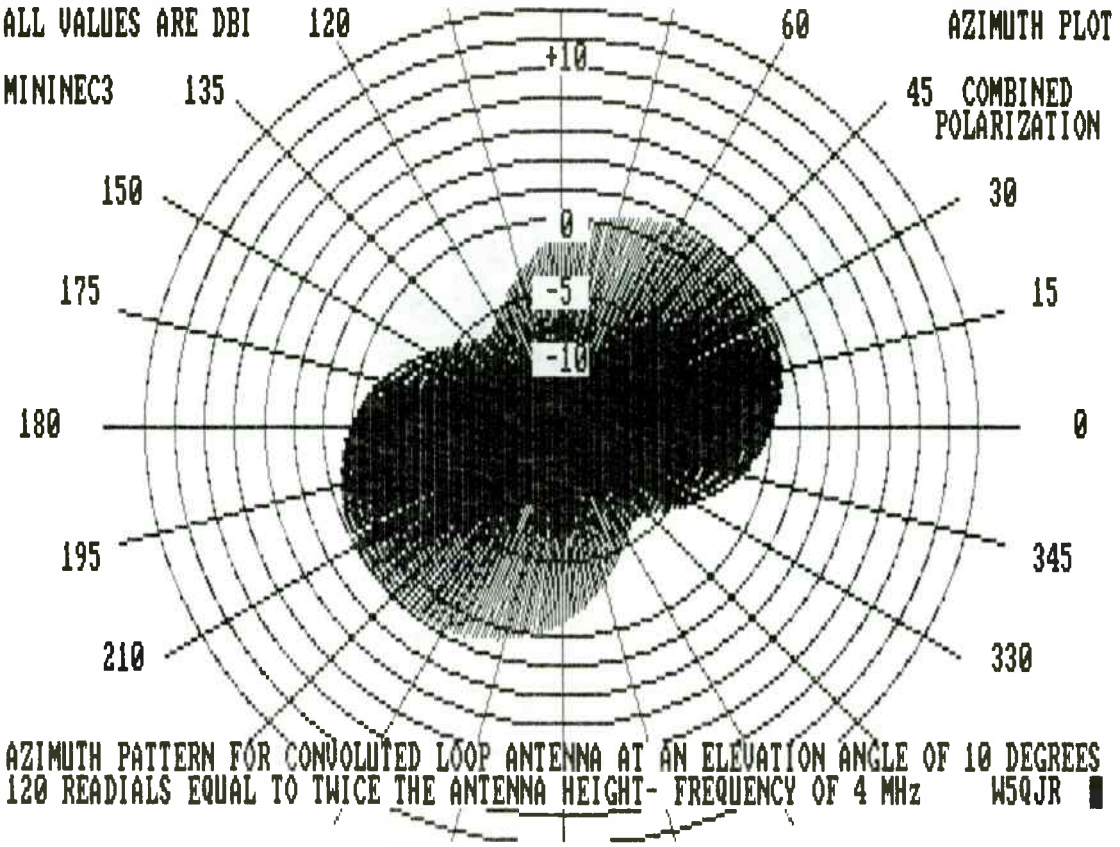


Figure 3

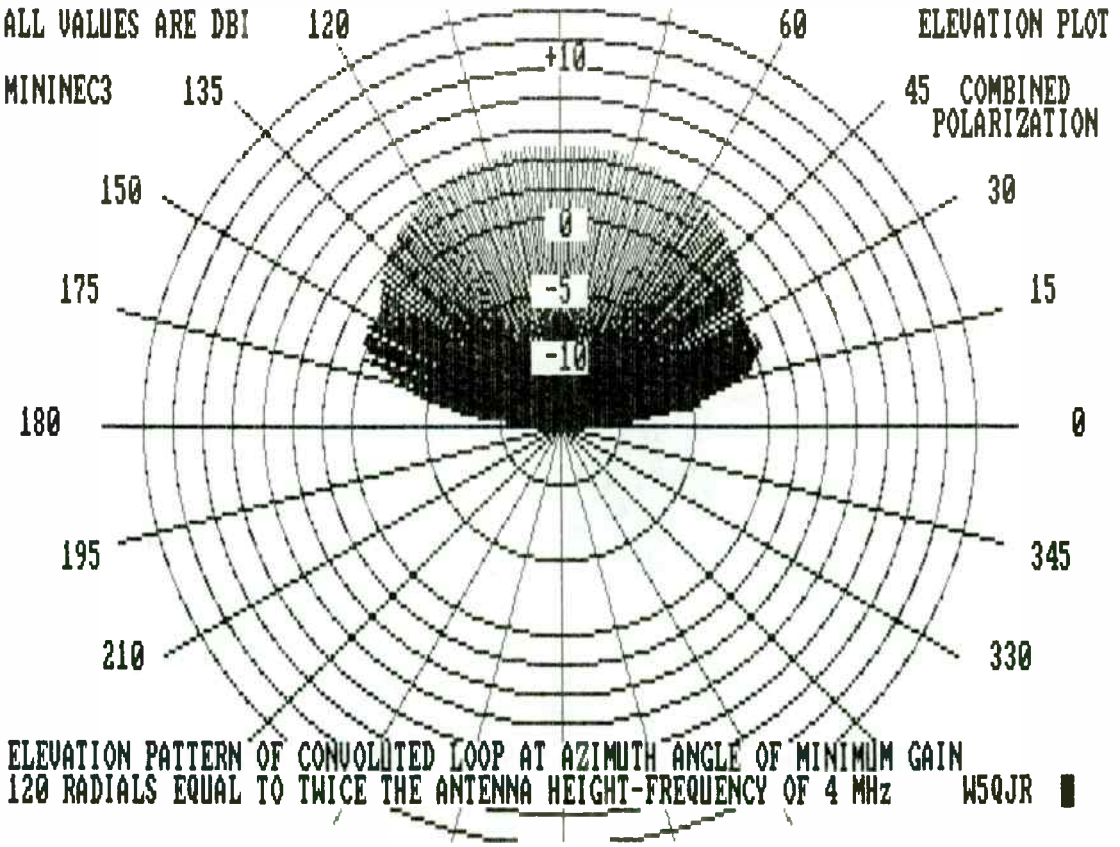


Figure 4

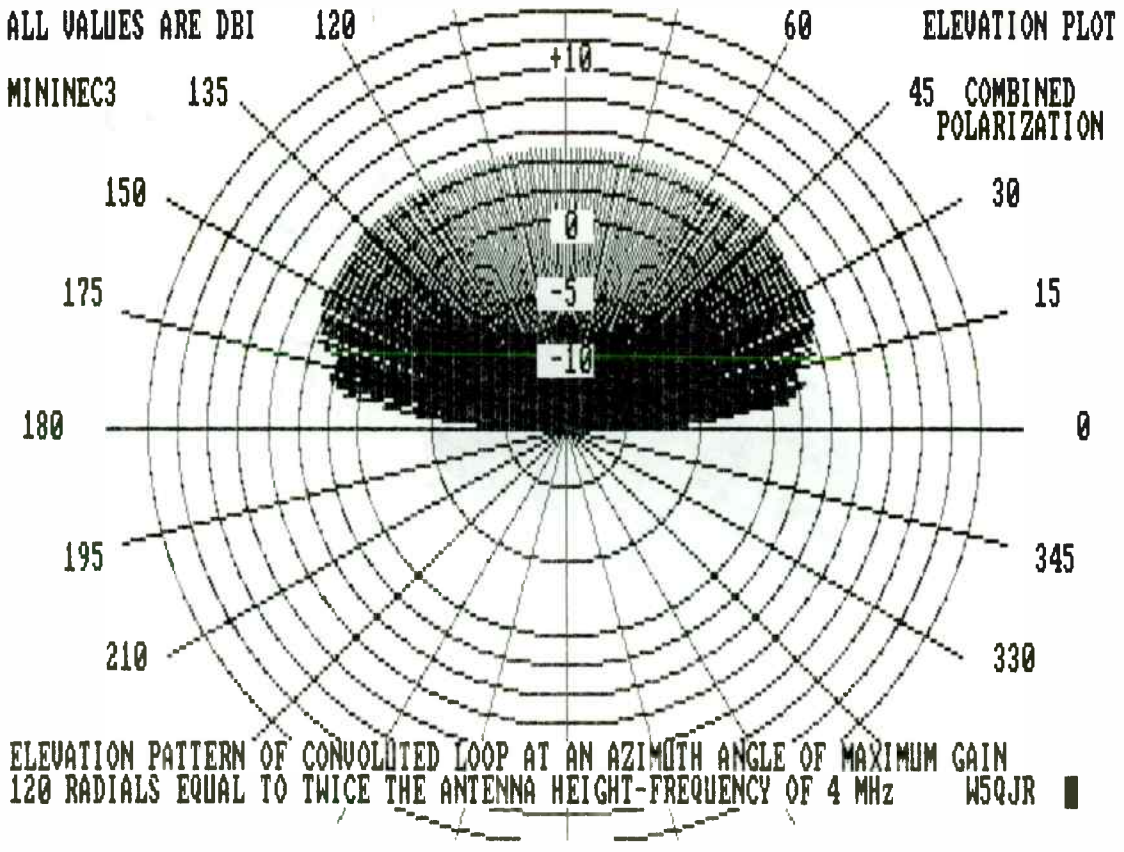


Figure 5

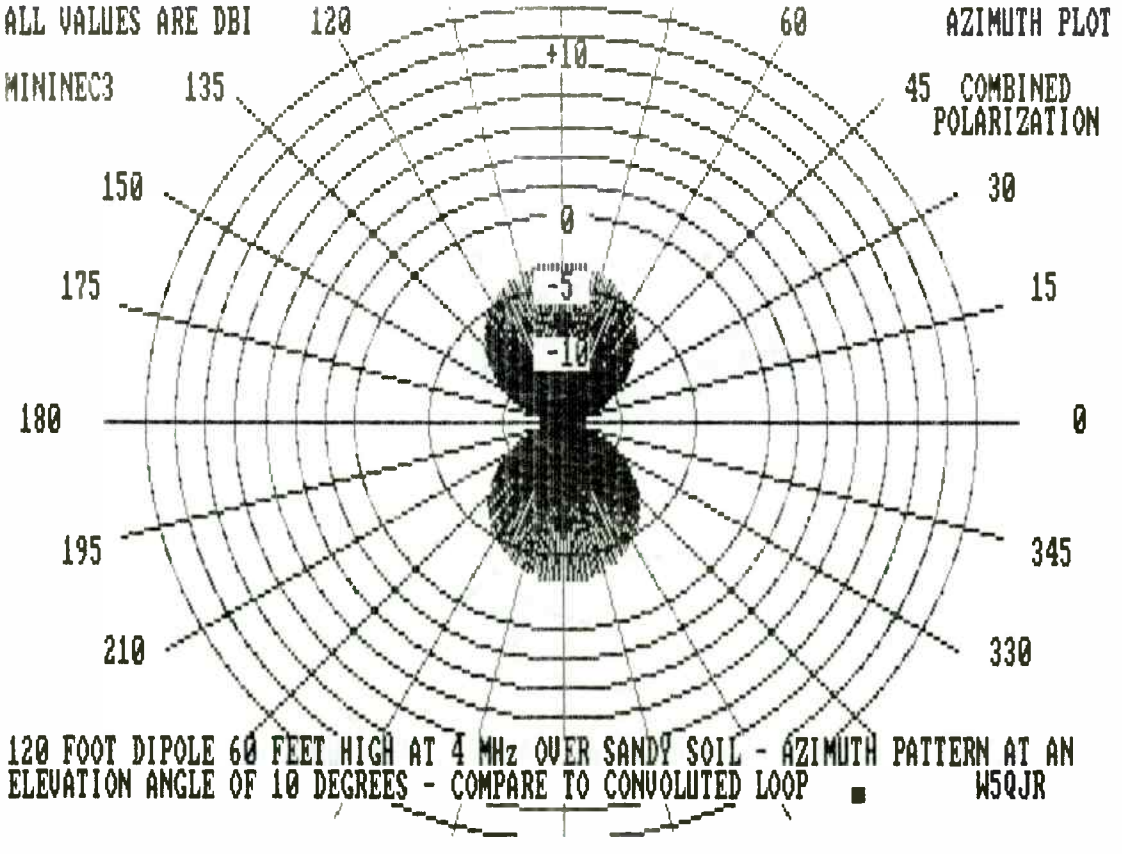


Figure 6

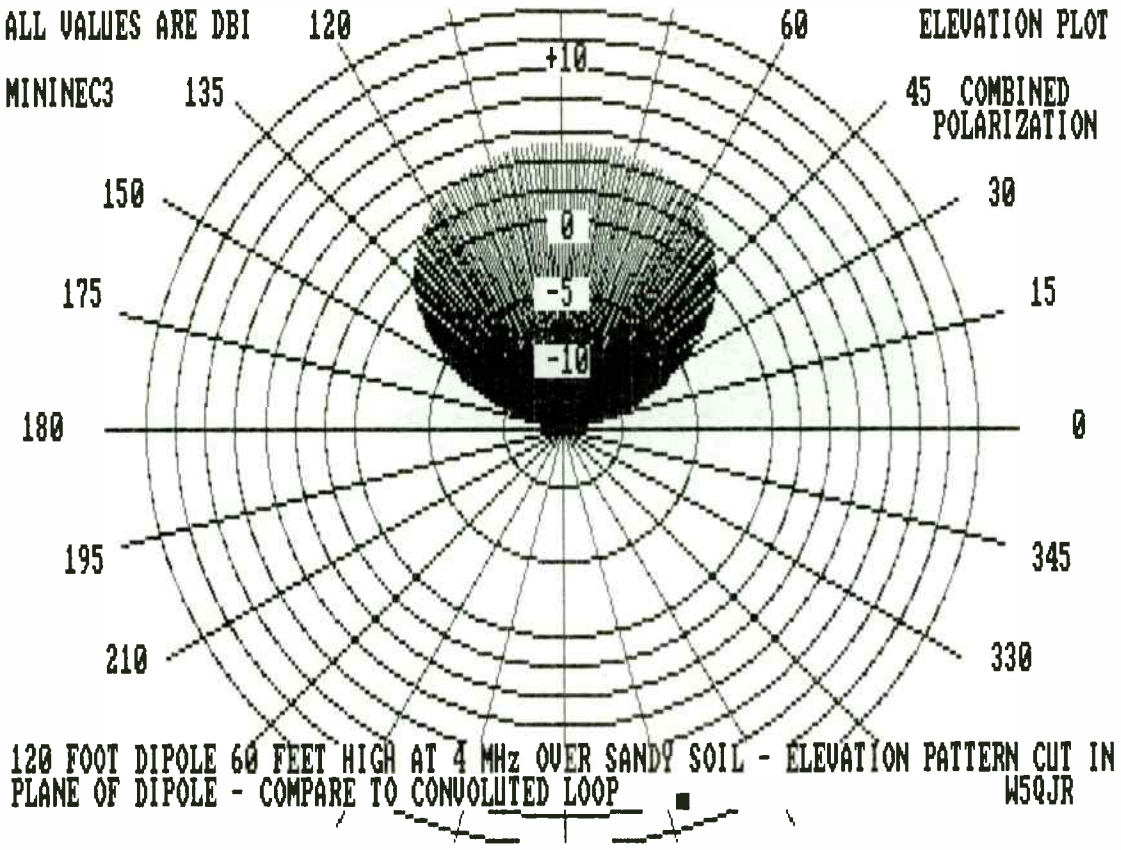


Figure 7

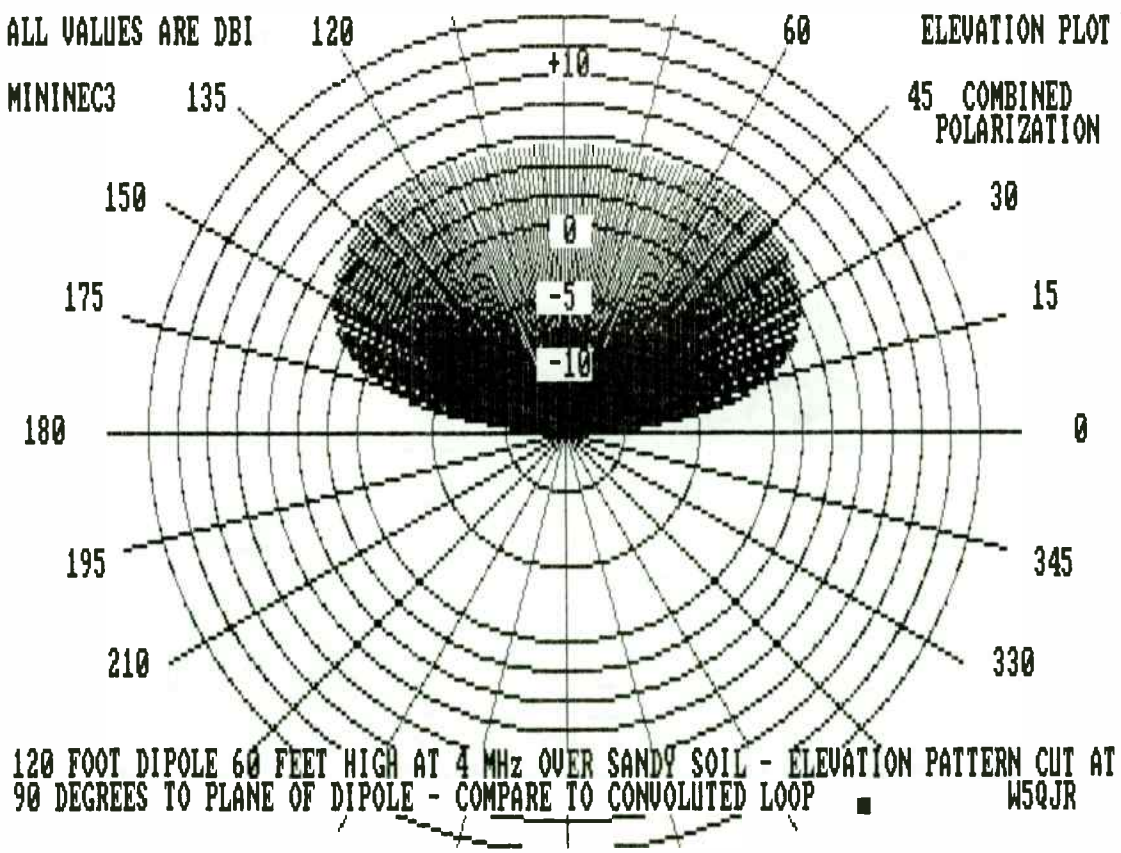


Figure 8

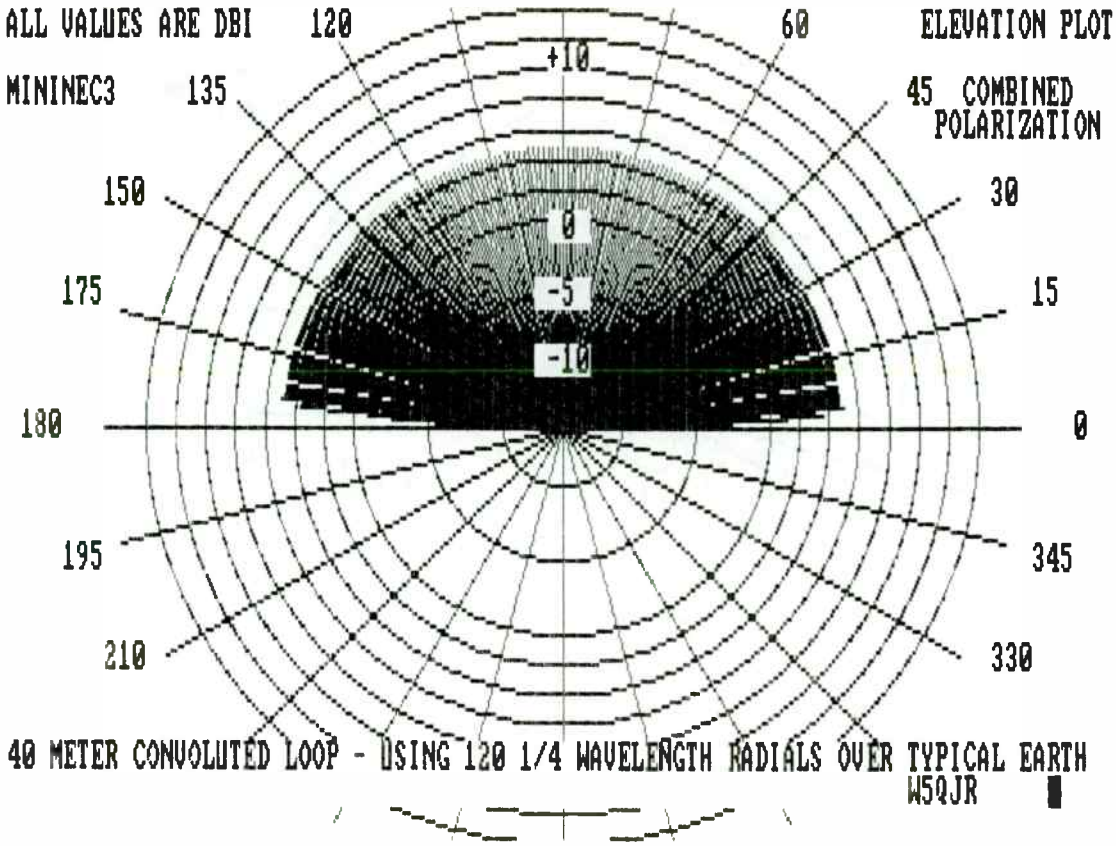


Figure 9

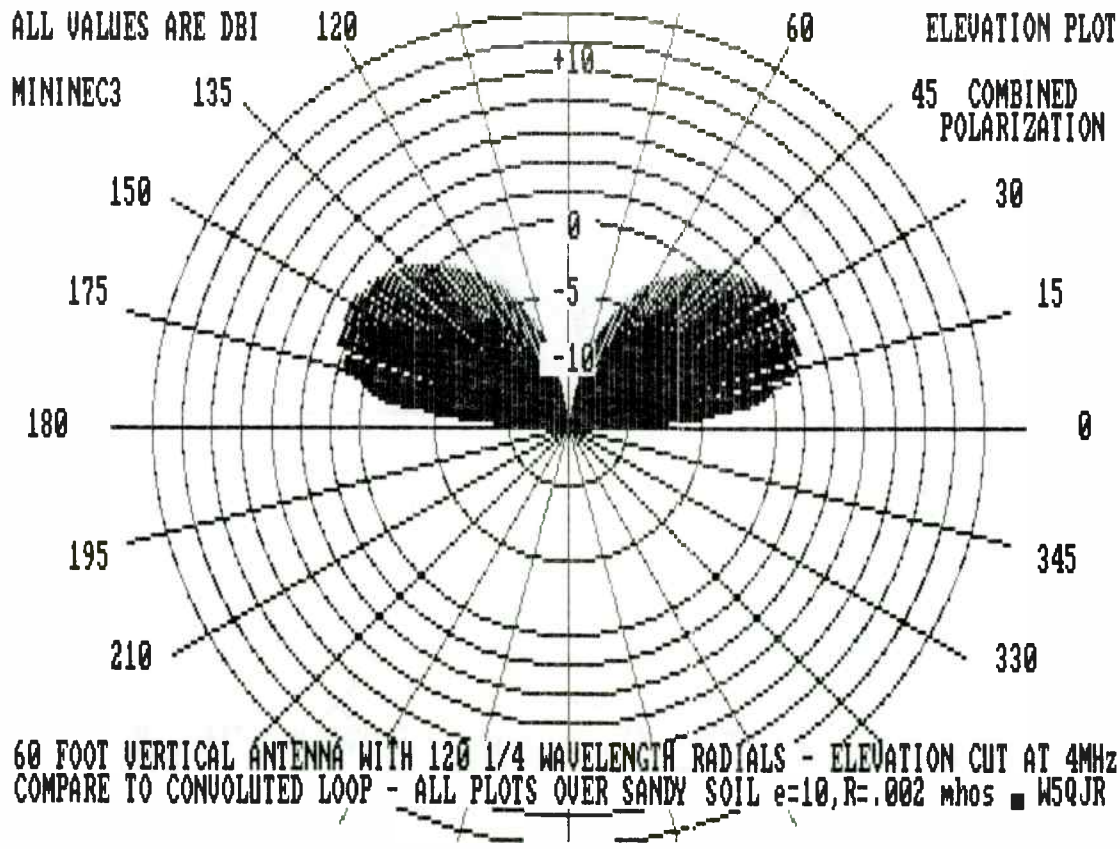


Figure 10

```
        PARAMETERS OF A CONVOLUTED LOOP ANTENNA

INPUT VALUES AS REQUESTED FROM THE FOLLOWING PROMPTS:

SPECIFY OPERATING FREQUENCY IN MHz -- ? 4
SPECIFY TOTAL LENGTH OF PIPE IN FEET -- ? 80
SPECIFY TRANSMITTER PEAK OUTPUT POWER IN WATTS --? 1000

RESULTS

FOR A  80  FOOT PIPE,  .9  INCH DIAMETER AT  4  MHz ---
RADIATION RESISTANCE OF THE ANTENNA  =  1.3844  OHMS
THE LOSS RESISTANCE OF THE ANTENNA =  .177  OHMS
EFFICIENCY =  88  PERCENT, WHICH IS (RELATIVE TO 100%) = -.5  dB
THE Q OF THE ANTENNA =  109
THE BANDWIDTH OF THE ANTENNA =  36.697  KHz
THE VALUE OF THE TUNING CAPACITOR  =  67  pfd
FOR  1000  WATTS, TUNING CAP VOLTAGE =  6115  VOLTS
THE CAPACITOR PLATE SPACING MUST BE >  .203  INCHS
THE HEIGHT (AND WIDTH) OF THIS LOOP IS  10  FEET
 - - - - - - - - - - - - - - - - - - - - - - - - -

WANT TO TRY A DIFFERENT FREQUENCY ? (Y/N)
```

Figure 11. Example Of Computer Run

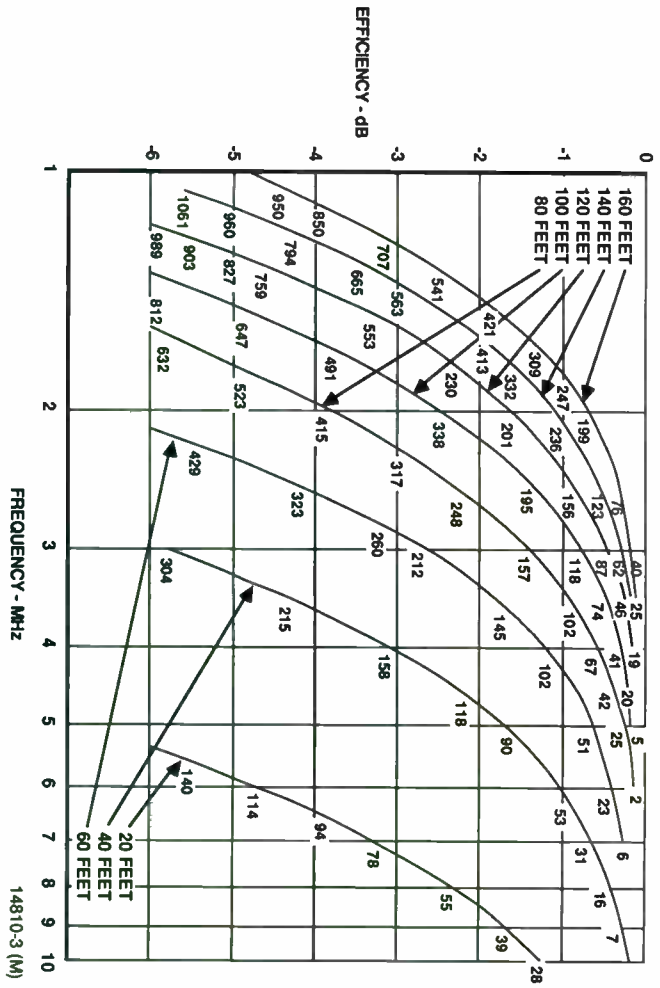


Figure 12. Antenna Characteristics Versus Frequency For Various Length Conductors

```
10 CLS:PRINT"          THIS PROGRAM DEVELOPED BY W5QJR FOR USE ON
 AN IBM-PC"
20 LOCATE 10,10
30 PRINT"PARAMETERS OF A CONVOLUTED LOOP ANTENNA": PRINT
40 PRINT"INPUT VALUES AS REQUESTED FROM THE FOLLOWING PROMPTS:": PRINT
50 INPUT "SPECIFY OPERATING FREQUENCY IN MHz -- ":F
60 INPUT "SPECIFY TOTAL LENGTH OF PIPE IN FEET -- ";S
70 INPUT "SPECIFY TRANSMITTER PEAK OUTPUT POWER IN WATTS --":P
80 D=.9    'THIS IS THE OD FOR 3/4 INCH COPPER PIPE
90 A=4*(S/8)^2 'THIS IS THE AREA OF 2 SQUARES MULTIPLIED BY THEIR IMAGES"
100 RR=3.38E-08*(F^2*A)^2 'RADIATION RESISTANCE
110 RR=FIX(RR*10000)/10000 'METHOD USED TO REDUCE TO 2 DECIMAL PLACES"
120 RL=.000996*F^.5*S/D ' LOSS RESISTANCE
130 RL=FIX(RL*10000)/10000
140 E=RR/(RR+RL) ' EFFICIENCY, DECIMAL
150 E=FIX(E*100)
160 DB=10*LOG(E/100)/LOG(10)  'EFFICIENCY IN dB
170 DB=FIX(DB*10)/10
180 L=(5.6624E-07*S^.6984)*1.13 ' SELF INDUCTANCE OF LOOP
190 XL=2*3.1416*F*1000000!*L    ' INDUCTIVE REACTANCE
200 Q=XL/(2*(RR+RL))    ' Q OF THE ANTENNA
210 Q=FIX(Q)
220 BW=F*1000000!/Q  ' BANDWIDTH OF THE ANTENNA IN MHz
230 BW=FIX(BW)/1000  ' BANDWIDTH IN KHz
240 CS=1.1374E-11+4.6841E-13*S 'DISTRIBUTED CAPACITY OF THE ANTENNA
250 CT=1/((2*3.1416*F*1000000!)^2*L)-CS ' REQUIRED TUNING C-CS
260 CT=FIX(CT*1E+12)
280 VC=(P*XL*Q)^.5 'VOLTAGE ACROSS THE TUNING CAPACITOR
290 VC=FIX(VC)
300 SC=VC/30000! ' PLATE SPACING OF TUNING CAPACITOR IN INCHES
310 SC=FIX(SC*1000)/1000
320 PRINT:PRINT"RESULTS": PRINT
330 PRINT"FOR A ";S;" FOOT PIPE, ";D;" INCH DIAMETER AT ";F;" MHz ---"
340 PRINT"RADIATION RESISTANCE OF THE ANTENNA  = ";RR;" OHMS"
350 PRINT"THE LOSS RESISTANCE OF THE ANTENNA = ";RL; " OHMS"
360 PRINT"EFFICIENCY = ";E;" PERCENT, WHICH IS (RELATIVE TO 100%) = ";DB;" dB"
370 PRINT"THE Q OF THE ANTENNA = ";Q
380 PRINT"THE BANDWIDTH OF THE ANTENNA = ";BW;" KHz"
390 PRINT"THE VALUE OF THE TUNING CAPACITOR  = ";CT;" pfd"
400 PRINT"FOR ";P;" WATTS, TUNING CAP VOLTAGE = ";VC;" VOLTS"
410 PRINT"THE CAPACITOR PLATE SPACING MUST BE > ";SC;" INCHS"
420 H=S/8
430 PRINT "THE HEIGHT (AND WIDTH) OF THIS LOOP IS ";H;" FEET
440 PRINT" - - - - - - - - - - - - - - - - - - - - - - - - - -": PRINT
450 PRINT"WANT TO TRY A DIFFERENT FREQUENCY ? (Y/N) "
460 X$=INPUT$(1)
470 IF X$="Y" THEN 480 ELSE 500
480 INPUT"SPECIFY THE NEW FREQUENCY IN MHz - ":F
490 GOTO 90
500 PRINT:PRINT"WANT TO CHANGE THE CONDUCTOR LENGTH ? (Y/N) "
510 X$=INPUT$(1)
520 IF X$="Y" THEN 530 ELSE 550
530 INPUT "SPECIFY THE NEW PIPE LENGTH IN FEET - ";S
540 GOTO 90
550 END
```

HIGH SENSITIVITY APPLICATIONS OF LOW POWER INTEGRATED CIRCUITS

DONALD ANDERSON
Strategic Marketing Manager
RF Communications Products
Signetics Corporation
811 E. Arques Avenue
Sunnyvale, CA 94086

ABSTRACT

This paper discusses four high sensitivity IF strips which utilize intermediate frequencies of 10.7MHz or greater. Each circuit utilizes a low power VHF mixer and high performance low power IF strip. The circuit configurations are (1) 45 or 49MHz to 10.7MHz narrowband, (2) 90MHz to 21.4MHz narrowband, (3) 100MHz to 10.7MHz wideband, and (4) 152.2MHz to 10.7MHz narrowband. Each circuit is presented with an explanation of component selection criteria, (to permit adaptation to other frequencies and bandwidths). Optional configurations for local oscillators and data demodulators are summarized.

INTRODUCTION

Traditionally, the use of 10.7MHz intermediate frequencies has been an attractive means to accomplish reasonable image rejection. However, applying significant gain at a high IF has required extensive gain stage isolation to avoid instability and very high current consumption to get adequate amplifier gain bandwidth. By enlightened application of two relatively new low power ICs, Signetics NE602 and NE604A, it is possible to build highly producible IF strips and receivers with input frequencies to several hundred megahertz, IF frequencies of 10.7 or 21.4MHz, and sensitivity less than 2μV (in many cases less than 1μV). The Signetics new NE605 combines the function of the NE602 and the NE604A. All of the circuits described in this paper can also be implemented with the NE605. The NE602 and NE604A were utilized for this paper to permit optimum gain stage isolation and filter location.

First let's look at why it is relevant to use a 10.7 or 21.4MHz intermediate frequencies. 455MHz ceramic filters offer good selectivity and small size at a low price. Why use a higher IF? The fundamental premise for the answer to this question is that the receiver architecture is a hetrodyne type as shown in Figure 1.

A pre-selector (bandpass in this case) precedes a mixer and local oscillator. An IF filter follows the mixer. The IF filter is only supposed to pass the difference (or sum) of the local oscillator (LO) frequency and the pre-selector frequency.

The reality is that there are always two frequencies which can combine with the LO: The pre-selector frequency and the "image" frequency. Figure 2 shows two hypothetical pre-selection curves. Both have 3dB bandwidths of 2MHz. This type of pre-selection is typical of consumer products such as cordless telephone and FM radio. Figure 2A shows the attenuation of a low side image with 10.7MHz. Figure 2B shows the very limited attenuation of the low side 455kHz image.

If the single conversion architecture of Figure 1 were implemented with a 455kHz IF, any interfering image would be received almost as well as the desired frequency. For this reason, dual conversion, as shown in Figure 3, has been popular.

In the application of Figure 3, the first IF must be high enough to permit the pre-selector to reject the images of the first mixer and must have a narrow enough bandwidth that the second mixer images can be attenuated. There's more to it than that, but those are the basics. The multiple conversion hetrodyne works well, but, as Figure 3 suggests, compared to Figure 2 it is more complicated. Why, then, don't we use the approach of Figure 2?

Historically there has been a problem: Stability! Comercially available IF amplifiers have been limited to about 60dB of gain. Higher discrete gain was possible if each stage were carefully shielded and bypassed, but this can become a nightmare on a production line. With so little IF gain available, in order to receive signals of less than 10μV it was necessary to add RF gain and this, in turn, meant that the mixer must have good large signal handling capability. The RF gain added expense, the high level mixer added expense, both added to the potential for instabilities, so the multiple conversion started looking good again.

But why is instability such a problem in a high gain high IF strip? There are three basic mechanisms. First, ground and the supply line are potentially feedback mechanisms from stage-to-stage in any amplifier. Second, output pins and external components create fields which radiate back to inputs. Third, layout capacitances become feedback mechanisms. Figure 4 shows the fields and capacitances symbolically.

If Z_F represents the impedance associated with the circuit feedback mechanisms, (stray capacitances, inductances and radiated fields) and Z_{IN} is the equivalent input impedance, a divider is created. This divider must have an attenuation factor greater than the gain of the amplifier to remain stable.

- If gain is increased, the input-to-output isolation factor must be increased.

- As the frequency of the signal or amplifier bandwidth increases, the impedance of the layout capacitance decreases thereby reducing the attenuation factor.

The layout capacitance is only part of the issue. In order for traditional 10.7MHz IF amplifiers to operate with reasonable gain bandwidth, the amount of current in the amplifiers needed to be quite high. The CA3089 operates with 25mA of typical quiescent current. Any currents which are not perfectly differential must be carefully bypassed to ground. The higher the current, the more difficult the challenge. And limiter outputs and quadrature components make excellent field generators which add to the feedback scenario. The higher the current, the larger the field.

THE SOLUTION

The NE602 is a double balanced mixer suitable for input frequencies up to 500MHz. It draws a 2.5mA of current. The NE604A is an IF strip with over 100dB of gain and a 25MHz small signal bandwidth. It draws 3.5mA of current. The circuits in this paper will demonstrate ways to take advantage of this low current, and 75dB or more of the NE604A gain in IF strips that would not be possible with traditional integrated circuits. No special tricks are used, only good layout, impedance planning and gain distribution.

THE MIXER

The NE602 is a low power VHF mixer with built-in oscillator. The equivalent circuit is shown in Figure 5.

The basic attributes of this mixer include conversion gain to frequencies greater than 500MHz, a noise figure of 4.6dB @ 45MHz, and a built-in oscillator which can be used up to 200MHz. LO can be injected.

For best performance with any mixer, the interface must be correct. The input impedance of the NE602 is high, typically 3kΩ in parallel with 3pF. This is not an easy match from 50Ω. In each of the examples which follow, an equivalent 50:1.5k match was used. This compromise of noise, loss, and match yielded good results. It can be improved upon. Match to crystal filters will require special attention, but will not be given focus in this paper.

The oscillator is a single transistor with an internal emitter follower driving the mixer. For best mixer performance, the LO level needs to be approximately $220mV_{RMS}$ at the base of the oscillator transistor. A number of oscillator configurations are presented at the end of this paper. In each of the prototypes for this paper, the LO source was a signal generator. Thus, a 51Ω resistor was used to terminate the signal generator. The LO is then coupled to the mixer through a DC blocking capacitor. The signal generator is set for 0dBm. The impedance at the LO input (Pin 6) is approximately 20k. Thus, required power is very low, but 0dBm across 51Ω does provide the necessary $220mV_{RMS}$.

The outputs of the NE602 are loaded with 1.5k internal resistors. This makes interface to 455kHz ceramic filters very easy. Other filter types will be addressed in the examples.

THE IF STRIP

The basic functions of the NE604A are ordinary at first glance: Limiting IF, quadrature detector, signal strength meter, and mute switch. However, the performance of each of these blocks is superb. The IF has 100dB of gain and 25MHz bandwidth. It is this feature which will be exploited in the examples. The signal strength indicator has a 90dB log output characteristic with very good linearity. There are two audio outputs with greater than 300kHz bandwidth and 180° phase relationship (one can be muted greater than 70dB). The total supply current is typically 3.5mA. This is the other factor which permits high gain and high IF.

Figure 6 shows an equivalent circuit of the NE604A. Each of the IF amplifiers has a 1.6kΩ input impedance. The input impedance is achieved by splitting a DC feedback bias resistor. The input impedance will be manipulated in each of the examples to aid stability.

BASIC CONSIDERATIONS

In each of the circuits presented, a common layout and system methodology is used. The basic circuit is shown symbolically in Figure 7.

At the input, a frequency selective transformation from 50Ω to 1.5k permits analysis of the circuit with an RF signal generator. A second generator provides LO. This generator is terminated with a 51Ω resistor. The output of the mixer is high impedance (1.5k). In order to keep the circuit stable, the first IF filter acts as an impedance down converter. The input impedance of the first limiter is modified with an external resistor. In most of the examples, a 430Ω external resistor was used to create a 330Ω input impedance (430//1.5k). The same basic treatment was used between the first and second limiters. However, in each of the 10.7MHz examples, this interstage filter is not an L/C tank; it is a ceramic filter. This will be explained in the first example. After the second limiter, a conventional quadrature detector demodulates the FM or FSK information from the carrier and a simple low pass filter completes the demodulation process at the audio outputs.

As mentioned, a single layout was used for each of the examples. The board artwork is shown in Figure 8. Special attention was given to: (1) Creating a maximum amount of ground plane with connection of the component side and solder side ground at locations all over the board; (2) careful attention was given to keeping a ground ring around each of the gain stages. The objective was to provide a shunt path to ground for any stray signal which might feed back to an input; (3) leads were kept short and relatively wide to minimize the potential for them to radiate or pick up stray signals; and finally, (4) bypass was done as close as possible to supply pins and inputs.

EXAMPLE: 45MHz to 10.7MHz NARROWBAND

As a first example, consider conversion from 45MHz to 10.7MHz. There are commercially available filters for both frequencies so this is a realistic combination for a second IF in a UHF receiver. This current can also be applied to cordless telephone or short range communications at 46 or 49MHz. The circuit is shown in Figure 9.

The 10.7MHz filter chosen is a type commonly available for 25kHz channel spacing. It has a 3dB bandwidth of 15kHz and a termination requirement of 3k//2pF. To present 3k to the input side of the filter, a 1.5k resistor was used between the NE602 output (which has a 1.5k impedance) and the filter. Layout capacitance was close enought to 2pF that no adjustment was necessary. This series resistance approach introduces an insertion loss which degraded the sensitivity, but it has the benefit of simplicity.

The secondary side of the crystal filter is terminated with a 10.7MHz tuned tank. The capacitor of the tank is tapped to create a transformer with the ratio for 3k:330. With the addition of the 430Ω resistor in parallel with

the NE604A 1.5k internal input resistor, the correct component of resistive termination is presented to the crystal filter. The inductor of the tuned load is adjusted off resonance enough to provide the 2pF capacitance needed. (Actual means of adjustment was for best audio during alignment).

If appropriate or necessary for sensitivity, the same type of tuned termination used for the secondary side of the crystal filter can also be used between the NE602 and the filter. If this is desired, the capacitors should be ratioed for 1.5k:3k. Alignment is more complex with tuned termination on both sides of the filter. This approach is demonstrated in the fourth example.

A ceramic filter is used between the first and second limiters. It is directly connected between the output of the first limiter and the input of the second limiter. Ceramic filters act much like ceramic capacitors, so direct connection between two circuit nodes with different DC levels is acceptable. At the input to the second limiter, the impedance is again reduced by the addition of a 430Ω external resistor in parallel with the internal 1.5k input load resistor. This presents the 330Ω termination to the ceramic filter which the manufacturers recommend.

On the input side of the ceramic filter, no attempt was made to create a match. The output impedance of the first limiter is nominally 1k. Crystal filters are tremendously sensitive to correct match. Ceramic filters are relatively forgiving. A review of the manufacturers' data shows that the attenuation factor in the passband is affected with improper match, but the degree of change is small and the passband stays centered. Since the principal selectivity for this application is from the crystal filter at the input of the first limiter, the interstage ceramic filter only has to suppress wideband noise. The first filter's passband is right in the center of the ceramic filter passband with constant attenuation factor and phase characteristics. (The crystal filter passband is less than 10% of the ceramic filter passband). This passband relationship is illustrated in Figure 10.

After the second limiter, demodulation is accomplished in the quadrature detector. Quadrature criteria is not the topic of this paper, but it is noteworthy that the choice of loaded Q will affect performance. The NE604A is specified at 455kHz using a quadrature capacitor of 10pF and a tuning capacitor of 180pF. (180pF gives a loaded Q of 20 at 455kHz). A careful look at the quadrature equations (REF) suggests that at 10.7MHz a value of about 1pF should be substituted for the 10pF at 455kHz.

The performance of this circuit is presented in Figure 11.

EXAMPLE: 90MHz to 21.4MHz NARROWBAND

This second example, like the first, used two frequencies which could represent the intermediate frequencies of a UHF receiver. This circuit can also be applied to VHF single conversion receivers if the sensitivity is appropriate. The circuit is shown in Figure 12.

Most of the fundamentals are the same as explained in the first example. The 21.4MHz crystal filter has a 1.5k//2pF termination requirement so direct connection to the output of the NE602 is possible. With strays there is probably more than 2pF in this circuit, but the performance is good

nonetheless. The output of the crystal filter is terminated with a tuned impedance-step-down transformer as in the previous example. Interstage filtering is accomplished with a 1k:330 step-down ratio. (Remember, the output of the first limiter is 1k and a 430Ω resistor has been added to make the second limiter input 330Ω). A DC blocking capacitor is needed from the output of the first limiter. The board was not laid out for an interstage transformer, so an "XACTO" knife was used to make some minor mods. Figure 13 shows the performance.

EXAMPLE: 100MHz to 10.7MHz WIDEBAND

This example represents three possible applications: (1) low cost, sensitive FM broadcast receivers, (2) SCA receivers and (3) data receivers. The circuit schematic is shown in Figure 14. While this example has the greatest diversity of application, it is also the simplest. Two 10.7MHz ceramic filters were used. The first was directly connected to the output of the NE602. The second was directly connected to the output of the first IF limiter. The secondary sides of both filters were terminated with 330Ω as in the two previous examples. While the filter bandpass skew of this simple single conversion receiver might not be tolerable in some applications, to a first order the results are excellent. (Please note that sensitivity is measured at -20dB in this wideband example.) Performance is illustrated in Figure 15.

EXAMPLE: 152.2MHz to 10.7MHz NARROWBAND

In this example (see Figure 16) a simple, effective, and relatively sensitive single conversion VHF receiver has been implemented. All of the circuit philosophy has been described in previous examples. In this circuit, tuned-transformed termination was used on the input and output sides of the crystal filter. Performance is shown in Figure 17.

OSCILLATORS

The NE602 contains an oscillator transistor which can be used to frequencies greater than 20MHz. Some of the possible configurations are shown in Figures 18 and 19.

L/C

When using a synthesizer, the LO must be externally buffered. Perhaps the simplest approach is an emitter follower with the base connected to Pin 7 of the NE602. The use of a dual-gate MOSFET will improve performance because it presents a fairly constant capacitance at Gate 1, and because it has very high reverse isolation.

CRYSTAL

With both of the Colpitts crystal configurations, the load capacitance must be specified. In the overtone mode, this can become a sensitive issue since the capacitance from the emitter to ground is actually the equivalent capacitive reactance of the harmonic selection network. The Butler oscillator uses an overtone crystal specified for series mode operation (no parallel capacitance). It may require an extra inductor (L_o) to null out C_o of the crystal, but otherwise is fairly easy to implement (see references).

The oscillator transistor is biased with only 220μA. In order to assure oscillation in some configurations, it may be necessary to increase

transconductance with an external resistor from the emitter to ground. Too small a resistance can upset DC bias (see references).

DATA DEMODULATION

It is possible to change any of the examples from an audio receiver to an amplitude shift keyed (ASK) or frequency shift keyed (FSK) receiver or both with the addition of an external op amp(s) or comparator(s). A simple example is shown in Figure 20. ASK decoding is accomplished by applying a comparator across the received signal strength indicator (RSSI). The RSSI will track IF level down to below the limits of the demodulator (-120dBM RF input in most of the examples). When an in-band signal is above the comparator threshold, the output logic level will change.

FSK demodulation takes advantage of the two audio outputs of the NE604A. Each is a PNP current source type output with 180° phase relationship. With no signal present, the quad tank tuned for the center of the IF passband, and both outputs loaded with the same value of capacitance, if a signal is received which is frequency shifted from the IF center, one output voltage will increase and the other will decrease by a corresponding absolute value. Thus, if a comparator is differentially connected across the two outputs, a frequency shift in one direction will drive the comparator output to one supply rail, and a frequency shift in the opposite direction will cause the comparator output to swing to the opposite rail. Using this technique, and L/C filtering for a wide IF bandwidth, NRZ data at rates greater than 4Mbaud have been processed with the NE602/604A and the new NE605.

References

1) R. Matthys: "Survey of VHF Crystal Oscillator Circuits", *RF Technology Expo Proceedings*, pp 371-382, February, 1987.

2) Krauss, Raab, Bastian: *Solid State Radio Engineering*, p. 311, Wiley, 1980.

3) Signetics: "NE/SA604A High Performance Low Power FM IF System", *Linear Data and Applications Manual*, Signetics, 1987.

4) Signetics; "NE/SA602 Double Balanced Mixer and Oscillator",*Linear Data and Applications Manual*, Signetics, 1985.

5) Signetics: "AN1982-Applying the Oscillator of the NE602 in Low Power Mixer Applications",*Linear Data and Applications Manual*, Signetics, 1985.

6) Anderson, D.: "Low Power ICs for RF Data Communication", Machine Design, pp 126-128, July 23, 1987.

Special thanks to Craig Hritz, Ali Fotowat, and Alvin Wong for their assistance in the development of this paper.

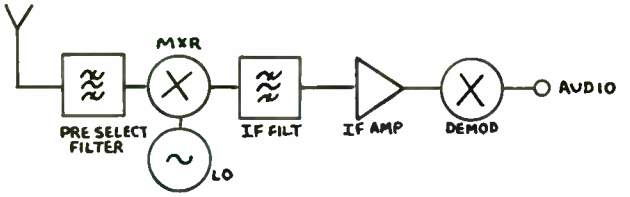


Figure 1. Basic Hetrodyne Receiver

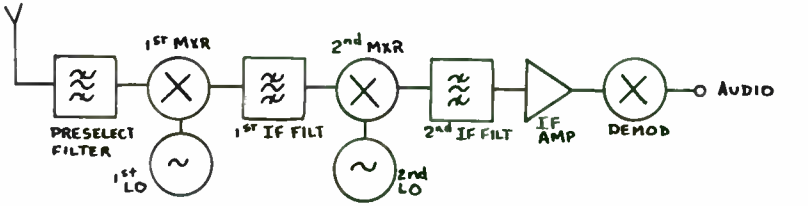


Figure 3. Dual Conversion

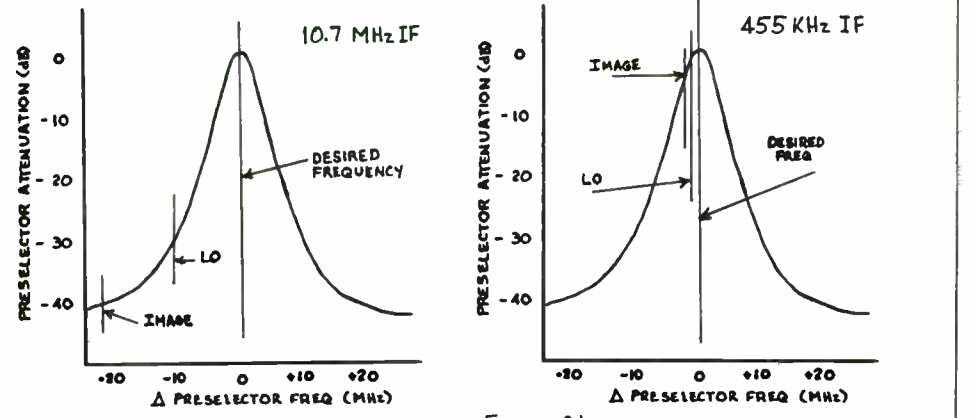


FIGURE 2a

FIGURE 2b

Figure 2. Effects of Preselection on Images

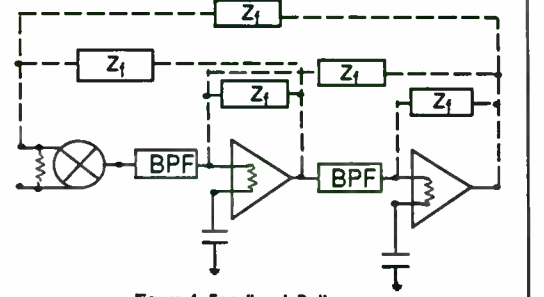


Figure 4. Feedback Paths

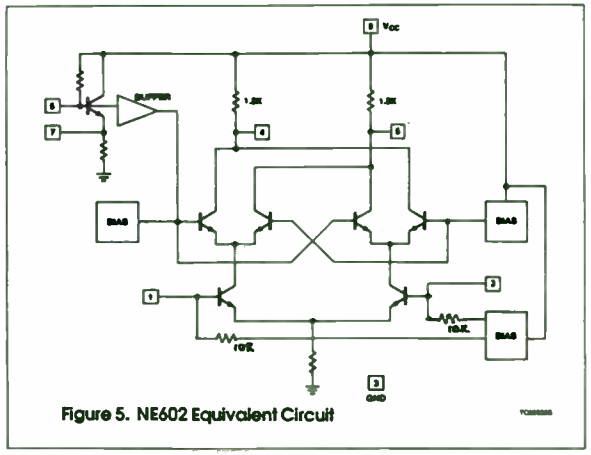


Figure 5. NE602 Equivalent Circuit

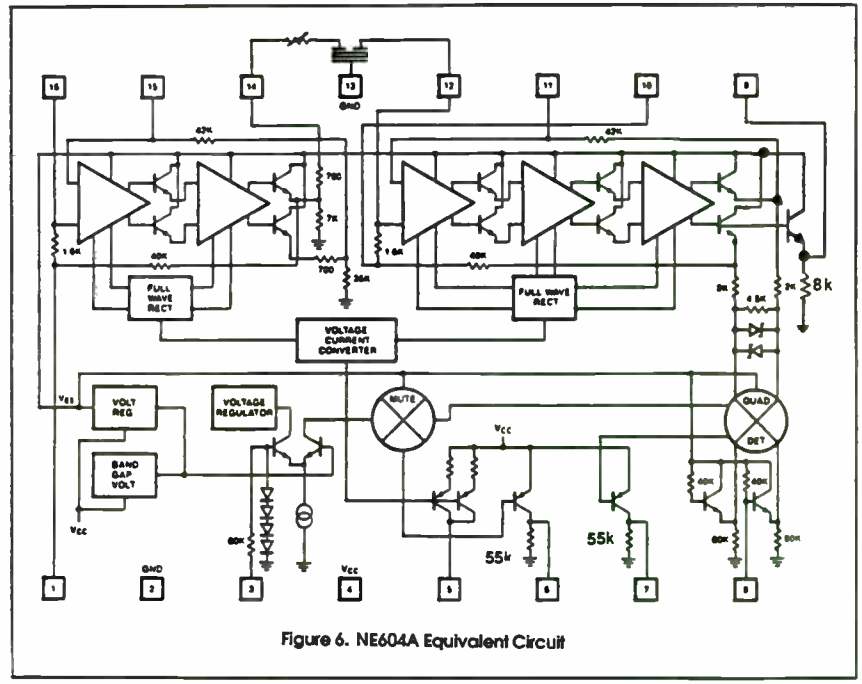


Figure 6. NE604A Equivalent Circuit

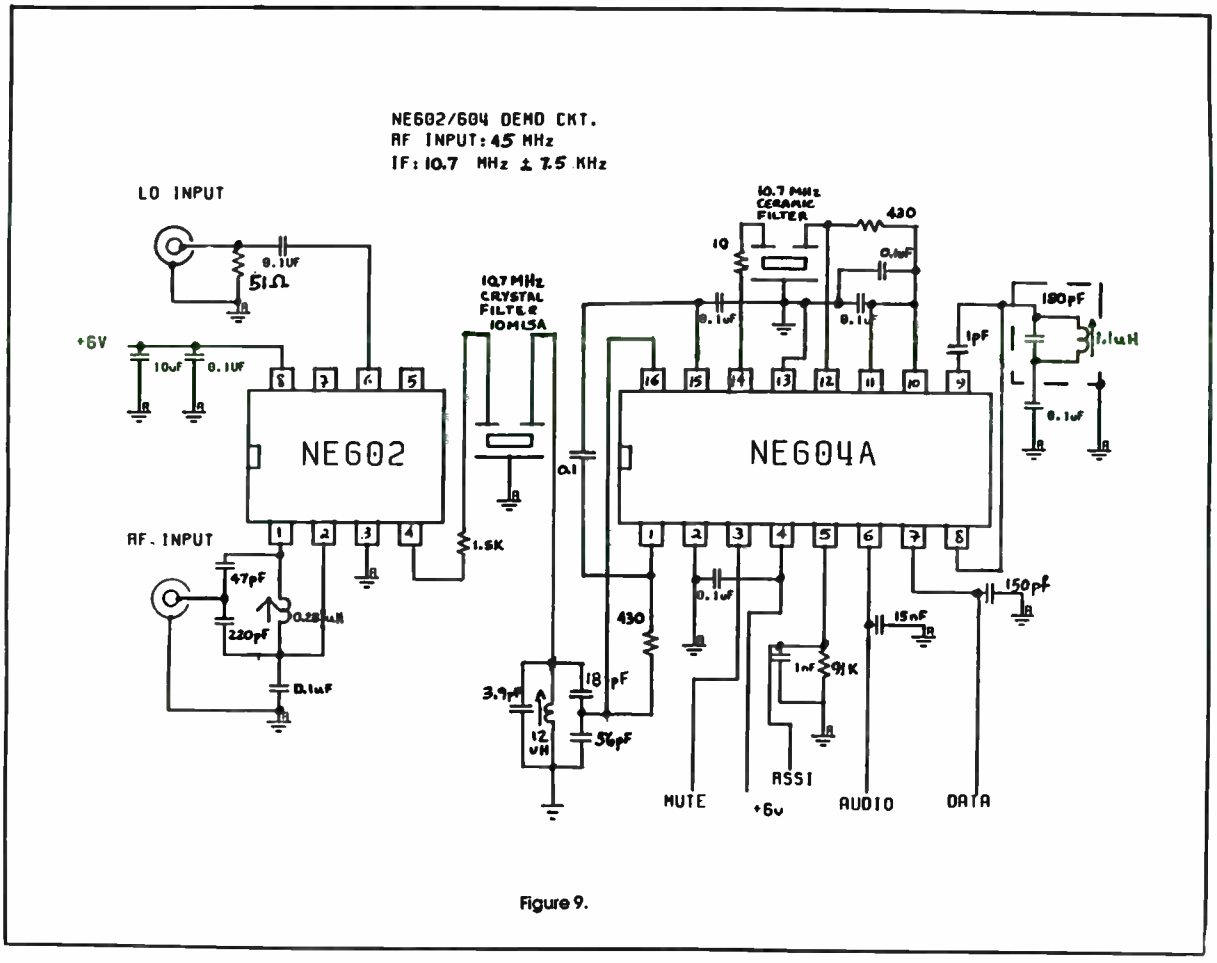


Figure 9.

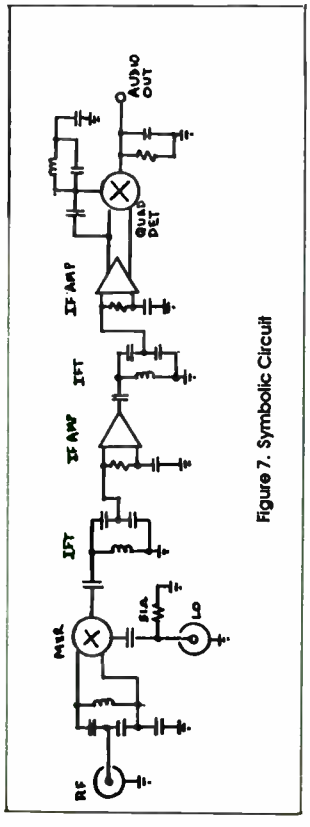
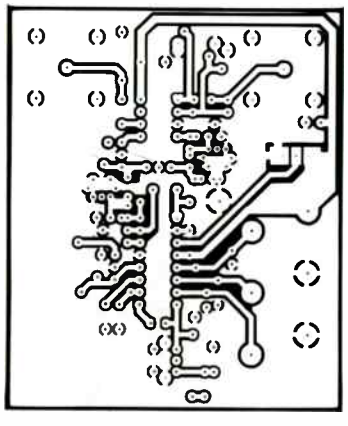


Figure 7. Symbolic Circuit



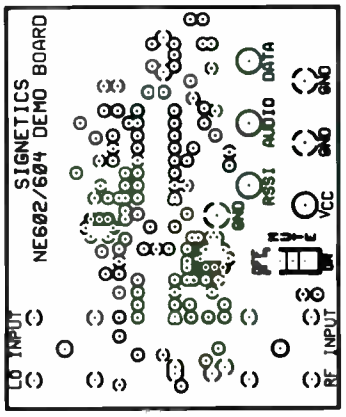


Figure 8. Circuit Board Layout

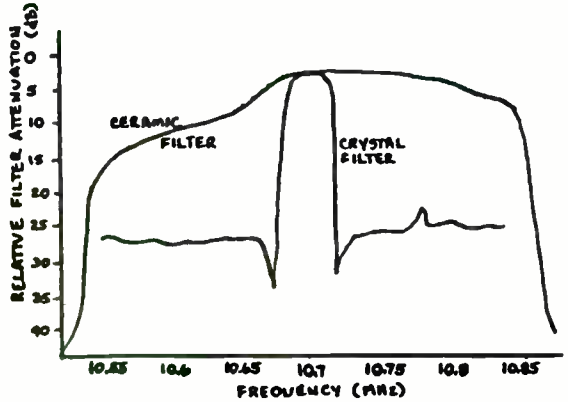
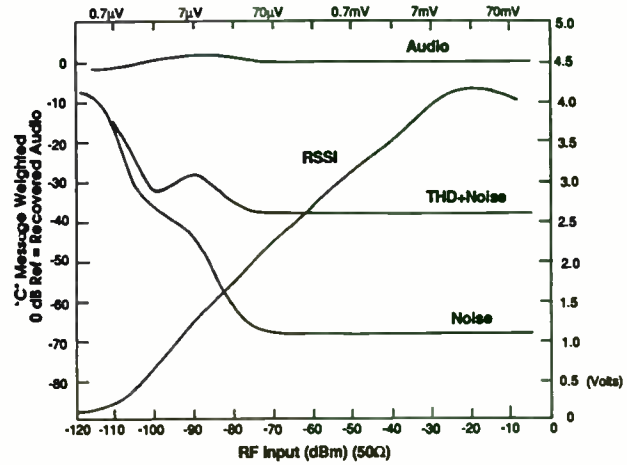


Figure 10. Passband Relationship



- 45MHz RF Input
- 55.7MHz LO (0 dBm)
- 10.7MHz IF
- 15kHz IF BW
- 7kHz deviation

Figure 11. VHF or UHF 2nd Conversion (Narrow Band)

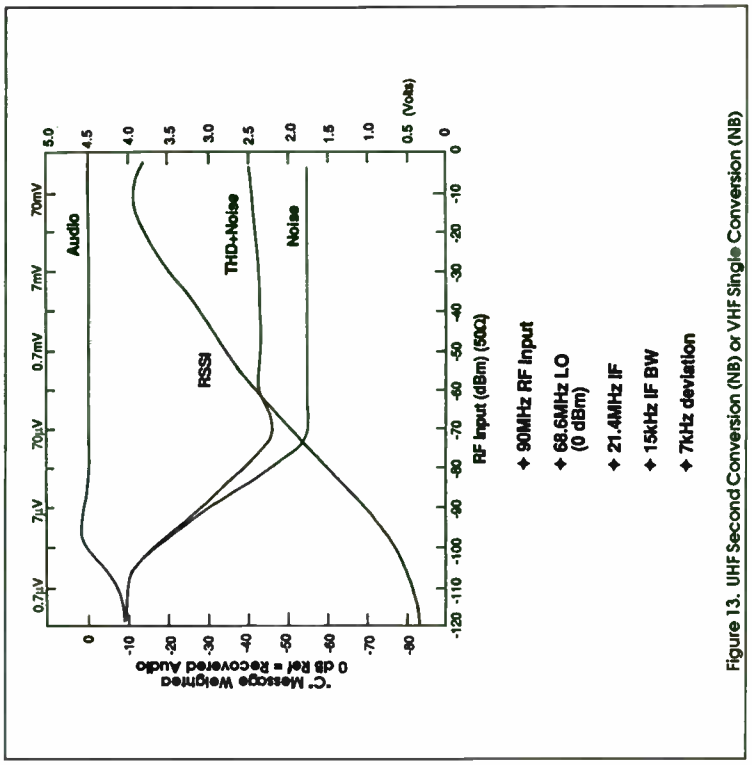


Figure 13. UHF Second Conversion (NB) or VHF Single Conversion (NB)

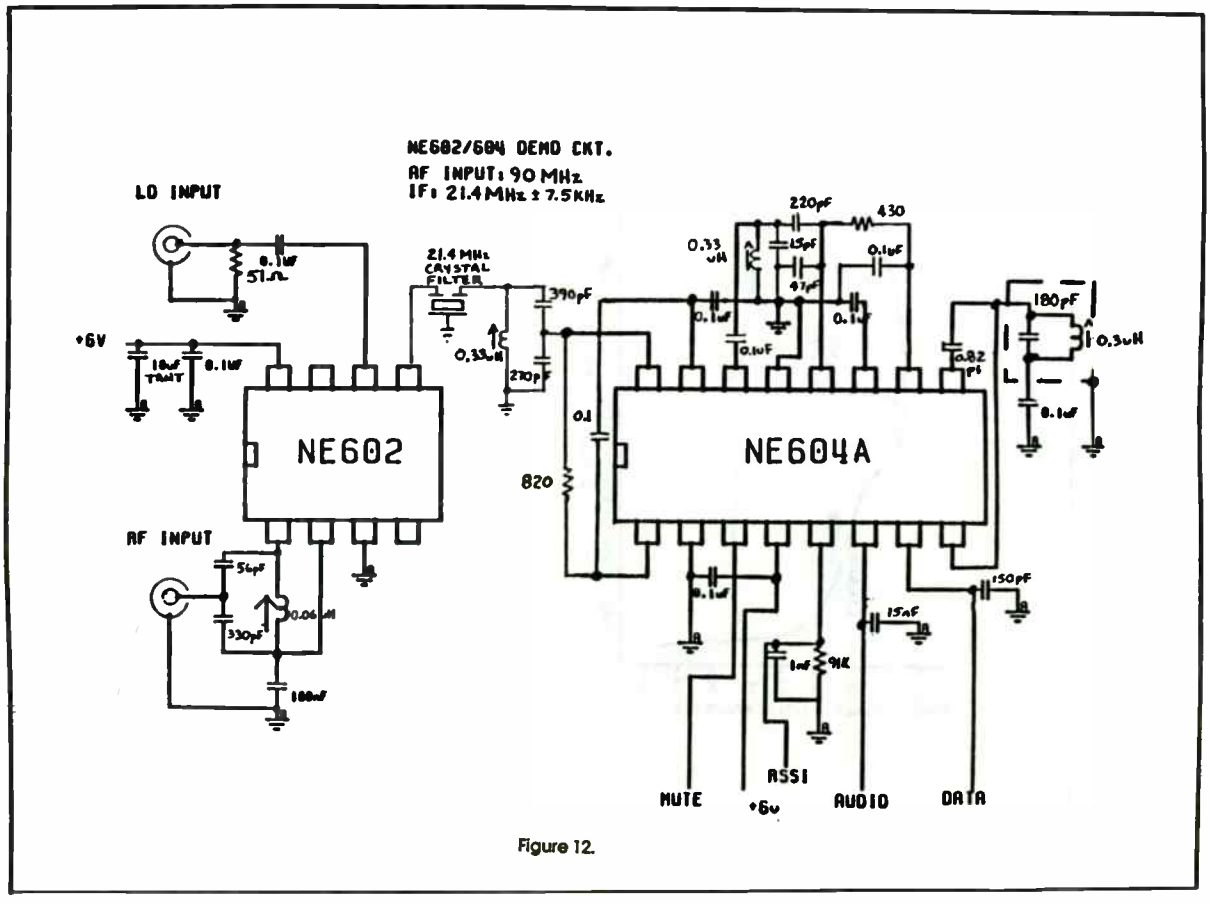


Figure 12.

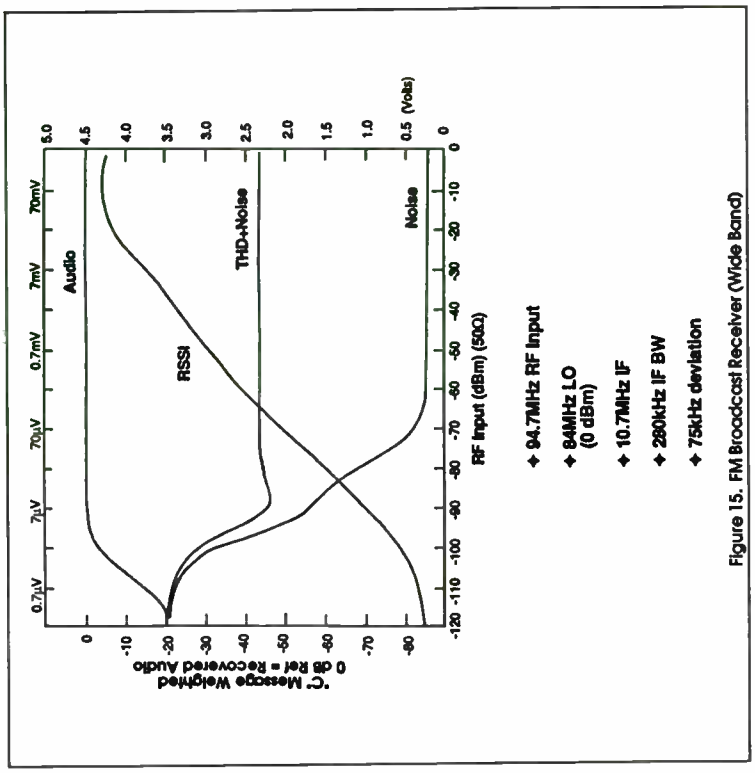


Figure 15. FM Broadcast Receiver (Wide Band)

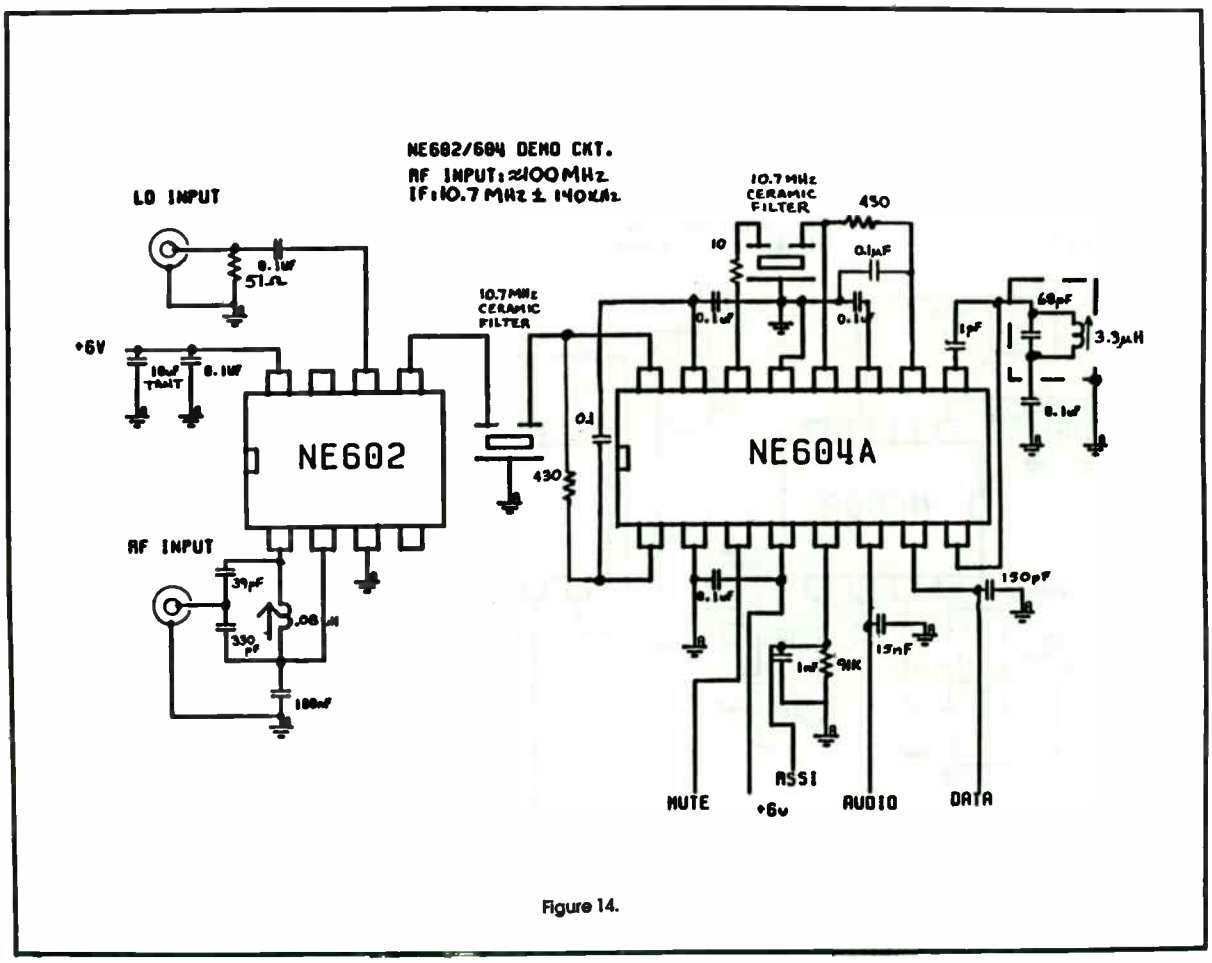


Figure 14.

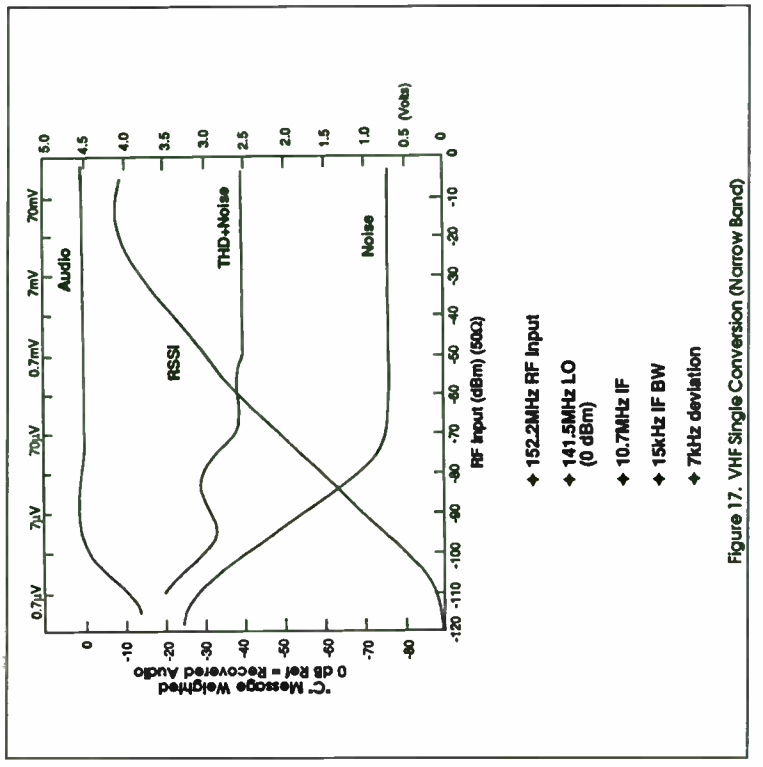


Figure 17. VHF Single Conversion (Narrow Band)

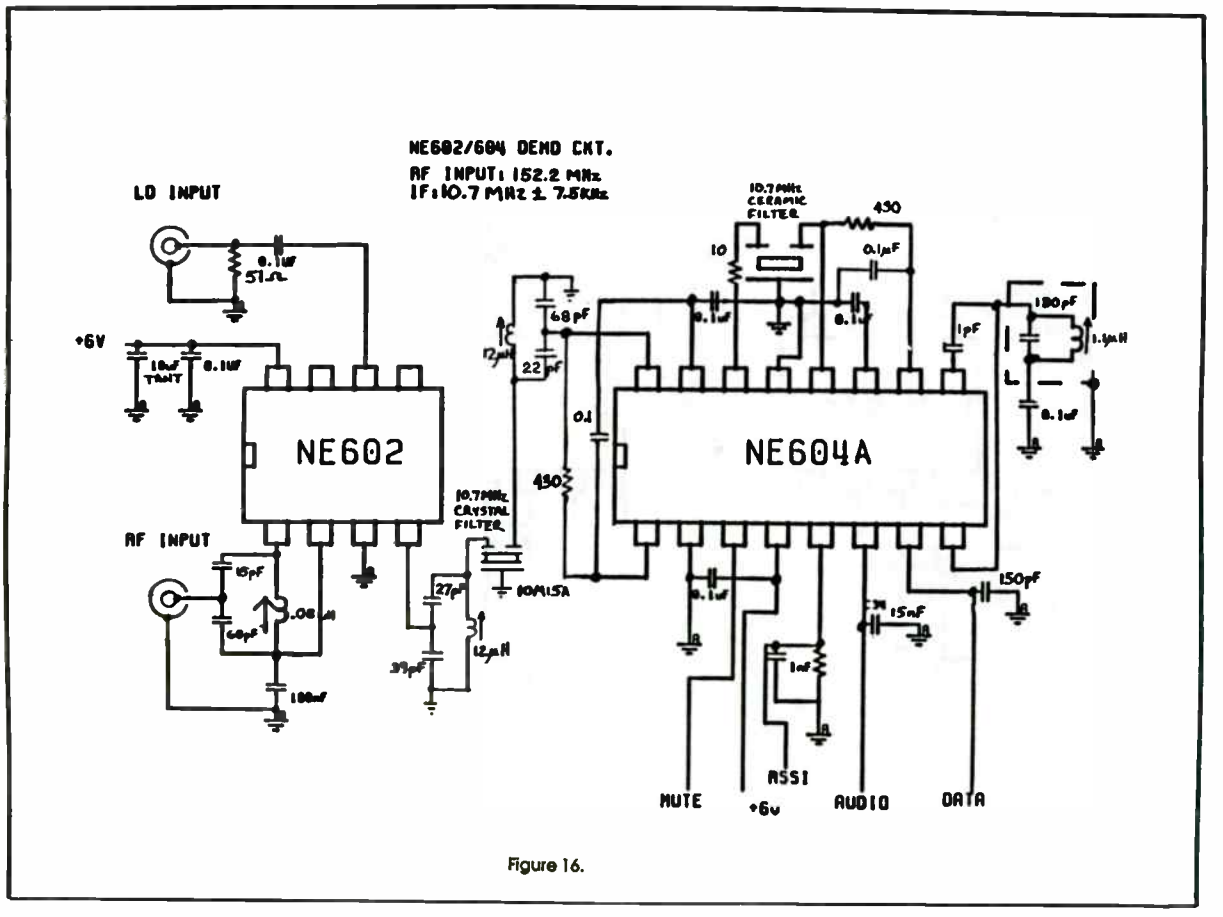


Figure 16.

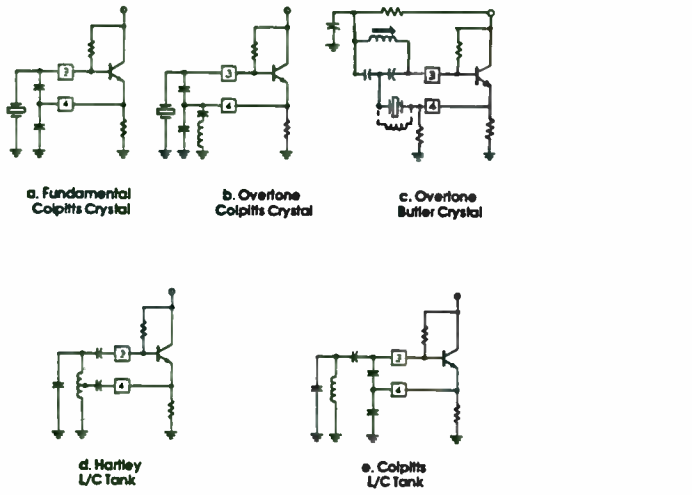


Figure 18. Oscillator Configurations

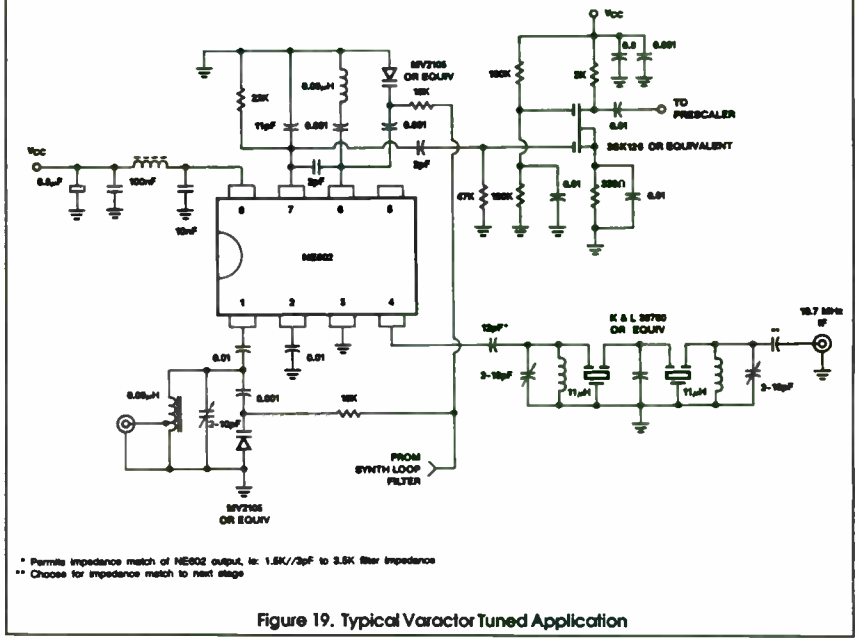


Figure 19. Typical Varactor Tuned Application

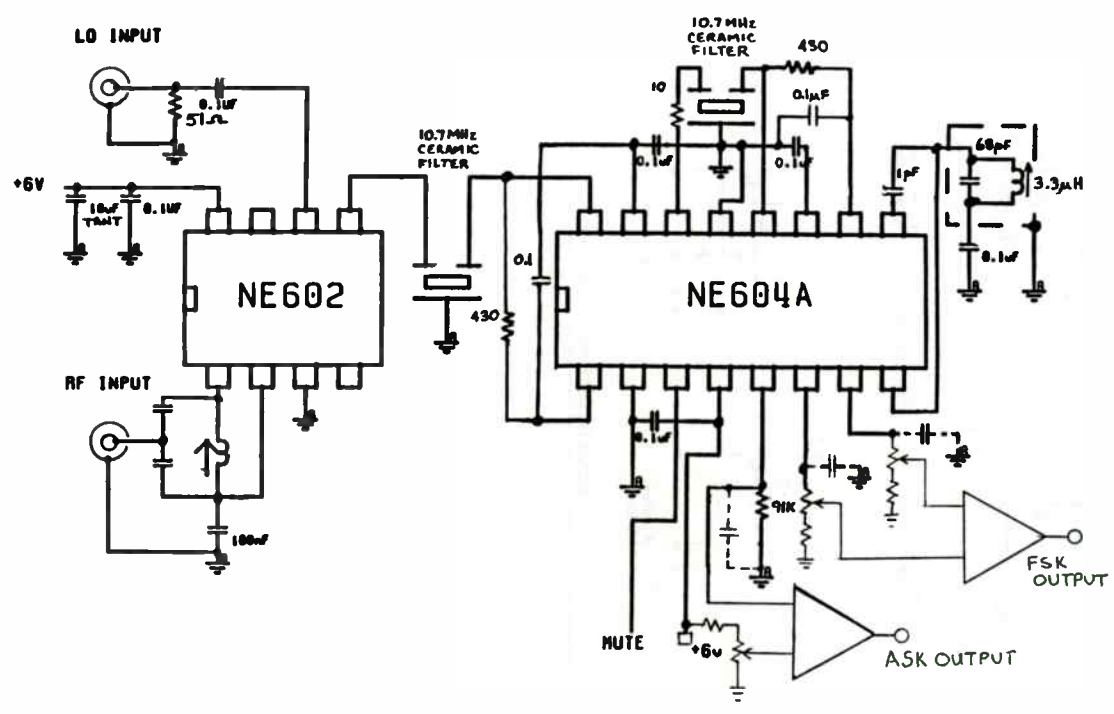


Figure 20. Basic NE602/604 Data Receiver

PHASE SHIFT MINIMIZED MICROSTRIP PIN DIODE ATTENUATORS

by

Robert J. Baeten*

T. Koryu Ishii

James S. Hyde**

Department of Electrical, Computer and Biomedical Engineering
Marquette University
1515 West Wisconsin Avenue
Milwaukee, Wisconsin 53233

ABSTRACT

Spurious phase shift minimized variable attenuators are important components in microwave instrumentation and measurements. In waveguide technology, a spurious phase shift of 0.1 degree/dB is achievable by using a rotary vane type attenuator at 9 GHz with 50 dB attenuation. Published data exists on microwave PIN microstrip attenuators having a spurious phase shift of 1.0 degree/dB at 9 GHz with 20 dB attenuation. The method described in this paper produces 0.14 degree/dB phase shift at 14 dB attenuation and possibly at 42 dB attenuation at 9 GHz. The design method employed consists of a series of sectionalized ABCD matrices CAD. The design results were confirmed through measurements. It was found that by adjusting the inter-component distances in the microstrip PIN diode circuit board, spurious phase shift could be reduced.

---

*R. J. Baeten is currently with Rockwell-International, Cedar Rapids, Iowa.
**J. S. Hyde is with the Medical College of Wisconsin, Milwaukee, Wisconsin.
This work was supported in part by Grants GM 27665 and RR 01008 from the National Institutes of Health.

INTRODUCTION

Attenuators are one of the most essential parts of microwave impedance bridges. For compact light weight variable microwave attenuators, microstrip PIN diode attenuators are preferred over heavy and bulky waveguide type mechanical variable attenuators in microwave instrumentation and measurement. Problems associated with microstrip PIN diode attenuators are however its inherent phase shift. Ideally, zero phase shift is desirable at any attenuation setting. In reality, greater phase shift is experienced at higher attenuation settings.

Current waveguide microwave impedance bridges utilize rotary-vane attenuators whose phase shift versus attenuation is typically less than 5 degrees over a 0-50 dB range [1]. The phase shift per dB attenuation is 0.1 degree/dB at the maximum attenuation. Such a small phase shift results in minimum bridge imbalance and yields excellent results. In order to construct a microwave impedance bridge on microstrip transmission line, a low-phase-shift of 5 degrees or less, 0-40 dB, variable attenuator is desired for optimum results.

PIN diodes have been used in the past to attenuate microwave signals and the results are well documented in the literature [2]-[14]. Only four authors have mentioned or discussed phase shift versus attenuation of PIN diode attenuators [6], [12]-[14].

Imai [6] examines PIN diodes mounted in waveguide over the 9.0-9.6 GHz frequency range. Over a 0.4 GHz bandwidth the worst case phase shift was 20 degrees for 20 dB of attenuation. This is 1°/dB phase shift per dB attenu-

ation. A maximum attenuation of 35.5 dB was achieved at 9.35 GHz at a forward bias current of 150 mA. No phase shift measurements were presented for attenuation values greater than 20 dB.

Erkinge and Hedstrom [12] constructed a stripline PIN diode attenuator using three-port hybrids from 8.0-12.0 GHz. They did not measure the phase shift versus attenuation but merely mentioned that if the PIN diode's impedance contained a reactive part there will be a phase shift along with the attenuation. The maximum attenuation range using one diode in each branch of the hybrid was 28 dB.

Ananasso [13] realized a low phase shift step attenuator with PIN diode switches and resistive pads in a Microwave Integrated Circuit (MIC) at S-band. The phase shift was less than 11 degrees from 0 to 70 dB attenuation from 2 to 4 GHz. In this example, the PIN diodes are used for switching and not for the attenuation elements.

Parris [14] realized a broad-band MIC attenuator from 0.5 to 3.0 GHz using two conventional tee-section resistive attenuators. This type of attenuator has constant input and output impedances versus attenuation. Three PIN diodes are used in each tee-section. The phase shift was 20 degrees for 25 dB of attenuation. In this frequency range the phase shift per dB attenuation is 0.8°/dB.

Though mechanical waveguide type variable attenuators enjoy low phase shift of 0.1 °/dB at 50 dB, current record of PIN diode attenuator phase shift is 1 °/dB at 20 dB at 9 GHz.

The objective of this paper is to present methods and results of computer aided analysis and design of microstrip PIN diode reflection-type attenuators. The attenuator utilizes one or more packaged PIN diodes mounted in microstrip transmission line. Relationship among the bias current, attenuation, phase shift and reflections were formulated. Based on this formulation, the PIN diode attenuators were designed under desirable frequency range and the designs were confirmed experimentally.

DESIGN PRINCIPLES

A schematic diagram of an example of packaged PIN diode microstrip attenuator circuit studied in this work is shown in Fig. 1. The PIN diodes [15]-[17] employed were CSB-7401-01 (Package 375). A computer program was developed using cascade ABCD matrices which consists of various transmission circuits [18]-[41]. First the attenuator circuit shown in Fig. 1 is sectionalized to many elementary sections. These sections include sections of the signal source, coax to microstrip transition [37]-[39] at the input, input microstrip [18]-[33][40],[41] two packaged PIN diode mount sections [15]-[17] three microchip capacitor sections, output microstrip, the output microstrip to coax transition and the output circuit section. [40] Various ABCD matrices were developed for each section and finally an over-all ABCD matrix was developed by cascading all matrices of each section. In this analysis, all loss [32][33] and dispersion [19],[24]-[26],[29] associated with the microstrip sections and the effect of bias circuit parameters [35] were incorporated. Once the over-all ABCD matrix is formulated in computer programming, the rest of the work is to adjust the circuit parameters to produce the desired phase shift with desired attenuation value at a desired operating frequency [42].

SECTIONAL ABCD MATRIX FORMULAE

Most ABCD matrices of each sections of the microstrip PIN diode attenuator circuit are trivial. For example, in Fig. 1, ABCD matrices of signal source section, co-axial to microstrip transition section, microstripline section, chip capacitor section and the load impedance section are trivial. These matrices can be found elsewhere. [18]-[42]

The ABCD matrix of the PIN diode section can be obtained based on an equivalent circuit of the PIN diode section [15]-[17] as shown in Fig. 2. In this figure, C_P is the package capacitance, L_S is the lead inductance and R_F is the forward biased diode resistance at operating frequency. The ABCD matrix of the packaged PIN diode section can be written as follows by observation.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ j\omega C_P & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/(R_F+j\omega L_S) & 1 \end{bmatrix} \quad (1)$$

DESIGN EXAMPLES

From the over-all system ABCD matrix attenuation and phase shift of this PIN diode microstrip attenuator can be calculated. Calculated attenuation of the PIN diode attenuator circuit in Fig. 1 is shown in Fig. 3 and the phase shift is shown in Fig. 4. In this calculation, many circuit parameters are limited by the commercial circuit components available on the market. [15]-[17] For example, the microstrip to coax transition employed was an SMA connector type 50-645-4547-31. All chip capacitors were ATC type. Microstrip was formed on RT Duroid 6010.2 in which ϵ_r = 10.2, h = 0.635 mm, τ = 0.036 mm, tan δ = 0.002. As seen from Figs. 3 and 4, this attenuator produces 46 dB attenuation with 20 degree phase shift at 8.9 GHz. This is only 0.43°/dB. This exceeds previously reported results.

FABRICATED PIN DIODE ATTENUATOR AND TEST RESULTS

Based on these calculations, the actual circuit schematically shown in Fig. 5 was built. The maximum attenuation observed at each frequency was plotted in Fig. 6 and the relationship between the attenuation setting and the observed phase shift was plotted in Fig. 7.

Comparing Figs 1 and 5, an isolator was added in the actual circuit to reduce backward coupling between the two PIN diodes. PIN diode biasing circuits are also shown in Fig. 5.

DISCUSSION

Due to the additions of bias circuits and the isolator, the maximum attenuation characteristic of the actual attenuator was not exactly as calculated as can be seen in Figs 3 and 6. However, the actual performance was not too far from the calculated values. By comparing Figs 4 and 7, the amount of the phase shift at maximum attenuation measured to be approximately equal to the calculated value. According to Fig. 7, this fabricated attenuator produced 46 dB attenuation with a phase shift of 31 degrees at 8.5 GHz. This is 0.7°/dB phase shift. This phase shift per dB attenuation is still an improvement over the previously reported results which were 1 °/dB at 9 GHz and 0.8°/dB at 3 GHz.

Further improvement of the phase shift per dB attenuation of this PIN diode can be obtained by inspecting a characteristic of a single PIN diode phase shifter as shown in Fig. 8. This phase shifter was designed and fabricated using the same procedure of the two PIN diode phase shifter described before. The measured results are shown in Fig. 9. According to these results, the phase shift up to 14 dB attenuation is 2 degrees at 9 GHz. This is only 0.14°/dB phase shift per dB attenuation. Therefore, if the attenuator is truncated at 14 dB and two, three or more stages of identical single PIN diode attenuators are employed in cascade, a variable attenuator of maximum attenuation up to 28 dB, 42 dB and more with 0.14°/dB phase shift per dB attenuation is possible.

CONCLUSION

As seen from the example, a method for designing and analyzing minimized phase shift microstrip PIN diode attenuator was developed. It was found that among other parameters, some of the most critical design factors affecting attenuation and phase shift of this microstrip PIN diode attenuator were the distance between the PIN diodes, and the distance between each PIN diode and the chip capacitors.

REFERENCES

[1] A. L. Lance, Introduction to Microwave Theory and Measurements, New York: McGraw-Hill Book Company, p. 115, 1964.

[2] R. V. Garver, Microwave Diode Control Devices, Dedham: Artech House, Inc., Chapter 8, 1976.

[3] J. K. Hunton and A. G. Ryals, "Microwave Variable Attenuators and Modulators Using PIN Diodes," IRE Trans. on Microwave Theory and Tech., vol. MTT-10, no. 7, pp. 262-273, July 1962.

[4] Hewlett Packard, "Application of PIN Diodes," Application note 922.

[5] L. N. Dworsky, "Computer-Optimized Design of PIN Diode Absorption Attenuators," Microwave Journal, pp. 39-41,49, Feb. 1976.

[6] Hiroshi Imai, "Microwave PIN Diode Attenuator," Mem. Chubu Inst. Technology (Japan), vol. 10A, pp. 47-53, Sept. 20, 1974.

[7] Idem, "Quarter-Wavelength Coupled PIN Diode Attenuator," Mem. Chubu Inst. Technology (Japan), vol. 11A, pp. 43-47, 1975.

[8] Idem, "Design of Microwave PIN Diode Attenuator," Mem. Chubu Inst. Technology (Japan), vol. 12A, pp. 55-61, 1976.

[9] Idem, "Characterization of Microwave 3-PIN Diode Attenuator," Mem. Chubu Inst. Technology (Japan), vol. 13A, pp. 61-66, 1977.

[10] Idem, "Microwave 3-PIN Diode Attenuator by Optimum Design," Mem. Chubu Inst. Technology (Japan), vol. 14A, pp. 61-65, 1978.

[11] Idem, "Design of a Broadband Microwave PIN JDiode Attenuator II," Mem. Chubu Inst. Technology (Japan), vol. 19A, pp. 41-48, 1983.

[12] R. Erkinge and T. Hedstrom, "A New Variable Microwave Attenuator," IEEE Trans. on Microwave Theory and Tech., vol. MTT-18, no. 9, pp. 661-662, Sept. 1970.

[13] F. G. Ananasso, "A Low Phase Shift Step Attenuator Using PIN Diodes," IEEE Trans. on Microwave Theory and Tech., vol. MTT-28, no. 7, pp. 774-776, July 1980.

[14] W. J. Parris, "PIN Variable Attenuator With Low Phase Shift," IEEE Trans. on Microwave Theory and Tech., vol. MTT-20, no. 9, pp. 618-619, Sept. 1972.

[15] Hewlett Packard, "Selection and Use of Microwave Diode Switches and Limiters," Application Note 932, May 1973.

[16] Hewlett Packard, "PIN Diode RF Resistance Measurement," Application Bulletin 6.

[17] Hewlett Packard, Diode and Transistor Designer's Catalog 1982-83, pp. 139.

[18] G. I. Zysman and D. Varon, "Wave Propagation in Microstrip Transmission Lines," IEEE G-MTT Int. Symp., 1969 Digest, pp. 3-9, 1969.

[19] R. Mittra and T. Itoh, "A New Technique For The Analysis of The Dispersion Characteristics of Microstrip Lines," IEEE Trans. on Microwave Theory and Tech., vol. MTT-19, no. 1, pp. 47-56, Jan. 1971.

[20] E. C. Jordan and K. G. Balmain, Electromagnetic Waves and Radiating Systems, Englewood Cliffs: Prentice-Hall, Inc., pp. 218-219, 1950.

[21] E. O. Hammerstad, "Equations for Microstrip Circuit Design," Proc. European Microwave Conf., pp. 268-272, 1975.

[22] I. J. Bahl and Ramesh Garg, "Simple and Accurate Formulas for a Microstrip With Finite Strip Thickness," Proc. of the IEEE, vol. 65, no. 11, pp. 1611-1612, Nov. 1977.

[23] W. J. Gestinger, "Microstrip Dispersion Model," IEEE Trans. on Microwave Theory and Tech., vol. MTT-21, no. 1, pp. 34-39, Jan. 1973.

[24] T. C. Edwards and R. R. Owens, "2-18 GHz Dispersion Measurements on 10-100 Ohm Microstrip Lines on Sapphire," IEEE Trans. on Microwave Theory and Tech., vol. MTT-24, no. 8, pp. 506-513, Aug. 1976.

[25] O. P. Jain, et al., "Coupled Mode Model of Dispersion in Microstrip," Electronic Letters, vol. 7, pp. 405-407, July 1971.

[26] M. V. Schneider, "Microstrip Dispersion," Proc. IEEE, vol. 60, pp. 144-146, 1972.

[27] H. J. Carlin, "A Simplified Circuit Model for Microstrip," IEEE Trans. on Microwave Theory and Tech., vol. MTT-21, no. 9, pp. 589-591, Sept. 1973.

[28] G. Kompa and R. Mehran, "Planar Waveguide Model for Calculating Microstrip Components," Electronic Letters vol. 11, pp. 459-460, Sept. 1975.

[29] B. Bianco, et al., "Frequency Dependence of Microstrip Parameters," Alta Frequenza, vol. 43, pp. 412-416, 1974.

[30] K. C. Gupta, Microstrip Lines and Slotlines, pp. 20-28, Artech House, Dedham, MA, 1979.

[31] H. A. Wheeler, "Transmission Line Properties of a Stripline Between Parallel Plates," IEEE Trans. on Microwae Theory and Tech., vol. MTT-26, no. 11, pp. 866-876, Nov. 1978.

[32] R. A. Pucel, et al., "Losses in Microstrip," IEEE Trans on Microwave Theory and Tech., vol. MTT-16, no. 6, pp. 342-350, June 1968. Also see "Correction to Losses in Microstrip," vol. MTT-16, no. 12, p. 1064, Dec. 1968.

[33] J. D. Welch and H. J. Pratt, "Losses in Microstrip Transmission Systems for Integrated Microwave Circuits," NEREM Rec., vol. 8, pp. 100-101, 1966.

[34] G. Matthaei, L. Young, E. M. T. Jones, Microwave Filters, Impedance-

Matching Networks, and Coupling Structures, Artech House, Dedham, MA, 1980.

[35] B. A. syrett, "A Broad-Band Element for Microstrip Bias or Tuning Circuits," IEEE Trans. on Microwave Theory and Tech., vol. MTT-28, no. 8, pp. 925-927, Aug. 1980.

[36] American Technical Ceramics, ATC VHF/Microwave Capacitors Performance Curves (Catalog), pp. 1-2, 1983.

[37] M. L. Majewski, R. W. Rose, and J. R. Scott, "Modeling and Characterization of Microstrip-to-Coaxial Transitions," IEEE Trans. on Microwave Theory and Tech., vol. MTT-29, no. 8, pp. 799-805, Aug. 1981.

[38] J. S. Wight, et al., "Equivalent Circuits of Microstrip Impedance Discontinuities and Launchers," IEEE Trans. on Microwave Theory and Tech., vol. MTT-22, no. 1, pp. 48-52, Jan. 1974.

[39] S. O. Ajose, "Equivalent Circuit of Coaxial-to-Microstrip Connectors Over the 8-12 GHz Range," Electronic Letters, vol. 13, no. 16, pp. 195-196, Aug. 4, 1977.

[40] K. C. Gupta, Ramesh Garg, and Rakesh Chadha, Computer-Aided Design of Microwave Circuits, p. 34, Artech House, Dedham, MA, 1981.

[41] J. Faulkner, et al., "Soft Microstrip With Integral Ground Plane Aids in Super Component Integration," Microwave Journal, pp. 105-106, 108, 110, 112-115, Nov. 1983.

[42] Robert J. Baeten, "Analysis and Chracterization of Microstrip PIN Diode Attenuators", MS Degree Thesis on file in Memorial Library, Marquette University, Milwaukee, WI 53233, May 1986.

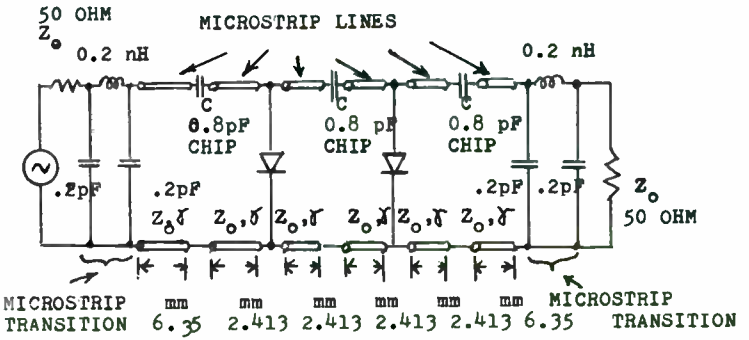


Fig. 1 A microstrip PIN diode attenuator circuit incorporating two minimum parasitic package PIN diodes with a center frequency of 9 GHz.

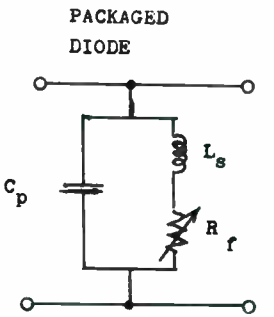


Fig. 2 Equivalent circuit of a PIN diode mount section.

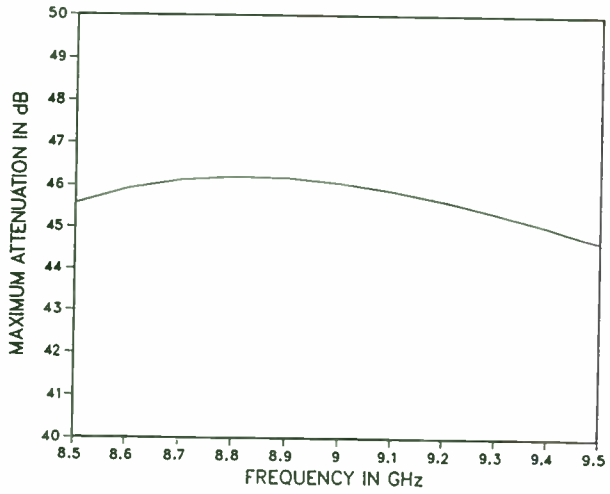


Fig. 3 Maximum attenuation versus frequency for the microstrip circuit shown in Fig. 1 including the microstrip-to-coaxial transitions. The maximum attenuation was determined when the resistance of the two PIN diodes were 3 ohms.

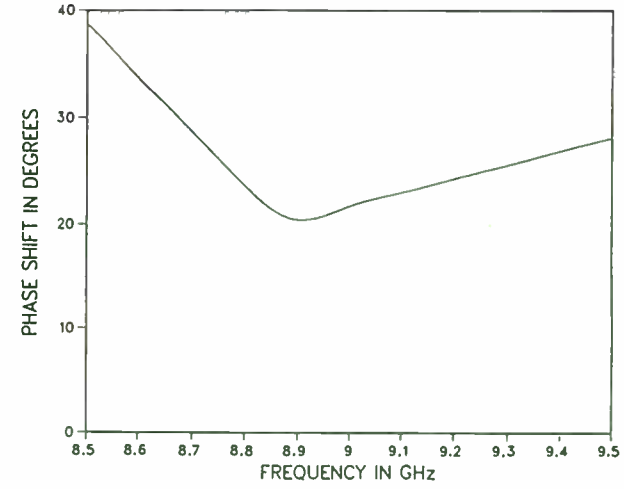
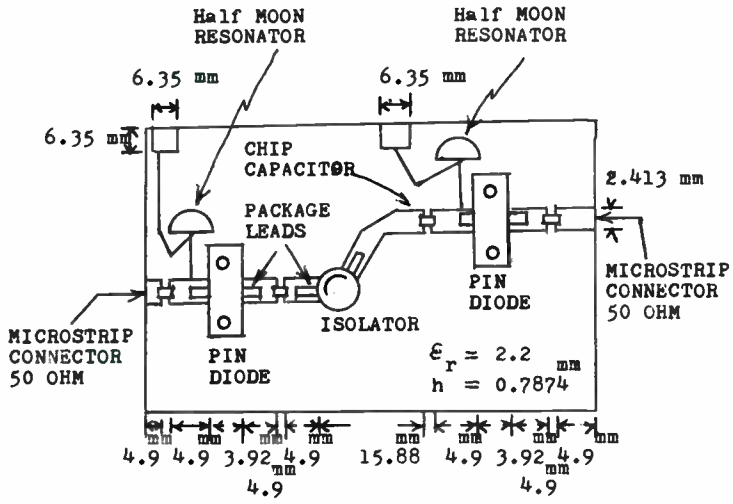


Fig. 4 Phase shift versus frequency for the microstrip circuit shown in Fig. 1 including the microstrip-to-coaxial transitions. The resistance of the PIN diodes were varied simultaneously from 10,000 ohms to 3 ohms.



-PIN DIODES ARE ALPHA CSB-7401-01, PACKAGE 375
-ISOLATOR IS TRAK-89A9101, 8 - 12 GHz.
-MICROSTRIP CONNECTORS ARE SAELECTRO #50-645-4547-31
-ALL CHIP CAPACITORS ARE ATC-0.9 pF

Fig. 5 A two PIN diode microstrip attenuator.

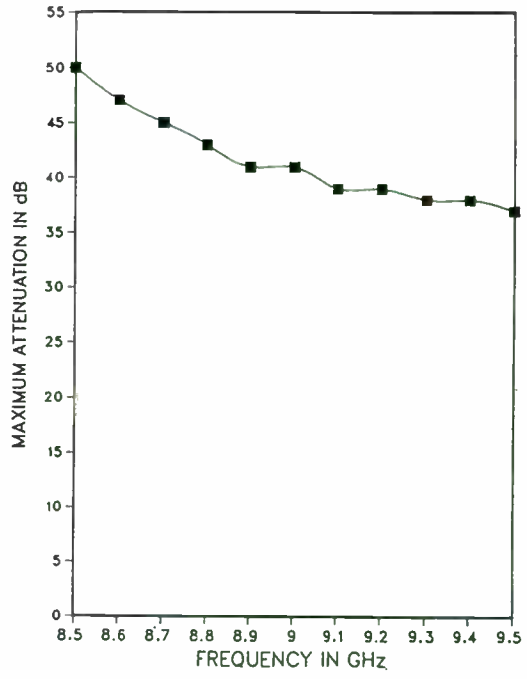


Fig. 6 Maximum attenuation versus frequency for two PIN diodes separated by an isolator on microstrip.

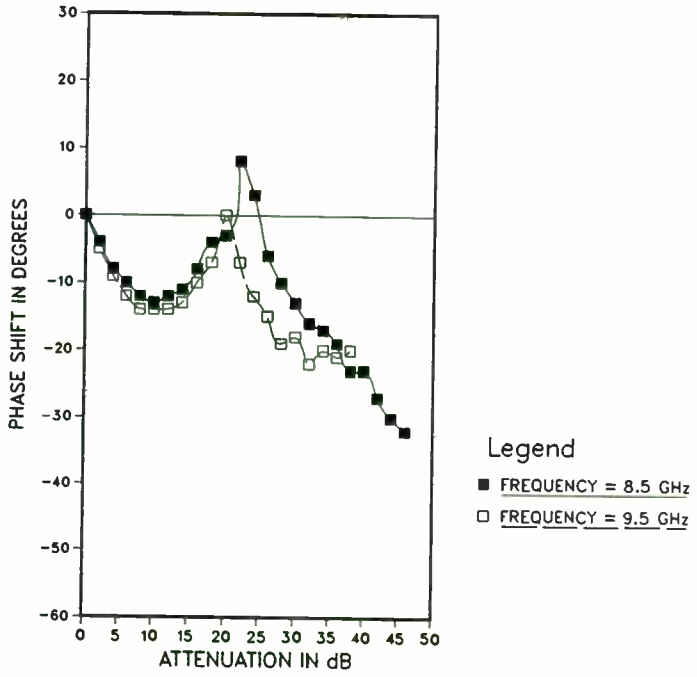


Fig. 7 Transmission phase versus attenuation for two PIN diode microstrip circuit.

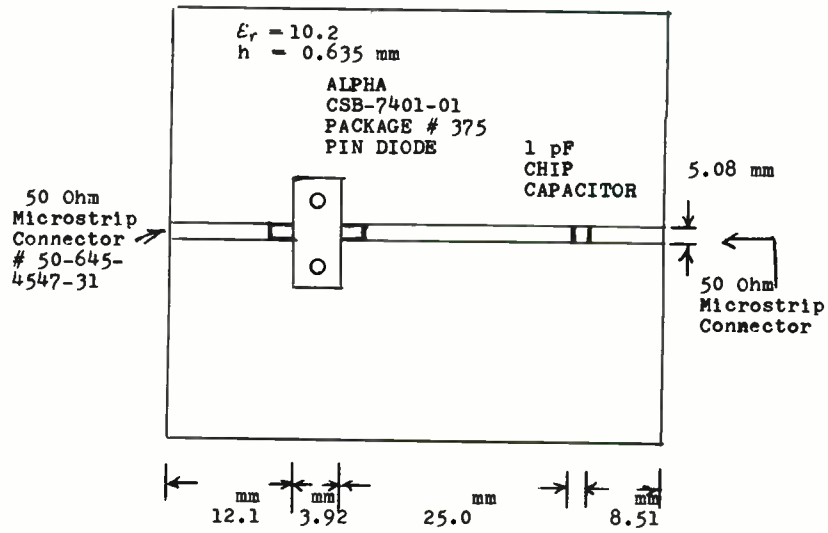


Fig. 8 A top view of the microstrip board for the single PIN diode attenuator circuit.

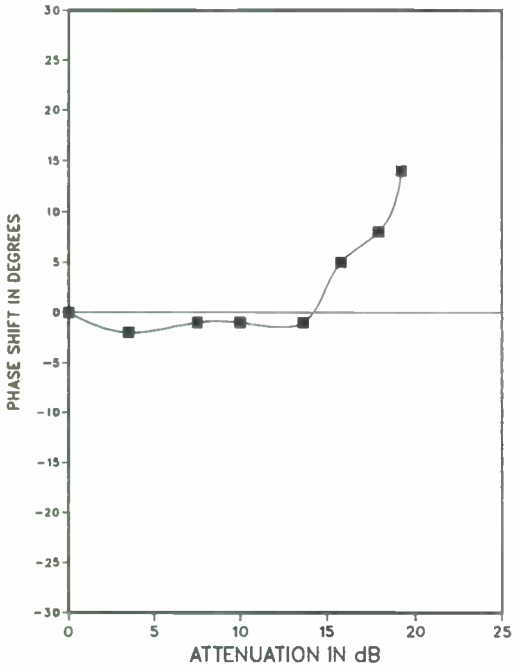


Fig. 9 Normalized phase shift versus attenuation for one PIN diode on microstrip as the bias current varies from 0 mA and 3 mA. Operating frequency is 9.0 GHz.

Proper Application of Power Dividers - Understanding Power Relationships

by Robert S. Larkin
Janel Laboratories, Inc.
33890 Eastgate Circle
Corvallis, OR 97333

INTRODUCTION - The power divider is a component which is widely used for the distribution of signals to multiple loads. For the most part the devices are straightforward and easy to apply. But there are some subtleties that need to be considered if the full performance of the divider is to be achieved. This paper examines the power divider in terms of its power relationships. The interaction between the divider and the external circuit is then expressed in terms of these power relationships.

Power dividers come in four basic types:

1) Resistive
2) Quadrature isolated
3) In-phase isolated
4) In-phase non-isolated

Resistive power dividers offer very wide bandwidths, often from dc to the millimeter wave region, all in a single device. They also generally offer excellent balance between outputs. The drawback is that these devices suffer from high insertion loss and relatively low isolation between output ports. This limits their application to instrumentation where the wide bandwidth is important.

Quadrature hybrids, with one port terminated, can be used as two way dividers. They have a 90° phase difference between the outputs that results in having a matched input as long as both output loads are identical. These devices find application in specialized applications such as amplifiers and mixers.

In-phase power dividers can be made to operate over a wide frequency range, can be made with isolated outputs and are available at reasonable cost. They are available with two or more outputs. These devices are widely used for signal distribution applications. The remainder of this paper will be concerned only with the in-phase power divider along with its operation as a power combiner.

ISOLATED AND NON-ISOLATED - The simplest power divider is formed by connecting the input port and all output ports together. This forms a primitive non-isolated divider. The obvious first improvement is to connect a transformer between the input and the parallel output ports. This allows the input port to be matched and provides full power transfer from the source to the output loads. This is the basic non-isolated power divider and it can be seen to have no internal dissipation. There are a number of configurations that the divider can take but they will all share a common network description, such as their scattering matrices.

Figure 1 shows the scattering matrices for the in-phase power dividers. The port numbering convention is shown in Figure 2. The scattering matrix for the non-isolated divider has

$$S_{ij} = \begin{cases} 0 & i=1, j=1 \\ \frac{1}{\sqrt{n}} & i=1 \text{ or } j=1,\ i \neq j \end{cases} \qquad (1)$$

indicating a match at the input port and the n-way power split. The other terms for the non-isolated divider result directly from the fact that the network is lossless with conservation of energy necessitating the following conditions [1] [2]:

$$\sum_{i=1}^{n+1} S_{ij} S_{ij}^{*} = 1 \quad \text{for } j = 1 \text{ to } n+1 \qquad (2)$$

$$\sum_{i=1}^{n+1} S_{ij} S_{ik}^{*} = 0 \quad j \neq k \qquad (3)$$

where the asterisk refers to the complex conjugate. It can be seen that the non-isolated divider has an isolation of 1/n (12 dB for a four-way divider).

Also shown in Figure 1 is a functional circuit for an isolated in-phase power divider along with the associated scattering matrix. The isolated divider includes dissipative networks that allow the isolation condition:

$$S_{ij} = 0 \qquad i>1 \quad j>1 \quad i \neq j \qquad (4)$$

Note that since the network is no longer lossless we can also achieve $S_{ii} = 0$, that is, all ports are matched. There are a number of electrical configurations that are used for isolated power dividers [3] [4]. Because of their wide usage, the following discussion will be limited to the isolated divider.

INSERTION LOSS - The power insertion loss of the ideal n-way isolated power divider is n or 10 $\log_{10}$n dB. This states that the power has been divided into n equal portions. In addition there is an actual power loss that represents heat dissipation in the divider. Reflections at both the input and output ports also produce added insertion loss. These three sources of insertion loss are generally well understood and can be determined from divider specifications and standard formulas. There is, however, an additional source of insertion loss that results when unterminated output ports exist, as illustrated by the following experiment.

Figure 3 shows the measured insertion loss of a four way divider with up to three unterminated output ports. With all the ports terminated, the insertion loss is seen to be about 6.5 dB in the 800-960 MHz region where the divider was designed to operate. Of this loss, 6 dB is that of an ideal four way divider and 0.5 dB is due to dissipative losses. Note that when ports are unterminated, the loss in the 800 to 960 MHz region varies, but only by about 0.5 dB or less.

At lower frequencies the story is quite different. With all ports terminated the insertion loss is relatively well behaved. With unterminated ports, however, the insertion loss goes through major excursions. To explain this behavior, two items must be noted. First, the unterminated ports were left with three foot lengths of RG223 coaxial cable attached. This causes the phase angle of the reflection to vary with frequency. Second, the isolation of the divider, shown in Figure 4, was close to ideal in the 800-960 MHz range but dropped to about 15 dB at 500 MHz. The variations in loss at the lower frequencies will be shown to be the result of reflections from the unterminated ports coming through the finite isolation of the divider, and adding and subtracting from the signal in the transmission path.

The scattering matrix for the n-way power divider with finite isolation can be approximated by

$$S = \begin{bmatrix} 0 & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & 0 & \cdots & \delta \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{n}} & \delta & \cdots & 0 \end{bmatrix} \qquad (5)$$

where δ is a small number. The 0 entries on the main diagonal represent a perfect impedance match at all ports.

We will now examine the forward transmission from the input port (port 1) to an output port, chosen to be port 2. We assume that m of the output ports (ports n-m+1 to n+1) are terminated in identical loads with reflection coefficient Γ_L. Thus n-m output ports are terminated in the characteristic impedance and have no reflection back to the divider. Writing out the expression for the reflected waves $\mathbf{b} = \mathbf{S}\,\mathbf{a}$ we have

$$b_1 = \frac{1}{\sqrt{n}}(\Gamma_L b_{n-m+1} + \ldots + \Gamma_L b_{n+1}) \qquad (6)$$

for the input port and

$$b_k = \frac{a_1}{\sqrt{n}} + \delta \sum_{\substack{i=n-m+1 \\ i \neq k}}^{n+1} a_i \qquad (7)$$

for the output ports (2 through n+1). The forward transmission from port 1 to port 2 is given by the ratio of the reflected wave out of port 2, b_2, to the incident wave at port 1, a_1:

$$\left| \frac{b_2}{a_1} \right| = \frac{1}{\sqrt{n}} \left[1 + \frac{m\,\delta\,\Gamma_L}{1 - (m-1)\,\delta\,\Gamma_L} \right] \qquad (8)$$

The quantity inside the brackets is the error ratio relative to the ideal case. Without loss of generality, one can assume a shift in reference plane so that δ is real, with the result that the maximum errors will occur for $\Gamma_L = \pm 1$, corresponding to open and short circuits. We can then express the maximum and minimum values of the error in dB as

$$e_{max} = 20 \log_{10}\left[1 + \frac{m\,\delta}{1 - (m-1)\,\delta}\right] \text{ dB} \qquad (9)$$

$$e_{min} = 20 \log_{10}\left[1 - \frac{m\,\delta}{1 + (m-1)\,\delta}\right] \text{ dB} \qquad (10)$$

These two equations are plotted in Figure 5 for a range of isolation δ. The error value is zero when the isolation is perfect, i.e., $\delta=0$. The error depends only on the isolation and the number of unterminated ports, m. The total number of output ports does not enter into the calculation.

It can be seen that it is important that dividers with high isolation be used if ports are to be left unterminated, or if the divider loads can be highly reflective. This latter case is common for many divider applications.

If the divider has a large number of outputs, and the phase of the reflection at the output ports can be assumed to be randomly distributed, then the variations in insertion loss may not be severe. This is due to the reflected signals both adding to and subtracting from the transmission path signal. In signal distribution applications, with highly reflective loads, this situation should be sought out. One should avoid having identical length output cables feeding identical loads, if possible.

ISOLATION - The isolation that is specified for commercial power dividers is measured with all ports terminated. In the case of output ports, this is not a major factor. However, the input port source impedance does have a major effect on the isolation. Figure 6 illustrates this effect using measured data. Again, this is a four way divider with isolation measured between two of the output ports. The input port source impedance was set with a return loss of 0, 6, and 12 dB, along with the perfectly terminated case. For the latter case, the isolation is always better than 30 dB. With the imperfect terminations, however, the isolation is almost constant with frequency and as will be shown below has a value of $20 \log_{10} (|\Gamma_s| / n)$ where Γ_s is the reflection coefficient of the source impedance.

The effect of source impedance on isolation can be evaluated by considering the scattering matrix of the isolated power matrix

$$\mathbf{S} = \begin{bmatrix} 0 & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{n}} & 0 & \cdots & 0 \end{bmatrix} \qquad (11)$$

The reflected power relationship $\mathbf{b} = \mathbf{S}\,\mathbf{a}$ for the first three rows expands to

$$b_1 = \sum_{i=2}^{n+1} \frac{a_i}{\sqrt{n}} \qquad (12)$$

$$b_2 = \frac{a_1}{\sqrt{n}} \qquad (13)$$

$$b_3 = \frac{a_1}{\sqrt{n}} \qquad (14)$$

Assuming that port 1 is terminated in an impedance with reflection coefficient Γ_s ($a_1 = \Gamma_s b_1$), that a signal is placed into port 2 so that

$$a_i = \begin{cases} a_2 & i = 2 \\ 0 & i > 2 \end{cases}$$

and that all other ports are terminated, we can solve for the isolation magnitude

$$\left|\frac{b_3}{a_2}\right| = \frac{|\Gamma_s|}{n} \qquad (15)$$

or, in dB, the isolation is $20 \log_{10} (|\Gamma_s| / n)$.

It can be seen that in order to achieve high isolation, it is necessary to not only use a power divider with high inherent isolation, but to also provide a well matched source impedance. In the previous section it was shown that for the insertion loss to be independent of unterminated output ports it was necessary to have high isolation. Here we have found that this poor isolation, and the resulting variations in insertion loss, may be the result of a poor source impedance match.

USE OF DIVIDERS AS COMBINERS - All power dividers can be used in reverse to combine multiple signals. When used in this fashion the insertion loss of the isolated power divider depends on the nature of the signals being combined. This is the result of power being dissipated internally in the balancing circuitry. The following three cases cover the normally encountered situations.

CASE 1 - Equal, same-frequency in-phase power sources. This case results in almost all of the power from the sources being delivered to the common (output) port. The only losses involved are the dissipative losses of the transformers inside the combiner along with some minor losses due to imbalances. For 2- or 4- Way combiners, these losses can generally be kept to less than 1 dB. In order to simplify the discussion we will ignore this class of losses for this and the following cases.

CASE 2 - Same as Case 1 but with some of the sources disabled. This results in a portion of the power being dissipated inside the combiner. In general, for an n-way combiner with m sources operating, the internal dissipation is

$$\frac{P_{diss}}{P_{max}} = \frac{m}{n} - \left(\frac{m}{n}\right)^2 \qquad (16)$$

where P_{max} is the power when all sources are operating. It can be shown that the greatest combiner dissipation occurs when half of the sources are disabled. Here the combiner dissipation is 25% of P_{max}. Figure 7 is a plot of this function for n = 8.

CASE 3 - Uncorrelated sources. This case results in an insertion loss of 1/n (or $-10 \cdot \log_{10} n$ dB) for each source applied to an n-way combiner. It follows that if m uncorrelated sources are applied, each with power level P (so that $P_{max} = nP$), the total output power is (mP/n) and the internal dissipation is

$$\frac{P_{diss}}{P_{max}} = m\left(\frac{1}{n} - \frac{1}{n^2}\right) \qquad (17)$$

In terms of internal dissipation, this is the most severe situation for power combiners. The power dissipated from each source is (n-1)/n times the available power from that source. For large n, the divider dissipates most of the available power. There are situations, however, where the isolated divider must be used and this penalty cannot be avoided.

CONCLUSIONS - It has been shown that when isolated power dividers are used to feed reflective loads, that it is important that the divider have high inherent isolation, in order to avoid major variations in the divider insertion loss. As an example, an isolation of 20 dB can result in as much as about ±2.5 dB variation for three mismatched output ports. It was further shown that if mismatched output loads are to be used, that steps should be taken to randomize the phase of the load reflections, by such means as varying the lengths of the attaching cables. Finally, it was shown that high isolation of the power divider requires that a good source match be provided to the divider input.

REFERENCES

[1] Ghose, R. N., *Microwave Circuit Theory and Analysis,* McGraw-Hill, New York, 1963, pp. 206-210.

[2] Collin, R. E., *Foundation for Microwave Engineering,* McGraw-Hill, New York, 1966, pp. 174-175.

[3] Wilkinson, E. J., "An N-way Hybrid Power Divider," *IEEE Trans. Microwave Theory and Techniques,* MTT-8, no. 1, 1960, pp116-118.

[4] Kim, D. I., "Design of Generalized n-way Power Divider for CATV and/or MATV Systems," *IEEE Trans. Consumer Electronics,* CE-32, no. 2, 1986, pp116-121.

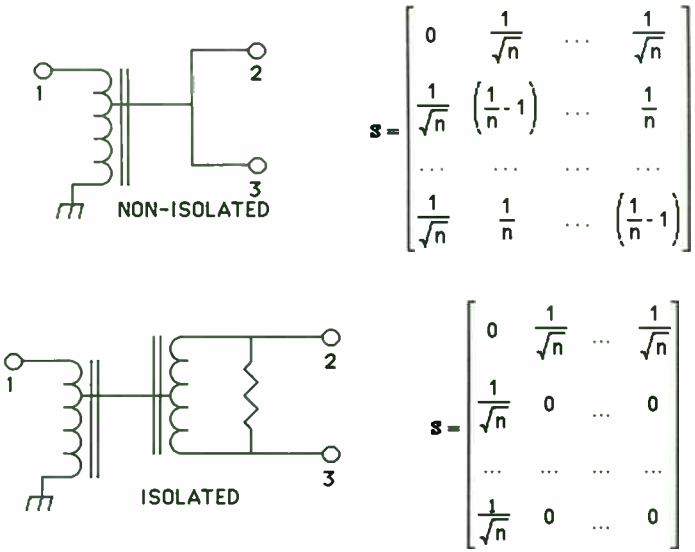


Figure 1. Non-isolated and isolated power dividers showing functional schematics (for n=2) and associated scattering matrices.

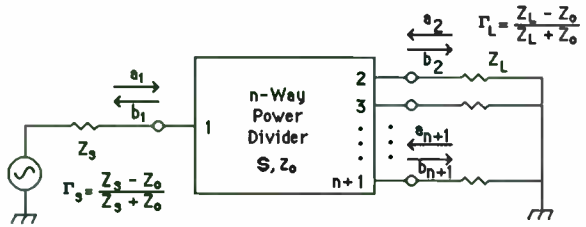


Figure 2. Power divider nomenclature.

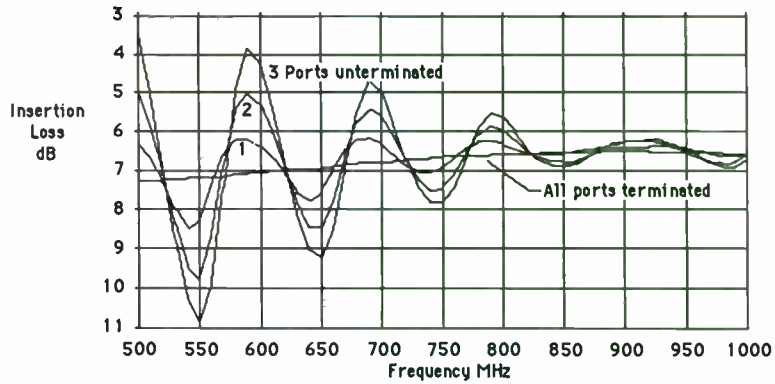


Figure 3. Measured insertion loss of four-way power divider showing combined effect of reflections at output ports and finite isolation.

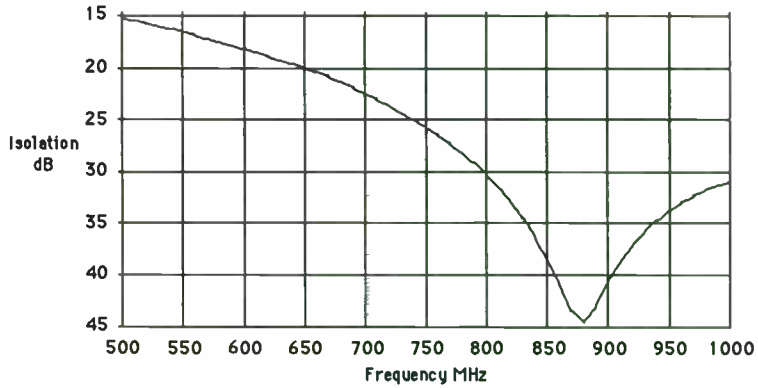


Figure 4. Measured isolation of the power divider used in the insertion loss experiment shown in Figure 3.

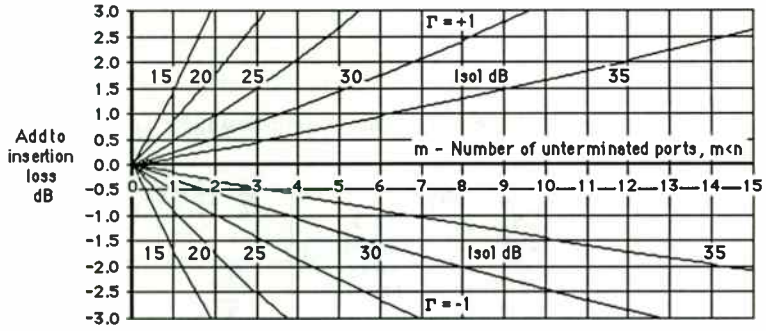


Figure 5. Change in insertion loss due to unterminated output ports and imperfect isolation.

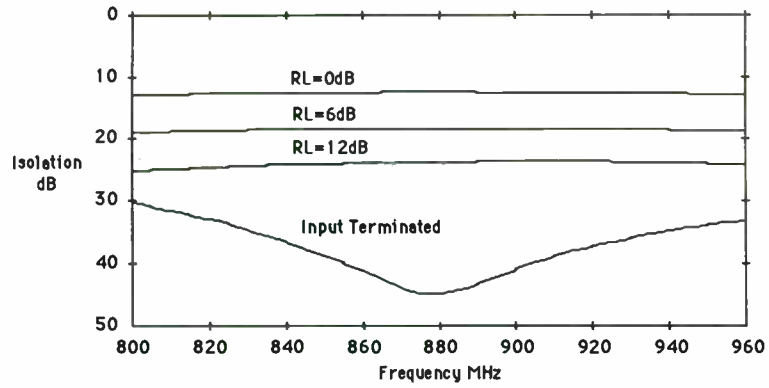


Figure 6. Isolation of four-way power divider with the input port terminated and with return losses of 0, 6 and 12 dB on the input.

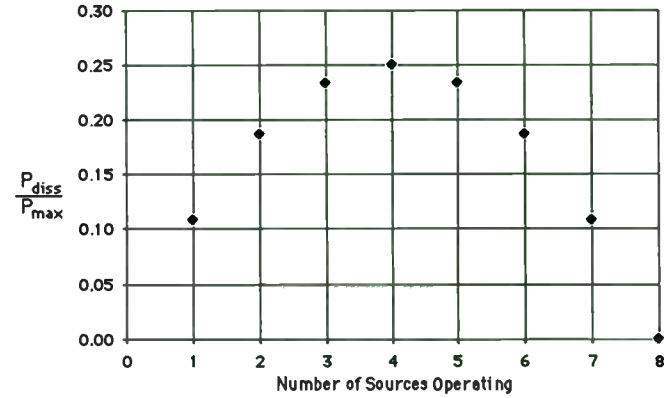


Figure 7. Power dissipated in an eight-way power combiner with some of the sources disabled. Sources are coherent.

System Approach to the Design of Automatic Gain Control

by

John G. Mohr
Engineer Scientist
Magnavox Electronics Systems Company
1313 Production Road
Fort Wayne, Indiana 46808

ABSTRACT

This paper is based on an original analysis by Victor and Brockman[1] entitled "The Application of Linear Servo Theory to the Design of AGC Loops.". Their analysis is expanded to include the effects of fixed gain within an automatic gain control (AGC) loop and the effects of group delay in components within the loop. The theory is then applied to an audio compressor circuit, a radio receiver circuit, and a transmitter automatic level control circuit.

INTRODUCTION

Historically, automatic gain control loops have often been designed without a strong theoretical basis. As a result, when problems such as instabilities are encountered, they are solved by "increasing the time constant" or some other technique based on "years of RF experience".

But even though the problem may be solved, the impact of the solution on the loop is often unknown. Further, it is still not known if the circuit will become unstable again as components vary from unit to unit or with temperature and age.

This analysis treats the AGC loop as a system wherein each element is assigned a transfer function. Then, using conventional servo analysis, the loop is designed to meet particular performance parameters with a full knowledge of the limiting factors.

The difficulty encountered with this analytical approach is that the gain control element usually has a transfer function that is linearly proportional to a dB/volt characteristic. This results in a transcendental loop equation that is not easily solved. However, by using the linearizing technique described by Victor and Brockman, the AGC loop can be modeled by a linear equation that is fairly accurate over a useful range of conditions. Once the loop has been modeled by a linear equation, it becomes relatively easy to design it to meet specific objectives and to perform stability and transient analyses.

[1] W.K. Victor and M.H. Brockman, "The Application of Linear Servo Theory to the Design of AGC Loops, "Proceedings of the IRE, Volume 48, pp 234-238, February 1960.

ANALYSIS

Figure 1 is a block diagram of a typical AGC loop. The function of the loop is to maintain the signal at the output of the voltage controlled attenuator at a constant level. Since the voltage controlled attenuator is followed by fixed gain, it follows that the level at the amplitude detector is also maintained at a constant level.

In Figure 1:

VCA = voltage controlled attenuator with numeric attenuation of a(r); gain = 1/a(r).

$r(t)$ = input signal with amplitude that varies with time.

r_{th} = the voltage output of the VCA which is to be maintained constant over all variations in r(t).

K_{IF} = fixed IF gain following the VCA.

K_A = attenuation characteristics of the VCA (linear in dB/volt).

K_D = amplitude detector gain.

V_R = reference voltage at which the amplitude detector output is to be maintained.

From this diagram:

$$\frac{r(t)}{a(r)} = \text{output of VCA} \quad (1)$$

$$\frac{r(t)}{a(r)} \cdot K_{IF} \cdot K_D = \text{output of amplitude detector} \quad (2)$$

$$\left\{ \frac{r(t)}{a(r)} \cdot K_{IF} \cdot K_D - V_R \right\} = \text{input to VCA Control} \quad (3)$$

$$K_A \cdot \left\{ \frac{r(t)}{a(r)} \cdot K_{IF} \cdot K_D - V_R \right\} = \text{VCA attenuation in dB} \quad (4)$$

Figure 2 is a modified block diagram to reflect the preceding equation.

The amount by which r(t)/a(r) differs from r_{th} is the error in the loop. The error expressed in dB is:

$$\text{Error} = [20 \log \{\frac{r(t)}{a(r)\, r_{th}}\}] \text{ dB} \qquad (5)$$

Since this is a unity feedback loop, the error is also equal to the output of the loop divided by the open loop gain (K_{OL}).

$$\text{Error} = [\frac{K_A\{\frac{r(t)}{a(r)} K_{IF} K_D - V_R\}}{K_{OL}}] \text{ dB} \qquad (6)$$

Equations 5 and 6 are both equal to the error and may be set equal to each other.

$$[20 \log \{\frac{r(t)}{a(r)\, r_{th}}\}] = \frac{K_A \{\frac{r(t)}{a(r)} K_{IF} K_D - V_R\}}{K_{OL}} \qquad (7)$$

The preceding equation is an exact model of an AGC loop. It is a nonlinear equation containing a logarithmic term, which makes it difficult to use. However, if the logarithmic term is expanded in a series approximation of the form:

$\log X = .434 \ln X = .434 [(X-1) - 1/2 (X-1)^2 + 1/3 (X-1)^3 + - - - -]$ for $O < X < 2$,

and in the approximation only the first term of the expansion is considered, the resulting equation is linear. The validity of excluding higher order terms can be shown. In the AGC loop it is desired that the ratio $\frac{r(t)}{a(r)\, r_{th.}}$ equal unity, i.e., $\frac{r(t)}{a(r)} = r_{th}$. For this condition, the first term of the expansion is an exact expression of ln X. It can be shown that if the ratio $\frac{r(t)}{a(r) r_{th}}$ is held within ± 30% of unity, the maximum error, considering only the first term of the expansion, is less than 20 percent.

When the approximation for log X is substituted in equation (7), the results are:

$$20 \left[.434 \left\{\frac{r(t)}{a(r)\, r_{th}} - 1\right\}\right] = \frac{K_A [\frac{r(t).}{a(r)} K_{IF} K_D - V_R]}{K_{OL}} \qquad (9)$$

$$8.7\, K_{OL} \left[\frac{r(t)}{a(r)\, r_{th}} - 1\right] = K_A \left[\frac{r(t)}{a(r)} K_{IF} K_D - V_R\right] \qquad (10)$$

By rearranging terms, the following equation results:

$$K_{OL} \left[\frac{r(t)}{a(r)\, r_{th}} - 1\right] = .115\, K_A V_R \left[\frac{r(t)}{a(r)} \frac{K_{IF} K_D}{V_R} - 1\right] \qquad (11)$$

From this equation it follows that:

$$K_{OL} = 0.115\, K_A V_R \text{ since } V_R = K_{IF} K_D r_{th} \qquad (12)$$

Equation (12) is the desired linear model for the open loop gain of an AGC loop with a logarithmic gain control element. Based on this equation, Figure 3 is a block diagram of the linearized AGC loop. Note that the loop filter has been separated from the other terms. It should be apparent that the loop filter need not be an integrator as shown, and that the designer has the freedom of selecting the loop filter type that best meets specified requirements.

The preceding results assume that all of the elements have broad frequency responses relative to the loop filter. If this is not the case in practice, these band limiting effects must also be considered in the model. For example, the gain control element (PIN diode attenuator or transconductance amplifier) is usually a wideband element, but circuit designers often decouple the control line. The effects of this decoupling must be included in the model. The ideal, of course, is to design the decoupling such that it only introduces minimal shift within the loop bandwidth. To avoid problems it is good practice to measure the transfer function (both amplitude and phase) of each element in the loop before actually performing closed loop tests.

Another often overlooked factor is that AGC loops often include elements with group delay, such as crystal filters. This delay can drastically effect AGC loop performance or at least certainly place a limiting constraint on the maximum AGC loop bandwidth. However, these effects can also be included in the model.

Generally, the amplitude response of an element with group delay is not a factor because its bandwidth is much wider than the AGC loop bandwidth. Therefore, only the group delay introduced by the element is a factor. The effects of group delay can be

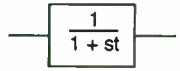
introduced into the model by a gain block with the following transfer function:

where t is the filter group delay. A simple method of including the first order effects of filter group delay in the loop analysis is to use the approximation:

$$e^{-st} = \frac{1}{1 + st + \frac{(st)^2}{2!} + \text{- - - - -}} \qquad (13)$$

$$\sim \frac{1}{1 + st} \qquad (14)$$

This leads to the following model for an element with group delay.



One advantage of modeling an AGC loop by the preceding methods is that it is helpful in identifying other problems. In one application, a loop that was being used for transmitter level control appeared to be unstable. After carefully measuring the transfer characteristic of each element in the loop and developing a linear model for the loop, it was concluded that the loop should have been stable. Based on this result, the circuit designer then proceeded to search elsewhere for the problem and discovered an unstable RF amplifier. The point is that because the loop had not been originally modeled, it was assumed to be the problem. Had the loop been originally designed based on this model, it is likely that the true problem would have been discovered in a more timely manner. The overall result is that many tedious hours of troubleshooting and selection of components by trial and error could have been avoided.

AUDIO COMPRESSOR CIRCUIT

The requirement was to design an audio compressor circuit compatible with the widely varying audio signals from a standard handset, such as the H-250, and provide a relatively constant output level to the following circuits.

The signal level from the handset varies over the range of 1 millivolt to 100 millivolts; thus, at least 40 dB of gain control was required. Fairly rapid response is required to allow the circuit to respond to changing voice patterns.

The circuit selected is shown in Figure 4. The gain control element is an integrated transconductance amplifier. The amplitude detector is a fullwave rectifier configured with an operational amplifier. The loop filter was chosen to be a perfect integrator to minimize error in the loop, thus maintaining a more constant output level. The reference level was set at the input to the loop filter by the resistor divider network.

The pertinent parameters for the loop design are:

30 dB/volt $\leq K_A \leq$ 100 dB/volt (measured).

V_R = .65 volts

Therefore, the nominal loop gain is:

K_{OL} = (.0115) (60) (.65) = 4.5

The model of the loop is shown in Figure 3. From this model, the open loop gain including the loop filter is $K_{OL}(s) = \frac{4.5}{sRC}$. Since the open loop gain expression has only a single pole at the origin, the closed loop will be unconditionally stable, if there are no other band-limiting elements in the loop, which was the case for this example.

The closed loop gain is given by the following equation, $\frac{C}{R}(s) = \frac{\frac{4.5}{RC}}{S+4.5/RC}$.

From this equation, it is evident that the closed loop 3dB bandwidth is $f = 4.5/(2\pi RC)$. Thus, R and C may now be selected to give the desired closed loop response.

Since the loop was not to track the audio signals from the handset (roughly 300 Hz to 3000 Hz) and only respond to the average value, it was necessary to set the loop bandwidth to less than 100 Hz. Because the gain control characteristic varies over a fairly wide range, the loop corner was set at 20 Hz for the nominal gain. The measured performance of the circuit is shown in Table 1.

From the test data it can be seen that the loop provided a fairly constant output over the specified input range, had very low distortion, and had the desired frequency response. The attack time was roughly 10 milliseconds which was adequate.

Since this circuit was designed, audio compressor circuits have become available in single integrated circuits, and the design has become much simpler. However, the analysis presented in this paper is still applicable, and it can be used to gain a good understanding of the performance of even the single integrated circuit audio compressors.

TABLE 1

PERFORMANCE OF AUDIO COMPRESSOR CIRCUIT

COMPRESSOR CHARACTERISTIC

Input (at 1 KHz)	Audio Output
-76 dBm	-7.8 dBo
-66 dBm	0
-56 dBm	Reference = 0 dBo
-46 dBm	0
-36 dBm	0
-26 dBm	0
-16 dBm	+.3 dBo
-6 dBm	+.6 dBo

OUTPUT HARMONIC DISTORTION

Input Frequency	For -56 dBm in	For -36 dBm in
300 Hz	1.8%	7.2%
1000 Hz	1.1%	1.05%
3000 Hz	1.05%	0.7%

FREQUENCY RESPONSE (-56 DBM IN)

Output Level	Frequency
-6.0 dBm low	17.5 Hz
-0.5 dBm low	35 Hz
-0.5 dB high	24.2 KHz
-6.0 dB high	36.8 KHz

RADIO RECEIVER CIRCUIT

This example is based on a receiver designed to receive satellite signals. The basic requirement was to control the receiver gain for input signal variations from -150 dBm to -90 dBm with a maximum rate of change (due to signal fading) of 20 dB/second. The AGC loop's function was to control the receiver gain such that a relatively fixed signal level was supplied to the receiver demodulator.

Two conflicting requirements made performance trade-off considerations necessary. The AGC loop bandwidth had to be narrow to maintain an AGC control that was relatively noise free for the very low signal levels, but the loop bandwidth also had to be wide enough to track the 20 dB/second signal fades. The approach was to design for low noise operation as a first consideration and then evaluate the dynamic response to signal fades. If necessary, design trade-offs could then be made based on the relative priority of the conflicting requirements.

To assure good low signal level performance, the bandwidth of the loop was designed to achieve a +10dB signal-to-noise ratio at the minimum input signal level. Since the receiver noise figure was approximately 5 dB, a double-sided loop noise power bandwidth of 8 Hz was required to achieve a 10 dB signal-to-noise ratio for a -150 dBm input.

For an AGC loop with an integrator loop filter and an open loop gain characteristic (See Figure 3) of $K_{OL} = \frac{.115\ K_A V_R}{sCR}$ it can be shown that the loop double-sided noise bandwidth is related to the open loop gain characteristic by the following equation:

$$\text{NPBW (double-sided)} = \frac{.115\ K_A\ V_R}{2RC}\ \text{Hz}^{\ 2} \qquad (15)$$

Thus, for a given noise bandwidth, and known K_A and V_R, the elements of the loop filter can be calculated.

The receiver gain control characteristic was measured to be approximately 60 dB/volt with a frequency response that was greater than 3 KHz including decoupling of the AGC control. The loop reference (V_R) was set to be .2 volts to provide the desired signal level to the receiver demodulator. From equation 15 and with the following parameters:

K_A = 60 dB/volt

V_R = .2

[2]F.M. Gardener, "Phaselock Techniques," 2nd Edition, John Wiley and Sons, New York, 1979, P.31.

NPBW (2 sided) = 8 Hz,

the loop time constant is determined to be

$RC = .115\ K_A\ V_R/(2 \cdot NPBW\ (2\ sided))$

$= (.115)(60)(.2)/(2)(8)$

$= .086$ seconds

To determine if an 8 Hz double-sided noise bandwidth is sufficient to track signal fading rates of 20 dB/second, the final value theorem of Laplace transform analysis is used to evaluate the steady state error in the loop due to a ramp input.

For the model shown in Figure 3, the error in the loop is given by the following equation:

$$\text{Error} = R(s)\left[\frac{SCR}{SCR + .115\ K_A V_R}\right] \qquad (16)$$

where R(s) is the transform of the loop input. For a ramp input $R(s) = \frac{V}{S^2}$ where V is the slope of the ramp. Therefore, the error resulting from a ramp input is:

$$\text{Error} = \frac{V}{S^2}\left[\frac{SCR}{SCR + .115 K_A V_R}\right].$$

The final value theorem states that the steady state error can be calculated by evaluating the following equation:

$$\text{Error (steady state)} = \lim_{S \to 0}\left\{S \cdot \frac{V}{S^2}\left[\frac{SCR}{SCR + .115\ K_A V_R}\right]\right\}$$

$$= \frac{VCR}{.115\ K_A V_R}$$

For the following parameters associated with the receiver AGC loop:

$V = 20$ dB/sec

$CR = .086$ seconds

$K_A = 60$ dB/volt

$V_R = .2$ volts

the steady state error due to a 20 dB/second rate of change input is 1.25 dB. This was an acceptable error in amplitude at the detector and no trade-offs in performance were necessary.

TRANSMITTER LEVEL CONTROL CIRCUIT

Often AGC loops are used for automatic leveling of high power transmitters. Figure 5 is an example of such a system. The modulator/carrier control circuitry is a feedback loop that controls the cascaded gain of the power amplifiers. In the case of amplitude modulated transmitters, as shown, the modulating signal can be input directly to the loop. The feedback action of the loop will cause the RF envelope of the RF output of the transmitter to follow the modulation input. The net effect is that the transmitter is linearized by the action of the AGC or automatic level control (ALC) loop.

This particular example is of an ALC loop that was used to linearize a 100 watt solid state transmitter. The problem was to make the loop as wide as possible because the modulation input was multiple tones over a wide bandwidth. Without the linearizing action of the ALC loop, the multiple tone modulation envelope would be distorted due to intermodulation in the high power stages, which are inherently non-linear.

To achieve the widest bandwidth possible each gain block in the loop was analyzed. Basically, the reflectometer with its associated amplitude detector was wideband. The detector bandwidth was in excess of 10 MHz. The PIN diode attenuator frequency response was measured by applying a low level audio signal to the control line and increasing its frequency until the response as observed on the output RF envelope decreased by approximately 3 dB. The result was a measured bandwidth of approximately 2.5 MHz.

The loop filter design was of the configuration shown in Figure 6. The operational amplifier was a high-speed, wideband design. The DC gain of the amplifier was 30. Thus, for a voltage controlled attenuator characteristic of 40 dB/volt, a voltage change of approximately 1.15 volt at the loop filter output provides the roughly 46 dB of attenuation range required. The carrier reference voltage was 0.15 volts.

With these parameters, it would seem that the loop bandwidth could have been made in excess of 1 MHz. However, in practice, loop stability could only be achieved for bandwidths of less than 500 KHz. The limiting factor was determined to be group delay in the transmitter power

amplifier.

Even though the transmitter was broadband, the signal was delayed by virtue of the number of stages, the power splitters and combiners. The measured delay was approximately 230 nanoseconds. Based on the model of equation 14, this delay results in a phase shift of approximately 45^{o} at a frequency of 590 KHz. Thus, to assure loop stability the capacitor in the loop filter must be selected to reduce the gain to less than 1 at a frequency of less than 590 KHz. This results in a total open loop phase shift of 135^{o} (90^{o} from the loop filter and 45^{o} due to the group delay) at the unity gain frequency.

For the given parameters the open loop gain is calculated to be:

$K_{OL} = .115\ K_A\ K_{DC}\ V_R = .115 \times 40 \times 30 \times 0.15$

$= 20.7$

$20 \log K_{OL} = 26.3$ dB

With the knowledge that the capacitor in the loop filter will introduce a gain roll off of 20 dB/decade (6 dB/octave), the RC time constant of the loop filter was selected to reduce the gain to unity (0 dB) at a frequency of 500 KHz. Thus, the 3 dB frequency corner of the loop filter was set at approximately 25 KHz. The use of Bode plots is often helpful in the selection of the loop corner frequency for designs such as this.

The resulting loop was stable and had an ALC loop bandwidth of approximately 500 KHz. This was the widest bandwidth possible based on the limitations of the circuitry.

CONCLUSION

The theory presented shows that AGC loops can be modeled by a linear equation based on known loop parameters, the voltage attenuator characteristic and the loop reference. Based on this model and using relatively simple analysis, it was shown that AGC loops can be designed to achieve a number of desired performance characteristics. The important thing is that the designer has a grasp of what elements in the loop will impact stability and performance.

Three widely divergent examples were shown to demonstrate the use of the model.

The examples used relatively simple analytical techniques, but the model can be used for much more detailed analysis of loops with more complex filters. For example, the receiver AGC loop was modeled on a computer to compare the dynamic response with different types of loop filters and different rates of change on the input signal.

In all respects, the model has been a powerful tool for analyzing a wide range of problems. It has been especially helpful in achieving desired loop characteristics and in separating loop problems from circuit problems.

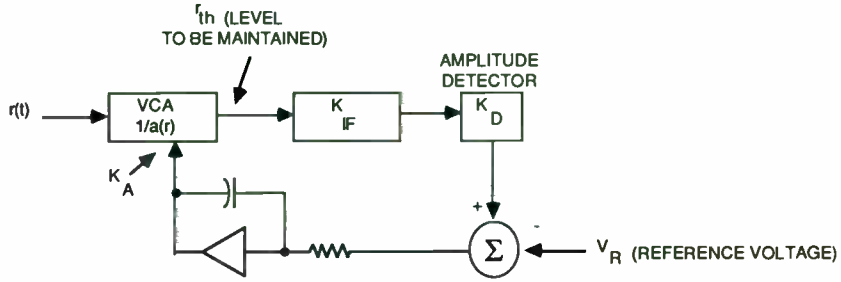


FIGURE 1. AUTOMATIC GAIN CONTROL LOOP BLOCK DIAGRAM

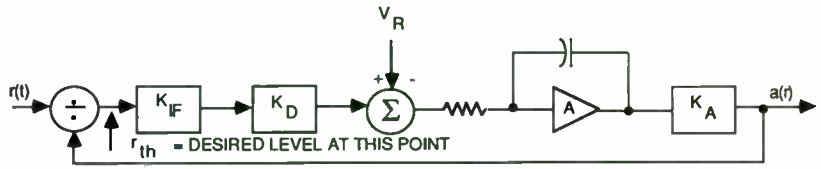


FIGURE 2. MODIFIED BLOCK DIAGRAM OF THE AGC LOOP

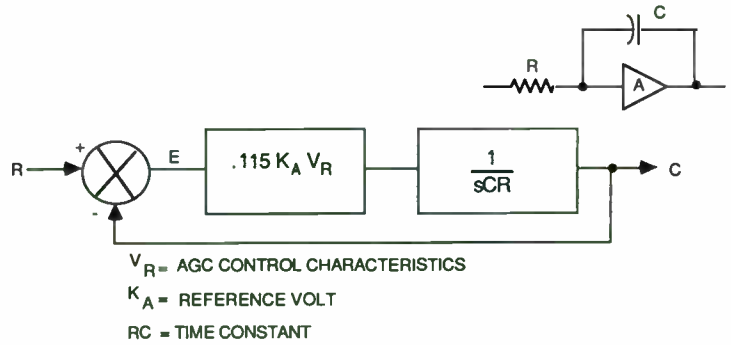


FIGURE 3. ANALYTICAL MODEL OF AGC LOOP

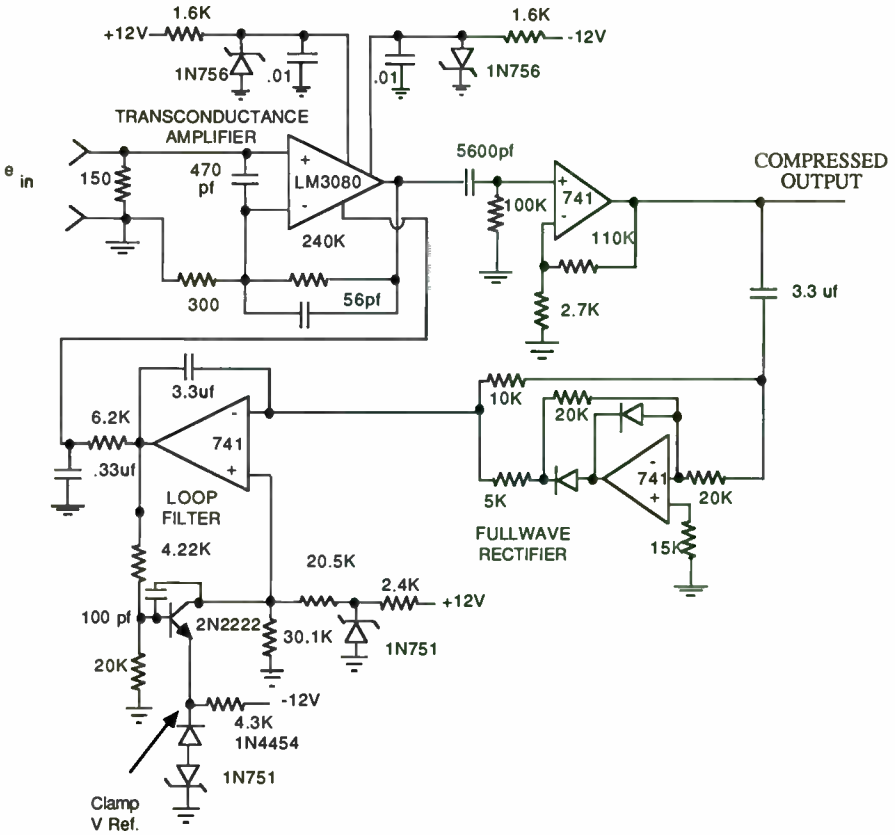


Figure 4. Audio Compression Circuit Schematic Diagram

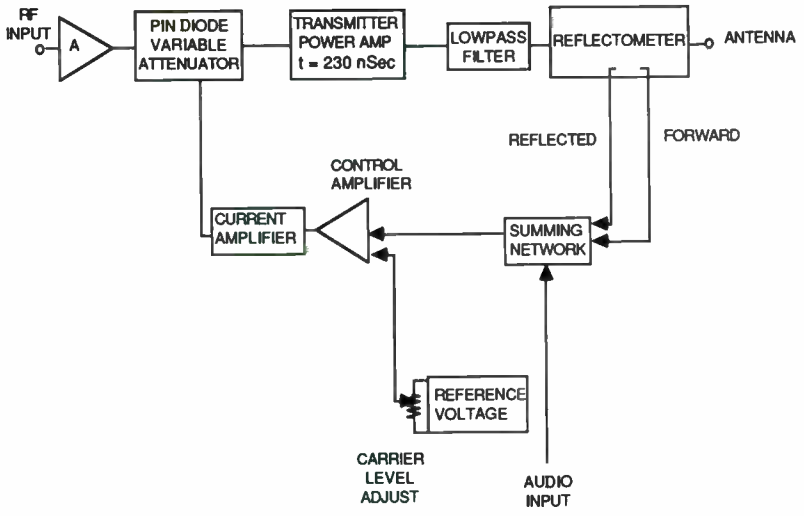
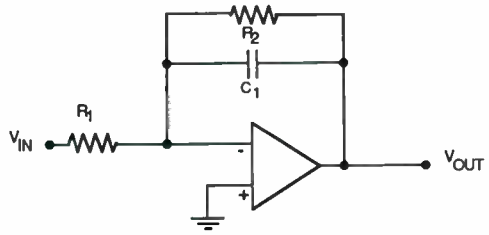


FIGURE 5 . MODULATOR/CARRIER CONTROL CIRCUITRY



$$\frac{V_{OUT}}{V_{IN}} = \frac{R_2}{R_1}\left(\frac{1}{1 + SC_1R_2}\right)$$

FIGURE 6. ALC LOOP FILTER

Design and Test Standardization for Cost Effeotive Hybrid Integrated Assemblies

by

Stephan Van Fleteren
Senior Member Technioal Staff
FEI Microwave Ino.
825 Stewart Drive
Sunnyvale, CA 94086

Abstraot

An E/J band dual ohannel miorowave integrated reoeiver was developed using a modular fixturing test system which simplified and shortened the breadboarding prooess. Good oorrelation was aohieved between the performanoe of the breadboard and the final engineering model, whioh inoluded stringent speoifioations for ohannel to ohannel amplitude and phase matohing.

The Modular Test System

The purpose of the modular test system is to simplify and speed the development of miorowave integrated oomponents by reduoing the effort spent on basio meohanioal aspeots of the design. The system allows the designer to oonoentrate on the more significant eleotrioal requirements and at the same time affords the maximum meohanioal versatility and flexibility. Traditionally, component development follows a process flow path as shown in Figure 1. The engineer has had to design speoialized oarriers, substrates, and test fixtures for eaoh oomponent and then design oustom system test fixtures to oombine the component funotions. If something did not work, or a oomponent needed to be added to meet system speoifioations, the test fixtures may have had to be oompletely redesigned.

Implementing the modular test system required the standardization of oarrier size, interoonnection looations, and oomponent/subsystem test fixtures. The result was the streamlining of the component development process (Figure 2). The engineer now designs and manufaotures the required substrate, assembles and tests his oirouit, with pre-stocked standard carriers and test fixture hardware.

The basis of the modular test system is a standardized geometry of equally-spaoed (.120") perpendioular grid lines. Lying on the interseotions of the grid lines are the oarrier mounting holes (meohanioal) with the miorostrip interoonneotions lying along the grid lines (eleotrioal) (Figure 3). Notioe the module edges lie between the grid lines. The oarriers are dimensioned so when the mounting holes of two adjaoent modules are lined up on the intersections of the grid lines, there is no meohanioal interference between the oarriers.

Designing a MIC module now beoomes a matter of following a few simple layout design rules. The designer is less oonoerned with dimensions, toleranoes and interoonneotions beoause all dimensioning has been pre-referenoed to the grid lines. After ohoosing a standard modular oarrier from the available stooked sizes (Figure 4) the layout is oompleted oonforming to the rules

shown in Figure 5. Assembly follows using standard techniques with substrates being positioned onto the carriers by means of an alignment frame (Figure 6).

Testing is performed using one "universal" test fixture which allows the designer to test the individual modules, the entire system, or any part of the system. In addition, the effects of an enclosure can be evaluated using standard walls and covers without building a housing. The hardware requirements are listed below and shown in Figure 7.

1. Base plates (1"x1", 2"x2", 2"x4")
2. Coax to microstrip transitions
3. Walls, with and without DC feedthroughs
4. Covers
5. Breadboard Substrates
6. Miscellaneous 0-80 screws

The most critical component the modular test system is the coax to microstrip transition. Great care has been taken in its design because its performance greatly affects component and system measurement accuracy. The transition is made-up of four parts, including the base, compression cover, cable assembly, and 50Ω microstrip line (Figure 8). The base and compression cover hold the cable rigid and ensure ground contact at the transition point. The cable assembly itself consists of a .047 inch coax cable assembled to a two hole flange female SMA connector. Time domain reflectometry measurements were made on the HP-8510 network analyzer in order to compensate for the reactance of the coax to microstrip transition (Figure 9). After tuning two transition assemblies they were placed back to back and measured for insertion loss and return loss. The results show less than 1.2 dB insertion loss and greater than 13dB return loss (1.58:1) to 26.5GHz for the two transitions. This corresponds to a 19 dB return loss (1.26:1) for a single transition (Figure 10).

The Integrated Front End

With the modular test system in place for breadboard development the design of the integrated front end could begin. The receiver developed is shown in block diagram form in Figure 11 along with some of the major specifications. Certain design goals were agreed to at the beginning in order to strongly focus attention on lowering the unit cost while maintaining good electrical performance. Some of the major design rules are listed below.

1. Minimize Parts:
   (a) Create multifunctioned substrates.
   (b) Use MMIC's and custom IC's wherever possible.
   (c) Reduce mechanical parts.
2. Minimize processes and process steps.
   (a) Wire and ribbon bonding.
   (b) Conductive epoxy.
   (c) Soldering.
   (d) Laser welding.
3. Minimize microwave discontinunities and the distances between discontiniuties.

(a) Single carrier

(b) Single assembly connector/glass feedthrough/.047" cable.

4. Minimize test time.

(a) Tight interface and component specifications-no tuning.

(b) HP-8510 and NF automatic tests.

Working within the design rules and customer specifications the breadboard shown in Figure 13 was developed and tested. Some of the electrical design highlights are listed below.

1. All beamlead switch design.
2. 30dB ± 1dB attenuator design.
3. Low VSWR (<2:1 all ports) mixer design utilizing lossy microstrip IF LPF for spurious rejection.
4. Broadband, high isolation (>18 dB), seven section Wilkinson power divider.
5. Low cost custom integrated driver chip replacing a 35 part discrete version. (Figure 12).
6. 3 month start to finish breadboard development time.

One of the best examples of how the modular test system saves time was when it was discovered a customer specification was not verified during breadboard development. The specification required 40dB of L to I isolation which meant adding a LPF after the IF MMIC amplifier. Using the modular test system all the engineer had to do is choose an appropriate modular carrier from stock, design and fabricate the circuit following the modular test system rules, mount it on a test fixture and evaluate. At the system level the breadboard (Figure 13) output transitions were moved to allow space for the required lowpass filter, the system performance was then tested and passed the isolation requirement. The same process took place when the mixer and power divider were lengthened and the driver circuit was changed.

With the successful completion of the breadboard the individual breadboard substrate designs were mounted on a single .120" thick kovar carrier as shown in Figure 14. This RF assembly was then mounted in an aluminum housing (whose outline was customer specified) as shown in Figure 15 for final RF tests. The test results show excellent correlation between the breadboard results and the engineering model results. Figures 16 through 18 compare the performance of both assemblies.

Conclusion

The modular test system allows the electrical engineer to bypass costly and time consuming mechanical considerations in order to quickly and easily evaluate different components and system configurations during breadboard development. The integrated front end described in this paper benefited from the modular test system. The result is a clean, simple low cost design in 8 months because the design team concentrated on the electrical design, improving electrical performance, while minimizing the number of parts and processes (Figure 19).

Reference

"Standardization Modules Speed Hybrid Layout and Testing", R.K. Kirschman, Microwaves and RF, June 1983, pp. 76-80.

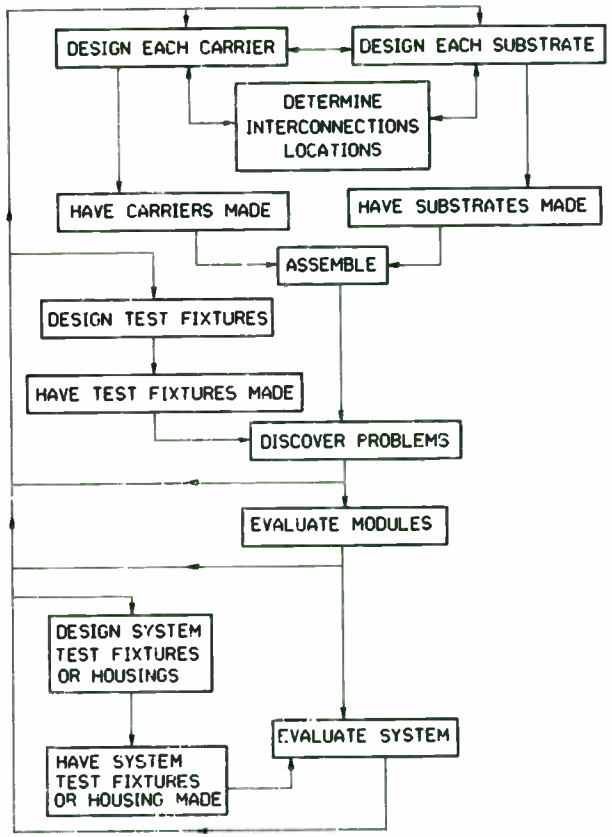


FIGURE 1: Traditional component/system development process.

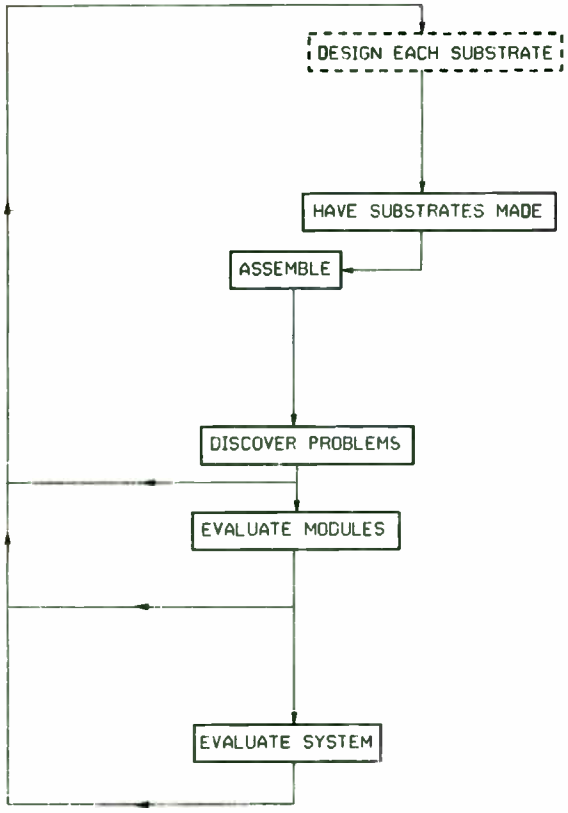


FIGURE 2: Modular test system component/system development process.

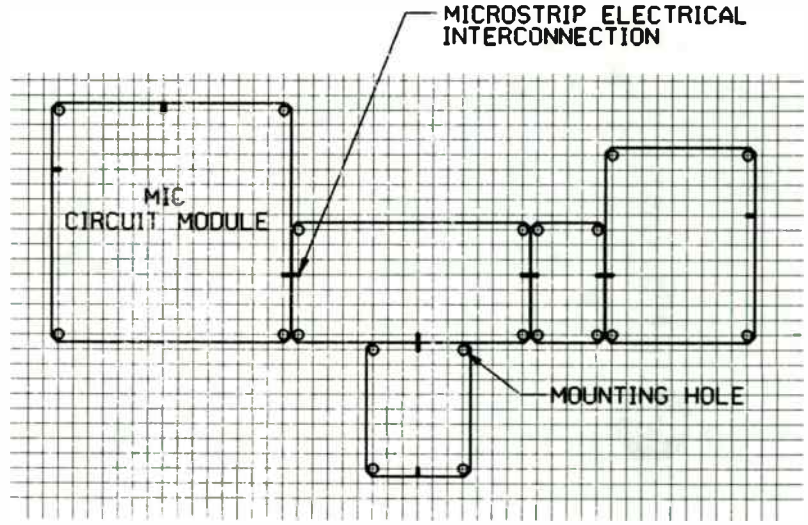


FIGURE 3: The modular test system places mounting holes on grid intersections and electrical interconnections along grid lines. Carrier spacing is .004" nominal.

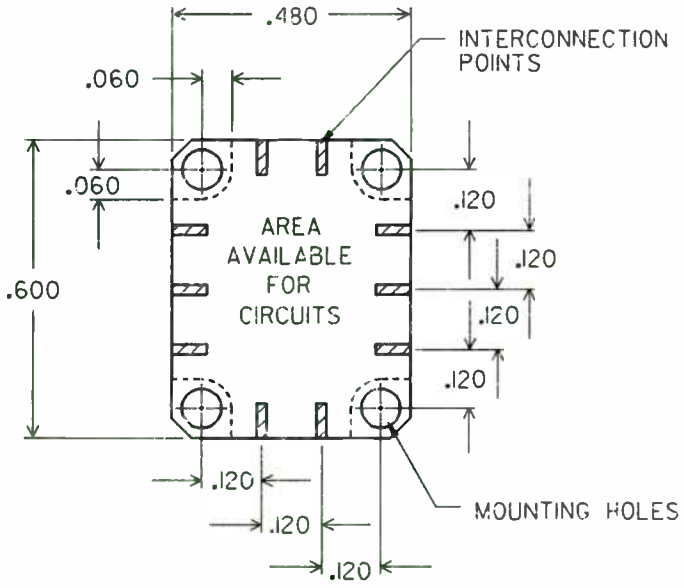


FIGURE 5: The layout of a typical module shows the possible interconnection locations and substrate limits. The broken line at each corner indicates the area reserved for mounting screws.

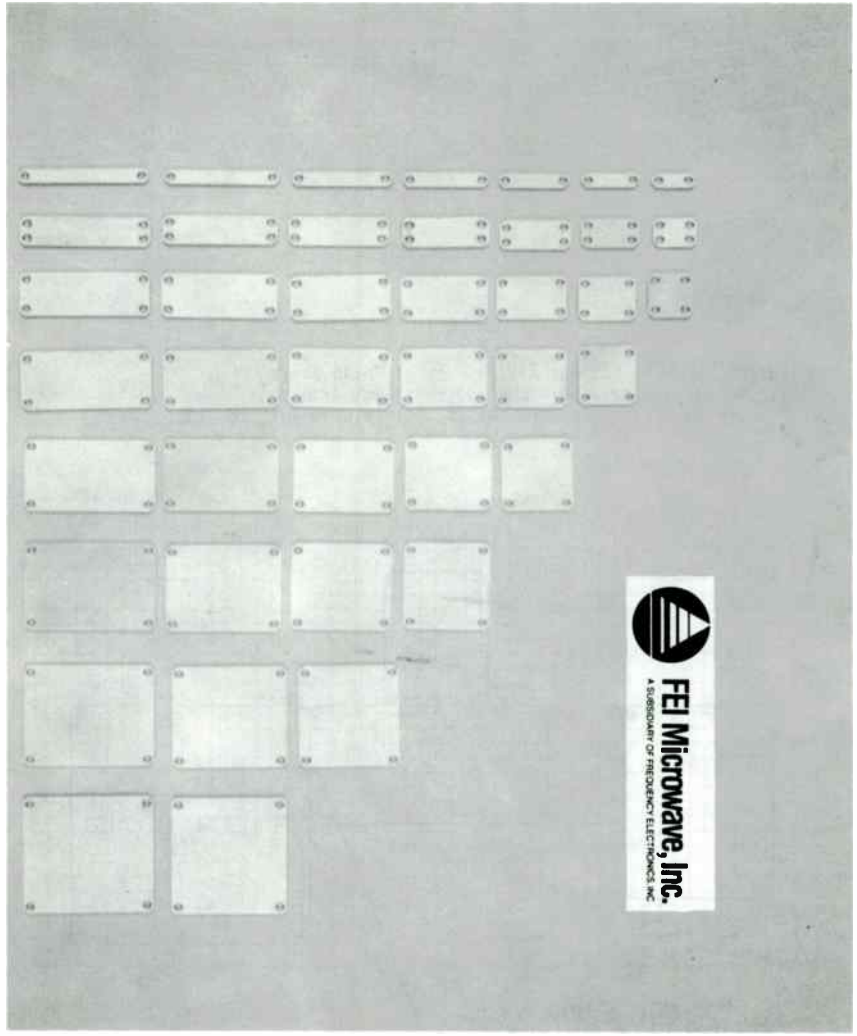


FIGURE 4: Standardized stocked carriers for modular test system.

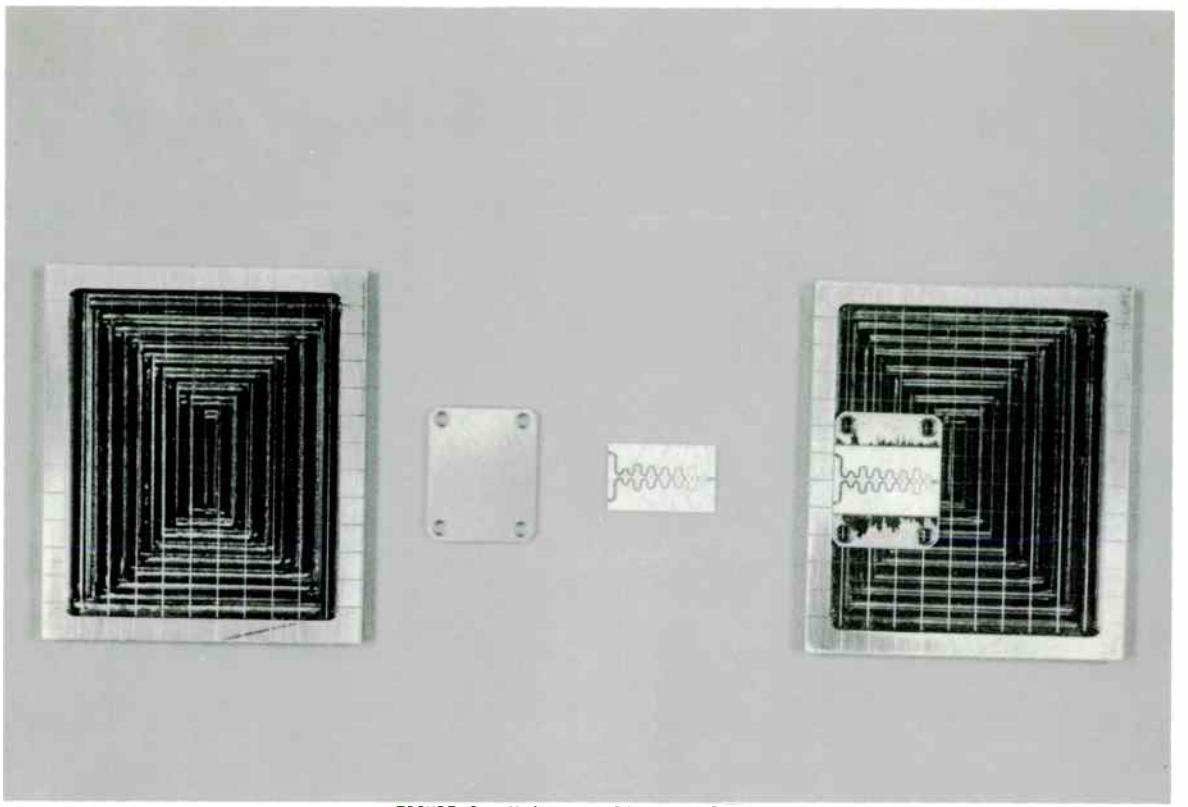


FIGURE 6: Using an alignment frame ensures accurate positioning of interconnection points to gridlines during substrate assembly.

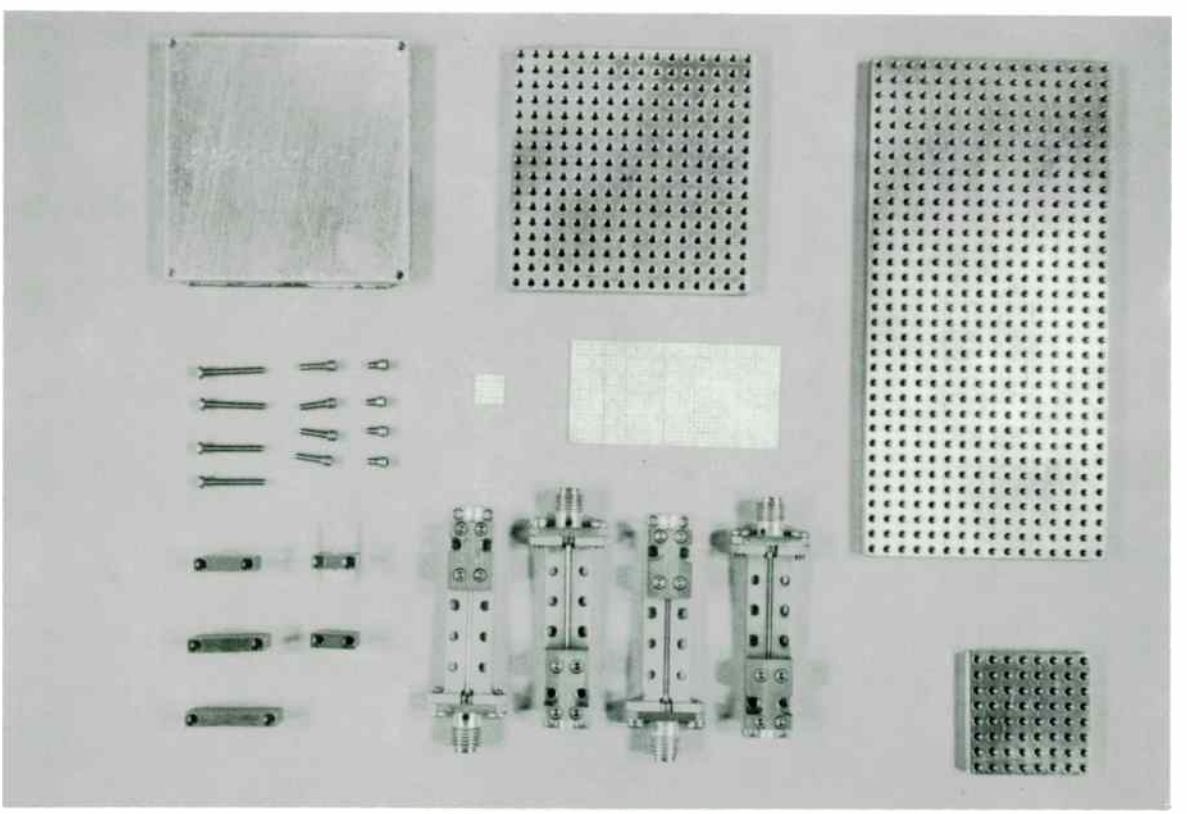


FIGURE 7: These component parts of the modular test system allow the engineer to RF test any device or subsystem to 26.5GHz.

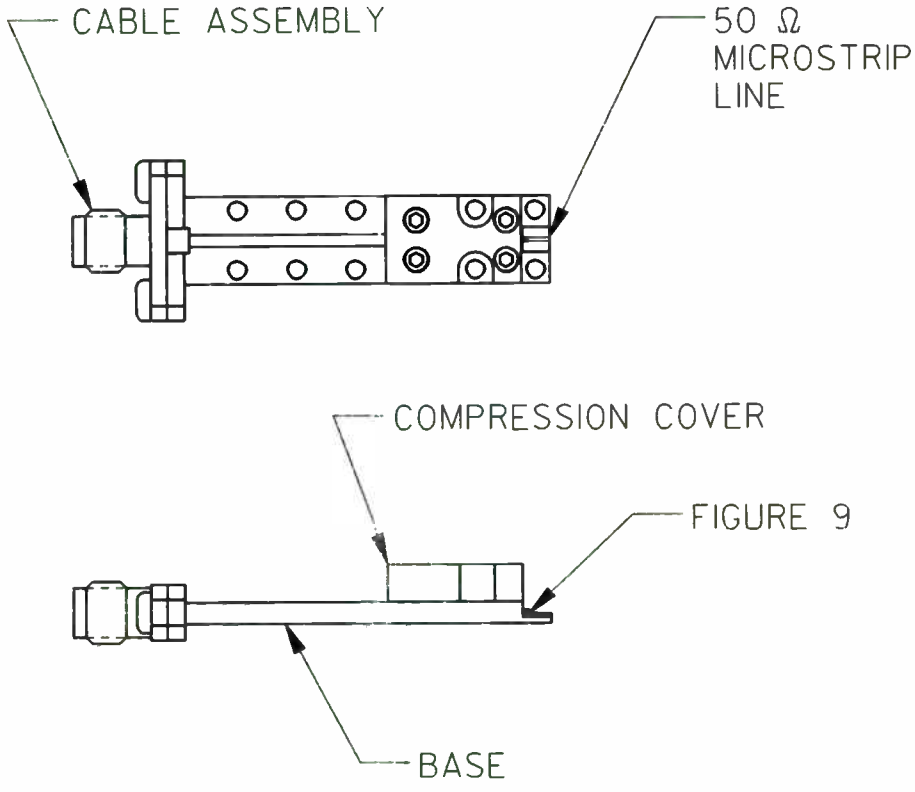


FIGURE 8: Coax to microstrip transition for the modular test system.

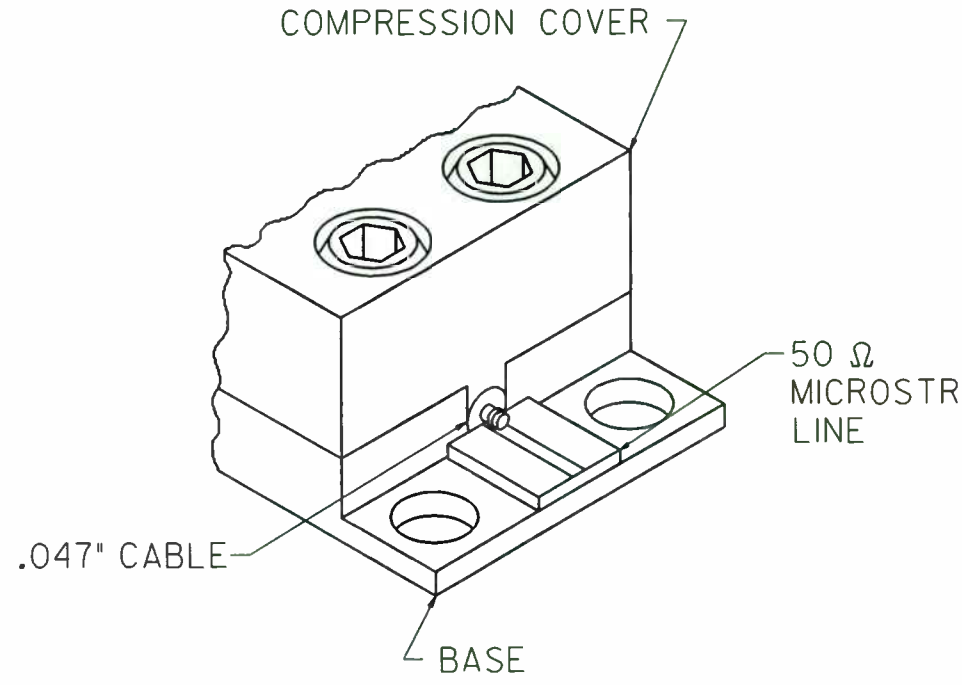


FIGURE 9: Detail of the coax to microstrip transition.

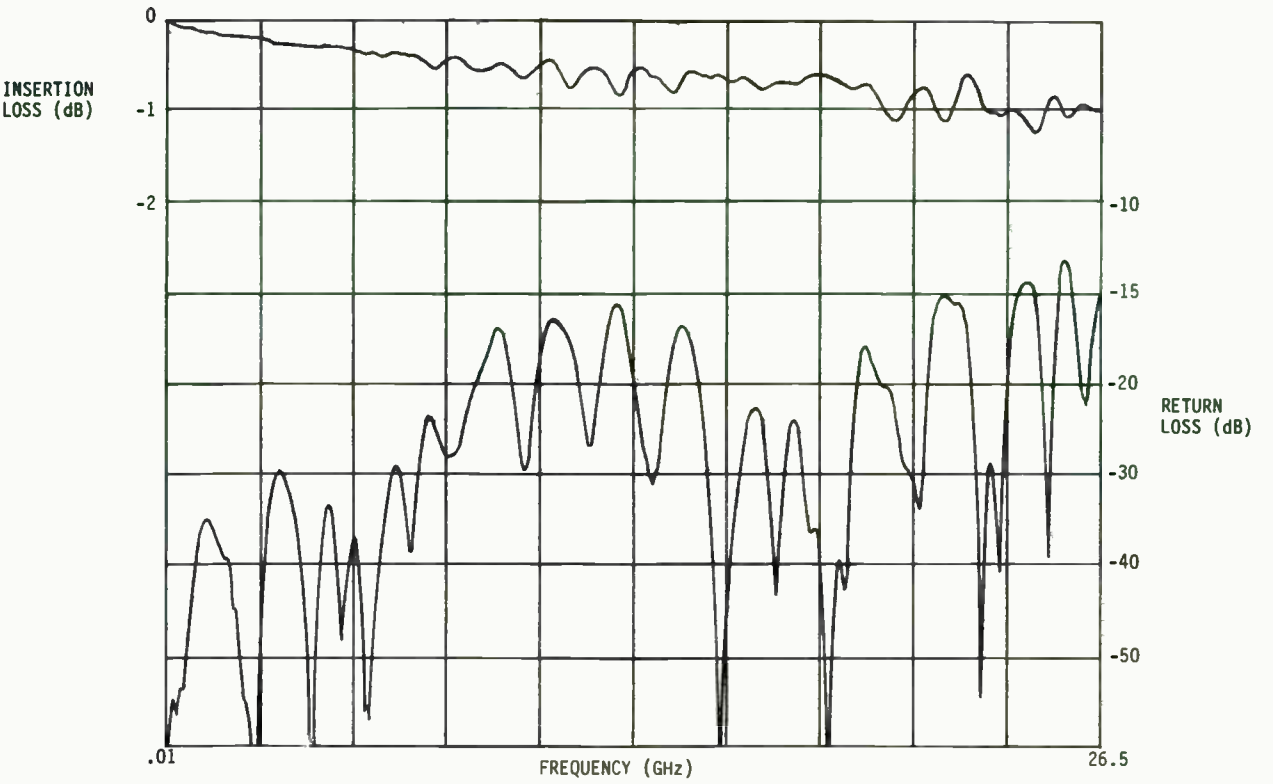


FIGURE 10: Insertion loss and return loss for two modular test system coax to microstrip transitions.

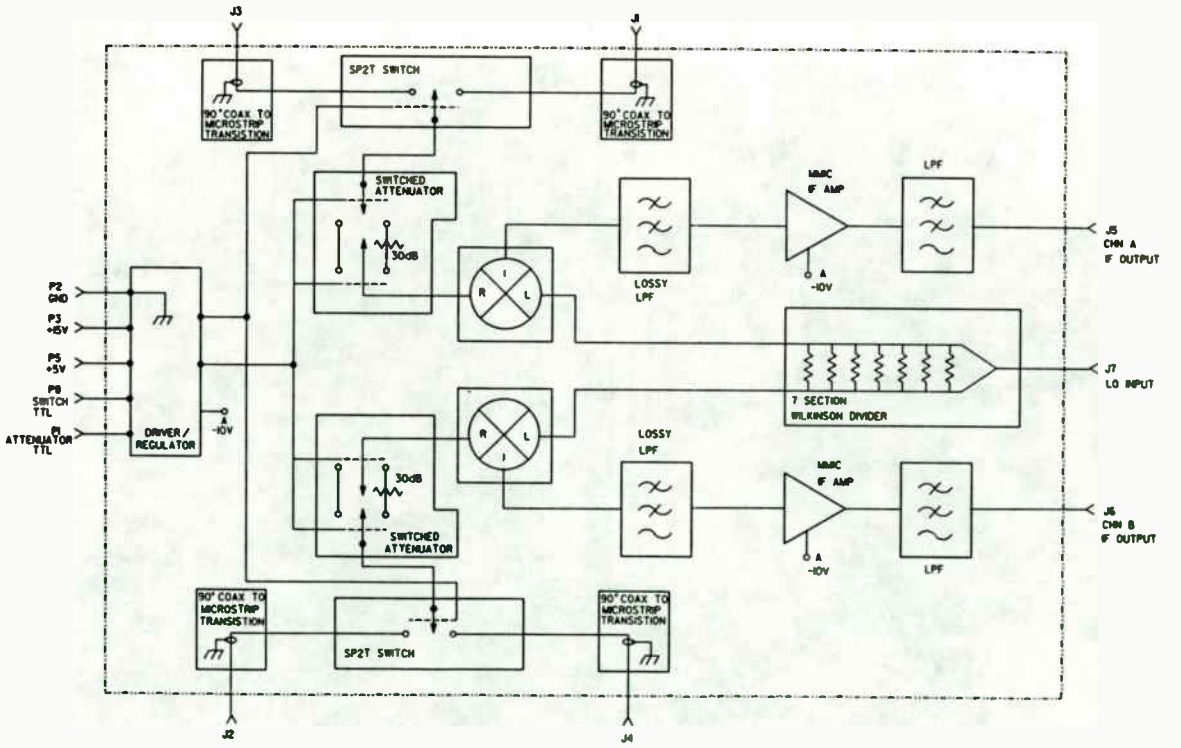


FIGURE 11: Dual Channel Integrated Front End Receiver.

Frequency range	:	E/J Bands
Conversion loss	:	4.5dB
Noise figure	:	18dB
Channel isolation	:	30dB
Amplitude match	:	±1.3dB
Phase match	:	±40°
Switching speed	:	35nS

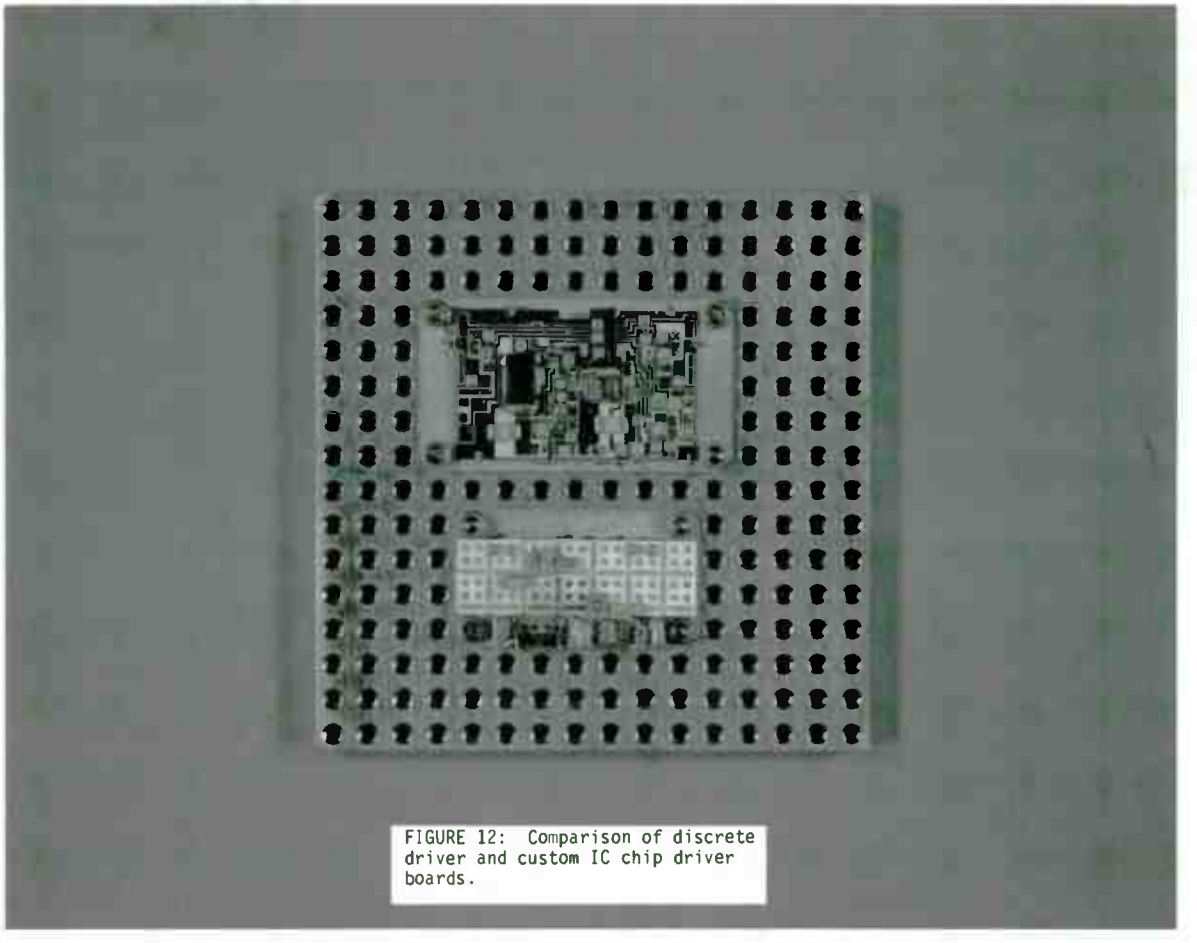


FIGURE 12: Comparison of discrete driver and custom IC chip driver boards.

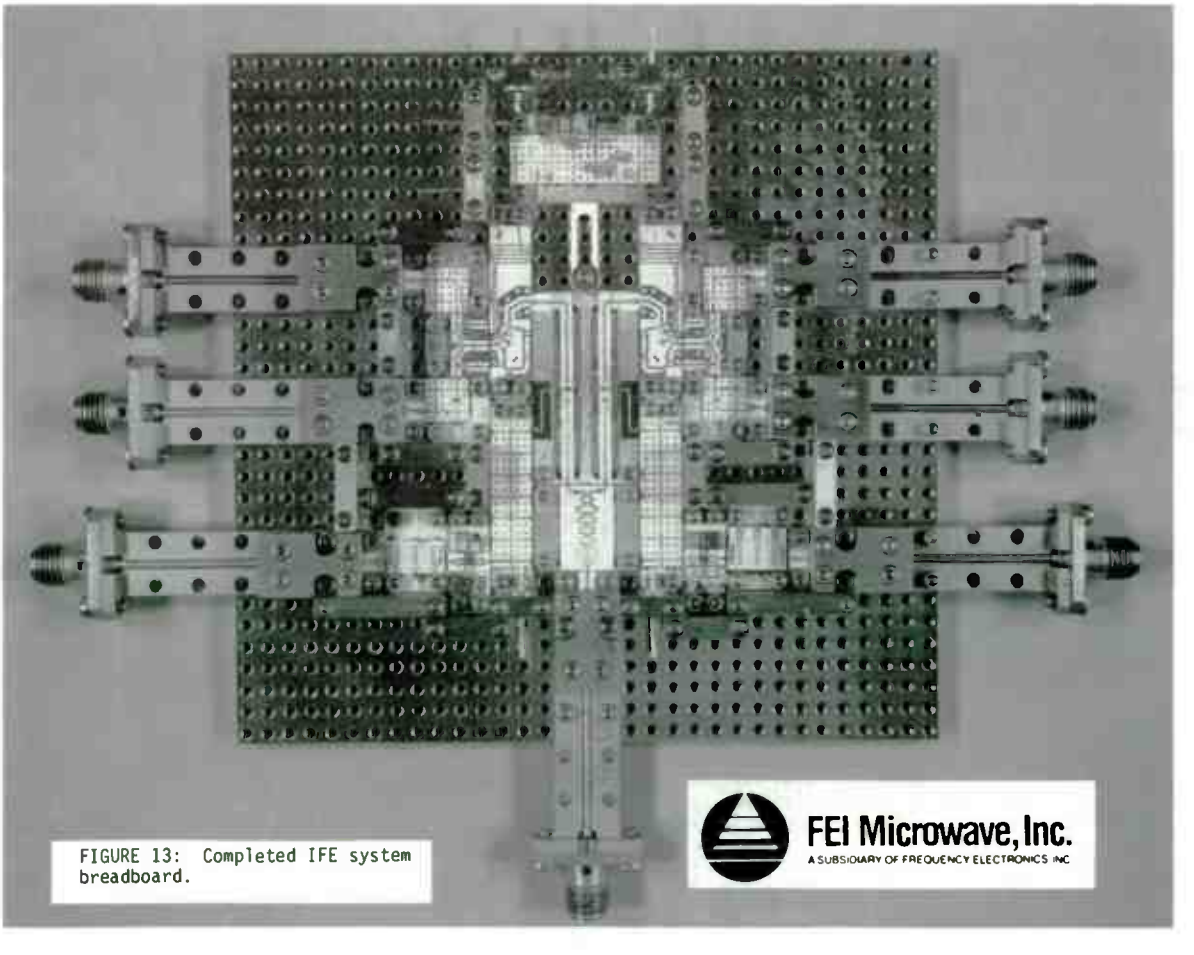


FIGURE 13: Completed IFE system breadboard.

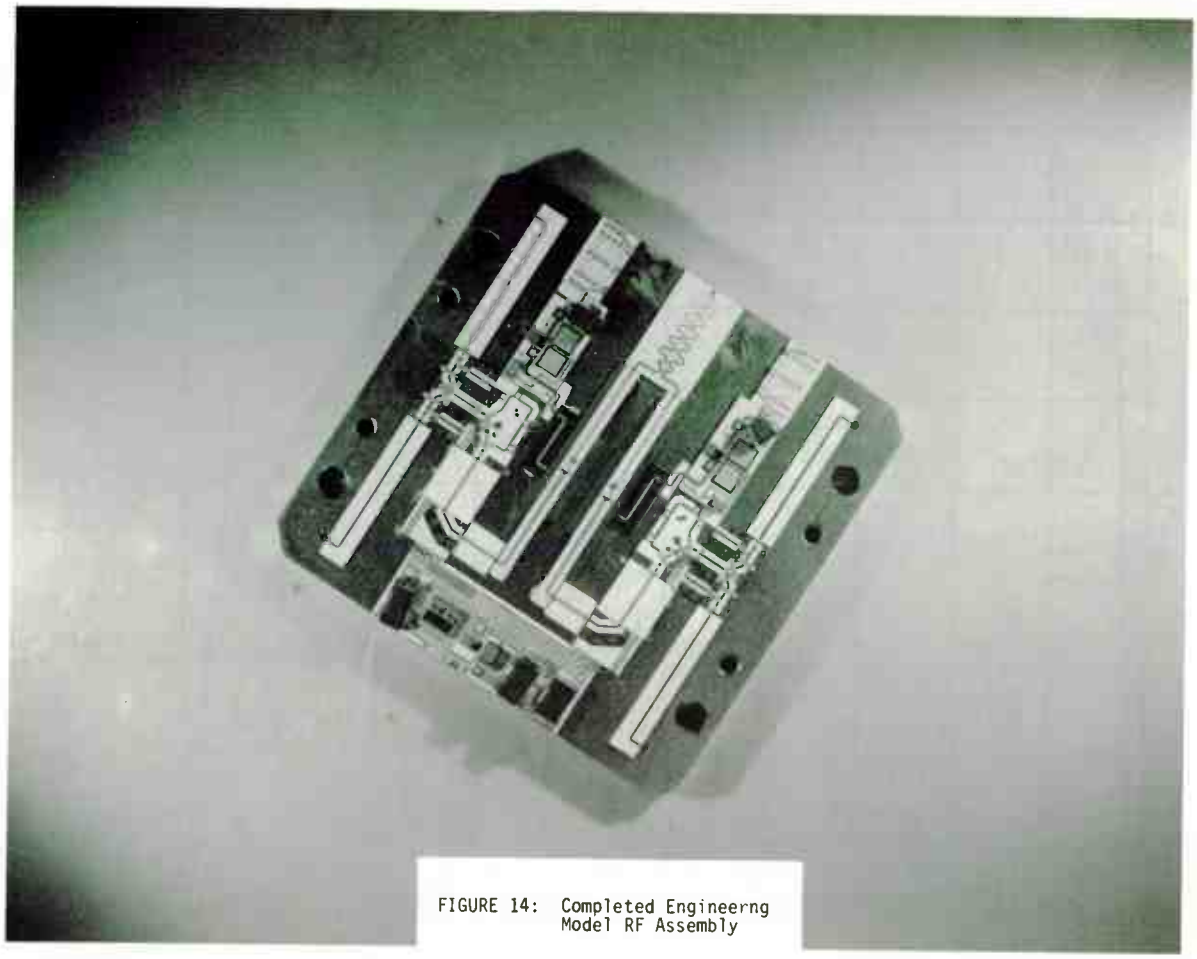


FIGURE 14: Completed Engineerng Model RF Assembly

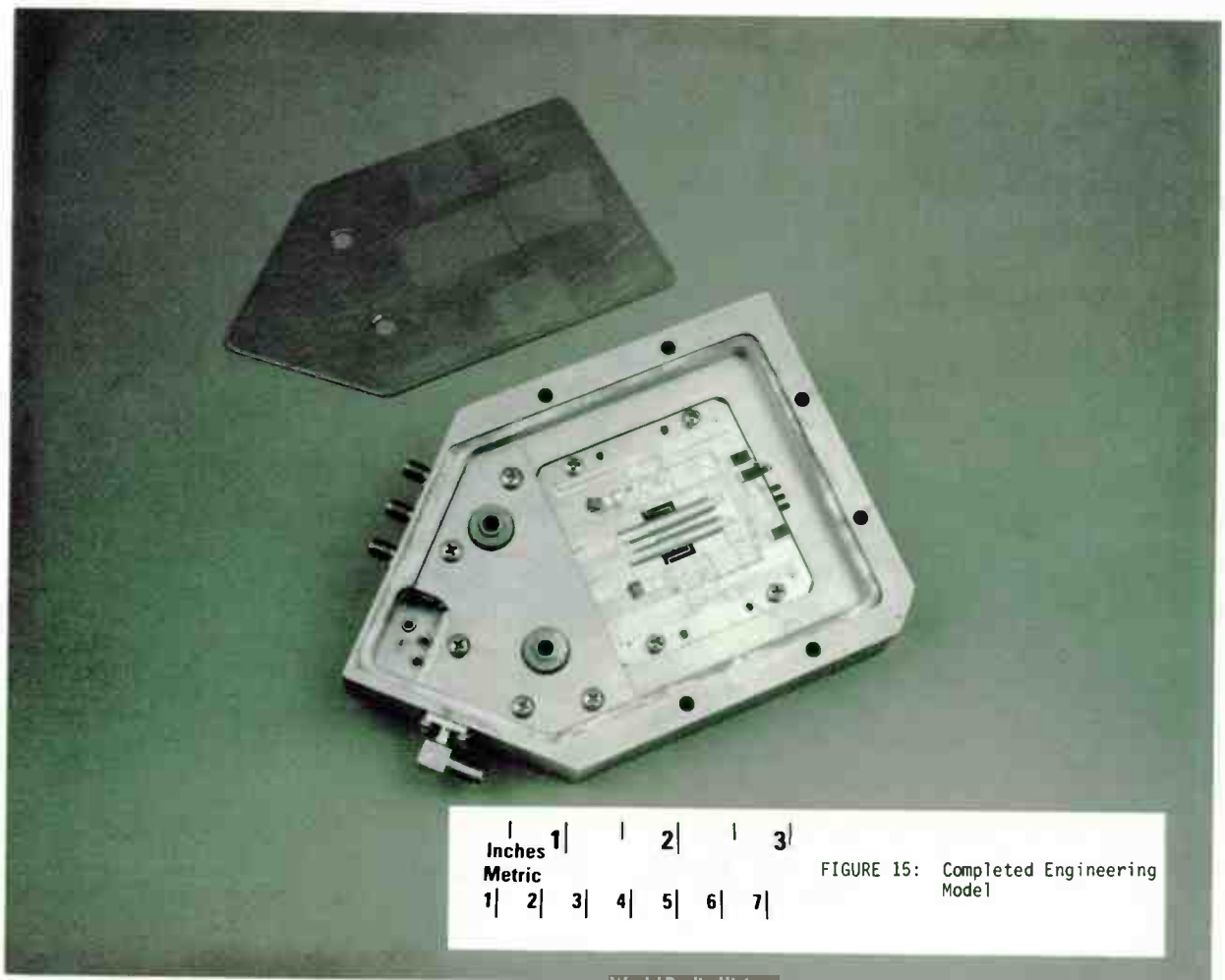


FIGURE 15: Completed Engineering Model

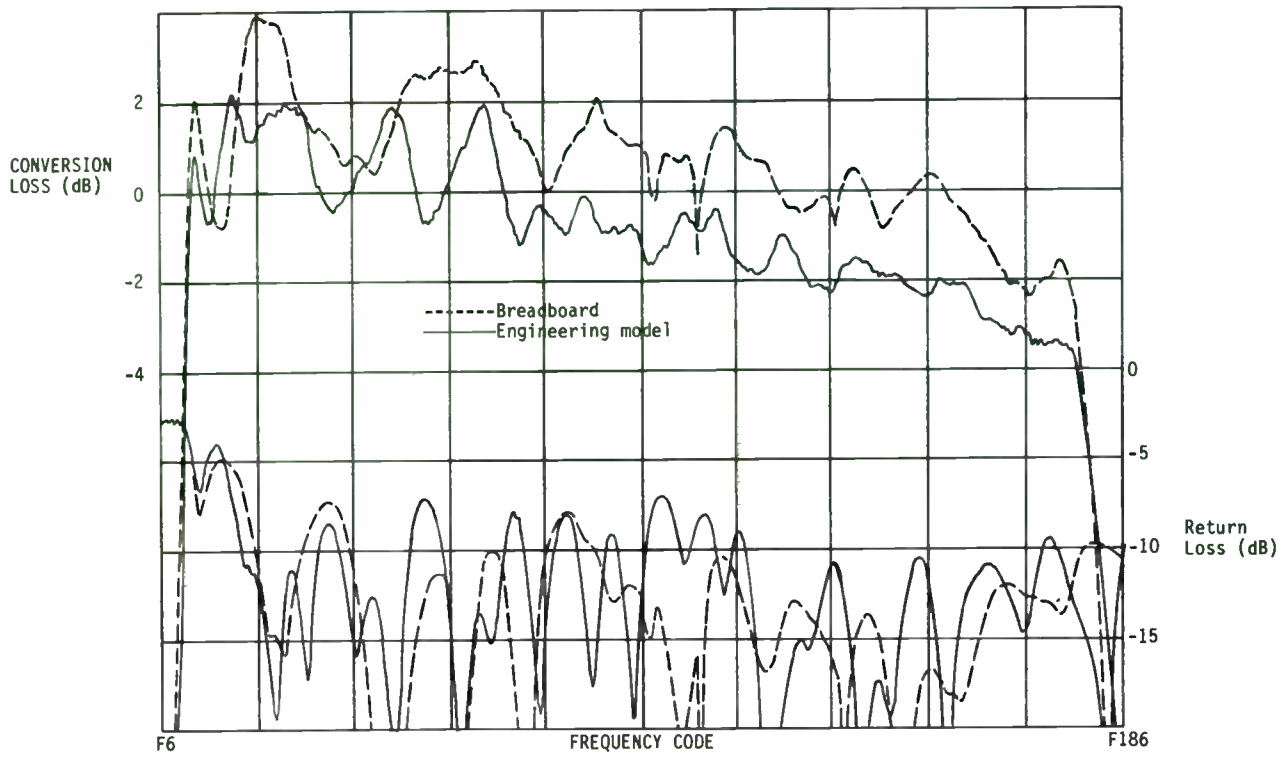


FIGURE 16: Comparison of breadboard and engineering model conversion loss and return loss.

FEI Microwave, Inc.
A SUBSIDIARY OF FREQUENCY ELECTRONICS INC

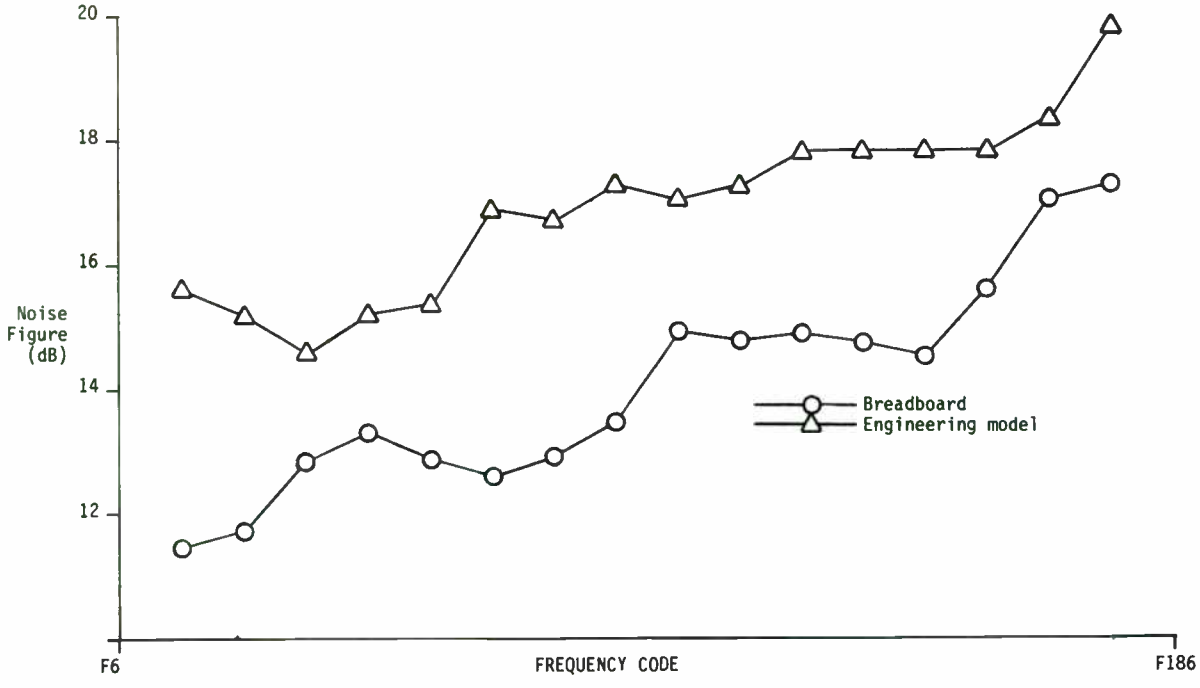


FIGURE 17: Comparison of breadboard and enineering model noise figure.

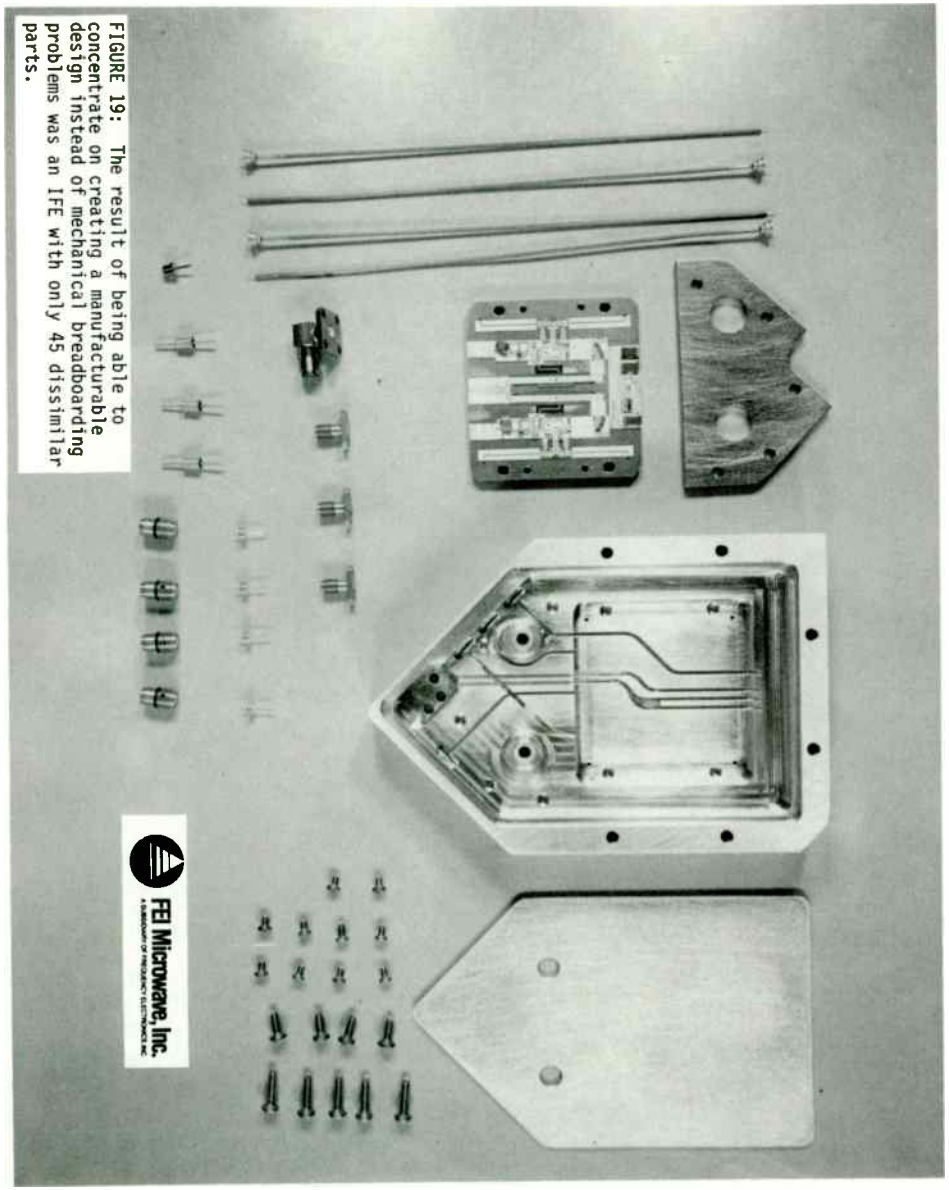


FIGURE 19: The result of being able to concentrate on creating a manufacturable design instead of mechanical breadboarding problems was an IFE with only 45 dissimilar parts.

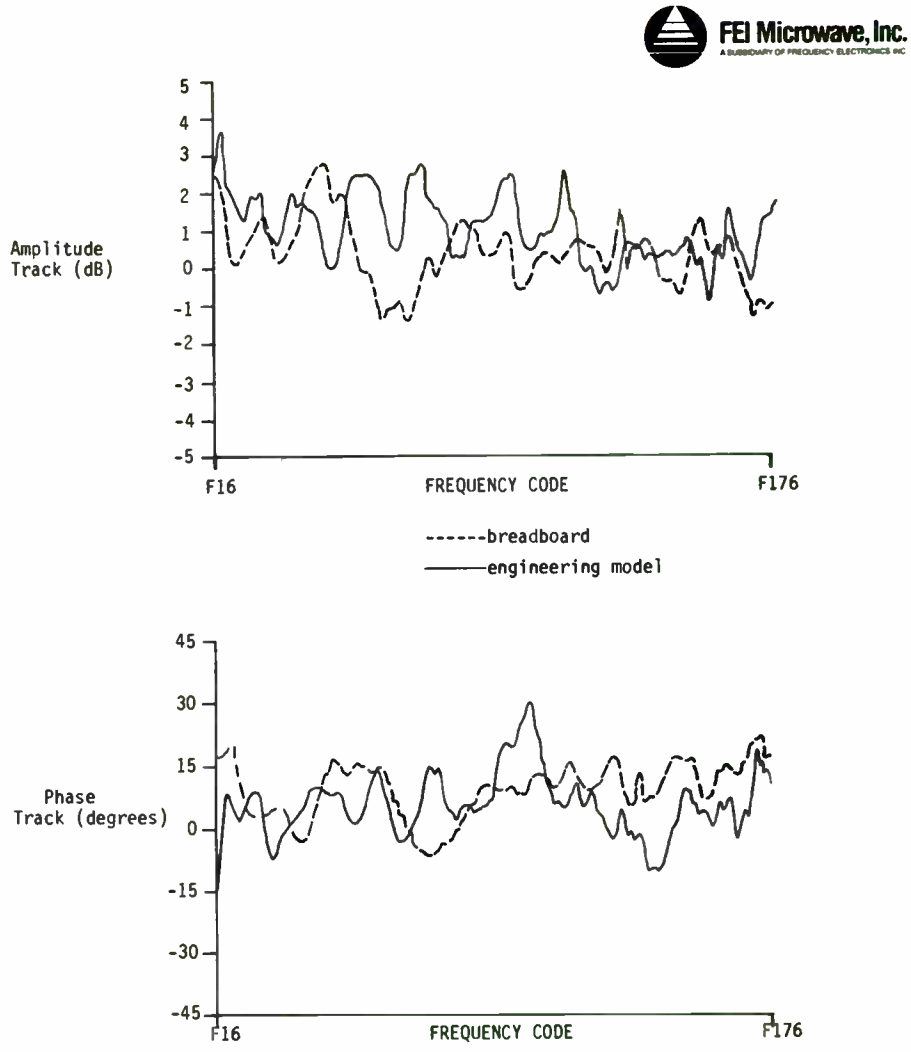


FIGURE 18: Comparison of breadboard, and engineering model, channel to channel amplitude and phase matching.

New Topology Multiplier Generates Odd Harmonics

by Charles Wenzel
Wenzel Associates, Inc.
14050 Summit Drive - Suite 119
Austin, TX 78728

The original purpose of this circuit was to take advantage of the superior noise and switching characteristics of Schottky barrier diodes to make a high performance odd-order frequency multiplier. Modern quartz oscillators have reached a level of performance where it has become difficult to multiply the fundamental frequency without degrading the phase noise by more than the unavoidable 20 dB per decade of multiplication. Thanks to the Schottky barrier diode's extremely low flicker noise, this circuit adds little excess noise to even the best sources. Additionally, the upper frequency limit of this configuration should be quite high: Schottky diodes are not slowed by minority carriers in the junction region and exhibit switching speeds measured in the picoseconds.

This particular application performs the difficult first stage multiplication of an ultra-low noise 10 MHz reference oscillator. Such low noise multiplication is necessary in constructing state-of-the-art synthesizers, radars and microwave communications equipment that rely on the reference oscillator for spectral purity near the carrier. I use the multipliers to make the job of measuring phase noise easier.

Topology

Figure 1 shows the voltage sinewave to current squarewave converter which forms the heart of the multiplier. This unique converter is basically a full-wave bridge with an inductor short-circuiting the DC terminals. The inductor is chosen to have a high impedance at the operating frequency so that an AC input results in DC in the inductor. This DC flows through alternate pairs of diodes due to the commutating action of the input voltage. So if one AC terminal of the bridge is driven with a low impedance sinewave, the other AC terminal will supply a squarewave to a low impedance load. The load must have a low impedance since the compliance of this current source is exactly equal to the input voltage. The Figure 2 shows the waveform obtained from the circuit in Figure 1 at 10 MHz with a 470 uH inductor. The bottom trace is the voltage across a 10 ohm resistor and represents a 10 mA p-p current squarewave. Notice that the diodes switch as the input signal's zero crossing. Consequently, this circuit produces a minimum of troublesome AM to PM conversion.

The Fourier expansion of a squarewave is:

$$f(x) = A \cdot \frac{4}{\pi} \sum_{n=1,3,5\ldots} \frac{\sin(n\,\omega\,x)}{n}$$

Notice that only odd harmonics are present with an amplitude inversely proportional to the harmonic number. All that is necessary to make a frequency multiplier is to design input and

output circuits to select and enhance the desired harmonic without disturbing the sine to square conversion.

Circuit Description

The circuit of Figure 3 meets these requirements to provide multiplication from 10 MHz to 30 MHz. The input matching network consisting of the 1.5 uH choke and the 150 pF capacitor does three jobs: it steps up the input voltage to overcome the diodes' barrier potential with series resonance; it provides a low impedance to ground for the switching current; and it isolates the input from the switching current. It is interesting to note that the input impedance of this series tank would be quite low except that the diodes' conduction spoils the Q with a resulting input impedance near 50 ohms for a 2 Vp-p input. for smaller input signals the input impedance drops, which explains the conversion efficiency peak near 0.5 Vp-p for the 25 ohm source (Fig. 4). This variable Q provides a degree of feedback to help ensure that the multiplier has a usable output over a wide range of input voltage.

The output network presents the required low impedance to the bridge while directing the desired harmonic to the output. The 0.22 uH choke and the 100 pF capacitor provide the low impedance away from 30 MHz and the series tank formed by the 2.2 uH choke and the 30 pF trimmer provide a low impedance path to the load at 30 MHz. Either higher Q or additional filtering may be used here if harmonic rejection better than about 30 dB is required.

Schottky diodes are mandatory for best performance. Passivated diodes without the p-n guard ring like the H-P 5082-2835 are good for higher frequencies but reverse voltage breakdown must be avoided. Hybrid diodes which include a guard ring are desirable due to the higher breakdown voltage, but they do exhibit more capacitance (about 1 pF at 0 V). The 5082-2811 has a reverse breakdown of 15 V nd exhibits 1.2 pF at 0 V. The 5082-2813 is a matched quad version.

The load inductor should exhibit a high impedance at the input and output frequencies. At low frequencies almost any large choke will suffice, but at higher frequencies a slight improvement is obtained by selecting an inductor to resonate with the diodes capacitance at the input frequency.

Test Results

Figure 4 shows the performance over a range of input voltage and Figure 5 shows the input and output waveforms. The conversion gain is good over a wide range considering that no active gain stages are employed. The conversion efficiency is as high as diode frequency doublers even though the multiplication factor is higher.

In order to check the phase noise, a second multiplier was constructed and grounded-gate amplifiers were added to boost the output level (Figure 6). Figure 7 shows the measured phase noise of the multipliers and of the oscillator used to make the measurement. Appendix A describes the phase noise measurement technique. The multipliers' noise is significantly better than

the oscillator so this test relies on noise cancellation in the mixer which should occur since both multipliers receive the same noise. Substituting a noisier oscillator increased the measured noise, suggesting that the measured noise may be in part due to dispersion in the signal paths. Such a measurement error would make the multipliers appear noisier than they actually are, but the indicated noise is already below most "ultra-low noise" oscillators.

Cost

The cost of this multiplier is quite low since the component count is low and no exotic parts are used. The trimmer capacitor should be high quality since the riveted connection to the rotor of cheap trimmers can become noisy. The prototype used molded chokes and NPO dielectric capacitors.

Applications

A half-wave version is easily constructed by eliminating the bottom two diodes and connecting the inductor to ground. This configuration is convenient for high frequency layout and has been used to multiply 100 MHz to 500 MHz with about 20 dB of loss. Add a $0.98 MMIC amplifier for a cost effective high performance multiplier. Stripline techniques would prove interesting above 500 MHz.

For high order multiplication the constant current inductor can be reduced so that the output becomes a short pulse instead of a square wave, reducing the power in the lower harmonics. This change also makes a nice bipolar pulse generator.

Summary

A high performance, low cost frequency multiplier has been described. The phase noise performance is sufficiently low to not seriously degrade the best commercially available oscillators and the conversion loss is good even for low level inputs. The new topology will provide state-of-the-art odd-order frequency multiplication over a wide range of frequencies.

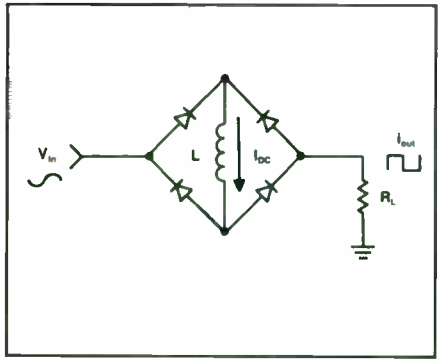


Figure 1. Sinewave to squarewave converter circuit.

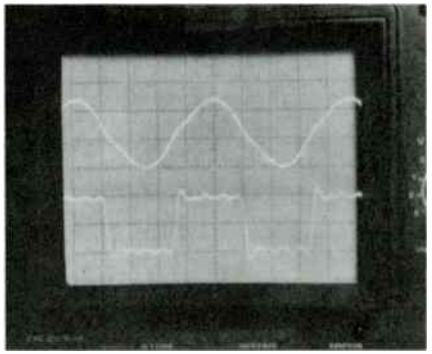


Figure 2. Top trace is signal into diode bridge (2V/div). Bottom trace is current in load resistor (10mA/div).

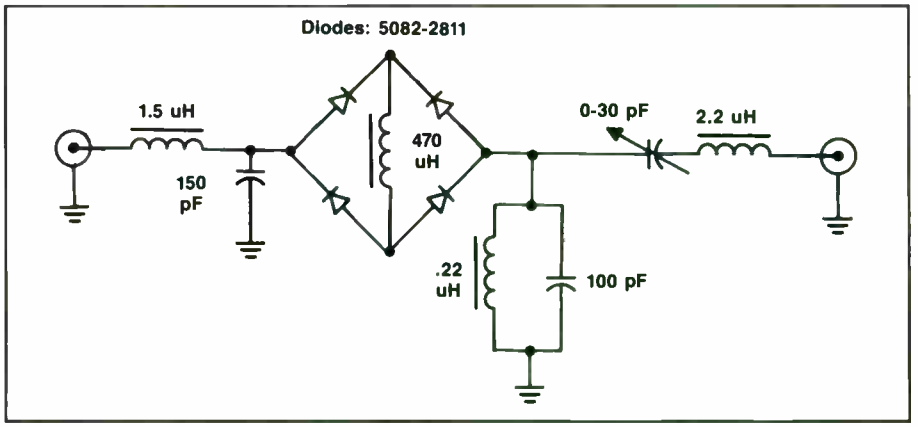


Figure 3. 10 MHz to 30 MHz multiplier circuit.

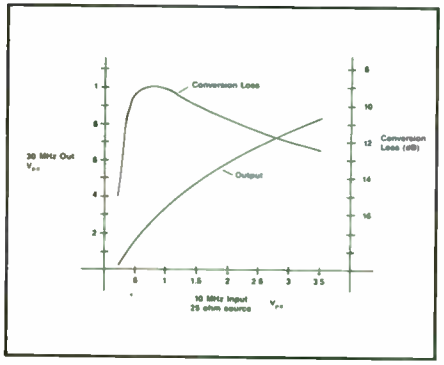


Figure 4. Prototype test results.

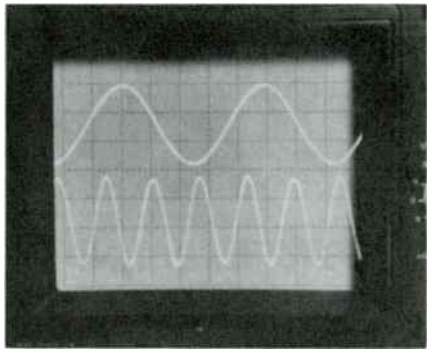


Figure 5. Top trace is input (1V/div) and bottom trace is output (.5V/div).

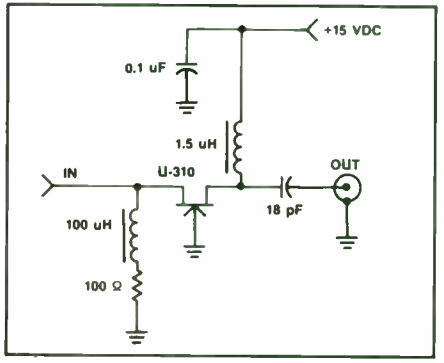


Figure 6. 30 MHz amplifier added to multipliers.

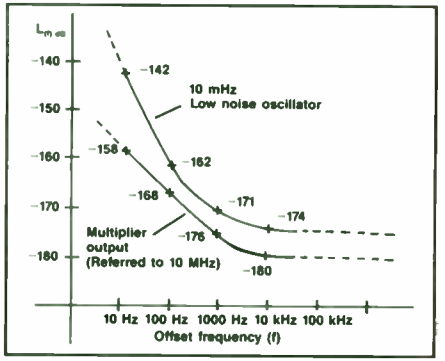
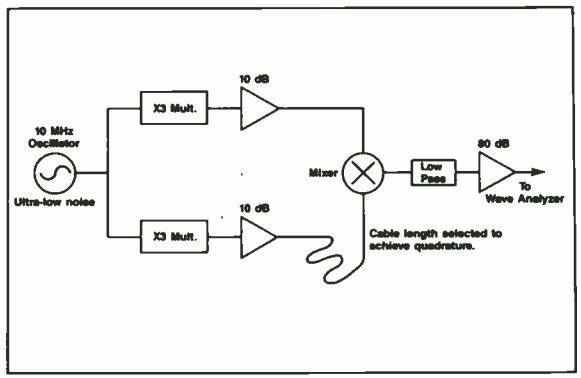


Figure 7. Phase noise measurement.

APPENDIX A

Phase noise is usually expressed as either a phase or frequency spectral density, $S_\phi(f)$ or $S_\nu(f)$ where f is the frequency offset from the carrier (Fourier frequency). The commonly used $\mathcal{L}(f)$ is the single-sideband phase noise-to-signal ratio and is one half of $S_\phi(f)$ for the low modulation index associated with low noise signals. Frequency multiplication causes sideband power (including phase noise) to increase by 20 dB per decade (i.e. the sideband power increases by n). Thus an ideal 3x multiplier will increase the noise by 9.5 dB (20 log 3 = 9.5).



The key to ultra-low phase noise measurements is the Schottky diode double balanced mixer (see Figure). When two quadrature signals are applied to the RF ports, slight variations in the relative phase of these signals are converted into voltage variations in the DC component of the IF port signal. Thus the mixer has a characteristic conversion sensitivity expressed in volts per radian. For low noise work the IF signal is filtered, amplified, and applied to a spectrum or wave analyzer. This noise voltage measurement is converted to phase noise as follows:

$$\mathcal{L}(f) = 10 \log[(V/K)^2/BW] - 3\ dB - 3\ dB$$

where: V is the measured voltage
K is the conversion gain (volts/radian)
BW is the measurement bandwidth

The first 3 dB correction assumes that the two devices under test contribute equal noise. The second 3 dB is removed because this hetrodyne technique detects both sidebands and since these are uncorrelated noise signals one sideband will have half the power.

A HIGH-PERFORMANCE NARROW-PULSE UHF TRANSMITTER

Steven M. Harrison
RF Engineer
System Planning Corporation
1429 N. Quincy Street
Arlington, Virginia 22207

The design of a very high-performance, narrow pulse, high-power VHF transmitter was previously described by M. Clark[1]. This paper presents the design and implementation of a UHF version with similar performance goals. Slightly different design techniques were used to achieve those goals, and the emphasis of this paper will be upon those techniques.

INTRODUCTION

Radar Cross Section measurement using pulsed RF signals requires extremely high pulseshape fidelity. Very fast rise- and falltimes are required to allow time-gating of clutter, and the pulse must be amplitude- and phase-stable (flat) in order that CW-equivalent measurements may be obtained. The bandwidth of the pulse is related to risetime as:

$$\text{Risetime} = \frac{0.7}{BW}$$, where BW = 3dB full RF bandwidth

Thus, nanosecond risetimes imply instantaneous transmitter bandwidths of several hundred Megahertz. Noise generated by the transmitter during receive periods must be kept very low in order to avoid compromising system instantaneous dynamic range; this can be achieved by pulsing the bias.

THE TRANSISTORS

In the past, wide-bandwidth high-power solid-state amplifiers have been designed using transistors intended for narrow-band Class-C operation. Few, if any, devices existed for use above 500 MHz, and high-power in that frequency range was usually achieved by combining many low-power transistors.

Devices are available today that enable multi-octave VHF and UHF linear amplifier designs with several hundred watts output power. In the 225 to 400 MHz region, both bipolar and MOSFET devices will deliver up to 100 watts at 500 MHz. Bipolar transistors are available for the 500 to 1000 MHz band with power ratings to 50 watts; these transistors are usually intended for low-power UHF television transmitters, and achieve high-order intermodulation distortion suppression at both high output power and wide operating bandwidth.

Even with these advances in technology, however, a compromise (usually in gain flatness and/or maximum available gain) must occasionally be accepted when an application requires either greater bandwidth or operation beyond (or below) the specified operating

frequency of a transistor. That situation was encountered when System Planning Corporation was required to build a pulsed UHF transmitter to deliver 100 watts peak power over the 365 to 600 MHz band; there did not seem to be any high-power RF transistors on the market that would cover that band. Experiments with TRW MRA0105-75 and MRA0105-100 transistors in a VHF transmitter showed that those devices did not have usable gain beyond about 550 MHz. Conversations with TRW engineers indicated that the MRA0510-50H bipolar transistor might meet our requirements, even though that device would be operating somewhat below its design frequency range. The concern was that since the transistor incorporates internal input matching to operate from 500 to 1000 MHz, gain ripple and group delay might be unacceptable for the proposed application of high-fidelity pulse amplification. However, the device performs very well, and reproducible amplifiers have been built with acceptable performance down to at least 200 MHz.

The MRA0510-50H is a dual bipolar common-emitter RF power transistor, designed for Class AB amplifiers operating over the 500 to 1000 MHz band with 6.5 dB minimum broadband power gain. It consists of a pair of 25-watt output transistors with diffused emitter ballast resistors and internal input-matching networks. At the time design began on the UHF amplifier submodules, TRW did not have impedance data available, so we attempted to measure the transistors with an in-house-built test fixture and the HP8510A Vector Network Analyzer. However, repeatable input impedance measurements could not be made using the in-house transistor fixture. The large-signal S-parameter data that appeared valid indicated that the transistors apparently had an internal low-Q resonance around 375 to 425 MHz which made input-matching below 500 MHz difficult to achieve; fortunately, all transistors seemed to have the same input resonance effect at the same frequency. Attempts to model the transistor for use in Touchstone software was made, but eventually we resorted to a breadboard design by Lance Wilson of TRW. This breadboard was then optimized to cover the frequency band of interest. Several Acrian samples of the same transistor were obtained, and performed in much the same manner. The final choice of manufacturers was made based upon the increased ruggedness and ability to withstand high output VSWR of the TRW versions. Gain ripple due to VSWR reflection effects back to the preceding stage could be masked by the use of Quadrature Hybrids[A], and that was the key which made this transistor usable below its design frequency.

APPLICATION OF THE TRANSISTORS

Solid-state amplifiers operating up to 1 GHz intended for SPC's MK III RCS measurement systems have traditionally operated Class A with no noise blanking employed. The major disadvantages of these amplifiers are extreme inefficiency (particularly with very low duty factors such as normally encountered in RCS transmitters) and weight; but another concern is the high noise level output during receive periods. These disadvantages were successfully overcome in the design of the VHF and UHF transmitters by the use of bias pulsing, the same technique used for NMR (Nuclear Magnetic Resonance)

high-power RF amplifiers. The bias-pulsing in the VHF and UHF MK III transmitters turns the RF amplifiers on and off within 100 nanoseconds; this speed is limited by the need to prevent the radiation of extraneous pulse energy, particularly within the operating frequency bands of the RCS system.

In the UHF transmitter, each amplifier submodule utilizes one MRA0510-50H transistor package. The two transistor chips are operated in push-pull, and biased into Class A by the application (to the transmitter) of a TTL-compatible gating pulse called the Transmit Bracket. The base-bias pulse is generated on the control logic board by a time-delay circuit. The primary function of this circuit is to time the turn-on and turn-off of the final amplifier module so that video switching transient noise generated in the intermediate amplifier is not radiated. Thus, the final amplifier module is gated on after the intermediate amplifier is on, and off before the intermediate amplifier is gated off. The RF transistor bias pulses in the VHF transmitter exhibited a rise- and falltime of less than 50 nanoseconds. Because the emitter-follower used in the VHF transmitter does not pull the RF transistor bases to ground quickly, the base-to-collector capacitance of the transistor tends to stretch out the on-time another 50 nanoseconds, resulting in a trailing tail on the output pulse. This tail is not evident when the transmitter is operated with a shaped RF input pulse, but does become obvious when the output pulse is generated by gating the amplifier modules while applying a CW RF input signal. In the UHF transmitter, the emitter-follower was changed by adding a pulldown transistor, and this resulted in reversing the rise- and falltimes to the extent that the transistors turn off faster than they turn on (risetime about 50 nanoseconds, falltime about 25 nanoseconds). See Figure 2. Figure 3 shows the timing relationship between the Intermediate Amplifier and the Final Amplifier brackets.

Ringing that had been observed on the pulse envelope of the VHF transmitter was traced, during the development of the UHF transmitter, to the relationship of the amplifier bandwidth compared to the pulse bandwidth. The shaped RF pulse generator used during testing of the VHF transmitter exhibited rise- and falltimes of less than 3 nanoseconds; this required a bandwidth of at least 233 MHz. During the last development stages of the VHF transmitter, the amplifier modules were modified to operate from about 120 MHz to over 500 MHz; the ringing problem disappeared with this increase in bandwidth. The UHF amplifier submodules, on the other hand, were initially intended to operate from 200 MHz to 700 MHz, which allowed rise- and falltimes approaching 1.5 nanoseconds. The VHF transmitter design has been discussed by M. Clark [1]; the following describes the UHF version.

THE UHF TRANSMITTER SYSTEM

The performance requirements of the UHF transmitter are listed in Table I. These requirements are essentially the same as those of the VHF transmitter except for the operating frequency. The UHF Transmitter System block diagram is shown in Figure 4. The

transmitter consists of three seperate amplifier modules, a control logic board, power supplies and directional couplers for signal sampling. The RF input signal is first sampled by a 10 dB directional coupler; the sample is made available on the rear panel for monitoring. The signal is then amplified by a Trontech P635A-2 preamplifier which provides about 30 dB gain, raising the signal level to about +30 dBm across the band. The gain flatness of the Trontech preamplifier is about +/- 0.25 dB or better up to 700 MHz where the gain peaks about 1 dB higher. The output of the preamplifier is then fed to the Intermediate Amplifier module.

The output of the preamplifier is isolated from the first MRA0510-50H amplifier stage by an 8-dB attenuator; this is necessary to prevent excessive amplitude ripple caused by the high input VSWR of the single-ended MRA0510-50H stage below 500 MHz. It was interesting to observe the effect upon the pulseshape of this VSWR; the pulse appeared to be seperated in time into a 25 nanosecond first portion which was relatively stable. The remainder of the pulse would flop up and down as frequency was changed, and the ripple on the envelope of the pulse would decrease as the frequency was increased to the rolloff of the stage. Correlation of the pulse bandwidth (set by the rise- and falltime of the input RF pulse) and the amplifier bandwidth was evident by the amplitude and time domain of the overshoot on the output RF pulse.

The second amplifier in the Intermediate Amplifier module consists of two similar MRA0510-50H submodules combined with Sage Laboratories Wireline used as a Quadrature Hybrid power splitter and combiner; the 3 dB bandwidth of the hybrids is greater than that of the amplifier submodules. The gain of the combined submodules is in excess of the TRW minimum specification of 6.5 dB broadband, approaching 10 dB. Had the input VSWR of the single-ended amplifier submodule been acceptable across the band, +50 dBm output power could have been obtained from the Intermediate Amplifier module alone. However, the design output power goal was +50.5 dBm minimum at the amplifier system output connector, and as is usually the case, an additional couple dB of amplification was necessary to attain that level. The peak output power of the second stage is about +45 dBm (32 watts).

The Intermediate Amplifier module is housed in a 9 1/4" X 6 5/8" X 1 1/2" aluminum box. Figure 5 shows the Intermediate Amplifier; note that the photograph was taken before the design was finalized. The photograph shows the RF Input connected to the combined submodules, whereas the final implementation actually used the single-ended submodule for the first Intermediate Amplifier stage. A dip in gain at about 450 MHz appeared when the top cover was installed on the box; judicious application of RAM (carbon-impregnated foam called Radar Absorbent Material) within the box broke up the resonance, eliminating the dip in frequency response.

The output of the Intermediate Amplifier is fed to the Final Amplifier, which consists of four identical amplifier submodules similar to the second stage amplifiers in the Intermediate Amplifier. See Figure 6. Two sets of two Quadrature-Hybrid- combined submodules in the Final Amplifier are combined with in-phase "Magic-Tee" Hybrid power splitter-combiners. This type of splitter-combiner was chosen over the Wilkerson type because of the much smaller physical size. Extremely wide bandwidth performance of the "Magic-Tee" hybrids is a plus, although not entirely necessary since the amplifiers are the bandwidth-limiting component in the system. The Final Amplifier submodules are enclosed within a 10 5/8" X 11 3/4" X 1 1/2" aluminum box. There was no resonance apparent when the top cover was installed on the Final Amplifier module box. The output of the Final Amplifier is fed to a 30 dB directional coupler for output monitoring, and then to the rear panel. There is no foldback VSWR protection circuit utilized. The MRA0510-50H transistors are rated to withstand a 5:1 output VSWR under Class AB CW conditions; in a low-duty factor, narrow-pulse application such as this transmitter, the ability to withstand high output VSWR should be enhanced. Individual amplifier submodules, as well as the entire transmitter, have been operated into both open and shorted loads without apparent damage.

Seperate power supplies are used for each of two submodules in the Final Amplifier, for the Intermediate Amplifier, and for the Preamplifier. The Control Logic board of the UHF Transmitter System performs two primary functions; it interfaces the transmitter to the operating controls, and provides protection to the amplifier submodules by inhibiting the TTL Transmit Bracket in the event of excessive transmit bracket pulsewidth (nominally set at 12 +/- 1 microseconds maximum) or duty cycle (nominally 5 +/- 1 percent maximum); either situation initiates a RESET condition, which, after the cause has been corrected, can be RESET from the operating controls. A secondary function, as previously described, is to time the Final Amplifier Transmit Bracket (bias pulse) to eliminate noise amplification when there is no RF pulse output. The control logic also detects under- and overvoltage conditions of the +28 VDC power supplies, and overvoltage of the +5 VDC power supply. If any of these events occur, the control logic shuts off primary power to those five supplies and initiates the RESET condition; the control logic can be RESET from the operating controls.

Design Factors

As stated earlier, the key to making the MRA0510-50H transistors work well in cascaded amplifier stages was the extensive use of Quadrature Hybrids to form balanced amplifier stages. The primary cause of pulse distortion turned out to be the aforementioned VSWR reflection problem from the single-ended Intermediate Amplifier submodule back to the preceding stage. The design of the Intermediate Amplifier originally placed the hybrid-combined stages at the output of the Trontech rather than the single-ended stage. This provided excellent pulse fidelity across the band due to the low input VSWR of the quadrature hybrid; however, the single-ended amplifier stage tended to operate at or just below saturation, and the pulseshape suffered

as a result. The final design operates the single-ended stage at very low power compared to its capability, and obtains high output power from the second stage with two combined amplifiers, both operating at about half of the maximum output power. Similarly, the Final Amplifier stage operates below its maximum output power capability. The base bias pulse circuit used in the VHF amplifier was an emitter follower. In the UHF transmitter, this circuit was modified with a pulldown transistor switch to pull the RF transistor bases back to ground faster than the emitter follower did. The base-to-collector capacitance of the transistor chip must be discharged quickly in order to turn the transistor off. This required a minimum of bypassing capacitance in the base bias RF decoupling network. Ferrite beads were used to help decouple the bases, and had little, if any, effect upon the pulseshape. The collector decoupling networks had to be a compromise between insufficient low-frequency decoupling (LF/HF region) and I^2R losses in the chokes due to the high peak collector current. Large electrolytic capacitors were placed at the power terminations of both the Intermediate Amplifier and Final Amplifier. At the same time, the base bias level was increased to pulse the transistors into Class A; the VHF transmitter submodules operated in Class AB. Peak collector current per transistor in the UHF transmitter is about 2.5 amperes. The high bias level increased the gain per submodule several dB, and was one reason the gain exceeded the TRW specification. Since the bias level puts the amplifier modules well into Class A, the output pulseshape is virtually identical to the input pulse. As in the VHF transmitter, each amplifier submodule has its own bias switch driver board to maintain isolation between amplifier boards. A TTL-compatible 50-ohm line driver is included in the bias pulse interface board within each module to drive the bias switches.

CONCLUSION

Extremely high quality VHF and UHF narrow-pulse Transmitter Systems have been designed, and the hardware implementation described. Their design required solving problems posed by the requirement to maintain pulseshape fidelity over narrow operating bandwidths relative to the pulse bandwidth, and at high output power levels. The transmitters utilized state-of-the-art techniques and components to provide state-of-the-art pulse amplification performance.

APPENDIX

A. Peak-to-Peak Gain Ripple (dB) =20 log (1 +/- $|\Gamma_A \Gamma_B|$), where

Γ_A is the reflection coefficient of the source, and

Γ_B is the reflection coefficient of the load

BIBLIOGRAPHY

1. "High Power Transmitters for Narrow Pulse, Wide Band Applications", presented at RF EXPO '87, Anaheim, California, February 1987

REFERENCES

1. RF DEVICES DATABOOK. RF Devices Division, TRW Electronic Components Group, TRW Inc., 1986.

2. Solid Circuits, Circuit Techniques. edited by J. Johnson, Communications Transistor Company, 1973.

3. "90 Degree Hybrid", ANZAC-RHG RF & Microwave Signal Processing Components, Adams-Russell Electronics Co. Inc., pp.229-233, 1987

4. DESIGNERS GUIDE TO WIRELINE & WIREPAC. Sage Laboratories, Inc., 1985

5. AMPLIFIER HANDBOOK. Miteq, 1986.

Table 1. 100 Watt UHF Pulse Transmitter Specifications

Operating Frequency Range:	365 to 600 MHz
Output Power (minimum):	+ 50 dBm peak (100 watts)
Gain:	50 dB minimum
Gain Flatness:	+/- 2 dB maximum
Pulse Amplification:	Mode 1 (application of gating pulse, CW RF input): transmitter to be capable of forming RF pulse with maximum risetime of 50 nanoseconds and minimum falltime of 100 nanoseconds
	Mode 2 (application of shaped RF input pulse): transmitter to amplify RF input pulse with rise- and falltimes as short as 5 nanoseconds with minimum additional distortion
	Either mode:
	a. Pulse-to-pulse amplitude variation: no more than 0.1 dB
	b. Pulse droop: no more than 0.2 dB/microsecond
	c. RF On-to-Off ratio: 60 dB minimum
	d. RF output pulse to be amplitude- and phase-stable within 10 nanoseconds after reaching 10 % point of RF envelope

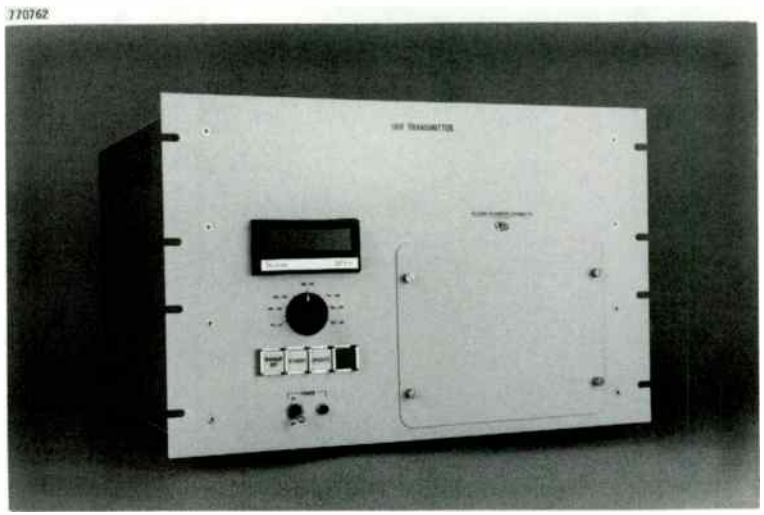


Figure 1

UHF Transmitter System

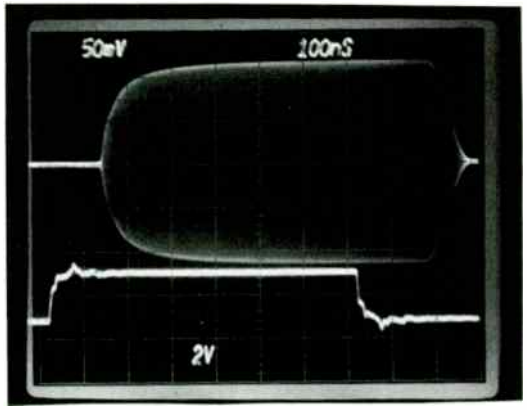


Figure 2. RF Output Pulse vs. Input Gating Pulse, CW RF Input

Top Trace: RF Output Pulse Bottom Trace: TTL Gating Pulse

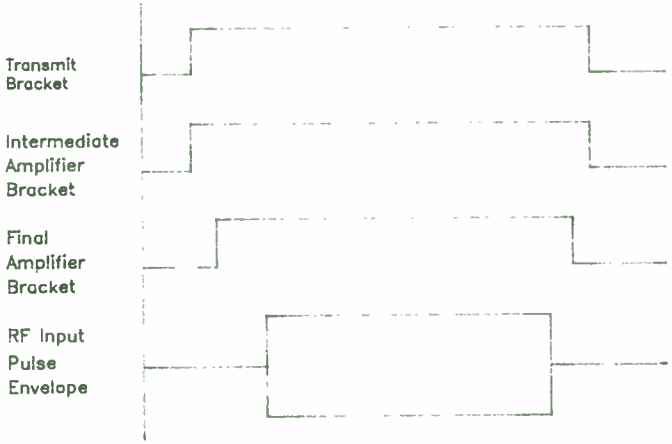


Figure 3. Timing Relationships, Gating Brackets and RF Pulse

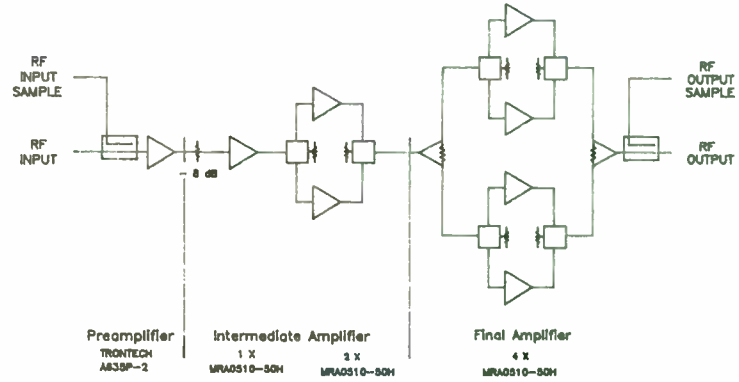


Figure 4

UHF Transmitter System

RF Block Diagram

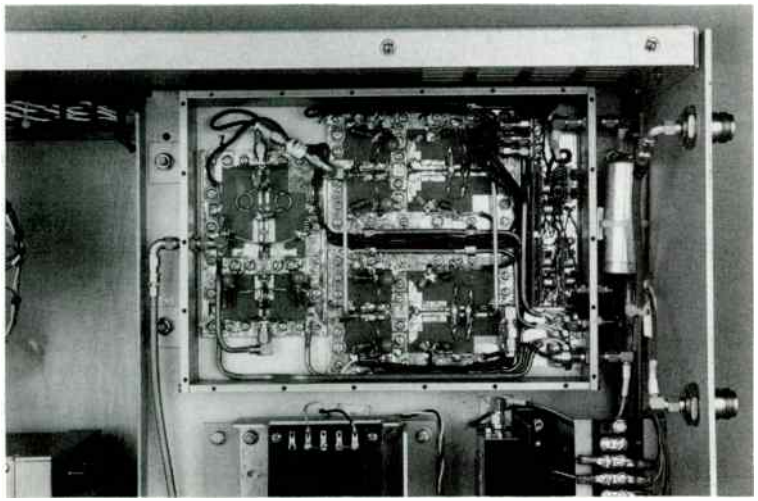


Figure 5

UHF Transmitter System

Intermediate Amplifier

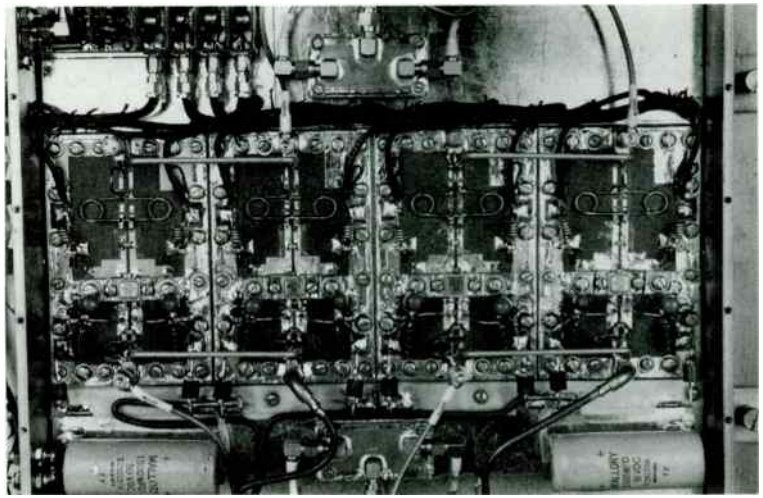


Figure 6

UHF Transmitter System

Final Amplifier

THERMAL CONSIDERATIONS IN AMPLIFIER DESIGN

by

Gregg A. Hollingsworth
Design Engineer
Acrian, Inc.
490 Race Street
San Jose, CA 95126

The big push today in the RF amplifier market is for higher power in smaller packages. While the electrical design of these amplifiers is given rigorous attention, many times the thermal design is just briefly touched upon. This paper will explore the thermal design considerations of the die, package, baseplate, and system heat exchanger elements of RF power amplifiers.

As the power output in these amplifiers increases, so does the amount of heat dissipated. Failure to design adequate heat sinking will result in an unreliable amplifier. As can be seen in figure 1, the thermal resistance of the amplifier is the summation of each of the parts in the thermal path (there are several computer programs which will model the thermal path and compute the thermal resistances of the system). There are several guide lines to follow which will produce a good thermally designed amplifier.

BASIC HEAT CONDUCTION

The basic heat conduction formula applies to designs where the cross-sectional area of the heat flow path is constant over the entire length. This is not the case with a transistor, since the die is usually smaller than the substrate it is attached to. This causes the heat to spread out at the edges of the die at some angle ϕ. Figure 2 shows how this makes the cross-sectional area of the heat flow path grow as it travels through the length of the material. Using the basic formula for thermal resistance

$$\theta jc = X/kA \ ^{o}C/watt \qquad (4)$$

where

X = length of thermal path (inches)

k = thermal conductivity of material (watts/inch)

A = cross-sectional area of thermal path (inches^2)

and the results of Rensted and Surty[2], assuming that the heat flow spreads uniformly from the heat source at a constant spread angle of 45 degrees, the average steady state thermal resistance can be calculated. For a square heat source

$$\theta jc = X/kL(L + 2X) \qquad (5)$$

and for a rectangular heat source

$$\theta jc = [1/2k(L-W)]\ln[(L/W)(2X+W)/(2X+L)] \qquad (6)$$

where L equals the length of the heat source and W equals the width of the heat source.

SEMICONDUCTOR DESIGN

The first step in the thermal design is the development of the die. In order for the transistor designer to develop increasingly higher power die, the active area of the die must be increased. Increasing the active area without altering its geometry will cause the junction temperature to rise

approximately in proportion to the square root of the area. The resulting large collector-base area devices are more prone to premature failure because they exhibit non-uniform current and temperature distributions, leading to "hot-spots".

In order to achieve the higher power levels, the active areas are split into multiple-isolated power generating areas. This increases the base periphery or heat-spreading edge leading to a reduced thermal resistance on a unit area basis.

The most common method of increasing the active periphery-to-area ratio, is by increasing the number of paralleled cells and making the length much greater than the width of each cell. Phase differences associated with this design (including package parasitics and physical die dimensions) can result in unequal power drive to the individual cells. As seen in Figure 3, the temperature distribution within the die will be unequal, leading to "hot" inner areas on the die while the cells on the outer edges remain "cool".

The unequal drive to the cells can be reduced by using emitter ballasting. As an emitter-base junction site becomes excessively hot, it will try to draw more than its proportionate share of drive current. Discrete resistive elements are introduced, in series with the emitter contact metallization, to divide the input voltage in relation to the I^2R drop required by the relative impedance levels (both the series resistance and the emitter-base junction).

PACKAGE DESIGN

The first consideration in the design of the package is which substrate material to use. Most packages use beryllia (BeO) substrates because of their high thermal conductivity (5 watts/inch ^{o}C compared with 0.7 watts/inch ^{o}C of alumina). BeO has been proven in volume production and can be easily metallized, making it a good material choice for transistor packages.

One of the potential problems associated with Beo is that it is mechanically weaker than alumina and can be prone to cracking and breaking during processing. Increasing the thickness of the BeO to 40 to 60 mils (compared to 25 mils needed for alumina) will solve this problem. As seen in equation 4, the added thickness (X) will increase the thermal resistance of the package.

The next step in the package design is the metalization process. There are two processes used in metalizing the substrate, thick-film and thin-film. The most often used method is a thick-film process where a conductive adhesion layer, usually tungsten, is screened onto the substrate (1 mil thick) and fired at 1100 ^{o}C. A thin layer of gold is then plated onto the tungsten. This thick-film process is inexpensive and, because of its ability to withstand high temperatures, lends itself to brazing of the substrate to the flange(750 to 950 ^{o}C). This creates a more reliable package than alternative solder technology.

In the alternative thin-film process, the conductive adhesion layer is sputtered on to the substrate, and the gold layer sputtered on top of it.

The process yields a total metallization thickness of less than 10 microinches, producing a lower themal resistance than the thick-film process. However, the thin-film process withstands only about 550°C, too low for brazing.

There are two methods of improving the maximum temperature of the thin-film process to a level suitable for brazing. One method adds a layer of platinum between the gold and adhesion layers. Another requires the use of ruthenium as the adhesion layer material. Both methods are substantially more costly than the thick-film process.

The choice of flange material depends on the size of the substrate. In packages where the braze attachment (measured diagonally) is less than .32 inches, copper is usually selected. However, when the braze attachment is larger than .32 inches, the difference in thermal expansion between the copper and BeO during the brazing process causes breakage. Under these conditions, an alloy called Elconite (80% tungsten, 20% copper) can be used. But Elconite has about twice the thermal resistance of copper, leading many designers to use two smaller BeO substrates on a copper flange, instead of one larger BeO substrate on an Elconite flange.

BASEPLATE DESIGN

Once the packaged transistor has been designed, the next step is to develop the most efficient thermal conductor for the base plate. The first consideration for the base plate is which material to use. Table 1 shows the thermal conductivity of materials commonly used in electronic equipment. One characteristic of the materials is that the thermal conductivity will change with temperature as seen in figures 4 & 5.

The two most commonly used materials for the base plate are copper and aluminum. Although copper has almost twice the thermal conductivity of aluminum, other requirements such as cost, weight and mechanical strength need to be considered. Aluminum is less expensive than copper and weighs substantially less per unit volume than copper. The base plates can be made thinner since it is mechanically stronger than copper also. Table 2 shows the temperature differences, normalized to copper, that would be required to conduct one unit of heat through a one inch length of material with a one $inch^2$ cross-sectional area.

The next consideration is the geometry and how it will effect the thermal resistance of the base plates. The most commonly used base plate geometry is a rectangle whose thermal resistance can be found by using equation 6. Equation 6 can also be used for thermal resistance calculations on complicated geometries by dividing the total length into several segments, each with a different cross-sectional area, and taking the sum of their thermal resistances.

The mounting interfaces of the system are key areas of concern. There are several problems involving the interfaces of the system where the surfaces of parts are not perfectly smooth and flat on a microscopic basis. As in Figure 6A, heat is conducted through the interface only at the points of contact. The effective cross-sectional area for the heat conduction is greatly reduced to as little as one percent of the total available cross sectional area.

There are four basic methods used to maximize the contact area of the interface. The first is to improve the surface finish of the two interfacing surfaces. The second is to apply large clamping pressures to force the two surfaces against each other. The clamping pressure must be uniform across the interface or the interfaces can bow, as in figure 6B, resulting in less surface area contact. The effects of surface finish and clamping pressure are shown in figure 7. (The data presented is for aluminum surfaces. If a different material with a different hardness were used, the clamping pressure and surface finished needed to achieve the same results would vary.)

The third way to improve surface area contact is to solder the two surfaces together. This is the best means to eliminate the mounting interface problems as it fills the voids in the interface with solder metal. However, this method has three distinct disadvantages:

- The replacement of components is difficult.
- The two surfaces must be plated so that the solder will wet the surfaces.
- Care must be taken to avoid damaging the components during soldering.

The fourth method of improving surface interfaces is an alternate to solder. The two surfaces are clamped together and a thermally conductive paste or heat sink compound is spread between the surfaces. These compounds are generally made of silicones loaded with thermally conductive metal oxides. The paste will fill the void between the mating surfaces and will not leak out of the interface, dry out, harden or melt, even under long exposure to temperatures up to 200 ^{o}C. The thermal conductivity of the pastes are approximately .01 to .04 watts/inch ^{o}C, which is 1/1000th the thermal conductivity of copper.

SYSTEM HEAT EXCHANGER DESIGN

Once the amplifier has been designed, the next step is to integrate it into the system. The amount of heat dissipated by the amplifier will determine the method of cooling required for the system. This paper will discuss two basic methods of cooling; radiation and natural convection, and forced air.

Radiation and Natural Convection

Of all the heat transfer methods, radiation and natural convection is the simplest to use. When designing the system, the heat transferred by radiation and the heat transferred by natural convection must be calculated separately, and then combined.

The amount of heat transferred by radiation depends on the following factors:

- Temperature of the radiating surface
- Temperature of the surrounding air
- Surface characteristics
- Shielding effects of adjacent fins

The formula for the amount of heat transferred, Q, is

$$Q = Q_r \varepsilon Aeff\ As \qquad (10)$$

where

Q_r = heat per unit area transferred by perfect, unshielded radiator (watts/Inch)

ε = emissivity of surface

Aeff = reduction of effective surface area caused by shielding

As = surface area ($inch^2$)

Figure 8 shows the dependency of Qr on surface temperature. The data shown is the amount of heat transferred by radiation from a "perfect" black body radiator. The factor which specifies how good a particular surface is in relationship to a "perfect" radiator is the emissivity, ε . Table 3 shows the importance of the surface characteristic on heat transfer by radiation. From the table, it can be seen that any surface can be made to approach the ideal radiator by painting the surface, especially with any color of oil paint.

The effect of shielding by adjacent fins on the effective radiating fin area is seen in figure 9. If the geometry of the heat sink fins is poorly designed, the effective radiating fin surface can be greatly reduced.

At the same time that heat is being transferred by radiation, natural convection is also taking place. The amount of heat that can be transferred by natural convection depends on:

- Temperature difference between the surface and the surrounding air
- Dimensions of the surface
- Orientation of the surface
- Spacing between adjacent surfaces
- Altitude (air density)

The temperature dependence of the amount of heat transferred for natural convection is approximately the same as that of radiation. The surface dimensions and the orientation of the heat sink in the system are key factors in the amount of heat transferred by natural convection. The surface dimension most critical to natural convection is the length along which the air moves. It is equal to the vertical height of a vertical surface, or for a horizontal orientation, since air flows in from all sides, it is equal to the geometric mean of the actual length and width of the surface.

$$\text{Effective Surface} = \frac{\text{Width X Length}}{\text{Width + Length}} \qquad (11)$$

Cooling by natural convection is more efficient for vertical surfaces or for horizontal surfaces facing upward than for horizontal surfaces facing downward. In the system, the heat sink should be oriented with the fins (or heat transfer surface) vertical rather than horizontal. Horizontal mounting (facing down) will reduce the heat transfer capability by one-half.

The spacing between two heat transfer surfaces must also be examined. If the surfaces are too close together, the air flow from each surface will interfere with the air flow of the other. In most cases the surfaces or fins can be brought to about one-half an inch apart without much interference.

The amount of cooling by natural convection is also dependent on the air density (a function of altitude). As the altitude increases, the amount of heat transferred goes down. At 10,000 feet, natural convection is only 80% as effective as it is at sea level.

Forced Air Cooling

If the heat dissipated by the amplifier exceeds 50 watts, the size of the fins required for natural convection and radiation becomes excessively large. The heat transfer capability can be increased by an order of magnitude by using forced air cooling. The price that is paid for this improvement is increased system complexity, reduced system efficiency and increased vibration and acoustical noise.

The design of a forced air cooling system consists of two parts; the design of the cooling fins and the choice of the fan or blower. The two steps must be done concurrently because the amount of heat transfer obtained and the pressure required to force air through the cooling fins depends on the air flow and fin geometry. By using previously discussed methods, the temperature of the cooling fin surface can be computed for a given power dissipation. The equation below is the basic formula for determining the surface temperature of the cooling fin.

$$T(Fin) = \Delta T(Air) + \Delta T(Fin\text{-}Air) + T(Ambient) \quad (12)$$

where

$T(Fin)$ = temperature of cooling fin surface (^{o}C)

$\Delta T(Air)$ = temperature rise of air as it absorbs heat from cooling fins (^{o}C)

$\Delta T(Fin\text{-}Air)$ = temperature rise of cooling fin above air in cooling duct (^{o}C)

$T(Ambient)$ = temperature of inlet air (^{o}C)

The temperature rise of the air as it absorbs heat from the cooling fins, $\Delta T(Air)$ is independent of the geometry or dimensions of the fin. It depends only on the power transferred and the total air flow.

$$\Delta T(Air) = 1.73Q/f \quad (13)$$

Q = total power being transferred (watts)

f = total air flow through the fins (CFM)

The temperature rise required to transfer a given amount of heat from the fins depends on the velocity of the air moving past the fins. The air velocity depends on fin geometry, dimensions and air flow rate.

$$\Delta T(Fin\text{-}Air) = 140WQ/n^2Z^2f^8L \quad (14)$$

where

W = width of ducts (inch)

Z = length of fins above base (inch)

L = length of fins along direction of air flow (inch)

n = number of ducts in the air flow

see figure 10

Several assumptions have been made to simplify the forced air flow

calculations. The average air temperature is at 20 °C, the system is at sea level, the air flow is turbulent, and the ducts are narrow compared to the height above the base.

As stated before, the fin geometry and dimensions effect the amount of pressure required. Minimizing the air pressure drop is an important step in reducing fin temperature. The total pressure into which the fan must work is determined not only by the cooling fins, but by other components of the system. These include the entry port, duct filter, inlet ducting, outlet ducting and the exit port. Only the cooling fins contribute to the cooling, but all contribute to the pressure drop. By using the following relationships, the pressure drops can be minimized.

- Pressure drop is proportional to 1/(cross-sectional area)$^{1/2}$
- Pressure drop is proportional to f^2 (air flow)
- Pressure drop is proportional to L (length of air flow path)

In a well designed system, the major part of the pressure drop is due to the cooling fins. The geometry and the dimensions of the fins effect both heat transfer and pressure drop, and generally a compromise between the two is used. The approximate formula for determining the pressure drop through the cooling fins as a function of the fin dimensions is:

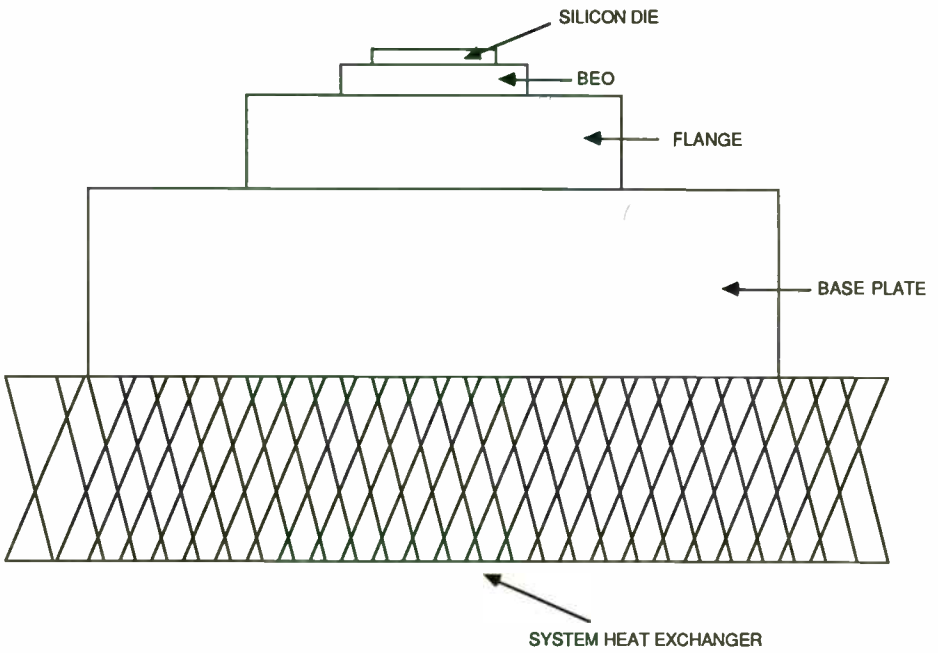
$$P = ((f/n)/(WZ))^2[1 + .01(L/W)] \times 10^{-3} \quad \text{(inches of water)} \quad (15)$$

See figure 10

Comparing equations 14 and 15, the fin temperature depends directly on the duct width but only slightly on the duct height. However, the pressure drop is equally effected by both duct width and height. Therefore, the fin height can be increased (to the extent of the available space) to reduce the pressure drop without significantly degrading the heat transfer capability of the fins. The above equation should be used as a guideline to determine the effects of fin dimensions on pressure drop.

Good amplifier performance is dependent upon the thermal design of the amplifier. Excessive junction temperatures will reduce the performance and shorten the life of the amplifier. Utilizing the above guidelines will help to ensure a thermally well designed system.

1. M. Flahie, "Reliability and MTF - The Long and Short of it," Microwaves July 1972
2. R. Frey, M. Kane, "Temperature Effects Examined for Microwave Power-Transistors Performance and Thermal Design Consideration, MSN & CT November 1985
3. "Thermal Time Constant for High Power Pulsed Transistors - Power Flow Calculations", Acrian Application Note
4. Allan W. Scott, Cooling of Electronic Equipment (New York: Wiley, 1974)
5. Gordon, N. Ellison, Thermal Computations for Electronic Equipment (New York: Van Nostrand Reinhold, 1984)



$$\theta_{Jc} = \theta\,(\text{Silicon}) + \theta\,(\text{BeO}) + \theta\,(\text{Flange}) + \theta\,(\text{Baseplate}) + \theta\,(\text{System})$$

Figure 1: Thermal Path of Amplifier

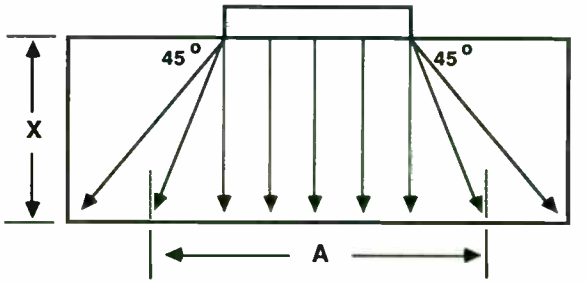


Figure 2: Themal Path with Constant Spread Angle

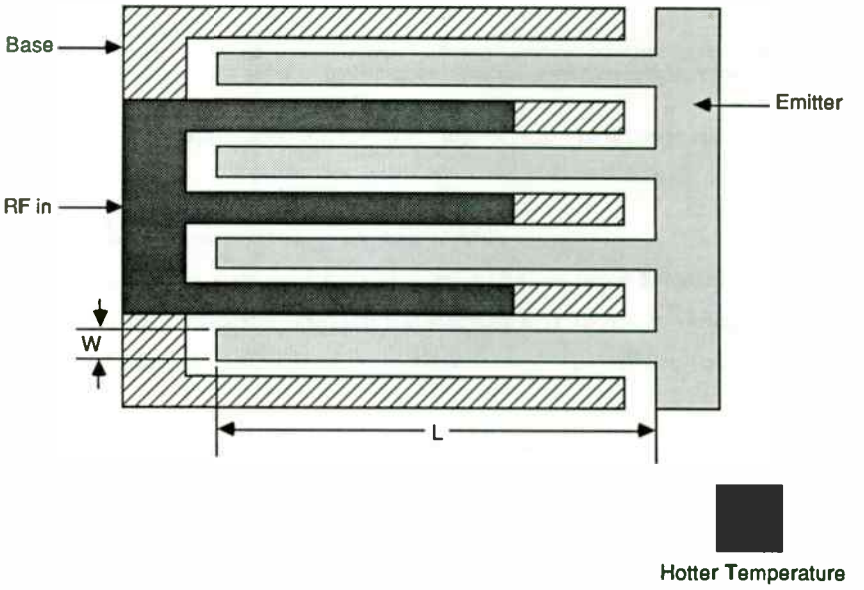


Figure 3: Non Uniform Temperature Gradients Due To Physical Size of Die

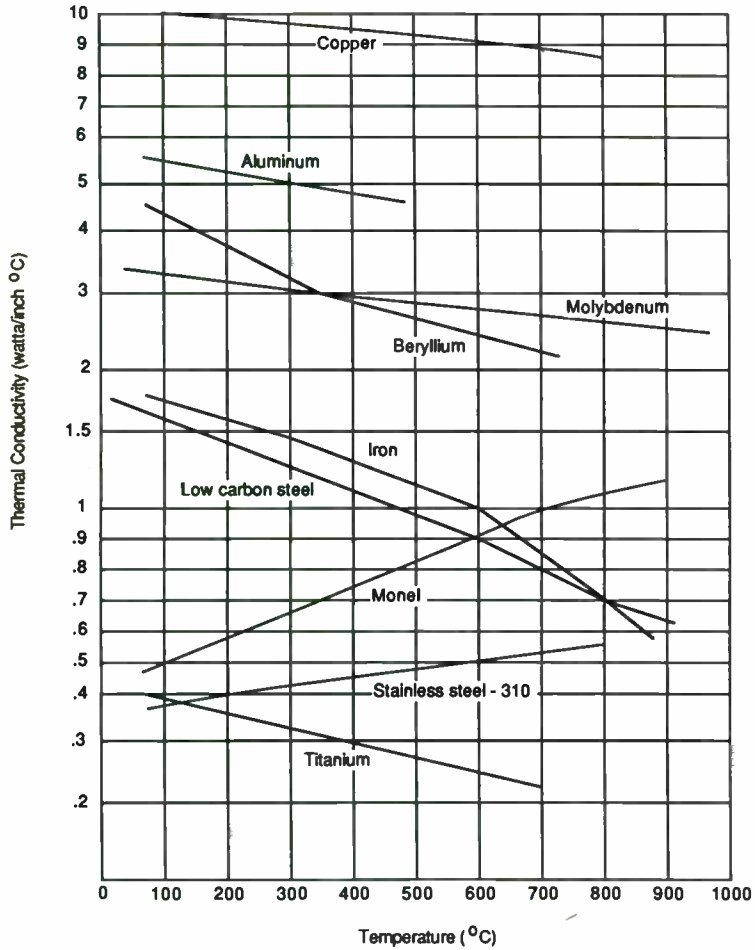


Figure 4:
Thermal conductivity of metals as a function of temperature

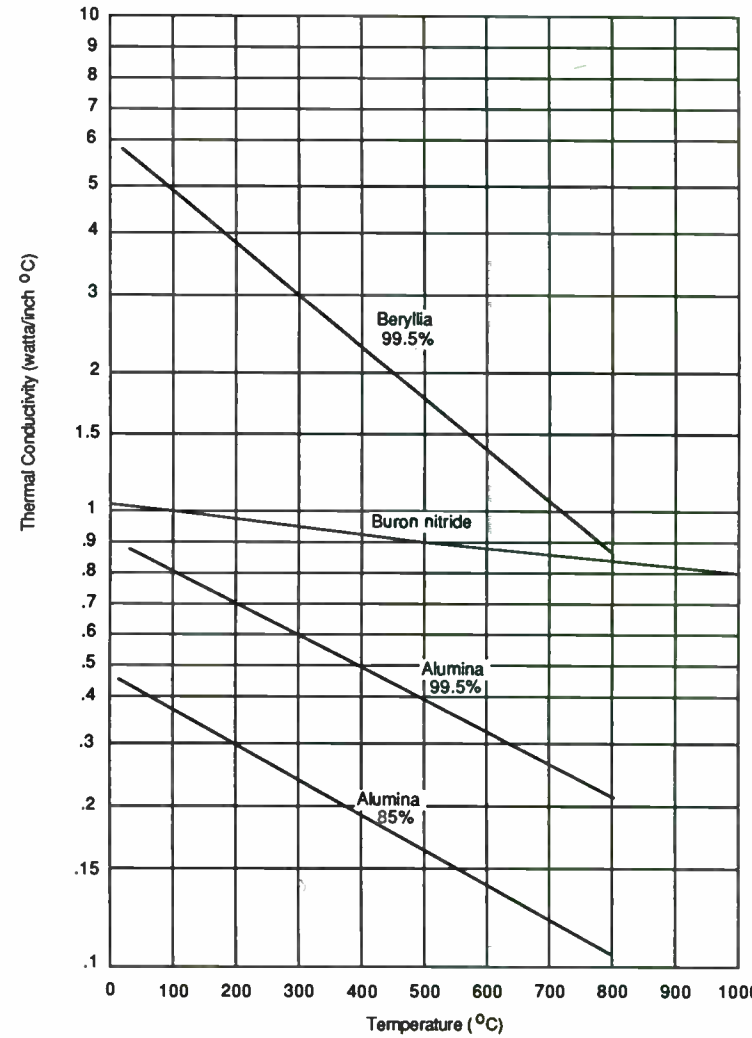
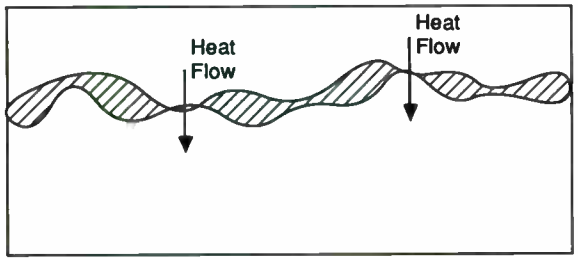
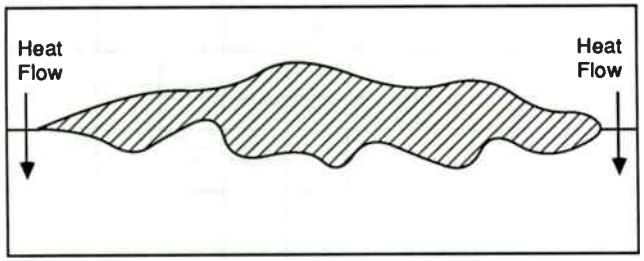


Figure 5:
Thermal conductivity of ceramic insulators as a function of temperature



A: SMOOTHNESS



B: FLATNESS

Figure 6: Microscopic View of Surface Interfaces

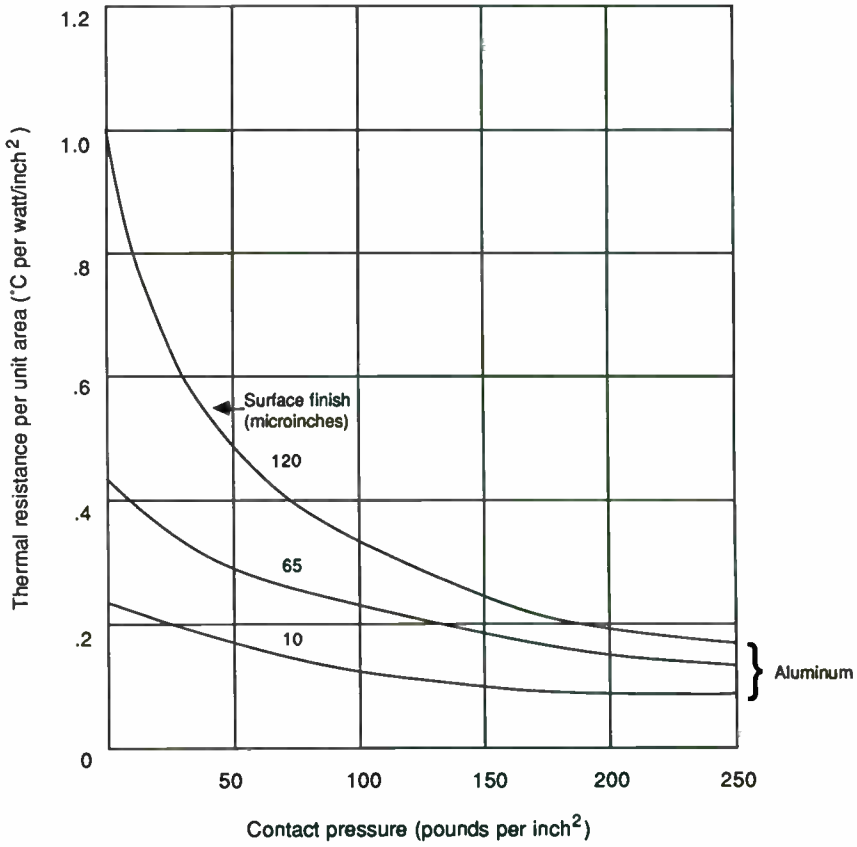


Figure 7: Thermal Resistance VS Contact Pressure

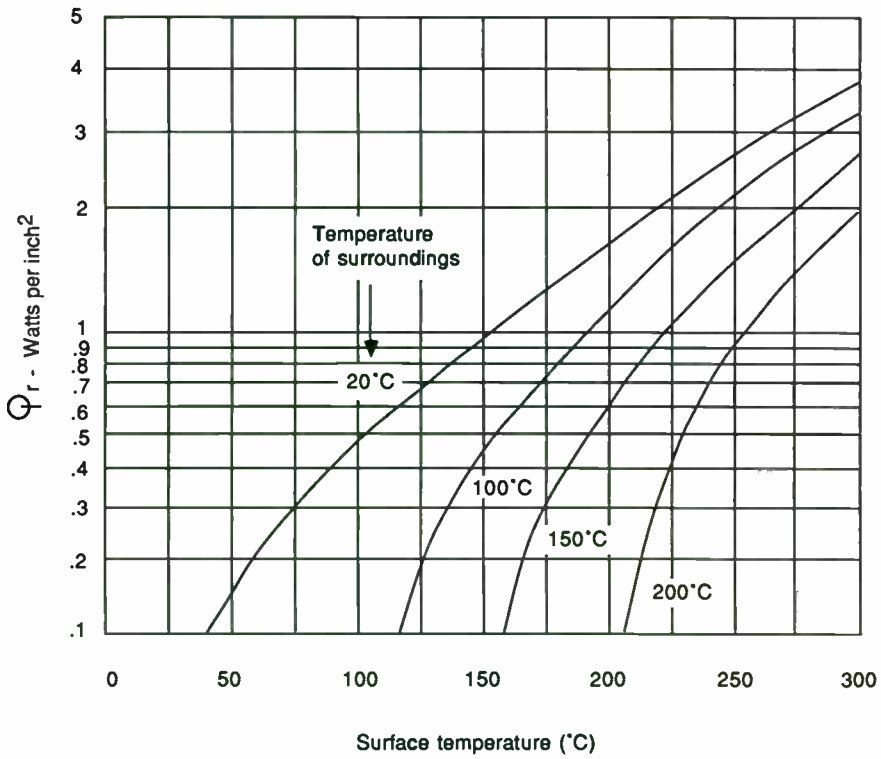


Figure 8: Power Transferred by Radiation VS Surface Temperature

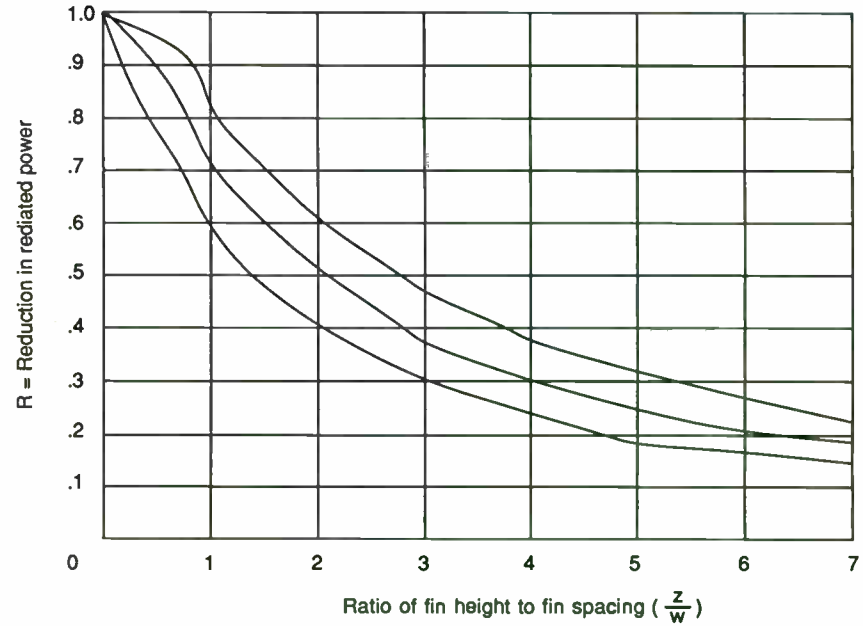


Figure 9: Radiation Shielding Effect of Adjacent Surfaces

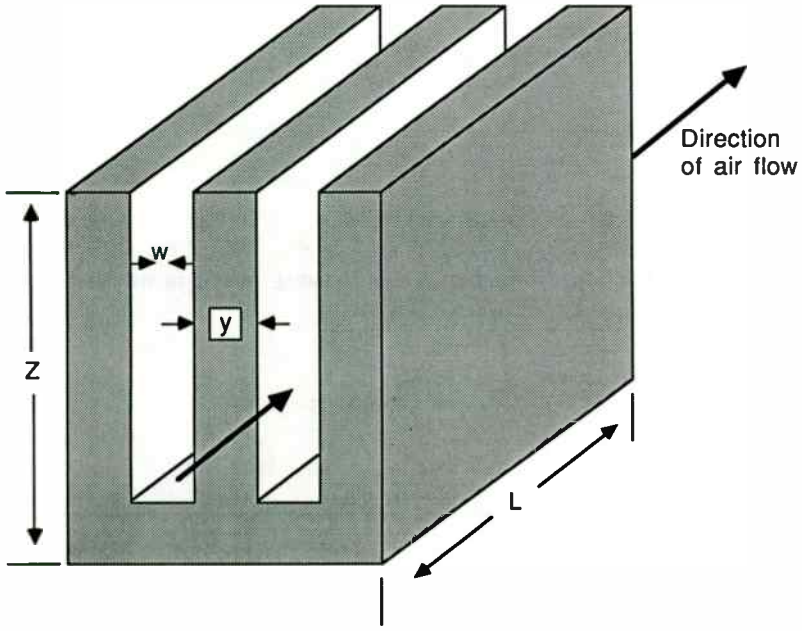


Figure 10: Geometry of Forced Air Cooled Heat Sink

Table 1:
Thermal Conductivity of Materials Commonly Used in Electronic Equipment

Material	Thermal Conductivity (Watts/Inch °C) at 100°C
Metals	
Aluminum	5.5
Beryllium	4.5
Beryllium copper	2.7
Brass (70% copper - 30% zinc)	3.1
Copper	10.0
Gold	7.4
Iron	1.7
Kovar	0.42
Lead	0.87
Magnesium	4.0
Molybdenum	3.3
Monel	0.50
Nickel	2.3
Silver	10.6
Stainless steel 321	0.37
Stainless steel 410	0.61
Steel, low carbon	1.7
Tin	1.6
Titanium	0.40
Tungsten	5.0
Zinc	2.6
Semiconductors	
Gallium arsenide	1.5
Silicon (pure)	3.7
Silicon (doped to resistivity of .0025 ohm-cm)	2.5
Insulators	
Still air	0.0007
Alumina (99.5%)	0.70
Alumina (85%)	0.30
Beryllia (99.5%)	5.0
Beryllia (97%)	4.0
Beryllia (95%)	3.0
Boron nitride (hot pressed)	1.0
Diamond	16.0
Epoxy	0.005
"Thermally conducting" expoxy	0.02
Glass	0.02
"Heat sink compound" (metal oxide loaded epoxy)	0.01
Mica	0.018
Mylar	0.005
Phenolic	0.005
Silicone Grease	0.005
Silicone Rubber	0.005
Teflon	0.005

Table 2:
Temperature difference required to conduct 10W of heat through a 1 in. length of material with a cross section area of 1 in ^{2}.

Material	Temperature Difference (°C)
Copper	1
Aluminum	2
Low carbon steel	6
Stainless steel #321	25
Beryllia ceramic	2
Alumina ceramic	15
"Thermally conducting" epoxy	500

Table 3: Emissivity of Materials

Surface	Emissivity
Commercial aluminum (polished)	.05
Anodized aluminum	.80
Aluminum paint	.27 to .67
Commercial copper (polished)	.07
Oxidized copper	.70
Stainless steel (polished)	.17
Stainless steel (with heavy oxide)	.85
Rolled sheet steel	.66
Air drying enamel (any color)	.85
Oil paints (any color)	.92
Lamp black in shellac	.95
Varnish	.90
Zirconium coating on molybdenum	.65

A Discussion on Test Methods for Shielding Effectiveness Measurements of Large Shielded Facilities

by

Antonio L. Cardenas
President
Advanced Measurement Systems, Inc.
1115 West Broad Street
Falls Church, VA 22046

Today's RF shielded enclosure testing standards were originally established many years ago. MIL STD 285, "Attenuation Measurements for Enclosures, Electromagnetic Shielding, for Electronic Test Purposes, Method of", dated 25 June 1956 is one of the original standards. This standard specified equipment to be used, measurement procedures, and attenuation levels required. Other shielding requirements and test procedures were provided in NSA 65-6, "National Security Agency Specification for R.F. Shielded Enclosures for Communications Equipment, General Specification" dated 30 October 1964 and NSA 73-2A "National Security Agency Specification for Foil RF Shielded Enclosures" dated 15 November 1972. These standards have been used extensively in the preparation of specifications for procurement of RF enclosures and their testing. The context of these standards have often been misinterpreted or misused by architectural engineering firms designing shielded facilities, general contractors installing the facilities and testing companies certifying their compliance. This has lead to improper construction and testing of many shielded enclosures.

Originally, the majority of the RF shielded enclosures were small rooms constructed as test labs and communications equipment enclosures. Testing of these smaller enclosures was fairly straight forward when using the old methods. Many of today's enclosures are large facilities or entire buildings. Their applications vary from test chambers for low level sophisticated electronic measurements to protection from Nuclear Electromagnetic Pulse (EMP) phenomena. There are a variety of materials and construction methods being used for these facilities, such as foil systems (Aluminum or Copper), steel demountable systems, all welded systems, and various lapped and/or taped systems. Each of these systems can have unique testing requirements depending on their size, location, application and method of construction. Regardless of the size or application of the shield, proper and complete testing is the only way to guarantee that the shielding system is providing the proper attenuation. This paper will discuss alternate methods for testing today's facilities. To give you an idea how rooms have been tested in the past, the following background information is provided.

MIL STD 285 was originally part of a government specification (MIL-S-4957A) for screen mesh enclosures. It was written in

the early 1950's to procure some wire mesh screen rooms for a research project. The testing portion of this specification was later published as MIL STD 285 in 1956. The test methods specified by this document became the standard for determining the performance of all RF shielded enclosures. These test methods provide the technician with antenna placements for conducting reference level measurements, a detailed measurement procedure and a description of test equipment to be used. The attenuation requirements for these enclosures were 70 dB for magnetic field from 150 KHZ to 200 KHZ, 100 dB from 200 KHZ to 18 MHZ in the electric field, and 100 dB at 400 MHZ. Note that these enclosures were only tested at 400 MHZ in the planewave.

In 1964 the National Security Agency published NSA 65-6. This document was a general specification that the agency issued in order to standardize the requirements for its RF shielded enclosures. A test procedure for measuring the shielding effectiveness (SE) of the enclosures was also included. The procedures are basically the same as those in MIL STD 285 except for some slight variances. The test equipment in NSA 65-6 is specified only by the frequency and dynamic range requirements, and the antenna positioning for the magnetic field measurements are coaxial instead of co-planer. The attenuation requirements for NSA 65-6 are illustrated in Figure 1A. Note that NSA 65-6 requires testing at several frequencies in the planewave, up to 10 GHZ.

NSA 73-2A, published in October of 1972 is another shielded enclosure specification with testing requirements included. This document specifies aluminum foil for an RF shielding system. The test methods described in this document are basically the same as for NSA 65-6, except for the distances between the test antennas and the shield. Also, the attenuation requirements for the foil system are less than that specified in NSA 65-6. These requirements are shown in Figure 1B.

The following paragraphs will present statements made in these documents referencing the methods in which the shielding measurements should be performed. The interpretation of the statements and their application to today's shielding systems and test procedures will be reviewed.

MIL STD 285 states that, for the magnetic field, a measurement shall be taken on all four sides of the enclosure, and the minimum attenuation recorded. The electric field requirement is similar except that "each side" replaces "all four sides." For the planewave measurement, it simply states that several readings shall be taken.

Measuring the SE of the original wire mesh screen rooms at one location per side may have been sufficient, especially if the side or wall was one single sheet of mesh material. However, measuring an enclosure with multiple seams or joints using this

test method would not provide sufficient data to make this assessment. For example, a single RF tight seam on a wall with 10 vertical seams does not imply that all the seams are going to be leak-free. Most panel systems usually have a vertical seam every 4 feet and, depending on the height of the room, as many as 2 or 3 horizontal seams. Each of these joints or seams is a potential leakage point for RF energy. This holds true for most systems, although the distances between seams may vary with design.

NSA 65-6 states, "Magnetic fields shall be measured with the loops parallel to a wall panel...", and, "Planewave measurements shall be taken through a wall panel...". It also states that "Leakage checks must be made all around the door frame, through accessible joints, around the filters and all around the air ducts. In addition, the magnitude and location of the maximum signal level emanating from the enclosure should be found by moving the antennas to at least four locations, preferably on different walls." Depending on how one wants to interpret these statements, the number of test points could be the total number of seams plus penetrations, or only the number of penetrations and 1 point on each wall.

The test methods and specified test points implied by both of these documents would not provide adequate representation of a large shields' performance. When specifying attenuation requirements and methods for verifying the shields compliance, the documents by themselves are not sufficient. Delineation of locations to be tested and specific testing requirements must be added in order for the shield to be tested properly.

Why do these documents fall short of their intended use? The reason is, they were written for two specific types of small enclosures. One is the wire screen mesh room and the other is the demountable panel system. For smaller rooms (i.e. 10' x 10' x10'), the testing methods specified would probably be sufficient. Once the size of the enclosure increases, the number of test points measured must also increase. All seams, joints, penetrations, doors and access panels must be tested to ensure that the shield is providing the required attenuation. Sometimes there are extenuating circumstances that prohibit testing of all areas of the shield. This could be anything from inaccessibility of the test points to time constraints dictated by construction schedules. Regardless of the circumstances, the enclosure must be tested at as many locations as necessary to ensure the shield is RF tight.

Next we discuss the actual test methods and the type of test equipment used to conduct SE measurements.

Today's more sophisticated receivers and synthesized signal sources make testing much more accurate and timely. The test equipments specified in MIL STD 285 are certainly outdated and are lucky to be found, but, I have seen some of equal technology in use for these measurements. Figure 2-1A through

2-1D are illustrations of the equipment configuration for a typical NSA 65-6 test. Table 1 is a list of new equipment meeting the requirements for this test. The system configuration for the magnetic field and electric field measurements is reversed from that of the planewave in this illustration. This is because the test locations that these figures represent had high level ambient signals outside the shield. This is common for most installations. The two NSA documents prefer the transmitter inside the room for all measurements, except in the case mentioned above. MIL STD 285 requires the transmitting system to be outside the shield, and the receiver inside for all measurements. Both methods are accurate. The one drawback to using the MIL STD 285 configuration is that standing waves are developed inside the room when testing high frequency E-field and Planewave. The test signal leaks into the room and is reflected around the shield, which makes it difficult to locate the source of the leak, and may affect measurement accuracy. The configurations shown here provide the best results.

When performing magnetic field measurements it is virtually impossible to cover every inch of seams involved. This is because the magnetic field only illuminates a very small portion of the shield and does not propagate very far from its source. To insure the shield is performing properly, the following locations should be tested.

In the magnetic field, every vertical seam on the walls should be tested at one location if the wall is less than ten feet high. If over 10 feet high it should be checked at a minimum of every 10 feet. A similar sampling of the seams on the roof and floor will be checked if readily accessible. Electric field measurements shall be performed at the same test locations as those for the magnetic field.

The planewave signals propagate very readily and illuminate a larger area of the shield. Therefore, fewer test points are required to scan the entire shield. The test points for the planewave measurements should be located a minimum of every 20 feet along a wall that is less than 20 feet high. The transmitting antenna will be positioned at the center of this area while the receiving antenna is scanned over every seam. The ceiling and floor will be tested in the same manner if readily accessible.

The following paragraphs describe an NSA 65-6 test procedure using the equipment listed and how the documentation of these measurements should be reported.

Magnetic Field

In accordance with NSA 65-6 the transmit and receive magnetic loop antennas are placed parallel to each other at a distance of twenty-four inches plus the thickness of the shield under test. The transmit antenna is connected to the signal source

and power amplifier, which make up the transmitting system, and the receive antenna is connected to the pre-amp and the spectrum analyzer, which make up the receive system. The output of the transmit system is a CW signal at the frequency under test. The spectrum analyzer is tuned to this frequency and the signal level measured. This level, measured in dBm, is recorded along with the settings of the transmit system. The settings of the systems are carefully noted to insure that the necessary dynamic range of 6 dB more than the minimum attenuation to be measured is achieved. This procedure is repeated at each of the test frequencies. With all reference levels recorded, and system settings noted, the test antennas are placed 12 inches on either side of the shield under test. The enclosure is sealed and the receive system tuned to the test frequency. The maximum level measured is recorded and subtracted from the reference level to obtain the attenuation in dB. If the signal is not measurable, a reading equal to 3 dB above the noise floor of the receive system is entered on the data sheet and used to calculate the shielding effectiveness. (Note: Both the transmitter and the receiver in this measurement have memories for storing front panel settings. Instead of dialing up each test frequency for the measurements, the settings can be stored and then recalled anytime during the entire measurement process.)

Electric Field

The same procedure used for the magnetic field measurements is used for the electric field measurements. The 41 inch electric field vertical whip test antennas are substituted for the magnetic loops. All distances and testing procedures remain the same.

Planewave

The reference levels for the planewave test frequencies are measured in the following manner. The test antennas are positioned at least two wavelengths of the test frequency from each other plus the thickness of the shield, plus 2 inches. A minimum of 72 inches of separation distance from the transmit antenna to the shield under test is required (where practical). The transmit and receive systems are adjusted in the same manner as with the other two measurements. The reference levels are recorded and the transmitter output settings noted. The receiving antenna is then positioned opposite the transmit antenna on the other side of the shield. The shielded enclosure is sealed. The receive antenna is oriented for the maximum signal level measured over the area being illuminated by the transmitter, and is never positioned closer than 2 inches from the shield. The received level is measured and that number subtracted from the reference level to obtain the RF attenuation of the shield at that location. These test procedures are repeated for all frequencies and test locations specified.

Data Presentation

The data and test results can be presented in the final report in either tabular or graphic form. A table should contain the test frequencies and the minimum attenuation measured throughout the enclosure along with the attenuation requirements of NSA 65-6. Copies of the data sheets containing all the test locations with attenuation factors for each frequency and field tested should be included in the report. A diagram of the enclosure with all penetrations, filters, waveguide air vents and test locations should also be included. The graphic data compares the attenuation factors measured and the attenuation factors specified by NSA 65-6 for each of the frequencies tested. The minimum attenuation measured at a test point in an area such as a wall or door section is used to represent the overall attenuation of that area. This minimum measured attenuation for an area should be the data presented in the graphs.

Accuracy of Measurement

Using the calibration procedures outlined, and the direct method of measurement, the only factors affecting the accuracy of the measurement will be the antenna positioning and the reading of the signal level directly from the analyzer display. The antenna gains, cable losses and amplifier gains are calibrated out of the measurement system when the reference levels are established.

This test method can provide you with the required data to determine a shields compliance with a certain specification,, and can be used for standard enclosures.

Another thing to consider is, what are you actually testing for? The shielding effectiveness of the material that the shield is constructed of, or are you trying to find out if there are any RF leaks in the seams and penetrations. If you are testing a larger facility or one on an upper floor, how do you gain access to the walls? The engineer who designed or specified the shielding material undoubtedly knows that the material in itself will meet the specified attenuation. For example, lets take a shielded facility designed to provide EMP protection for a large telecommunications center. The shield is constructed of 1/8" plate steel and measures 100' x 100' x 80' high. Eighth inch plate steel is not going to leak RF at any point on the sheet. The joints, seams, and penetrations are the items actually under test.

To use the standard methods described above, it is necessary to gain access to the outside walls. Short of using a window washing bucket or portable elevator this is virtually impossible. Locating and positioning scaffolds and "cherry pickers" is another way, but can be time consuming and expensive. To test this facility properly and get an overall indication of its SE, and find the RF leaks, the following procedures can be used. (Note: For EMP purposes the frequencies

under consideration are from 100 MHZ to 500 MHZ planewave.) This procedure will describe a planewave measurement of this shield.

Position the transmit antenna at ground level at a distance from the building equal to the height of the building. In this case 80 ft. Direct the antenna to the center of the area you would like to test. For a building this size, 4 test points, one in each quadrant of the wall, should be sufficient. The large beam width of the log periodic or log spiral antennas should cover this area with an equal amount of field intensity. Prior to this you must take a reference level measurement as described in the test procedures above, except that you must perform them with the antennas at the appropriate distance (80 ft). You must also be aware that a certain amount of dynamic range may be lost due to the distance involved and you may have to compensate for it by boosting your transmit power.

Once the transmitter is set, you may scan the area under test for RF leaks and document your results. This test method can cover more area in less time than the standard method, and provide the same results. This method can also be used for testing at microwave frequencies and on facilities that are located above the 1st floor.

Magnetic field and electric field measurements of a facility like this using conventional SE measurement techniques is almost impossible. In lieu of testing in these fields at places that are not accessible from the outside, the Shielded Enclosure Leak Detection System (SELDS) tester is a very reliable testing method. This method drives a modulated test current on the surface of the shield. The test leads of the SELDS transmitter are attached to two opposing points on the shield so the current is driven across the area under test. The maximum surface area to be tested should be kept below 4000 square feet. It is recommended that the leads be attached to the outside of the shield so measurements can be made inside the shield. Once the transmitter is attached and operating, the detector, which is a hand-held battery operated receiver, can be probed along each welded seam to determine if any magnetic field anomalies are present. The receiver detects these anomalies when a high level magnetic field is generated around holes or cracks in the shield. The demodulated signal is metered on the detector to indicate SE in dB. This type of test is not as accurate as an actual SE measurement, so if a leak is detected your best bet is to fix it.

In summary, one of the most important functions that will be required when an RF shielded facility is installed, is the shielding effectiveness measurements. These measurements will certify that the room will meet the intended attenuation requirements. If the specification for testing is not written

properly, the validity of the test is automatically in question. The standard documents that were written for specific enclosures and which are now outdated, must be supplemented to provide testing that will control the quality of shielding systems being installed and ensure the continuing performance of those that already exist.

TABLE 1

TEST EQUIPMENT LIST

MODEL	DESCRIPTION	FREQUENCY RANGE	CAL. DATE
TEK-494P	SPECTRUM ANALYZER	1 KHZ - 21 GHZ	
MA-2022	SIGNAL SOURCE	10 KHZ - 1 GHZ	
HP-8672A	SIGNAL SOURCE	2 GHZ - 18 GHZ	
WT-188	SIGNAL SOURCE	1 HZ - 4 MHZ	
ENI 2100	POWER AMPLIFIER	10 KHZ - 12 MHZ	
KH 7500	POWER AMPLIFIER	DC - 1 MHZ	
BPA 1000	PRE-AMPLIFIER	10 KHZ - 1 GHZ	
EMCO-6505B	MAGNETIC LOOP(T)	1 KHZ - 30 MHZ	
EMCO-6505A	MAGNETIC LOOP(R)	1 KHZ - 30 MHZ	
RVR-25	E-FIELD ANTENNA(	1 KHZ - 30 MHZ	
WA-75	E-FIELD ANTENNA(T)	1 KHZ - 30 MHZ	
EMCO-3101	LOG SPIRAL ANT.(T)	100 MHZ - 1 GHZ	
EMCO-3101	LOG SPIRAL ANT.(R)	100 MHZ - 1 GHZ	
SA-12.4	HORN ANTENNA(T)	8.0 GHZ - 12.4 GHZ	
SA-12.4	HORN ANTENNA(R)	8.0 GHZ - 12.4 GHZ	
RG-214	CABLES		
RG-223	CABLES		
TR-1	TRIPODS(2 EACH)		
TFLS	Times Fiber Low Loss Cables		

SHIELDING EFFECTIVENESS MEASUREMENTS

Test Specifications NSA 65-6 (100 DB)

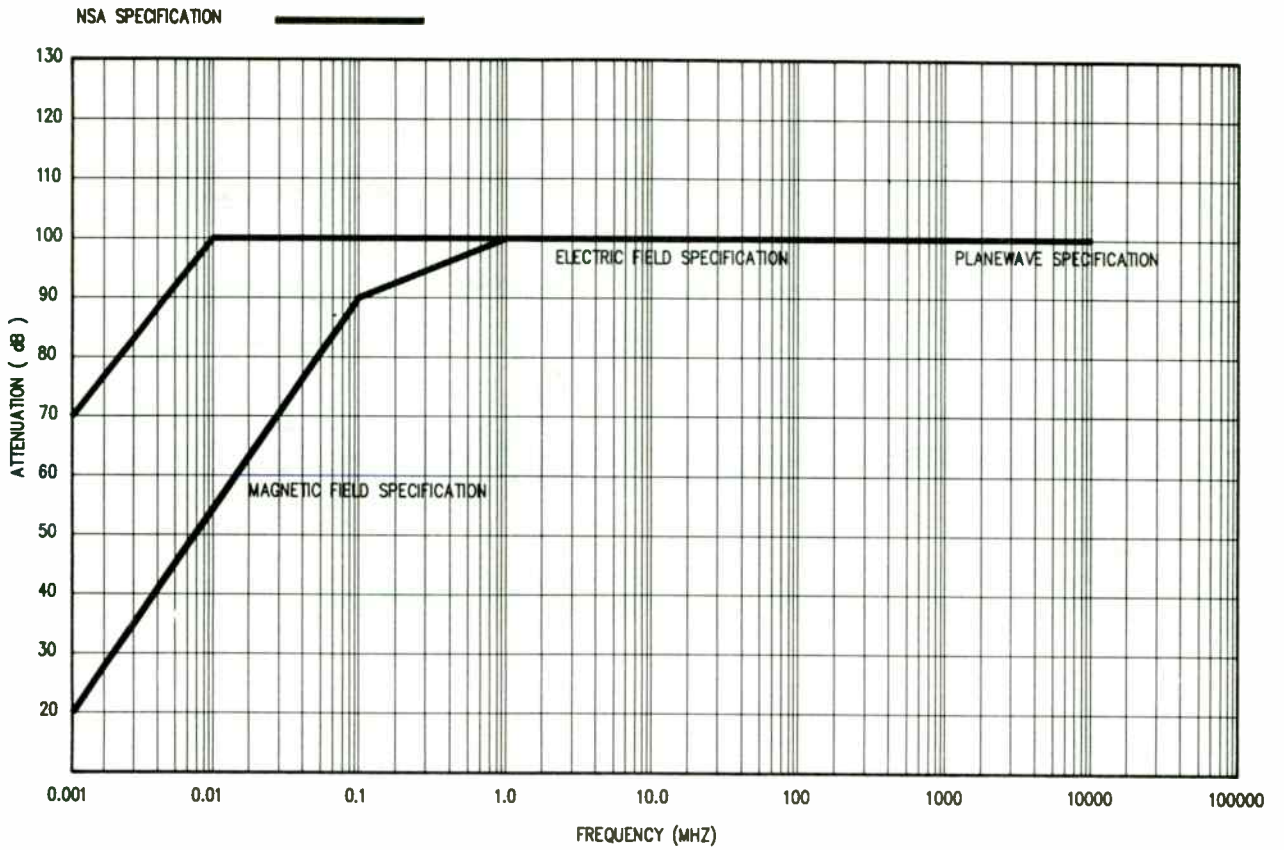


FIGURE 1A

421

SHIELDING EFFECTIVENESS MEASUREMENTS

Test Specifications NSA 73-2A (50 DB)

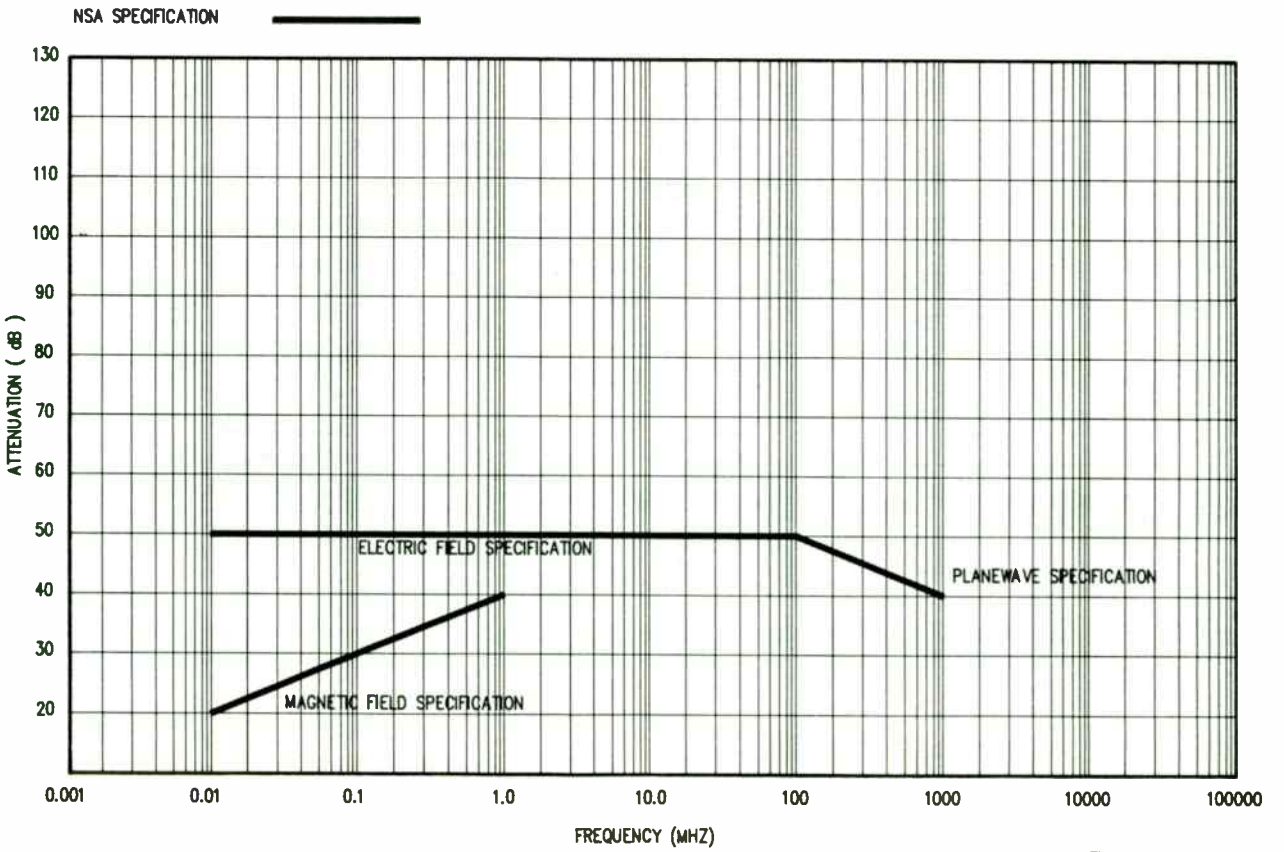


FIGURE 1B

MAGNETIC FIELD TEST EQUIPMENT

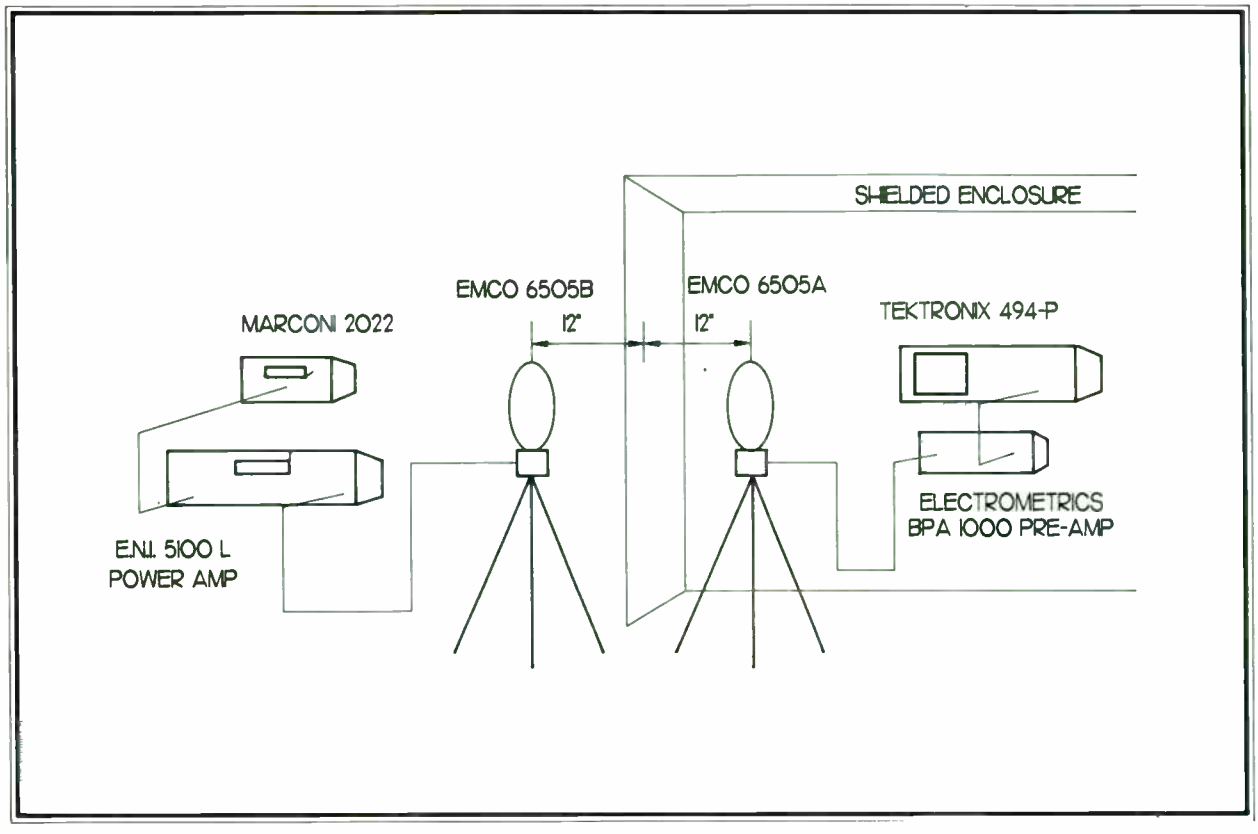


FIGURE 2-1A

ELECTRIC FIELD TEST EQUIPMENT

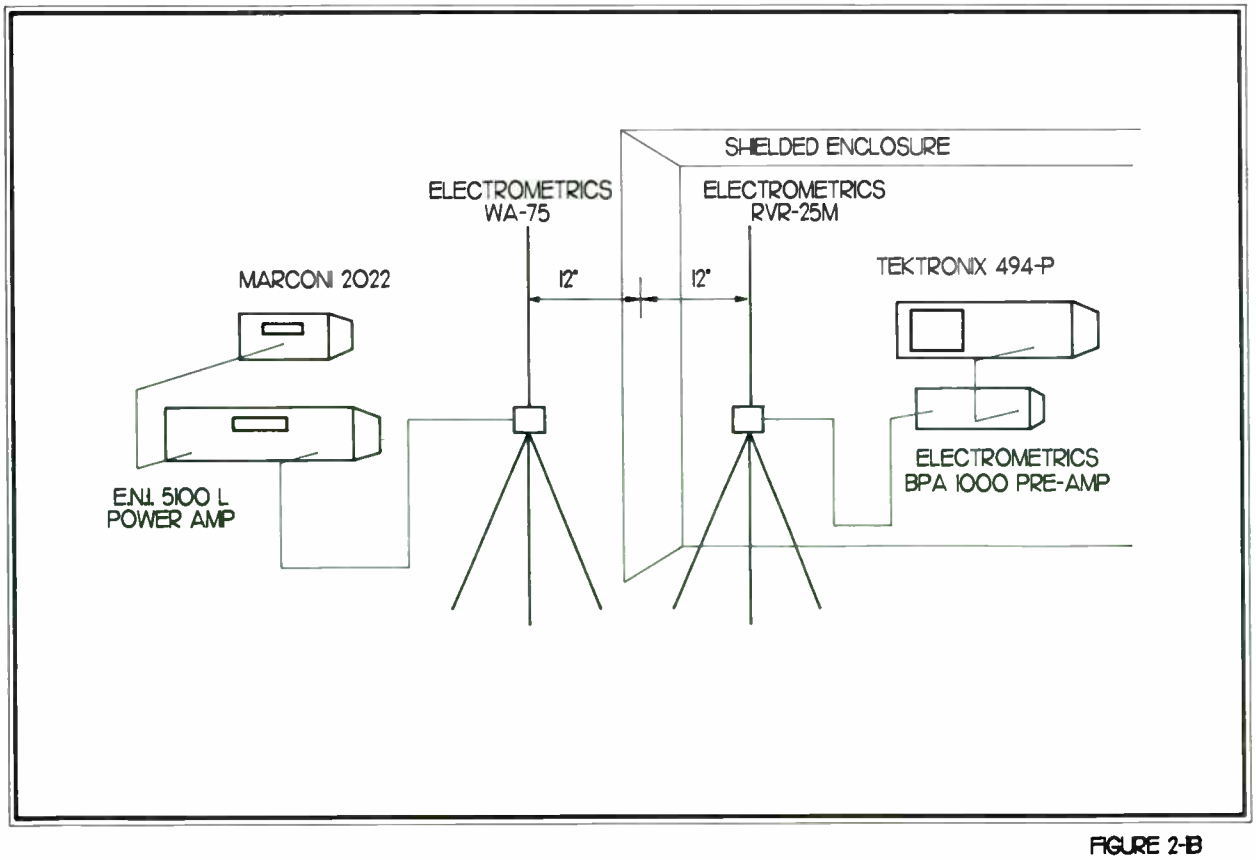


FIGURE 2-1B

PLANEWAVE (THRU 1 GHZ) TEST EQUIPMENT

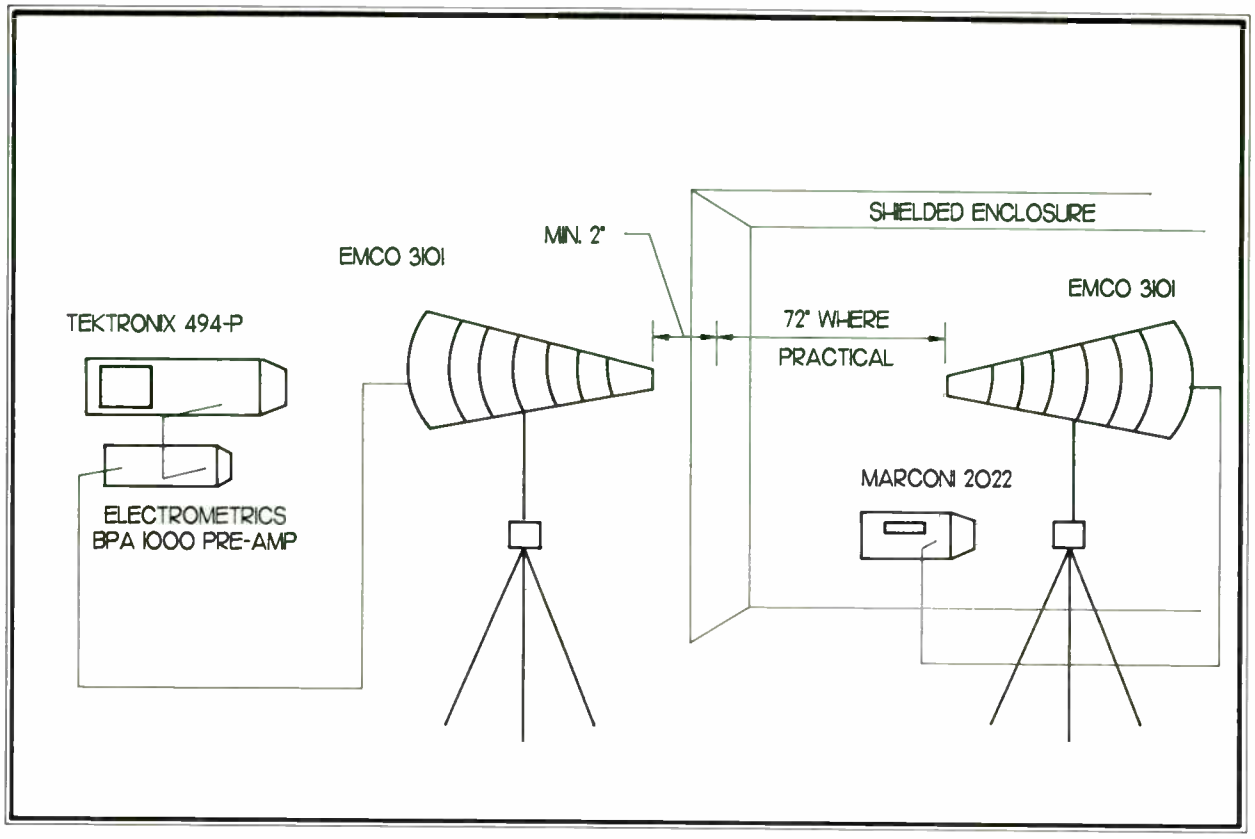


FIGURE 2-1C

PLANEWAVE (10 GHZ) TEST EQUIPMENT

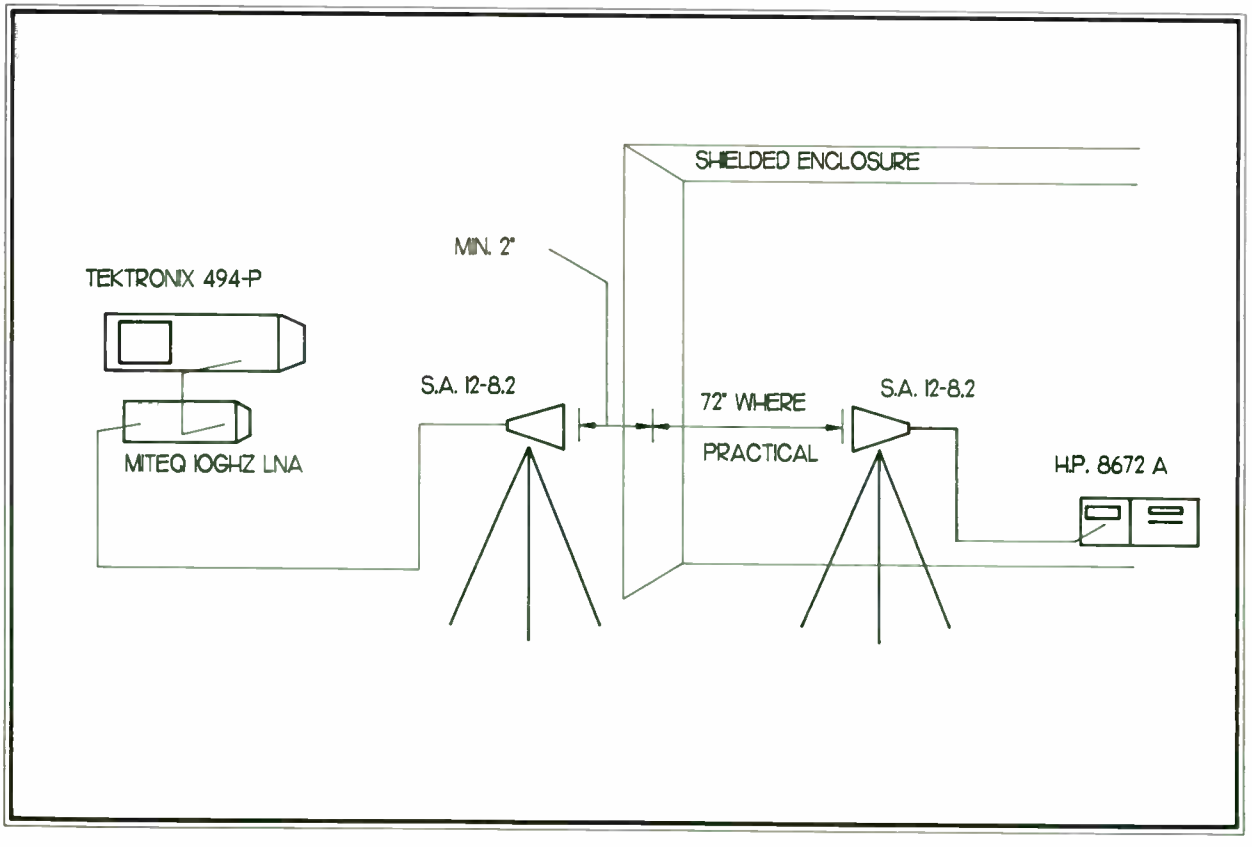


FIGURE 2-1D

MICROPROCESSOR INTERFERENCE TO VHF RADIOS

Daryl Gerke, PE
Kimmel Gerke & Associates, Ltd.
1544 North Pascal
St. Paul, MN 55108
612-330-3728

ABSTRACT

Interference to communications systems from microprocessors is not new. In fact, this interference is the basis for the current FCC regulations that limit emissions from digital equipment. Unfortunately, these regulations were designed to protect television receivers, and are often inadequate to protect more sensitive land mobile communications receivers.

This paper examines the current FCC regulations and compares them to typical threshold sensitivity levels for communications equipment. The paper then discusses a case history of a microprocessor control system interfering with a land mobile communications receiver. Finally, several recommendations are made.

INTRODUCTION

When analyzing and trying to understand an electromagnetic interference (EMI) situation, it is helpful to first divide the problem into categories. At the highest level, the three categories are the source, the victim, and the coupling path. Secondary categories typically address the coupling path: radiated, conducted, cross-talk, common impedances, etc.

In this paper, the assumptions are as follows: the sources are microprocessor based digital computers, the victims are VHF/UHF communications receivers, and the primary coupling path is direct electromagnetic radiation which is picked up by the receiver antenna.

MICROPROCESSORS - THE SOURCE

In characterizing microprocessor based systems as a source of emissions, several alternate approaches are available. One can make predictions, one can make measurements, or one can use take a "systems approach" and use existing equipment limits as baseline.

In this paper, we will use the "systems approach", and we will use the Federal Communications Commission limits for digital computers, as specified in Part 15J of the FCC rules and regulations. This is a reasonable starting point, since "personal computer" systems must meet those requirements, and since these "personal computers" are being incorporated as controllers in larger systems.

In the late 1970s, it was apparent that the proliferation of home and business computers was creating a serious interference problem to licensed communications systems. Thus, the Federal Communications Commission established minimal emission requirements for digital computer equipment. Since the goal was to establish realistic requirements that were representative of typical problems, without being overly restrictive, several assumptions were made.

First, it was noted that below 30 MHz, the primary coupling path was conduction through the power lines, while above 30 MHz the primary coupling path was electromagnetic field radiation. This resulted in conducted emission limits from 450 KHz to 30 MHz, and radiated emission limits from 30 MHz to 1 GHz.

Second, it was noted that most of the complaints were regarding interference to television receivers. This resulted in the radiated emission limits set to protect television receivers. The values were determined by assuming typical signal levels, typical signal/noise ratios, and typical antenna factors. The goal was to limit undesired "noise" to levels below those that would cause interference in "normal" metropolitan situations.

Third, it was noted that computers were being installed in two environments, in businesses and in homes. In the home, it was assumed that a computer could be installed close to a television receiver, which resulted in a distance assumption of 3 meters. In business, it was assumed that the nearest television receiver would not be as close, so a distance was assumed as 30 meters.

The current emission requirements for digital computers, as specified in the FCC rules and regulations, Part 15J, are shown in figure 1. Although the Class A limits appear more stringent, when scaled to the same distance (assuming electric field intensity varies as 1/distance), it can be seen that Class A allows higher emanations. Thus, we can use the Class A limits as a "default" for computers.

THE VICTIM -THE VHF RECEIVER

The frequency and sensitivity characteristics of land mobile receivers are well defined. Land mobile receivers operate in three discrete frequency bands throughout the VHF/UHF range: 30-50 MHz, 150-170 MHz, and 450-470 MHz. Typical input sensitivities are 0.25 uv to 1 uv at the antenna input terminals, and typical antennas are 1/4 or 1/2 wavelength antennas.

Figure 2 shows the electric field intensities that result in 0.25 uv at the antenna terminals, assuming both 1/4 wavelength and 1/2 wavelength antennas with 50 ohm inputs. These were derived using the formulas:

E=eF/33 for a 1/4 wavelength antenna

E=eF/39 for a 1/2 wavelength antenna

where E is the field intensity in uv/meter, e is the input voltage in uv, and F is the frequency in MHz. The assumption is that if the interference is at or below the input threshold, then interference to even the weakest signals can not occur.

A COMPARISON OF SOURCE AND VICTIM LEVELS

By comparing the "source" characteristics of a digital computer meeting the Class A FCC requirements in terms of field intensity generated, and the "victim" characteristics of a VHF receiver at the antenna in terms of field

intensity tolerated, we can gain some insights into the potential for problems.

Figure 3 shows the limits for field intensities, as derived in Figure 2, versus the anticipated field intensity emissions from a Class A computer at several distances, based on a 1/distance decrease in field intensity. It can be plainly seen that the potential for interference between the computer and the receiver exists if the two are within one kilometer of each other! (This ignores attenuation effects due to buildings or other materials. Nevertheless, it does show the gravity of the situation.)

An alternate way of viewing this is to observe that additional shielding will be needed at the closer distances. This is shown in Figure 4. At one meter, up to 72 dB of additional shielding would be needed to provide the necessary margin between the source and the receiver.

A CASE STUDY

As EMI/EMC engineers, we were retained to assist in the system integration of a system using a microprocessor based controller, located three feet from a VHF receiver. The controller was originally designed using FCC Class A limits as a guideline. Needless to say, there were problems.

Comments on the Source

The microprocessor based system used an 8 MHz clock, divided internally by four. As a result, harmonics appeared at 2 MHz intervals with enough amplitude to cause problems on the 150 MHz-170 MHz band, as observed on both a radio receiver and a spectrum analyzer.

The computer system was housed in a rack, which provided little shielding at 150 MHz. The computer consisted of three units, a central processor, a memory unit, and a display/control unit. The three were interconnected with parallel high speed digital input/outputs, plus there were over a hundred low speed input/output lines used for control and sensing.

Comments on the Victim

The communications receiver had a sensitivity of 0.25 uv, and was connected to a 1/4 wavelength whip antenna. The signal received from the computer was enough to "break squelch", and was considered a major nuisance by the end user. The system was often used in fringe areas, so the maximum sensitivity was needed. Furthermore, the communications transceiver was installed for safety reasons, so its use was required.

Moving the communications receiver and/or antenna was not an option. For safety reasons, it was required that the operator have a VHF radio readily accessible. We were able to at least guarantee that the radio antenna would not be located any closer than three feet.

Only a very narrow band of frequencies needed to be guarded, approximately 2 MHz at 150-170 MHz, and 1 MHz at 450-470 MHz. Fortunately, there were no discernable problems in the 450-460 MHz band.

Comments on the solutions

First, an analysis was undertaken to assess the severity of the situation. Using the techniques described in this paper, it was determined that a Class A design still needed over 60 dB of isolation. Since isolation by distance was not an option, this meant additional shielding was needed. Incidentally, we also compared this with MIL-STD 461 levels, which was quite helpful in demonstrating the gravity of the situation to "management" (ie, this was even beyond normal military requirements.)

Second, the following fixes were needed:

-Solid copper shields on all high speed I/O lines

-Routing of I/O lines as close as possible to case

-Shielded bulkhead connectors on all I/O, with circumferential shield terminations

-Filter pins on all I/O and power lines

-Solid shielded cabinet on three subassemblies, with rf gaskets at seams

-Screen shielding over the display unit

In addition, the clock frequency was changed to move the clock harmonics off one critical frequency. This was not considered a long term solution, however, since harmonics still fell within the overall band to be used.

The results of these retrofits were a quasi-militarized system, at no small expense. It was definitely not a commercial "personal computer" type of design.

RECOMMENDATIONS AND CONCLUSIONS

-Do not assume that a computer meeting the FCC Part 15J requirements will work with a sensitive VHF radio receiver. The FCC limits were designed primarily to protect television receivers at prescribed distances.

-Analyze the situation, and try to predict both the sensitivity level of the receiver and the emission level of the computer. If the computer meets the FCC limits, use that as a default, and adjust for distance.

-Use space separation between the source (computer) and victim (receiver) where possible.

-Use frequency separation (changing clock frequencies) carefully. This is successful only if you are guarding a single frequency; otherwise it is likely to fail, since digital systems are rich sources of many signals.

-Be prepared to use much shielding and filtering if a microprocessor and a VHF receiver are co-located in a system.

BIBLIOGRAPHY

Daryl Gerke is a principal in Kimmel Gerke & Associates, Ltd., an engineering consulting firm that specializes in electromagnetic interference control. His career spans twenty years, and he has held engineering positions with Rockwell Collins, Sperry Defense Systems, Tektronix, and Intel Corporation.

Mr. Gerke received his BSEE from the University of Nebraska. He is a Registered Professional Engineer in Minnesota, and he holds an FCC Commercial license.

Figure 1

FCC LIMITS

	Frequency MHz	Class A (30 meters)	Class B (3 meters)
RADIATED	30-88	30 uv/m	100 uv/m
	88-216	50 uv/m	150 uv/m
	216-1000	70 uv/m	200 uv/m
CONDUCTED	0.45-1.6	1000 uv	250 uv
	1.6-30	3000 uv	250 uv

Figure 2

Electric Field Intensity Levels
For 0.25 uv at Antenna

	30 MHz	50 MHz	150 MHz	460 MHz
1/4 WAVELENGTH	0.23 uv/m	0.38 uv/m	1.1 uv/m	3.5 uv/m
1/2 WAVELENGTH	0.19 uv/m	0.32 uv/m	0.96 uv/m	2.9 uv/m

Figure 3

Class A Computer Electric Field Intensity Levels
VS
Receiver Sensitivity Levels

	30 MHz	50 MHz	150 MHz	460 MHz
1/4 Wave Antenna	0.2 uv/m	0.4 uv/m	1.1 uv/m	3.5 uv/m
Class A -1 meter	900 uv/m	900 uv/m	1500 uv/m	2100 uv/m
3 meters	300	300	500	700
10 meters	90	90	150	210
30 meters	30	30	50	70
100 meters	9	9	15	21
300 meters	3	3	5	7
1000 meters	0.9	0.9	1.5	2.1
3000 meters	0.3	0.3	0.5	0.7

Figure 4

Isolation Required Between Class A Computer and Receiver

	30 MHz	50 MHz	150 MHz	460 MHz
1/4 Wave Antenna				
-uv/m	0.2	0.4	1.1	3.5
-dB uv/m	-13	-8.5	0.8	11
Class A Computer-1 meter				
-uv/m	900	900	1500	2100
-dB uv/m	59	59	63	66
-Isolation dB	72	68	62	55
Class A Computer-3 meters				
-uv/m	300	300	500	700
-dB uv/m	50	50	54	57
-Isolation dB	63	59	53	46

NEXT GENERATION LOW NOISE EMC DESIGN

Bruce C. Gabrielson PhD
Vice President TEMPEST/EMC Division
Sachs/Freeman Associates

What does the next generation hold in store for extremely low noise, highly sensitive circuit design? In particular, as industry pushes to get greater capability into less space at lower cost, how will the EMC designer cope with the limiting factor of device susceptibility to self generated and locally generated interference. This paper first reviews current information and then examines new and emerging technologies which may help engineers keep up with the increasing noise related issues of the fast moving high tech environment. The following discussion describes basic noise source.

Where Does Noise Come From?

Noise can be generated due to internal device sources, other nearby devices, circuit basing and physical parameters, power sources, and grounding conditions. Internal noise sources include intrinsic, flicker, popcorn, pink, Schottky, and white or thermal noise.

Intrinsic noise is that noise which is characteristic of the semiconductor material itself rather than the content of any impurities it contains. All noise power from each of the various intrinsic noise sources is additive. There are three types of intrinsic noise sources prevalent at low frequency, with one type (flicker) dominant below 100Hz. Since considerable research is currently underway to develop new semiconductor materials, only flicker noise continues to be the major concern compared to other intrinsic noise sources.

Flicker noise in a semiconductor is also called 1/f, excess, or modulation noise. In a resistor it is called contact noise. Flicker noise exhibits a 1/f power spectrum where n can vary between $.8<n<1.5$. Flicker noise density e_n is defined as

$$e_n = Kf^{-n/2}$$

where K is a constant dependent on the type of material and its geometry, and f is frequency in hertz.

Popcorn noise (burst) is generated at semiconductor junctions as a result of modulation of the barrier height caused by metallic impurities in the semiconductor. It is greatest in high impedance devices such as op-amp input circuits. Popcorn noise is reduced by improving manufacturing techniques and reducing the metallic impurities of the semiconductor device, and is characterized by the constant amplitude of the individual bursts. However, the duration of the popcorn noise varies from microseconds to seconds and the repetition rate is not periodic. The amplitude is generally 2 to 100 times thermal noise. Popcorn noise exhibits a 1/f power spectrum, where n is generally 2, and appears to be related to flicker noise.

Pink noise is the term used to describe any noise that has a 1/f power spectrum. The total flicker noise in bandwidth B falling between frequencies f_1 and f_2 has been shown [1] to be $E_n = 1.5DK$ where D is the frequency separation between f_1 and f_2 in decades. Pink noise, therefore, increases 3dB for every doubling of decade separation.

Shot or **Schottky noise** is caused by the fluctuations in the rate of arrival of charge at semiconductor junctions. Junction current for Schottky noise can be calculated from:

$$I_n = [2qI_{dc}B]^{1/2} \text{ amps}$$

where

I_{dc} = quiescent junction current (amps)
B = device bandwidth and
q = election charge.

White noise (thermal) is dominant at frequencies above 100 Hz. White noise, also referred to as Johnson noise, is the result of resistances internal to the component. Normal temperature variations have only a small effect on the value of thermal noise. For example, according to Ott[2], an increase in temperature from 17 degree C to 117 degree C produces only a 16% increase in thermal noise for op-amps.

Research Into Low Noise Technologies

Probably the most research being done to reduce internal noise sources relates to superconducting materials. The major thrust is to develop materials which exhibit superconducting properties at higher temperatures. Although, many materials offer promise, the step between laboratory and practical application is a large one. Currently Barium yttrium copper oxide (Y_1Ba_2C0), which becomes superconducting at 90°K, and Niobium Titanium ($N6S_nNbTi$), which becomes superconducting at about 23°K, are applicable in thin films for applications in Josephson junctions and other electronic devices.

Other areas of materials research are showing promise as well. Currently, Gallium arsenide (GaAs) offers the greatest available potential for very fast and stable operation at low power. As shown in Figure (1), electrons travel through gallium arsenide much faster then through silicon because of their conduction band shapes. Of particular interest is current research [3] into depositing extremely pure gallium arsenide thin films on a silicon wafer. In the longer term, advances in this area could lead to the development of new kinds of devices capable of coupling the optoelectronic and fast switching properties of GaAs with the low cost and higher density of silicon.

One promising area of research which may eventually succeed gallium arsenide is the high-electron-mobility transistor (HMET). HMETs outperform GaAs FET's at high frequencies having extremely high gain (down to about 1 volt) and low noise characteristics [4]. Toshiba has developed a HMET with a noise figure of only 1.3dB at 18 GHz and a gain of 9.5dB.

Another recent development is the research into current based logic. Such a device, a typical current-mode logic (CML) gate, is shown in Figure (2). The CML gate can take both true and false values for its inputs [5]. Since it has a multilevel structure, each input must be shifted to the proper level. Another scheme, complementary current-mirror logic (CCML), has gate power consumption close to CMOS, yet operates at frequencies close to ECL. CCML takes advantage of complementary bipolar devices connected in a way to keep power drain low in the absence of signals, and reduces power consumption about 75% over ECL when active.

Crosstalk

Noise from nearby devices usually appears as a result of crosstalk. Crosstalk can appear at the device itself, such as a dip package containing more than one logic element, or, as is often the case, crosstalk is capacitively coupled between adjacent PC board traces.

Looking first at package designs, Figure (3) shows the results of isolation testing between two package types for a typical logic device. As can be seen, the dip package with its long adjacent leads [shown in Figure 4] exhibits less isolation then does a flat package with shorter lead lengths. Additionally, the physical layout of individual logic devices on a package again effects circuit isolation. Figure (5) shows the lead configuration of a 7404 and a 6N140.

Belisle and Jackson [6] have shown that an average increase in adjacent device isolation of around 20dB can be achieved simply by proper pin out selection.

How are the problems related to device crosstalk being addressed?

Chip carriers currently offer the greatest potential for achieving an ultimate solution to the problem of accommodating increased speed, maximum density devices with a high level of lead isolation. On a chip carrier device, all four sides of the package contain leadless nearly equal length traces to the internal die. Packages are currently ceramic, and are mounted directly to the PC board on a small exposed board trace for each package trace. This technique is known as surface mounting. A 24 pin ceramic chip carrier package is shown in Figure (6).

Maerschalk [7] has looked at the future trends in increasing circuit density in packages beyond chip carriers. He suggests that further increases in circuit density can be expected with the implementation of individual chips directly on boards (COB) for nonhermetic applications, and the use of multi-chip, hermetically sealed packages when required. Regardless of the ultimate packaging technique perfected, however, it is expected optimum EMC design for the next several years will be achieved at the chip carrier level.

Beside packaging, the ability of a single chip carrier to be designed to carry more sections of individualized design elements is being enhanced through the use of ASIC. ASIC stands for application specific integrated circuit. A typical ASIC design cycle will involves integrating several data libraries to quickly produce one specifically tailored device.

PC Board Advances

Biasing is a well understood issue to most analog circuit designers. However, as the need for fast digital processing and high frequency analog signal propagation increases, circuit designers must re-look at equivalent circuits in terms of feed back coupling paths and sensitivity to coupled impulse noise at high frequencies. A digital signal has harmonics existing considerably above its operating bit rate. Obviously these signals will couple to other circuits in their immediate vicinity due to capacitive effects.

This brings us to advances in PC board design techniques to reduce crosstalk and radiated noise problems. While the single layer board is the most inexpensive and popular, newer designs will force more and more organizations to adopt the multilayer approach. Figure (7) from Walter [8] shows a typical multilayer board configuration.

Multilayer circuit boards offer a significant increase in EMC protection at high operating frequencies when compared to single or double sided circuit boards for three reasons. First, since interconnecting traces can be routed on layers other than the layer which has the components mounted on it, the components can be located closer together enabling greater component density, and thus reducing interconnecting trace length and associated coupling. Secondly, critical or sensitive traces can be routed perpendicularly on different boards, reducing the length which critical traces have in common and, in turn, reducing capacitive coupling. Third, ground, signal return, and power planes are included in the multilayer design. Not only do these planes provide a low impedance signal return or ground, but they also perform the important function of providing isolation between the trace/interconnect layers of the multilayer board and also isolation to the power plane which is generally the bottom layer.

Another PC board area being developed is the use of backplanes for high speed emitter-coupled logic (ECL) designs. Backplanes transmit energy like any other wiring. Wiring runs effect propagation delay, but since this type has lower impedance then PC board traces, ECL can transmit much faster digital signals, and is less affected by noise. Also, using backplane drivers that are terminated, as are ECL drivers, significantly reduces circuit crosstalk and transmission line reflections.

CAE Advances

Besides multiple layers, tremendous advances have been made recently in PC board computer aided engineering programs [9]. Unfortunately, most programs do not interact between physical design efforts and electrical evaluations addressing EMC susceptibility or noise generation. Unless the circuit designer takes a direct role in most instances, sensitivity, critical paths, and special placements or routings will not be addressed.

Recent advances in artificial intelligence and rule driven design will hopefully rectify these problems. Rule driven design, when coupled with interactive circuit analysis programs such as SPICE, and with antenna modeling programs, should allow PC cards to be designed for maximum interference suppressions, thus allowing increased usage of more sensitive devices within circuits. Coupling rule driven design decisions with ASIC could ultimately produce maximum source suppressed circuitry on an extremely protected circuit card.

Power Supply Noise

Power supply noise, particulary noise related to switchers, has been one of the leading causes of reduced sensitivity ever since this type of electronics was first developed. To address the power supply issue, switching regulator IC's are being developed which will greatly aid in reducing the noise problems. Diane Tunick [10] recently published an informative article describing various devices and how they work.

For years bipolar switchers kept regulator operating frequencies below 80kHz. Now, with the development of MOSFETs, regulator manufacturers are reaching for operating frequencies considerably higher then previous designs. Already rates of 200 to 300 kHz are common, and a few chips are operating at 1 MHz, with advanced design schemes affording more speed and less power dissipation.

Before leaving the subject of power supplies, one additional point is important to note. The path through which power supply noise reaches a particular device is primarily its powerline. Some devices such as op-amps have limited built in power supply noise rejection. This characteristic is specified in terms of the Power Supply Rejection Ratio (PSRR) and is defined as the ratio of the equivalent voltage change on the input to a change in the supply voltage on the power input. Since PSRR deteriorates with increasing frequency, a designer might at first think higher frequency noise on the input power is going to create more rather than less problems. However, since it is much easier to filter higher frequencies using either passive devices or ferrites, the evolution to high frequency power supplies will be of considerable advantage to EMC designers.

Grounding Noise

The final source of noise problems relates to improper grounds. Although the solution to improper grounds is intertwined with previously discussed issues, a few points merit further discussion.

By far the most often identified problem encountered by the EMC circuit designer relates to common mode signals. Figure (8) shows the relationship between common mode and differential model noise. As can be seen, if a ground condition exists whereby a noise source appears on both the ground and the input to a device, this noise will propagate through the device regardless of the internal protection employed. Such a condition is often created through capacitive coupling when the ground return of a device consists of a narrow circuit trace routed near to the device input trace.

To effectively reduce the common mode noise problem, eliminating ground loops and creating a single point ground for circuit functional nets is most effective. Using ground plane on the circuit card, and providing a low impedance path to the single system grounded point for this plane is usually effective where common mode problems are suspected.

Rule driven PC card layout programs which use functional nets have recently been used to solve circuit grounding problems. Functional nets divide the circuity into logical blocks, so grounds can be more effectively controlled. The golden rule of thumb is to always try to provide a ground return path as close to the noise or signal source as possible. Using this as a guide, rule driven layout programs can specifically address issues of the type described.

Conclusion

Industry advances in higher speeds, lower power, greater capacity, and increased sensitivity have required increased emphasis by circuit designers in the area of EMC. So long as designers allow their perspective of EMC related issues to evolve with each new technology, the chances for developing maximum sensitivity and minimum susceptibility circuits with each technology advance are excellent.

References:

1. Gabrielson, B.C. and Walter, R.L. An EMC Designers Guide to Operational Amplifiers, ITEM, 1986

2. Ott H.W. Noise Reduction Techniques in Electronic Systems, Wiley Interscience, New York, 1976

3. Iverson, W.L. Researchers Tilt Silicon to Grow Pure GaAs On It, Electronics, Feb 10, 1986

4. Cohen, C.L., HEMTs Finally Emerge to Compete With GaAs FET, Electronics, Dec 9, 1985

5. Gabrielson, B.C., Future EMC Trends in PC Board Design, T18.4, EMC EXPO 86, Washington, D.C.

6. Belisle, K. and Jackson, M., EMI Design Techniques for Decoupling and Isolation of Microcircuits, IEEE EMC Symposium, 1983, CH1838 - 2/83/0000-0207

7. Maerschalk, J., Chip Carriers Today and Tomorrow, Journal of Electronic Defense, Feb, 1986

8. Walter, R.L., Identifying EMI Parameters On Printed Circuit Boards, T18.2, EMC EXPO 86, Washington, D.C.

9. Gabrielson, B.C., EMC Intercircuit Isolation, Design in Early or Risk Finishing Late and Over Budget, EMC Technology, July, 1984

10. Tunick, D., Power-supply Regulator ICs are Switching into High Gear, Electronic Design, October, 1987

11. Gabrielson, B.C. and Reimold, M.J., Suppression of Power Line Noise With Isolation Transformers, T14.16, EMC EXPO 87, San Diego

12. Eden, R.C., Livingston, A.R., and Welch, B.M., Integrated Circuits - a case for gallium arsenide, IEEE Spectrum, Dec, 1983

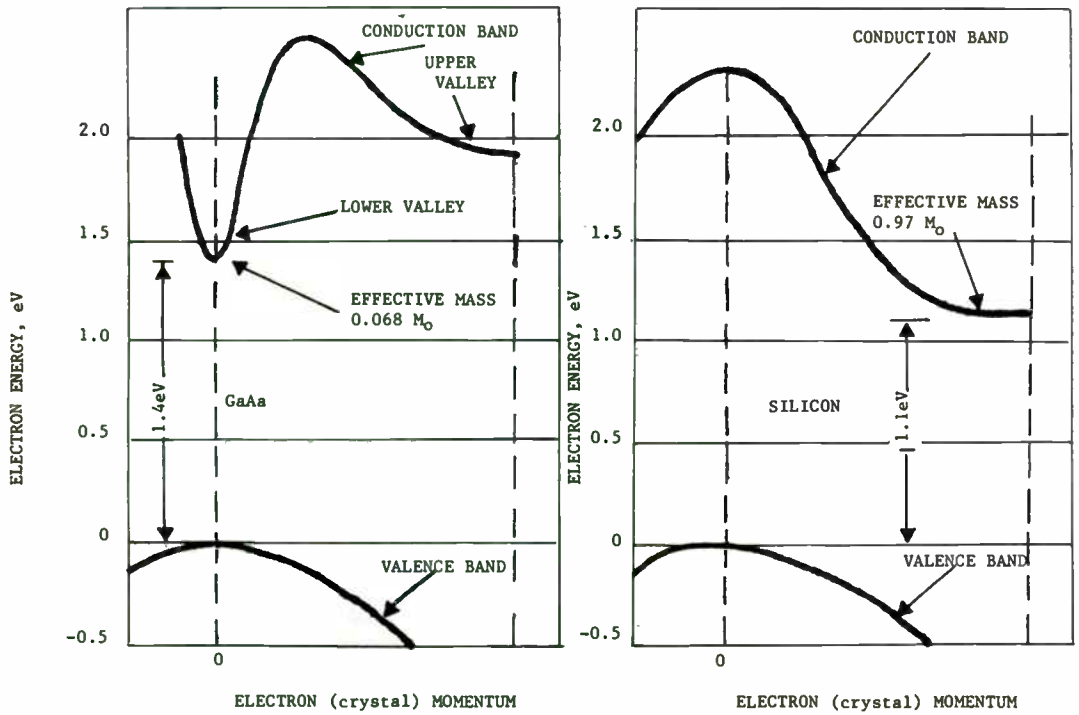


FIGURE 1. Electrons travel through gallium arsenide faster than through silicon because of their conduction band shapes. [12]

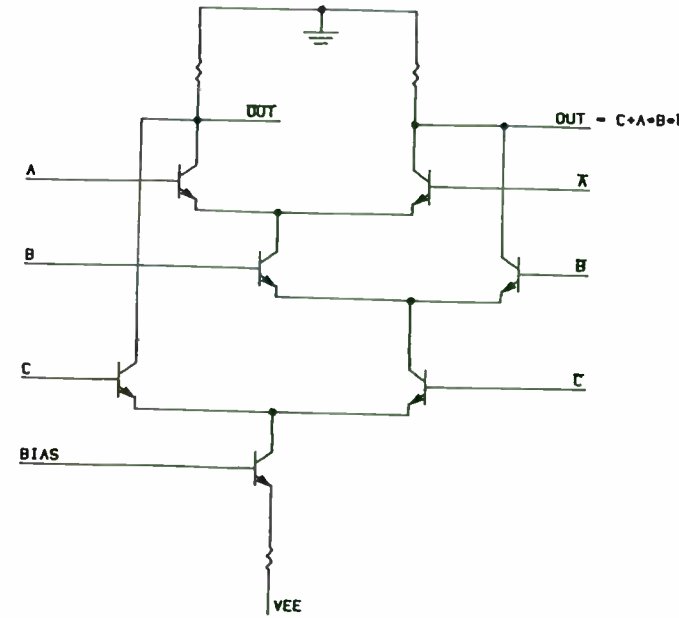


FIGURE 2. A Typical Current-Mode Logic (CML) Gate

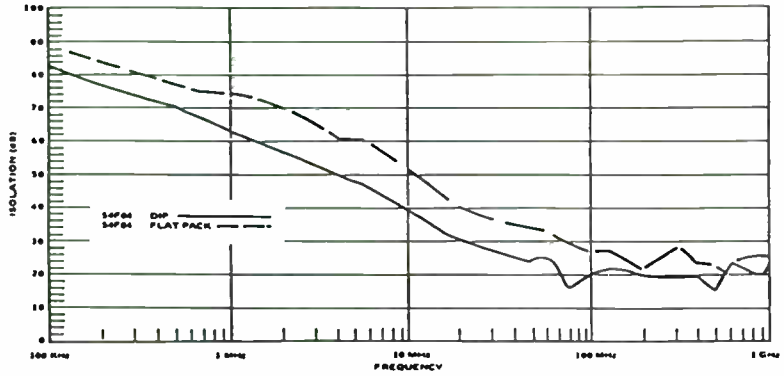


FIGURE 3. Isolation of a Flatpack Versus a DIP

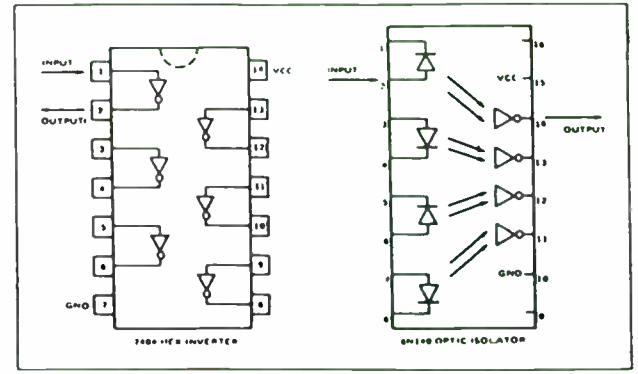


FIGURE 5. Pin Configurations for Typical and Isolated Devices (8)

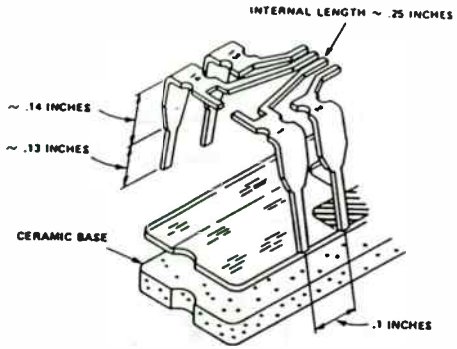


FIGURE 4. Dual In-Line Construction Showing Lead Configuration At One End

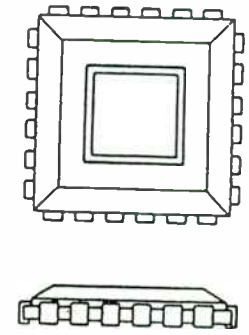


FIGURE 6. 24-Pin Ceramic Chip Carrier

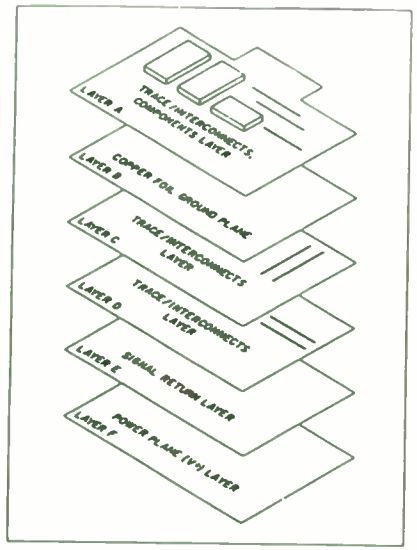


FIGURE 7. Possible Multilayer PC Board Configuration

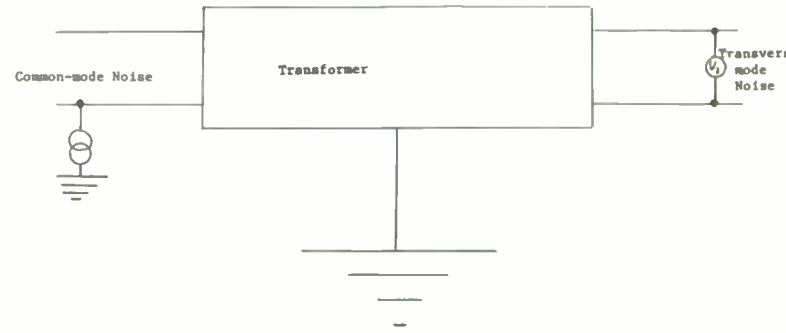


Figure 8. Common-Mode and Transverse-Mode Noise

Practical Considerations in Specification of High Stability Crystal Oscillators

Glenn R. Kurzenknabe
Sales Manager
Piezo Crystal Company
100 "K" Street, P.O. Box 619
Carlisle, PA 17013
Phone: 717-249-2151

ABSTRACT

Problems for the user and manufacturer of high-stability crystal oscillators can be minimized by increased understanding of parameters affecting design and cost.

This paper will address system design considerations and critical performance parameters for oscillators utilizing "SC" cut crystals in the 5-10 MHz and 90-110 MHz regions.

Parameters discussed will be:	
Frequency Stability	5×10^{-7} to 1×10^{-10}
Aging	1×10^{-8} to 1×10^{-11} per day
SSB Phase Noise	to −165 dBc/Hz
Allan Variance	to 5×10^{-13} per second
Environmental Conditions	Temperature and vibration and their affects on Frequency Stability, Aging and SSB Phase Noise.

INTRODUCTION

It is very important for the systems design engineer to have a basic understanding of the parameters affecting the design of high-stability crystal oscillators. This paper will focus on oscillators utilizing SC cut crystals. The SC cut crystal exhibits tremendous advantages over more common AT cut crystals in areas such as aging, warm-up, phase noise, frequency stability and vibrational sensitivity. Under or over specification by the user can add additional costs and jeopardize systems performance.

The term **Frequency Stability** is a generic term which means a variety of things to different people depending on their individual requirements and interests. In its broadest concept, it means the degree of constancy of the frequency of an oscillator under a particular set of operating conditions. In crystal oscillator applications, there are several different types of frequency stability:

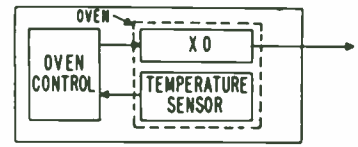
A. Frequency stability as affected by environmental changes, primarily temperature.
B. Long-term frequency drift (generally called aging) as affected by the aging of the quartz crystal resonator or oscillator components.
C. Short-term stability (Time Domain) sometimes called Allan Variance
D. Frequency stability as affected by outboard circuit changes such as supply voltage changes—called "pushing" and load variations called "pulling."

A: Frequency vs Temperature

Frequency Stability is associated with changes in output frequency of an oscillator compared to the effects of a changing ambient temperature. This is expressed for example as $\pm 5 \times 10^{-9}$, 0° to +70°C or 2×10^{-9}, +15° to +60°C. You will note that the latter is "a total" rather than "±" which is obviously more difficult.

In order to achieve frequency stabilities over temperature in the region 5×10^{-7} to 1×10^{-10}, a well-designed proportional oven is required to "shield" the quartz crystal, oscillator stage(s) and buffer amplifier from a changing ambient such as 0° to +70°C. A basic oven is shown in Figure 1.

Figure 1.

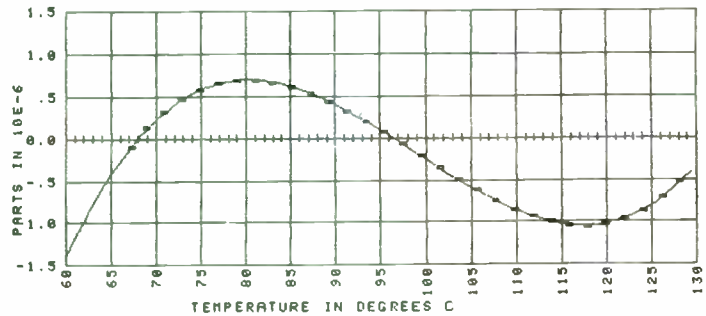


The housing or cavity, as it is called, is wrapped with many turns of resistance wire forming at least one and maybe two heater windings. This wire will heat the cavity as current flows through the wire. The oven controller will sense both the ambient temperature and the cavity temperature. A well-designed proportional oven always operates 10-15 degrees C higher than the highest ambient seen by the oscillator module. This 10-15 degree C "guard-zone" allows the oven controller to maintain constant control even at the highest ambient temperature. Should the outside ambient rise higher than the 10-15 degree C "guard zone", the oven sensor (usually a thermistor) will become confused and ultimately turn "off" resulting in greatly reduced performance.

Specification of the operating temperature range should take into consideration not only the operating ambient temperature of the overall system, but any heat rise inside the system due to heat dissipating devices such as power supplies and power amplifiers. To achieve stabilities over temperature of 5×10^{-7} to 1×10^{-10}, the cavity temperature of the oven must be matched to the turn-over temperature of the crystal. The turn-over temperature is the point where the crystal's frequency vs temperature curve begins to "flatten" and eventually reverse direction. It is at this point that changes in cavity temperature will have the least effect on the resulting frequency stability. An SC cut crystal has a lower turn point temperature (LTPT) in the range of +50° to +85°C.

Figure 2 indicates a typical frequency vs temperature curve for a 10 MHz, 3rd overtone SC cut crystal. You will notice tht the LTPT is +81°C. This point can be varied by adjusting the Z angle prior to cutting the crystal blank. Temperature changes in the operating environment can affect frequency stability. As the oven temperature sensor begins to see a change in ambient, it will call for more current if the ambient is decreasing as the oven will need to generate more heat to maintain the same cavity temperature. It will call for less current if the ambient is rising as the oven will need to generate less heat to maintain the proper cavity temperature.

Figure 2.



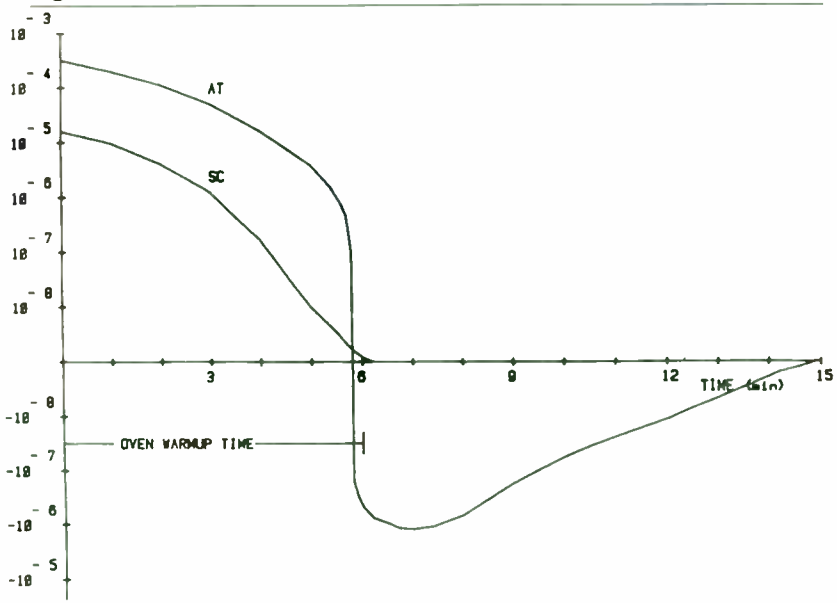
The design of this oven must be such that very exacting thermal stabilities are achieved. For example, using previously mentioned 0° to +70°C to hold 5×10^{-7}, ± 3°C will be required and for 1×10^{-10}, ± 0.2°C will be required. Oscillators of this type utilize oven cavities machined from a solid block of aluminum with an HC-37 or HC-40 crystal holder embedded in the aluminum. This affords a large thermal mass thereby yielding excellent thermal performance.

Warm-up characteristics are greatly improved by the use of SC-cut crystals. Figure 3 indicates that no "over-shoot" is present with SC cut crystals as is with AT cut crystals. This permits more rapid and exacting warm-up performance. This is due primarily to the fact that SC cut crystals are virtually stress free.

Double ovens or an oven inside another oven can be used to achieve stabilities beyond 1×10^{-10}, however, they are physically large, require more DC power, are higher in cost and reduced MTBF as compared to a single oven. Dewar (vacuum bottles) flasks are large, costly and very fragile.

The user should specify the supply voltage and amount of power available. If restrictions exist, such as time limits on power consumption or frequency characteristics, they should be included in the specification. Ripple characteristics should be defined as they may affect the design of the regulator circuit in the oven controller. Most high performance oscillators are designed with a "split input", i.e., the oven and oscillator supplies have separate inputs thereby allowing different supply voltages or quality of supply voltages.

Figure 3.



Ovens are sometimes operated from raw aircraft +28 VDC that may vary from +22 to +30 VDC. This source can usually produce large amounts of power but with relatively poor quality, i.e., ripple, regulation, etc. This requires the manufacturer to design in additional regulation and decoupling.

If switching DC power supplies are used, the switching frequency and spur amplitude should be specified. Ovens operating on AC inputs are not recommended due to 60 Hz modulation that may appear on the output of the oscillator.

B: Long-Term Drift (Aging)

Aging of a crystal oscillator is usually specified on a daily, monthly or yearly basis and defines the amount of frequency shift permitted over the time specified, i.e., 5×10^{-10}/day or 1×10^{-7}/year. As a point of reference here, it might be appropriate to indicate the relative magnitude of "1×10^{-10}". In specific terms, it is: 1 part in 10 billion

0.0001 Hz at 1 MHz

0.001 Hz at 10 MHz
0.01 Hz at 100 MHz
0.25 second in a human lifetime of 80 years
0.25 inch out of the circumference of the earth

The primary cause of aging is the aging of the quartz crystal. Of course, it is impossible to build a perfect crystal; however, with care and state-of-the-art manufacturing equipment, near-perfect crystals can be produced.

The major causes of aging are contamination of the quartz resonator and its housing, poor finishing techniques of the quartz blank, and stresses placed upon the blank by poor mounting. These problems can be minimized by very strict cleaning of crystal blanks under cleanroom conditions, the use of gold plated housings, low outgassing adhesives to bond the crystal blank to the mounting structure and established manufacturing processes.

Aging is also a function of the mass or size of the blank. Therefore, a 10 MHz 3rd overtone SC cut crystal will have superior aging compared to a 100 MHz 3rd overtone SC cut crystal as the 10 MHz blank is physically larger.

Comparison of Aging and Crystal Frequency

Time	5 MHz	10 MHz	100 MHz
1 day	$1x10^{-10}$	$5x10^{-10}$	$5x10^{-9}$
1 month	$3x10^{-9}$	$1.5x10^{-8}$	$1.5x10^{-7}$
1 year	$3x10^{-8}$	$1x10^{-7}$	$1x10^{-6}$

*All crystals are 3rd overtone SC cut. Levels represent current production yields.

Aging is measured after a warm-up time (usually 24 to 72 hours) during which the oscillator is allowed to thermally stabilize. Data is then collected regularly during the measurement period either manually or with a computer based automatic system. During this period, the oscillator should be maintained at a relatively constant (±2°C) temperature to insure that the aging data will not be "masked" by frequency shifts due to temperature variations.

Figure 4.

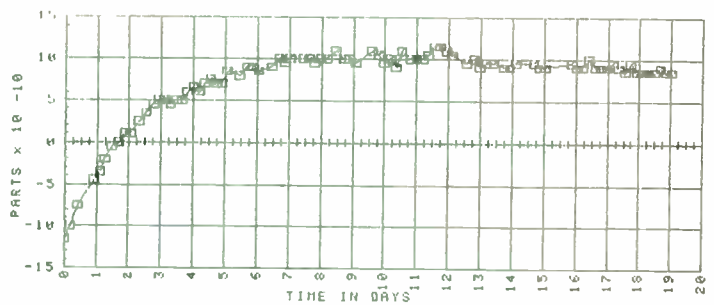
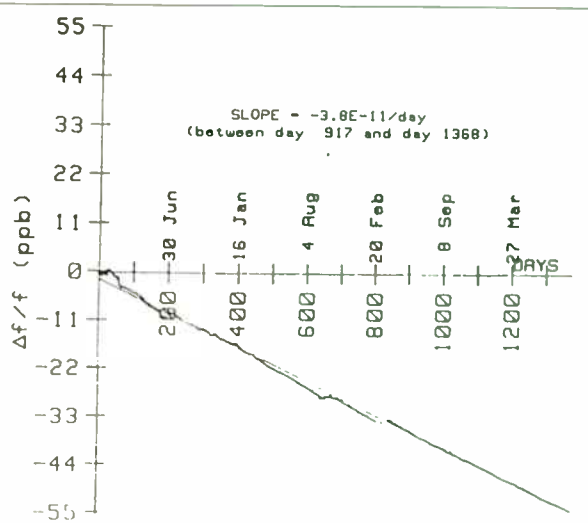


Figure 4 indicates actual aging data collected by a computerized system. You will note that the test required 19 days and the daily aging rate at time of shipment was $1x10^{-10}$/day.

Figure 5 indicates data collected over a 1368 day period on a Piezo manufactured oscillator at 10 MHz utilizing a 3rd overtone SC cut crystal. This test was conducted by the U.S. Army–LABCOM at Fort Monmouth, NJ. The specification for aging was 5×10^{-10}/day. The user is cautioned to avoid specifying aging only on a monthly or yearly basis. These specifications are "uncheckable" as delivery cycles and customer needs usually dictate shipment before data could be collected over month(s) or a year. These specifications should be reduced to daily requirements and the vendor should be required to submit test data to substantiate the results as shown in Figure 4. Data gathered on a daily basis can be easily computed for a month or year simply by multiplying by the number of days. This calculation assumes a linear progression which does not generally apply to aging. Aging is normally an exponential progression whereby the second month will be better than the first and the third better than the second, etc. By using a linear progression, it is not only very easy to calculate but also provides a "worst case" result.

Figure 5.

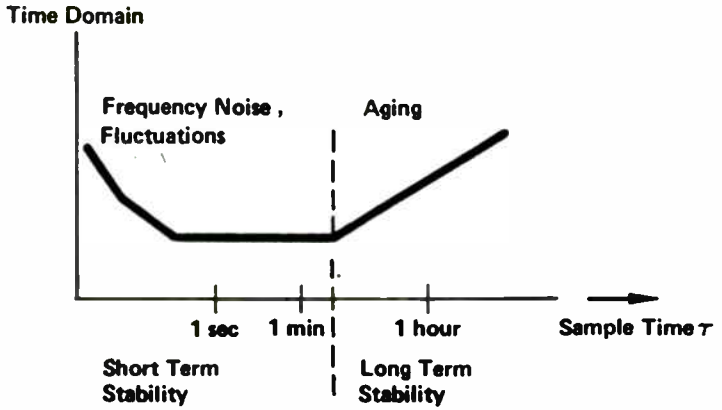


Once an aging (long term) slope has been established, power can be removed and the oscillator left dormant for a period of time. After the reapplication of power and a re-stabilization period, the oscillator will assume the original aging slope.

C: Short Term Stability

Short term stability or Allan Variance (referring to Dr. David Allan of the National Bureau of Standards, who pioneered much of the work on measurement techniques for short term stability), is characterized by random, noise-like changes in frequency measured in the time domain. It is not meant to include frequency variations due to component aging or ambient temperature changes. In most cases, it is specified as the RMS fractional frequency deviation for some specified measurement time, i.e., 5×10^{-12}/ 100 msec. Short term stability is best measured by using a Hewlett Packard computing frequency counter where many measurements are taken and averaged using software contained in the frequency counter.

Figure 6.



The time domain (Allan Variance) method of specifying short term stability is useful for counting intervals ranging from microseconds to 10 seconds. Figure 6 shows the typical relationship between Allan Variance and aging. If measurement times exceed about 10 seconds, their accuracy can be "masked" by the effects of aging and temperature variations. Allan Variance is a function of crystal Q, the higher the Q of the crystal, the better the Allan Variance. Therefore, due to the fact that a 10 MHz 3rd overtone SC cut crystal has a Q of about 1.2 million compared to a 10 MHz 3rd overtone AT cut crystal with a Q of about 100 thousand, the resulting short term stability will be better.

Comparison of Allan Variance and Crystal Frequency

Time	5 MHz	10 MHz	100 MHz
1 second	7×10^{-12}	5×10^{-11}	5×10^{-10}

*All crystals are 3rd overtone SC cut. Levels represent current production yields.

Short term stability is not degraded or affected in any way by frequency multiplication. Therefore oscillators at frequencies above 10 MHz which must provide excellent short term stability must incorporate multiplier stages.

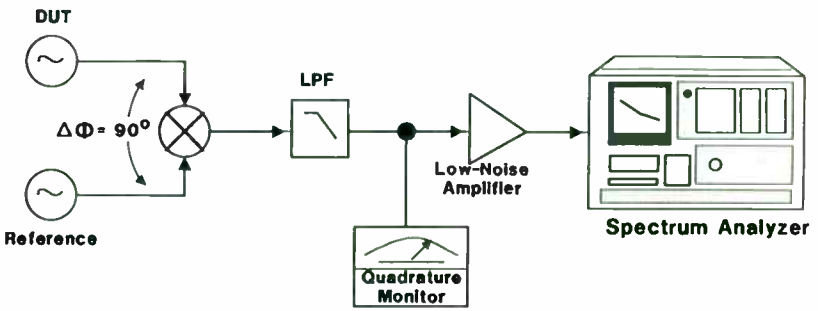
D: Pushing and Pulling

"Pushing" refers to changes in the oscillator frequency caused by changes in the supply voltage. This can be reduced or eliminated when the user clearly defines the conditions of the supply voltage such as level, regulation and "spike" content. The manufacturer must incorporate the correct type and amount of regulation to ensure that the actual oscillator circuit will see the most highly regulated voltage possible. Generally accepted voltage regulation methods can be incorporated in most oscillator designs.

"Pulling" refers to changes in the oscillator frequency caused by changes in the load of the oscillator being reflected back to the actual oscillator circuit. This can be reduced by the use of emitter follower stages and isolation/buffer amplifiers between the oscillator circuit and any class C stages. Multiplier stages are especially suspect when "pulling" problems are identified in an oscillator. The user should clearly indicate the impedance of the load and what VSWR, if any, will be present. Testing an oscillator for "pulling" with VSWR variations will create additional cost.

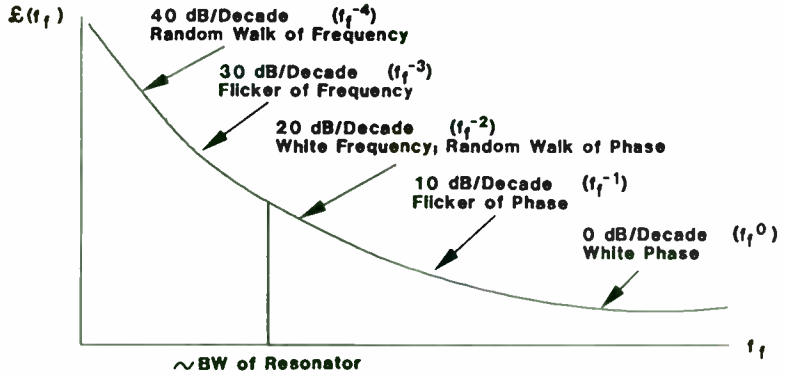
Phase noise is measured in the frequency domain and is the random phase modulation which is present on all crystal oscillators. This modulation is caused by a combination of the crystal and the noise figure of the transistors used. Phase noise is measured by beating two oscillators of like design against each other in a double balanced mixer (see Figure 7) in phase quadrature. A low frequency wave analyzer or spectrum analyzer is swept along the carrier envelope. This system measures the power relative to the peak power of the carrier in a very narrow bandwidth—normally 1 Hz.

Figure 7.



Close-in phase noise (5 Hz to 500 Hz from the carrier) is generally a function of the crystal. The phase noise floor (5 kHz and out from the carrier) is generally a function of the signal-to-noise ratio of the oscillator transistor(s) and the drive level on the crystal (see Figure 8). An SC cut crystal can operate well with 5-6 mw of drive level without degraded aging or spectral breakup. An equivalent AT crystal must operate with drive levels in the microwatt range, thus yielding a higher (worse) noise floor.

Figure 8.



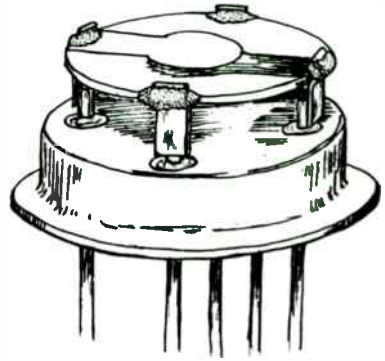
Phase noise is degraded by frequency multiplication. This degradation is at the rate of 20 log (N) with N being the multiplication factor. Therefore, with a doubler (x2) the degradation would be 6dB and for a x10 multiplication scheme the degradation would be 20dB. This factor equates to 6dB/octave. Phase noise is very critical in many systems such as radar, ECM and communications systems, and should be addressed in detail with potential manufacturer.

Environmental conditions can have a great impact on the performance of crystal oscillators. Vibration can affect an oscillator by physically damaging it. This is due to inaccurate specification of the vibration conditions by the customer and/or poor or improper mechanical design and construction techniques by the manufacturer. It is necessary for the user to clearly specify in detail the vibration spectrum the oscillator will see, and the g levels it will experience.

Operating and non-operating conditions must be specified as they both must be addressed during the design phase. Proper design dictates that the customer must allow adequate volume for secure mounting of the oscillator. Typical ruggedized mounts include threaded studs, welded flanges or ears, threaded inserts or an actual vibration isolation system utilizing sealastic rubber mounts. The manufacturer must incorporate ruggedized design and manufacturing techniques such as PC boards supported on four sides by a system of rails, plus additional mounting points near the center of the board to prevent "oil-canning" of the PC board. Individual component parts such as resistors and capacitors may be "spot-bonded" with an approved bonding agent, or the entire assembly can be foam filled for the most severe environments. Outgassing characteristics of these bonding agents and foams should be reviewed to prevent any long term caustic effects on the component parts of the oscillator.

Mechanical assemblies within the oscillator, such as the oven, must be physically secured to prevent damage and subsequent failure. Interconnecting leads between PC boards or assemblies must be secured to prevent fatigue and breaking. The crystal mounting structure must be a 4 point design (4 supports spaced at 90 degrees), (see Figure 9) utilizing the HC-35 at 100 MHz, the HC-37 and HC-40 at 5 and 10 MHz for maximum rigidity.

Figure 9.



Up to this point, our focus has been on the survivability aspects of vibration. We must now turn our attention to a greatly misunderstood phenomenon: the effects of vibration on phase noise. A quartz crystal is often thought of as an electrical device. To some degree it is, as it has C, L and R components (see Figure 10). In reality, it is a mechanical device. Therefore, basic FM theory says that when a crystal is vibrated, FM sidebands will be produced at the vibration frequency. These sidebands appear in the form of degraded phase noise. When subjected to sinewave vibration spectrums, the sidebands appear only at the vibration frequency (assuming the absence of mechanical resonances in the crystal mounting structure or oscillator assembly).

Figure 10.

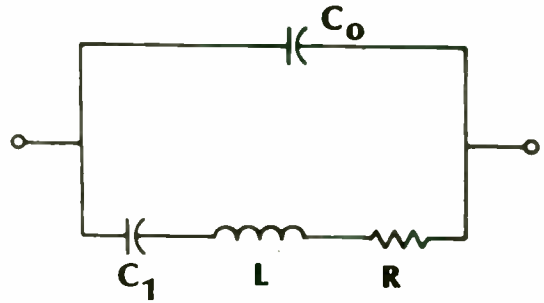


Figure 11 shows sinewave vibration induced sidebands at 100 Hz on the output spectrum of an oscillator. Common sources of sinewave vibration are cooling fans and propellers on a ship. During random vibration, where the frequency and amplitude of the vibration spectrum change "randomly", the phase noise spectrum will rise overall, possibly causing serious system degradation.

Figure 11.

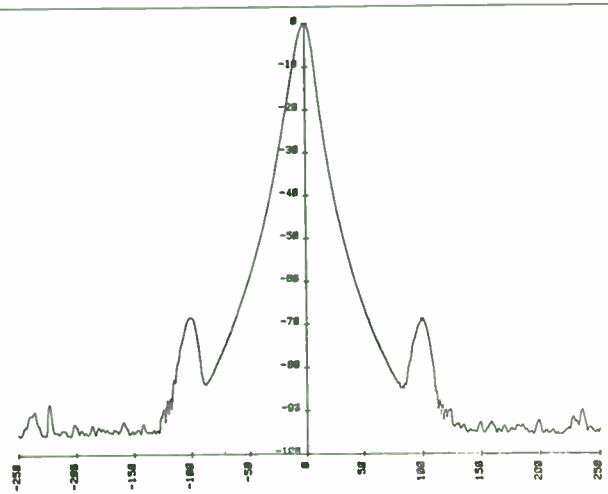


Figure 12 shows data for an oscillator subjected to a typical random vibration environment. Common sources of random vibration include aircraft and helicopters. The effects of vibration on phase noise can be better predicted by using SC cut crystals which exhibit improved g sensitivity. Mechanical vibration isolation systems can also be used, although they are large and bulky. Experimentation into electronic compensation has been carried out using accelerometers and microprocessor technology. The results of active compensation are mixed and not conclusive.

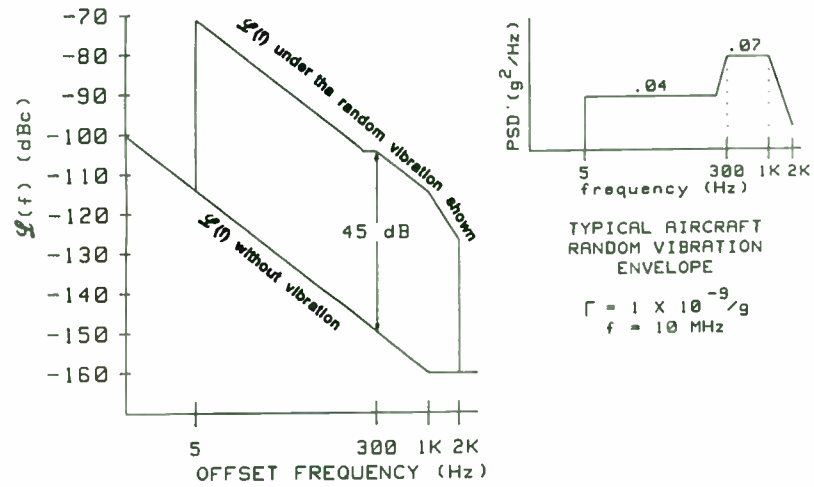
The effects of sinewave vibration on a crystal oscillator are calculated using the following equation:

$$SSB = 20 \log \left[\frac{\Gamma \cdot g \text{ level} \cdot f_o}{2 \cdot f_v} \right]$$

Using this equation and assuming a "perfect oscillator" (noise-free) at rest with a g sensitivity (Γ) of 1×10^{-9} at 10 MHz, the SSB phase noise due solely to a sinewave vibration level of 1 g would be:

Vibration Frequency	SSB Phase Noise
1 Hz	−46 dBc/Hz
10 Hz	−66 dBc/Hz
100 Hz	−86 dBc/Hz
1 kHz	−106 dBc/Hz
10 kHz	−126 dBc/Hz

Figure 12.



The effects of random vibration on a crystal oscillator are a bit more complex. The power spectral density (PSD) normalized to a specific frequency must be taken into consideration for this calculation. The equation is:

$$SSB = 20 \log \left[\frac{\Gamma \cdot A \cdot f_o}{2 \cdot f} \right]$$

Using this equation and again assuming a "perfect oscillator" with a g sensitivity (Γ) of 1×10^{-9}/g and a power spectral density (PSD) of 0.1 g^2/Hz at 10 MHz, the resulting SSB phase noise would be:

Offset Frequency (from carrier)	SSB Phase Noise
1 Hz	−53 dBc/Hz
10 Hz	−73 dBc/Hz
100 Hz	−93 dBc/Hz
1 kHz	−113 dBc/Hz
10 kHz	−133 dBc/Hz

Note: SSB = Phase Noise in dBc
Γ = g sensitivity in Hz/g
f_o = carrier frequency in Hz
f_v = Frequency of vibration in Hz
A = $\sqrt{2 \cdot PSD}$ in a 1 Hz bandwidth
f = Frequency at point of interest on PSD curve in Hz.

Software is available for most computers which speed these computations and "model" proposed designs. The Γ or g sensitivity of the oscillator can be determined by using the following examples:

Crystal Frequency	Overtone	Γ or g-sensitivity
5 MHz	3rd	4×10^{-9}/g
10 MHz	3rd	2×10^{-9}/g
100 MHz	3rd	8×10^{-10}/g

*All crystals are SC cut utilizing 4 point mount holders.

Generally speaking, there are no operational effects on stability and aging caused by vibration. The user is cautioned that vibration induced problems can be minimized by providing as much detail as possible to the manufacturer during specification.

The user should specify as many requirements about a given oscillator application as are known. Manufacturers can be of help during the early phases of specification preparation, if asked.

Critical parameters to be addressed are:

1. Nominal output frequency
2. Setability
3. Aging
4. Output, including level, harmonics, subharmonics and spurious
5. Load
6. Short term stability
7. Phase noise
8. Electrical tuning
9. Warm-up time
10. Input supply voltage
11. Input supply current
12. Operating temperature range
13. Storage temperature range
14. Frequency stability versus temperature
15. Frequency stability versus load change
16. Frequency stability versus input voltage change
17. Package size
18. Specific military requirements
19. Environmental specifics
20. Semi-conductor screening or levels
21. Passive component screening or levels
22. Testing requirements
23. Data requirements
24. Quality/Inspection requirements
25. Details of end use

This may appear to be an extensive list. However, when dealing in a military or spacecraft application, it will be impossible for the manufacturer to adequately quote, design and manufacture a given oscillator without all of this information. The risks are high for the customer, manufacturer, and, of course, the ultimate customer. MIL-0-55310, which is the military specification for oscillators, can be helpful in the definition and understanding of performance parameters. Users are encouraged to ask questions.

REFERENCES

1. Dr. John Vig, "Quartz Crystal Resonators and Oscillators, a Tutorial," U.S. Army Electronics Technology and Devices Laboratory (LABCOM), 1986.
2. M. E. Freking, "Crystal Oscillator Design and Temperature Compensation," Van Nostrand, 1970.
3. B. Parzen, "Design of Crystal Oscillators and Other Harmonic Oscillators," John Wiley & Sons, 1983.
4. Hewlett Packard Application Note 207, "Understanding and Measuring Phase Noise in the Frequency Domain."

ACKNOWLEDGMENTS

I wish to thank my colleagues at Piezo Crystal Company for their guidance and support, and my wife, Barbara, for her support and freely giving of her literary skills.

DESIGN IN THE NONLINEAR WORLD -- THE METHOD OF HARMONIC BALANCE

Rowan J. Gilmore
Compact Software, Inc.
483 McLean Boulevard
Paterson, New Jersey 07504

ABSTRACT

Linear analysis programs are unable to quantitatively predict the behavior of nonlinear microwave components such as mixers, oscillators, frequency convertors, or amplifiers which are harmonically driven. The method of harmonic balance opens up a new realm of components that may be designed and whose performance may be predicted accurately and efficiently.

A tutorial overview of nonlinear computer-aided design techniques is presented, with emphasis given to the harmonic balance method. The distinction between linear and nonlinear analysis programs is made, and device modeling, convergence, and accuracy issues are touched. Examples of oscillator, mixer, and power amplifier design illustrate the use of currently available nonlinear CAD methods.

INTRODUCTION

The nonlinear analysis of electronic circuits has been an active area of study for many decades now. Nonlinear circuit analysis differs from linear circuit analysis in that the transfer characteristic between input and output is a function of drive level. The implications of this are enormous, and result in significant changes between the circuit description of a linear circuit compared to a nonlinear one. Firstly, all nonlinear components in the circuit must have large-signal models which describe their transfer function over the desired range of operating levels. Secondly, the analysis must account for any new frequency components which are generated. Thirdly, information on the bias sources and drive level must be conveyed to the analysis. Finally, the solution may not be unique, stable, or even exist.

SPICE is a general time-domain tool for the analysis of transient effects in a circuit, and although it can indeed analyze steady-state operation by using its transient analysis out to steady state, it is very inefficient when it does so. Although SPICE has been used for nonlinear analysis for over fifteen years, it is generally unsuited to the design of microwave and distributed RF nonlinear circuits. This class of circuits is inefficient for SPICE-like analysis because most circuits operating at RF frequencies and higher are steady-state, or the pulse widths employed are long enough relative to the RF frequency that the circuit can be treated as steady state. SPICE is not an effective tool for optimization or statistical analysis, as this requires sensitivity calculation and repetitive analyses. In addition, many microwave circuit elements are best modeled in the frequency domain (e.g. dispersive transmission lines, microstrip discontinuities) and cannot be easily incorporated into SPICE.

Prompted principally by the emergence of device models for the GaAs FET, and more readily available computing power, it was only in the early 1980's that effort was applied towards increasing simulation efficiency for steady-state analog

circuits. The harmonic-balance procedure has emerged as a practical and efficient tool for the design and analysis of steady-state circuits with sinusoidal excitation. Although these circuits are a subclass of the circuits which can be analyzed by SPICE, it includes most circuits of interest to the microwave engineer. The harmonic-balance simulator is up to two orders of magnitude more efficient than SPICE, and lends itself well to optimization as well as analysis of circuits including amplifiers, oscillators, mixers, frequency converters, and numerous types of control circuits such as limiters and switches, if transient effects are not of concern. Another major advantage of the harmonic balance method is that it is capable of representing linear circuits of practically any size, with no significant decrease in speed if additional internal nodes are added, or if elements of widely varying time-constants are used (such is not the case with SPICE).

The purpose of this paper is to describe the harmonic-balance method of analysis and to present examples illustrating its usefulness and power.

BACKGROUND TO NONLINEAR ANALYSIS

The harmonic balance technique has been reported in many previous papers [1,2,3]. For purposes of illustration, an example of an FET amplifier driven at the gate and drain by externally applied bias and drive voltages is presented.

In this example, the harmonic balance technique is presented here in an iterative form. Although not implemented in exactly this form in commercial CAD programs such as Microwave Harmonica [4], the iterative procedure presented more readily illustrates the principle. The method seeks to match the frequency components (harmonics) of current in a set of edges joining two subcircuits. The edges are chosen in such a way that nonlinear elements are partitioned into one subcircuit; and linear elements into the other. The edges at the linear - nonlinear interface connect the two circuits and define corresponding nodes; current flowing out of one circuit must equal that flowing into the other. Matching the frequency components in each edge satisfies the continuity equation for current. The current at each edge is obtained by a process of iteration so that the dependencies are satisfied for both the linear and nonlinear sides of the circuit.

Suppose the nonlinear circuit is represented by a nonlinear set of equations in the time domain

$$i_J(t) = g(v_1(t),\ldots,v_N(t)) \qquad (1)$$

where g is an arbitrary nonlinear function (and can include differentiation and integration), and i_J and v_J are the J th edge current and voltage, respectively. The dependent variables i_J are nonlinear functions of the independent variables v_J at some

point in time T_s. Periodic, steady state operation is assumed so that integrals and derivatives at T_s may be determined.

The linear circuit may be represented by an N x (N + M) matrix, obtained (internally) by standard linear circuit analysis programs such as SUPERCOMPACT [5]. The M additional variables are the additional external nodes (or edges) at which applied voltages (or currents) are present. The linear circuit matrix is calculated at each frequency component present in the circuit. In the case of an applied input signal which contains harmonically related components at w,2w,...qw, there will be (q + 1) matrices relating the independent variables at each edge to the dependent variables

$$\begin{pmatrix} v_1(k\omega) \\ \vdots \\ v_N(k\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(k\omega) & H_{12}(k\omega) & \dots & H_{1(N+M)}(k\omega) \\ H_{21}(k\omega) & H_{22}(k\omega) & \dots & H_{2(N+M)}(k\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1}(k\omega) & H_{N2}(k\omega) & \dots & H_{N(N+M)}(k\omega) \end{pmatrix} \begin{pmatrix} i_1(k\omega) \\ \vdots \\ i_N(k\omega) \\ v_{N+1}(k\omega) \\ \vdots \\ v_{N+M}(k\omega) \end{pmatrix} \quad (2)$$

for k = 0,1, ...q, where the $H_{ij}(kw)$ are impedance or transfer ratios depending on which of the variables are voltages and which are currents. The purpose of the harmonic balance program is to find a simultaneous solution to (1) and (2) for $v_1, v_2, \dots v_N$, so that $i_1, i_2, \dots i_N$ may be determined. Figure 1 illustrates the application of the technique to a three terminal device such as an FET. Two edges constitute the FET gate input and the FET

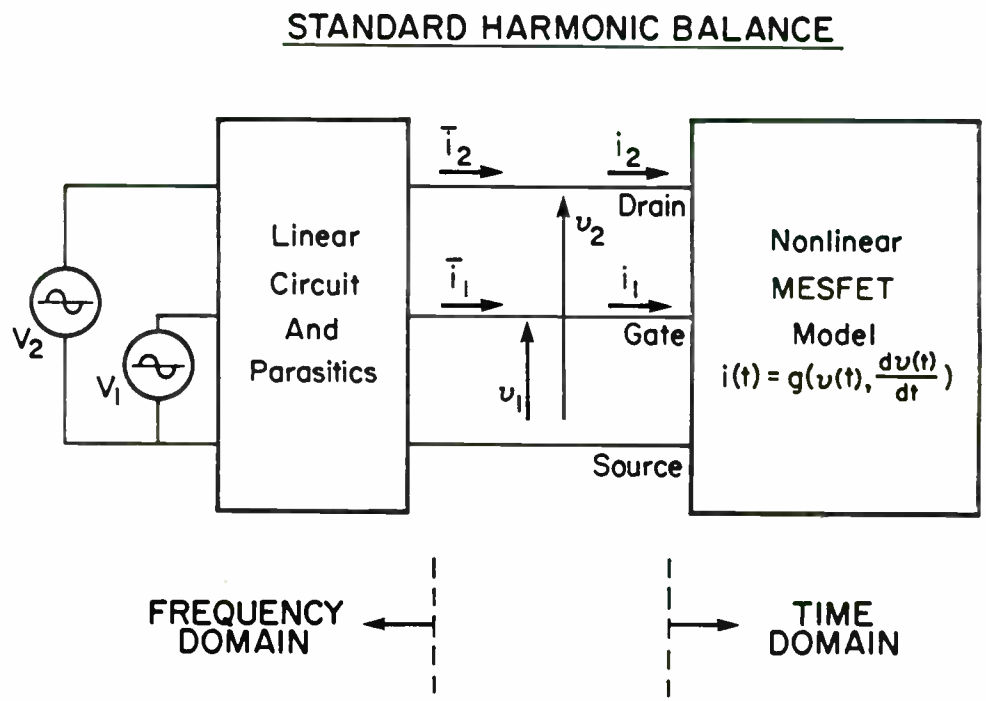


Figure 1: Analysis of an FET by the harmonic balance method, showing the partitioning of the circuit into linear and nonlinear subcircuits, and the definition of the variables at the linear-nonlinear interface.

drain output, separating the nonlinear FET elements into one subcircuit and the parasitics, matching, and output networks into another (linear) subcircuit. The third edge is the source of the FET and is chosen as the reference, so that N = 2. v_1 and v_2 are the independent variables; i_1 and i_2 are the dependent variables. Additional applied inputs are the external voltages V_1 and V_2. The desired output variables such as the current and voltage in the load can be found once i_1 and i_2 are determined.

Equation (1) is stated in the time-domain; (2) in the frequency-domain. Time to frequency conversion is achieved using the discrete Fourier transform (DFT). If estimates of $v_J(t)$ for J = 1, ...N at some time T_S are substituted into (1), i_J can be found at time T_S. If this is done at time instants $T_S, 2T_S, \ldots LT_S$ an L-point sequence of time samples of i_J results. The Nyquist sampling theorem states that if a sequence of points is obtained by sampling a waveform at a rate that is at least twice the highest frequency component contained, the original waveform can be reconstructed. If the waveform contains only discrete frequencies which are spaced by integral multiples of w, up to qw, one can set $T_S = (2\pi)/[(2q + 1)w]$ with L = (2q + 1) to satisfy the Nyquist criterion, and can extract the desired frequency components at w from the L-point sequence by using the discrete Fourier transform.

An initial estimate must be made for i_J and v_J because they are not known a priori. Iteration between (1) and (2) is performed using the DFT to obtain the frequency components from the time samples obtained from (1) until a self-consistent set of variables (i.e. those which satisfy the current continuity equations) is attained. The algorithm used in the analysis is given below.

1. Initial guesses are established for the current phasors $\overline{i_J}(kw)$ at the interface edges at the DC, fundamental, and harmonic frequencies (k = 0,1,...q). The overbar refers to the current flowing in the linear 'side' of the interface edges.

2. The hybrid matrix for the linear circuit H(kw) is calculated at DC, the driving frequency w, and each harmonic. This is used with $\overline{i_J}(kw)$ and the applied external voltages in (2) to calculate the unknown phasor components of voltage at each of the N edges.

3. Using an expression $v_J(t) = Real \sum_{k=0}^{q} v_J(k\omega) e^{jk\omega t}$ to derive the time value of the edge voltages at time $t = T_S, 2T_S, \ldots LT_S$, and a similar expression for derivatives, the time samples of voltage and its derivatives may be calculated at each of the N edges.

4. Values of $i_J(t)$ in the nonlinear 'side' of the interface edges may be obtained at corresponding time instants by substitution of the time samples of

voltages $v_J(t)$ and its derivatives into (1).

5. Using the discrete Fourier transform, the harmonic phasor components $i_J(kw)$ may be extracted from the L-point sequence of $i_J(t)$, if the sequence consists of samples obtained at the Nyquist rate.

6. An error function is formed to compare the 'nonlinear' current estimates i_J with the 'linear' estimates $\overline{i_J}$ so that

$$E(i_1, i_2, ..i_N, \overline{i_1}, \overline{i_2}, ...\overline{i_N}) = \sum_{k=0}^{q} (|i_1(k\omega) - \overline{i_1}(k\omega)|^2 + ... + |i_N(k\omega) - \overline{i_N}(k\omega)|^2) \quad (3)$$

In programs such as Microwave Harmonica [4], weighted terms corresponding to the difference between the desired optimization goal and the actual circuit performance can be added to the error function to allow simultaneous optimization of additional circuit elements or driving functions to achieve desired circuit performance.

7. The continuity equation for current states that the 'nonlinear' currents must equal the 'linear' currents. This corresponds to zero error function as a solution (if circuit optimization is not desired). The error function is minimized by forming new initial guesses for the current phasors $\overline{i}(kw)$ from the old estimates, and repeating steps (2) - (7) until the error function lies below some threshold. At this point, the linear and nonlinear partitions give self-consistent results, since the currents on each 'side' of the interface edges are equal. The quantities i_J and v_J are thus determined, and the voltage (or current) can be found at any desired node in the circuit (e.g. at the load) by linear analysis. If optimization criteria are added to the error function, the variable circuit elements are also adjusted to force the error function to zero on successive iterations.

The fixed-point method of Hicks and Khan (6) can be used in this example to achieve convergence and force the error function to zero by allowing the phasor currents to more closely approximate their true values on successive iterations. After the rth iteration of the loop, consider the current in the Jth edge $i_{J_r}(t) = \sum_k i_{J_r}(k\omega)e^{jk\omega t}$, with corresponding $\overline{i_{J_r}}(t) = \sum_k \overline{i_{J_r}}(k\omega)e^{jk\omega t}$. The next iteration is then carried out with $\overline{i_{J(r+1)}}(kw)$ formed by $\overline{i_{J(r+1)}}(k\omega) = p\, i_{J_r}(k\omega) + (1-p)\overline{i_{J_r}}(k\omega)$ where p is determined by convergence considerations and $0 < p \leq 1$. Hicks and Kahn and other authors [7,8] have investigated various criteria for convergence.

The example above illustrates the partitioning between the linear and nonlinear sides of the circuit. The linear side is analyzed efficiently in the frequency domain, and the nonlinear

side in the time domain. The discrete Fourier transform was used to "mesh" the two domains, although we must use other methods where noncommensurate sinusoidal signals (those with no integral multiple of a common period) are present in the system. A hybrid "H" matrix allows either voltage or current to be the independent variable and allows nonlinearities which are either voltage controlled or current controlled to be incorporated.

A general mathematical treatment of the harmonic balance method is found in Kundert et al [3]. Although that formulation uses only a "Y" matrix to represent the linear circuit i.e. Kirchoff's current law is used, it can be extended to the more general case of the hybrid matrix in which both KCL and KVL are applied, allowing any functional form of nonlinearity to be considered.

In this formulation, if the total circuit has N nodes, and if v is the vector of node voltage waveforms, then applying KCL to each node yields a system of equations

$$f(v,t) = i(v(t)) + \frac{d}{dt}q(v(t)) + \int_{-\infty}^{t} y(t-\tau)v(\tau)d\tau + i_s(t) = 0 \qquad (4)$$

where we have chosen the nonlinear circuit to contain only (voltage controlled) resistors and capacitors for representational ease. The quantities i and q are the sum of the currents and charges entering the nodes from the nonlinearities, y is the matrix impulse response of the linear circuit with all the nonlinear devices removed, and i_s are the external source currents.

In the frequency domain, the integral maps into YV, where V contains the Fourier coefficients of the voltage at each node and at each harmonic, and Y is a block node admittance matrix for the linear portion of the circuit. The system of equations (4) then becomes, on transforming into the frequency domain

$$F(V) = I(V) + j\Omega Q(V) + YV + I_s = 0 \qquad (5)$$

where Ω is a matrix with frequency coefficients representing the differentiation step. The notation used here is small letters to represent the time domain waveforms and capital letters the frequency domain spectra. This equation is, then, just KCL in the frequency domain for a nonlinear circuit. We have implicitly used Parseval's theorem in implying that the total energy of the system is equivalent when calculated in either the time or frequency domain. Harmonic balance seeks a solution to (5) by matching harmonic quantities at the linear/nonlinear interface. The first two terms are spectra of waveforms calculated in the time domain via the nonlinear model i.e.

$$F(V) = \Im i(\Im^{-1}V) + j\Omega\Im q(\Im^{-1}V) + YV + I_s \qquad (6)$$

and $\Im$ is the Fourier transform.

In practical CAD implementations of the harmonic balance technique [4], the discrete Fourier transform (DFT) is only used for circuits driven with fundamental and harmonic excitations. Many circuits, including mixers and amplifiers driven by intermodulation test inputs, have frequencies present that are not harmonically related. For circuits driven by incommensurate frequencies such as these, alternative numerical techniques can be used in place of the DFT to convey the time domain signal at the nonlinear interface into a frequency domain signal for application to the linear network. These include repetitive sampling at the band-pass rate and continued use of the DFT [9], or multi-dimensional Fourier transform techniques [10, 11]. The efficiency of the standard harmonic balance method is retained, because the linear circuit is still analyzed in the frequency domain by use of its Y- or hybrid matrix at each frequency component present. This is possible because all frequency components present in the circuit are known a priori so that an assumed solution form may be imposed in the analysis. A SPICE-like analysis, on the other hand, is not efficient because the form of the steady-state solution is not assumed prior to analysis. It should also be pointed out that SPICE has very poor efficiency for the intermodulation analysis because although the number of cycles necessary to reach steady state does not change, a very large number of cycles must be specified in order to analyze the slowly-varying output waveform, which will consist of many RF cycles beating slowly at the intermodulation rate.

DEVICE MODELING

Equation 1 is a time domain representation of the nonlinearity. The ability to use a time-domain representation is particularly appropriate because most physically-based models of solid-state devices involve the solution of semiconductor transport equations which are functions of time, not frequency. For example, models for a GaAs MESFET, such as those of Materka and Kacprzak [12], or Madjar and Rosenbaum [13], involve expressions which are instantaneous functions of the gate-source voltage, the drain-source voltage, and their derivatives, and are well suited for use with the harmonic balance method because the time samples of these quantities are readily calculable. The model is called many times within each iteration of the harmonic-balance loop, once at each value of time. The derivatives at the same time instants are needed for calculation of the currents in the nonlinear FET capacitances, but these are easily found by the program from the phasor components of voltage.

Programs such as Microwave Harmonica allow the user to define state variables with a user-supplied nonlinear device model. This is an extremely powerful approach and allows very general user definition of nonlinearities. The state-variable approach allows the user to arbitrarily define voltage or current (or combinations of them) as the nonlinear controlling variable. This is important not only for defining voltage-controlled nonlinearities, but also current-controlled nonlinearities. In

addition, a state variable can be embedded within a nonlinear model and need not be a terminal quantity i.e. one at the nonlinear-linear interface. The program then iterates on the state-variables (rather than interface currents, as we have described above), and calls the nonlinear models with time-samples of the state-variables. The user-defined model then returns to the program the interface (terminal) voltage and current at the same instants of time, which it calculates from the state-variable and model values. The interface quantities are used within the linear subcircuit, but the program now adjusts the values of the state-variables to force the error-function to zero.

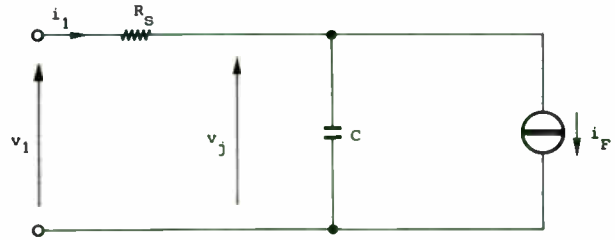
For example, consider the nonlinear model of Figure 2, which has a nonlinear capacitor and nonlinear resistor in series, and which may be embedded within a larger circuit. Choosing the junction voltage V_j as the state-variable allows us to write a very simple user-defined model as:

$$i_1 = C\frac{dv_j}{dt} + i_F$$

$$v_1 = v_j + R_s i_1$$

where i_1 and v_1 are the interface current and voltages calculated within the model to feed back to the harmonic balance algorithm. This model is called repeatedly from within the harmonic-balance program with V_j as input, and i_1 and v_1 as output. The state-variable approach of Microwave Harmonica has allowed two nonlinearities to be connected in series, which would not be possible if the program required i_1 to be calculated as a function of v_1. This situation can arise even in very simple device models, such as a Schottky diode which has a nonlinear space-charge capacitance in series with a nonlinear spreading resistance.

Generally, however, the user will utilize built-in library device models for which the assignment of state-variables is made automatically by the program.



All components of the equivalent circuit are assumed to be nonlinear functions of the voltage across the junction, denoted by v_j in the figure.

Figure 2. Simple nonlinear model for a microwave Schottky diode. i_F is the ideal diode current.

ACCURACY

The accuracy of the method is limited by

(i) the final value of the error function

(ii) the number of harmonics included in the analysis (i.e. truncation error). Truncation error arises because, theoretically, an infinite number of harmonics can be generated by the interaction of even a simple nonlinearity with the linear circuit. The error can be reduced, of course, by specifying additional harmonic components qw e.g. as the drive level is increased.

(iii) the dynamic range of the time-to-frequency conversion. As the resolution between two closely-spaced frequency components is reduced, the ability of the transform to distinguish between them becomes limited by the numerical accuracy of the computer. This error is small for most practical applications.

Naturally, errors also arise due to the quality of the device and linear circuit models, but these are not dependent on the type of analysis used.

EXAMPLES

Several circuit examples will be presented at the conference to illustrate the power which nonlinear analysis gives to the designer. Whereas linear analyses have in the past restricted designers to "frequency sweeps," "power sweeps" of circuits are now possible. The ability to quantitatively predict output power levels at all frequency components present in the circuit is likely to revolutionize microwave circuit design. Furthermore, optimization to achieve specified transducer gain, spectral purity, conversion gain, and DC to RF efficiency is possible because of the enhanced efficiency of the harmonic balance method. An example of optimization of an FET frequency doubler will be demonstrated. Examination of the output phase-plane (I-V) characteristics will be used to provide insight into device operation.

CONCLUSION

The harmonic balance method is an efficient technique for analyzing analog circuits under steady-state sinusoidal excitation. It is useful at microwave frequencies because most circuits are of this type, and also because the device models are generally time-domain models while the linear circuit elements are often frequency domain representations (e.g. through spectral domain analysis). This allows all existing linear circuit element models to be used. The efficiency gained by this representation, and by imposing the known solution form onto the circuit before the analysis commences, results in speed improvements that are several orders of magnitude better than for time-domain

techniques such as SPICE. The dual-frequency analysis capability is also difficult to implement effectively in SPICE if the frequencies are closely spaced.

The improvement in efficiency now makes it possible to optimize circuits for given nonlinear responses. Mixers, oscillators, power amplifiers, limiters and PIN attenuators are among the types of circuits than can be quantitatively designed using this commercially available technique. The ability to predict compressed output powers, spectral purity, DC to RF and RF to RF efficiencies, and transducer and conversion gains under varying drive are likely to result in tremendous productivity gains as the design and optimization of nonlinear components is transferred from the workbench to the computer.

BIBLIOGRAPHY

[1] R. J. Gilmore, "Nonlinear Circuit Design Using the Modified Harmonic Balance Method," IEEE Trans. on Microwave Theory and Tech., Vol. MTT-34, No. 12, pp. 1294-1306, December 1986.

[2] V. Rizzoli, A. Lipparini, and E. Marazzi, "A General Purpose Program for Nonlinear Microwave Circuit Design," IEEE Trans. on Microwave Theory and Tech., Vol. MTT-31, No. 9, pp. 762-770, September 1983.

[3] K. S. Kundert and A. Sangiovanni-Vincentelli, "Simulation of Nonlinear Circuits in the Frequency Domain," IEEE Transactions on Computer-Aided Design, Vol. CAD-5, No. 4, October 1986.

[4] Microwave Harmonica, Compact Software, Inc., Paterson, New Jersey.

[5] SUPERCOMPACT, Compact Software Inc., Paterson, N.J.

[6] R. G. Hicks and P.J. Kahn, "Numerical Analysis of Nonlinear Solid-State Device Excitation in Microwave Circuits," IEEE Trans. on Microwave Theory and Tech., Vol. MTT-30, PP. 251-259, March 1982.

[7] C. Camacho-Penalosa, "Numerical Steady-State Analysis of Nonlinear Microwave Circuits with Periodic Excitation," IEEE Trans. on Microwave Theory and Tech., Vol. MTT-31, pp 724-730 September 1983.

[8] A. R. Kerr, "Noise and Loss in Balanced and Subharmonically Pumped Mixers: Parts I and II: Theory and Application," IEEE Trans on Microwave Theory and Tech., Vol. MTT-27, pp. 938-950, December 1979.

[9] R. J. Gilmore and F. J. Rosenbaum, "Circuit Design to Reduce Third Order Intermodulation Distortion in FET Amplifiers," IEEE 1985 Int. Microwave Symposium Digest, pp 413-416.

[10] A. Ushida and L.O. Chua, "Frequency-Domain Analysis of Nonlinear Circuits Driven by Multi-Tone Signals," IEEE Trans. on Circuit and Systems, Vol. CAS-31, No. 9, September 1984.

[11] G. B. Sorkin, K.S. Kundert, and A. Sangionvanni-Vincentelli, "An Almost-Periodic Fourier Transform for use with Harmonic Balance," IEEE 1987 Int. Microwave Symposium Digest, pp. 717-720.

[12] A. Materka and T. Kacprzak, "Computer calculation of large-signal GaAs FET amplifier characteristics," IEEE Tans. of Microwave Theory and Tech., Vol. MTT-33, pp. 129-135, February 1985.

[13] A Madjar and F. J. Rosenbaum, "A large-signal model for the GaAs MESFET," IEEE Trans. on Microwave Theory and Tech., Vol. MTT-29, pp. 781-788, August 1981.

A PLL Synthesizer Utilizing a New GaAs Phase Frequency Comparator

by

Dan T. Gavin
Avantek, Inc.

and

Ronald M. Hickling
GigaBit Logic, Inc.

INTRODUCTION

This paper reviews the tradeoffs in the design of a new GaAs phase/frequency comparator and presents results for several PLL configurations employing this device (GigaBit Logic model 16G044 phase frequency comparator). Comparisons are made with the functionally similar Motorola ECL MC12040. Methods for measuring phase detector noise in a PLL and ultimate reference suppression are covered, including measured data for the 16G044 and MC12040. Maximum operating frequency of the 16G044 is also investigated.

I. TRADEOFFS IN PHASE/FREQUENCY COMPARATOR DESIGN

The 16G044 is also similar to the phase detector 1 of the TTL MC4044 from Motorola. The 16G044, phase detector 1 of the MC4044 and the MC12040, are all Type 4 phase detectors.

One of the traditional concerns about the digital phase/frequency comparator has been the reported crossover distortion in the neighborhood of zero phase difference between the R and V input ports. Certain types of crossover nonlinearities are fatal to the proper operation of a loop. Three such nonlinearities are shown in Fig. A1: (A1a) a region in which the slope changes sign; (A1b) a region in which the small signal gain drops to zero; and (A1c) a discontinuous characteristic around zero phase which results in drastically different small signal gains between the "R leads V" and "R lags V" cases. The respective results of these nonlinearities are: (1) an unpredictable lock due to the existence of more than one valid lock point; (2) a "wandering" in phase due to a small signal loop gain of zero; and (3) an unpredictable acquisition characteristic due to loop parameters which vary as a function of phase. Items (1) and (3) are particularly dangerous since they present the potential for loops to possess regions of small signal instability.

On the other hand, "weakly" nonlinear characteristics in the phase detector are acceptable for many (perhaps even most) PLL applications. Qualitatively, an acceptable characteristic should possess these properties (besides not possessing the qualities outlined above!):

1. Any nonlinearities should be confined to a small region around zero phase.
2. The transfer characteristic should be symmetrical about zero phase.
3. The ratio of the maximum to the minimum small signal sensitivity should be small (in general less than five; less than two for high performance loops).

As is mentioned in Reference (6), both the Motorola MC4044 and Fairchild 11C44 suffer from an unacceptable level of crossover distortion for most precision PLL applications. However, this distortion may be due mainly to the asymmetry in the dynamic response of the charge pump circuits of those devices rather than

an inherent flaw in their logical design as a casual reading of Reference (6) might suggest. In fact, it will be shown shortly that the linearity of a phase/frequency comparator is almost exclusively a function of the rise and fall time limitations in its output buffers compared with the width of the pulse on the inactive (U or D) output. With this in mind, it is appropriate to review the architectures of various phase/ frequency comparators that have been developed to date.

Figure A2(a) shows the logic diagram of a "conventional" phase/frequency comparator. It can be shown by a logic timing analysis that the maximum operational speed of such an architecture is 1/(10td), where td is the internal gate propagation delay of the device. Furthermore, the width of the pulse on the inactive output is 1 td. An analogous architecture was utilized by NTT (Reference [7]) which simply replaced all of the NAND gates with NOR gates. This is shown in Figure A2(b). By utilizing GaAs technology, the authors of this paper were able to achieve a maximum operating frequency of 600 MHz implying a td of 167 ps. For the MC 12040 (a silicon device that is used as a benchmark in this paper), td is approximately 1.5 ns; from the typical operating frequency of 80 MHz, one can anticipate that it utilizes a special architecture which operates at 1/8 td.

Authors at Hughes (Reference [6] utilized a less conventional architecture to create a 1/(14td) architecture GaAs device operating at 480 MHz. With a pulse width of 5 td on the inactive output, the architecture essentially "throws away" maximum operating speed performance in favor of perfect linearity. Parenthetically, it is interesting to note that a "compounded logic" equivalent of the Hughes device would be capable of operating at 1/(10td) with an inactive output pulse width of 3 td. Finally, the 16G044 utilizes a special 1/(8td) architecture which makes it capable of operating 1.25 times faster than a conventionally architectured device using identical process technology. The inactive output pulse width is again 1 td. Laboratory tests have shown devices to operate as high as 1.2 - 1.3 GHz, indicating a td of 96 - 104 ps.

It is possible to estimate the crossover distortion by using a simple slew rate approximation to rise and fall time. Figure A3 shows the waveforms and equations used to compute the average output voltage from a driver circuit having rise and fall times of Tr and Tf respectively. The analysis assumes that the buffer completely isolates the driven load from the internal logic such that internal logic propagation delays are independent of output load. Furthermore the internal slew rates of the part are assumed to be much faster than those of the output driver. Note that in Case 1 where the pulse width (Tp) appearing at the output is greater than average between the rise and fall times, the output buffer has perfect fidelity in representing the average output voltage associated with that pulse. However, as shown in Case 2, when the pulse width is less than the average between the rise and fall times, distortion causes the output voltage to follow a "square law" characteristic.

The results of the foregoing analysis are plotted in Figure A4. The model agrees fairly well with actual measurements in the

laboratory. The worst case distortion occurs when td is negligibly small compared to the rise time of the output buffers. An illustrative example is the case of a GaAs phase/frequency comparator utilizing a CMOS output driver. The majority of the devices available today have a ratio between output rise time and internal propagation delay of approximately 2-3; thus the nonlinear behavior of these devices is confined to a region where the skew between R and V is less than 2/5 of the output rise time. For the MC12040 this translates into 1.3 ns and for the 16G044 this number is about 100 ps. The worst case small signal sensitivity occurs in the neighborhood of zero skew and is equal to the slope of the square law characteristic. With the same ratio of 2-3 between rise time and propagation delay, the reduction in phase detector gain is also about 2-3, which is acceptable for a variety of PLL applications.

One other issue which must be considered is reference rejection (see Section III). One of the disadvantages of achieving perfect linearity through the widening of the inactive output pulse width is the increase in output "spurious" content at the reference frequency. Thus, the "linear" architecture is best suited to loops with sufficiently narrow loop bandwidths to reject the spurious reference signal content while the faster, weakly nonlinear architectures are best suited to wide band loops.

II. PHASE NOISE MEASUREMENT

The approach used is to lock the phase detector (16G044) in a PLL with sufficient bandwidth to obtain phase noise data out to 1 MHz, and configured so the phase detector is the dominant noise source in the loop. Two measurements are made, N=110 and N=220. If there are no other noise sources degrading the measurement, there will be 6 dB difference in the two measurements. From these measurements the SSB spectral noise density data can be obtained.

The configurations used for these two measurements are shown in Figure 1 with their respective open loop gains plotted in Figure 2. (For a detailed schematic, contact the author). The phase detector gain, Kd, of .11 to .14 volts/radian for the 16G044 yields loop bandwidths consistent with standard PLL design techniques. The first zero is chosen to reduce the phase noise contributions due to the VCO and op amps for frequency offsets less than 15 KHz and to minimize closed-loop peaking. (The predicted closed-loop response is shown in the appendix.)

The predicted noise of all sources in the loop; i.e., reference, op amp input noise voltage and cascaded op amp output noise voltage for N=110 and N=220 are shown in Figures 3 and 4, respectively. (This data is calculated from two basic programs, OPPNOISE AND PLNOISE, included in the appendix.) These two figures show that the noise contribution due to the op amps will add approximately 1dB to the measurement at 1 MHz which can be accounted for. Test results are shown in Figure 5 for the 16G044 loop for N=110 and N=220. The same measurements are made for the MC12040. However, the measurement system is limited by the noise floor of the HP8566 spectrum analyzer and is only valid for frequency offsets greater than 10 KHz for the MC12040, therefore data from earlier measurements is used in this region. From this

data the SSB spectral noise density is extracted and plotted in Figure 6. The higher noise for the 16G044 (inherent in GaAs D-MESFET devices) is 1/f noise and begins at 200 KHz.

III. ULTIMATE REFERENCE SUPRESSION

In high performance PLLs, one of the largest problems is to obtain maximum loop bandwidth and simultaneously achieve very low reference sidebands at the output. The 16G044 has a theoretical 17dB improvement over the MC12040 due to its 100ps propagation delay and 300ps output rise and fall times.

Consider the near ideal case: voltage offsets in the loop caused by variations in phase detector outputs, loop filter components and VCO modulation sensitivity are very small. Also, the increase in pulse width of the phase detector U or D outputs required to compensate this offset error is less than the propagation delay of the device. Under these conditions, phase detector pulse width, which is equal to the propagation delay of the device, becomes the dominant cause of reference feedthrough at the output. For a given phase detector, the minimum pulse width is measured by injecting identical signals into the R and V inputs as shown in Figure 7. Typical output pulses for the MC12040 and 16G044 are shown in Figure 8. These pulses will produce sidebands at the output of the VCO according to the formula in Figure 9. (This does not take into account higher order effects of the op amp.) The ratio of Vpk for the MC12040 and the 16G044 is 17.2 dB. This means that 17 dB less filtering is required for 16G044-based loops and the phase margin will have improved by an amount dependent on the loop configuration.

To verify the above analysis, the loop in Figure 10 is used and reference sidebands for both the MC12040 and the 16G044 are measured. The same loop is used for both phase detectors since their gain, Kd, is approximately .14 volts/radian. Any offsets are nulled out by injecting the appropriate current through the 100k resistor at the inverting input of A1. The results shown in Figure 11 reveal that the 16G044 is 19dB better than the MC12040 at 10MHz which is 2dB more than calculated.

IV. MAXIMUM FREQUENCY OF OPERATION

The configuration used to determine maximum operating frequency is shown in Figure 12. This measurement is also used to verify the predicted op amp noise in Figures 3 and 4. The 16G044 used is a -3, 750MHz version, which works well up to 789MHz. The typical maximum operating frequency for the MC12040 is 80MHz. The phase noise is plotted in Figure 13 and shows that the op amp noise is in agreement with that predicted in Figures 3 and 4.

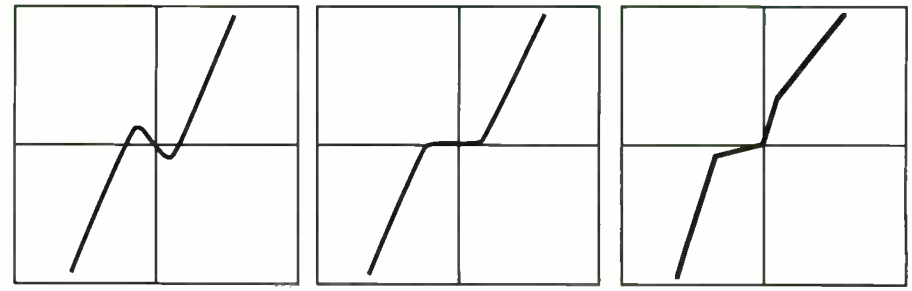
V. CONCLUSIONS

The 16G044 is an impressive device with an operating frequency up to 1GHz and a theoretical improvement of 17dB of ultimate reference suppression over the MC12040. Phase noise may limit its use in applications that require low close in phase noise, wide bandwidth and high divide ratios. However, for single frequency PLL's, or low divide ratios, improved noise performance can be obtained over the MC12040 for frequency offsets greater than 1kHz due to the high operating frequency of the 16G044. For

example, if a 960MHz signal is required from a single loop PLL, the MC12040 would require N=12 with a reference of 80MHz which would increase its contribution to the output noise by 21.5dB. If we used the 16G044 with N=1 and a reference frequency of 960MHz, the noise at 500kHz would be 19.5dB better than that of the loop using the MC12040. This assumes, of course, that all other noise sources are better than the noise of the 16G044. Other applications for the 16G044 include: cleanup loops, replacement for double balanced mixers and associated search circuitry and anywhere an MC12040 is used.

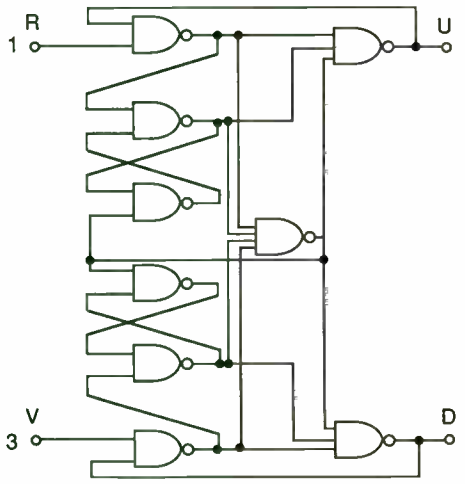
VI. REFERENCES

1. Best, Roland E., "Phaselocked Loops Theory, Design and Applications".
2. GigaBit Logic 16G044 Data Sheet.
3. Martin, Larry, "Noise-Property Analysis Enhances PLL Designs," EDN, September 16, 1981.
4. ECL MC4044 Application Note, ECL Data Book, page 7-34, 1983 Motorola, Inc.
5. Rohde, Ulrich L., "Digital PLL Frequency Synthesizers Theory and Design," Prentice Hall, New Jersey, 1983.
6. Shariary, I., et al, "GaAs Monolithic Digital Phase/Frequency Discriminator," 1985 GaAs IC Symposium Proceedings.
7. Osafune, K., et al, "Ultra-High Speed GaAs Monolithic Prescaler and Phase/Frequency Comparator IC" IEEE Transactions on MTT, 7/87.

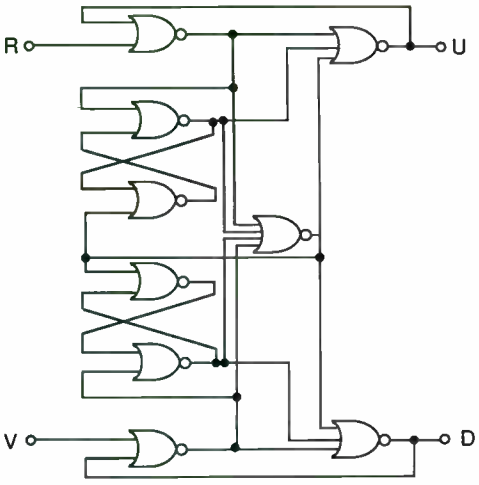


(a) Slope changes sign (b) Small signal gain goes to zero (c) Discontinuous gain characteristic

FIGURE A1: TYPES OF PHASE DETECTOR CROSSOVER NONLINEARITIES



a) MC 4044 ALL NAND ARCHITECTURE

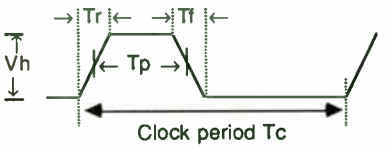


b) NTT ALL NOR ARCHITECTURE (Reference [7])

FIGURE A2: CONVENTIONAL PHASE/FREQUENCY COMPARATOR ARCHITECTURES

FIGURE A3: PHASE/FREQUENCY COMPARATOR DISTORTION DUE TO FINITE SLEW RATES IN THE OUTPUT BUFFERS

CASE 1: The represented pulse width (Tp) is greater than the average between the rise and fall times.

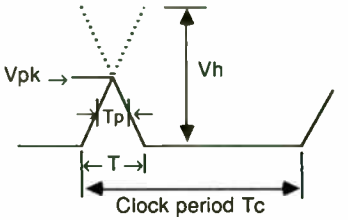


$$Tp > \frac{Tr + Tf}{2}$$

$$\overline{V} = \frac{1}{Tc}\left[\left(\frac{Tr + Tf}{2}\right)Vh + \left(Tp - \frac{Tr + Tf}{2}\right)Vh\right]$$

$$= Vh\frac{Tp}{Tc}$$

CASE 2: The represented pulse width (Tp) is less than the average between the rise and fall times.



$$Tp < \frac{Tr + Tf}{2}$$

$$T = 2Tp$$

$$Vpk = \frac{2TpVh}{Tr + Tf}$$

$$\overline{V} = \frac{1}{Tc}\left(\frac{1}{2}TVpk\right) = \left[\frac{1}{2}\left(2Tp\right)\left(\frac{2TpVh}{Tr + Tf}\right)\right]\frac{1}{Tc}$$

$$= Vh\ \frac{2Tp^2}{Tc\,(Tr + Tf)}$$

For the inactive output, Tp = 1 td; for the active output Tp = Tskew* + 1 td.
Thus the comparator output is:

$$Vo = \frac{2Vh}{Tc\,(Tr + Tf)}\left[(Tskew + td)^2 - td^2\right] = \frac{2Vh}{Tc\,(Tr + Tf)}\left[Tskew^2 + 2Tskew\ td\right]$$

Therefore, the small signal sensitivity for Tskew < $\frac{Tr + Tf}{2}$ is:

$$kd = \frac{\partial Vo}{\partial Tskew} = \frac{4Vh}{Tc\,(Tr + Tf)}\left[Tskew + td\right]$$

* Tskew is the difference in phase between phase detector R and V inputs.

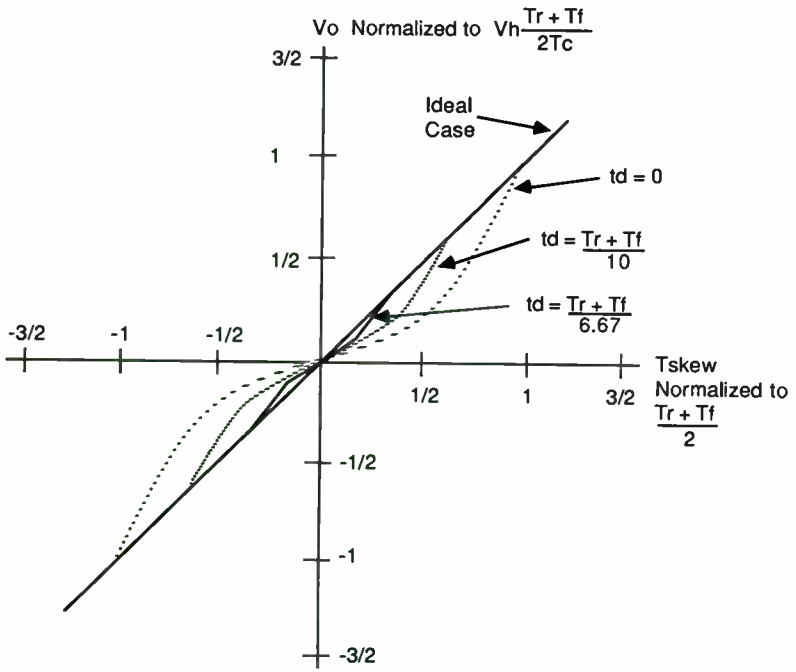
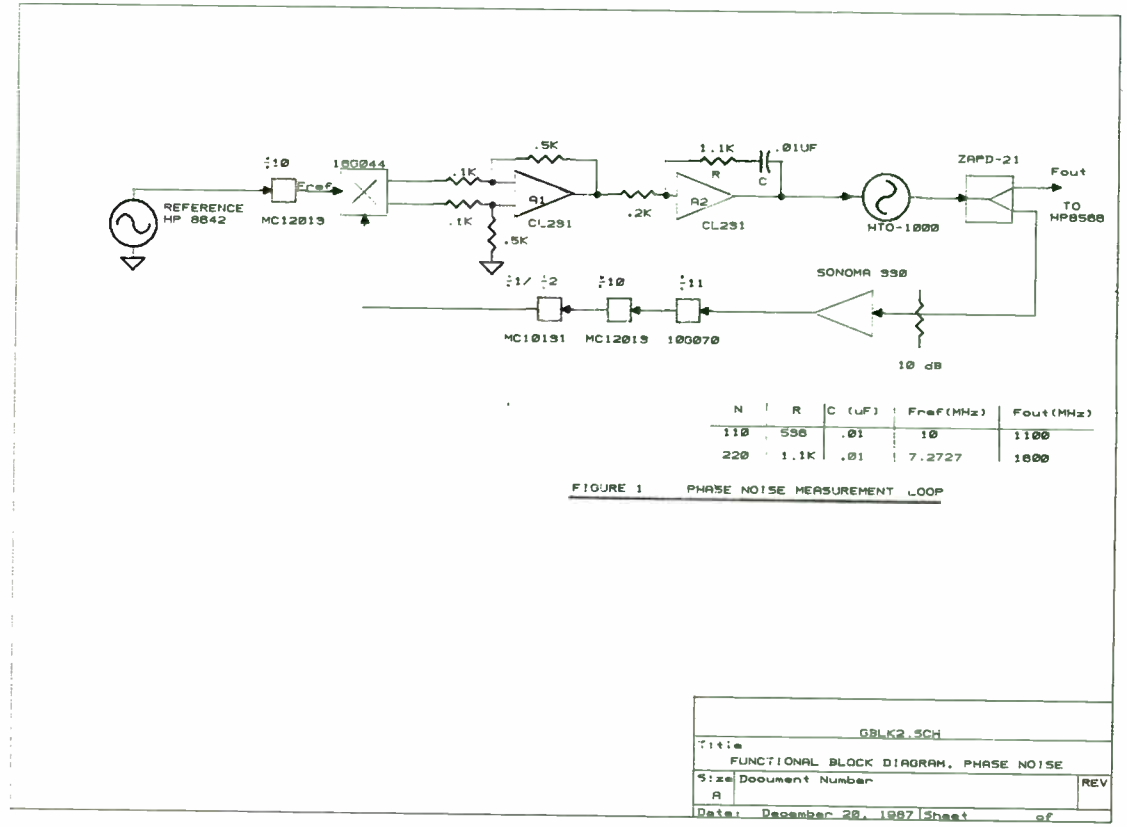
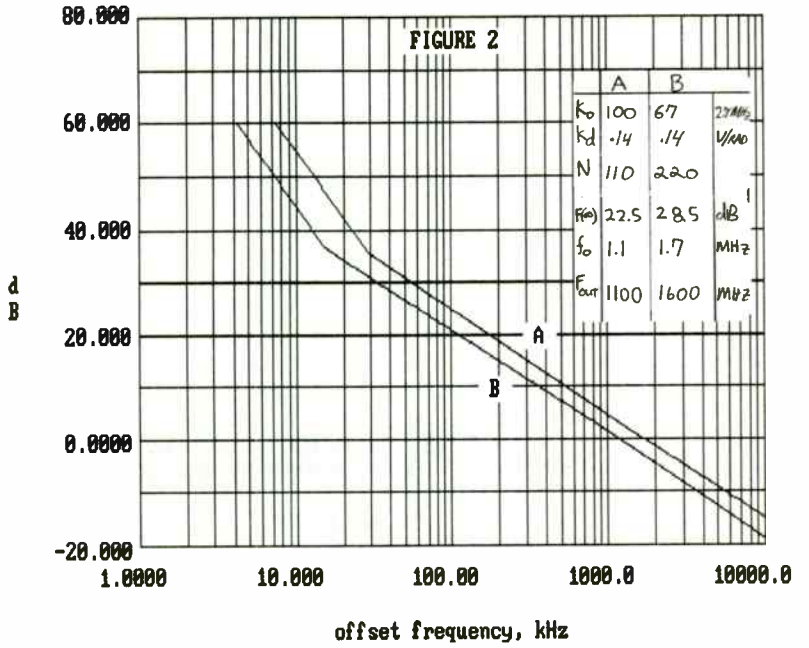


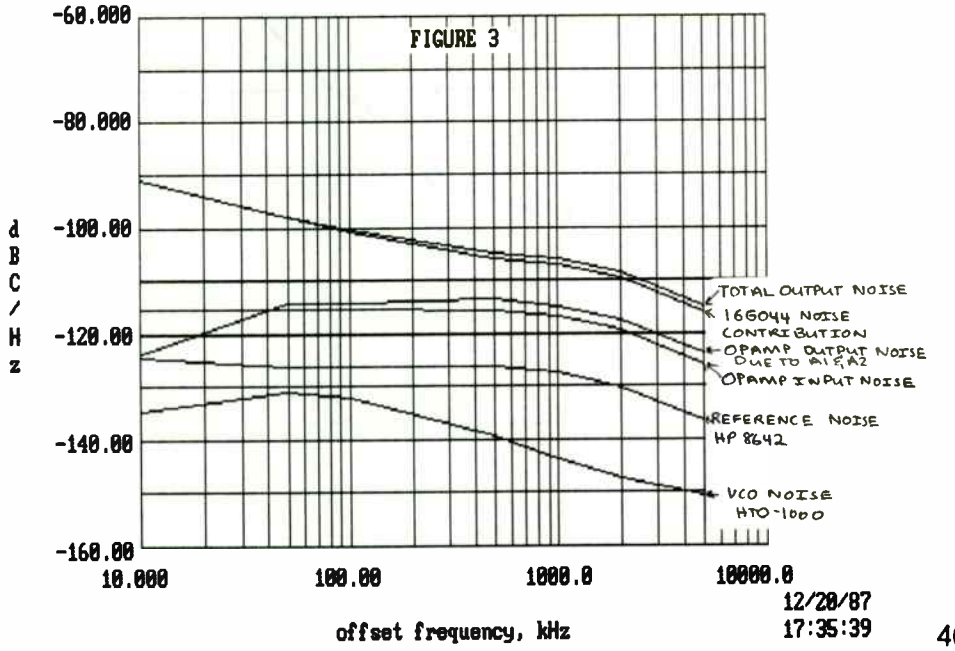
FIGURE A4: A PLOT OF COMPARATOR OUTPUT VERSUS SKEW FOR ARCHITECTURES WITH AN INACTIVE OUTPUT PULSE WIDTH OF 1 td



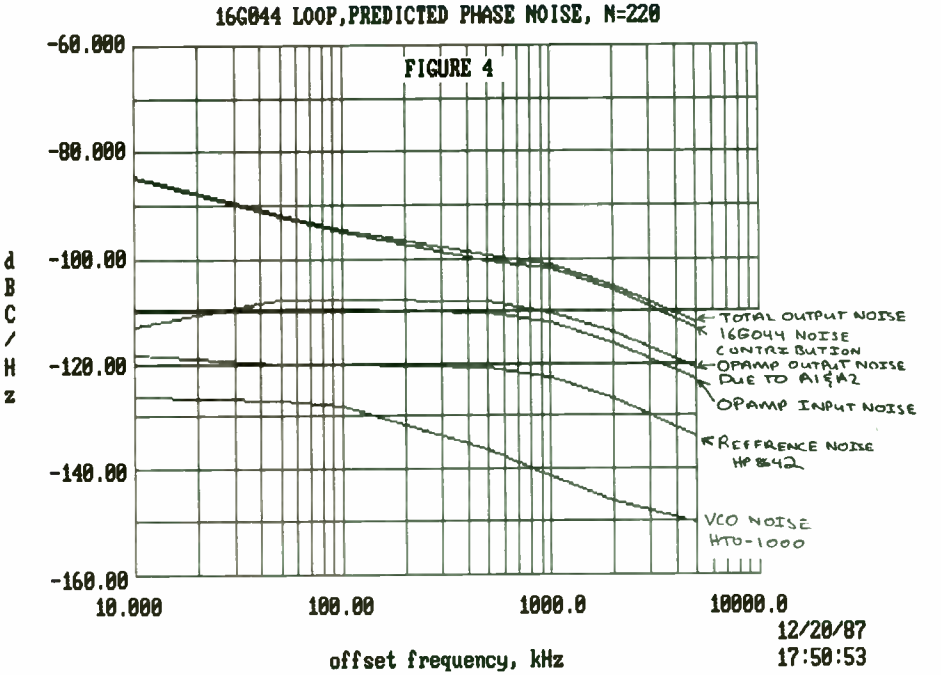
OPEN LOOP GAIN, 16G044 LOOP



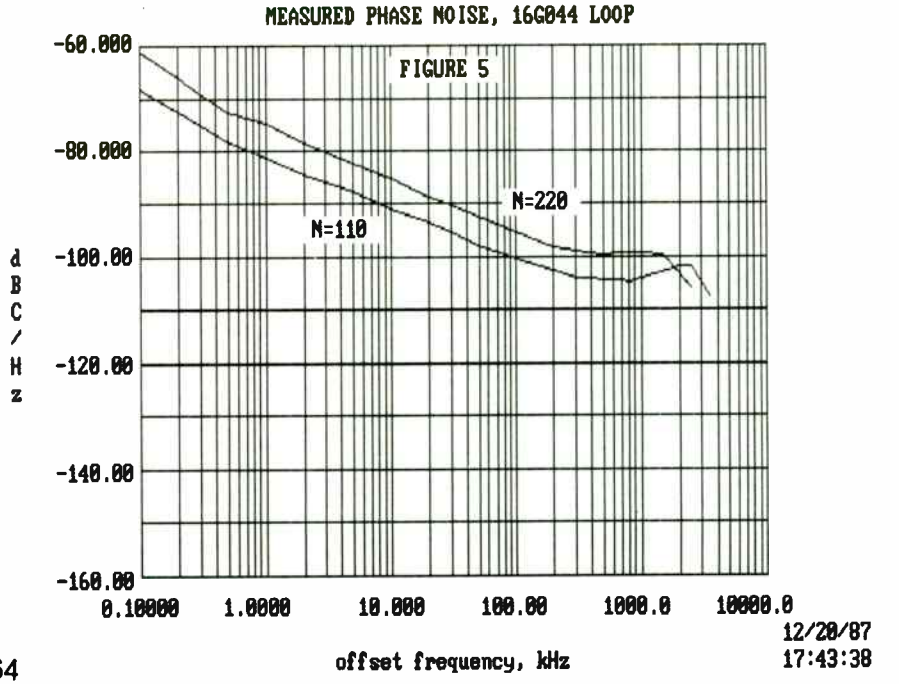
16G044 LOOP, PREDICTED PHASE NOISE, N=110

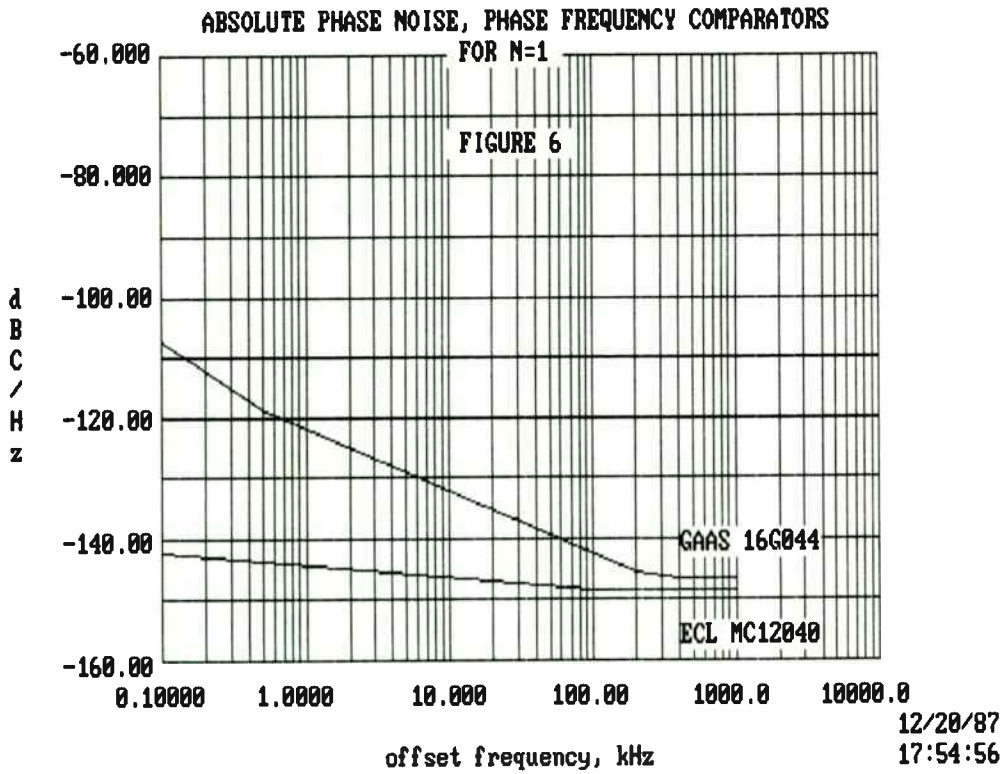


16G044 LOOP, PREDICTED PHASE NOISE, N=220



MEASURED PHASE NOISE, 16G044 LOOP





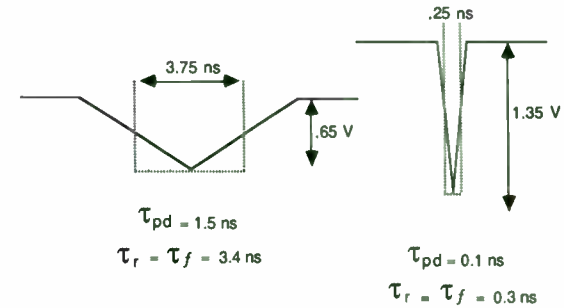


FIGURE 8:
TYPICAL PULSE WIDTHS FOR
MC12040 AND 16G044

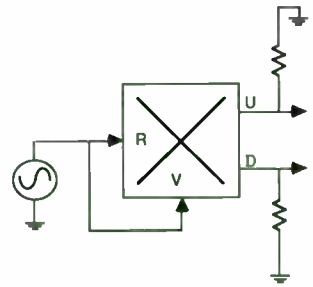
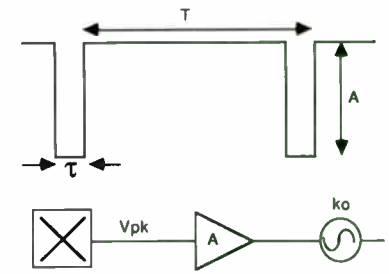


FIGURE 7:
PHASE DETECTOR MINIMUM PULSE
WIDTH DETERMINATION



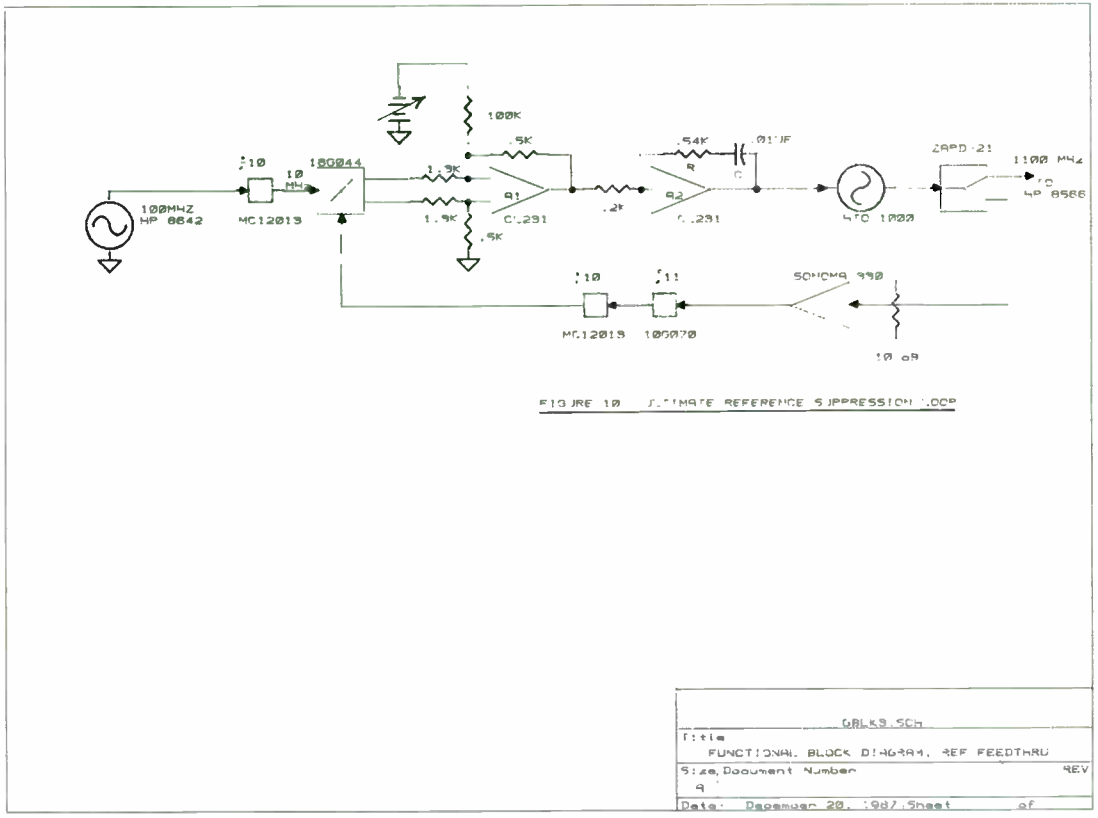
$$\text{SIDEBAND} \atop \text{CARRIER} = 20 \log \left[\frac{Vpk \; A \; Ko}{2 \; \omega m} \right] \text{ in dBc}$$

$$Vpk = 2 A \frac{\tau}{T}$$

Ko = VCO MODULATION SENSITIVITY (RADS/Hz)

$\omega m = 2\pi \cdot$ OFFSET FREQUENCY (RADS)

FIGURE 9:
CALCULATION OF SIDE BANDS



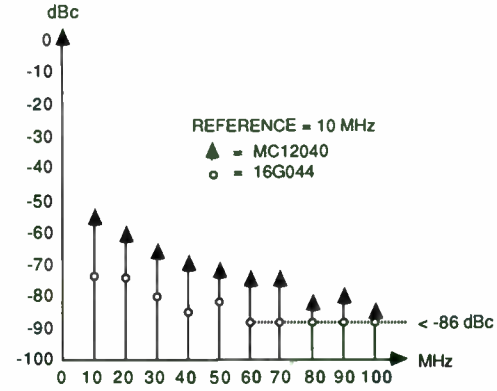


FIGURE 11: ULTIMATE REFERENCE SUPPRESSION

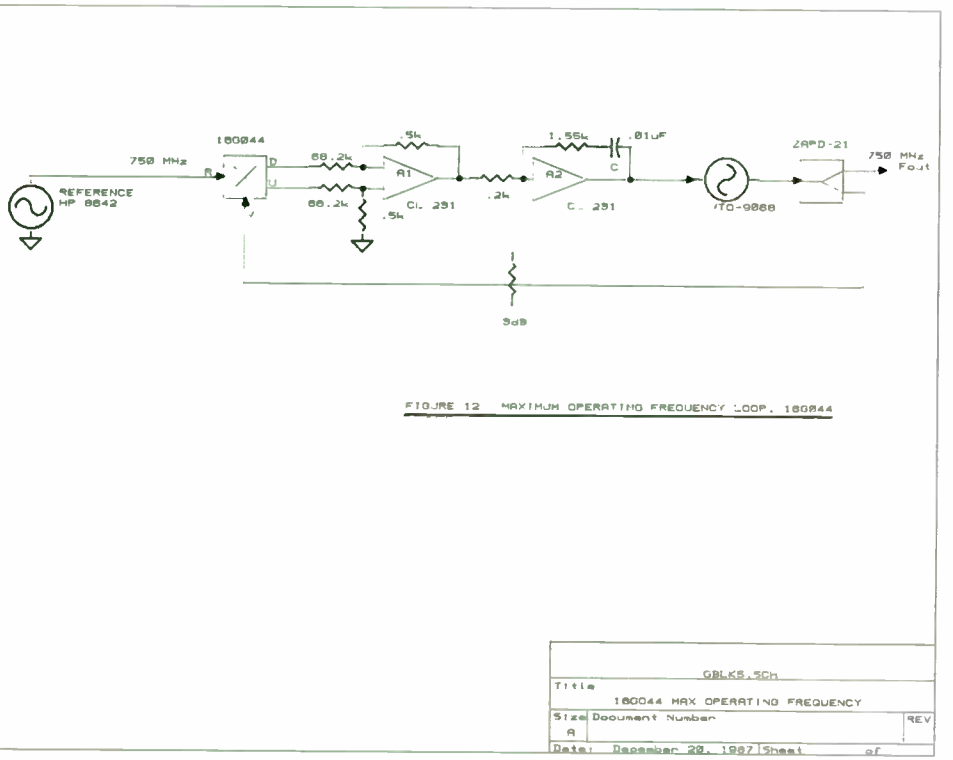
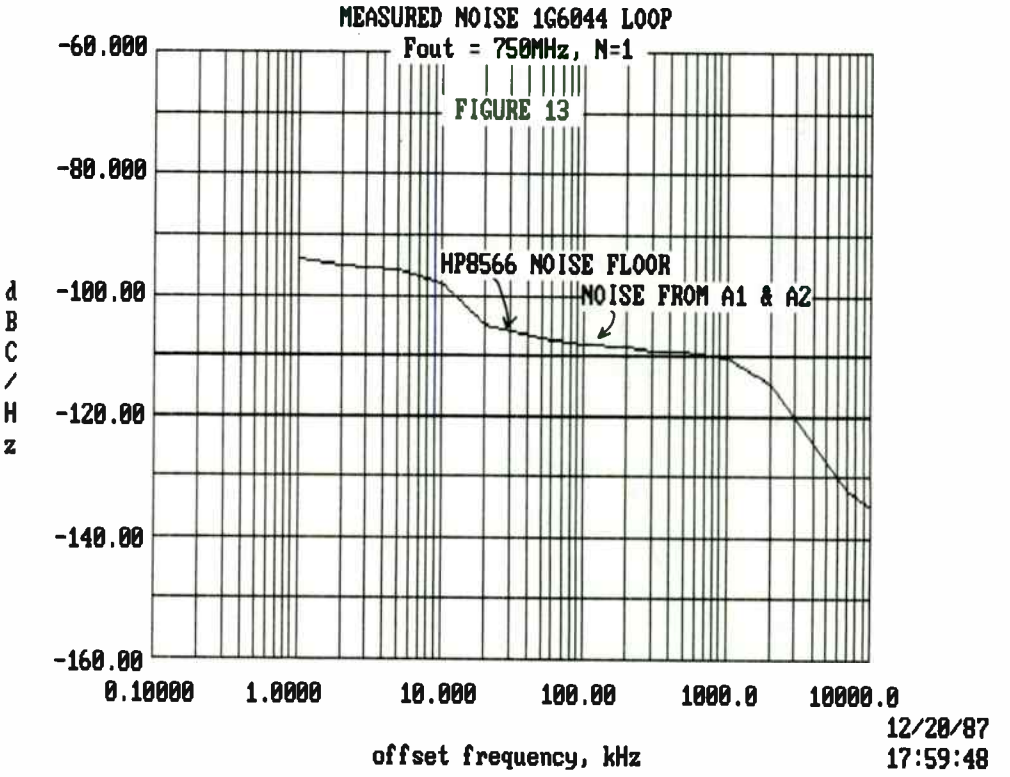
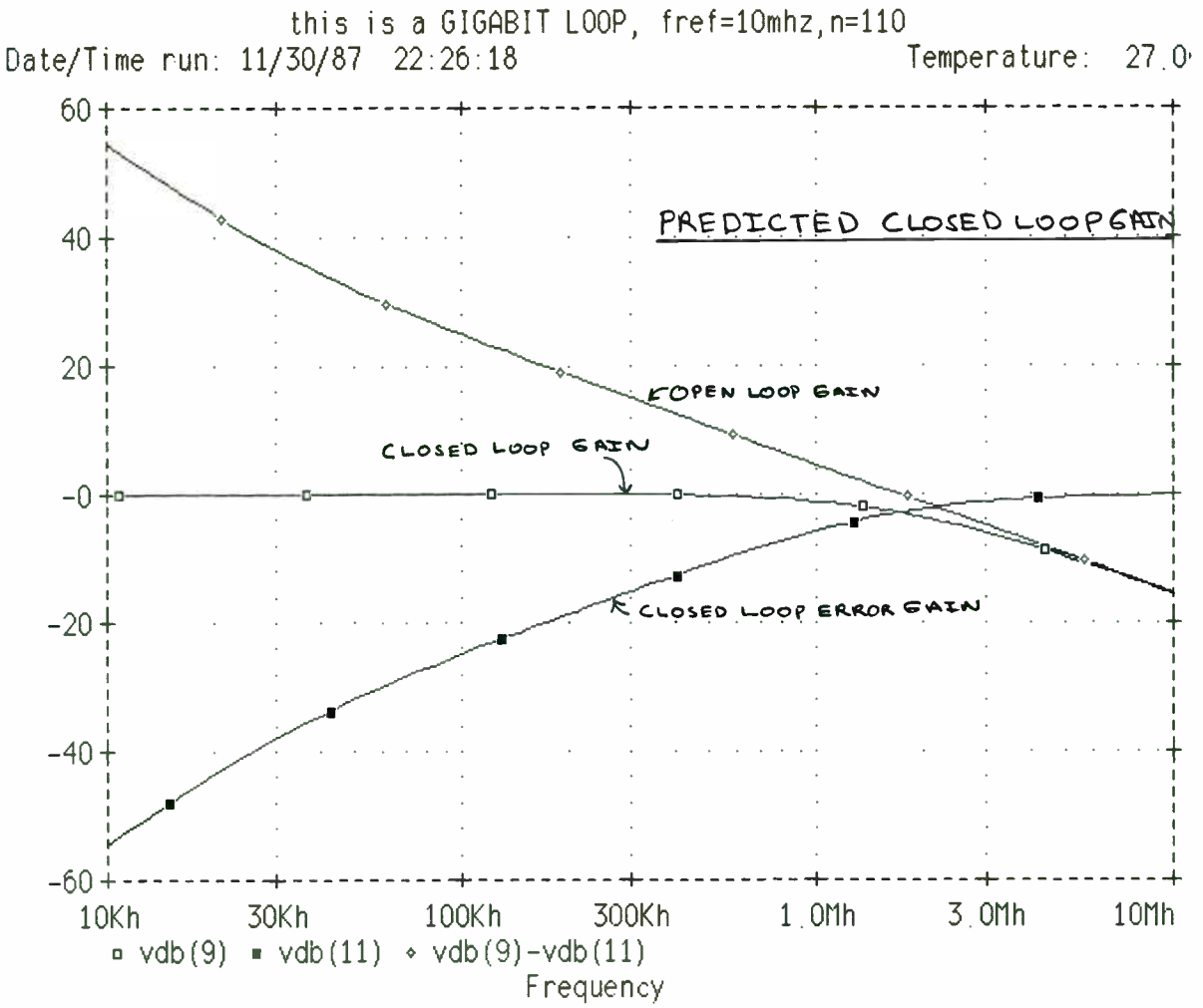
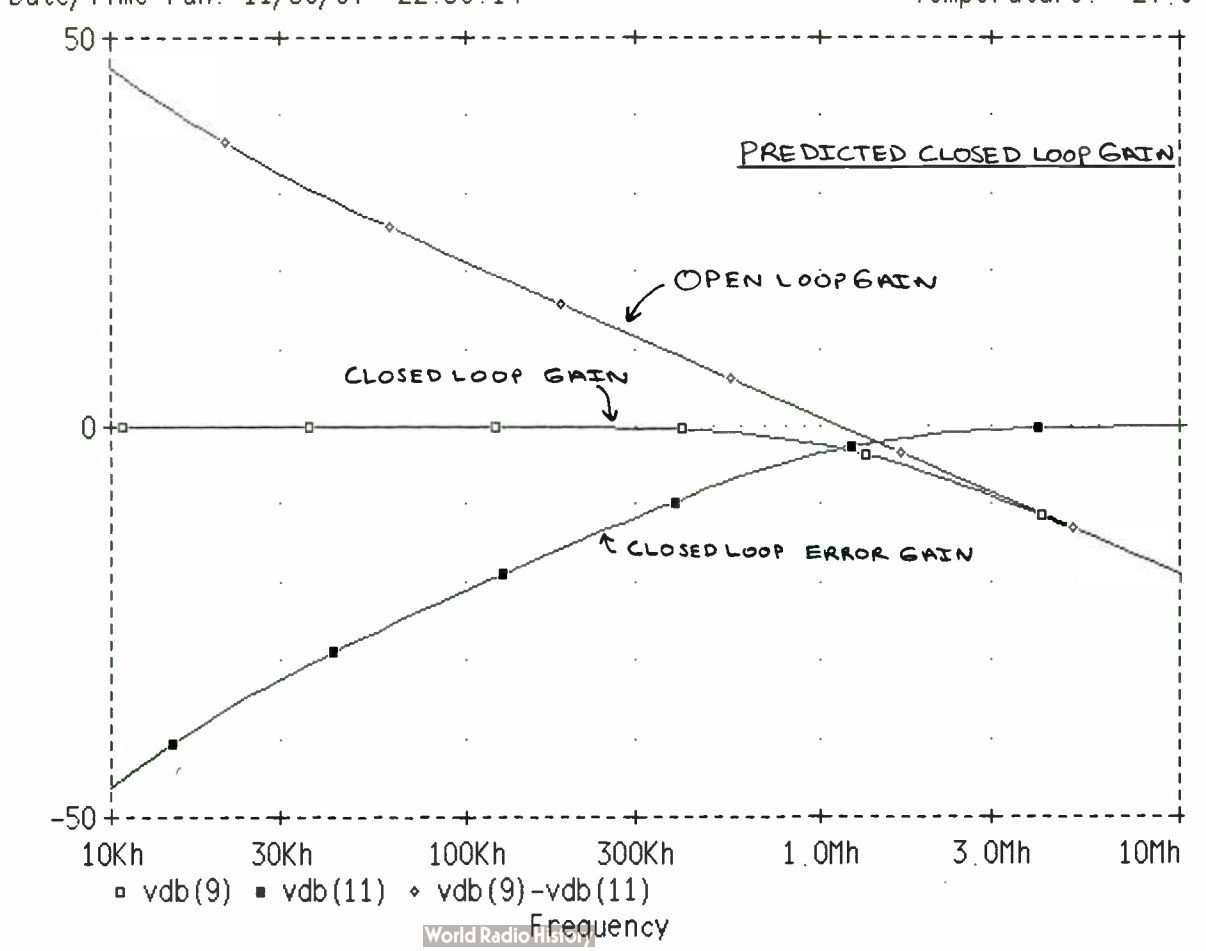


FIGURE 12 MAXIMUM OPERATING FREQUENCY LOOP, 180044







```
10 REM OPPNOISE.BAS
20 REM THIS PROGRAM COMPUTES THE INPUT AND OUTPUT NOISE VOLTAGE
30 REM IN RMS VOLTS/(ROOT HERTZ) OF ONE OPAMP OR TWO CASCADED,
40 REM GIVEN THE VOLTAGE NOISE DENSITY/ROOT HZ, CURRENT DENSITY
50 REM /ROOT HZ, RESISTORS AROUND THE OPAMPS AND ACLI2 (THE
60 REM ACTUAL INVERTING GAIN OF THE SECOND STAGE). ONCE THIS NOISE HAS BEEN DETE
RMINED IT MAY BE USED TO DETERMINE THE PHASE NOISE SPECTRUM OF A VCO.
70 REM PROGRAM UPDATED 12/19/87
80 PRINT"              PROGRAM WRITTEN BY DAN T. GAVIN 1/8/87
90 PRINT
100 PRINT
110 PRINT
120 PRINT
130 PRINT
140 PRINT "                 |                GND        |
150 PRINT " GND--R2-----|  +    *          |---R22--|  +     *
160 PRINT "             |     A1     *-|            |  A2          *---------*
170 PRINT " GND--R1--|--|  -    *       |----R12-|--|  -     *      |
180 PRINT "          |  |               |          |  |              |
190 PRINT "          |------RF-------|          |------RF2------|
200 PRINT "
210 PRINT "
220 INPUT "En= (NANO VOLTS/ROOT HZ)";EN
230 INPUT "In= (PICO AMPS/ROOT HZ)" ;IN
240 INPUT "R1= (KOHMS)";R1
250 INPUT "RF= (KOHMS)";RF
260 INPUT "R2= (KOHMS)";R2
270 INPUT "R12= (KOHMS)";R12
280 INPUT "RF2= (KOHMS)";RF2
290 INPUT "R22= (KOHMS)";R22
300 INPUT "ACTUAL INVERTING CLOSED LOOP GAIN OF A2 ACLI2 V/V " ;ACLI2
310 EN=EN*1E-09
320 IN=IN*1E-12
330 R1=R1*1000
340 R2=R2*1000
350 RF=RF*1000
360 R12=R12*1000
370 R22=R22*1000
380 RF2=RF2*1000
390 REM CALCULATE INPUT NOISE OF A1
400 RS11=R1*RF/(R1+RF)
410 ETI1=SQR(EN^2+(IN*RS11)^2+(IN*R2)^2+1.638E-20*RS11+1.638E-20*R2)
420 ACLN1=1+RF/R1
430 ETO1=SQR(ETI1^2*ACLN1^2)
440 REM COMPUTE INPUT NOISE OF A2 BY ITSELF
450 RS12=RF2*R12/(RF2+R12)
460 ETI2=SQR(EN^2+IN^2*RS12^2+IN^2*R22^2+1.638E-20*RS12+1.638E-20*R22)
470 REM COMPUTE TOTAL OUTPUT NOISE DUE TO A2 BY ITSELF
480 ACLN2=1+RF2/R12
490 ETO2=SQR(ETI2^2*ACLN2^2)
500 REM COMPUTE TOTAL CASCADED NOISE OF A1 AND A2
510 ECTON=SQR(ETO2^2+ETO1^2*ACLI2^2)
520 PRINT
530 LPRINT"En="EN/1E-09"NANO VOLTS/ROOT HZ","In="IN/1E-12"PICO AMPS/ROOT HZ
540 LPRINT
550 LPRINT"R1="R1/1000"K","R2="R2/1000"K","RF="RF/1000"K","R12="R12/1000"K","R22
="R22/1000"K","RF2="RF2/1000"K","RS11="RS11/1000"K","RS12="RS12/1000"K","ACLI2="
ACLI2
560 LPRINT
570 LPRINT"                        NOISE CALCULATIONS"
580 LPRINT
590 LPRINT
600 LPRINT"NOISE REFFERED TO THE INPUT"
610 LPRINT"         OF A1 ETI1="ETI1/1E-09"nV/root Hz"
620 LPRINT"         OF A2 ETI2="ETI2/1E-09"nV/root Hz    (R12 GROUNDED)"
630 LPRINT
640 LPRINT"TOTAL RMS OUTPUT NOISE "
650 LPRINT"         OF A1 ETO1="ETO1/1E-09"nV/root Hz"
660 LPRINT"         OF A2 ETO2="ETO2/1E-09"nV/root Hz    (R12 GROUNDED)"
670 LPRINT
680 LPRINT
690 LPRINT"THE TOTAL CASCADED NOISE AT OUTPUT OF A2 DUE TO BOTH A1 AND A2"
700 LPRINT
710 LPRINT"         ETCON="ECTON/1E-09"nV/root Hz"
720 GOTO 20
730 STOP
```

```
10 REM PLNOISE.BAS
20 REM THIS PROGRAM TAKES FIVE SOURCES OF NOISE IN A PLL AND CALCULATES THE INDI
VIDUAL CONTRIBUTION FROM EACH AND THE SUM OF ALL OF THEM. TwO FORMS OF GAIN ARE
INPUT USED TO CALCULATE THE NOISE SPECTRUMS.
30 REM THE FIRST TAKES THE MAGNITUDE AND ANGLE OF THE OPEN LOOP RESPONSE AND CAL
CULATES THE CLOSE LOOP AND CLOSE LOOP ERROR RESPONSE FROM THE EQUATION HE(S)=1/(
1+G(S)*H(S)) AND H(S)=G(S)H(S)*HE(S).
40 REM THE SECOND TECHNIQUE IS TO USE CLOSE LOOP AND CLOSE LOOP ERROR RESPONSE D
ATA IN DB WHICH CAN BE OBTAINED FROM ANY AC CIRCUIT ANALYSIS PROGRAM.
50 REM IN THIS VERSION OF THE PROGRAM THE MAGNITUDE OF THE OPEN LOOP GAIN AND PH
ASE ANGLE HAVE NOT BEEN DEBUGED 12/19/87
60 REM PROGRAM WRITTEN BY DAN GAVIN 11/7/87
70 REM
80 REM
90 REM         DEFINE INPUT VARIBLES
100 REM NUMBER OF INPUT FREQUENCIES SPECIFIES HOW MANY FREQUENCY POINTS YOU HAVE

110 REM FM IS THE OFFSET FREQUENCY (KHz) WHERE YOU SPECIFY CLOSED LOOP GAIN AND
CLOSED LOOP ERROR GAIN AND NOISE DATA
120 REM CLDB IS THE MAGNITUDE OF THE CLOSED LOOP GAIN IN DB
130 REM CLEDB IS THE MAGNITUDE OF THE CLOSED LOOP ERROR GAIN IN DB
140 REM
150 REM             INPUT NOISE DEFINITIONS
160 REM
170 REM LRI IS THE REFERENCE ABSOLUTE SSB NOISE SPECTRUM IN DBC/HZ
180 REM EN1 IS THE NOISE CONTRIBUTION FROM THE OPAMPS REFERRED TO THE INPUT OF T
HE OPAMPS IN NANOVOLTS/HZ
190 REM EN2 IS THE NOISE CONTRIBUTION DUE TO THE OUTPUT NOISE OF THE OPAMPS OR N
OISE PRESENT AT THE VARACTOR OF THE VCO NANOVOLTS/HZ
200 REM LVI IS THE ABSOLUTE SSB NOISE CONTRIBUTION DUE TO THE VCO IN DBC/HZ
210 REM LDI IS THE ABSOLUTE SSB NOISE CONTRIBUTION OF A DOWN CONVERTING LO OR TH
E NOISE OF OF A DIGITAL DIVIDER REFERRED TO THE INPUT OF THE DIVIDER IN DBC/HZ.
220 REM
230 REM              OUTPUT NOISE DEFINITIONS IN DBC/HZ
240 REM
250 REM LRO IS SSB NOISE CONTRIBUTION DUE TO THE REFERENCE NOISE, LRI
260 REM LIO IS SSB NOISE CONTRIBUTION DUE TO THE INPUT NOISE VOLTAGE, EN1
270 REM LOO IS SSB NOISE CONTRIBUTION DUE TO OUTPUT NOISE OF THE OPAMPS, EN2 OR
A NOSIE VOLTAGE PRESENT AT THE VARACTOR
280 REM LVO IS THE SSB NOISE CONTRIBUTION DUE THE VCO INPUT NOISE, LVI
290 REM LDO IS THE SSB NOISE CONTRIBUTION DUE THE A DOWNCONVERTING LO OR DIGITAL
 DIVIDER NOISE REFERRED TO IT'S INPUT.
300 REM LOT IS THE TOTAL OUTPUT SSB NOISE WHICH IS THE RMS SUM OF THE FIVE ABOVE

310 REM
320 REM
330 Q=0
340 INPUT "PHASE DETECTOR GAIN (VOLTS/RADIAN)";KD
350 INPUT "VCO GAIN (MHZ/VOLT)";KO
360 INPUT "DIVIDER RATIO N (NO UNITS)";N
370 INPUT "DO YOU WANT NEW NOISE AND CLDB AND CLEDB DATA YES=1,NO=0";REDO
380 IF REDO=0 GOTO 510
390 PRINT
400 PRINT
410 INPUT "DO YOU WANT TO CALCULATE THE NOISE SPECTRUM FROM MAG AND ANGLE OF THE
 OPEN LOOP GAIN (1) OR FROM MAG OF CLOSE LOOP GAIN AND CLOSE LOOP ERROR GAIN (0)
 (IN THIS VERSION OF THE PROGRAM 0 IS THE ONLY CHOICE HERE)";D
420 IF D=0 THEN GOTO 460
430 INPUT "FREQUENCY OFFSET FROM THE CENTER IN KHZ, FM";FM
440 INPUT"OPEN LOOP GAIN GH IN DB,AND PHASE ANGLE AT OFFSET FM";GHDB,PHI
450 GOTO 510
460 INPUT "N of freqs";X
470 DIM FM(X),CLDB(X),CLEDB(X),LRI(X),EN(X),ENO(X),LVI(X),LDI(X),LRO(X),LIO(X),L
OO(X),LVO(X),LDO(X)
480 IF REDO=0 GOTO 510
490 INPUT "FM,CLDB,CLEDB";FM(Q),CLDB(Q),CLEDB(Q)
500 INPUT "LRI,EN1,EN2,LVI,LDI";LRI(Q),EN(Q),ENO(Q),LVI(Q),LDI(Q)
```

```
510 FM(Q)=FM(Q)*1000!
520 EN1=EN(Q)*1E-09
530 EN2=ENO(Q)*1E-09
540 REM CALCULATE G2 NORMALIZED TO 1 HZ
550 L=LOG(10)
560 G21=(KO*1000000!)/FM(Q)
570 G2=20/L*LOG(G21)
580 IF D=1 GOTO 700
590 REM CALCUALTE CONTRIBUTION OF NOISE DUE TO REFERENCE NOISE
600 LRO(Q)=CLDB(Q)-LRI(Q)+20*LOG(N)/L
610 REM CALCULATE NOISE CONTRIBUTION DUE TO NOISE INJECTED BETWEEN THE PHASE DET
ECTOR AND LOOP FILTER. minus 3db is used here due to the fact the noise is rms
620 LIO(Q)=CLDB(Q)-20*LOG(KD)/L+20*LOG(EN1)/L+20*LOG(N)/L-3
630 REM CALCULATE NOISE CONTRIBUTION DUE NOISE INJECTED AT VARACTOR
640 LOO(Q)=CLEDB(Q)+G2+20*LOG(EN2)/L-3
650 REM CALCULATE NOISE CONTRIBUTION DUE VCO NOISE SPECTRUM
660 LVO(Q)=CLEDB(Q)-LVI(Q)
670 REM CALCULATE NOISE CONTRIBUTION DUE DIVIDER NOISE REFFERED TO ITS INPUT OR
ABSOLUTE NOISE OF UP OR DOWN COVERTING LO
680 LDO(Q)=CLDB(Q)-LDI(Q)
690 GOTO 810
700 GH=10^(GHDB/20)
710 PI=3.14159265#
720 PHI=2*PI/360*PHI
730 B=SIN(PHI)
740 A=COS(PHI)
750 CLE=1/(SQR((1+GH*A)^2+(GH*B)^2))
760 CLEDB=20*LOG(CLE)/L
770 CL=GH*CLE
780 CLDB=20*LOG(CL)/L
790 GOTO 590
800 REM CALCULATE SUM OF NOISES
810 LOT(Q)=-10*LOG(10^(LVO(Q)/10)+10^(LRO(Q)/10)+10^(+LIO(Q)/10)+10^(+LOO(Q)/10)
+10^(LDO(Q)/10))/LOG(10)
820 FM(Q)=FM(Q)/1000!
830 Q=Q+1
840 IF Q<X GOTO 480
850 Q=Q-1
860 PRINT "      KHZ       dB        dB"
870 PRINT "      FM       CLDB     CLEDB"
880 PRINT USING "########.##"; FM(Q),CLDB(Q),CLEDB(Q)
890 Q=Q-1
900 IF Q>-1 THEN GOTO 880
910 Q=X-1
920 PRINT
930 PRINT
940 PRINT
950 PRINT
960 PRINT "                                  INPUT          "
970 PRINT"     KHZ    DBC/HZ      nV/HZ      nV/HZ    DBC/HZ    DBC/HZ"
980 PRINT"     FM       LRI       EN1        EN2       LVI       LDI  "
990 PRINT
1000 PRINT USING "########.";FM(Q),LRI(Q),EN(Q),ENO(Q),LVI(Q),LDI(Q)
1010 Q=Q-1
1020 IF Q>-1 THEN GOTO 1000
1030 Q=X-1
1040 PRINT
1050 PRINT
1060 PRINT
1070 PRINT
1080 PRINT "                                  OUTPUT         "
1090 PRINT"       KHZ                          DBC/HZ                                -D
BC/HZ"
1100 PRINT
1110 PRINT"       FM         LRO         LIO         LOO         LVO         LDO
LOT"
1120 PRINT USING "########.#"; FM(Q),(LRO(Q)),(LIO(Q)),(LOO(Q)),(LVO(Q)),(LDO(Q)
),(LOT(Q))
1130 Q=Q-1
1140 IF Q>-1 THEN GOTO 1120
1150 PRINT
1160 PRINT
1170 PRINT
1180 PRINT
1190 PRINT
1200 PRINT
1210 PRINT
1220 PRINT
1230 IF D=0 GOTO 330
1240 GOTO 430
1250 STOP
```

Transmitter Notch Filter

by
Gregory F. Kinnetz
Electrical Engineer, R & D
Microwave Filter Company, Inc.
6743 Kinne Street
E. Syracuse, NY 13057

ABSTRACT:

A filter has been developed with the function of deleting unwanted transmitter output signals. The notch filter utilizes resonant cavity technology to achieve the required electrical response. This report covers the design basis and component fabrication techniques of high power (up to 10 kW) notch filters for transmitter applications in the VHF spectrum. A practical implementation of a three section bandstop filter is described, outlining technical considerations which must be observed in order to handle the large currents and voltages present in the filter. Phase matching of high power cables to maintain low loss and VSWR in the passband is also described. Considerations for heat transfer, connector types to handle power, and other often overlooked items are covered as well.

INTRODUCTION:

Transmitter notch filters are a commonly used component at sites with both transmit and receive capabilities in the VHF band. One limitation of high power transmitters is that they produce, along with the desired output, harmonics and other spurious radiation. Thus, the need for this type of filter arises when the transmit and receive frequencies are close in the radio frequency spectrum and spurious transmitter output is causing receiver interference. The notch filter presented, which connects in line on the transmitter output, alleviates this type of interference.

THEORY:

The required filter for this application basically must be capable of passing the full transmitter power at the transmit frequency and attenuate any transmitter output at the receive frequency (see Electrical Response, Figure 1). One method of realizing these specifications is the three resonant cavity section filter described here (see figure 2). The length of the center rod shown in Figure 3 is approximately 1/4 wavelength (resonant) at the desired notch frequency (receiver frequency). In practice, the length of the rod will be somewhat less than the free space 1/4 wavelength due to capacitive fringing at the free end of the rod. A 1/4 wavelength is also the approximate length of the interconnecting jumper cables, so that full diplexing occurs. However, both lengths are refined in the "Design" section to follow - the cables to aid in impedance matching and the center rod to set the resonant frequency. Coupling into the cavity is inductive and occurs at the shorted end of the center rod with a loop. Inductive coupling at the shorted end was selected over capacitive coupling at the open end to avoid the high voltages present at that end of the resonator.

DESIGN:

There are basically five topics that need to be addressed in designing high power notch filters, including: inner and outer conductor diameters, center rod length, loop dimensions, jumper cables, and other special considerations.

The first subject is determining the inner and outer conductor diameters. The most important parameter in this area is the quality factor, Q. Ao, the notch attenuation, is related to Q by the following equation:

$$Ao = 20 \text{ Log } (Qu/ QL + .5) \quad \text{(Eqn 1)}$$

where Qu = unloaded Q
QL = loaded Q
= fo/BW

The passband insertion loss, on the other hand, is directly proportional to Q. It is, therefore, in the interest of optimum filter performance to design for maximum Q. In general Q = XL/r
where XL = inductive reactance = 2 Pi fL
r = RF resistance

In all cases one can see that Q can be increased by decreasing r. The resistance, in this particular instance, is predominately a function of the center rod's material, its skin depth at a specific frequency and diameter.

$$d = \text{Skin depth} = (\lambda / \text{Pi suc})^{1/2} \quad \text{(Eqn 2)}$$

(meter)

c = velocity of light = 2.998×10^8 m/s

1/s = 1.724×10^{-8} R/Rc ohm-meter

u = 4 Pi x 10^{-7} ur H/m

ur = Relative permeability of the conductor (ur = 1) for all non-magnetic materials

R = Resistivity of the conductor

Rc = Resistivity of copper

$$Rsq = \text{Resistance per square} = \frac{1}{ds} \quad \text{(Eqn 3)}$$

(Ohm)

$$Rsq = (2.61 \times 10^{-7} (f^{1/2})) (ur\ R/Rc)^{1/2} \text{ Ohm} \quad \text{(Eqn 3A)}$$

$$Rac = \text{RF resistance} = r = (12/\text{Pi } D)\ Rsq \quad \text{(Eqn 4)}$$

(Per Foot)

For cylindrical rod

Therefore, for highest Q and consequently lowest loss the resistivity of the outer surface of the center rod material must be small and the diameter as large as possible.

Q is also dependent on the characteristic impedance (Zo) of the cavity. For optimum Q, it is well recognized [1] that a characteristic impedance of 76 Ohms is necessary, where

$$Zo = Zo\ (\text{Coaxial Line}) = 60/(er)^{1/2} \ln (b/a) \quad \text{(Eqn 5)}$$

er = relative dielectric constant (See Figure 4)

The center rod length is easily determined. Initially the center rod should be cut to one-quarter wavelength at the notch frequency where

$$\lambda\ (\text{in.}) = \frac{11802.8}{f\ (\text{MHz})} \quad \text{(Eqn 6)}$$

"Common RF Sense" tells us that the 1/4 λ inside cavity will be somewhat less than the free space wavelength calculated above, due to fringing. Therefore, by experimentation we can determine the actual length and remove the excess. Measure the initial resonant frequency, fo, of the filter and use Eqn 7 to alter the center rod length.

$$\triangle \text{Length} = 1/4\ \lambda\ (\text{Obtained}) - 1/4\ \lambda\ (\text{desired}) \quad \text{(Eqn 7)}$$

Perhaps the component that requires the most design time is the coupling loop. General guidelines in design include minimizing the loop length and ensuring the wire diameter is great enough to handle the expected RF current. With reference to the latter, an estimate of the expected RF current can be found given the transmitter power level and the system impedance. $I = (P/R)^{1/2}$. The current could double if the output load of the filter was an open or a short (purely reflective load), therefore, double the amount calculated above.
Equation 8 is used next to calculate the minimum wire size.

$$\text{Minimum wire diameter} = \frac{Irf\ (f)^{1/4}}{1000} \quad \text{(Eqn 8)}$$

Where f = frequency in MHz

As a rule of thumb the wire size should be doubled for a margin of safety. Because this is a minimum value, it is recommended that the largest diameter physically possible be used.

The loop length previously mentioned should be kept as small as possible so that the impedance mismatch caused by it will be minimal. The following procedure is used to calculate the loop dimensions:

1.) Based on the electrical requirements (transmit and receive frequencies), decide on the allowable notch bandwidth.

2.) From earlier calculations on the inner and outer rods as well as these on the loop, calculate the ratio shown in Figure 5. Keep in mind that the inter rod spacing is inversely proportional to the loop length. As a rule of thumb make the inter rod spacing equal to the loop diameter.

3.) Using equations 9 through 11, solve for L cos $\emptyset$

$$Q = [(\text{Pi } Zo\ N)/4]\ [(r^2 Rg)/(A^2 uo^2 f^2 \cos^2 \emptyset)] \quad \text{(Eqn 9)}$$

[3]

where Q = loaded Q (QL) of the cavity

uo = 4 Pi x 10^{-7} H/m

N = number of 1/4 wavelengths in the length of the rod = 1

f = frequency in hertz

Rg = internal generator resistance = 50 ohms

r and $\emptyset$ are the coordinates of the effective center of the loop (see figure 3)

A = loop area

Zo = characteristic impedance

Simplifying Eqn 9 yields:

$$Q = [(770.91\ ZoRg)/f^2]\ [r/A]^2\ [1/\cos^2 \emptyset] \qquad \text{(Eqn 9A)}$$

Eqn 9A assumes a constant magnetic field over the loop area and that the loop length is small, so it can be considered a point from the shorted end of the rod. Those approximations can be refined using the following equation:

$$\frac{A}{r} = \int_{ra}^{rb} \frac{Ldr}{r} = L \ln (rb/ra) \qquad \text{(Eqn 10)}$$

(See Figures 3 & 5)

$$Q \text{ (loaded)} = fo/BW \qquad \text{(Eqn 11)}$$

4.) Now an iterative process begins where L Cos $\emptyset$ avg is solved for varying loop lengths until L Cos $\emptyset$ avg approximates L Cos $\emptyset$.

$$L \cos \emptyset \text{ avg} = L\ [(\cos \emptyset\ y + \cos (\emptyset\ y + L))/2] \qquad \text{(Eqn 12)}$$

$$\cos \emptyset\ y = \cos\ [(Y / \lambda)(360)] \qquad \text{(Eqn 13)}$$

$$\cos (\emptyset\ Y + L) = \cos\ [((Y + L)/\lambda)(360)] \qquad \text{(Eqn 14)}$$

At this point all the loop dimensions are known.

5.) Varify and refine with testing.

In theory, a 1/4 wavelength transmission line between notch filters is required to ensure full diplexing (addition of the individual notch attenuations). At this length the filters electrically do not "see" each other, are not loaded down by one another and notch loss summation occurs. This arrangement is fine in a 50 ohm system, but when coupling into a cavity with an unknown impedance transmission line (loop) a mismatch occurs and further action is required. Once an individual notch is operational, one can measure its impedance at the transmit frequency using a network analyzer with Smith chart capabilities. Next all that is necessary is to determine the required transmission line length to match the impedance of the cavity to that of the system. It is essential to keep in mind that some attenuation is sacrificed in matching the filter since we are deviating from 1/4 λ . Once a filter is at this stage one can determine the number of sections required to meet the notch attenuation specifications plus a margin of safety.

So far, the topics discussed have dealt with trap design in general. There are several special considerations when designing high power notch filters. These have to do with loss and heat.

It should be reiterated that most of the loss associated with this particular filter is located in the center rod. Therefore, the largest portion of the dissipated power occurs there. Obviously, to minimize heating, the RF resistance of the center rod must be minimized. Two means of accomplishing this are to select low resistivity center rod material and to ensure a sound mechanical connection at the shorted end of the rod. From a design standpoint, the RF current maximum of the inner rod is at its shorted end. Since the dissipated power is equal to the current squared times the resistance, the heat generated by the filter predominately is at this end. To facilitate heat transfer, the end plate and the outer tube should be constructed out of an excellent thermal conductive metal (most likely aluminum) and the surface contact area between the two should be generous. Temperatures of the center rod will become elevated regardless, because there will always be some loss, however small, associated with the filter. Whether or not this will present a problem (filter de-tuning or failure) can best be tested by actually powering up the filter in incremental steps and monitoring its electrical performance. Other test possibilities include heating up the inner conductor with an DC or AC source and a heating element and monitoring the response. Also the use of Invar (A steel alloy with an extremely small linear thermal expansion coefficient) as a center rod will, for all practical purposes, eliminate electrical response variations due to temperature.

Both the jumper cables and the connectors must be selected to handle the required power at the frequency of interest. Figure 6 [4] shows most common connectors and their operating voltages. For higher power applications 1 5/8", 3 1/4" or 6 1/8" EIA connectors are recommended.

Cables should be selected on the basis of low attenuation, power rating, physical construction, and flexibility characteristics. Most manufacturers give plots of power rating and attenuation versus frequency so candidate cables can be selected. A minimum bending radius and other mechanical considerations are also described and usually will narrow the field.

CONCLUSION:

Transmitter notch filters with the function of attenuating specific unwanted transmitter output frequencies have been developed and tested. Their design differs from conventional transmission line filters in that special emphasis is placed on high Q, low loss, and power handling design. The procedure presented is applicable for transmitter frequencies in the VHF frequency band.

REFERENCES:

[1] Microwave Filters, Impedance - Matching Networks, and coupling structures, G. Matthaei, L. Young, E.M.T. Jones, Artech House Books, Dedlam, MA

[2] Reference Data For Radio Frequency Engineers, Howard W. Sams & Co., Inc., New York

[3] Very High Frequency Techniques, S. Cohn, R.A. Soderman, S.J. Griffin, R.O. Petrich, New York, McGraw-Hill

[4] A Short Reference Guide To R.F. Coaxial Connectors, Automatic Connector, Inc., Commack, New York

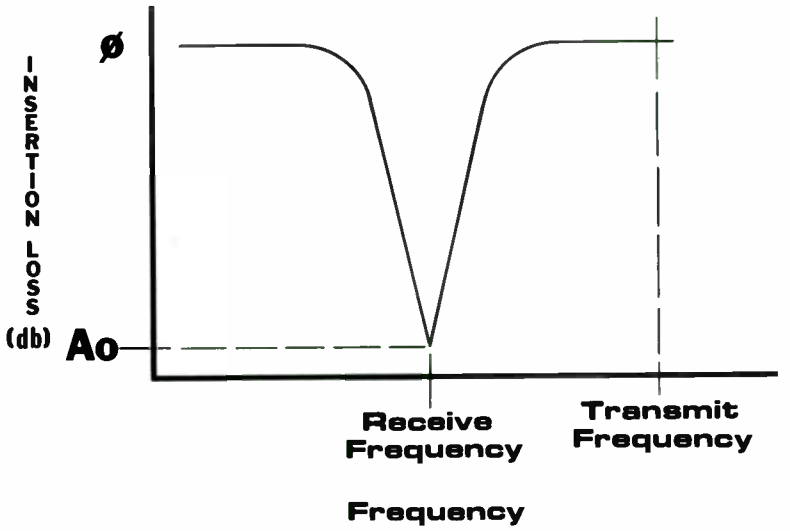
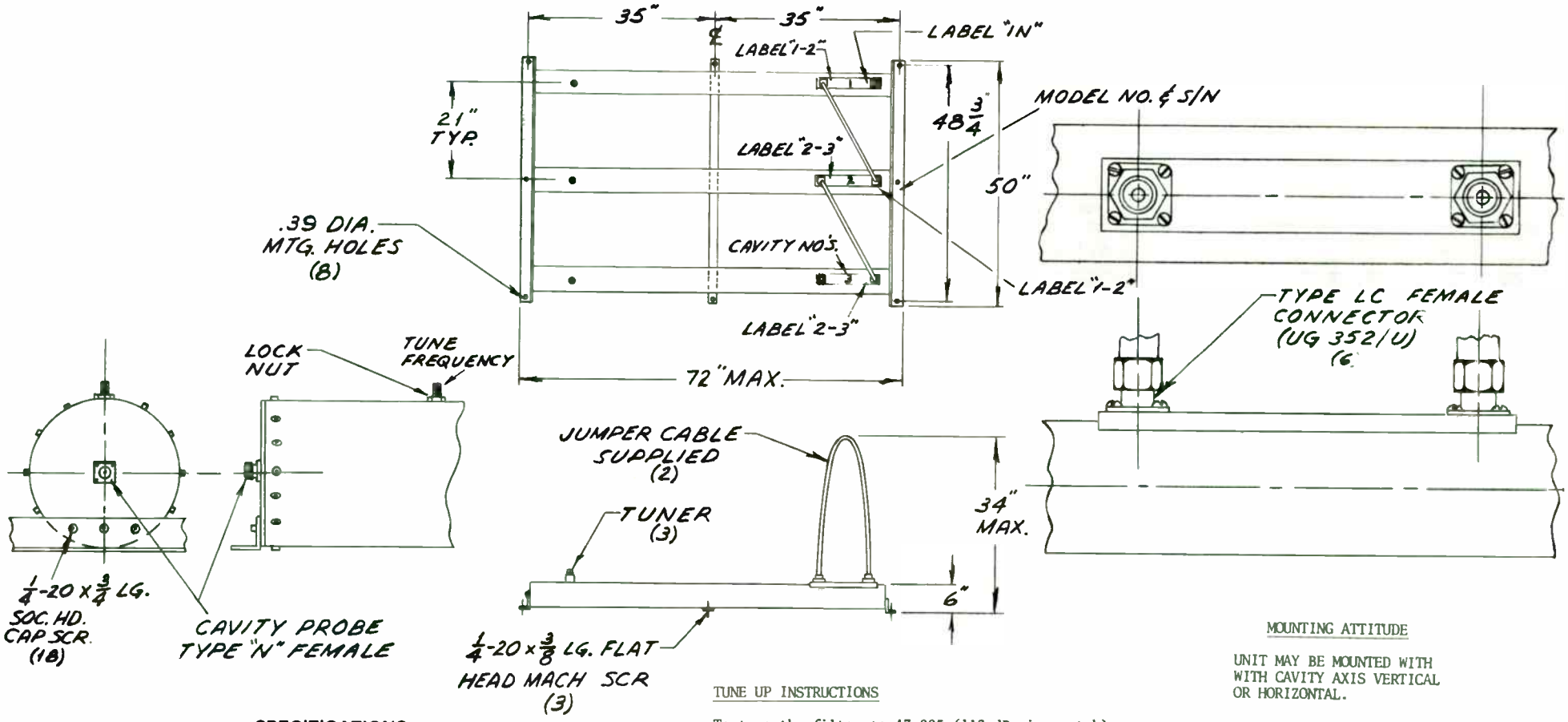


Figure 1: Typical Electrical Response



SPECIFICATIONS

Notch Frequency: 47.005 MHz
Notch Attenuation: 112 dB minimum
Insertion Loss At 49.595 MHz: .5 dB max. (.20 dB max. typ.)
Power Handling: 10 KW CW (At 49.595)
Connectors: Type LC Female

TUNE UP INSTRUCTIONS

To tune the filter to 47.005 (112 dB min. notch) under running conditions, proceed as follows:

(1) Loosen tuning screw nuts on all three cavities and turn tuning screws fully counterclockwise (until internal stop).
(2) Connect analyzer to cavity probe #3 and tune its frequency to 47.005 MHz & lock the tuning screw.
(3) Repeat step (2) for cavity #2.
(4) Repeat step (2) for cavity #1.

MOUNTING ATTITUDE

UNIT MAY BE MOUNTED WITH WITH CAVITY AXIS VERTICAL OR HORIZONTAL.

Figure 2: Filter Outline

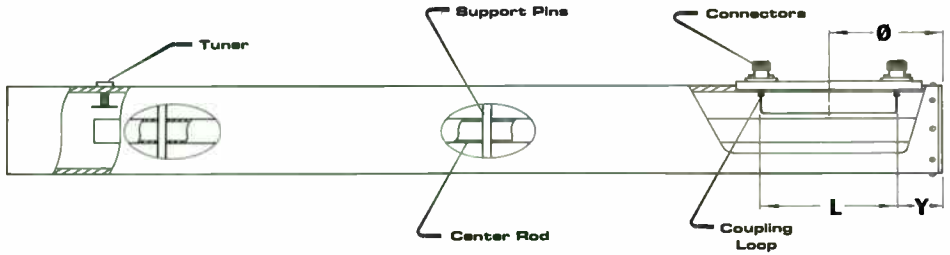


Figure 3: Cut Away Of Single Cavity & Loop Dimensions

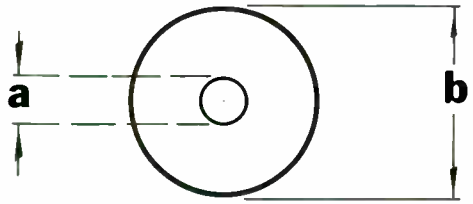


Figure 4: Coaxial Transmission Line

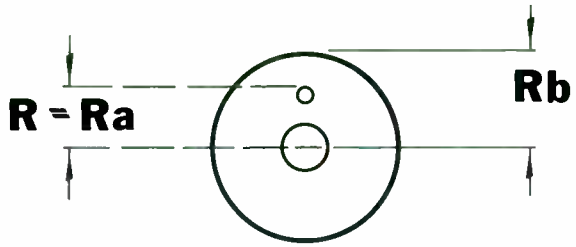


Figure 5: End View Of The Cavity

CONNECTOR TYPE (SERIES)	COUPLING METHOD	SIZE	CABLE O.D.	MAX. FREQ. (GHz)	VSWR	WORKING VOLTAGE	REL. COST
LC	Thread	Large	1/2-7/8	3	1.3	5000V RMS	High
LT					1.3		
UHLC					50 Ohms		
Comm N	Thread	Medium	3/8-7/8	11	1.3	1000	Low
N	Thread					1000	Medium
C	Bayonet					2000	High
SC	Thread					2000	
HN	Thread			4	1.4	3500	
BNC	Bayonet	Minature	1/8-3/8	4	1.3	500	Low
TNC	Thread			11	1.3	500	Medium
SHV	Bayonet			—	—	3500	High
MHV*	Bayonet				—	3500	High
SMA	Thread	Submin.	5/64-1/4	18	1.3	335	Medium
SMB	Snap-on			4	1.4		Low
SMC	Thread			10	1.6		Low
PMA	Push-on			12.4	1.3		High
BSM	Bayonet			.5	1.4		Low
TSM	Thread			.5	1.4		Medium
75 ohm SMB	Snap-on			—	—		Medium
75 ohm SMC	Thread						Medium
TPS	Bayonet			10	1.35	500	High
SSMA	Thread	Ultra-min	5/64-9/64	18	1.3	250	High
PMMA	Push-on				1.2		
ADM	Push-on			1	1.3		Medium

* Inactive for new design per Mil-C-39012.

Figure 6: Connector Data

HIGH FREQUENCY TEST FIXTURING FOR MICROWAVE HYBRID CIRCUITS AND LEADLESS DEVICES

by

John K. Logan
Inter-Logic Systems Company, Inc.
7201-E Garden Grove Blvd.
Garden Grove, CA 92641

It is the author's observation that most technical papers and articles on high frequency testing focus on electrical test results, rather than on the test fixturing used. This is understandable because most of the time it is possible to package the device so that it is necessary only to make attachment to the package leads, or to I/O connectors from a hybrid circuit or printed circuit board.

However, two significant trends in high frequency technology make it necessary to pay increasing attention to contacting semiconductor chips and leadless packaged devices for test purposes. These trends are:

Commercial production of high speed silicon and gallium arsenide semiconductors, which require high frequency probing at the wafer level. In a laboratory environment, it is possible to do wafer test at relatively low frequency and then accept considerable yield loss of both chip and package at final test. This high degree of yield loss is not acceptable in price-competitive commercial production.

Second, the increasing popularity of surface mount technology requires the testing of devices packaged in leadless enclosures. Additionally, these same leadless packaged devices must be tested after mounting on the SMT board in cases where malfunction or final test failures occur.

Thus, it is necessary to develop wafer probe cards that function in the gigahertz range. It is necessary to develop hybrid circuit and surface mount test fixtures that will pass acceptable signals in the high megahertz to low gigahertz ranges.

It is the purpose of this paper to present a number of fixturing techniques that are used to interface with semiconductor chips, hybrid circuits, leadless devices, and surface mount assemblies.

This paper will focus on practical examples of test fixtures that have been built and used for high frequency testing. Where available, test results and frequency response curves for the test fixtures will be presented. Attention will be centered on test fixturing for frequencies in the range below 5GHz, since this is the region of the frequency spectrum where the largest volume of production and test activity is centered at present.

When a device has no leads, it is necessary that the test fixture supply the missing leads. This is done by means of probes or contactors. Two types of probes are commonly used for contacting small geometry devices: the cantilever, or needle-type probe, and the axially spring-loaded probe, commonly called a "Pogo" probe, or "bed-of-nails" probe. The needle type probe is used for semiconductor wafer probing, and contacting small-geometry thin-film hybrid circuits. The spring-loaded probe is used for larger area devices, such as printed circuit boards and SMT boards.

Figure 1 shows a typical cantilever needle probe. This is a metal blade probe and has a bandwidth extending from DC to about 30 MHz. However, the same mechanical parameters are applicable to high frequency microstrip probes.

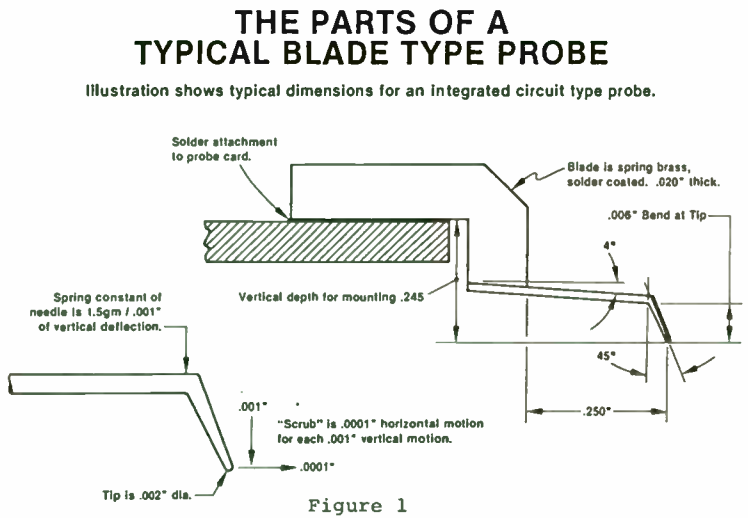


Figure 1

Figure 2 shows a typical spring-loaded "Pogo" probe. These probes

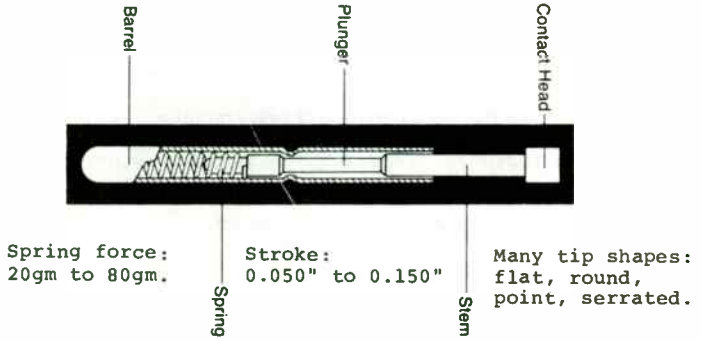


Fig. 2: SPRING-LOADED PROBE

were developed for printed circuit board testing, generally on 0.1" centers. As device geometries became smaller, these probes were reduced in size and can now accomodate center-center spacings as small as 0.020"

Figure 3 shows a typical high frequency wafer probe card. (Picture at end of paper.) This probe card has been used at 2.5 GHz test frequency. This probe card uses ceramic microstrip probe blades with beryllium-copper needles.

DESIGN PARAMETERS OF CERAMIC MICROSTRIP PROBE BLADES:

The microstrip probe blades provide a 50 ohm path to within 0.050" of the device pad. A 50 ohm chip resistor may be placed at the base of the needle in order to terminate the transmission line at that point. Figure 4 shows a typical ceramic microstrip probe:

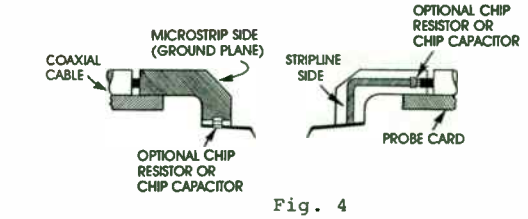


Fig. 4

The probe blades are made of 96% alumina ceramic. The metallization is done by conventional thick-film screen and fire techniques. Solder attachment is used to mount the probe blade to a solder-

plated copperclad board. Coaxial cable carries the signal from the blade to an SMA connector located elsewhere on the card.

Probe needle length is critical at high frequencies, because even a few nanohenries of inductance created by the unshielded needle can result in a high inductive impedance at high frequencies. A normal DC probe needle is 0.250" long. A typical high frequency needle is from 0.050" to 0.150" long. Needles as short as 0.015" have been used (although with great care!)

Probe needle material becomes critical because of the high currents and high switching speeds encountered. Tungsten is widely used for DC probes. However, it is less desireable for high frequencies because of its higher contact resistance. Also, its stiffness - which is a virtue in a DC probe - is a problem in the very short probes. The probes become excessively rigid, and tend to scratch the delicate pad surfaces.

Chip components may be attached to the ceramic blade at various points, as shown in Figure 4. Signal channels are terminated with chip resistors of proper value. Power and V_{CC} channels may be bypassed at several points with chip or leaded capacitors.

TEST FIXTURES FOR THREE PRODUCT AREAS:

We will consider test fixtures for: small geometry semiconductor chips; ceramic hybrid circuits, and surface mount components and assemblies.

SEMICONDUCTOR PROBE CARDS:

Figure 5 shows a high density double-end probe card with 140 probes. Figure 6 shows a close-up view of the probe area. Signal channels were terminated with 50 ohm chip resistors at the inner ring. V_{CC} channels were bypassed with 100pf chip capacitors. Small semi-rigid coaxial cable was trimmed to length for each signal channel, to provide equal overall runs of coaxial cable . Flexible coaxial cables were 5 feet in length. The probe card contains two internal planes, providing a 50 ohm path for signals on the card traces.

Figure 7 shows a high frequency probe card with a "ring mount" to support the SMA connectors. A cover may be placed over the ring to provide RFI/EMI shielding of the device. A sliding door opens to provide for probe alignment to the device pads.

HYBRID CIRCUIT PROBE CARDS:

Figure 8 shows a large-area probe card utilizing ceramic microstrip probes to contact a 1¼" x 2½" ceramic hybrid circuit. Metal blade probes are used for ground contacts. This probe card can accomodate a substrate up to 3" x 3" in size.

Figure 9 shows an adaptation of the spring-loaded "Pogò" probes to a probe card. A clear plastic plate covers the hole in the probe card and supports the probes. An advantage of this arrangement is that direct viewing of the device pads is possible. This makes alignment to small pads much easier.

SURFACE MOUNT TEST FIXTURE:

Figure 10 shows a test fixture designed to permit the contacting of leadless devices that have been reflowed on to an SMT board. A large area may be covered: up to 16" x 36". A turret supports eight (8) test heads. Each head utilizes the spring-loaded probes to contact pads on the SMT board, just outboard of of the device. Heads with 14, 16, 24, 48, 64, and 84 pin package capability are provided. Two heads are mounted to handle packages mounted at both 0^{o} and 90^{o} to the length of the board. The test fixture utilizes buffer packages close to the probe heads, and can pass 200 MHz signals.

SINGLE POINT PROBE:

Figure 11 shows a unique adaptation of the microstrip probe for diagnostic and trouble-shooting work on small geometry devices. The probe is mounted on a precision X-Y-Z micro-positioner. The probe tip may be located to an accuracy of 0.0001" in all three axes. This is valuable where critical signals must be injected or removed from an integrated circuit or hybrid circuit.

It is the purpose of this paper to give the test engineer a picture of various fixturing techniques that are available to him for contacting small geometry leadless devices and surface mount boards. The author trusts that these examples will be useful and will spark additional ideas on the part of the reader.

****

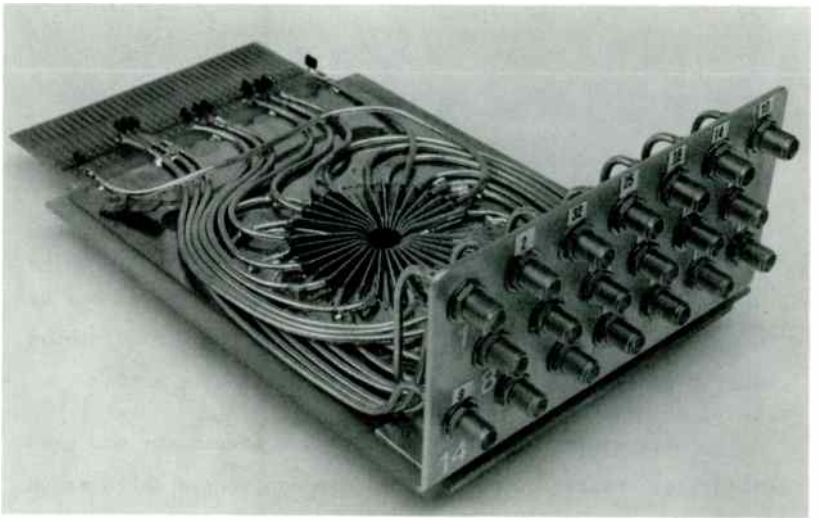


Figure 3. Wafer Type Probe Card.

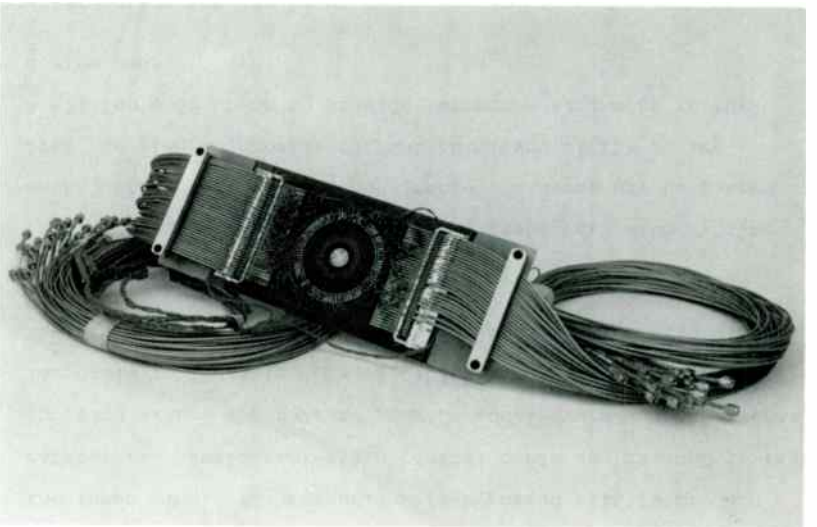


Figure 5. High-Density Epoxy Ring Card.

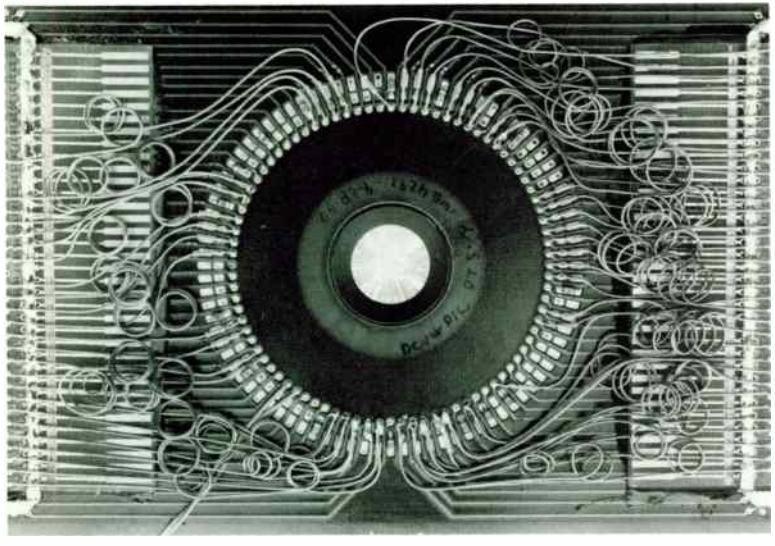


Figure 6. Epoxy Ring Probe Card - Closeup View.

Figure 7. Ring Mount High Frequency Probe Card.

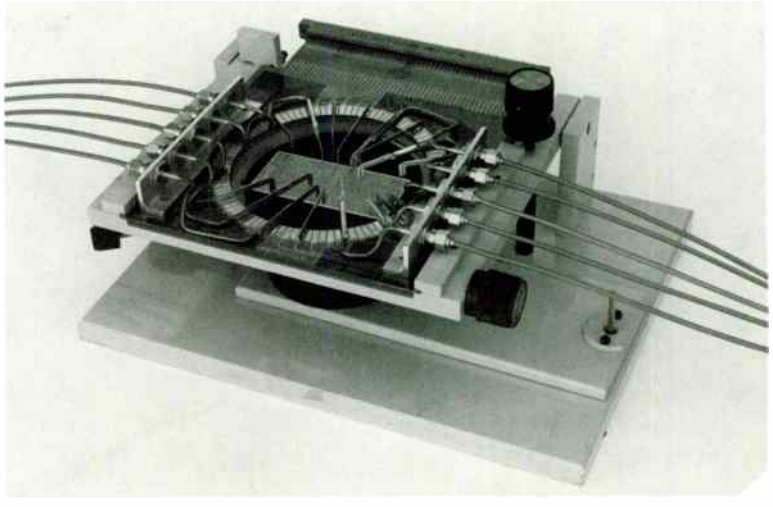


Figure 8. Hybrid Circuit Probe Card.

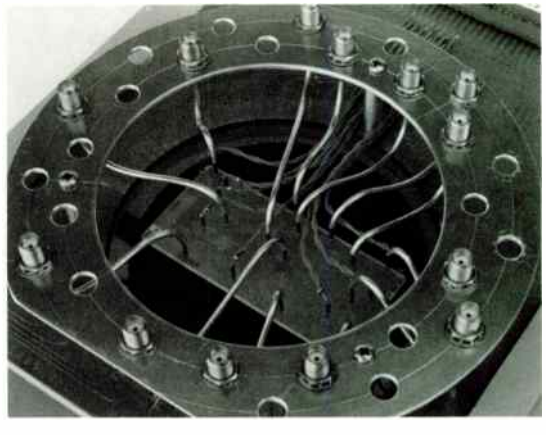


Figure 9. Hybrid Circuit Probe Card With Spring-Loaded Probes.

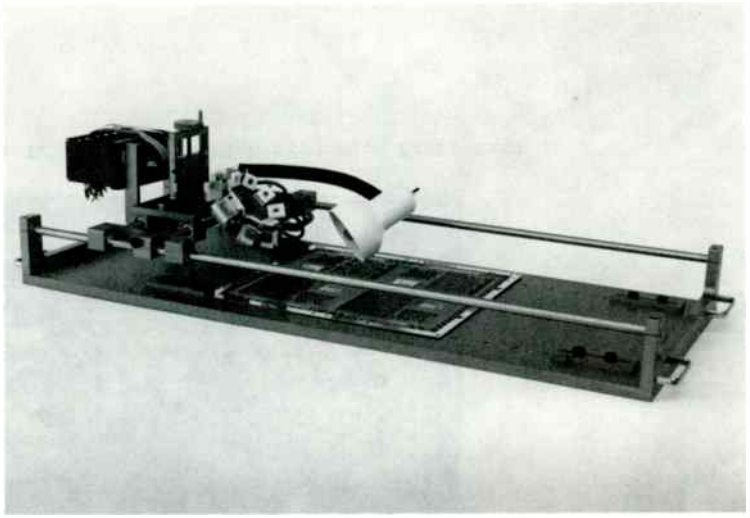


Figure 10. Surface Mount Test Fixture.

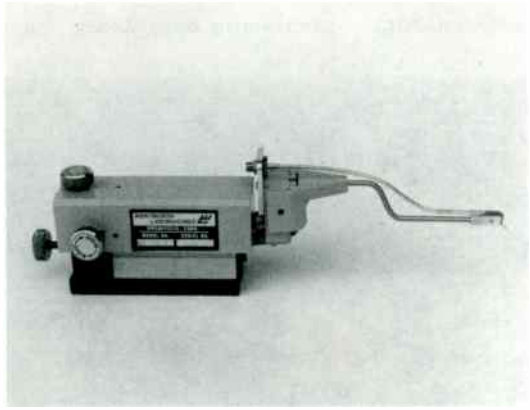


Figure 11. Single-Point High Frequency Test Probe.

A High Performance Synchronous Detector and Phase Noise Measurement System Using a C-QUAM (R) AM Stereo Decoder IC

by

Jon GrosJean
Woodstock Engineering
Box 110, Pulpit Rock Road
S. Woodstock, CT 06267

Abstract:

In this age of high-tech microwave integrated circuits, computer aided design and simulation, more mundane subjects like AM radio receivers seem like uninteresting topics. In fact, the whole consumer electronics business is considered passé by many engineers even though they listen as much to their radios as those of 30 years ago did. Nevertheless, most of the new innovations in the consumer industry have come from inventors in the USA while most of the receivers are produced in other countries. AM stereo is one of these inventions, and although there are 5 competing systems which can be used by broadcasters, the Motorola C-QUAM (R) system is dominant because Motorola spent a lot of time and money developing transmitter exciters and receiver integrated circuits which could produce a high quality demodulated stereo signal.

Introduction to AM Stereo:

The Sprague ULN3820A / Motorola MC13020, Motorola MC13022 and MC13024 integrated circuits are designed to decode the C-QUAM

C-QUAM (R) is a trademark of Motorola, Inc.

AM stereo signal. Since the L-R component of the stereo signal is phase modulated, these ICs contain a very quiet phase locked loop which can use an LC, ceramic or crystal resonator in the oscillator. They all include envelope and synchronous in-phase (I) and quadrature (Q) detectors as well as other circuitry for detecting the 25 Hz phase modulated pilot signal. Because the IC's are designed to provide an output signal-to-noise ratio of at least 50 dB in the stereo mode, the loop is very quiet, and the Q detector output can be used to measure the noise of the receiver local oscillator or any other received signal as well as for evaluation of signal generators and other oscillators for phase noise. Equation 1 describes the C-QUAM AM Stereo signal:

$$Ec=Ac[1+(L+R)\cos 2\pi Fmt]\cos[2\pi Fct+\theta] \quad (1)$$

$$\theta=\arctan\{[(L-R)\cos 2\pi Fmt+.05\sin 50\pi t]/[1+(L+R)\cos 2\pi Fmt]\}$$

L,R = left and right audio signals Fc = carrier frequency

Fm = modulating frequency

The L+R audio signal modulates the carrier amplitude, and the L-R audio signal phase modulates the carrier in quadrature, but it is modified so that an envelope detector produces an undistorted signal to make the system compatible with existing receivers. Thus, the L+R signal can be recovered in a receiver with an envelope detector, and a phase-locked loop can be used to recover the L-R signal. However, the L-R signal will be distorted unless it is corrected for the phase dependent attenuation introduced in the transmitter. This is

done in the receiver with the so called cos Φ corrector in the decoder IC. This corrector compares the undistorted envelope detected signal with a synchronously detected L+R signal and modulates the incoming signal before it is applied to the L+R and L-R synchronous detectors. Figure 1 shows the phasor relationships for a quadrature modulated and a C-QUAM signal. Note that in the C-QUAM signal, the amplitude term follows an arc of constant amplitude so that the quadrature terms will not produce a signal from an envelope detector.

The Decoder ICs:

Figure 2 shows the block diagram of the MC13020/ULN3820 decoder together with the components normally used for stereo operation. (Sprague, Motorola and Toshiba manufacture the same IC under a cross-licensing agreement.) Figure 3 shows the block diagram of the newer Motorola MC13022 and its external components.

The 25 Hz .05 radian phase modulated signal is transmitted to indicate the presence of a stereo signal. The MC13020 detects this pilot by passing the output of the Q detector through an active filter and then to a threshold detector. The MC13022 detects the pilot using a digital detection scheme and does not need an active filter.

Included in all the ICs is some method for detecting excessive signals in the Q output and either blending the L and R channels or forcing the decoder into the monaural mode to eliminate the interference.

The phase-locked-loop used in all the decoders has an internal oscillator and phase-shift network and can use either an LC, crystal or ceramic tuned element. The phase detector output is a current source. ECL internal dividers are used to insure good phase accuracy and low noise for the 0° and 90° signals used in the I and Q phase detectors. The oscillator runs at 8x the input frequency and is designed for an IF input of 455 or 450 KHz. The loop is normally designed for a natural frequency of about 10 Hz so that the 25 Hz pilot can be recovered from the quadrature detector.

To get some idea of the requirements for these decoders, consider that they have an audio S/N of about 55 dB at 30% AM stereo modulation. The peak phase modulation of the carrier at 30% L only modulation is 23.2°, so the residual phase modulation of either the decoder PLL or incoming signal must be less than .073° RMS or .106° peak. The 25 Hz level at the Q detector is typically 20 mV peak, and the noise is typically about 1 mV peak. This gives a typical phase noise of about .07° peak at the low frequencies.

The MC13020/ULN3820 ICs are the first version of C-QUAM decoders, and the MC13022 is a newer version with added features and some improvements.

Using the MC13020/ULN3820 Decoders for other purposes:

Application notes available from both Motorola and Sprague give all the necessary parameters for designing the PLL of this decoder. The VCO and phase detector parameters are

given for LC, Ceramic and Crystal VCOs. The output current of the phase detector is about 10 times the locked value when unlocked so that lock-up is very fast. Referring to Figure 2, note that there is an operational amplifier between pins 13 and 14 which can be used as an independent amplifier in other applications. Pin 11, the AGC'd Q output can be used for phase noise measurements. In this application, the 430 Ω resistor (and other components) connected to this pin should be omitted. A 4.7 nF capacitor from pin 11 to ground will reduce the residual RF. 200 mV RMS carrier level is an ideal IF input level, but the internal AGC circuit is very good, so the input level can be between 100 and 350 mV RMS if the carrier level is constant. With the components shown, the loop frequency is about 10 Hz, so the Q output will be a phase detector with a flat response from 25 hZ to about 7 KHz depending on the bandwidth of the circuits on the input of the IC and the value of the Q detector filter capacitor on pin 20. Of course, a lower loop natural frequency could be used if desired. It may be necessary to adjust the resistor in series with the loop filter capacitor on pin 19 to make the Q output flat down to the loop cutoff frequency. The phase detector constant at pin 11 is 636 mV peak/radian. Any noise produced by the IF source or receiver LO will show up here.

Note that the I,Q, and Envelope detector outputs are available on pins 1,20, and 2 respectively. The output impedance on these pins is 4.3 KΩ. The bad news is that, the corrector circuit inside the IC forces the I output to be the same as the envelope detector so that it is not a real synchronous detected output. The good news is that, the corrector can be disabled by increasing the size of the error amp bypass capacitor on pin 5. Anything greater than 10 µF works well unless the carrier fades completely because the corrector is DC coupled internally. The IC also includes a loop lock detector which causes pin 10 to go to Vcc when the loop locks. Be careful when using this pin, because it can also go high when the 450 KHz input to the IC is less than 40 mV RMS. The lock detect signal could be used to control a gate to switch from envelope to synchronous detector when the loop has locked. The audio output level of both the envelope and I detectors is the same as the level at pins 7 and 8, the audio output pins, but the DC level is not, so some provision for removing the DC component at these pins should me provided if an audio amplifier is switched between them.

The components on pins 11,12,13,14 and 15 can be omitted if the IC is to be used as a phase or synchronous detector. They are used to prevent the decoder from going into the AM stereo mode if the received signal is marginal and for detecting the pilot signal and turning on the LED stereo indicator.

The MC13022 decoder:

This IC was designed to have more features and better performance than the first generation decoder and has some

features which make it ideal for synchronous detection. The loop itself is essentially the same as that in the MC13020 except that it has a little less gain and the internal multipliers are more carefully balanced. Referring to Figure 3, note the following:

1. An IF amplifier is included. It has its own AGC, plus there is an internal variable shunt attenuator at the input. A resistor should be used in series with pin 5 if the input level is to exceed 10 mV RMS. The input signal level range is 3 mV to more than 1V RMS if 10 KΩ is used. The input signal can also be applied to pin 2 if a coil is used between pins 2 and 3 for DC biasing. In this case, the ideal input level for lowest audio distortion is 150 to 170 mV RMS. The IF amplifier does add a small amount of distortion, so the pin 2 input will produce the lowest audio distortion. This is somewhat academic because the overall distortion is lower than most signal generators.

2. The L and R outputs are available on pins 10 and 11. If these pins are used, an RC lowpass output filter should be added to eliminate IF signals on these pins from the audio circuitry. The DC voltage on pins 10 and 11 normally biases the controlled amplifiers and buffers on pins 7,8,9 and 12,13,14. The controlled amplifiers are intended to be used as a variable notch/output filter but can be used independently of the rest of the decoder if they are supplied with a DC bias to turn them on. The control pin is 15. When V15 is low, the controlled amplifier gain is -.85 . When V15 is high, the gain is +.85. At 2.4 V the gain is 0.

3. The decoder will operate using a supply voltage of 4 to 10 V and uses only 18 mA.

Note also that the filter capacitor on pin 24 does not have a resistor in series with it because there is a 330 Ω resistor internal to the IC in series with pin 24. The loop with these components is slightly underdamped and has about a 1 dB peak at 8.5 Hz. 100 Ω in series with the loop filter capacitor will flatten out the phase detector response at pin 26, the Q out. Under these conditions, the phase detector constant is 735 mV peak/radian The internal resistance at pin 26 is 11KΩ, so the 2.2 KΩ/10 µF network coupling the Q out to the pilot detector in affects the Q out response. If the IC is used for measuring phase noise, pin 26 should be bypassed with a 1 nF capacitor to ground and the 2.2 KΩ removed. The high frequency response of the phase detector for these conditions is -3 dB at 6 KHz.

The decoder normally changes from mono to stereo by changing the L-R level. In the mono mode the L-R output is zero and it reaches full output when pin 23, the DC blend voltage reaches 2.1 V. The cos Φ corrector is not enabled until pin 23 exceeds approximately 2.8 V.

The best part about this decoder is the performance of the detectors. They are better than any others available at this

time and can be used to test the distortion of most signal generators. Another nice feature is that the I and Q detectors remain in the synchronous mode unless a stereo pilot is received. The correction circuit is not active until the blend voltage on pin 23 exceeds 2.8 V, and this is controlled by the "Signal Quality Detector" which needs the 25 Hz pilot to operate. The envelope detector feeds the L and R outputs in mono and is a quasi-synchronous type where the input signal is hard limited and applied to a multiplier. The THD of this envelope detector is .22 % at 100 % modulation. The output of the I detector on pin 28 is even better and measures only 2.7 % THD at 120 % modulation. (In the stereo mode with a 50 % L only signal, the THD is only .3 %) The envelope detector could, obviously, be used at other frequencies higher than 450 KHz. The PLL, has an internal phase-shift network optimized for 450 KHz, but it could also be used at higher frequencies.

Since the phase locked loop must be very accurate in order to obtain good AM stereo separation, typically greater than 30 dB, it produces very accurate I and Q signals. The oscillator is designed to produce 90° within +/-1° phase accuracy. A good test of this is to apply an AM signal to the input of the IC and check the Q output. It should be zero. Then apply a 1 KHz FM signal modulated to 1 KHz deviation and check the I output. It should be a double frequency sine wave with symmetrical peaks. Move the carrier 1 KHz off the center frequency and then repeat the test to see how well the loop maintains a 0° phase error.

Summary:

It would be very difficult and much more expensive to build a synchronous detector using building block ICs having I and Q outputs and the performance of either of the C-QUAM AM stereo decoders. The total component cost of either of the two circuits would be much more than that using either the MC13020 for MC13022. Even more impressive is the fact that these parts are designed to be used in mass produced consumer radios and will operate over a temperature range of -40° to +85° C.

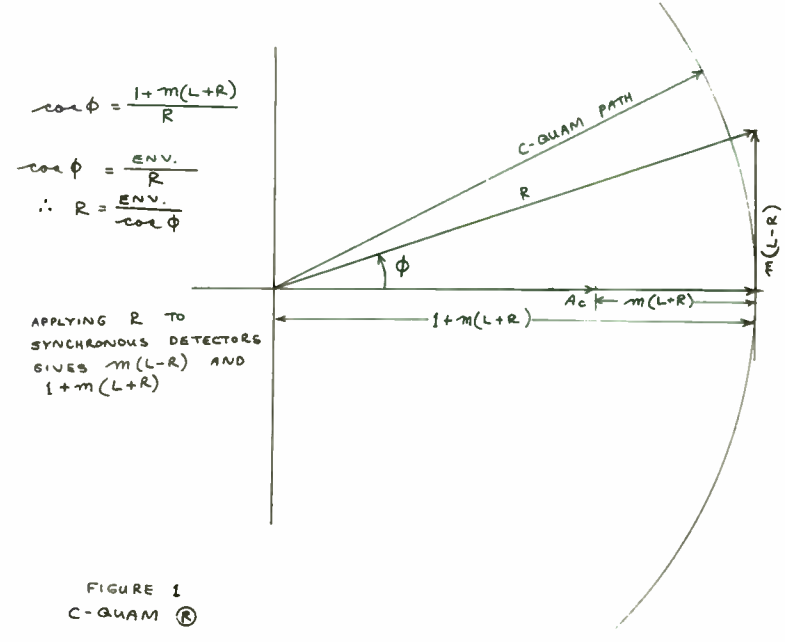


FIGURE 1
C-QUAM ®

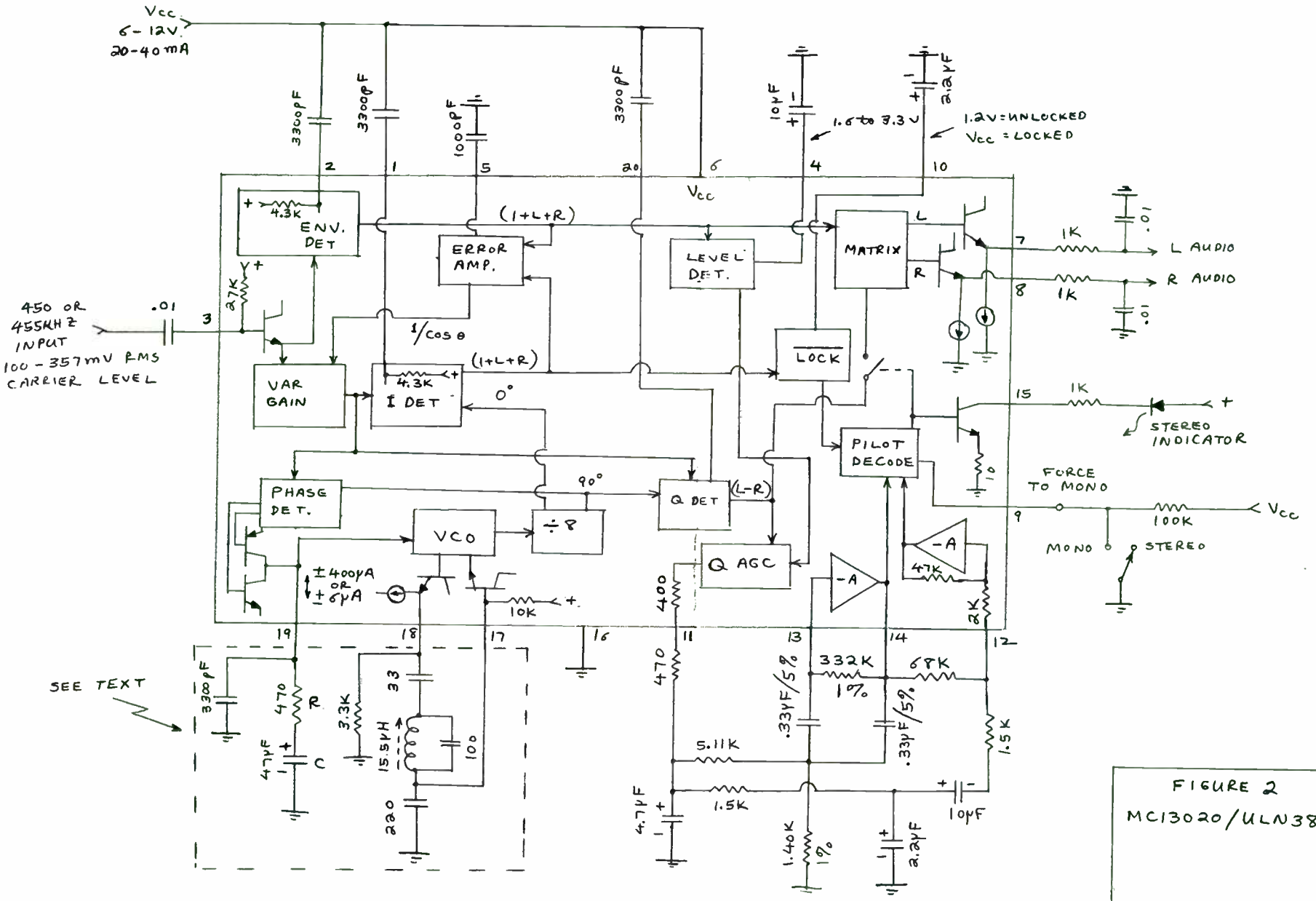


FIGURE 2
MC13020/ULN3820

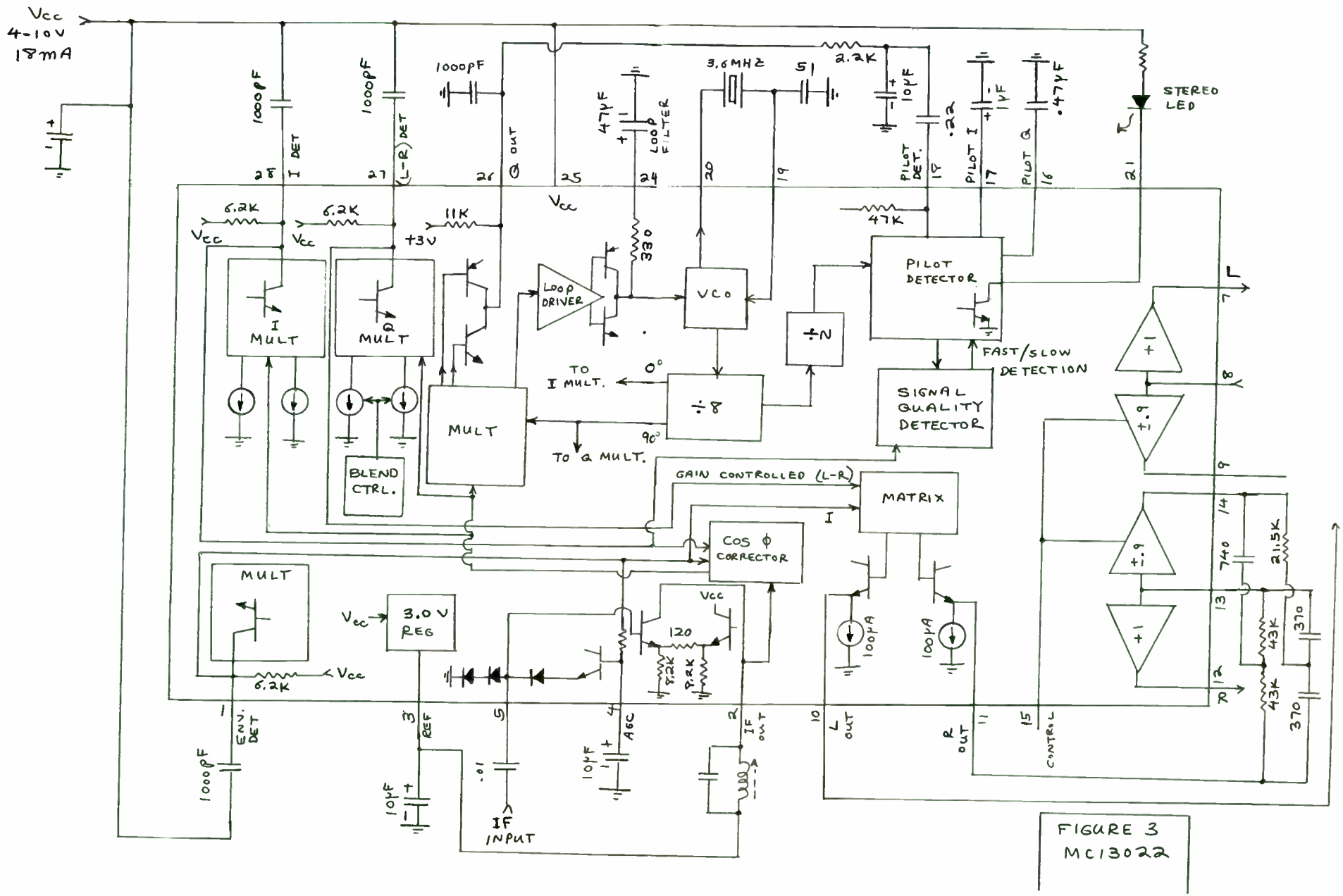


FIGURE 3
MC13022

A Single Chip 2 Gbit/s Clock Recovery Subsystem for Digital Communications

by

Ronald M. Hickling
GigaBit Logic, Inc.
1908 Oak Terrace Lane
Newbury Park, CA 91320

Introduction

A self-contained clock recovery/data resynchronizer phase locked loop (PLL) for use in microwave and fiber optic digital communications has been fabricated using GaAs integrated circuit technology. The IC contains the analog and digital components for the PLL: an edge-triggered phase detector based on a 1.2 GHz phase/frequency comparator, an op amp for creating the loop filter, and a VCO based on a differential source-coupled pair amplifier.

A photomicrograph of the "Clock and Data Recovery Subsystem" (CDR) appears in Fig.1. The die measures 74 X 130 mils. The VCO circuit is in the upper left-hand corner of the chip, and the op amp is in the lower left-hand corner. The reclocking flip-flop lies in the middle of the chip with the phase detection circuitry occupying the right-hand side. For maximum application flexibility, the loop filter components and VCO resonator are attached to the IC externally. The die has been packaged in two forms: (1) in a leadless chip carrier (LCC) which, in turn, is mounted on a duroid circuit board and (2) on an alumina substrate which is housed by a "plug-in" type hybrid package. The first approach is for applications less than 1 Gbit/s while the latter is used for maximum data rate operation in a compact package.

Circuit Operation

A block diagram of the CDR loop appears in Fig. 2. A delayed version of the incoming data is fed into the R input of the phase/frequency comparator; the V input is fed by a VCO-reclocked version of the same data. The reclocking operation effectively delays the rising edge of the V input data until the rising edge of the VCO generated clock. The skew between the two rising edges is sensed by the phase/frequency comparator and used to control the VCO through a loop filter. The external delay line is required by the loop so that it is possible for the clock rising edge to "precede" the data rising edge. If this delay were zero, an erroneous phase reading would occur each time the clock rising edge leads the data since the previous bit would be reclocked.

The phase detector is a "sample and hold" variety since when no transitions occur in the incoming data, no transitions occur in either the R or V inputs of the phase/frequency comparator, keeping both the U and D outputs at zero. Thus no error voltage is injected into the loop filter. When the loop is out of lock, the rising edge of the clock "slides" past the data waveform, causing beat notes to occur in the phase/frequency comparator output. These beat notes trace out the phase detector characteristic in time and thus have a sawtooth shape. Representative oscillographs of the U and D outputs of the phase detector are shown in Fig. 3.

Loop Performance

In order to test the ability for the loop to function in the presence of "missing" transitions, the CDR was tested using a 2^23-1 bit long pseudorandom pattern. A block diagram of the test set up appears in Fig. 4. The test was performed using the LCC version of the part at approximately 600 MHz. This frequency is sufficiently low to ensure that the loop components will behave close to theoretically ideal so that the loop can be characterized from an architectural standpoint.

The jitter of the resynchronized data output (referenced to the transmitter clock) was found to be somewhat independent of pattern length and is about 25 ps. The recovered clock and resynchronized 2^23-1 long pseudorandom data were both fed into a bit error rate receiver in order to count total bit errors; none were detected.

The spectrum of the VCO output is shown in Fig. 5. Fig. 5(a) shows the spectral purity of the VCO running at 583 MHz in an open loop configuration. Fig. 5(b) shows the clock recovered from a 2^7-1 bit long pseudorandom pattern clocked at 583.5 MHz. The presence of 1 MHz FM sidebands indicate a tendency for the loop to drift slightly; however, these "spurs" are approximately 40 dBc below the clock. When a 2^23-1 long pattern was used, these sideband levels increased to -30 dBc. However, both of these values are quite acceptable for a variety of high speed point-to-point LAN-type applications. For low-phase noise applications such as long-haul repeaters, these sidebands can be dropped through the use of a loop filter with switchable bandwidth, a VCXO, or both.

Circuit Performance

Each of the loop components were characterized individually in order to obtain the appropriate loop parameters and to determine the maximum frequency of operation. The phase detector section was found to operate up to 2.4 Gbps with a "large signal" gain constant of approximately 95mV/radian (300 mV per π radians) with an alternating 1-0 pattern input. Note that this is half of what one would expect from the phase/frequency comparator alone. The reason for this 50% degradation in the detector gain is that a bit transition occurs in the R and V inputs only once per clock period instead of twice as would be the case when using the phase comparator in its "normal" operating mode. One should note that this detector gain further degrades in the absence of missing transitions. This is given by the following formula:

$$Kd(\text{effective}) = 0.5 * \frac{\text{number of transitions}}{\text{number of bits}} * Kd\ (\text{phase/frequency comparator})$$

The data input sensitivity is approximately 300 mV for data rates less than 500 Mbps rising to 600 mV for maximum data rate operation.

The VCO circuit diagram is shown in Fig. 6. An on chip varactor was used in order to minimize interconnection parasitics when achieving 2.0 GHz operation. The feedback capacitor is an MIM capacitor. The LCC package has been used to achieve 1.2 GHz operation; above this frequency, however, package and inter-connect parasitics seem to dominate. The microstrip resonator on the hybrid has been used to achieve 2.08 GHz operation, although inadequate AC grounding has resulted in poor resonator Q

which limited the output levels to approximately 150 mV. Redesign of the alumina substrate is expected to correct this problem. Satisfactory performance has been achieved in the hybrid at 1 GHz, however. The overall performance of the VCO is summarized below:

Resonator Approach*	Maximum Frequency	Peak Output Voltage	Tuning Range**
microstrip (duroid)	1.2 GHz	800 mV	80 MHz
semirigid coax	600 MHz	1.2V	20 MHz
microstrip (alumina)	2.08 GHz	150 mV	~200 MHz
microstrip (alumina)	1.08 GHz	1.2V	~100 MHz

* The duroid and semirigid coax approaches used the LCC part; the microstrip on alumina results were obtained with dice.

** For the first two approaches, the tuning range was measured closed loop, for the latter two the tuning range was measured open loop with a tuning voltage of +/- 2.5V.

Conclusions

Results for a new type of single-chip clock recovery loop have been presented. Potential applications for this component are high speed LANs, long haul networks, and digital microwave communications.

Acknowledgement

The author wishes to thank Ray Tasker of Tasker Associates for development of the duroid board and generation of the microstrip on duroid test results.

References

(1) Floyd M. Gardner, Phaselock Techniques, Wiley, 1979.

(2) GigaBit Logic, 16G040 and 16G044 data sheets, 1987.

(3) Vijay K. Bhargava, et al, Digital Communications by Satellite, Wiley, 1981.

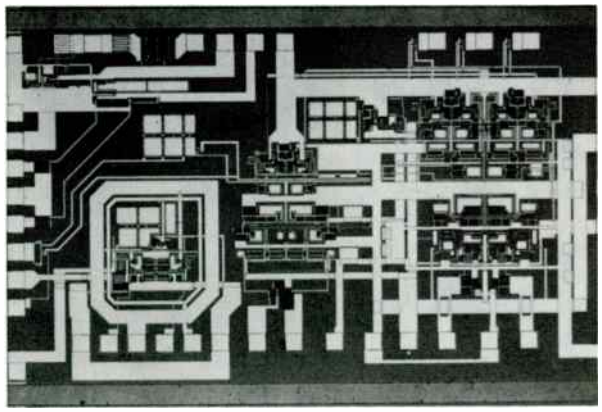
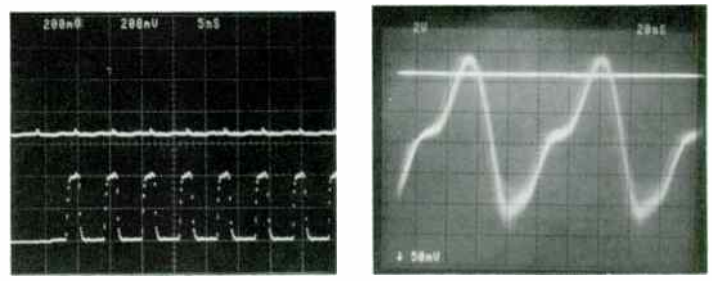


FIGURE 1: PHOTOMICROGRAPH OF THE 16G040 CLOCK AND DATA RECOVERY CIRCUIT



(A) (B)

FIGURE 3: PHASE DETECTOR OUTPUT WAVEFORMS

(A) 340 Mbps alternating pattern with an approximate 2.1 ns leading clock edge (U is upper waveform, D is lower; corrected vertical (10 dB) = 632 mV/division).
(B) Beat note waveform at 650 MHz measured at the loop filter output.

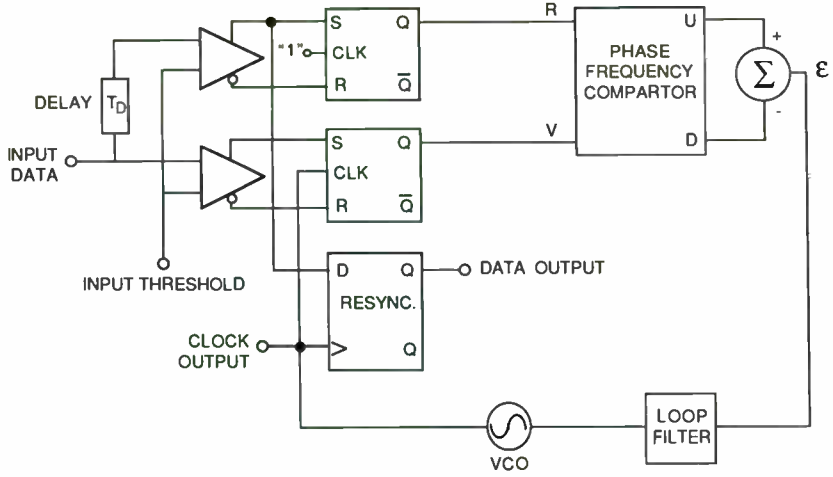


FIGURE 2: BLOCK DIAGRAM OF THE 16G040 CLOCK/DATA RECOVERY CIRCUIT (PATENT PENDING)

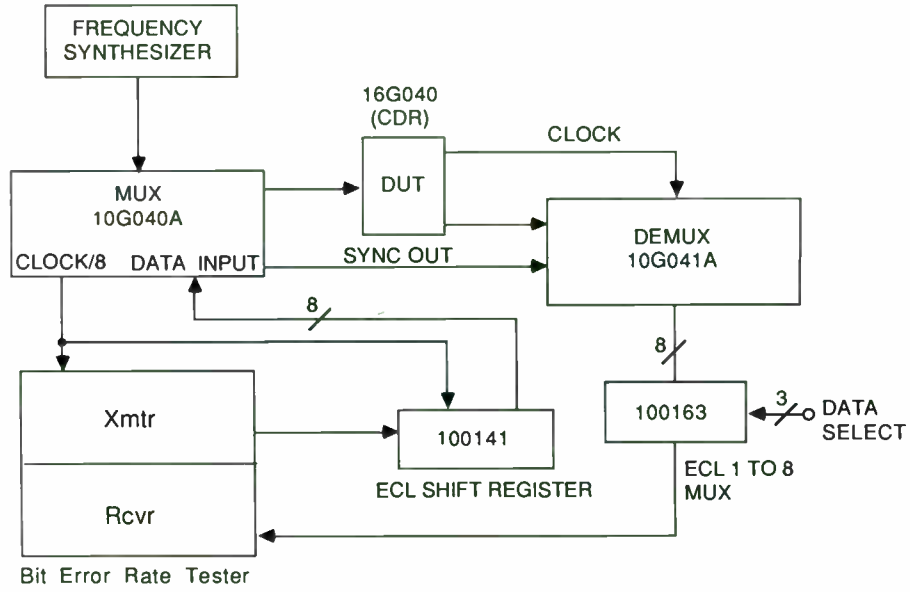
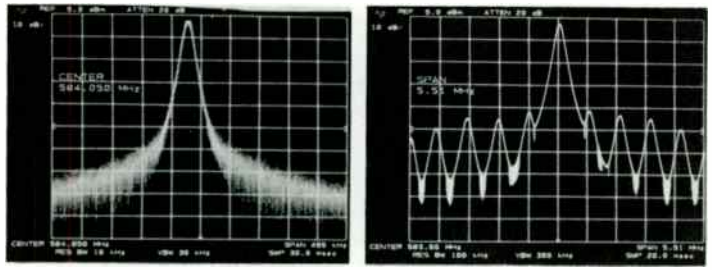


FIGURE 4 : LOOP PERFORMANCE TEST USING PSEUDORANDOM DATA



(A) (B)

FIGURE 5: RECOVERED CLOCK SPECTRUM

(A) The inherent purity of the VCO with the loop open.
(B) Recovered clock (input pattern = 2 exp 7 - 1) showing drift spurs approximately 40 dB down.

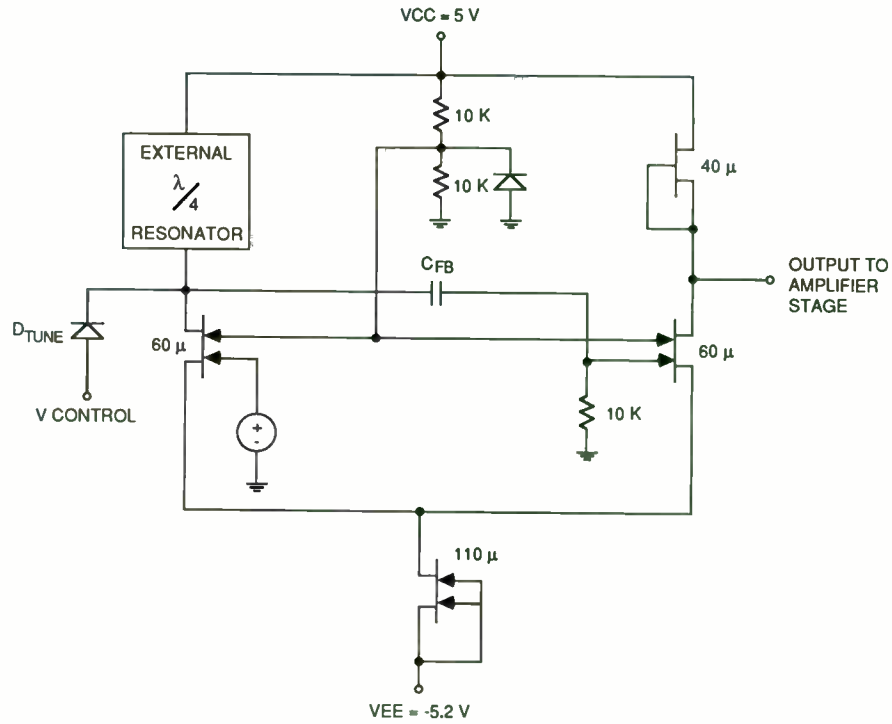


FIGURE 6: SCHEMATIC OF THE 16G040 VCO

Envelope-elimination-and-restoration system requirements

by

Frederick H. Raab, Ph.D.
Green Mountain Radio Research Company
50 Vermont Avenue, Fort Ethan Allen
Winooski, Vermont 05404

ABSTRACT

Envelope elimination and restoration (EER) is a technique through which highly efficient but nonlinear RF power amplifiers can be combined with highly efficient AF amplifiers to produce a high-efficiency linear-amplifier system. The linearity of an EER system depends not only upon the linearity of its components, but also upon system characteristics such as AM-PM conversion, envelope bandwidth, carrier bandwidth, phase bandwidth, and differential delay. This paper derives system requirements as functions of system parameters for both multicarrier and two-tone RF signals. Numerical Fourier-series techniques are used for multicarrier signals, while semianalytical techniques are used for two-tone signals.

1. INTRODUCTION

Envelope Elimination and Restoration (EER) is a technique through which highly efficient but nonlinear RF power amplifiers (PAs) can be combined with highly efficient audio amplifiers to implement a high-efficiency linear RF power amplifier. A limiter (Figure 1) *eliminates* the envelope, allowing the constant-amplitude, phase-modulated carrier to be amplified efficiently by class-C, -D, -E, or -F RF PAS. The detected envelope is amplified efficiently by class-S or -G AF PAs. Amplitude modulation of the final RF PA *restores* the envelope to the phase-modulated carrier, creating an amplified replica of the

input signal.

EER was originally developed by Kahn [1] in the 1950s as a means of improving short-wave broadcast transmitters. It has been applied [2] to high-powerMF and HF broadcasting, experimental HF and VHF transmitters, and has also been used in both amateur and commercial satellite repeaters operating at VHF and UHF.

Amplitude-modulated signals (including multicarrier signals) spend a considerable portion of the time at relatively low signal levels [4]. The efficiency of a class-B PA decreases linearly with decreasing signal amplitude, resulting in a low average efficiency for typical amplitude-modulated signals. In contrast, the efficiency of switching-mode amplifiers remains high for both large and small signal amplitudes, resulting in a much higher average efficiency for typical amplitude-modulated signals. Consequently, an EER system that employs a class-D RF PA and a class-S modulator can be several times as efficient as a class-B linear RF PA for typical AM signals.

The intermodulation-distortion characteristics of an EER system can be improved by feedback techniques [2], including:

- Envelope feedback,
- Polar feedback, and
- Cartesian feedback.

Envelope feedback uses an audio-frequency feedback loop to force the output envelope to track the input envelope. Polar feedback combines envelope feedback with phaselock, while Cartesian feedback locks the I and Q components of the system output to those of its input.

However, the linearity of an EER system depends not only upon the linearity of its components, but also upon system characteristics such as AM-PM conversion, envelope bandwidth, carrier bandwidth, phase bandwidth, and differen-

tial delay. This paper derives system requirements as functions of system parameters for both multicarrier and two-tone RF signals.

Table 1 presents system requirements based upon two measures of linearity. The carrier-to-intermodulation (C/I) ratio compares the IM products to the amplitude of the unmodulated (PEP) carrier. The signal-to-intermodulation ratio (S/I) compares the IM products to the desired output signal; for a two-tone signal, S/I is lower than C/I by 6 dB.

2. MULTICARRIER SIGNALS

The IMD produced by an EER system driven by a deterministic signal can be determined numerically by transforming between time and frequency domains to accommodate nonlinear processes and linear filters. The specific set of steps in such a procedure depends upon the model of the EER system. For a simple EER system with finite envelope- and phase-channel bandwidths, nonzero differential delay, and AM-PM conversion, the procedure consists of the following steps (Figure 2):

- Specify Fourier coefficients of the RF signal,
- Compute Fourier coefficients of the I and Q waveforms,
- Compute E and ϕ waveforms,
- Compute Fourier coefficients of E and ϕ,
- Modify E and ϕ Fourier coefficients for bandwidth and delay,
- Generate resultant E and ϕ waveforms from the modified coefficients,
- Modify the waveforms for amplitude nonlinearity and AM-PM conversion,
- Compute resultant I and Q waveforms,
- Compute Fourier coefficients of the resultant I and Q waveforms,
- Compute Fourier coefficients of the resultant RF signal.

GMRR TP87-4

This section derives the applicable mathematical relationships.

The spectrum of the RF signal is illustrated in Figure 3. The signal comprises up to $2M + 1$ components centered at frequency f_c and contained within bandwidth B_{RF}. The frequencies of the components differ from the center frequency by multiples of f_m, hence f_m is the lowest frequency of interest in subsequent computations. All frequency components generated by subsequent processes are therefore multiples of f_m.

The temporal representation of the RF signal is

$$e_{RF}(t) = \sum_{m=-M}^{+M} [a_{RFm} \cos(\omega_c t + m\omega_m t) - b_{RFm} \sin(\omega_c t + m\omega_m t)] \quad , \qquad (1)$$

where a_{RFm} and b_{RFm} ($-M \leq m \leq +M$) are its Fourier coefficients.

Fourier Coefficients of I and Q Waveforms

EER is intended for use with bandlimited signals ($f_c >> f_m$), hence the RF signal can be described by in-phase (I) and phase-quadrature (Q) modulation functions. Since f_m is the lowest frequency that can occur in the modulation functions, it is convenient to define *angular time* as $\theta = \omega_m t$, where $0 \leq \theta < 2\pi$.

The in-phase modulation waveform $e_I(\theta)$ can be obtained by mixing the RF signal with a cosinusoid of frequency f_c. It is convenient to represent the I waveform as the Fourier series

$$e_I(\theta) = a_{I_0} + \sum_{m=1}^{M} [a_{Im} \cos m\theta - b_{Im} \sin m\theta] \quad . \qquad (2)$$

Trigonometric manipulation produces

$$a_{Im} = \begin{cases} a_{RF_0} & , m = 0 \\ a_{RF(+m)} + a_{RF(-m)} & , 1 \leq m \leq M \end{cases} \qquad (3)$$

and

$$b_{Im} = b_{RF(+m)} - b_{RF(-m)}, \quad 1 \le m \le M \quad . \tag{4}$$

The phase-quadrature modulation waveform $s_Q(\theta)$ is similarly obtained by mixing the RF signal with a negative sinusoid of frequency f_c. The Fourier series for the Q waveform is similarly defined by

$$s_Q(\theta) = a_{Q_0} + \sum_{m=1}^{M} [a_{Qm} \cos m\theta - b_{Qm} \sin m\theta], \tag{5}$$

where

$$a_{Qm} = \begin{cases} + b_{RF_0} & , m = 0 \\ + b_{RF(+m)} + b_{RF(-m)}, & 1 \le m \le M \end{cases} \tag{6}$$

and

$$b_{Qm} = - a_{RF(+m)} + a_{RF(-m)}, \quad 1 \le m \le M \quad . \tag{7}$$

Envelope and Phase Waveforms

The envelope and phase waveforms are related to the I and Q waveforms by the nonlinear functions

$$E(\theta) = [s_I^2(\theta) + s_Q^2(\theta)]^{1/2} \tag{8}$$

and

$$\phi(\theta) = \arctan \left[\frac{s_Q(\theta)}{s_I(\theta)}\right] \quad , \tag{9}$$

respectively [1]. Equation (9) requires a four-quadrant inverse tangent.

Envelope and Phase Fourier Coefficients

The Fourier series for the envelope and phase waveforms are defined by

$$E(\theta) = a_{E_0} + \sum_{n=1}^{\infty} [a_{En} \cos n\theta - b_{En} \sin n\theta] \tag{10}$$

and

$$\phi(\theta) = a_{\phi_0} + \sum_{n=1}^{\infty} [a_{\phi n} \cos n\theta - b_{\phi n} \sin n\theta] \quad . \tag{11}$$

Since these waveforms are generated by nonlinear processes acting upon the I and Q waveforms, they are not in general bandlimited.

The Fourier coefficients in (10) and (11) can be obtained by standard Fourier integrals of the relevant waveforms. For example:

$$a_{E_0} = \frac{1}{2\pi} \int_0^{2\pi} E(\theta)\, d\theta \tag{12}$$

$$a_{En} = \frac{1}{\pi} \int_0^{2\pi} E(\theta) \cos n\theta\, d\theta \tag{13}$$

$$b_{En} = \frac{-1}{\pi} \int_0^{2\pi} E(\theta) \sin n\theta\, d\theta \quad . \tag{14}$$

Since $E(\theta)$ and $\phi(\theta)$ are in general awkward functions, their Fourier coefficients must in general be determined numerically.

Effects of System Characteristics

Finite envelope- and phase-channel bandwidths are modelled by truncation or alteration of the corresponding Fourier series. Nonlinear gain and AM-PM conversion are introduced by distorting the envelope and/or phase waveforms. Differential delay can be modelled either by time-shifting the envelope waveform or by phase rotation of the Fourier coefficients of the envelope.

Resultant I and Q Waveforms

The resultant I and Q waveforms are obtained from the distorted envelope and phase waveforms by the relationships

$$\hat{s}_I(\theta) = \hat{E}(\theta) \cos \hat{\phi}(\theta) \quad . \tag{15}$$

and

$$\hat{s}_Q(\theta) = \hat{E}(\theta) \sin \hat{\phi}(\theta) \quad . \tag{16}$$

These equations are the inverse of equations (8) and (9). The Fourier coefficients of $\hat{s}_I(\theta)$ and $\hat{s}_Q(\theta)$ are computed from the waveforms by analogy to (12) - (14).

Resultant RF Signal

The Fourier coefficients of the resultant RF signal are directly related to the Fourier coefficients ($\hat{a}_{Im}$, $\hat{b}_{Im}$, $\hat{a}_{Qm}$, and $\hat{b}_{Qm}$) of the resultant I and Q waveforms. Since the RF signal is the result of modulation of cosine and sine carriers by $\hat{s}_I(\theta)$ and $\hat{s}_Q(\theta)$,

$$\hat{s}_{RF}(t) = \hat{s}_I(\theta) \cos \omega_c t - \hat{s}_Q(\theta) \sin \omega_c t \tag{17}$$

Collection of the corresponding terms in (46) then yields the Fourier coefficients of the resultant RF signal:

$$a_{\hat{RF}m} = \begin{cases} \frac{1}{2}(a_{\hat{I}m} + b_{\hat{Q}m}), & m < 0 \\ a_{\hat{I}_0}, & m = 0 \\ \frac{1}{2}(a_{\hat{I}m} - b_{\hat{Q}m}), & m > 0 \end{cases} \tag{18}$$

$$b_{\hat{RF}m} = \begin{cases} \frac{1}{2}(-b_{\hat{I}m} + a_{\hat{Q}m}), & m < 0 \\ a_{\hat{Q}_0}, & m = 0 \\ \frac{1}{2}(b_{\hat{I}m} + a_{\hat{Q}m}), & m > 0 \end{cases} \quad . \tag{19}$$

Intermodulation Products

The IMD products can now be determined from

$$a_{\tilde{RF}m} = a_{\hat{RF}m} - a_{RFm} \tag{20}$$

and

$$b_{\tilde{RF}m} = b_{\hat{RF}m} - b_{RFm} \quad . \tag{21}$$

3. TWO-TONE SIGNALS

The relative simplicity of the commonly used two-tone (two-carrier) test signal allows the derivation of analytic or semianalytic relationships between each of the system parameters and the IMD. This note derives such relationships for abrupt-cutoff filters and pure delay, and the technique can be expanded to include other filter and delay characteristics.

The input and output signals are described by

$$v_i(t) = E_i(\theta) \cos[\omega_c t + \phi_i(\theta)] \quad . \tag{22}$$

and

$$v_o(t) = E_o(\theta - \tau) \cos[\omega_c t + \phi_o(\theta)] \quad , \tag{23}$$

where $\theta = \omega_m t$.

The envelope and phase functions for the two-tone test signal [RN86-34] are

$$E_i(\theta) = |\cos\theta| \tag{24}$$

and

$$\phi_i(\theta) = \frac{\pi}{2}[1 - c(\theta)] \quad , \tag{25}$$

respectively, where $c(\theta)$ is a +1/-1 squarewave with cosinusoidal phasing.

Expansion of the envelope function as a Fourier series produces

$$E_o(\theta) = [a_0 + \sum_{m=2,4,6,\ldots} a_m \cos m\theta] \quad , \tag{26}$$

where

$$a_m = \begin{cases} \frac{2}{\pi}, & m = 0 \\ \dfrac{4(-1)^{(m-2)/2}}{\pi(m^2-1)}, & m > 2 \end{cases} . \tag{27}$$

Similar expansion of the switching function yields

$$c(\theta) = \sum_{n=1,3,5,\ldots} c_n \cos n\theta \quad , \tag{28}$$

where

$$c_n = \frac{4}{\pi}\frac{(-1)^{(n-1)/2}}{n} \quad . \tag{29}$$

4. AM-PM CONVERSION

The relationship between the amount of AM-PM conversion and the resultant IMD levels can be determined by approximation for the relatively low IMD levels that are of practical interest. This relationship can then be used to specify the maximum AM-PM conversion allowable in the limiter of an EER system.

The AM/PM conversion in the limiter can be represented by the phase-modulation function $\phi_a(t)$. The IMD produced by this function is then

$$x_d(t) = \cos[\omega_c t + \phi_a(t)] - \cos\omega_c t \tag{30}$$

$$= [\cos\phi_a(t) - 1]\cos\omega_c t - \sin\phi_a(t)\sin\omega_c t \quad . \tag{31}$$

Since the distortion is assumed to be relatively small, $|\phi_a(t)| \ll 1$ and

$$x_d(t) \cong -\phi_a(t)\sin\omega_c t \quad . \tag{32}$$

The AM-PM phase deviation can be related to the envelope through a power series:

$$\phi_a(t) = c_1 E(t) + c_2 E^2(t) + c_3 E^3(t) + \ldots \quad . \tag{33}$$

However, for the cases of interest, $|\phi_a|$ is small and only the first term in (33) is significant, hence

$$\phi_a(t) \cong c_1 E(t) \quad . \tag{34}$$

For a two-tone signal, the phase deviation is conveniently represented by

$$\phi_a(t) = \phi_{max}\,|\cos\omega_m t| \tag{35}$$

$$= \phi_{max}\left(\frac{2}{\pi} + \frac{4}{3\pi}\cos 2\omega_m t - \frac{4}{15\pi}\cos 4\omega_m t + \ldots\right) \quad . \tag{36}$$

The dc and second-harmonic terms of (36) are the most significant. Insertion of the first two terms of (36) into (32) produces

$$x_d(t) \cong -\frac{2\phi_{max}}{\pi} \sin \omega_c t - \frac{2\phi_{max}}{3\pi} \sin (\omega_c t - 2\omega_m t) - \frac{2\phi_{max}}{3\pi} \sin (\omega_c t + 2\omega_m t) \quad . \tag{37}$$

These terms therefore produce zero-order (carrier-frequency) and second-order IMD products.

The maximum phase deviation ϕ_{max} due to AM-PM conversion can now be determined by comparing the amplitudes of the IMD products in (37) with the normalized carrier amplitude (1) or desired-sideband amplitude (1/2). Comparison of the amplitudes of the zero-order IMD product and the carrier produces

$$(C/I)_{dB} = 20 \log \frac{1}{2\phi_{max} / \pi} \tag{38}$$

$$\phi_{max} \leq \frac{\pi}{2} \cdot 10^{-(C/I,dB) / 20} \quad . \tag{39}$$

A 30-dB C/I ratio therefore requires that

$$\phi_{max} \leq 0.0497 \text{ rad} = 2.85° \quad . \tag{40}$$

5. ENVELOPE BANDWIDTH

For a two-tone signal, the phase changes are equivalent to changes in the carrier polarity. Consequently, the effects of finite envelope bandwidth can be determined by moving the switching function $c(\theta)$ outside of the cosine function in (23) and then combining it with the envelope function, thus

$$v_o(t) = E_o(\theta)\, c(\theta) \cos \omega_c t = y(\theta) \cos \omega_c t \quad . \tag{41}$$

If the output envelope is an exact replica of the input envelope, then $y(\theta) = \cos \theta$.

The modulation function $y(\theta)$ can be represented as the Fourier series

$$y(\theta) = b_1 \cos \theta + b_3 \cos 3\theta + b_5 \cos 5\theta + \ldots \quad . \tag{42}$$

The coefficient b_1 corresponds to the amplitude of the desired sidebands; the higher-order coefficients (b_3, b_5, ...) correspond to the amplitudes of the IMD products.

The spectral components of modulation function $y(\theta)$ are produced by mixing each of the spectral components of $E_o(\theta)$ with each of the spectral components of $c(\theta)$. The effect of an abrupt-junction filter in the envelope channel is modelled by truncating the Fourier series (26) after component M. Since the RF bandwidth of the two-tone signal is $2\omega_m$,

$$M = \text{int}\,(B_{RF}/2) \quad , \tag{43}$$

where int represents rounding to the nearest smaller integer.

Since

$$2 \cos n\theta \cos m\theta = \cos(m + n)\theta + \cos(m - n)\theta \quad , \tag{44}$$

the coefficients b_k in (44) have the forms

$$b_1 = \frac{1}{2} [2a_0 c_1 + a_2(c_1 + c_3) + a_4(c_3 + c_5) + \ldots] \quad , \tag{45}$$

$$b_3 = \frac{1}{2} [2a_0 c_3 + a_2(c_1 + c_5) + a_4(c_3 + c_7) + \ldots] \quad , \tag{46}$$

$$b_5 = \frac{1}{2}\,[2a_0c_5 + a_2(c_3 + c_7) + a_4(c_1 + c_9) + a_6(c_1 + c_{11}) + a_8(c_3 + c_{13}) + \ldots\,] \quad \text{, etc.} \tag{47}$$

The terms in (45) - (47) have the form $a_m(c_{|m-k|} + c_{m+k})$. Consequently, they eventually decrease as either $1/m^3$ or $1/m^4$, depending whether the two c_n are of the same or different sign. Determination of the values of the coefficients b_k is most readily accomplished numerically.

Since modulation of the carrier by $y(\theta)$ produces two components of amplitude $b_k/2$, the carrier-to-intermodulation ratio is

$$(C/I)_{dB} = 20 \log (2 / |b_n|) \quad . \tag{48}$$

The S/I ratio can be obtained by subtracting 6 dB from the C/I ratio.

The resultant C/I ratios are shown in Figure 4. Obtaining $C/I > 30$ dB requires $B_E > B_{RF}$, and obtaining $S/I > 30$ dB requires $B_E > 2\,B_{RF}$. It is interesting to note that order of the maximum-amplitude IM produce increases as the envelope bandwidth increases. Consequently, the third-order IM product is not generally the maximum IM product. For example, if $B_E = 5\,B_{RF}$, the ninth-order product is the largest.

The effects of envelope bandwidth B_E upon C/I ratio are shown in Figure 5 for multicarrier signals. It is apparent that the two-carrier (two-tone) signal produces the largest IM products for a given bandwidth.

6. CARRIER BANDWIDTH

The effects of finite carrier bandwidth are similarly analyzed. Since the bandwidth of the limiter output can be assumed to be effectively infinite,

the phase-switching function can be converted to an amplitude-switching function and therefore moved outside of the cosine function. The effect of carrier bandwidth is then modelled by truncating the Fourier series for $c(\theta)$ from (29) after coefficient N.

The truncated Fourier series for $c_o(\theta)$ is combined with the complete Fourier series for the envelope to produce a modulation function $y(\theta)$:

$$v_o(t) = E_i(\theta)\; c_o(\theta) \log \omega_c t = y(\theta) \cos \omega_c t \quad , \tag{49}$$

The spectral components of the modulation function are again produced as sums and differences of the spectral components of the E_i and c_o functions. The Fourier coefficients of the modulation function $y(\theta)$ therefore have the forms

$$b_1 = \frac{1}{2}\,[c_1(2a_0 + a_2) + c_3(a_2 + a_4) + c_5(a_4 + a_6) + \ldots\,] \quad , \tag{50}$$

$$b_3 = \frac{1}{2}[c_1(a_2 + a_4) + c_3(2a_0 + a_6) + c_5(a_2 + a_8) + \ldots\,] \quad , \tag{51}$$

$$b_5 = \frac{1}{2}[c_1(a_4 + a_6) + c_3(a_2 + a_8) + c_5(2a_0 + a_{10}) + \ldots\,] \quad \text{, etc.} \tag{52}$$

The terms in (51) - (53) have the form $c_n(a_{|n-k|} + a_{n+k})$, and are therefore readily evaluated numerically.

The resultant C/I ratios are shown in Figure 6. Obtaining $C/I > 30$ dB requires $B_C > 3.5\,B_{RF}$, and obtaining $S/I > 30$ dB requires $B_C > 7\,B_{RF}$. Again, the order of largest IM product increases as the carrier bandwidth increases, and the third-order product is not generally the largest. With $B_C = 7\,B_{RF}$ so that $S/I > 30$ dB, the 15^{th}- and 17^{th}-order IM products are most significant.

7. PHASE BANDWIDTH

The effects of finite phase bandwidth can be represented by the introduction of a phase error ϕ_e, producing the output signal

$$v_o = E_i \cos(\omega_c t + \phi_i + \phi_e) \tag{53}$$

$$= E_i [\cos(\omega_c t + \phi_i) \cos \phi_e - \sin(\omega_c t + \phi_i) \sin \phi_e] \quad . \tag{54}$$

For useable systems, $\phi_e << 1$, hence

$$v_o \approx E_i \cos(\omega_c t + \phi_i) - E_i \phi_e \sin(\omega_c t + \phi_i) \quad . \tag{55}$$

The distortion is the last term in (55) and therefore has the form

$$v_d = E_i \phi_e \sin(\omega_c t + \phi_i) \tag{56}$$

$$= - \phi_e \cos \phi \sin \omega_c t = y_d \sin \omega_c t \quad . \tag{57}$$

The phase error is the difference between the ideal switching function $c_i(\theta)$ and the actual phase response $c_o(\theta)$. The effect of an abrupt-cutoff filter in the phase channel is modelled by truncating the Fourier series (29) of the ideal switching function after the N^{th} term, hence

$$\phi_e = \phi_i - \phi_o = \tfrac{\pi}{2} [c_o(t) - c_i(t)] \tag{58}$$

$$= \tfrac{\pi}{2} \sum_{n=N+2}^{\infty} c_n \cos n\phi \quad . \tag{59}$$

The spectral components of ϕ_e are simply those components of ϕ_i that are not passed by the abrupt-cutoff filter.

The distortion-modulation function $y_d(\theta)$ therefore has the form

$$y_d = - \tfrac{\pi}{2} \cos \phi \sum_{n=N+2}^{\infty} c_n \cos n\phi \tag{60}$$

$$= - \tfrac{\pi}{4} \sum_{n=N+2}^{\infty} c_n [\cos (n - 1)\phi + \cos(n + 1)\phi] \tag{61}$$

$$= \sum_{n=0}^{\infty} b_n \cos n\phi \quad . \tag{62}$$

Its Fourier coefficients have the form

$$b_n = \begin{cases} 0 & , n < N \\ - \tfrac{\pi}{4} c_{n+1} & , n = N + 1 \\ - \tfrac{\pi}{4} (c_{n-1} - c_{n+1}) & \end{cases} \quad . \tag{63}$$

The maximum IM product is also the first IM product and occurs at $n = N + 1$. Its amplitude is simply

$$|b_n| = \tfrac{\pi}{4} |c_n| + 1 = \frac{1}{n + 1} \quad . \tag{64}$$

The magnitudes of the higher-order components are given by

$$|b_n| = \left[\frac{1}{n - 1} - \frac{1}{n + 1}\right] = \frac{2}{(n^2 - 1)} \quad . \tag{65}$$

The resultant C/I ratios are shown in Figure 7. Obtaining $C/I \geq 30$ dB requires $B_\phi \geq 7.5\ B_{RF}$, and obtaining $S/I \geq 30$ dB requires $B_\phi \geq 15\ B_{RF}$. Finite phase bandwidth produces even-frequency IMD components, in contrast to the odd-frequency IMD components produced by finite envelope and carrier bandwidths. Since the amplitudes of the higher-order components decrease with n^2,

the first IM component is the largest component and the others are often insignificant. It occurs just above $B_{RF}/2$.

The variation of C/I with phase bandwidth is shown in Figure 8 for several multicarrier signals. It is again apparent that the two-tone signal produces the maximum IMD products.

8. DIFFERENTIAL DELAY

Differential delay between the envelope and phase modulations is generally introduced into the envelope channel by the output filter of a class-S modulator. However, the magnitude of the resultant IMD is more readily determined by advancing the phase-switching waveform (Figure 9). The equivalent amplitude-modulation function for the output signal is therefore

$$y(\theta) = |\cos\theta|\, c(\theta + \tau) = \cos\theta + u(\theta) \quad , \tag{66}$$

where

$$u(\theta) = \begin{cases} -2\cos\theta\ , & \pi/2 - \tau < \theta < \pi/2 \\ -2\cos\theta\ , & 3\pi/2 - \tau < \theta < 3\pi/2 \\ 0\ , & \text{otherwise} \end{cases} \quad . \tag{67}$$

The differential delay τ in radians is related to the differential delay in seconds (Δt) by

$$\Delta t = \frac{\tau}{2\pi}\frac{1}{f_m} = \frac{\tau}{2\pi}\frac{2}{B_{RF}} = \frac{\tau}{\pi B_{RF}} \quad . \tag{68}$$

To determine the IMD products, $u(\theta)$ is expanded in a Fourier series, thus

$$u(\theta) = \sum_{k=1}^{\infty} (a_k \cos k\theta + b_k \sin k\theta) \quad . \tag{69}$$

Since only the magnitude of the IMD products is of interest, it is more convenient to analyze

$$u'(\theta) = \begin{cases} -2\sin\theta, & 0 < \theta < \tau \\ -2\sin\theta, & \pi < \theta < \pi + \tau \\ 0, & \text{otherwise} \end{cases} \quad . \tag{70}$$

Symmetry dictates that the even-order coefficients are zero. The odd-order coefficients are

$$a_k = (-4/\pi) \int_0^{\tau} \sin\theta \cos k\theta \, d\theta \tag{71}$$

$$= (-2/\pi) \left[\frac{1 - \cos(k+1)\tau}{k+1} + \frac{\cos(k-1)\tau - 1}{k-1}\right] \quad . \tag{72}$$

For $|x| \ll 1$,

$$\cos x \cong 1 - x^2/2 \quad . \tag{73}$$

Therefore for $k\tau \ll 1$,

$$a_k \cong (-2/\tau) \left[\frac{(k+1)^2\tau^2/2}{(k+1)} - \frac{(k-1)^2\tau^2/2}{k-1}\right] = (-2/\pi)\,\tau^2 \tag{74}$$

Similarly,

$$b_k = (-4/\pi) \int_0^{\tau} \sin\theta \sin k\theta \, d\theta \tag{75}$$

$$= (-2/\pi) \left[\frac{\sin(k-1)\tau}{k-1} - \frac{\sin(k+1)\tau}{k+1}\right] \quad . \tag{76}$$

For $k\tau \ll 1$,

$$b_k \cong (-2/\pi)\left[\frac{(k-1)\tau}{k-1} - \frac{(k+1)\tau}{k+1}\right] \cong 0 \quad . \tag{77}$$

It is therefore apparent that

$$c_k \cong |a_k| = (2/\pi)\,\tau^2 \quad . \tag{78}$$

The variation of C/I with delay is shown in Figure 10 for a 1-Hz RF bandwidth. Obtaining $C/I > 30$ dB requires $\Delta t < 0.1/B_{RF}$, and obtaining $S/I > 30$ dB requires $\Delta t < 0.07/B_{RF}$. For speech-bandwidth signals ($B_{RF} = 3$ kHz), a 30-dB S/I requirement implies a differential delay of no more than 23.7 μs. The IMD spectrum consists of odd-order products whose amplitudes decrease slowly with increasing order.

The variation of C/I with delay is shown in Figure 11 for the multicarrier signals. The two-tone test signal generally (but not always) produces the maximum IMD products.

9. REFERENCES

1. L. R. Kahn, "Single sideband transmission by envelope elimination and restoration," *Proc. IRE*, vol. 40, pp. 803 - 806, July 1952.
2. F. H. Raab, "Envelope-elimination-and-restoration concepts," *Proceedings of RF Expo East '87*, Boston, MA, pp. 167 - 177, November 11 - 13, 1987.
3. H. O. Granberg, "Applying power MOSFETs in class D/E RF power amplifier design," *R. F. Design*, vol. 8, no. 6, pp. 42 - 47, June 1985.
4. F. H. Raab, "Average efficiency of power amplifiers," *Proceedings of RF Technology Expo '86*, Anaheim, California, pp. 473 - 486, January 30 - February 1, 1986.

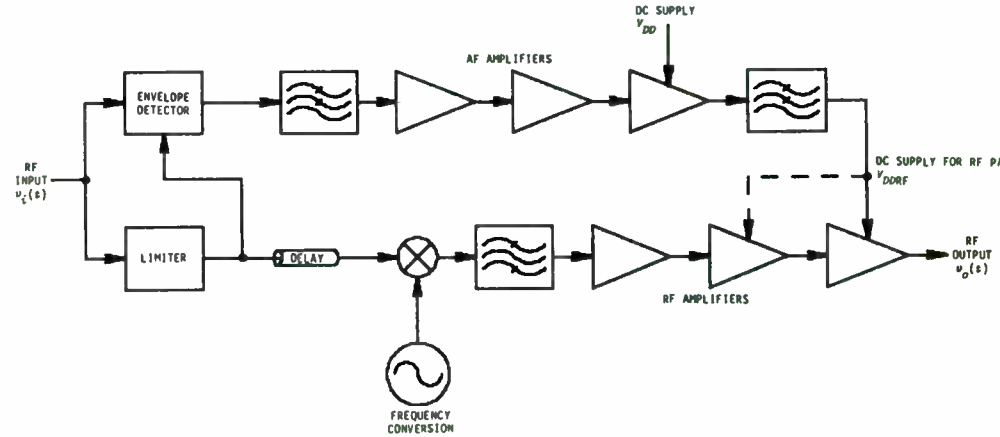


Figure 1. Basic EER system.

NUMBER OF CARRIERS	PHASING	ξ,dB	REQUIREMENTS FOR $C/I > 30$ dB			REQUIREMENTS FOR $S/I > 30$ dB		
			B_E/B_{RF}	B_ϕ/B_{RF}	$\Delta t \cdot B_{RF}$	B_E/B_{RF}	B_ϕ/B_{RF}	$\Delta t \cdot B_{RF}$
2 A	COHERENT	3.0	1.00	7.67	0.100	2.00	15.67	0.07
4 A	COHERENT	6.0	1.00	3.67	0.018	1.67	14.33	0.09
4 B	RANDOM	2.9	1.00	7.00	0.085	1.00	19.00	0.03
10 A	COHERENT	10.0	0.44	1.00	0.320	1.78	15.67	0.095
10 B	RANDOM	5.3	0.56	3.00	0.160	1.22	7.00	0.030

Table 1. Signals and requirements.

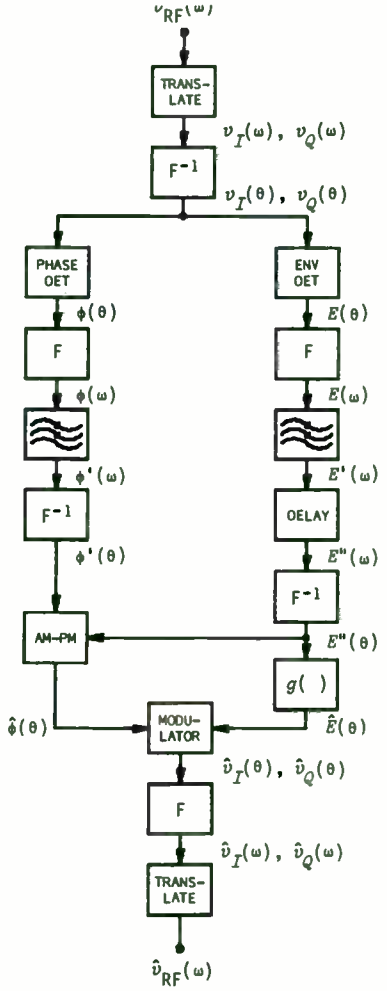


Figure 2. Computational procedure for multicarrier signals.

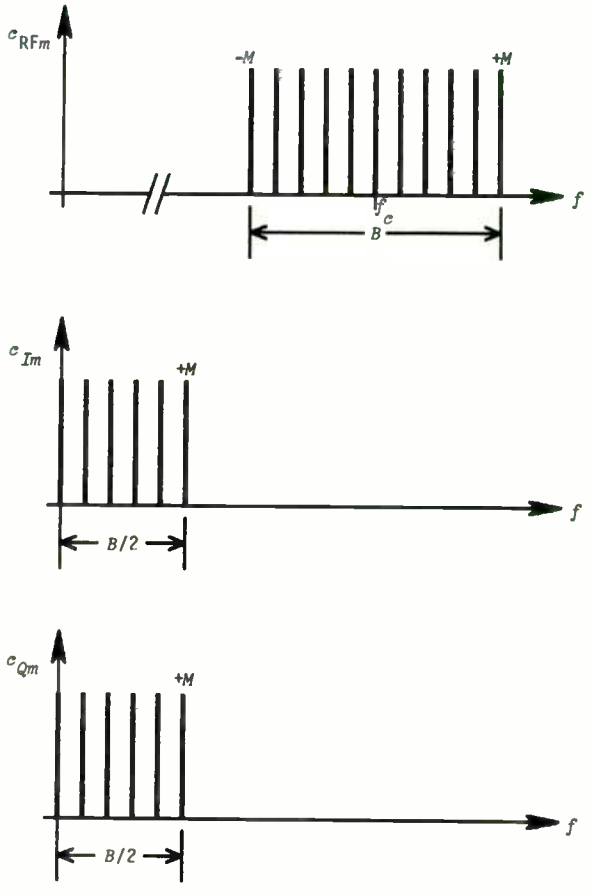


Figure 3. Fourier coefficients of RF, I, and Q signals.

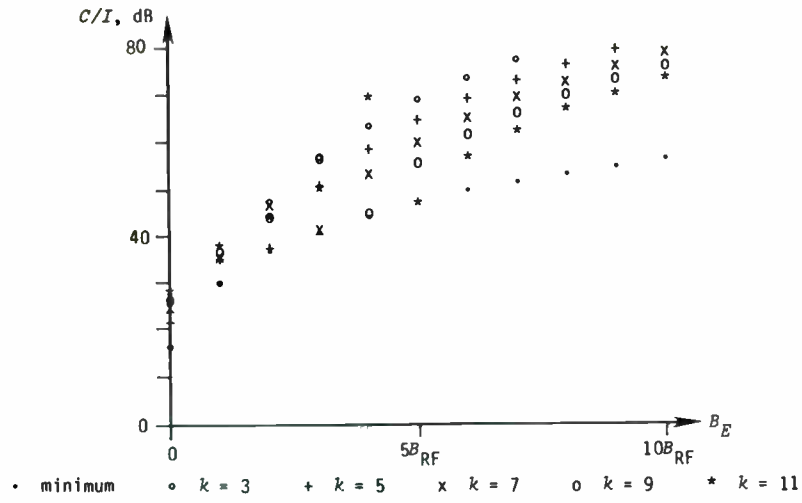


Figure 4. Effect of envelope bandwidth upon IMD products for two-tone signal.

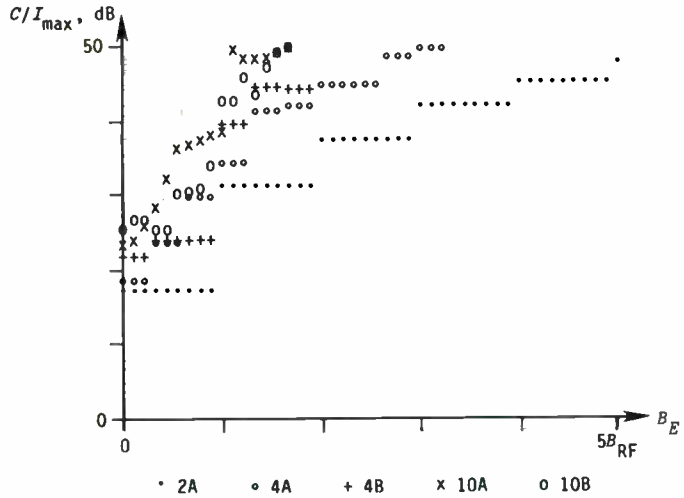


Figure 5. Effect of envelope bandwidth upon IMD for multicarrier signals.

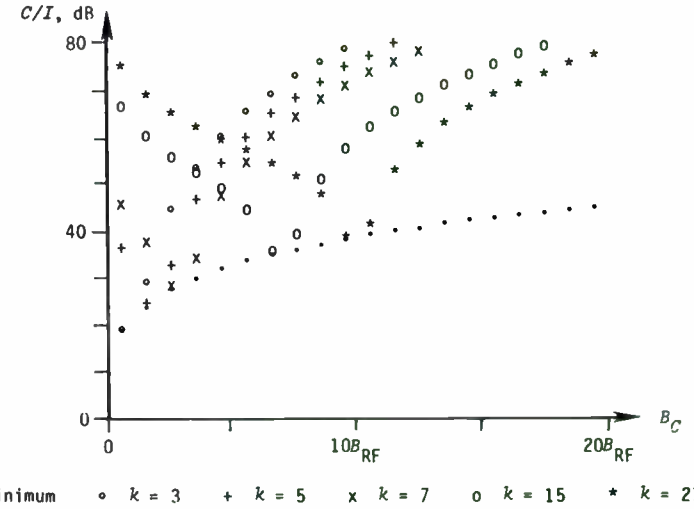


Figure 6. Effect of carrier bandwidth upon IMD products for two-tone signal.

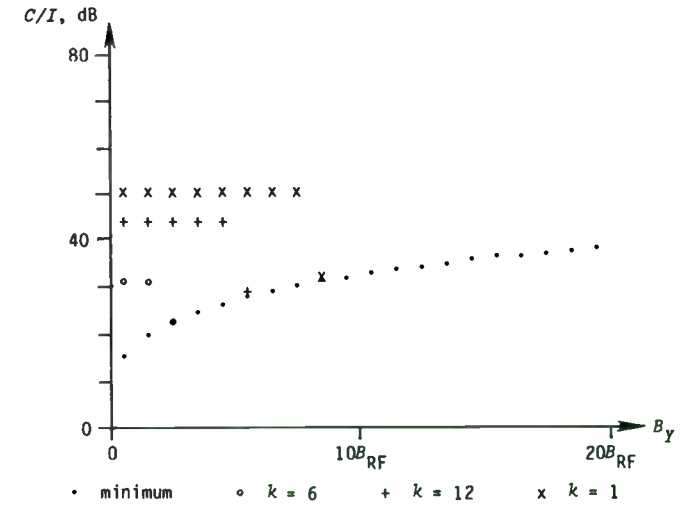


Figure 7. Effect of phase bandwidth upon IMD products for two-tone signal.

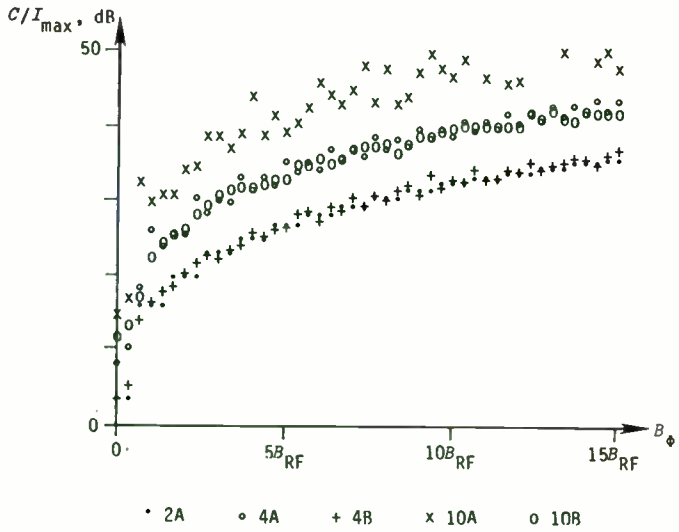


Figure 8. Effect of phase bandwidth upon IMD for multicarrier signals.

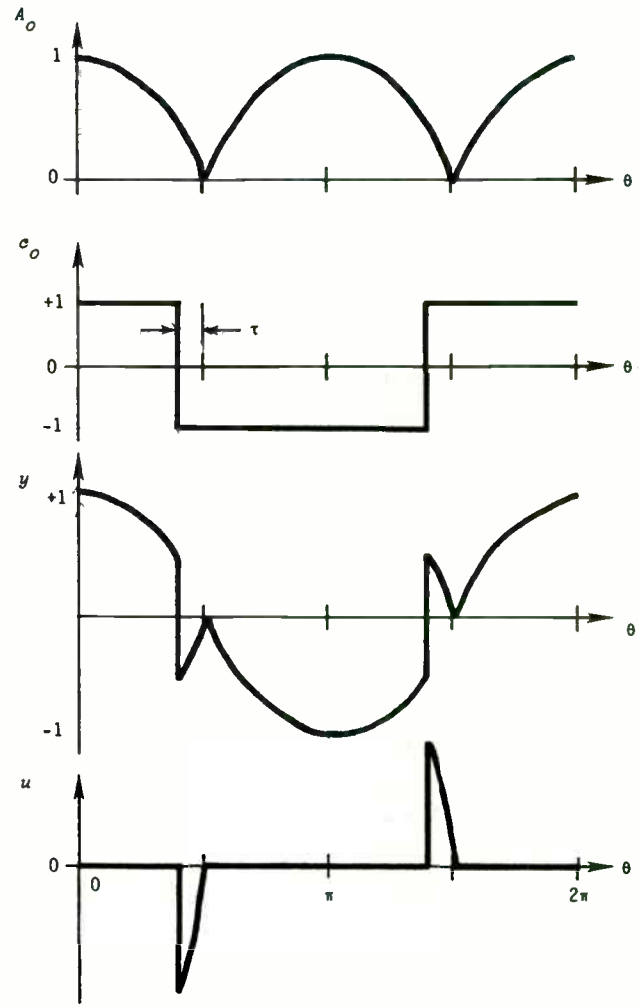


Figure 9. Model for unequal delay with two-tone signal.

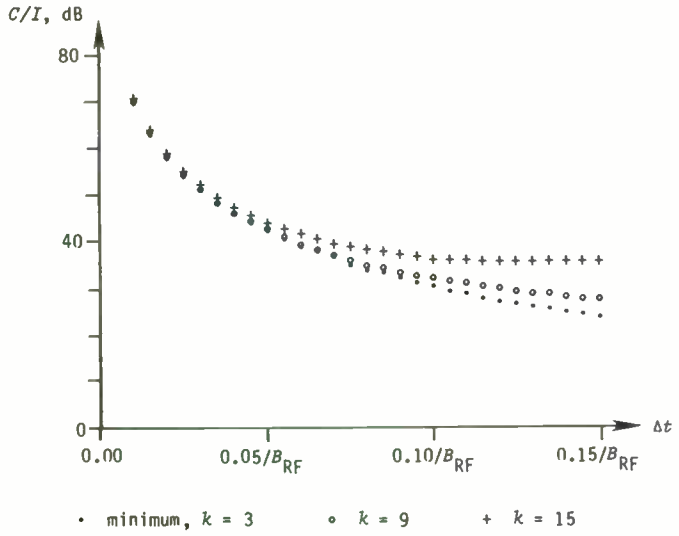


Figure 10. Effect of delay upon IMD products for two-tone signal.

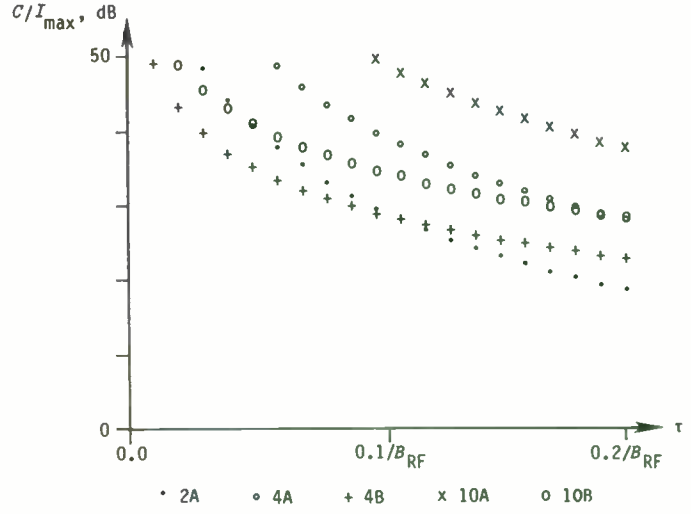


Figure 11. Effect of delay upon IMD for multicarrier signals.

Power MOSFETs in VHF Communications Amplifiers
A Practical Look

By
Daniel Peters
Falcon Communications

The bipolar transistor is the device of choice for most low and medium power RF communications power amplifiers in the VLF through UHF range; having replaced the vacuum tube in most applications some years back. The warm glow of the filament has been replaced by the black fins of the heat sink.

Another choice, the RF power MOSFET, is often overlooked. There are many reasons for this; some real and some imagined.

The discussion to follow will discuss comparisions between the RF power bipolar transistor and the RF power MOSFET (variously known as VMOS, TMOS, DMOS, etc.) and to point out some applications where you might find the RF power MOSFET superior.

Falcon Communications has been building both bipolar and MOSFET amplifiers in the 50-250 Watt range for 4 years; using devices made by several manufacturers.

We feel MOSFET devices deserve greater use in the industry. Unfortunately, some companies have produced very poor quality devices and have been guilty of terrible business practices. Thus, souring many people on MOSFETs.

This is a technical discussion and I won't dwell on this. However: a) Don't blame MOSFET technology for the lackings of some companies. There are companies making excellent devices. b) Before committing your company to production, based on a given company's devices, make doubly sure that company can perform.

MOSFET ADVANTAGES:

Advantages claimed for MOSFET, over bipolar, devices include:

a) Greater thermal stability - No "thermal runaway".

b) High input impedance.

c) Potentially lower cost.

d) Low noise.

e) Low spurious output.

f) Reduced feedback.

g) Reduced parameter changes.

h) Simplified circuit design.

i) Higher operating voltage.

j) Greater ruggedness.

Looking at each in a bit more detail:

a) THERMAL STABILITY - The current gain (beta) of a bipolar device increases with temperature. Thus, collector current can increase with temperature, resulting in higher temperatures, and yet higher currents. This can lead to the well known "thermal runaway". Careful circuit design will prevent this problem.

On the other hand, at high power levels the transconductance of a power FET decreases with temperature and higher temperature results in decreased current. This tends to be self stabilizing.

This stabilization also applies across the chip and serves to prevent destructive phenomena characteristic of the bipolar device: current hogging, hot spotting, and second breakdown. You generally won't find the "no-no" areas of the SOAR curves included with MOSFET data sheets.

SOAR is an acronym for Safe Operating ARea and refers to those regions of the collector curves where operation is not permitted. A bipolar power transistor has several areas where operation is not allowed, even though you are operating below the normal breakdown voltoage and within the power dissipation ratings, for fear of second breakdown.

As a practical matter, bipolar power devices are built with emitter ballasting resistors (small resistors in the emitter of each transistor cell) that greatly reduce thermal problems and any competent amplifier designer can design a bipolar amplifier that doesn't suffer from thermal runaway problems. Thus, thermal runaway problems are not sufficient reason to use MOSFET devices.

Emitter ballasting resistors come at a price. They reduce gain, increase parasitic capacitance and increase costs. MOSFET devices do not require ballasting resistors.

HIGHER INPUT IMPEDANCE - The power FET gate is essentially a MOS capacitor and at low frequencies this results in much higher input impedances than equivalent bipolar device. At VHF the ratio is less favorable. However, the FET looks capacitive to higher frequencies than an equivalent bipolar device; simplifying the input matching networks.

To raise the input impedance of bipolar devices, manufacturers build matching networks, consisting of internal MOS capacitors and the bonding wire inductance, into the transistors. This simplifies the design of external matching network to where they are not too much different than the same power level MOSFET. Although internal matching networks can raise the input impedance of bipolar power transistors to much more workable levels, they reduce the bandwidth over which the transistor is useful. Simple LC impedance matching networks have only a limited bandwidth over which they exhibit the desired impedance match.

MOSFET devices, designed for use through at least the low UHF range typically don't have internal matching networks.

POTENTIALLY LOWER COSTS - The lack of ballasting resistors and internal matching networks is sometimes cited as a reason why MOSFETs can be made more cheaply than bipolar devices.

However, a MOSFET device will use slightly larger chips than equivalent power bipolar device. In addition, to my knowledge, nobody is making MOSFET RF power devices in near the production quantities bipolar devices are being made.

Thus, for the immediate future at least, expect to pay considerably more for MOSFET devices.

LOW NOISE - Power FETs generate less broadband noise, 10 dB less typically, than a comparable bipolar device; partially due to the absence of a forward-biased junction and it's shot noise.

The lower noise is very beneficial for repeaters or any location where other equipment is operating in close proximity. This is a major advantage of MOSFET devices.

LOW SPURIOUS - The transfer characteristic of a typical FET displays no abrupt changes in shape meaning that, when biased for Class AB conditions, as in typical linear amplifier service, there will be lower high order intermodulation products than a similiarly operated bipolar transistor.

When comparing Falcon MOSFET amplifiers with their bipolar equivalents we find that 3rd order intermodulation products are roughly the same, while 5th order and higher are 10 dB, or more, better on the MOSFET amplifier. This can be significant in ACSB equipment. Figure 1 is a spectrum analyzer photo of the output of one of our MOSFET amplifiers, operating at 217 MHz., driven by an ACSB transmitter. Some of the spurious products are produced by the transmitter driving the amplifier.

The 10 dB difference is sometimes sufficient to allow a simple amplifier to do the job without resorting to complex, and difficult to adjust, feedback and feed-forward schemes.

REDUCED FEEDBACK - MOSFETs have reduced internal feedback. Also, the higher input impedance results in gate drive voltages higher than typical base drive voltages; with two benefits.

First, the voltage induced across the source inductance affects the input voltage of the FET proportionally less than the equivalent voltage across the emitter inductance in a bipolar transistor circuit. Because of the higher input impedances, the FET is operating at a higher input voltage (assuming the same gain) the the equivalent bipolar amplifier. Hence, any voltage developed across emitter/source inductance will have less effect.

Second, the effect of reverse transfer capacitance, already low in the FET, is further reduced by the lower voltage gain,

PARAMETER CHANGES - The transfer parameters of RF MOSFETs are quite insensitive to power level. This means smooth tuning of the input and output along with continuous input/output curves. Bipolar amplifiers often require retuning at each power level.

SIMPLIFIED CIRCUIT DESIGN - Gate leakage currents in the nanoampere or sub-nanoampere range result in essentially no bias power being used. Simple, low power bias circuits can be used.

The negative temperature coefficient allows the use of bias supplies without the complex temperature compensation schemes common to bipolar designs. In some higher power designs it is still desireable to use temperature compensation. However, the compensation is to reduce variations in circuit performance with temperature, not to protect the devices.

HIGHER OPERATING VOLTAGE - Higher voltage FETs now reaching the market present exciting possibilities.

As supply voltage increases, current decreases and impedance levels in the output matching networks increase. At a fixed power level doubling the voltage quadruples the impedance.

Higher impedance reduces the effect of parasitic inductance elements and makes the internal leads of the transistor a less critical part of the matching networks.

Capacitor values in matching networks become more reasonable and bypassing is easier.

Finally, increased impedance allows easier broadband design.

This might be a good opportunity for me to get on one of my favorite "soap-boxes".

There is a common misconception that the output network of a power amplifier is designed to "match" the load to the output impedance of the transistor; or to us "RF types", present a "complex conjucate match". Not true! And not true in most power work. If you build a toaster, you select the resistance of the

heater to develop the desired power at the available voltage. You do not select a resistance to "match" the source impedance of the power source. You would sure brown your toast in a hurry if you did.

The "matching" myth arises out of extending the "equivalent circuits" concept beyond where it should be extended. We are told that, no matter how complex, any source of power (assuming it is linear and in this case a transistor) can be represnted by a voltage source with a series impedance.

Next, we are told; "To obtain maximum power out of this circuit, we must supply a load with a real (resistive) part equal to the real part of the source impedance and an imaginary (reactive) part equal in magnitude and opposite in sign to the imaginary part of the source impedance".

However, and this is the key point, we don't always want to obtain maximum power. And, as strange as it may sound, in RF and audio power amplifiers we seldom want to obtain "maximum" power out. Remember the toaster.

Those familiar with "Pi" network vacuum tube amplifiers know you adjust for "rated" power output. If you adjust for "maximum" output the tube life can be very short indeed. You do not adjust the impedance of the "Pi" network to "match" the impedance of the tube, you adjust it to an impedance that results in the desired power output. There can be a big difference between "maximum" power output and maximum output without destruction.

Look at it another way. At the load for "maximum" power output, 1/2 the generator voltage will appear across the source resistor and 1/2 across the load resistor. The highest possible efficiency would be 50% and it would generally run much lower. Yet, we know power amplifiers consistently run better than 50% efficiency. Thus, the designers must not be "matching" the output impedance of the device.

GREATER RUGGEDNESS - FETs generally have higher breakdown voltage ratings than their bipolar counterparts. This, combined with their favorable temperature characteristics, can result in amplifiers with the potential for putting up with considerable abuse.

Though there are claims that FETs are more reliable than bipolar transistors, and there is a certain amount of logic to this, our early experience was just the opposite. Our MOSFET amplifier reliability was terrible.

We found the MOSFETs had a design flaw. The gate bonding wires were aluminum and of inadequate size and melted. It took over a year to shame the manufacturer into responding, at which point they increased the wire size; only to create another problem. The wires weren't properly bonded and broke loose. We gave up and changed vendors.

The new devices have proven to be very reliable. However, our bipolar amplifiers are also very reliable and we don't have enough data to draw conclusions about relative reliability.

The point of the foregoing story is to illustrate that, as is true of many components, design details and manufacturing quality often have more to do with reliability than the technology used.

A PRACTICAL AMPLIFIER:

Let's compare the schematic of a typical bipolar amplifier with that of a typical MOSFET amplifier.

Figure 2 is a schematic of a bipolar amplifier we make in several versions from 100 to 200 Watts; covering frequencies from 50 to 250 MHz. Figure 3 is the schematic of the MOSFET version. The particular schematics refer to 100 Watt 220 MHz. amplifiers. Please note that all variations of this basic amplifier use the same bare P/C board.

The most popular versions of these two amplifiers are sold for Amateur repeaters and are rated for 100% duty cycle on FM.

First, let's look at some similarities. Both amplifiers use two devices in parallel. They run on 13.8 Volts and there are no commercially available devices that can deliver 100 Watts at 220 MHz. with a single device at this voltage.

Both amplifiers have the same output matching network and low-pass filter. The matching network is essentially 3 "L" sections.

Likewise, the input matching networks are also very similar, except for C9 and C10, in the MOSFET amplifier. The primary purpose of these is dc blocking.

As might be expected, the first real difference is in the biasing. The bipolar version uses the forward drop across D4 to produce the bias. D4 operates as both a regulator and temperature sensor. D4 receives it's current through RFC4, the relay contacts, RFC1 and R2//R5.

The rectifying action of the base-emitter diodes upon the input drive signal acts to change the dc bias current; resulting in a reduction in gain at higher drive levels. To reduce this effect, a large current is pumped through D4. R2//R5 is as small as 25 Ohms in some of our amplifiers. The bias circuit uses about 8 Watts dc.

Individual transistor bias adjustment is inconvenient at this power level and we use dc beta matched transistors to keep a reasonable match in quiescent collector currents. We use no further transistor matching at this power level.

Using a diode as a regulator provides reasonable tracking between the desired and the actual bias voltage with temperature and we don't have temperature problems.

If we were to build a higher power bipolar amplifier we would consider a more complex bias scheme. At this power level we don't feel the added cost is worthwhile.

The gate of a MOSFET, being voltage operated requires a different type of bias.

Our regulator in this case is a zener diode (D4). The zener is fed through R2, the relay contacts, and RFC1. The BIAS source voltage is the 13.8 Volt supply fed through a transistor switch which is controlled by the DETECTOR. The reason for running the bias through the relay contacts is to assure that the amplifier is not biased up unless the relay is in the proper position. Even though the circuits on the CONTROL BOARD normally accomplish this, our customers sometimes modify the CONTROL BOARD to suit their individual needs and running the bias through the relay is a protection in case they goof.

For all practical purposes, the dc current drawn by the MOSFET gates is zero and the impedance levels in the bias circuits can be high; thus the power is minimal. We elected to use individual bias potentiometers for each transistor. Doing so has eliminated the need to match transistors. We do use devices from the same batch in a given amplifier.

A precaution to be observed when paralleling MOSFETs (or using them in push-pull) is to make sure that the failure of one FET does not cause the failure of the other one. One failure mode in a MOSFET is a gate-drain short. This raises the gate voltage to the drain voltage and if this can feed through the bias network you can draw excessive current in the other device and destroy it. The bias scheme shown is one way to prevent this.

This amplifier was temperature tested over the range of -40 to +60 degrees Centigrade. The room temperature quiescent current was set at 6 Amperes, a value we find to be the best compromise between gain, linearity and efficiency.

The quiescent current was slightly temperature dependent. As expected, the current is higher at lower temperatures. The total variation was 0.4 Amperes across the entire range.

The variations in RF parameters were so slight as to be difficult to measure. The gain appeared to be a few tenths of a dB higher at the lower temperatures. Current drawn at the 100 Watt point did not vary significantly with temperature. In all cases the spectrum was clean, with the second harmonic being over 65 dB down and higher harmonics being hard to measure on the equipment used.

Going back to the two schematics, we see another difference between the bipolar and MOSFET amplifiers, the feedback networks around the bipolar transistors (RFC2, C9, C10, R3 and RFC3, C12, C13, R4). Bipolar transistor gain increases markedly as frequency decreases and VHF bipolar amplifiers are very prone to oscillate at HF frequencies. Careful low frequency by-passing and the feedback network shown generally cures the problem. MOSFET transistors do not have this marked gain slope.

WHAT'S AVAILABLE:

By now, I don't think it is a secret that I prefer using MOSFETs over bipolar devices in power amplifiers. However, this preference doesn,t do me any good unless there are commercially produced devices readily available.

I was tempted not to include a summary of what's available becuse no matter how careful I am, I am going to miss somebody's latest magic device and someone else is going to introduce something here at the show that is going to make the list invalid. However, let's give it a try.

1 3 . 8 V O L T D E V I C E S :

For 13.8 Volt applications the choice of MOSFET vendors is very limited. In fact, there is only one that we will consider buying from. If your need is a mobile amplifier that will boost a 25 Watt mobile High Band FM transceiver to 150 Watts, go bipolar. They work fine and MOSFETs cost more than twice as much.

For High-Band communications applications with a 13.8 Volt supply both bipolar and MOSFET devices are available at up to 80 Watts output per device with about 10 dB gain.

If your application is a 13.8 Volt repeater amplifier, where low broadband noise is important, go for the MOSFET.

At 220 MHz. the MOSFET will still give you 10 dB of gain in a 60 Watt device while the bipolar will be down to 7 db.

There are no 13.8 Volt, 60 Watt, MOSFET devices specified for 400 MHz and above. You can get a 60 Watt 4+ dB device in bipolar for both the 407-512 MHz. and the 800-960 MHz. bands.

At 50 MHz. the bipolar and MOSFET devices compete for power output and gain.

At HF the bipolar devices have the edge, giving higher gain, more power and lower cost. 100 Watts per device is available.

HIGHER VOLTAGE DEVICES:

For 28 and 50 Volt supplies a much larger variety of devices are available.

At 30 MHz., using a 50 Volt supply, you can obtain 600 Watts per device with 10 dB of gain in bipolar and 17 dB in MOSFET.

Push-pull MOSFET devices are available producing 300 Watts up through 200 MHz. At 500 MHz., 100 Watt devices are available in both bipolar and MOSFET with the MOSFET having higher gain.

SUMMARY:

MOSFET power devices have been around for a number of years now and no longer can be ignored as "new unproven technology".

In some applications they offer distinct advantages and you should at least consider them for your next design.

For those of you who wish to explore the topic of RF power amplifiers further, I have included list of references.

Thank you for your time.

REFERENCES

1. R. Brand, P. Ledger, "Slicing S Parameters", RF Design Sept/Oct 1982 pp 12-21.

2. G. Gonzalez, "Microwave Transistor Amplifiers", Prentice Hall, 1984, Chapters 1 and 2.

3. C. Bowick, "RF Circuit Design", Howard W. Sams, 1982.

4. "R. F. Small Signal Design Using 2-Port Parameters" Motorola Application Note, AN-215.

5. H. Krauss, C. Bostian, F. Raab, "Solid State Radio Engineering", John Wiley & Sons, 1980, Chapters 4, 12 and 13.

6. A. Burwasser, "TI-59 Program Computes Values For 14 Matching Networks", RF Design, Nov/Dec 1983, pp 12-31

7. J. Hardy, "High Frequency Circuit Design", Reston Publishing Company (Prentice Hall) 1979, Chapter 5.

8. A. Przedpelski, "Special Low Pass Filters" RF Design Sept/Oct 1982 pp 48-54.

9. S. Houng, "A practical Approach to Amplifier Matching", Microwaves & RF, July 1985

10. D. Peters, R. Larken, "Power FETs: Trend for VHF Amplifiers", Ham Radio, Jan 1984 pp 12-23

11. B. Becciolini, "Impedance Matching Networks Applied to R-F Power Transistors" Motorola Application Note AN-721.

12. F. Davis, "Matching Network Designs With Computer Solutions", Motorola Application Note AN-267.

13 A. LaPenn, "BASIC Program Computes Values For 14 Matching Networks". RF Design, April 1985, pp 44-47.

14. H. Granberg, "A Simplified Approach to VHF Power Amplifier Design", Motorola Application Note AN-791.

15. W. Orr, "Radio Handbook", Howard W. Sams, 1981, Chapter 7.

16. H. Granberg, "Broadband Transformer and Power Combining Techniques for RF", Motorola Application Note AN-749.

17. R. La Rosa, "Hybrid Coupled Amps: Can They Weather a Mismatch?", Microwaves, February 1975, pp 44-49.

18. H. Swanson, B. Tekniepe, "A 100-Watt PEP 420-450 MHz Push-Pull Linear Amplifier" Motorola Engineering Bulletin EB-67.

19. D. Peters, The Basics od R. F. Power Amplifier Design", Proceedings of RF Technology Expo 86, pp 33-48.

20. D. Peters, "Some Common Myths About RF Power Amplifiers.", Mobile Radio Technology, December 1986, pp45-51

21. D. Peters, "R. F. Power Amplifiers.", Mobile Radio Technology, July, August 1987.

- REFERENCES END -

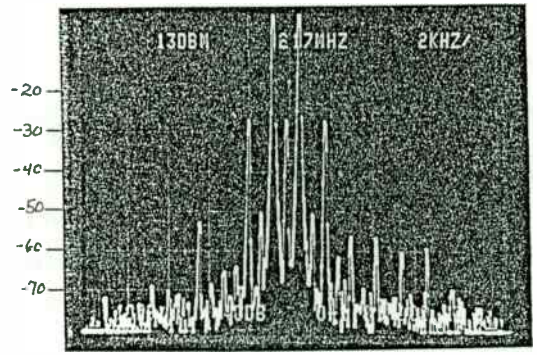


Figure 1

Falcon Model 8252 @ 75 W. PEP. Driver is Aerotron Pioneer 1000 ACSB amplifier

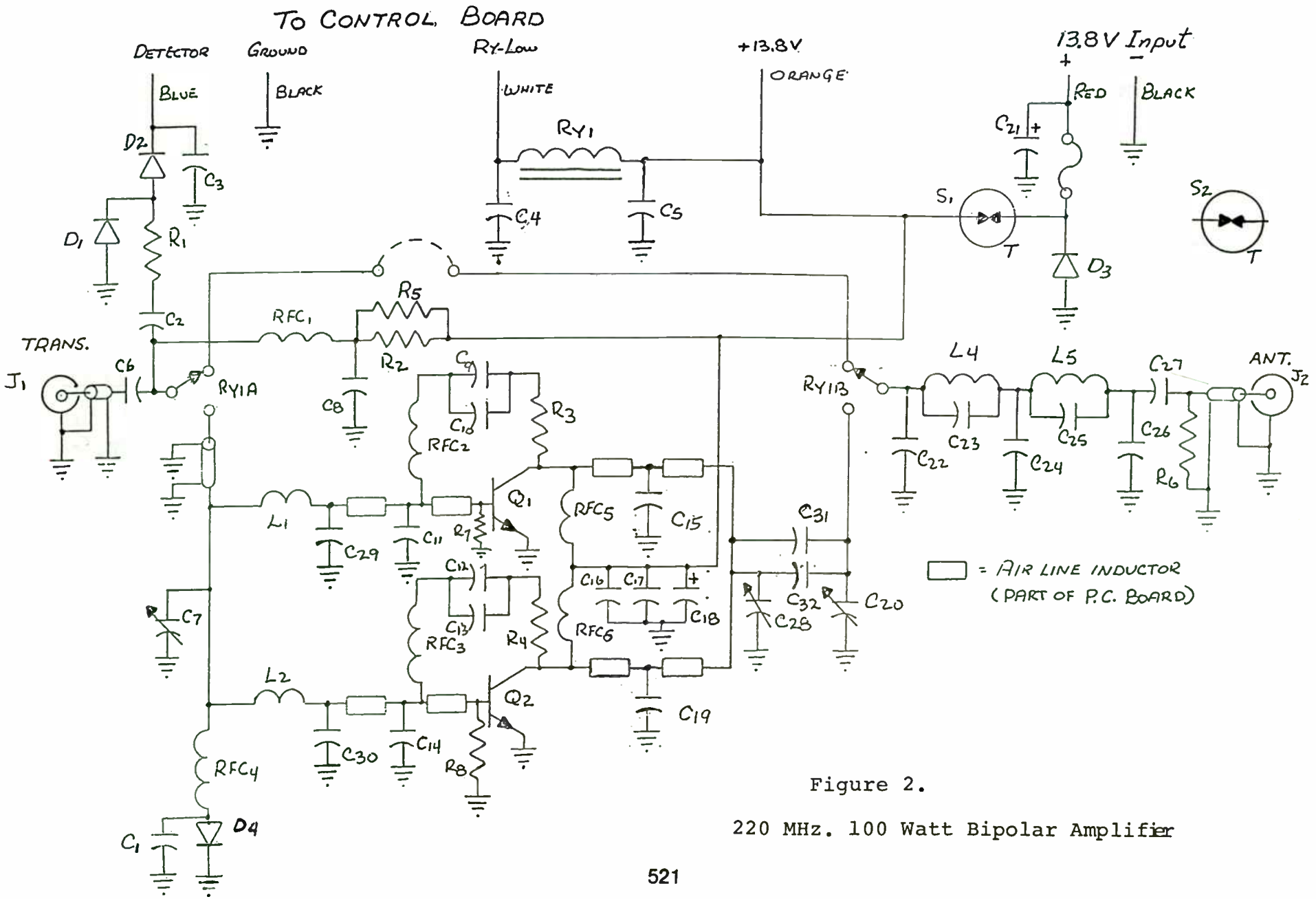


Figure 2.

220 MHz. 100 Watt Bipolar Amplifier

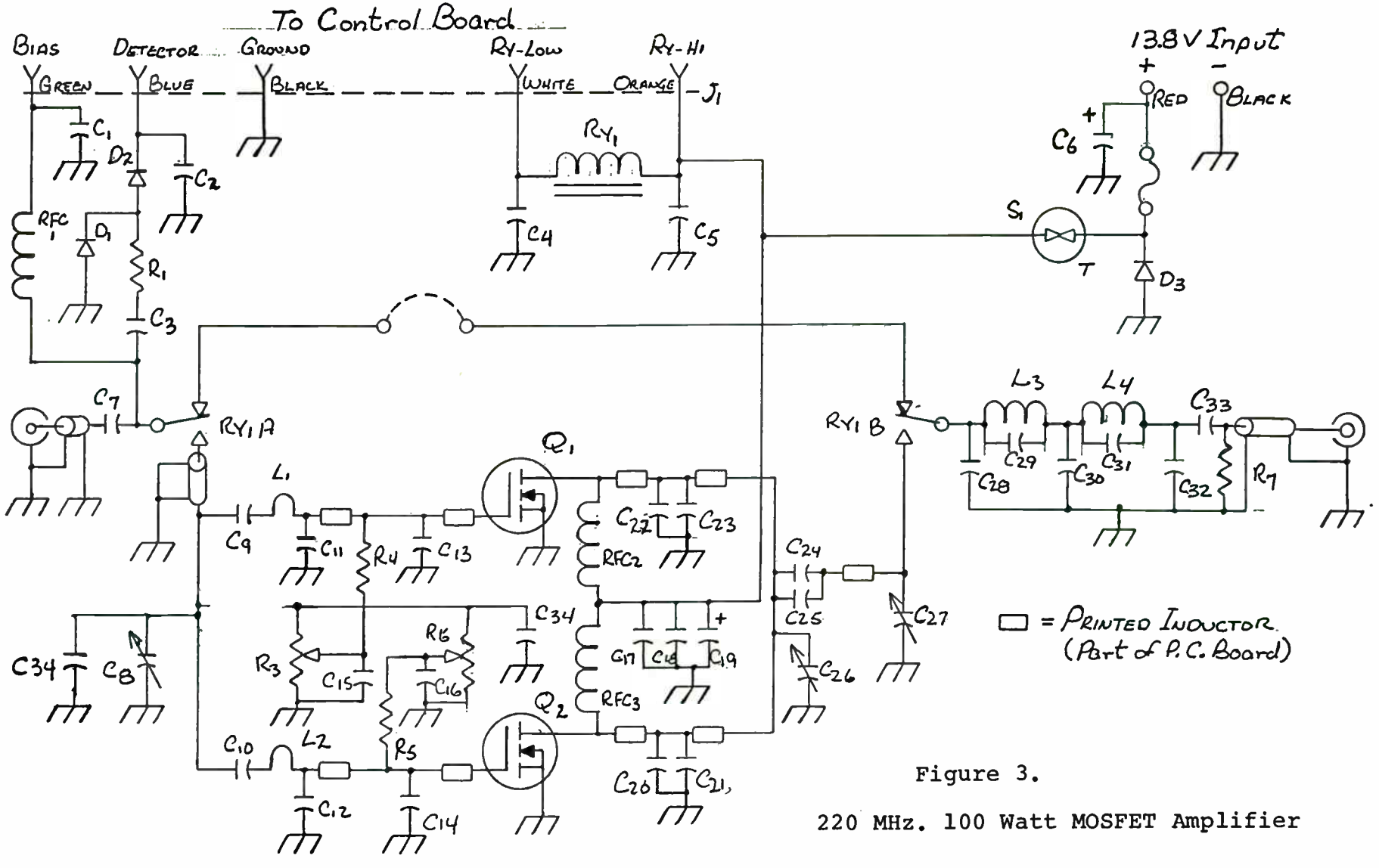


Figure 3.

220 MHz. 100 Watt MOSFET Amplifier

Control Of Spurious Signals In Direct Digital Synthesizers

by

Earl W. McCune Jr.

President

Digital RF Solutions Corp.

Abstract: The Achilles' heel of direct digital synthesis (DDS) has been the concurrent generation of undesired spurious signals. Recent research has shown that these spurious signals can be significantly suppressed, even further than conventional wisdom has held. Low spurious levels on direct DDS outputs has allowed this technique to move into high quality signal generation designs.

1 DDS Concept

1.1 Overview and history

Direct digital synthesis (DDS) have held much promise since the initial concept was proposed in the early 1970s. The possibility of direct, numeric control of frequency with extremely fine resolution and crystal stability has been the goal of many a radio designer.

Acceptance of DDS techniques has been severely slowed by these system's concurrent generation of unwanted spurious signals with significant amplitude. A 40-50 dB dynamic range is not acceptable for a great many communications applications.

Fortunately technology has progressed over the years, and this problem of spurious signals has been overcome. Spurious free dynamic ranges of 60-80 dB are now directly achievable, which allows DDS to move into high quality system design.

For a quick review of DDS techniques, refer to Figure 1. This shows a block diagram of a basic direct digital synthesizer. In effect this is a highly specialized calculator, optimized to repeatedly calculate amplitude values of periodic waveforms, such as sine waves. From a stable (clock) oscillator, a tuning number controls the rate at which a phase accumulator advances the phase of the output signal. This output signal phase information is converted to amplitude information by the waveform map, typically a wave shape stored as a table of numbers in a read-only-memory. These amplitude numbers are finally converted to the final analog signal in the digital to analog converter (DAC).

1.2 Quantized nature generates distortion

This digital representation of the output waveform is quantized by its very nature. Because only discrete, fixed values of amplitude are allowed in the output waveform, there are distortions from the ideal sinewave that to be expected. The questions then become: 1) how to predict these distortions, and 2) what can be done to minimize them.

2 Quantizing Effects

2.1 Sampling Theorem predictions

To get the first handle on predicting these distortions, first examine the process of sampling a sinewave. Shown in Figure 2, there are three basic steps from a pure sinewave to the sampled version as would be seen at the DDS DAC output.

The process begins from a pure sinewave in Figure 2a. A pure sinewave corresponds to a two-component spectrum centered around zero Hertz. Signal energy is tightly concentrated around the frequency of the sinewave. When the sinewave is repeatedly sampled, the pictures of Figure 2b result.

Amplitude samples are present only at the times of the sampling. At all other times the output waveform is zero. The effect on the spectrum is to replicate the original about the sampling frequency, and all its harmonics. This sampling distortion therefore does not generate harmonic distortion of the original sinewave, but rather arithmetic shifts of the original spectrum spaced according to the sampling frequency.

When the sinewave samples are held constant over a sample interval, the resulting waveform begins to return to sinusoidal shape, as shown in Figure 2c. This should correspond to less distortion, which the spectrum confirms. Signals are still present at their shifted positions, but amplitudes are suppressed for the shifted spectra. This suppression is effected by an envelope that nulls at the sampling frequency and all its harmonics.

This waveform is the ideal output of a direct digital synthesizer. The distortion from the sampling process is not harmonic, but rather the generation of image frequencies that move above and below the clock frequency, in addition to the desired fundamental. Theoretically, this is what will be present at the output of any DDS system. It should be noted that there are no spurious responses in this model. Likewise, the closer a DDS design can approach the ideal model, eventually it will be expected that non-ideal spurs will vanish.

2.2 Error types

2.2.1 Amplitude

Any real world system will possess some errors in its operation. The primary error sources of interest in DDS design are amplitude and phase mistakes in any step. Figure 3 shows the definition of a step amplitude error.

2.2.2 Phase

Similarly, Figure 4 shows the definition of a phase step error. In any actual system, both types of error are present to some degree. These errors can arise from several possible sources, which are discussed in detail below.

3 Waveform Map Effects

3.1 Amplitude Quantization

The waveform map translates quantized phase information into quantized amplitude numbers. As in any quantized system, errors occur whenever the ideal value of the sampled waveform does not exactly match a quantization value.

To get a conceptual understanding of this quantizing error, refer to Figure 5. This figure shows the effect on a sinewave subject to 8 phase position and 8 amplitude quantization. Errors are determined in amplitude relative to precise phase positions. In accordance with the amplitude error definition from Figure 3, the output errors are shown as crosshatched rectangles.

3.2 Phase Quantization

Error can also be measured with respect to precise amplitude sampling, which examines phase quantization. This approach is shown in Figure 6. The major realization from this presentation is that fixing amplitudes tends to generate far greater error than fixing phase quantization. This explains the historic tendency to fix phase sampling and take the resulting amplitude quantization. The fact that a fixed phase sampled approach is easier to realize further entrenches this method.

3.3 'Conventional Wisdom' discussion

An axiom exists in the DDS community that relates to synthesizer performance due to these quantizing errors. Stated simply, in frequency domain (spectrum) terms,

In a direct digital synthesizer, the non-harmonic spurs due to signal quantization will be suppressed 6 dB below the carrier for every bit used in the amplitude quantization.

This is the '6 dB per DAC bit' spurious signal generation 'rule'. The argument for its validity is amplitude quantization based: every additional bit applied in the amplitude quantization will improve the quantization quality by a factor of 2. In the dB world, this corresponds to almost exactly an additional 6 dB. Hence, the establishment of this 'rule' as a performance limit to be strived for.

For example, a DDS design using an 8 bit DAC should expect non-harmonic spurs 6*8 = 48 dB below the carrier. To achieve 60 dB of spurious suppression, 60/6 = 10 bits would be required of the DAC. The design efforts clearly pointed to the need for high speed, high resolution DACs for any hope of achieving high quality digital synthesizers.

Before further discussion on this 6 dB/bit axiom, examination of the effects brought by DAC non-idealities in the analog conversion are necessary.

4 Analog Conversion Effects

4.1 DAC as a multiplier

Any DAC (digital to analog converter) develops an analog output signal proportional to its input number. The DAC is therefore a multiplier (see Figure 8), according to

Analog Output = Input Number * DC Reference Level.

As such, all conventional theory regarding multipliers and effects of their non-idealities applies. For example, imbalance in the transfer function should result in even order output distortion, and cross modulation of all input signals will occur through high order nonlinear terms. Indeed, all these effects are seen in the DDS DAC output signal.

Physical realization of these various non-linearities is special to the study of DACs. Best performance will be achieved with a DDS design using a DAC exhibiting fewest non-linearities. Identification of the major non-linear processes and their solution is covered below.

4.2 Time Skew

Time skew is far and away the non-linearity of greatest magnitude. As shown in Figure 9, time skew is the result of each bit of the input number arriving at the DAC at a slightly different time. If this skew is large compared to the DAC settling time, then the DAC output will respond to the effectively changing input number and also vary until the input stabilizes. This generates extra energy in the output, which manifests itself as non-harmonic spurs.

Is a 1 nanosecond time skew important to a DAC with 100 nanosecond settling time? Using a worst case linear approximation, 1ns / 100ns = .01, which is -40 dB error energy. In actuality the error energy is lower due to the lower contributions from errors near waveform peaks and valleys. But it is still clear that this relatively small skew may well be dominant in an 8 bit DAC system.

Latching the data before application to the DAC will remove any system contribution to data time skew. Modern monolithic multi-bit digital latches exhibit timing skews under 100 picoseconds across each device. DAC internal time skews are a function of DAC design and manufacture, and subject to individual DAC type evaluation.

4.3 Integral Nonlinearity

Integral non-linearity is the overall variation of the DAC transfer function from the ideal straight line, and is depicted in Figure 10. Since integral non-linearity tends to be a smooth curve, the primary spur contribution is output signal harmonics. This harmonic distortion shows that the non-linearity affects each cycle equally, resulting in periodic distortion at the same rate as the primary output signal.

4.4 Differential Nonlinearity

Differential non-linearity is a small scale measurement of the transfer function variations from a straight line. It is essentially a bit by it output linearity measurement. Differential non-linearities by their nature are very fast moving with respect to the primary output signal.

These rapid variations are effectively high order non-linear terms, and therefore manifest themselves as cross-modulations and high order harmonics. Cross modulations may have either sign relationship with the primary output signal, and so can include 'cross over' signals. A cross over is where a spurious signal moves in opposite direction as the primary output and at some point crosses over, or shares, the same frequency.

The existence of cross over signals eliminates any chance of using filters to remove the spurs. Therefore once generated, cross over spurs are there to stay. Their only cure is to not be generated, which means using DACs with low differential nonlinearity.

4.5 'Glitch Energy'

Glitch energy is actually a misnomer, but has become ingrained in the DAC vernacular. Shown in Figure 12, this term refers to the 'spike' observed on the DAC analog output while a transition settles. The traditional unit of measure is the area of the circumscribed rectangle around the main pulse, and has units of volt-seconds, which is not an energy measure.

Two factors are dominant in the 'glitch pulse'. First is time skew, which has already been covered. It should be easy to see that the varying output amplitude due to time skew will have characteristics similar to the glitch pulse.

Second is the analog settling time of a linear system. Since at this point the system is no longer digital but analog and linear (hopefully), reactions of this linear system to the step input provided by the DAC switches must be some type of damped sinusoid. In the traditional analog tradeoff of settling time versus edge speed, a slightly underdamped response is usually chosen. Such is the glitch pulse.

Since this settling time/edge speed tradeoff is closely related to system bandwidth, lower settling times are achieved with wider bandwidths. Therefore, higher speed DACs with no time skew should also exhibit smaller glitch pulses.

4.6 Output stage balance

Balance in the output stage calls for symmetry in all equally sized and oppositely signed transitions. For a transition from one output level to another, the return transition should be its mirror image. Such performance is outlined in Figure 13.

Differences in transition times models as imbalance in the multiplier. From before, this imbalance is expected to produce even order distortion terms. Actual practice shows this to produce even order harmonics of the primary output signal, which can have appreciable amplitude (-25 dBc).

There are two immediate solutions. First is to find a DAC that exhibits excellent output balance. Second, a DAC exhibiting this distortion can be improved by including a well balanced sample and hold on its output. Timing the sample and hold to always wait for the longest settling time will clean up the transition imbalance. With this type of non-linear filtering, no net loss of system speed is realized. The main cost is increased complexity.

4.7 Output stage compliance

Many high speed DACs are of the current output type. Usual translation of this current to a voltage is done with a simple resistor to ground of the current source. This method varies the voltage across the current source, and with sufficient amplitude will change the characteristics of the source.

Output compliance limits appear as saturation effects, and therefore can be modeled as an integral non-linearity. The result is harmonic distortion of the primary signal. Avoidance of output stage compliance is primarily a matter of proper design, based on the recommendations of the DAC manufacturer.

4.8 Sample time SLOWING improves performance

Another popular conception on DDS design is that more samples for each output cycle is always better. If the sample rate is increased, there will be more samples per cycle for any particular output frequency. This will be a better approximation of the desired signal, and so should have lower distortion.

In general this is not the case with existing DACs, as Figure 14 and the earlier discussion of the sampling theorem will show. The sampling theorem demonstrated that with a good enough filter, only two samples are necessary to accurately define a sinewave. However, this assumes that transition times from one sample to another are zero. In any real DAC this is not possible.

Idealities can be approximated, and this is shown in Figure 14. By slowing the DDS sampling clock, as long as the two sample per cycle theorem requirement is maintained, the transition time effectively tends toward zero. What is actually happening is that the DAC settling time waveform error is being averaged over more time, reducing its overall contribution.

With this in mind, a counter-tendency is realized: fewer samples per cycle with a slower sample clock and a relatively fast DAC provides a close approximation to the sampling theorem ideality. This will reduce net distortion in the output signal.

5 Test Equipment Effects

5.1 Mask actual performance of synthesizer

One other major point needs to be discussed in the development of high performance digital synthesizers, which also holds for any synthesizer development: test equipment performance. If the test equipment exhibits non-linear behavior, as all do, then under some circumstances spurious responses will be internally 'created'. What is displayed on the screen may not be a truly faithful representation of what the synthesizer is doing.

It is important to keep this limitation in mind, especially in the development of digital synthesizers which by nature produce many output signals. Spectrum analyzer non-linearities can readily mix these signals together, resulting in an artificial mess on the display. This mask of actual synthesizer performance can be very aggravating in pursuit of those 'final dB'. Refer to Figure 15.

5.2 Second order intercept

For example, consider the two non-linearity specifications provided by spectrum analyzer manufacturers: second and third order intercept points. With no input attenuation, typical specifications for high quality bench analyzers are

second order intercept	+30 dBm
third order intercept	0 dBm

For the purposes of discussion, assume that input attenuation required to bring the nominal synthesizer output to the analyzer's specification amplitude, typically -30 or -40 dBm, is in place.

For second order distortion to be -60 dBc, then the input signal power must be half this specification below the second order intercept. This is +30 - 60/2 = 0 dBm. In other words, the spectrum analyzer mixer can see its specification amplitude and still meet this specification. If, however, a -80 dBc specification is required, then the input power must not exceed +30 - 80/2 = -10 dBm. Attenuators are therefore an essential part of high quality synthesizer proof of performance.

5.3 Third order intercept

In consideration of third order distortion, the rule changes to the greatest input signal amplitude must be below the intercept point by one third of the desired suppression. As before, for third order distortion to be 60 dB below the peak signal power, 0 - 60/3 = -20 dBm. This is even lower than the high quality second order requirement.

For -80 dBc performance, the input power must not exceed 0 - 80/3 = -27 dBm. Apparently, high quality signal measurements are still subject to low level signals, and cannot be enhanced by increasing signal to noise ratio with more power.

In these instances the analyzer internal distortion products will be low enough to make the traditional method of spur determination difficult. Assuming enough dynamic range is available, input attenuators provide the only demonstrable proof of internal (fake, the spectrum analyzer's problem) or external (real, the DDS designer's problem) spur origin.

6 Currently Available Performance Levels

Currently available performance levels from Direct Digital Synthesizers are shown in Figure 16. These measurements were taken from the DRFS-PCHPE demonstrator board. As the number of DAC bits increases the signal distortion decreases as expected, but performance is consistently below the 6 dB per DAC bit rule. Furthermore, the slope is not 6dB/bit, but closer to 8dB/bit.

The total reason for this result is not yet clear, but several points are determined. Primarily, phase errors have been shown to be significant contributors to the final distortion. By control of roundoff occurrence, phase errors have been minimized. This change in phase performance has shown a marked decrease in output distortion.

7 Conclusion

Direct Digital Synthesis (DDS) is now practical for use in high quality communications applications. DDS has been historically plagued with a high discrete spurious signal content, making its application in these applications impractical. Processes have now been identified and described that generate these spurious signals, along with several suggested means to eliminate or reduce their occurrence. And finally, current DDS performance levels have been presented.

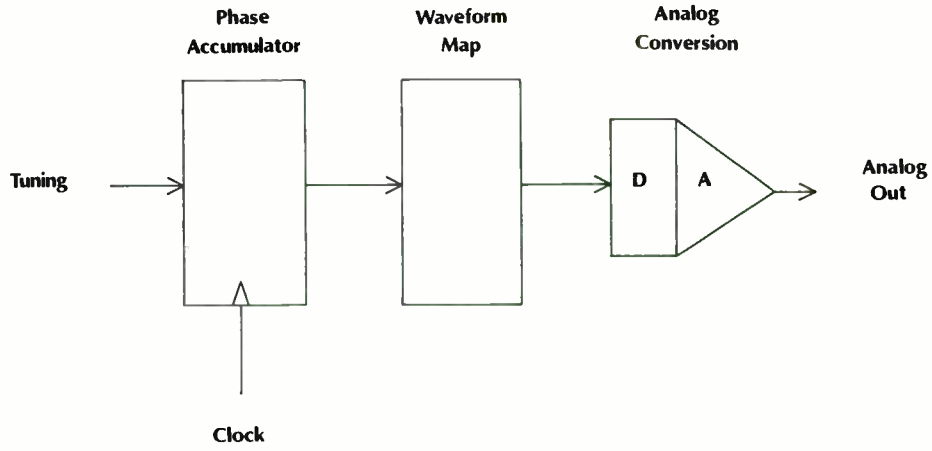


Figure 1. Basic Digital Oscillator

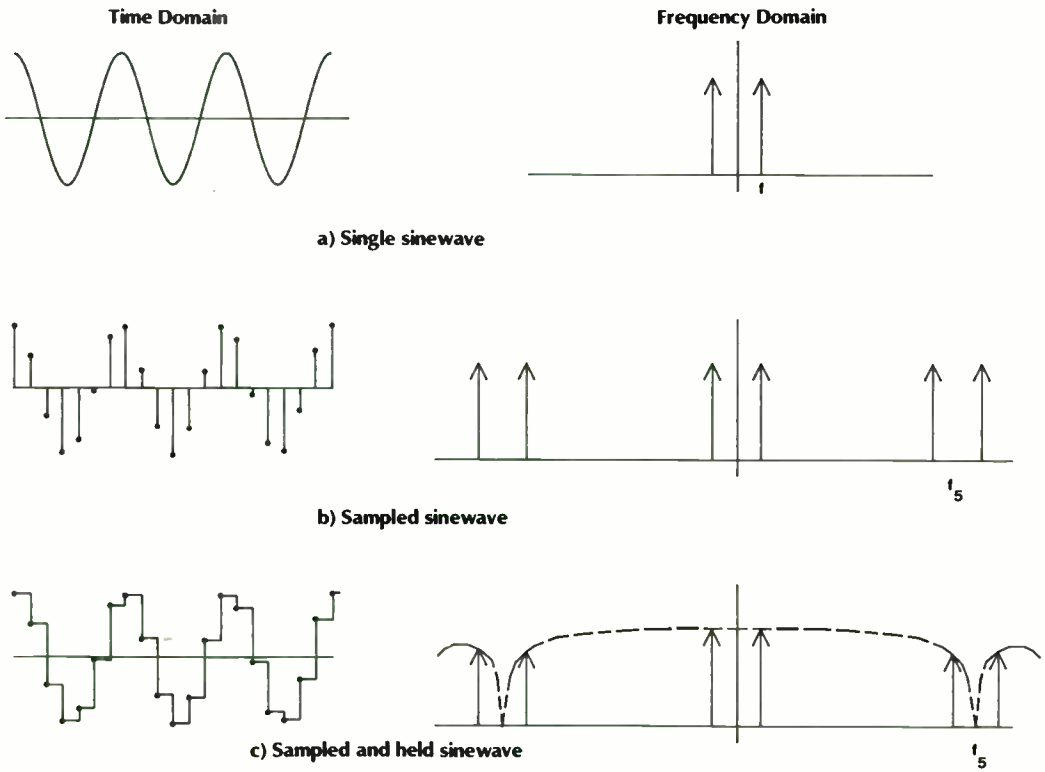


Figure 2. Sampling theorem applied to a sinewave

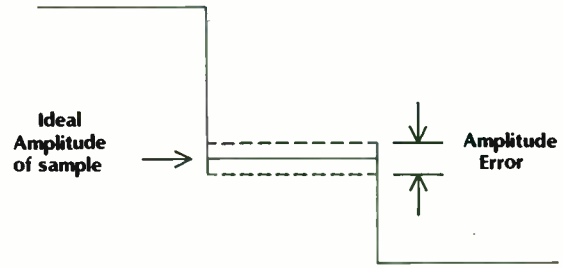


Figure 3. Amplitude error of a sample

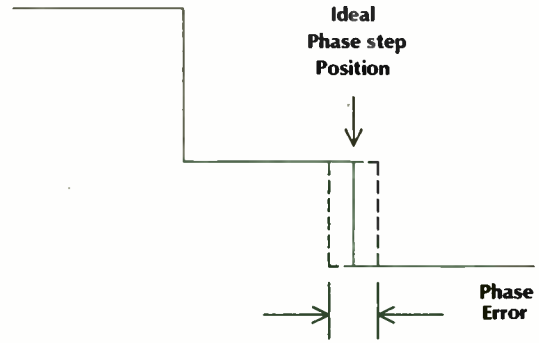


Figure 4. Phase error of a sample

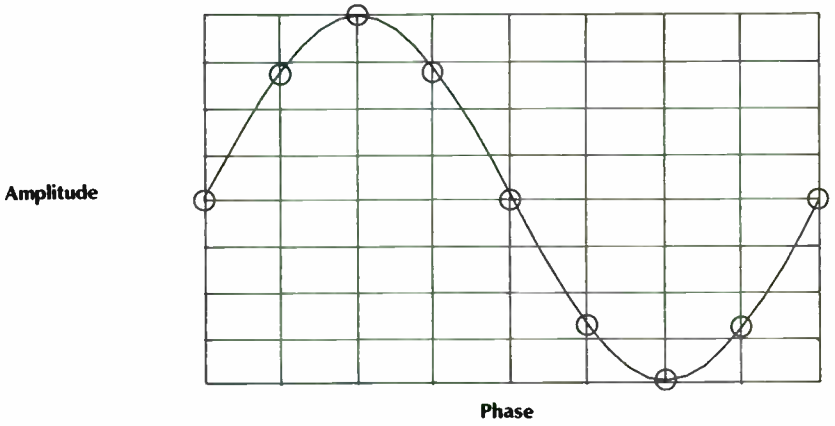


Figure 5. Amplitude errors in an 8 x 8 step (3 x 3 bit) Sinusoid waveform map

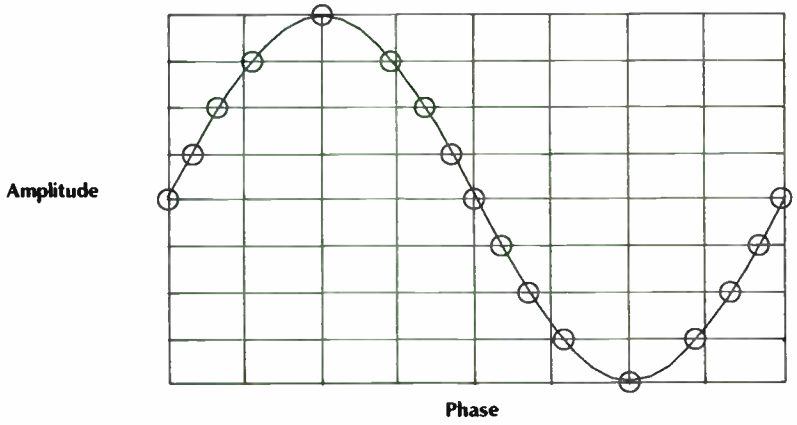


Figure 6. Phase errors in an 8 x 8 step (3 x 3 bit) Sinusoid waveform map

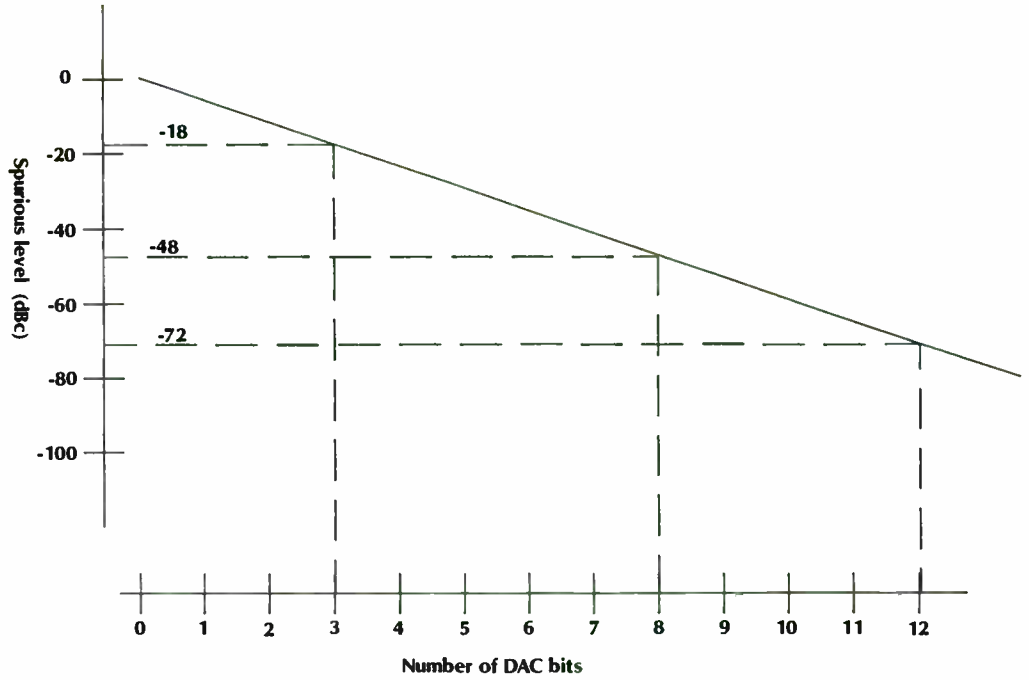


Figure 7. 'Conventional Wisdom' on spurious output levels as a function of DAC input bits

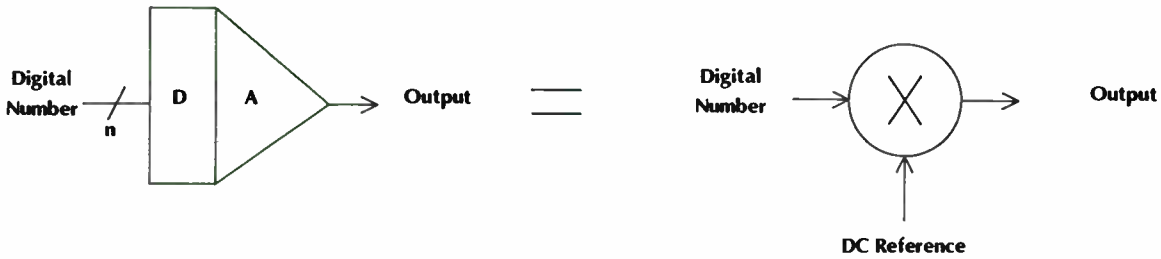


Figure 8. DAC operation as a multiplier

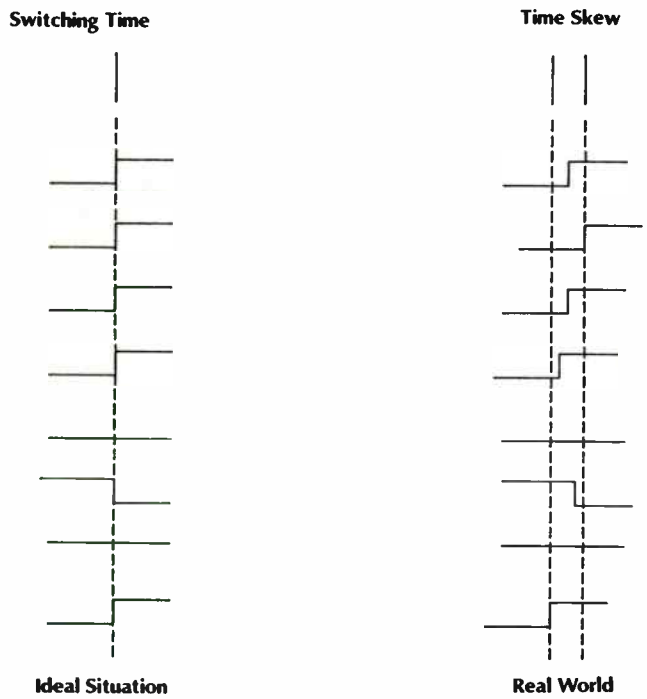


Figure 9. Time skew on the DAC digital input

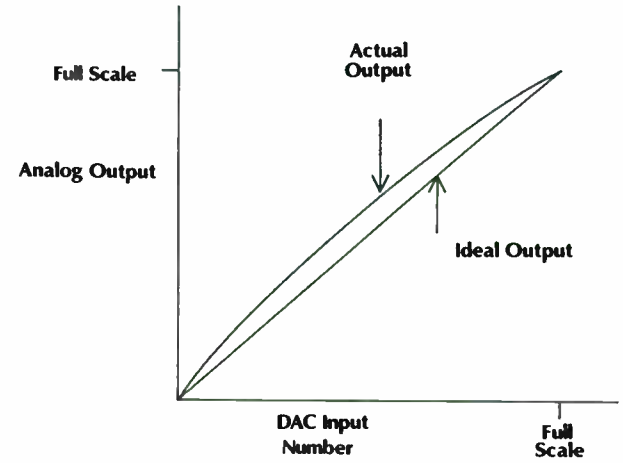


Figure 10. DAC Intergral nonlinearity

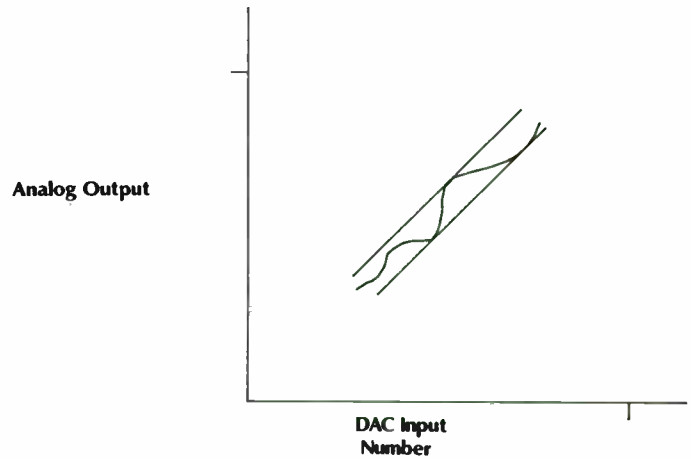


Figure 11. DAC Differential nonlinearity

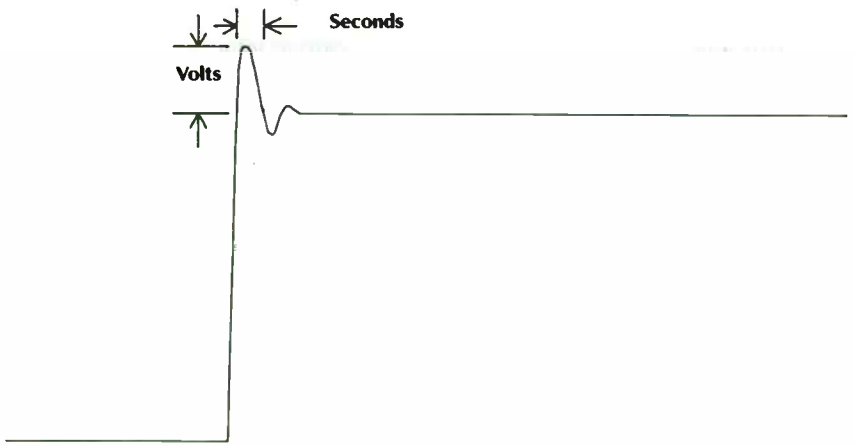


Figure 12. DAC output 'glitch energy' definition

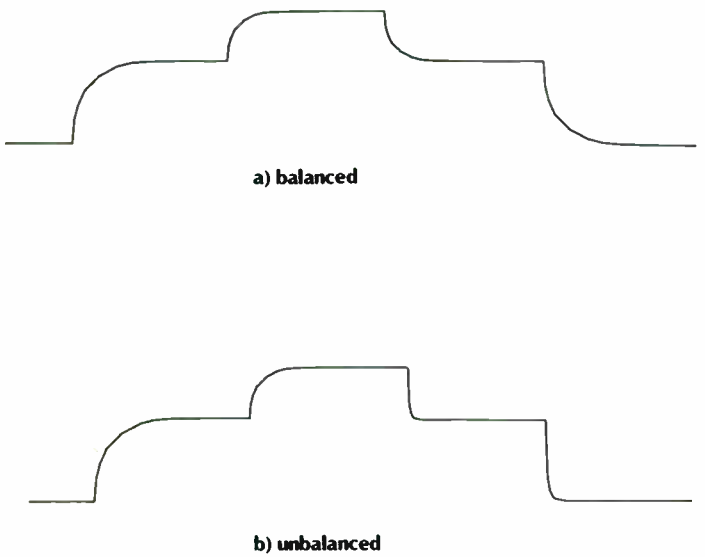


Figure 13. DAC output stage balance

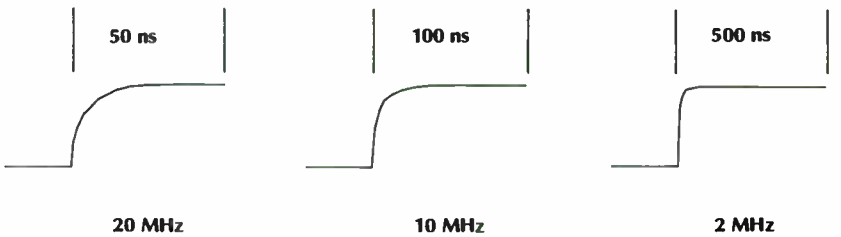


Figure 14. Effect of sample rate slowing on a high speed DAC output step

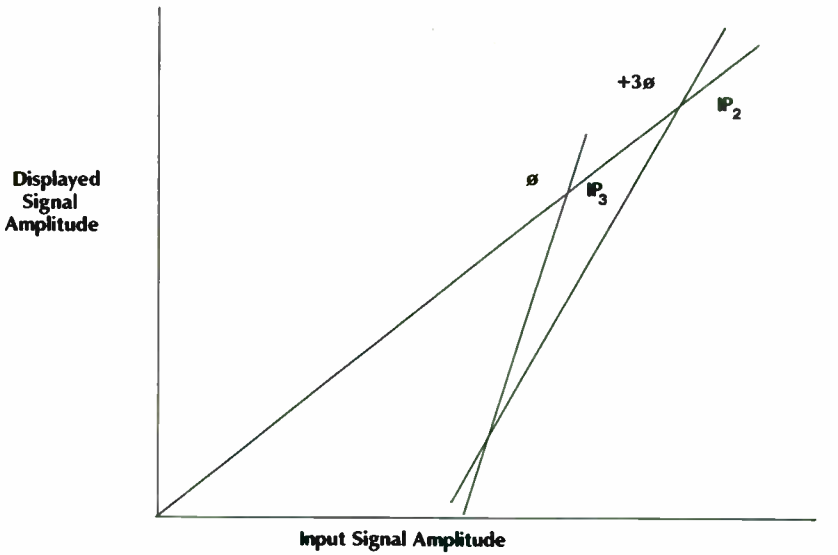


Figure 15. Measurement effects of spectrum analyzer input intercept points

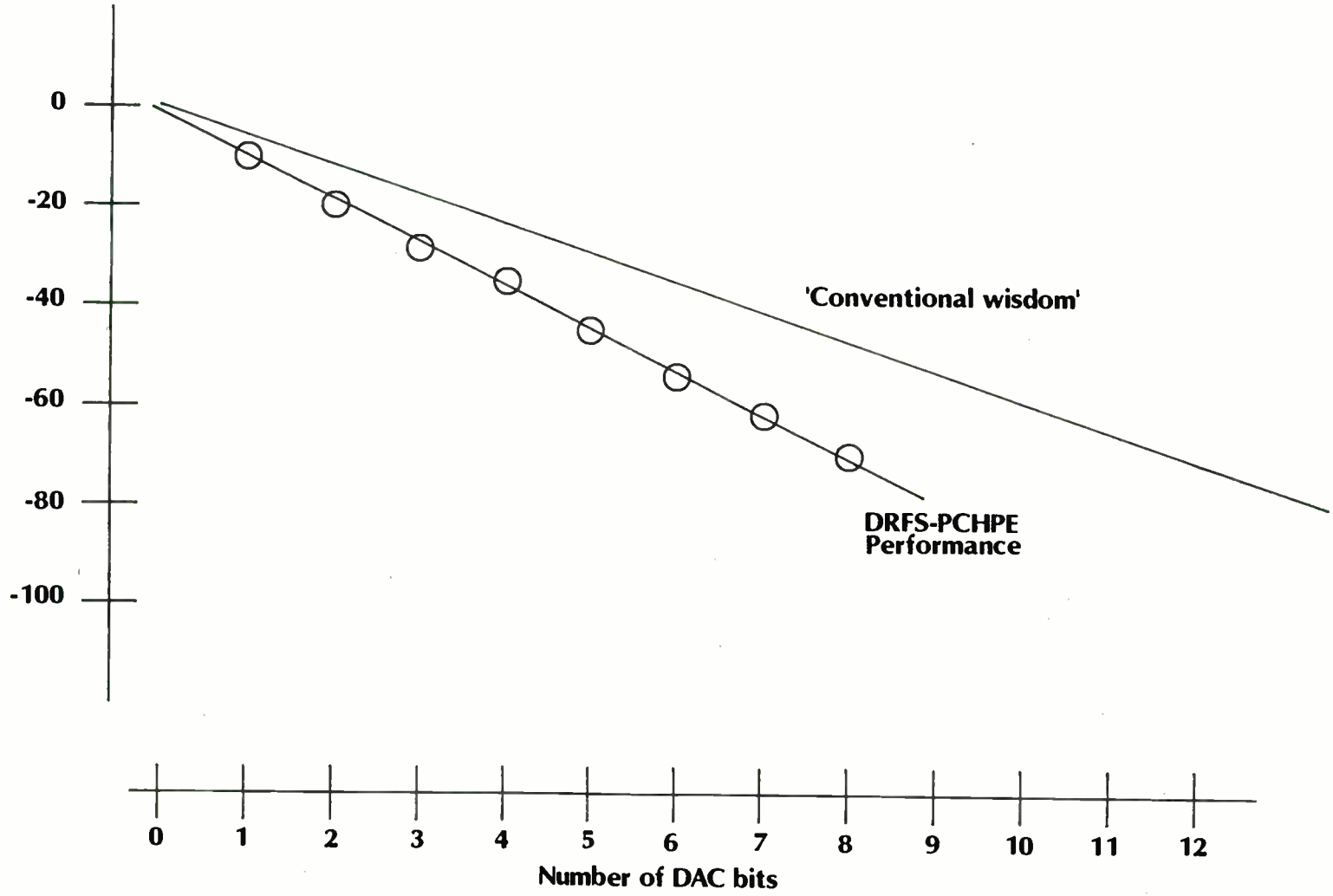


Figure 16. Typical performance of a modern direct digital sythesizer

rfdesign PROCEEDINGS ORDER FORM

PLEASE SEND ME THE FOLLOWING PROCEEDINGS:

☐ RF Tech 88 ☐ RF Tech 87 ☐ RF Expo East 87
☐ RF Tech 86 ☐ RF Expo East 86

Credit card: AE VI MC No. _____

Signature _____ **Exp. Date** _____

Name _____

Title _____

Company Name _____

Address _____

Bldg. No. _____ **M/S** _____

City _____ **State** _____ **ZIP** _____

1/88

RF DESIGN PROCEEDINGS are packed with all the timely, vital information presented at our trade shows. You'll find practical circuit design papers. . . papers on the leading edge of RF technology. . . papers you can't *afford* to be without.

Each costs only $95 ($125 outside the U.S.), and prepayment is required. Use your credit card or send your check to: **RF DESIGN PROCEEDINGS,** 6300 S. Syracuse Way, Suite 650, Englewood, CO 80111.

For a complete listing of the papers published in each, call (303) 220-0600, ext. 656.

rfdesign PROCEEDINGS ORDER FORM

PLEASE SEND ME THE FOLLOWING PROCEEDINGS:

☐ RF Tech 88 ☐ RF Tech 87 ☐ RF Expo East 87
☐ RF Tech 86 ☐ RF Expo East 86

Credit card: AE VI MC No. _____

Signature _____ **Exp. Date** _____

Name _____

Title _____

Company Name _____

Address _____

Bldg. No. _____ **M/S** _____

City _____ **State** _____ **ZIP** _____

1/88

RF DESIGN PROCEEDINGS are packed with all the timely, vital information presented at our trade shows. You'll find practical circuit design papers. . . papers on the leading edge of RF technology. . . papers you can't *afford* to be without.

Each costs only $95 ($125 outside the U.S.), and prepayment is required. Use your credit card or send your check to: **RF DESIGN PROCEEDINGS,** 6300 S. Syracuse Way, Suite 650, Englewood, CO 80111.

For a complete listing of the papers published in each, call (303) 220-0600, ext. 656.

rfdesign FREE SUBSCRIPTION REQUEST

To receive *RF Design*
Answer all questions, sign, date and mail today.

✓✓ I wish to receive *RF Design* free of charge ☐ yes ☐ no

✓✓ **Signature** _____ **Date** _____
required

NAME		
TITLE		
COMPANY NAME		
ADDRESS		
BLDG. NO.	M/S	
CITY	STATE	ZIP

The above is my ☐ company address. ☐ home.

1. Please check the one category which best describes your business.

25 ☐ CATV & broadcast equipment
45 ☐ Land mobile equipment
140 ☐ Video, audio equipment
142 ☐ Data transmission, computer systems
60 ☐ Instruments & test equipment
144 ☐ Medical electronics equipment
146 ☐ Industrial controls & power supplies
70 ☐ Consumer electronics equipment
80 ☐ Components & subsystems
15 ☐ Aviation, marine, navigation systems
112 ☐ Satellite & space systems
114 ☐ EW, ECM & radar systems
116 ☐ Military communications systems
118 ☐ Government agency (non-military)
90 ☐ Lab or consultant
150 ☐ User of electrical equipment
180 ☐ Library or school
Other _____ (please specify)

2. Please check the one category which best describes your title.

01 ☐ Engineering manager
02 ☐ Engineer (MTS)
03 ☐ Engineering Service Titles (Standards, Reliability, QC, Test Engineers)
04 ☐ President, Owner/operator, Vice President, General or Plant Manager
05 ☐ Other — please specify _____

9860

Place
Stamp
Here

rfdesign
Cardiff Publishing Company
6300 S. Syracuse Way, Suite 650
Englewood, CO 80111-9912

Place
Stamp
Here

rfdesign
Cardiff Publishing Company
6300 S. Syracuse Way, Suite 650
Englewood, CO 80111-9912

Place
Stamp
Here

rfdesign
Cardiff Publishing Company
P.O. Box 6317
Duluth, Mn. 55806-9954