COMSAT
Technical Review

COMSAT TECHNICAL REVIEW
Volume 14 Number 2, Fall 1984

A joint estimator-detector for QPSK data transmission

J. J. POKLEMBA

(Manuscript received January 19, 1984)

Abstract

A steady-state receiver structure for QPSK data transmission is developed, and its
performance is evaluated. The structure is derived from the principles of detection
and estimation theory. As the nature of optimum band-limited detection for QPSK is
well known, the thrust of this effort is to integrate a joint estimator, which is optimum
in a minimum mean square error sense, with the detector. The three parameters to
be estimated are the received signal amplitude, the received signal carrier phase, and
the symbol timing of the data modulation. The resulting implementation for the phase
and timing estimates consists of two interdependent, data-aided, phase-locked loops
(PLLS). The amplitude estimate also depends on the outputs from the other estimators
and the detector. Because each loop coherently aids the others, the best output signal-
to-noise ratio is obtained when the channel noise and the processes to be estimated
are jointly Gaussian.

The measurements performed may be divided into three basic categories: carrier
phase and symbol timing tracking loop jitter, cycle skipping, and bit error rate (BER).
The results show that at a 32-kbit/s symbol rate and narrow to moderate closed-loop
bandwidths ($\lesssim$ 320 Hz), the phase jitter on the recovered carrier reference is within
0.4° rms of that for a PLL operating on an unmodulated carrier. Furthermore, phase
jitter measurements for the recovered symbol timing reference with a 1010 . . . data
pattern fall directly on PLL predictions over the range of bandwidths tested, from
32 Hz to 1 kHz.

Introduction

This paper describes the development of a practical steady-state QPSK receiver in terms of joint multidimensional estimation and detection theory, and presents an assessment of its performance. When coherent carrier and symbol timing references are available, the structure of the optimum data detector for QPSK is well known. Thus, this paper focuses on the derivation of carrier and timing recovery implementations, and the manner in which they are combined within the decision-making portion of the receiver.

The approach described requires joint estimation of the amplitude, carrier phase, and symbol timing uncertainties according to a minimum mean square error (MMSE) cost function. Although the derivation is intended for QPSK, it degenerates readily for BPSK operation, and can be extended to M-ary signaling. It is also possible to adapt the joint estimator-detector structure to other 2-dimensional signaling formats, which can be modeled as amplitude-modulated, quadrature carrier data transmissions. These include minimum shift keying (MSK) and quadrature amplitude modulation (QAM).

Numerous techniques for carrier and clock information recovery for digital data transmission have found acceptance [1]–[8]. These are generally characterized by either serial or parallel reconstruction. The former method allocates separate time intervals to obtain carrier and clock synchronization, whereas the latter enacts simultaneous recovery. Their common characteristic is that carrier and clock synchronization and data detection do not coherently aid each other. Only recently have studies considered joint optimization [9]–[12], which has the advantage that the information assimilated in recovering symbol timing aids in reconstructing the carrier, and vice versa. In addition, decision-feedback data interact with the tracking loops to minimize the effect of pattern noise.

One of the earliest realizations of a joint estimator-detector was developed by Kobayashi [9], as depicted in Figure 1. When data sequences are represented by the complex variables, z_n and c_n, application of Kobayashi's realization extends over various modulation formats, including pulse amplitude modulation (PAM-DSB, VSB, or SSB), QAM, and coherent or differential phase-shift keying (PSK). Observe that the receiver feeds back the postdetected (sampled) data to correlators, which output the control signals in the carrier and clock recovery loops. An obvious advantage with this technique is that the multiplication entails a signal-plus-noise times a signal operation, whereas previous approaches have often been signal-plus-noise squared.

The structure shown in Figure 2, derived by Falconer and Salz [10], is similar to that proposed by Kobayashi. The primary difference is that Falconer and Salz's phase and timing estimates are updated on a symbol-by-symbol basis, rather than by an average taken over a data sequence.
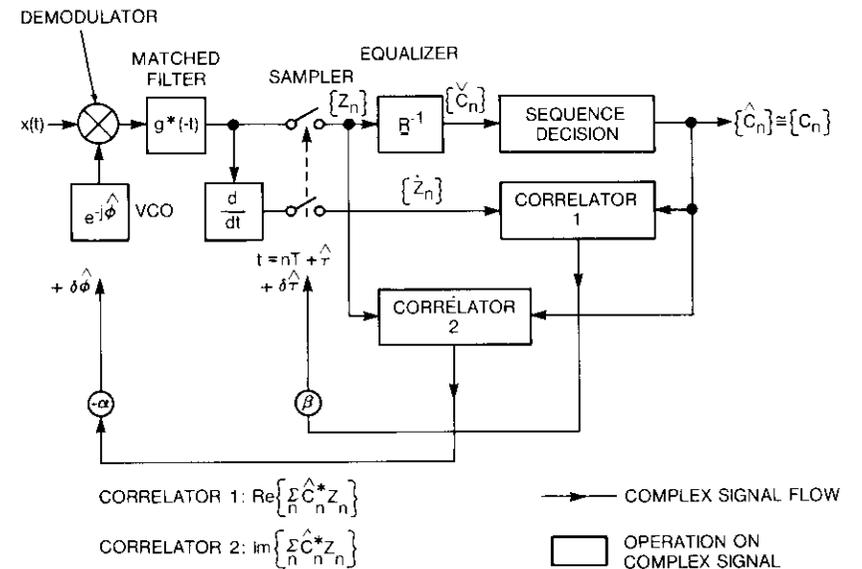
Figure 1. *Kobayashi Receiver* [9]
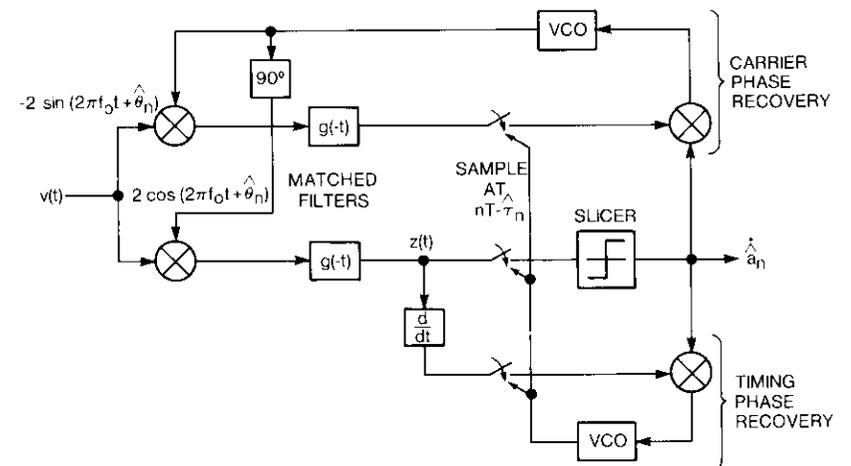
Figure 2. *Falconer and Salz Receiver* [10]

In Figure 3, the QAM demodulator developed by Franks [8] shows how the previous implementations can be extended to recover data transmitted on
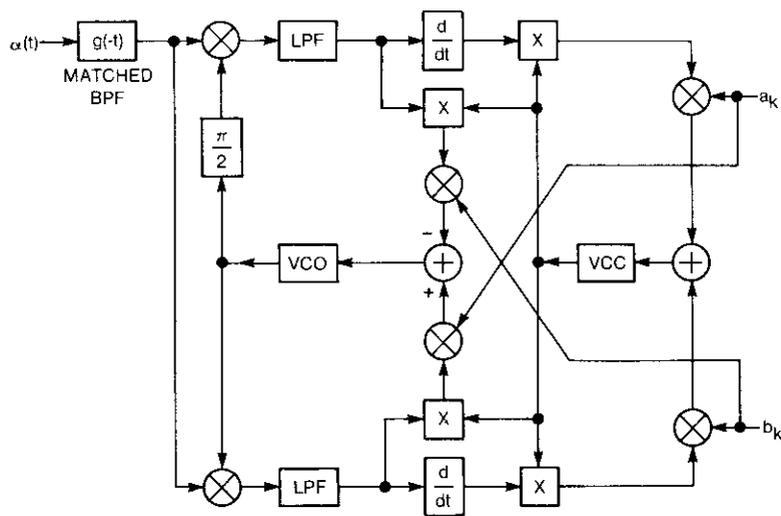
Figure 3. *Franks Receiver [8]*

quadrature carriers. Note that the correlator outputs for each channel are added. This increases the overall loop signal-to-noise ratio (S/N) because the in-phase and quadrature noises at baseband are independent.

A version of the joint estimator-detector developed herein is shown in Figure 4 [13]. It was derived from multidimensional waveform estimation principles, rather than by maximizing separate likelihood ratios. This approach treats the transmission impairments as random processes rather than random variables, and it allows additional parameters such as random amplitude to be integrated into the structure. The random process technique also provides a definition of the optimum loop filters for minimum steady-state tracking jitter. Moreover, the joint estimator-detector constitutes a more general and comprehensive approach.

The continuous waveform estimator is also more compatible with the input signal format. That is, the previous implementations utilized sampled and held signals at the incoming and feedback correlator inputs in both the carrier and clock recovery loops; whereas the structure of Figure 4 has continuous analog signal inputs. This is a consequence of multidimensional waveform estimation, which indicates that continuous analog (rather than sampled and held) signals should be used, because the receiver is operating on a continuous signal. The analog feedback estimates are derived by filtering the postdetected data again, using the same characteristic employed in the transmitter. As a result, the optimum data feedback signals are estimates of the baseband data
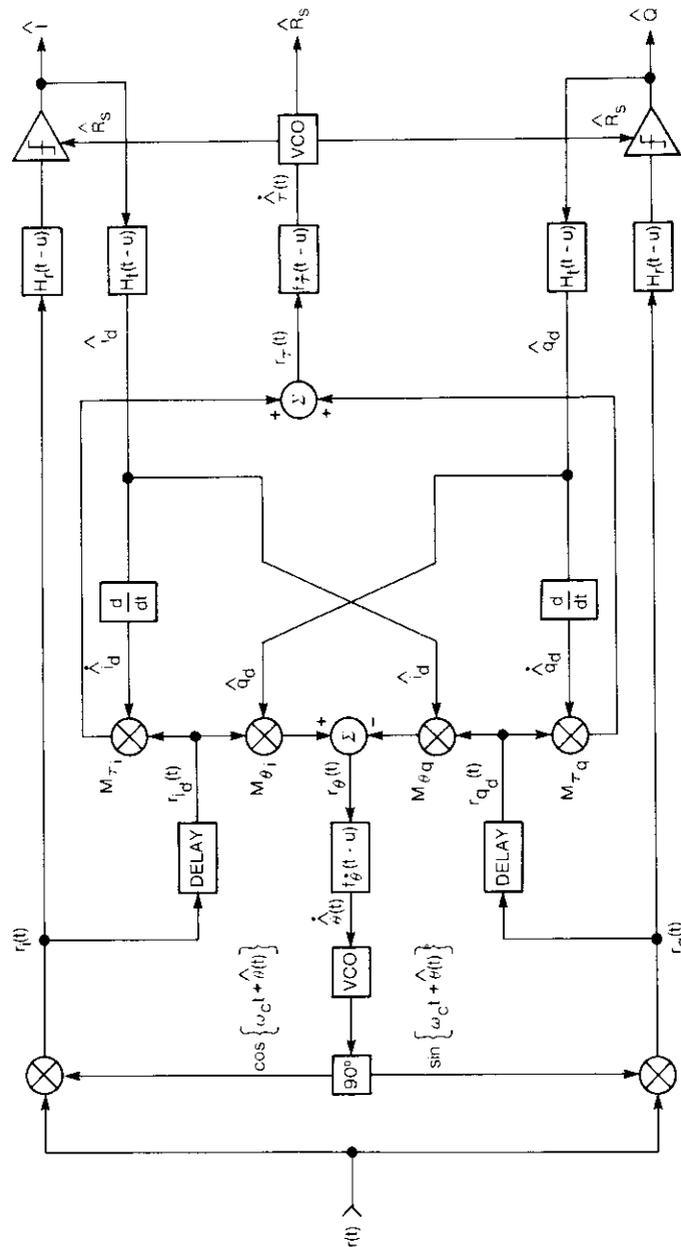
Figure 4. *Poklemba Receiver*

that were transmitted, and the data correlators yield a maximum output S/N, similar to a matched filter.

Another distinction concerns the signal processing in the symbol timing loop. The present derivation suggests that differentiation be performed on the data feedback, rather than on the incoming signal plus noise. With either method, the signal component of the timing loop control voltage basically remains the same. However, differentiating the input noise component gives rise to f^2 noise; whereas, with the multidimensional estimator approach, the noise spectrum remains white.

Problem formulation

Development of the optimum MMSE demodulator structure is based on the stochastic relationship between the received signal and noise, and the various parameters to be detected and estimated. Thus, the analysis is predicated on a description of a generalized received signal, $r(t)$, which has been passed through a linear additive noise channel,

$$r(t) = s[t, \underline{a}(t)] + n(t) \quad . \tag{1}$$

The noise, $n(t)$, is assumed to be zero mean, white, and Gaussian, with a double-sided spectral height of $N_o/2$. The vector, $\underline{a}(t)$, is composed of the transmission uncertainties embedded in the signal; that is,

$$s[t, \underline{a}(t)] = s[t, A, \theta(t), \tau(t)] \quad . \tag{2}$$

The components of $\underline{a}(t)$ are the received signal amplitude, A; the carrier phase, $\theta(t)$; and the symbol timing of the data modulation, $\tau(t)$.

It is well known that digital data may be transmitted without intersymbol interference in a bandwidth-efficient manner by using Nyquist filtering [14]. Nyquist has postulated that the necessary minimum, one-sided, half-amplitude bandwidth, B, is such that $BT_s = 0.5$; where T_s is the time duration of an information symbol. When Nyquist filtering is employed, the transmitted waveform for many PCM signaling schemes can take the form of amplitude-modulated carriers. However, to develop a detailed receiver structure, a QPSK modulation format will be employed because of its power-bandwidth efficiency and widespread use.

For filtered QPSK, the signal portion of equation (1) can be written in an amplitude-modulated format as

$$
\begin{aligned}
s[t, \underline{a}(t)] &= A \sum_{k=1}^{M} i_k[t - kT_s + \tau(t)] \cos [\omega_c t + \theta(t)] \\
&\quad + q_k[t - kT_s + \tau(t)] \sin [\omega_c t + \theta(t)] \\
&= A\{i(t, \tau) \cos [\omega_c t + \theta(t)] \\
&\quad + q(t, \tau) \sin [\omega_c t + \theta(t)]\}
\end{aligned} \tag{3}
$$

where

$$i(t, \tau) = \sum_{k=1}^{M} i_k[t - kT_s + \tau(t)]$$

$$q(t, \tau) = \sum_{k=1}^{M} q_k[t - kT_s + \tau(t)] \quad .$$

The signal consists of in-phase and quadrature carriers of frequency, ω_c, modulated by the filtered baseband data sequences, $i_k(t, \tau)$ and $q_k(t, \tau)$, respectively, which are bipolar, unit power, Nyquist pulse-shaping functions. For notational convenience, the data sequences have been written as continuous analog waveforms, $i(t, \tau)$ and $q(t, \tau)$. The sequences carry M information symbols over the observation interval, $T = MT_s$. The signal parameters are summarized in Table 1.

TABLE 1. WAVEFORM DESCRIPTION

SYMBOL	DESCRIPTION
$r(t)$	Received signal plus noise
$s[t, \underline{a}(t)]$	Received signal with uncertainties
$n(t)$	Received noise
A	Unknown amplitude variable
$\theta(t)$	Carrier phase uncertainty parameter
$\tau(t)$	Symbol timing uncertainty parameter
ω_c	Carrier frequency
T_s	Symbol time interval, $1/R_s$
$i_k(t, \tau)$	kth transmit filtered in-phase data pulse shape
$q_k(t, \tau)$	kth transmit filtered quadrature data pulse shape
$i(t, \tau)$	Entire transmit filtered in-phase data waveform
$q(t, \tau)$	Entire transmit filtered quadrature data waveform

For burst operation, M is divided into two parts: a preamble ($1 \leq k \leq P$) where time is allocated for receiver synchronization, and a message portion ($P + 1 \leq k \leq M$) to communicate information. In general, there are both acquisition and steady-state solutions to the problem. This investigation focuses on the steady-state solution.

Appendix A gives a derivation of an optimum MMSE joint estimator-detector receiver structure. Since this implementation results from the direct application of the estimator equations, it is designated as a "textbook" receiver structure. The textbook implementation can be greatly simplified when a Nyquist-filtered QPSK signal format is used. The particular QPSK receiver which emerges is referred to as an analog feedback structure, since it requires Nyquist-filtered, postdetection data feedback to coherently aid the processing in all three estimators.

Because the excess delay involved in refiltering the data has a negative impact on carrier recovery loop stability, a more practical suboptimum digital feedback implementation is presented. This approach utilizes decision feedback directly, thereby eliminating the postdetection data filtering stage. Thus delay is reduced, and the maximum operable bandwidth can be doubled.

Implementation description

A block diagram for the QPSK modulator is shown in Figure 5. In-phase and quadrature (I and Q) digital input data streams at a symbol rate $R_s = 32$ kbit/s are converted to bipolar levels to facilitate suppressed carrier modulation. Transmit data shaping is effected by cascading second-order RC active filter sections. In particular, each data filter consists of a 6-pole Butterworth magnitude response followed by a second-order group delay equalizer. A square-root reciprocal $(\sin x)/x$ section is also included so that the 3-dB bandwidth for the transmit Butterworth filter may be normalized to $R_s/2$ to enable operation at the Nyquist rate, $B/R_s = 0.5$ [15]. The I and Q
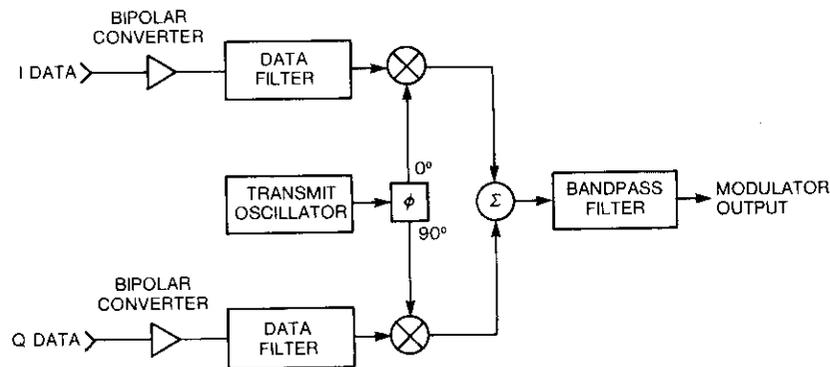
Figure 5. *QPSK Modulator*

data channels are then quadrature-phase modulated on an 18-MHz IF carrier. A 1-MHz bandpass filter removes spurious mixing products while passing the desired signal.

The digital feedback QPSK demodulator with data-aided joint carrier and clock recovery, functionally outlined in Figure 6, consists of two parallel I and Q data detectors whose carrier phase and symbol timing references are jointly estimated, along with the data, in an interdependent fashion. Post-detected (sampled) data feedback is used to coherently aid the processing in the carrier and clock recovery loops. Estimates derived in the receiver are denoted with ^s.

A 110-kHz-wide bandpass filter centered at 18 MHz is intended to limit the power of the input noise while being relatively transparent to the incoming signal spectrum. After coherently mixing down to baseband, the data filters complete the shaping of the minimum intersymbol interference signal with the same design parameters as their counterparts in the modulator. The coherent carrier reference is obtained from a crystal voltage-controlled oscillator (VCO) which operates at 18 MHz. With random or "10" data modulation, the carrier loop filter is configured to give a second-order frequency response with damping ratio $\zeta = 0.7$. The loop gain, which is the product of the phase detector, VCO, and loop filter DC gains, for the 32, 100, 320, and 1,000-Hz bandwidths is 360,000 (1-kHz frequency offset corresponds to a 1° phase error). The loop gain for an unmodulated carrier is 650,000, with a damping ratio of 0.95, which yields bandwidths of approximately 50, 150, 500, and 1,500 Hz.

The reconstructed symbol clock with embedded symbol timing reference is extracted by dividing down the output of a crystal-controlled VCO running at 1.92 MHz (60 R_s). The clock loop filter is also configured to operate as a second-order system with $\zeta = 0.7$. Its loop gain with "10" data modulation is 3,600 (10 Hz/1°), whereas with PN modulation, the average loop gain is 2,250 and the damping ratio is 0.55.

In Figure 7a, b, and c, waveforms from the detector and estimators are displayed for joint carrier and clock recovery bandwidths of 320 and 250 Hz with $E_b/N_o = \infty$, 10 dB, and 0 dB, respectively. Proceeding from top to bottom, the first frame is an x-y representation of the signal space after the receive data filters; the next displays the I-channel eye pattern along with the recovered clock; and the last shows the phase detector outputs (summed outputs from Figure 6) in the carrier (upper) and clock (lower) recovery loops. It can be seen that, with digital rather than analog data feedback, the pattern noise at the carrier phase detector output is not completely eliminated. Also, the symbol timing phase detector is discontinuous at $\pm T_s/2$. The three
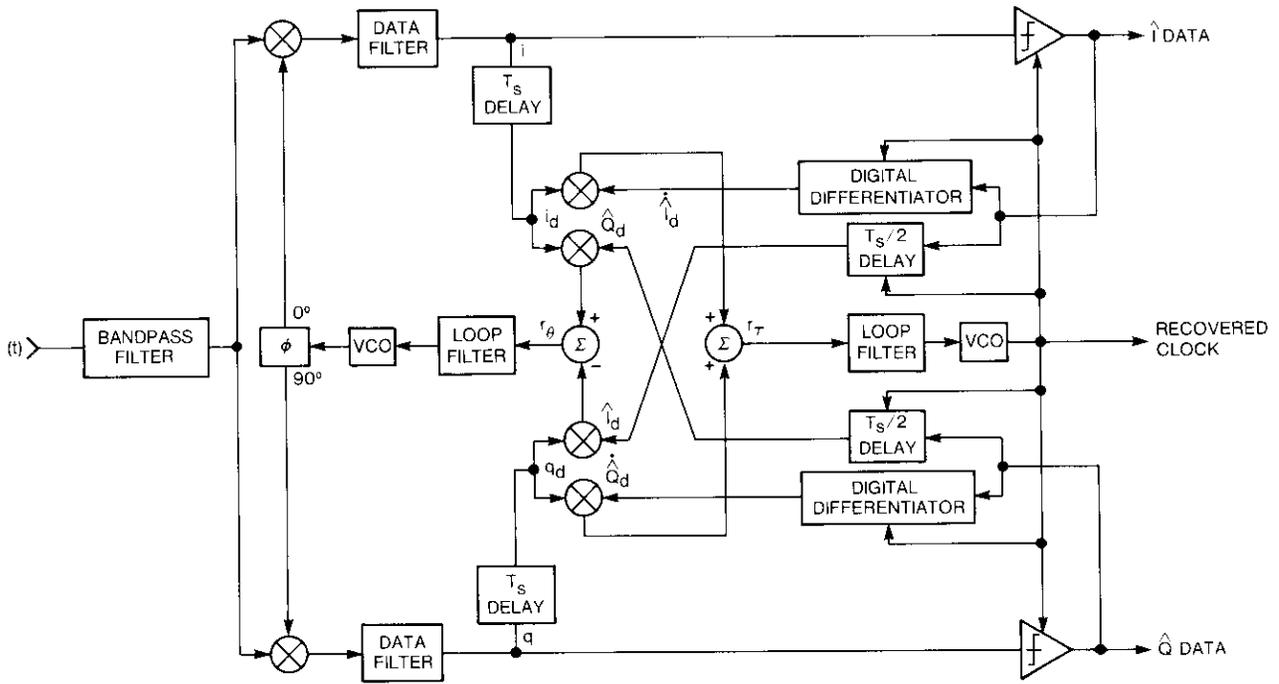
Figure 6. *QPSK Demodulator*

QPSK Signal Space

I-Channel Eye Pattern

Recovered Clock

Carrier Phase Detector Pattern Noise

Clock Phase Detector Pattern Noise

(a) $E_b/N_o = \infty$ (b) $E_b/N_o = 10$ dB   (c) $E_b/N_o = 0$ dB
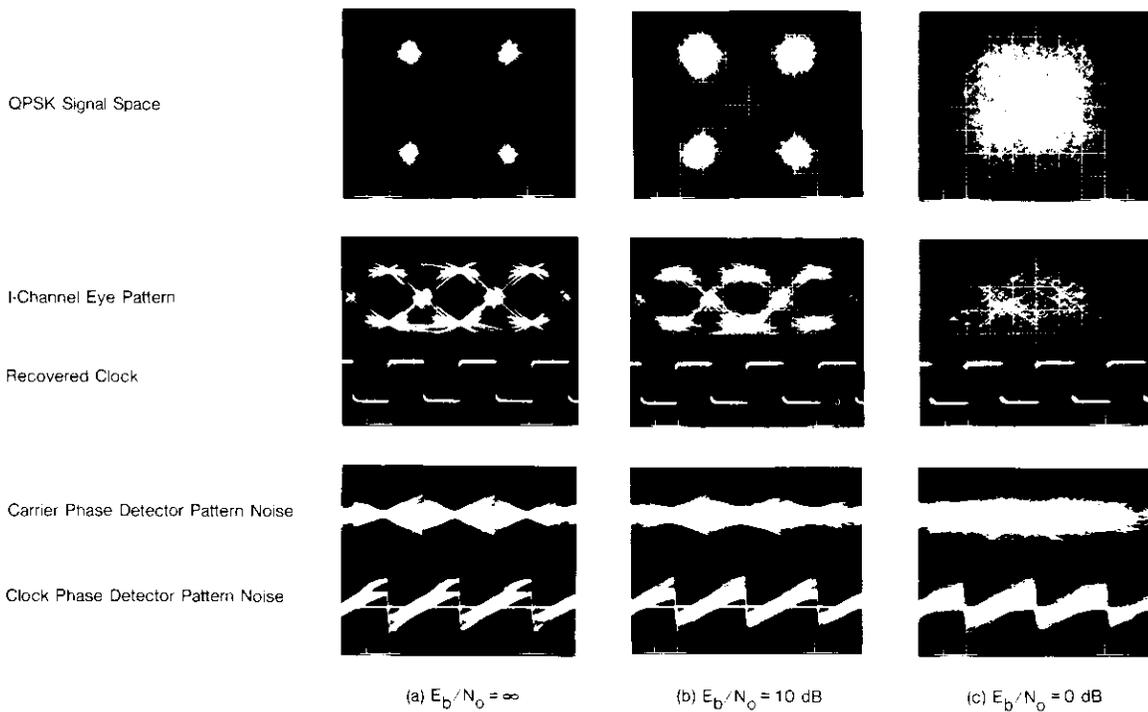
Figure 7. *Demodulator Waveforms*

levels in the symbol timing detector correspond to the cases of no data transitions (zero baseline), or transitions in one or both channels. Finally, observe in Figure 7c that even when the signal space appears to be completely obscured by noise $(E_b/N_o = 0$ dB), coherent control signals in the carrier and clock recovery loops are discernible; and the recovered clock is remarkably free of jitter.

Phase jitter performance

Carrier tracking loop

The rms phase jitter measured while recovering an unmodulated carrier with hardwired clock is plotted in Figure 8 vs loop bandwidth, B_L, and E_b/N_o. For comparison, the theoretical output jitter taken from equations (B-1) and (B-11) and Table B-1 in Appendix B, is drawn for $B_L = 50$ Hz. The theoretical curve has two components: phase jitter on the local oscillators, and phase jitter resulting from the input thermal noise modulating the carrier VCO. At lower loop bandwidths, the thermal noise power contribution is reduced and the local oscillator jitter component is dominant; whereas at larger loop bandwidths, local oscillator jitter is tracked and the thermal component dominates.

The measured phase jitter for $B_L = 50$ Hz is about 0.3° poorer than ideal at moderate-to-high E_b/N_o, and diverges more at low E_b/N_o where the PLL linearized analysis is no longer valid. Loss of synchronization typically occurs when the rms phase jitter on the carrier exceeds approximately 6° or 0.1-rad rms. The remaining curves for $B_L = 150$ Hz to 1.5 kHz are dominated by the thermal noise component of equation (B-11). Although the curves should be about 5 dB apart according to their bandwidth ratios, the $B_L = 1.5$ kHz curve is about 7 dB beyond that for 500 Hz. This is because the excess delay $(3.9T_s)$ incurred with postdetected, data-aided tracking causes the larger closed-loop bandwidths to become underdamped, and hence less stable. Thus, the noise equivalent bandwidth is, in effect, wider than intended. A subsequent section will reveal that the instability caused by the excess delay also adversely affects BER performance.

An analogous set of curves for output phase jitter with a PN data-modulated carrier and joint carrier and clock recovery is presented in Figure 9. The performance is similar to that of Figure 8 with no modulation. Additional measurements with hardwired clock rather than joint carrier and clock recovery showed no perceptible difference. Moreover, the inclusion of data modulation appears to cause a degradation of 0.1° in rms output jitter over most of the operable range of loop bandwidths and S/Ns. This small degradation is most likely due to the utilization of digital rather than analog data feedback.
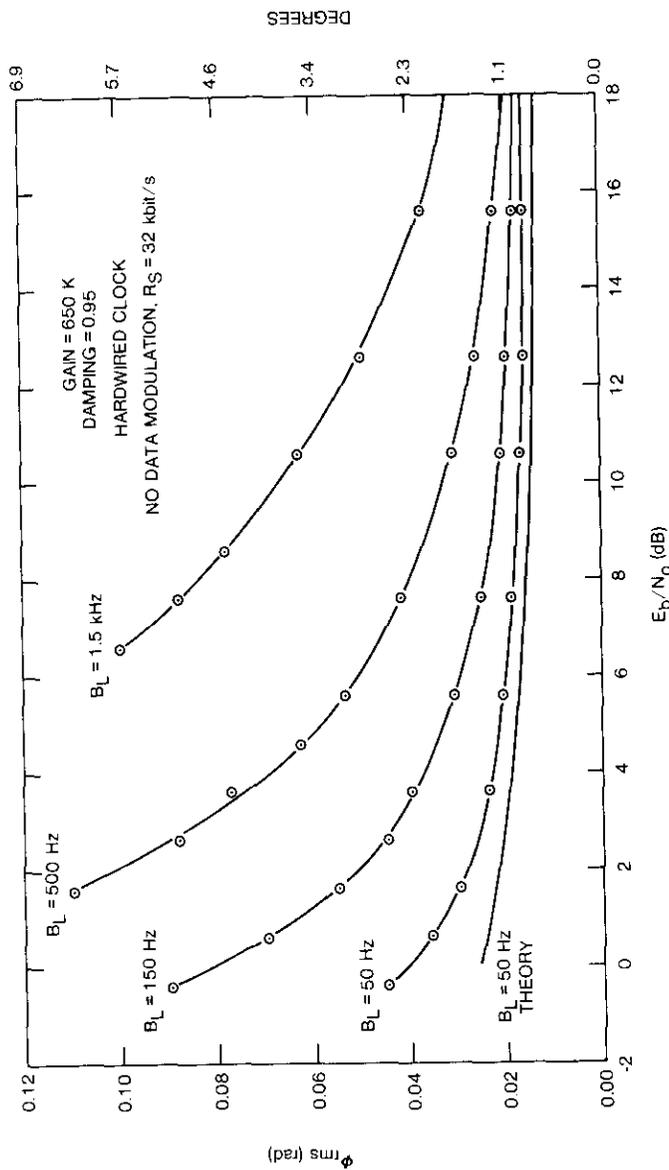
Figure 8. QPSK Carrier Tracking Loop Phase Jitter vs E_b/N_o and Carrier Recovery Loop Bandwidth, B_L
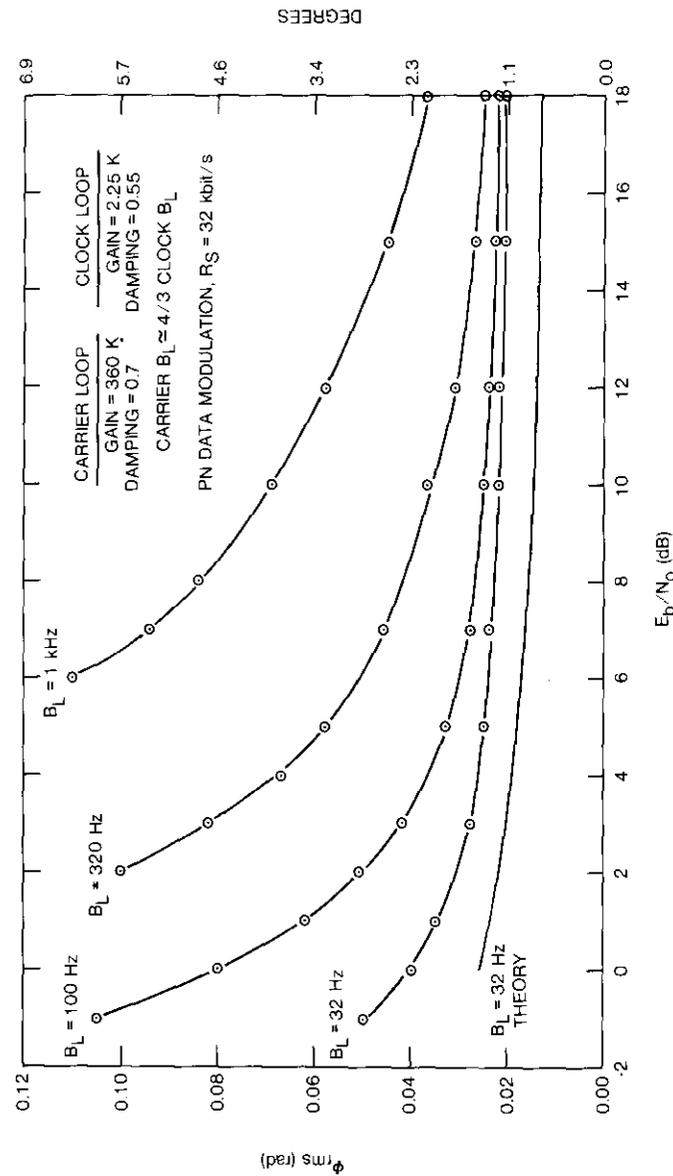
Figure 9. QPSK Carrier Tracking Loop Phase Jitter vs E_b/N_o and Joint Recovery Loop Bandwidth, B_L

Apart from this small degradation, the data-aided carrier recovery technique yields basically the same phase jitter performance as that of an unmodulated PLL, which is the best that can be expected.

Clock tracking loop

The ideal rms phase jitter on the reconstructed clock with "10" data modulation derived in Appendix B and given by equation (B-21) is plotted in Figure 10 vs loop bandwidth, B_L, and E_b/N_o. Because of the relatively low phase noise on the transmit and receive clock oscillators, the phase jitter is dominated by the thermal noise component. Measured data points are included for comparison. There was no discernible difference between joint recovery and hardwired carrier recovery at moderate-to-high E_b/N_o. However, with a hardwired carrier, the loss of lock threshold was greatly extended.

With B_L = 32 and 100 Hz, the points are very close to the predicted performance; however, when the loop bandwidth is widened, the measured points are actually better than expected. This is due to control voltage compression in the clock recovery loop and modulation rate limitations in the VCO. That is, for the wider loop bandwidths, the amplitude of the double frequency pattern content out of the phase detector grows proportionally larger than the desired low-frequency control signal. As a result, the low-frequency thermal jitter is limited prior to the VCO by an overdriven amplifier, and the VCO's 10-kHz maximum modulation rate suppresses the double frequency pattern content at R_s = 32 kHz. This problem can be avoided either by using a VCO with sensitivity that is more compatible with other elements which make up the loop gain, or by placing an additional low-pass filter in the loop to remove the clock component at R_s. The latter technique was experimentally verified subsequent to the measurements.

At low E_b/N_o, the measured data in Figure 10 diverge from predicted performance at the point where linearized PLL analysis is no longer valid. The rms jitter where loss of lock occurs (about 12°) is twice as large as that in the carrier recovery loop. This is because the maximum carrier phase detector excursions with a QPSK modulation format are ±π/4 (±45°), whereas the clock recovery phase detector is restricted by ±π/2 (±90°).

Although an analysis which predicts clock recovery loop output phase jitter with PN data modulation is beyond the scope of this paper, some comments regarding general loop performance in switching from "10" to PN modulation are in order. Recall that the phase detector in the clock recovery loop comprises a Nyquist-filtered baseband QPSK waveform multiplied by its time derivative. When no data transitions occur, the derivative
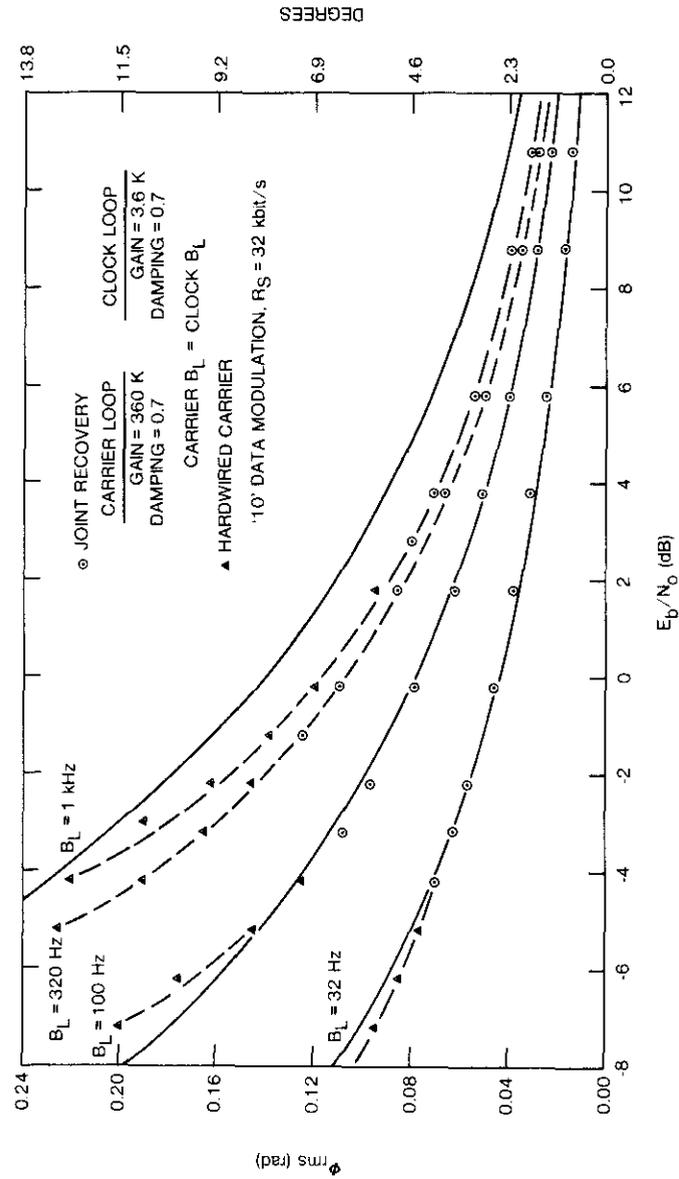
Figure 10. *QPSK Clock Tracking Loop Phase Jitter vs E_b/N_o and Loop Bandwidth, B_L*

is zero, and the symbol timing estimator behaves as a gated PLL. Because there are approximately half as many data transitions as symbols in a PN sequence, with a QPSK signaling format the probabilities of zero, one, or two transitions occurring at the phase detector output are $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively. This causes the average control signal power to be $\frac{3}{8}$ (-4.26 dB) that of "10" modulation. The average noise power in the loop is cut in half (-3.01 dB) as a result of the 50-percent duty cycle in the derivative feedback, and is reduced by an additional 1.76 dB due to a narrowing of the closed-loop bandwidth. That is, in switching from "10" to PN modulation, the average loop gain is halved, which lowers both the damping ratio and natural frequency of the loop, and reduces the bandwidth by approximately one-third. The net result is an increase in the average clock recovery S/N of 0.51 dB. In practice, the transmit power also changes as a function of the modulation pattern, so the measured results deviate slightly from these values.

The clock rms phase jitter measurements with PN modulation and hardwired carrier are given in Figure 11. The performance is significantly poorer than that for "10" modulation in Figure 10. Since the gated phase detector maintains loop S/N, the large increase in output jitter with PN modulation must be attributed to the zero crossing pattern jitter on the input waveform to the loop. Moreover, as the rolloff rate of Nyquist data filtering is increased, so is the spread of the zero crossings to which the clock must synchronize.

The B_L = 25, 80, and 250-Hz curves in Figure 11 are about 5 dB apart according to their bandwidth ratios; whereas, the B_L = 800 Hz data is much better than expected. This is again due to control voltage compression and modulation rate limitations in the clock recovery loop. With "10" modulation, phase jitter suppression occurred for both B_L = 320 Hz and 1 kHz. This is because the average amplitude of the double frequency component for "10" modulation is larger than that of PN, and consequently, it compresses sooner.

The measurements were repeated for joint carrier and clock recovery, as shown in Figure 12. The B_L = 25, 80, and 250-Hz curves have an additional 0.2° of rms output jitter relative to the hardwired carrier case. Thus, carrier recovery causes a slight degradation in the clock loop output jitter, whereas the previous section showed that clock recovery had no influence on carrier loop output jitter. The most significant impact of joint recovery on clock loop performance, as shown in Figures 11 and 12, is that loss of lock threshold is reduced by about 2 dB relative to a hardwired carrier.
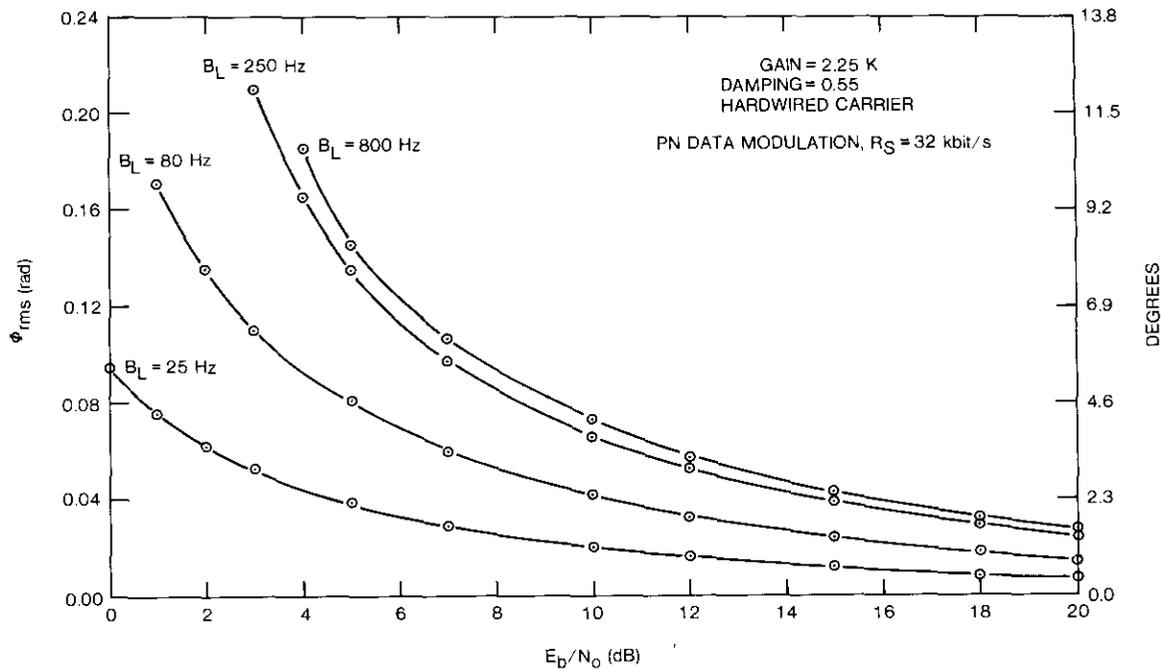
Figure 11. *QPSK Clock Tracking Loop Phase Jitter vs E_b/N_o and Clock Recovery Loop Bandwidth, B_L*
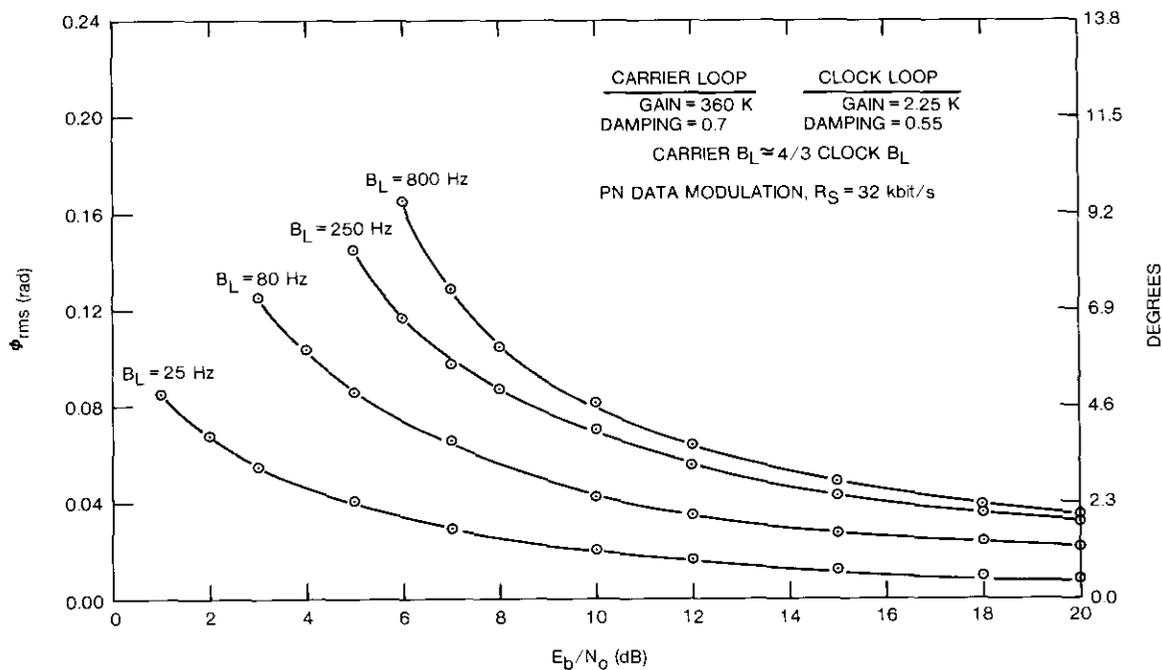
Figure 12. *QPSK Clock Tracking Loop Phase Jitter vs E_b/N_o and Joint Recovery Loop Bandwidth, B_L*

Cycle skipping performance

Carrier tracking loop

The average number of carrier cycle skips per bit for joint and hardwired clock recovery is plotted in Figure 13 *vs* recovery loop bandwidth, B_L, and E_b/N_o. These were obtained by observing the number of slips per unit time and normalizing by the bit rate. As expected, an exponential relationship exists between cycle skipping and E_b/N_o [16]. Joint and hardwired clock recovery data are virtually identical, except at very low E_b/N_o where joint recovery is a few tenths of a dB poorer. Note that as the loop bandwidth is lowered, the curves are progressively closer together than their bandwidth ratio of 5 dB. This is probably due to an effective lowering of the loop S/N caused by decision feedback when the error rate exceeds 10^{-1}. [For example, with BER $= 10^{-1}$, an average of 9 out of every 10 bits influence the loop control signal in the proper direction. However, one bit pushes in the opposite direction, and the recovery loop S/N is reduced by $(0.8)^2$ or 1.9 dB.] The impact of data feedback errors is more dramatic if there is a significant probability of a higher instantaneous error rate within a time interval $\sim 1/B_L$. (For example, if BER $= 4 \times 10^{-1}$, the loop S/N would be lowered by 14 dB.) The cycle skipping curves may also be affected by signal compression in the demodulator front end at low S/Ns. In particular, with $B_{IF} \approx 110$ kHz, $(C/N)_{IF} \approx -3$ dB when $E_b/N_o = 0$ dB.

Note that the cycle skipping curves of Figure 13 continue below the S/N which corresponds to the 6° rms phase jitter thresholds in Figure 9. This should not be misinterpreted as a lower threshold, since the endpoints were dictated by the limitations of the measurement techniques. Subsequent BER measurements will verify that the 6° rms phase jitter threshold in the carrier loop is appropriate.

Clock tracking loop

An analogous set of curves for cycle skipping in the clock recovery loop is shown in Figure 14. All three cases experienced a few tenths of a dB degradation in the cycle skipping performance with joint recovery, as opposed to hardwired carrier. This is consistent with other measurements which indicate that the carrier recovery loop has a dominant influence on overall performance. Again, the curves are not 5 dB apart. However, they were obtained in a region of even more severe data feedback error rate reduction and front-end compression, where $E_b/N_o \stackrel{<}{\sim} -7$ to $+1$ dB. Moreover, performance extrapolation above and below the bandwidths measured is not plausible.
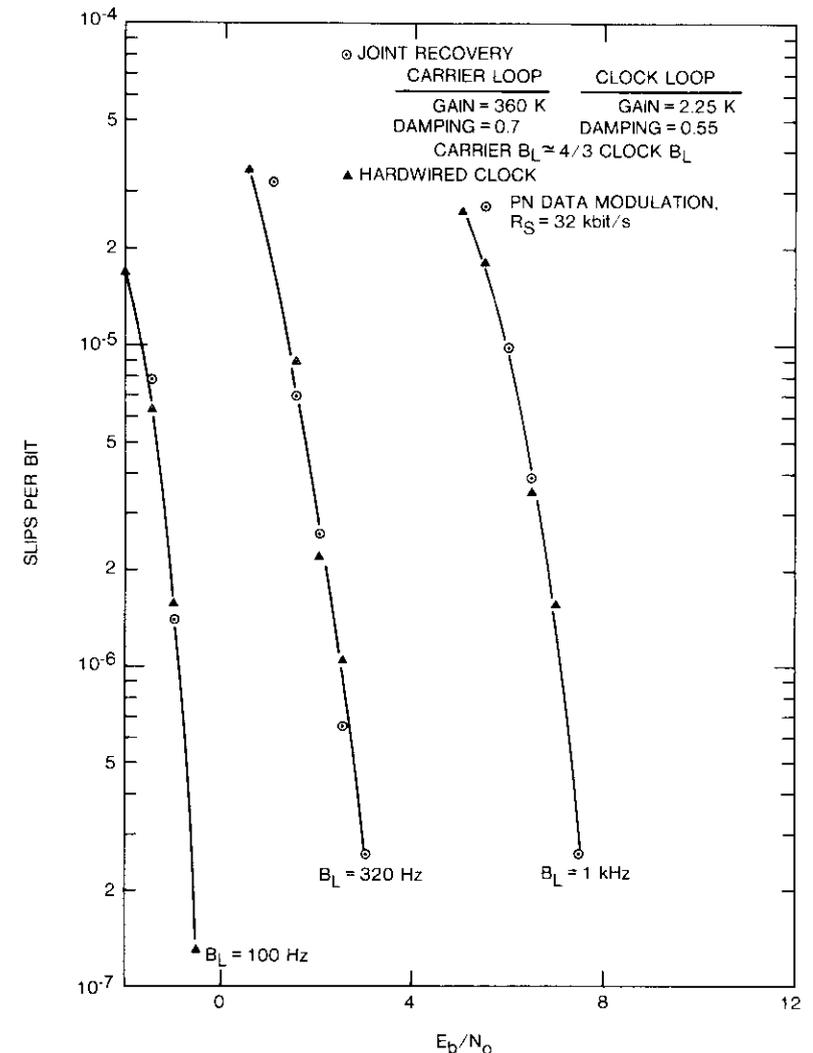
Figure 13. *QPSK Carrier Recovery Cycle Skipping vs E_b/N_o and Loop Bandwidth, B_L*

Bit error rate performance

In general, when coherent carrier and clock references with relatively low jitter are available, the degradation in BER is dominated by intersymbol
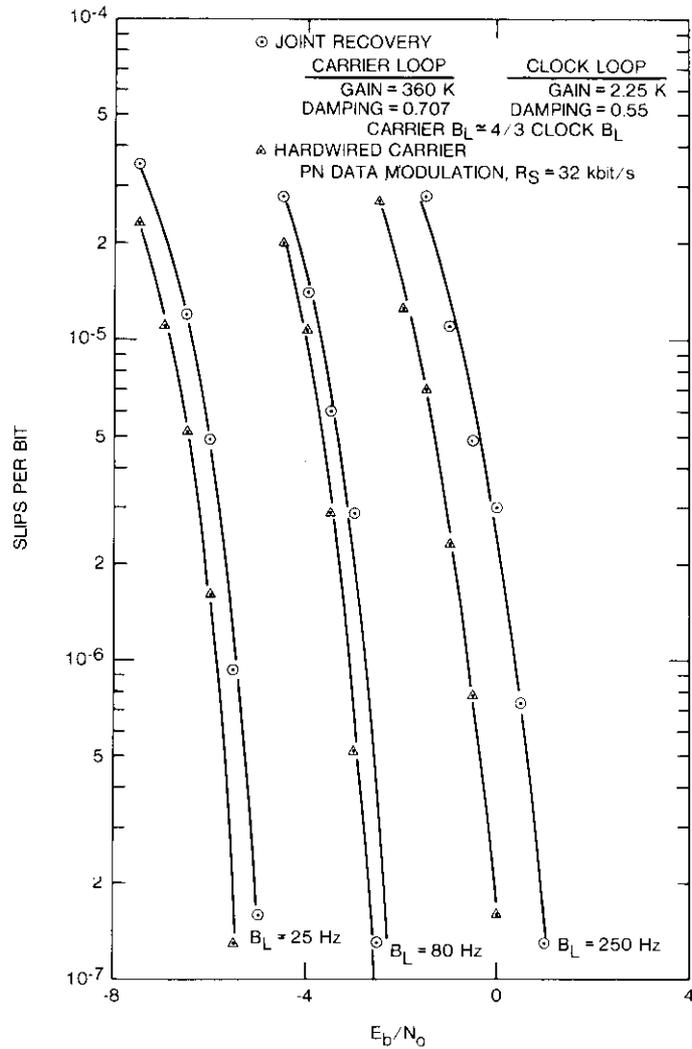
Figure 14. *QPSK Clock Recovery Cycle Skipping vs E_b/N_o and Loop Bandwidth, B_L*

interference (ISI). As it was not feasible to implement ideal square-root Nyquist data filters in this application, the BER data presented in this section should not be used to assess the performance of the joint estimator-detector. These data are included for completeness only.

Curves of BER performance *vs* carrier recovery loop bandwidth and E_b/N_o are presented in Figure 15. Curve 1 for the case of hardwired carrier and clock exhibits a degradation of 0.3 to 0.6 dB from ideal QPSK BER performance. Computer simulations indicate a nominal degradation of about 0.3 dB for
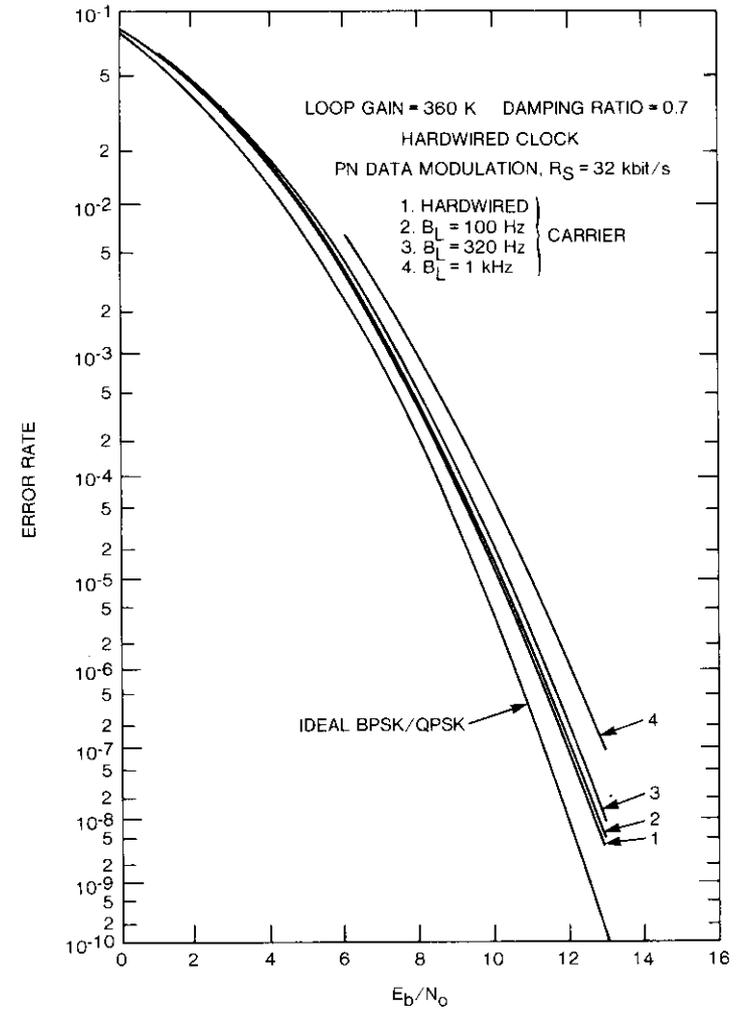
Figure 15. *QPSK Bit Error Rate vs E_b/N_o and Carrier Recovery Loop Bandwidth, B_L*

the data filtering previously described. Thus, the implementation margin is small and consistent with predicted results. The range of variation in BER performance between the hardwired case and the outermost curve where the carrier recovery loop bandwidth $B_L = 1$ kHz, is about 1 dB. Degradation between the $B_L = 320$ Hz and 1 kHz curves is larger than expected due to the excess delay of $3.9T_s$ with data feedback, which causes the closed-loop phase margin to diminish with increasing carrier recovery bandwidth. Hence, additional phase compensation is necessary at the larger bandwidths to achieve better performance.

To approach hardwired performance requires that $B_L \lesssim 100$ Hz; or equivalently, the ratio $R_s/B_L \gtrsim 300$. For this case, the loop filter averages the carrier phase estimate over 300 symbols. The breaks in the curve at low E_b/N_o correspond to those points beyond which cycle skipping is severe enough to affect BER.

In assessing BER performance vs the clock recovery loop bandwidth, no discernible difference in error rate was observed for B_L ranging from 25 to 800 Hz. Furthermore, the error rate was not appreciably different from that with hardwired carrier and clock. Thus, recovering the clock had a negligible impact on BER performance for the equalized 6-pole Butterworth data filtering utilized in the detector.

The BER curves in Figure 16 are for joint carrier and clock recovery. Note that they are nearly identical to those for recovered carrier with hardwired clock in Figure 15. Hence, joint recovery did not cause any significant degradation in error rate for the type of data filtering employed.

Conclusions

It has been shown that in the steady state, the digital feedback QPSK receiver structure developed in this paper is capable of providing a coherent carrier reference with phase jitter close to that obtained by a PLL operating on an unmodulated carrier, which is the best achievable. In particular, with PN data modulation, a degradation of only $0.1°$ rms relative to an unmodulated carrier over most of the operable range of loop bandwidths and S/NS was encountered. Because the suboptimum digital feedback structure does not subtract out the I and Q data pattern noise completely (see Figure 7a), it is expected that with analog feedback this degradation could be effectively eliminated.

The carrier phase jitter performance is achieved primarily for two reasons. Postdetected (sampled) decision feedback, which is the best available information in terms of S/N, is utilized in carrier phase estimate processing; and the references developed in each of the estimators (amplitude, phase,
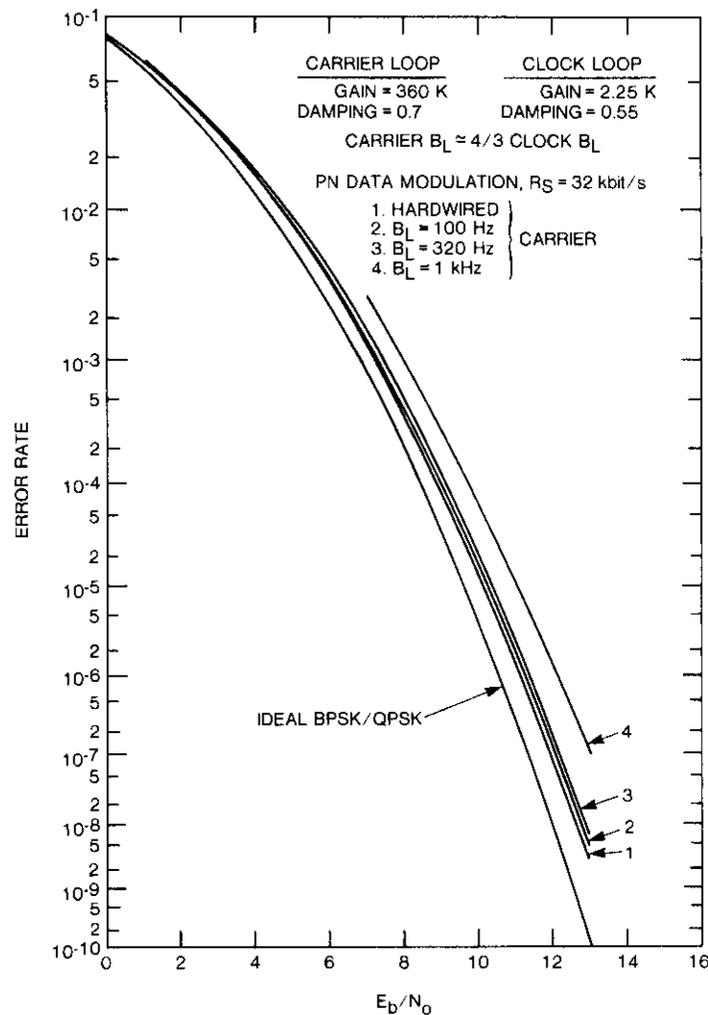
Figure 16. *QPSK Bit Error Rate vs E_b/N_o and Joint Recovery Loop Bandwidth, B_L*

and timing) coherently aid the others. This arrangement differs from a Costas or squaring loop, where knowledge of the symbol timing and postdetected data is not exploited. In addition, the postdetection decision feedback carrier recovery technique, which utilizes a signal-plus-noise times a signal operation

to remove the modulation, has a large S/N advantage over frequency quadrupling, whose inherent loss is about 12 dB.

The excess delay of approximately four symbols in the carrier recovery loop, due to postdetection data feedback, caused instability at the largest closed-loop bandwidth tested, B_L = 1 kHz. The delay reduces the loop damping ratio, which results in a noise equivalent bandwidth larger than intended. Future work should determine whether more elaborate loop phase compensation or a suboptimum, minimum-delay data filter path for carrier recovery could alleviate this problem. Lastly, recall that analog feedback would contribute additional delay.

The rms phase jitter measured in the clock recovery loop exactly matched predicted PLL performance curves for B_L = 32 and 100 Hz when "10" data modulation was utilized. Measured performance for B_L = 320 Hz and 1 kHz was better than expected due to control voltage compression and modulation rate limitations, which effectively reduced the design bandwidth.

With PN modulation, clock recovery phase jitter was significantly poorer than that at the same loop S/N as "10" modulation. This was attributed to the zero crossing pattern jitter on the Nyquist filtered data waveshape at the input to the loop. Future work could substantiate this contention if data filtering with minimal zero crossing dispersion, such as a 100-percent raised cosine response, were utilized.

The digital feedback implementation furnishes a coherent symbol timing reference with phase jitter close to that of a PLL operating on an unmodulated carrier, because differentiation is performed on the noise-free postdetected data, as opposed to other carrier phase and bit timing recovery schemes which differentiate signal plus noise. In addition, since the phase detector comprises input and feedback data multiplication, input noise is not injected into the loop when there are no data transitions. Hence, loop S/N is not degraded.

The cycle skipping threshold occurs at about E_b/N_o = -1 dB for the carrier recovery loop, and around E_b/N_o = -3 dB in the clock loop, for carrier and clock recovery loop bandwidths of 100 and 80 Hz, respectively. These S/Ns correspond to BERs just below 10^{-1}. Experience indicates that a practical threshold for data-aided carrier and clock recovery occurs when BER $\approx 10^{-1}$. As demonstrated previously, this error rate causes an effective reduction in the recovery loop S/Ns of 1.9 dB; and if there is a significant probability of a higher instantaneous error rate, the attendant loss in S/N is much greater.

The impact on BER performance of the joint estimator-detector was dominated by carrier recovery. No difference in error rate was observed in going from hardwired to recovered clock for any of the loop bandwidths

tested, B_L = 25 through 800 Hz. Thus, even though the carrier and clock recovery loops had the same S/N, the error rate was much more sensitive to carrier phase jitter.

To get within 0.1 dB of the hardwired carrier and clock BER performance required that $B_L \lesssim 100$ Hz; which for R_s = 32 kbit/s implies that the carrier loop filter must average over 300 symbols. An additional 0.1 dB was experienced for B_L = 320 Hz. However, with B_L = 1 kHz, the error rate degraded sharply by approximately 0.7 dB at 10^{-5}. This is again due to the excess delay incurred with data feedback.

To summarize, the fundamental utility of the optimum joint estimator-detector is to provide coherent references for digital data detection with minimum corruption by thermal and/or pattern noise contributions. In this manner, BER degradation is minimized, cycle skipping and loss of lock occur at lower S/Ns, and operation may be extended to higher uncoded error rates.

The measurements have confirmed that, for relatively narrow recovery loop bandwidths, the remodulator described in Appendix A is not required for QPSK operation. However, when multilevel signaling formats such as 8-ary PSK or 4-ary PAM are used, the remodulator may be needed. Moreover, the techniques used in this study to extract amplitude, phase, and timing information are not restricted to QPSK, but are also adaptable to other digital modulation formats. Thus, the receiver structure presented here can provide a general framework for near-optimum estimation and detection of the parameters associated with digital data transmission.

Acknowledgment

References

[1] J. P. Costas, "Synchronous Communications," *Proceedings of the IRE*, December 1956, pp. 1713–1718.

[2] F. D. Natali and W. J. Walbesser, "Phase-Locked Loop Detection of Binary PSK Signals Utilizing Decision Feedback," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-5, No. 1, January 1969, pp. 83–90.

[3] W. C. Lindsey and M. K. Simon, "Data-Aided Carrier Tracking Loops," *IEEE Transactions on Communications Technology*, Vol. COM-1, No. 2, April 1971, pp. 157–168.

[4] C. J. Wolejsza, "State of the Art of PSK Demodulation Techniques," Second International Conference on Digital Satellite Communications, Paris, 1972, *Proceedings*, pp. 154–163.

[5] Y. Matsuo et al., "A Study on Carrier and Bit-Timing Recovery for Ultrahigh-Speed PSK-TDMA Systems," Second International Conference on Digital Satellite Communications, Paris, 1972, *Proceedings*, pp. 277–286.

[6] F. M. Gardner, "Comparison of QPSK Carrier Regeneration Circuits for TDMA Application," IEEE International Communications Conference, Minneapolis, 1974, *Proceedings*, p. 43B-1.

[7] J. Snyder, "Digital Phase-Locked Loop Finds Clock Signal in Bit Stream" *Electronics*, Vol. 52, No. 18, August 30, 1979, pp. 126–130.

[8] L. E. Franks, "Carrier and Bit Synchronization in Data Communications—A Tutorial Review," *IEEE Transactions on Communications*, Vol. COM-28, No. 8, August 1980, pp. 1107–1121.

[9] H. Kobayashi, "Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems," *IEEE Transactions on Communications Technology*, Vol. COM-19, No. 3, June 1971, pp. 268–280.

[10] D. Falconer and J. Salz, "Optimal Reception of Digital Data over the Gaussian Channel With Unknown Delay and Phase Jitter," *IEEE Transactions on Information Theory*, Vol. IT-23, No. 1, January 1977, pp. 117–126.

[11] U. Mengali, "Joint Phase and Timing Acquisition in Data Transmission," *IEEE Transactions on Communications*, Vol. COM-25, No. 10, October 1977, pp. 1174–1885.

[12] M. H. Meyers and L. E. Franks, "Joint Carrier Phase and Symbol Timing Recovery for PAM Systems," *IEEE Transactions on Communications*, Vol. COM-28, No. 8, August 1980, pp. 1121–1129.

[13] J. J. Poklemba, "Concurrent Carrier and Clock Synchronization for Data Transmission Systems," U.S. Patent 4,419,759.

[14] W. Bennet and J. Davey, *Data Transmission*, New York: McGraw-Hill, 1965, Section 5-3.

[15] J. Poklemba, "An Approximation Technique for Realizing the Reciprocal sin x/x Frequency Response," *COMSAT Technical Review*, Vol. 12, No. 1, Spring 1982, pp. 121–156.

[16] F. Gardner, *Phaselock Techniques*, New York: John Wiley and Sons, 1966, Section 3-1, 2.

Appendix A. Joint MMSE Estimator-detector derivation

A textbook receiver structure

If the processes A, $\theta(t)$, $\tau(t)$, and $n(t)$ as defined in Table 1 are considered jointly Gaussian, a closed-form solution, which specifies the receiver structure, can be

extracted from estimation theory. Although in general, the processes do not satisfy this criterion, it has been found that MMSE estimators which are based on the Gaussian assumption work well for a broad class of problems.

To begin, the multidimensional estimator equation for jointly Gaussian random processes in additive white noise with power density $N_o/2$ is called upon [A-1, Part I].

$$\underline{\hat{a}}(t) = \frac{2}{N_o} \int_0^T K_{\underline{a}}(t,u)\, D[u,\underline{\hat{a}}(u)]\, \{r(u) - s[u,\underline{\hat{a}}(u)]\}\, du \quad . \tag{A-1}$$

This expression is contingent upon the signals being continuously differentiable. The ^s indicate where process estimates derived in the receiver must be substituted.

The vector and matrix elements that make up equation (A-1) are defined as follows. The estimate vector is selected as

$$\underline{\hat{a}}(t) = \begin{bmatrix} \hat{A} \\ \hat{\theta}(t) \\ \hat{\tau}(t) \end{bmatrix} \quad . \tag{A-2a}$$

Assuming that the processes are statistically independent, its covariance matrix has entries only along the diagonal:

$$K_{\underline{a}}(t,u) \triangleq E[\underline{a}(t)\underline{a}^T(u)] = \begin{bmatrix} \sigma_A^2 & 0 & 0 \\ 0 & K_\theta(t,u) & 0 \\ 0 & 0 & K_\tau(t,u) \end{bmatrix} \quad . \tag{A-2b}$$

To maintain continuity in the analysis, the parameter, A, will be treated as a random variable. Later it will degenerate to the status of unknown but nonrandom. $K_\theta(t,u)$ and $K_\tau(t,u)$ are the covariance functions of $\theta(t)$ and $\tau(t)$, respectively.

The derivative vector, which takes the partial derivative of the signal with respect to the function to be estimated, is

$$D[t,\underline{\hat{a}}(t)] = \begin{bmatrix} \dfrac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{A}} \\ \dfrac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{\theta}(t)} \\ \dfrac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{\tau}(t)} \end{bmatrix} \quad . \tag{A-2c}$$

After substitution with the QPSK signal format of equation (3), the derivatives in equation (A-2c) are more specifically written as

$$D[t,\underline{\hat{a}}(t)] = \begin{bmatrix} \hat{i}(t,\hat{\tau}) \cos [\omega_c t + \hat{\theta}(t)] \\ + \hat{q}(t,\hat{\tau}) \sin [\omega_c t + \hat{\theta}(t)] \\ \hat{A}\{\hat{q}(t,\hat{\tau}) \cos [\omega_c t + \hat{\theta}(t)] \\ - \hat{i}(t,\hat{\tau}) \sin [\omega_c t + \hat{\theta}(t)]\} \\ \hat{A}\{\hat{i}(t,\hat{\tau}) \cos [\omega_c t + \hat{\theta}(t)] \\ + \hat{q}(t,\hat{\tau}) \sin [\omega_c t + \hat{\theta}(t)]\} \end{bmatrix} \qquad \text{(A-2d)}$$

Equations (A-1) and (A-2) provide an idealized sketch for the MMSE joint estimator. The individual estimators are separated as follows:

$$\hat{A} = \frac{2\sigma_A^2}{N_o} \int_0^T \{\hat{i}(u,\hat{\tau}) \cos [\omega_c u + \hat{\theta}(u)]$$

$$+ \hat{q}(u,\hat{\tau}) \sin [\omega_c u + \hat{\theta}(u)]\}\{r(u) - s[u,\underline{\hat{a}}(u)]\} \, du \qquad \text{(A-3a)}$$

$$\hat{\theta}(t) = \frac{2\hat{A}}{N_o} \int_0^T K_\theta(t,u) \{\hat{q}(u,\hat{\tau}) \cos [\omega_c u + \hat{\theta}(u)]$$

$$- \hat{i}(u,\hat{\tau}) \sin [\omega_c u + \hat{\theta}(u)]\}\{r(u) - s[u,\underline{\hat{a}}(u)]\} \, du \qquad \text{(A-3b)}$$

$$\hat{\tau}(t) = \frac{2\hat{A}}{N_o} \int_0^T K_\tau(t,u)\{\dot{\hat{i}}(u,\hat{\tau}) \cos [\omega_c u + \hat{\theta}(u)]$$

$$+ \dot{\hat{q}}(u,\hat{\tau}) \sin [\omega_c u + \hat{\theta}(u)]\}\{r(u) - s[u,\underline{\hat{a}}(u)]\} \, du \quad . \qquad \text{(A-3c)}$$

All three estimators operate on the difference between the received input and the remodulated estimate of the signal component, that is, on $r(t) - s[t,\underline{\hat{a}}(t)]$. (This procedure requires that the entire transmitted signal be reconstructed from the recovered information.) In addition, each estimator coherently translates the IF difference signal down to baseband with recovered carrier references in phase quadrature, $\cos [\omega_c t + \hat{\theta}(t)]$ and $\sin [\omega_c t + \hat{\theta}(t)]$. This common processing is shown in block diagram form in Figure A-1. The resultant baseband signals, $r_i(t)$ and $r_q(t)$, are complicated. Their signal components will be evaluated in detail in a subsequent section. For the present, they can be viewed in the steady state as the transmitted data waveforms, $i(t,\tau)$ and $q(t,\tau)$, embedded in additive noise.

The baseband amplitude estimator is given in Figure A-2. Its input signals, $r_i(t)$ and $r_q(t)$, are multiplied by the respective reconstructed analog data waveforms $\hat{i}(t,\hat{\tau})$ and $\hat{q}(t,\hat{\tau})$. These are derived by filtering the digital outputs from the data detector, $\hat{I}$ and $\hat{Q}$, with the same characteristic that was utilized in the transmitter. The sum of the products, $r_A(t)$, is then integrated and scaled to yield the amplitude estimate, $\hat{A}$. Because it is the best S/N replica of the transmitted data, the estimator equations call for postdetection decision feedback.
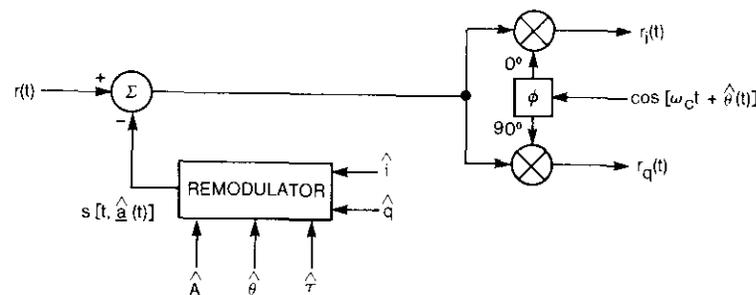
Figure A-1. *Common Estimator Processing*
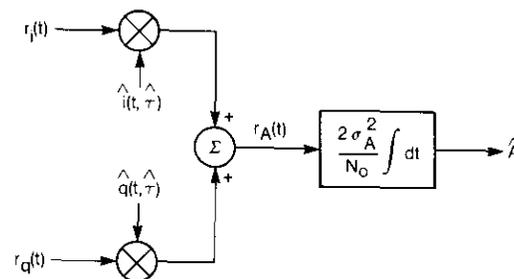
Figure A-2. *Baseband Amplitude Estimator*

For the carrier phase estimator of Figure A-3, $r_i(t)$ and $r_q(t)$ are cross-multiplied by $\hat{q}(t,\hat{\tau})$ and $\hat{i}(t,\hat{\tau})$, respectively. The difference of their products, $r_\theta(t)$, is filtered with the covariance function to give the phase estimate, $\hat{\theta}(t)$. When $\hat{\theta}(t)$ is injected
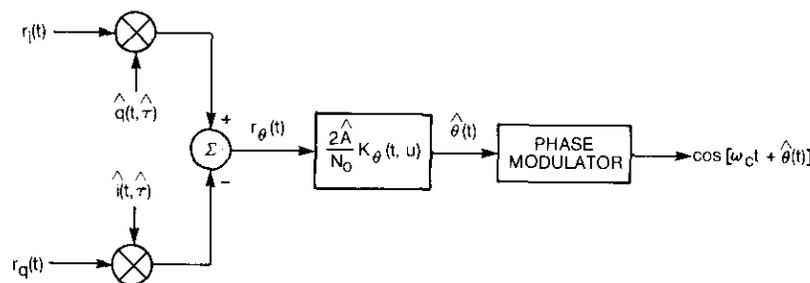
Figure A-3. *Baseband Carrier Phase Estimator*

into a phase modulator, a reconstructed, coherent carrier is fashioned, and the processing is complete.

Analogously, the timing estimate $\hat{\tau}(t)$ in Figure A-4 is generated when $r_i(t)$ and $r_q(t)$ are multiplied by the time derivatives of their respective data waveforms, $\dot{i}(t,\hat{\tau})$ and $\dot{q}(t,\hat{\tau})$. Summing, filtering, and driving a phase modulator yield a reconstructed version of the symbol timing clock, $\hat{R}_s$, which was used in the transmitter. Recall that with the receiver implementations of References 9 and 10, differentiation in the timing loop is performed instead on the predetected data.
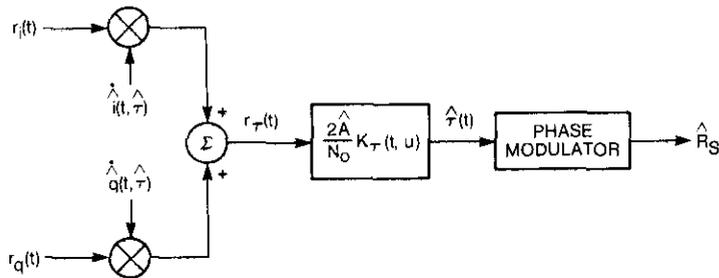
Figure A-4. *Baseband Symbol Timing Estimator*

As the previous discussion implies, many of the operations required in the estimator portion of the receiver are common. Furthermore, if decisioned-data waveforms are available, the estimator framework is well defined. Similarly, given coherent carrier and timing references, the structure of the QPSK detector is well established [A-1, Section 4.2.1]–[A-3]. Integrating the estimator and detector results in the implementation of Figure A-5.

With the estimator portions of the structure already described, consideration must now be given to how the input signal, $r(t)$, is processed for detection. With the coherent carrier references derived in the phase estimator, the in-phase and quadrature signal components are translated down to baseband, as indicated in the outermost paths of Figure A-5. Decisions are made after the signal passes through the receive data filter, $H_r(t - u)$, and is sampled with the reconstructed symbol clock, $\hat{R}_s$, from the timing estimator. Outputs $\hat{I}$ and $\hat{Q}$ are the post-detection digital information bits. The lowercase letters used throughout this paper denote analog waveshapes. This receiver is a textbook idealization of a jointly optimal estimator-detector. The following sections illustrate the necessary modifications for transposing it into a practical structure.

Data-aided feedback waveforms

The primary disadvantage with the data-aided carrier and clock recovery technique developed in this study is that analog delay must be inserted in the estimator path to compensate for the delay encountered in the detection process. This relationship is illustrated in Figure A-6 with an arbitrary data sequence shown as it progresses
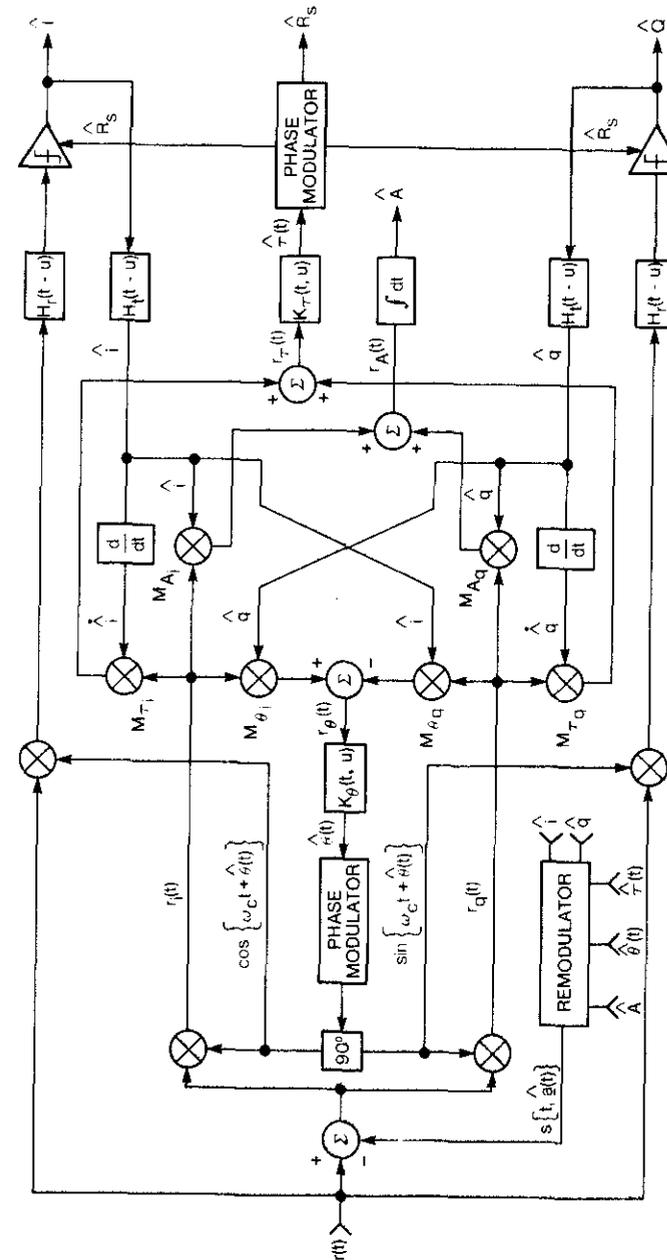
Figure A-5. *Idealized Estimator-Detector Receiver*
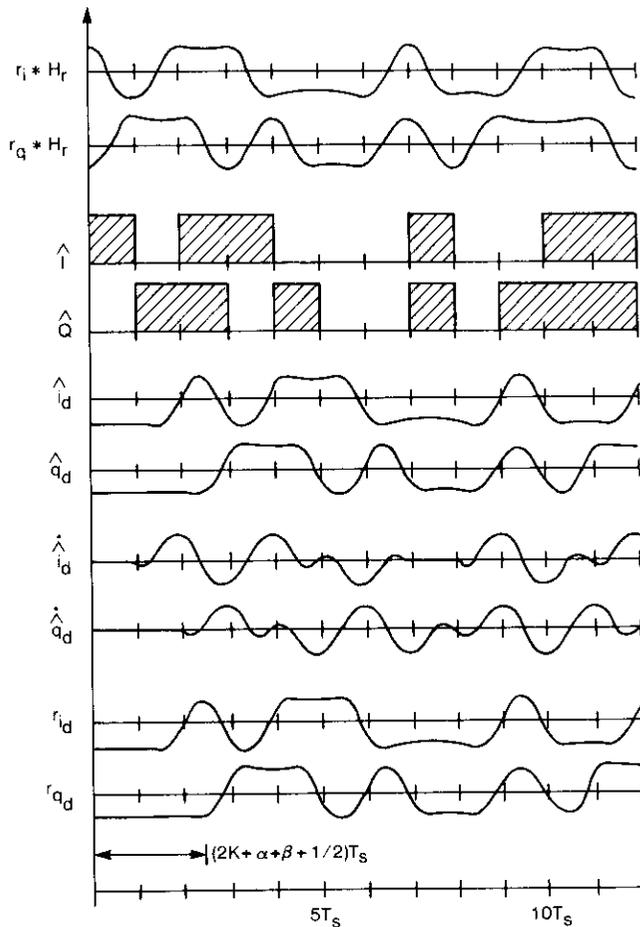
Figure A-6. *Steady-State Analog Feedback Waveforms*

through the system. The i and q subscripted waveforms at the top of the figure represent the receive data filter outputs under "ideal" conditions (when the recovery loops are perfectly synchronized and no detection errors are encountered). Decisions are made via sampling at the time of maximum S/N (i.e., maximum eye opening), which causes the digitized detected data, $\hat{I}$ and $\hat{Q}$, to be delayed by half a symbol time. Their analog counterparts, $\hat{i}_d$ and $\hat{q}_d$, are further shifted because of refiltering. The same shifting occurs with the derivatives $\dot{\hat{i}}_d$ and $\dot{\hat{q}}_d$, where the subscript "d" denotes delay. For alignment with the feedback, the input signals must be delayed, as indicated in the last two frames. The manner in which these waveforms are combined to yield estimates for A, $\theta(t)$, and $\tau(t)$ will be described in the sections that follow.

Loop control signals

To expedite the analysis, it will be helpful to write the multidimensional estimator of equations (A-1) and (A-2) in the following form. First, the estimated amplitude is

$$\hat{A} = \frac{2}{N_o} \int_0^T \sigma_A^2 \, r_A(u) \, du \qquad (A\text{-}4)$$

where

$$r_A(t) \triangleq s_A(t) - \hat{s_A}(t) + n_A(t)$$

$$r_A(t) = \frac{\partial s[t, \hat{\underline{a}}(t)]}{\partial \hat{A}} \{s(t, \underline{a}(t)) - s[t, \hat{\underline{a}}(t)] + n(t)\} \quad .$$

Next, the estimated carrier phase is

$$\hat{\theta}(t) = \frac{2}{N_o} \int_0^T K_\theta(t, u) \, r_\theta(u) \, du \qquad (A\text{-}5)$$

where

$$r_\theta(t) \triangleq s_\theta(t) - \hat{s_\theta}(t) + n_\theta(t)$$

$$r_\theta(t) = \frac{\partial s[t, \hat{\underline{a}}(t)]}{\partial \hat{\theta}(t)} \{s[t, \underline{a}(t)) - s[t, \hat{\underline{a}}(t)] + n(t)\} \quad .$$

Finally, the estimated symbol timing of the data modulation is

$$\hat{\tau}(t) = \frac{2}{N_o} \int_0^T K_\tau(t, u) \, r_\tau(u) \, du \qquad (A\text{-}6)$$

where

$$r_\tau(t) \triangleq s_\tau(t) - \hat{s_\tau}(t) + n_\tau(t)$$

$$r_\tau(t) = \frac{\partial s[t, \hat{\underline{a}}(t)]}{\partial \hat{\tau}(t)} \{s(t, \underline{a}(t)) - s[t, \hat{a}(t)] + n(t)\} \quad .$$

It is apparent that each of the three estimates is ultimately extracted by a simple filtering operation. However, their inputs, $r_{\underline{a}}(t)$, usually entail nonlinear processing. That is, both the received waveform, $s[t, \underline{a}(t)] + n(t)$, and its remodulated signal estimate, $s[t, \hat{\underline{a}}(t)]$, are multiplied by the derivative vector. The treatment that follows focuses on the input signal element in $r(t)$, along with that developed by the remodulator. The influence of the input and remodulated signals will be analyzed separately for each recovery loop.

AMPLITUDE ESTIMATION

In this section, the low-pass loop control signals required for amplitude estimation will be analyzed. The processing of the input signal component in $r_A(t)$ is developed

by utilizing the particular QPSK waveshapes from equation (3) and omitting double carrier frequency terms, as

$$s_A(t) = \frac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{A}} \; s[t,\underline{a}(t)] \; \Big|_{LP}$$

$$= \frac{A}{2} \{ [i\hat{i} + q\hat{q}] \cos(\theta - \hat{\theta}) + [q\hat{i} - i\hat{q}] \sin(\theta - \hat{\theta}) \} \quad . \tag{A-7}$$

The time dependence of the estimates and the "d" subscripts have been suppressed to simplify the notation. When the estimates are good, the cosine term in equation (A-7) dominates. Thus, the net effect has been to coherently translate the input signal down to baseband, where an operation like magnitude squared is carried out on the i and q data sequences to extract the unknown amplitude variable, A.

Analogously, the processing of the remodulated signal component in $r_A(t)$ is

$$\hat{s_A}(t) = \frac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{A}} \; s[t,\underline{\hat{a}}(t)] \; \Big|_{LP}$$

$$= \frac{\hat{A}}{2} \{ [\hat{i}^2 + \hat{q}^2] \cos(\hat{\theta} - \hat{\theta}) + [\hat{q}\hat{i} - \hat{i}\hat{q}] \sin(\hat{\theta} - \hat{\theta}) \}$$

$$= \frac{\hat{A}}{2} [\hat{i}^2 + \hat{q}^2] \quad . \tag{A-8}$$

In this instance, a true magnitude squared operation results regardless of the quality of the estimate. Also, because no contributions from carrier terms exist, the effect of the remodulator in amplitude estimation can be accounted for with simple baseband processing.

With the arbitrary data sequences of Figure A-6, the input component of the amplitude loop control signal, $s_A(t)$, is plotted in Figure A-7a. Its DC content is a measure of the relative signal amplitude, A. In the event of a carrier phase error, $\theta \neq \hat{\theta}$, the magnitude of the cosine (coherent) term in $s_A(t)$ will always be reduced. This is evident by inspection of equation (A-7). Analogously, symbol timing errors, $\tau \neq \hat{\tau}$, cause some excursions of the quantity $[i(t,\tau) \hat{i}(t,\hat{\tau}) + q(t,\tau) \hat{q}(t,\hat{\tau})]$ to become negative, again lowering the magnitude of $s_A(t)$.

CARRIER PHASE ESTIMATION

The processing of the input signal component in the phase estimation loop is described by

$$s_\theta(t) = \frac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{\theta}(t)} \; s[t,\underline{a}(t)] \; \Big|_{LP}$$

$$= \frac{A\hat{A}}{2} \{ [i\hat{q} - q\hat{i}] \cos(\theta - \hat{\theta}) + [i\hat{i} + q\hat{q}] \sin(\theta - \hat{\theta}) \} \quad . \tag{A-9}$$
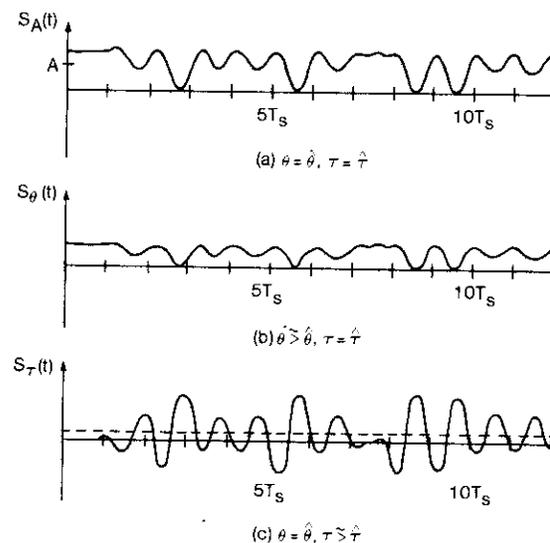
Figure A-7. *Steady-State Analog Feedback Control Signals*

If the estimates are good, the pattern noise, $[i\hat{q} - q\hat{i}]$, cancels out; whereas the DC component of $[i\hat{i} + q\hat{q}]$ allows the sine of the phase error to drive the loop in a conventional fashion.

Similarly, the remodulated signal component is written as

$$\hat{s_\theta}(t) = \frac{\partial s(t,\underline{\hat{a}}(t))}{\partial \hat{\theta}(t)} \; s[t,\underline{\hat{a}}(t)] \; \Big|_{LP}$$

$$= \frac{\hat{A}^2}{2} \{ [\hat{i}\hat{q} - \hat{q}\hat{i}] \cos(\hat{\theta} - \hat{\theta}) + [\hat{i}^2 + \hat{q}^2] \sin(\hat{\theta} - \hat{\theta}) \}$$

$$= 0 \quad . \tag{A-10}$$

Since equation (A-10) is exactly zero, the remodulator is irrelevant to phase estimation for cases in which double carrier frequency terms may be ignored.

The net loop control signal, $s_\theta(t)$, is sketched in Figure A-7b, assuming a small carrier phase offset. Note that in the absence of a phase error, the signal would be zero. The envelope of $s_\theta(t)$ has the same magnitude squared characteristic as the amplitude loop, scaled by the sine of the phase error. As before, if the estimates from the other loops (amplitude and timing) are not accurate, the drive signal is reduced. This is confirmed by an examination of equation (A-9).

SYMBOL TIMING ESTIMATION

The processing of the input signal component for symbol timing is

$$
\hat{s_\tau}(t) = \frac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{\tau}(t)}\, s[t,\underline{a}(t)] \Bigg|_{LP}
$$

$$
= \frac{A\hat{A}}{2} \{ [i\hat{i} + q\hat{q}] \cos(\theta - \hat{\theta}) + [q\hat{i} - i\hat{q}] \sin(\theta - \hat{\theta}) \} \quad . \qquad \text{(A-11)}
$$

When the estimates are good, the cosine term in equation (A-11) is dominant. Thus, the effect of the processing is to coherently translate the input signal down to baseband; where a second phase detector, constituted by $[i(t,\tau)\,\hat{i}(t,\hat{\tau}) + q(t,\tau)\,\hat{q}(t,\hat{\tau})]$, establishes a symbol timing control signal. In other words, whenever a timing error exists, $\tau \neq \hat{\tau}$, $[i\hat{i} + q\hat{q}]$ generates a nonzero DC component, which is the driving force in the loop.

The contribution of the remodulator to symbol timing is

$$
\hat{s_\tau}(t) = \frac{\partial s[t,\underline{\hat{a}}(t)]}{\partial \hat{\tau}(t)}\, s[t,\underline{\hat{a}}(t)] \Bigg|_{LP}
$$

$$
= \frac{\hat{A}^2}{2} \{ [\hat{i}\hat{i} + \hat{q}\hat{q}] \cos(\hat{\theta} - \hat{\theta}) + [\hat{q}\hat{i} - \hat{i}\hat{q}] \sin(\hat{\theta} - \hat{\theta}) \}
$$

$$
= \frac{\hat{A}^2}{2} [\hat{i}(t,\hat{\tau})\,\hat{i}(t,\hat{\tau}) + \hat{q}(t,\hat{\tau})\,\hat{q}(t,\hat{\tau})] \quad . \qquad \text{(A-12)}
$$

Because of continuity requirements and even time symmetry with the Nyquist-filtered QPSK data waveforms, $i(t,\tau)$ and $q(t,\tau)$, equation (A-12) has no DC component. The frequency content of $\hat{s_\tau}(t)$ is primarily around the symbol rate, to cancel out the pattern fluctuations in the input control signal, $s_\tau(t)$. This is evident when $s_\tau(t)$ is drawn in Figure A-7c with a small timing offset. It consists of a DC component proportional to the timing error, $\tau \neq \hat{\tau}$, which is the necessary drive signal in the loop. Again, its magnitude is reduced when errors are encountered in the other estimation paths. The signal $s_\tau(t)$ is also made up of sinusoidal fluctuations that occur at the symbol rate. These are analogous to the double frequency terms that were ignored in the carrier loop. The fluctuations are characterized by three possible conditions: no data transitions, transitions in one channel, or transitions in both channels. They correspond to the zero, and half or full amplitude excursions, respectively, in Figure A-7c. Since the loop response from the phase detector error signal to VCO output phase is typically an integrator, frequency content around the symbol rate is effectively filtered out. Therefore, the remodulator may also be discarded for timing recovery.

Estimate filtering

From a practical standpoint, it is desirable to substitute VCOs for the phase modulators of Figure A-5. Since the VCO is modeled as an integrator in a phase parameter feedback system, the transfer function of the loop filter must be modified by $j\omega$. Consequently, the filters in the phase and timing loops deliver the time derivative of the estimate to the VCO.

In addition, the optimum filters for estimating the amplitude, carrier phase, and symbol timing, as depicted in Figure A-5, are neither "stationary" nor "realizable." That is, they may be time varying, and must process the entire past and future of the input, respectively. The latter situation is typically approximated by filtering with delay. However, since the carrier phase and symbol timing filters are utilized in a feedback context, this is not possible. As a result, each must be replaced with a realizable counterpart, whose relationship to the original filter has been determined through the imposition of MMSE optimality constraints [A-1, Parts I and II], [A-4]–[A-7]. That is, when the processes are jointly Gaussian, the estimates have the smallest output jitter attainable with realizable filtering.

In the amplitude estimator, the optimal filter is merely an integrator. The optimum filters for the carrier phase and symbol timing loops are, respectively,

$$
F_\theta'(s) = \frac{s}{\hat{\gamma}} \{ \sqrt{\gamma}\,[S_\theta(-s^2) + 1/\gamma]^+ - 1 \} \qquad \text{(A-13)}
$$

$$
F_\tau'(s) = \frac{s}{\hat{\gamma}(\pi R_s)} \{ \sqrt{\gamma(\pi R_s)}\,[S_\tau(-s^2) + 1/\gamma(\pi R_s)]^+ - 1 \} \qquad \text{(A-14)}
$$

where $\gamma = 2A^2/N_o$ and $\hat{\gamma} = 2A\hat{A}/N_o$. The " $+$ " operator requires that the portion of the spectrum that yields an impulse response for positive time be factored out. Observe that $F_\theta'(s)$ is a function of both the spectrum of the phase process and the S/N ratio. Moreover, the loop filter seeks both to minimize the jitter on the estimate and to maintain tracking ability.

It should be emphasized that the filters in the carrier phase and symbol timing loops are intended for steady state tracking. Their acquisition behavior, in general, is not optimum because the Gaussian and small loop error assumptions upon which they were derived may not be warranted.

A practical receiver implementation

It will now be useful to summarize the results of the previous sections with respect to their effects on the transformation of the idealized receiver in Figure A-5 into a more viable structure, and also to identify the accompanying restrictions. The utilization of VCOs and realizable (causal) loop filters does not create significant limitations. However, the removal of the remodulator and insertion of analog delay elements should be qualified. With regard to the remodulator, its effect was incorporated into the amplitude loop by modifying the DC gain; and when A degenerated to the

status of an unknown, nonrandom variable, it made no contribution. With phase estimation, if double carrier frequency terms were ignored, the remodulator signal component cancelled out completely. For symbol timing, it was found that the control signal term resulting from the remodulator had little low frequency content; thus it was integrated out. Moreover, since the remodulator does not exert a controlling influence, it is possible to combine the detector and estimator into the more compact arrangement depicted in Figure A-8.

Recall that analog delay elements were required to balance out the delay caused by postdetection feedback. These delays are listed above their respective modules in Figure A-8. The variable K is an integer, and α, β, and $\frac{1}{2}$ are fractional symbol time multiples (e.g., $K = 3$, $\alpha = \beta = 0.4$). The amount of delay accumulated in each estimation loop is $(2K + \alpha + \beta + \frac{1}{2})T_s$. Depending on the complexity of the data filtering, this number may vary from T_s to $10T_s$; and the excess phase it contributes ranges from negligible to prohibitive, as the tracking loop bandwidth is increased.

The optimality of this structure is brought out heuristically, when the detector and estimator are viewed in the following way. An ideal low-pass matched filter detector requires both a coherent carrier to translate the received signal to baseband, and a symbol timing reference for sampling. In general, access to the transmit oscillators is not feasible. Therefore, reconstructed replicas, based on estimates derived from all available information, are the best substitutes. It might be thought that the optimum way to recover the carrier would be to "unmodulate" it and inject the resultant into a PLL. This is carried out in reverse order in the joint estimator-detector, where data multiplication effects an absolute value operation on the modulation at baseband. Because the transmit data, clock, and amplitude levels are not known, the signals used for carrier recovery are based on the estimates developed in the other loops. In a similar manner, optimality can also be rationalized for amplitude and symbol timing estimation. It becomes evident, therefore, that each processing loop coherently aiding the others is the best possible strategy.

There are two primary disadvantages with the implementation of the receiver shown in Figure A-8. The postdetection data filter, $H_i(t - u)$, adds not only a moderate degree of complexity, but also an additional delay, which constrains the bandwidth and stability of the estimation loops. These difficulties suggest that a suboptimum digital feedback approach may be a more practical alternative. The resulting receiver structure is given in Figure A-9. Its layout is similar to that of analog feedback except that the secondary data filter, $H_s(t - u)$, has been removed, and differentiation must be accommodated digitally. Note that the overall loop delay has been reduced from $(2K + \alpha + \beta + \frac{1}{2})T_s$ to $(K + \alpha + 1)T_s$. With typical filtering, this corresponds to trimming one to four symbol times of excess delay from each estimator.

Another implementation advantage of digital feedback is that the data multipliers M_A, M_θ, and M_τ can be fashioned as chopped gain amplifiers. That is, their outputs are either zero or ± 1 times the analog inputs. This capability is particularly significant because the DC stability and bandwidth of four quadrant analog multipliers are limiting factors in loop performance.
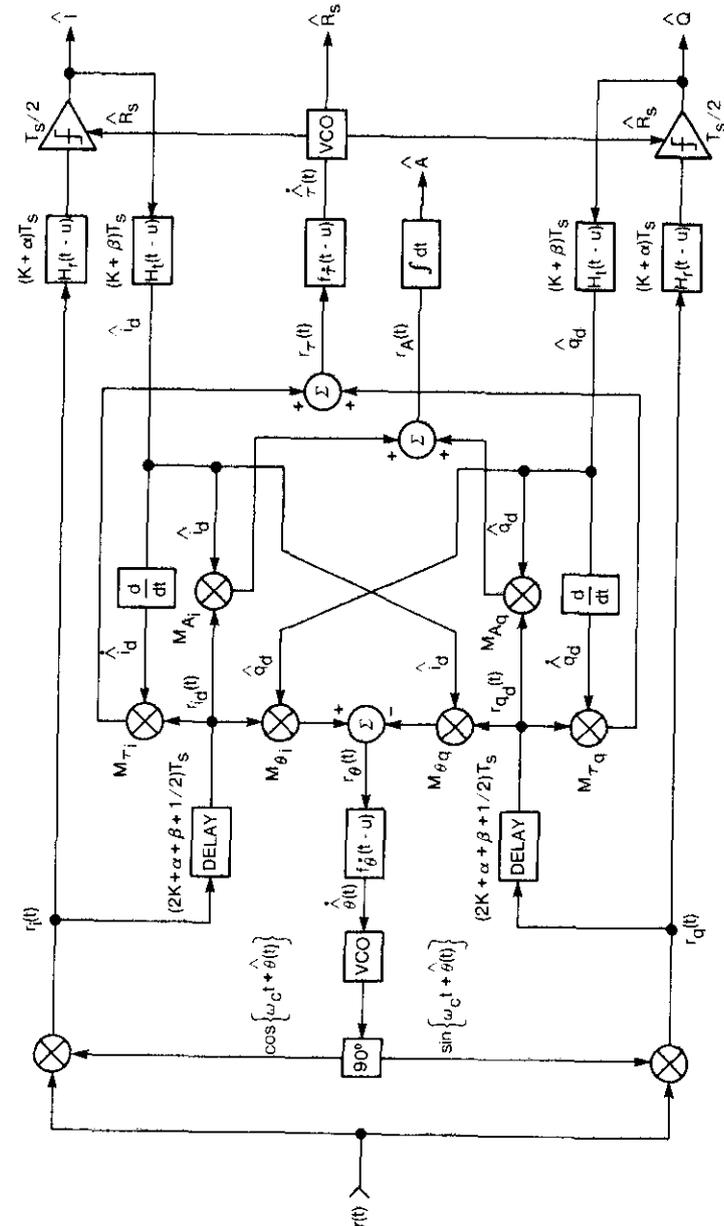
Figure A-8. Analog Feedback Receiver
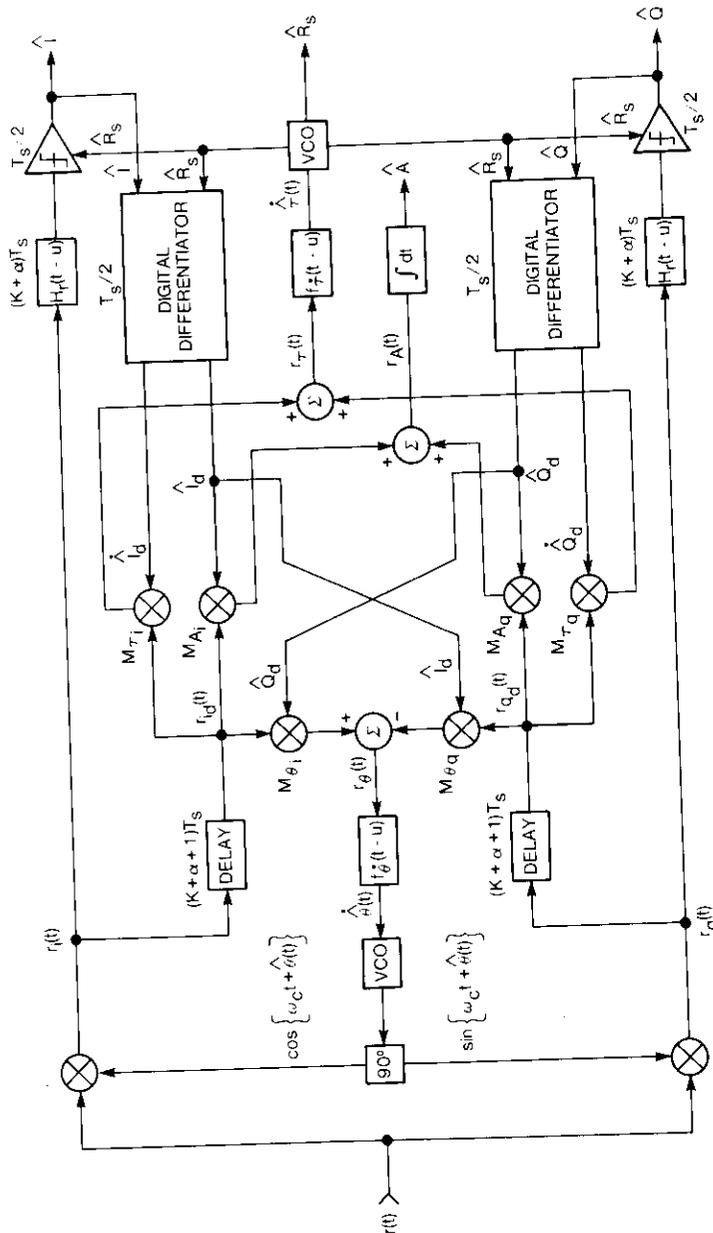
Figure A-9. *Digital Feedback Receiver*

In summary, digital feedback not only brings about a substantial reduction in complexity, but the accompanying decrease in delay allows for wider tracking bandwidths and better stability. A disadvantage with this implementation, as before, results from the limitations incurred by using analog delay elements. However, the effect can be minimized by tapping the estimation loop inputs after the receive data filter, $H_r(t - u)$. In this manner, the net analog delay required can be reduced to exactly one symbol time, T_s.

References

[A-1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: John Wiley and Sons, 1968.

[A-2] S. Stein and J. Jones, *Modern Communication Principles*, New York: McGraw-Hill, 1967, Section 12-1.

[A-3] J. J. Stiffler, *Theory of Synchronous Communications*, Englewood Cliffs, N.J.: Prentice-Hall, 1971, Sections 4-2 and 4-5.

[A-4] R. Jaffe and E. Rechtin, "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, March 1955, pp. 66–67.

[A-5] M. Yovits and J. Jackson, "Linear Filter Optimization With Game Theory Considerations," *IRE National Convention Record*, Part 4, 1955, pp. 193–199.

[A-6] D. Snyder, "Some Useful Expressions for Optimum Linear Filtering in White Noise," *Proceedings of the IEEE* (Correspondence), Vol. 53, June 1965, pp. 629–630.

[A-7] D. Snyder, "Optimum Linear Filtering of an Integrated Signal in White Noise," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-2, No. 2, March 1966, pp. 231–232.

Appendix B. Derivation of phase jitter expressions

Carrier tracking loop analysis

Phase jitter on the recovered carrier is attributable to two sources: local oscillator phase noise and thermal noise at the receiver input. In general, they are independent, so the total variance of the output carrier phase jitter is

$$\sigma^2_{\phi_o} = \sigma^2_{\theta_o} + \sigma^2_{\theta_{n_o}} \tag{B-1}$$

where $\sigma^2_{\theta_o}$ and $\sigma^2_{\theta_{n_o}}$ are the output variances caused by oscillator and thermal noise, respectively. Local oscillator phase noise is tracked by the carrier recovery loop, whereas thermal noise is not. As a result, the output phase variances are related as [16], [B-1], [B-2]

$$\sigma_{\theta_0}^2 = \int_{-\infty}^{\infty} |1 - H(f)|^2 \, S_{\theta_i}(f) \, df \tag{B-2a}$$

$$\sigma_{\theta_{n_0}}^2 = \int_{-\infty}^{\infty} |H(f)|^2 \, S_{\theta_{n_i}}(f) \, df \tag{B-2b}$$

Where $H(f)$ is the carrier recovery closed-loop transfer function, and $S_{\theta_i}(f)$ and $S_{\theta_{n_i}}(f)$ are the respective power spectral densities for the equivalent loop input oscillator phase noise and thermal noise components.

The local oscillator phase variance term in equation (B-1), $\sigma_{\theta_0}^2$, comprises jitter from both the 18-MHz transmit crystal oscillator of Figure 1, and the crystal-controlled vco in Figure 2. The resulting variances were measured with an HP8901A modulation analyzer. The results are summarized in Table B-1. The values are representative of what can be obtained with good quality crystal oscillators.

TABLE B-1. LOCAL OSCILLATOR
PHASE VARIANCE

B_L(Hz)	σ_{θ_0} (rad)
32	0.013
50	0.009
100	0.004
150	0.003
320	0.001
500	0.001
1,000	Negligible
1,500	Negligible

Since $|1 - H(f)|$ is highpass, $\sigma_{\theta_0}^2$ is proportional to the area under the tails of the phase noise power spectrum; consequently, it diminishes for progressively larger loop bandwidths.

The relationship for the phase variance caused by the thermal noise component in equation (B-1), $\sigma_{\theta_{n_0}}^2$, is derived as follows. Let the input signal to the demodulator be defined as in equations (B-1), (B-2), and (B-3).

$$r(t) = A\{i(t,\tau) \cos [\omega_c t + \theta(t)] + q(t,\tau)$$
$$\cdot \sin [\omega_c t + \theta(t)]\} + n(t) \quad . \tag{B-3}$$

The carrier recovery phase detector output is taken after the loop summer in Figure 2. In the steady state, with analog data feedback, the input signal of equation (B-3) results in the phase detector output from equation (A-9)

$$r_\theta(t) = \frac{A\hat{A}}{2}\{[i\hat{q} - q\hat{i}] \cos (\theta - \hat{\theta}) + [i\hat{i} + q\hat{q}]$$
$$\cdot \sin (\theta - \hat{\theta})\} + n_\theta(t) \tag{B-4}$$

where

$$n_\theta(t) = \hat{A}/\sqrt{2} \, [\hat{q} \cdot n_i(t) - \hat{i} \cdot n_q(t)] \quad .$$

The carrier phase detector noise term, $n_\theta(t)$, comprises the recovered data estimates, $\hat{i}$ and $\hat{q}$, which are cross-multiplied against the quadrature baseband equivalent noise processes, n_i and n_q. The signals n_i, n_q, $\hat{i}$, and $\hat{q}$ are statistically independent; and n_i and n_q are zero mean low-pass noise processes of strength $N_o/2$. Thus, the variance of the additive noise term in equation (B-4) is

$$\overline{n_\theta^2} = \frac{\hat{A}^2}{2}\left[\overline{\hat{q}^2 n_i^2} + \overline{\hat{i}^2 n_q^2} \right] \quad . \tag{B-5}$$

With $\hat{i}$ and $\hat{q}$ defined as unit power, this reduces to

$$\overline{n_\theta^2} = \frac{\hat{A}^2}{2}\left[\overline{n_i^2} + \overline{n_q^2} \right]$$
$$= \frac{\hat{A}^2}{2}\left[\frac{N_o}{2} + \frac{N_o}{2} \right] 2\left(\frac{B_{IF}}{2} \right) = \frac{\hat{A}^2 N_o B_{IF}}{2} \tag{B-6}$$

where B_{IF} is the noise equivalent bandwidth of the IF filter.

To determine the equivalent phase jitter at the input to the carrier recovery loop that results in this variance, equation (B-4) is rewritten as

$$r_\theta'(t) = \frac{A\hat{A}}{2}\{[\hat{i}i + q\hat{q}] \sin [(\theta - \hat{\theta}) + \theta_{n_i}]\}$$
$$\simeq \frac{A\hat{A}}{2}\{[\hat{i}i + q\hat{q}][(\theta - \hat{\theta}) + \theta_{n_i}]\} \tag{B-7}$$

where the additive amplitude noise term has been replaced with an equivalent additive phase disturbance. Its variance is

$$\overline{n_\theta'^2} = (A\hat{A})^2 \, \sigma_{\theta_{n_i}}^2 \quad . \tag{B-8}$$

Equating this result with that of equation (B-6) gives

$$\sigma_{\theta_{n_i}}^2 = \frac{N_o B_{IF}}{2A^2} \triangleq \frac{1}{2(C/N)_{IF}} \tag{B-9}$$

where

$$N = N_o B_{IF}$$
$$C = A^2$$
$$(C/N)_{IF} = C/N \text{ in IF bandwidth.}$$

Thus, the equivalent input phase variance is the reciprocal of twice the carrier-to-noise ratio measured in the IF bandwidth. The relationship in equation (B-9) is exactly the same as that found by Gardner [16] for PLLs without data modulation, where $i(t,\tau) = q(t,\tau) = 1$. Hence, the joint estimator-detector with analog data feedback yields the same tracking jitter performance as a PLL operating on an unmodulated carrier.

The output variance of the carrier recovery phase jitter thermal noise component is obtained by substituting equation (B-9) into (B-2b)

$$\sigma^2_{\theta_{n_o}} = \frac{(N_o B_{IF}/2A^2)}{2(B_{IF}/2)} \int_{-\infty}^{\infty} |H(f)|^2 df \quad . \tag{B-10}$$

This may be simplified to

$$\sigma^2_{\theta_{n_o}} = \frac{N_o B_L}{A^2} \stackrel{\triangle}{=} \frac{1}{(C/N)_L} \tag{B-11}$$

where $B_L \stackrel{\triangle}{=} \int_0^{\infty} |H(f)|^2 df$ is the noise equivalent bandwidth, and $(C/N)_L$ is the carrier-to-noise ratio in the loop.

The total output phase jitter defined in equation (B-1) is now completely specified, and a theoretical basis exists to evaluate the measured performance. Observe that the component resulting from oscillator phase noise listed in Table B-1 diminishes with increasing loop bandwidth, whereas that resulting from thermal noise given by equation (B-11) grows larger. This implies that there is an optimum loop bandwidth at a given C/N to achieve minimum jitter on the recovered carrier reference. This bandwidth is primarily determined by the spectral distribution of the local oscillator phase noises, which may vary widely depending on the application.

Clock tracking loop analysis

To enable uniform acquisition with the joint recovery scheme, the carrier phase and symbol timing estimation loop bandwidths were set equally. From a practical standpoint, this rendered the phase noise contribution from the clock oscillators negligible. Hence, the total output phase jitter on the reconstructed clock was dominated by thermal noise.

The symbol timing phase detector output is taken after the loop summer in Figure 2. With the demodulator input signal of equation (3), the phase detector output is

$$r_\tau(t) = \frac{A\hat{A}}{2} \{[i\hat{i} + q\hat{q}] \cos(\theta - \hat{\theta}) + [\hat{q}i - i\hat{q}]$$
$$\cdot \sin(\theta - \hat{\theta})\} + n_\tau(t) \tag{B-12}$$

where

$$n_\tau(t) = (\hat{A}/\sqrt{2})\,[\hat{i} \cdot n_i(t) + \hat{q} \cdot n_q(t)] \quad .$$

The symbol phase detector noise term, $n_\tau(t)$, comprises the time derivatives of the recovered data estimates, $\hat{i}$ and $\hat{q}$, which are multiplied against the quadrature baseband noise processes, n_i and n_q. The signals n_i, n_q, $\hat{i}$, and $\hat{q}$ are statistically independent; n_i and n_q are zero mean low-pass noise processes of strength $N_o/2$. Thus, the variance of the additive noise term in equation (B-12) is

$$\overline{n_\tau^2} = \frac{\hat{A}^2}{2}\left[\overline{\hat{i}^2 n_i^2} + \overline{\hat{q}^2 n_q^2}\right] \quad . \tag{B-13}$$

For simplicity, the remainder of the analysis will be carried out for the case of 1010 . . . data modulation. With Nyquist filtering, the data waveforms are then represented as

$$\hat{i}(t,\hat{\tau}) = \hat{q}(t,\hat{\tau}) = \sqrt{2}\sin[\pi R_s(t + \hat{\tau})] \tag{B-14a}$$

$$\dot{\hat{i}}(t,\hat{\tau}) = \dot{\hat{q}}(t,\hat{\tau}) = \sqrt{2}\cos[\pi R_s(t + \hat{\tau})] \tag{B-14b}$$

Substituting equation (B-14b) into (B-13) yields

$$\overline{n_{\tau_{10}}^2} = \frac{\hat{A}^2}{2}\left[\overline{n_i^2} + \overline{n_q^2}\right]$$

$$= \frac{\hat{A}^2}{2}\left[\frac{N_o}{2} + \frac{N_o}{2}\right]2\left(\frac{B_{IF}}{2}\right) = \frac{\hat{A}^2 N_o B_{IF}}{2} \quad . \tag{B-15}$$

To determine the equivalent timing jitter at the input to the clock recovery loop that would result in this variance, equation (B-12) is rewritten as

$$r'_\tau(t) = \frac{A\hat{A}}{2}[i(t,\tau + \tau_{n_i})\,\dot{\hat{i}}(t,\hat{\tau})$$
$$+ q(t,\tau + \tau_{n_i})\,\dot{\hat{q}}(t,\hat{\tau})] \tag{B-16}$$

where the additive amplitude noise term has been replaced with an additive phase disturbance. Substituting equation (B-14b) into (B-16) gives the noise component

$$n'_{\tau_{10}}(t) = A\hat{A}[\sin(\pi R_s \tau_{n_i})]$$

$$\simeq A\hat{A}(\pi R_s \tau_{n_i}) \quad . \tag{B-17}$$

Its variance is

$$\overline{n'^2_{\tau_{10}}} = (A\hat{A})^2 \, \overline{(\pi R_s \tau_{n_i})^2_{10}} \quad . \tag{B-18}$$

Equating this result with equation (B-15) yields

$$\overline{(\pi R_s \tau_{n_i})^2} = \frac{N_o B_{IF}}{2A^2}$$

$$\triangleq \frac{1}{2(C/N)_{IF}} \quad . \tag{B-19}$$

This is exactly the same relationship that was found for the carrier phase jitter in equation (B-9). Also note that the timing jitter in equation (B-19) corresponds to a frequency of $R_s/2$, because 1010 . . . data represents the maximum rate that can exist at the phase detector inputs. The timing jitter on the recovered clock operating at R_s is

$$\sigma^2_{\tau_{n_i}} = \overline{(2\pi R_s \tau_{n_i})^2} = \frac{2}{(C/N)_{IF}} \quad . \tag{B-20}$$

In terms of the loop S/N defined in equation (B-11), this is written as

$$\sigma^2_{\tau_{n_o}} = \frac{4}{(C/N)_L} \quad . \tag{B-21}$$

Equation (B-21), then, is the analytical basis upon which clock recovery loop phase jitter will be assessed.

References

[B-1] W. Lindsey, *Synchronization Systems in Communications and Control*, Englewood Cliffs, N.J.: Prentice-Hall, 1972, Section 4-3.

[B-2] C. Wolejsza, "Effects of Oscillator Phase Noise on PSK Demodulation," *COMSAT Technical Review*, Vol. 6, No. 1, Spring 1976, pp. 107–125.

John J. Poklemba received a B.S.E.E. from Drexel University in 1972, and an M.S. and E.E., concurrently, from M.I.T. in 1974. Prior to joining COMSAT Laboratories, he spent 2 years working with HF receivers and high-resolution video signal processing. Mr. Poklemba is currently a Staff Scientist in the Transmission Processing Department. His principal interests are the application of detection and estimation theory to practical receiver structures, and the development of various types of filtering relevant to information transmission.

Block orthogonal convolutional coding for the INTELSAT TDMA data channel

K. M. MACKENTHUN

(Manuscript received May 1, 1984)

Abstract

Because of its stringent bit error rate (BER) requirement, data traffic in the INTELSAT system may require more coding gain than that provided by the (127,112) Bose-Chaudhuri-Hocquenghem (BCH) code presently used. Economic constraints suggest that the additional coding gain be obtained by retaining the (127,112) BCH code and by introducing an additional code in the time division multiple access (TDMA) terminal. Herein, additional codes which can be used outside the (127,112) BCH code are discussed.

A block orthogonal convolutional code, which is a self-orthogonal convolutional code designed for the random block error channel produced by the BCH decoder, is shown to be effective in reducing BER. The code is easily implementable and can be added to the INTELSAT TDMA terminal without hardware or procedural modification.

Introduction

Data traffic requires a lower BER than that specified by the current INTELSAT BER service objective, which applies mainly to voice transmission. The BER may be economically lowered by using channel coding.

The current INTELSAT TDMA terminal, which uses a (127,112) BCH code [1], does not have adequate coding gain for data traffic. To minimize service interruptions and operational changes to the current INTELSAT system while

keeping costs as low as possible, the coding gain needed for data traffic can be obtained by retaining the (127,112) BCH code and using an additional code outside the BCH code. This paper presents several codes which are suitable for use as outer codes in the INTELSAT terminal.

The (127,112) BCH code corrects all single and double errors. Therefore, the BCH decoded channel is a sequence of 112-bit blocks in which block errors (typically three bit errors per block) occur randomly. Since the BCH decoded channel is not a random error channel, a random error-correcting code is not necessarily well suited as an outer code. However, interleaving a simple random error-correcting code can be effective for this channel.

A t error-correcting $(n - 1, n)$ self-orthogonal convolutional code,* bit interleaved to depth 112, can improve BER significantly for small t. Because of the small variation in BER performance with n, a large n can be used, reducing rate loss caused by the outer code. The threshold decoder for the code can be easily implemented with shift registers. For large n and $t > 1$, the shift register memory required by the interleaved self-orthogonal convolutional code becomes large. To reduce the memory requirement, a block orthogonal $(n - 1, n)$ convolutional code is constructed for the random block error channel produced by the BCH decoder. The code uses roughly $1/n$ of the memory of the $(n - 1, n)$ interleaved self-orthogonal convolutional code. The t block error-correcting $(n - 1, n)$ convolutional code is threshold decodable and has the same performance as the interleaved t error-correcting $(n - 1, n)$ self-orthogonal convolutional code.

The next section presents a discussion of the current INTELSAT system, including the (127,112) BCH code. Subsequent sections discuss the error statistics of the channel at the output of the BCH decoder, codes which perform well for this channel, and placement of the outer code in the terminal; the performance and implementation of the interleaved self-orthogonal and of the block orthogonal convolutional codes; and other code possibilities. The concluding section gives the recommended approach.

In this paper, raw channel errors are assumed to be independent and identically distributed; therefore, the codes presented here are not designed for channel degradations such as co-channel interference, adjacent channel interference, terrestrial interference, and other transmission anomalies.

The INTELSAT TDMA system

The INTELSAT terminal

In the present INTELSAT TDMA system, each burst contains up to 32 subbursts, with each subburst consisting of up to 128 satellite channels (SCs).

---

* Where t is the number of errors corrected, $n - 1$ is the number of information bits in a code subblock, and n is the subblock length [2].

Each SC corresponds to 128 consecutive bits in one subburst in successive frames. Since the frame size is 2 ms, each SC has a 64-kbit/s data rate.

The INTELSAT terminal interfaces terrestrial data and speech channels with satellite channels. Terrestrial data channels of rate 64 kbit/s are assigned to one satellite channel; those that are multiples of 64 kbit/s are transmitted by using several SCs. Terrestrial data channels smaller than 64 kbit/s are multiplexed together to form one SC. An SC used for data is called a data not interpolated (DNI) channel. An SC used for speech, which can be shared among several simultaneous voice conversations, is called a digital speech interpolation (DSI) channel.

The INTELSAT terminal consists of common TDMA terminal equipment (CTTE), a multiplexer/demultiplexer, and one or more DSI/DNI units, as shown in Figure 1. The DSI/DNI unit interfaces terrestrial channels with SCs. During transmission, each DSI/DNI unit forms one subburst in each frame. During each frame, the multiplexer multiplexes one subburst from each DSI/DNI unit into one data burst; the resulting burst is then sent to the CTTE where it is encoded and transmitted.

During reception, each burst is first decoded and then sent to the demultiplexer. The demultiplexer routes particular subbursts to the appropriate
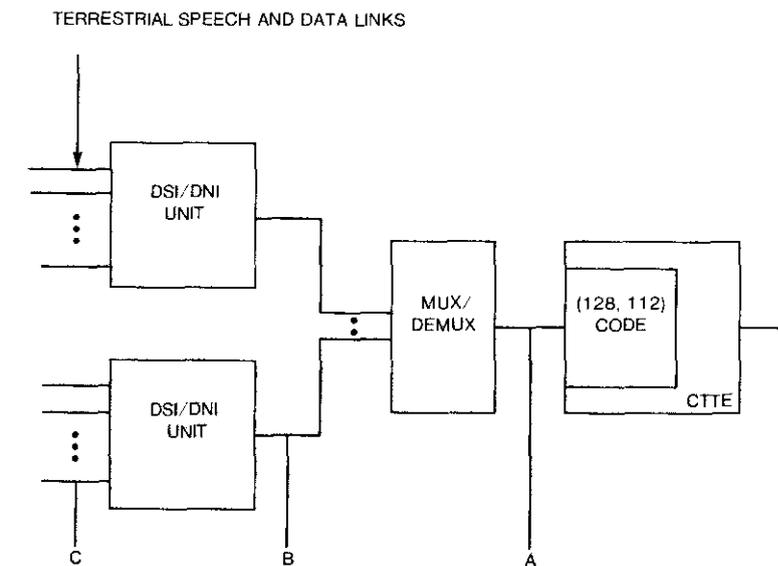
Figure 1. *Block Diagram of INTELSAT TDMA Terminal (A, B, and C are possible locations for additional coding)*

DSI/DNI unit. Each DSI/DNI unit can receive up to 8 subbursts per frame; these subbursts may occur in different bursts.

The (127,112) Code

The INTELSAT terminal uses a (127,112) BCH code that corrects all single and double errors and detects all triple errors. [By strict definition, the (127,112) code is a (127,113) BCH code which is shortened by deleting one information bit and extended by adding an overall parity bit. For convenience, however, the term (127,112) BCH code is used here.] The (127,112) BCH code has a minimum distance of 6. Since the TDMA terminal has an 8-bit clocking cycle, an additional dummy bit is added to each block, making the code (128,112), so that both the number of bits in each block and the number of information bits in each block are multiples of 8.

For each received word, the decoder calculates a 14-bit syndrome and an overall parity check. The 14-bit syndrome is used to address one of 2^{14} memory locations in ROM, which store two 7-bit numbers, p and q, indicating up to two error locations in the received word. There are $\binom{127}{k}$ syndromes and $\binom{127}{k}$ corresponding memory locations that indicate k errors in the received word, for $k = 0, 1, 2$. The decoder changes the p and q position of the received word if:

a. the syndrome corresponds to a single error and the parity check is odd, or

b. the syndrome corresponds to a single or double error and the parity check is even.

When the received word has more than three errors, the syndrome may determine a p, q value that does not correspond to any of the error positions in the received word; consequently, the decoded word will have more errors than the received word. Simulation results show that when more than three errors occur, the number of additional decoder errors is directly proportional to the fraction of ROM locations that specify whether 0, 1, or 2 error corrections are to be made [1]. Thus, $p(m,j)$, the probability of m errors in the decoded word and j errors in the received word, may be easily calculated:

$$p(j) \qquad\qquad m = 0, j < 3$$

$$p(j) \qquad\qquad m = j = 3$$

$$p(m,j) = \begin{cases} \dfrac{2^{14} - \binom{127}{1}}{2^{14}}\, p(j) & \begin{array}{l} m = j > 3, \\ j \text{ odd} \end{array} \\[2em] \dfrac{\binom{127}{1}}{2^{14}}\, p(j) & \begin{array}{l} m - 1 = j > 3, \\ j \text{ odd} \end{array} \\[2em] \dfrac{2^{14} - \binom{127}{1} - \binom{127}{2}}{2^{14}}\, p(j) & \begin{array}{l} m = j > 3, \\ j \text{ even} \end{array} \\[2em] \dfrac{\binom{127}{1}}{2^{14}}\, p(j) & \begin{array}{l} m - 1 = j > 3, \\ j \text{ even} \end{array} \\[2em] \dfrac{\binom{127}{2}}{2^{14}}\, p(j) & \begin{array}{l} m - 2 = j > 3, \\ j \text{ even} \end{array} \\[2em] 0 & \text{otherwise} \end{cases} \qquad (1)$$

where $p(j)$ is the probability of j errors in the received codeword, given by

$$p(j) = \binom{127}{j} p_b^j (1 - p_b)^{127-j} \qquad (2)$$

where p_b is the channel BER [the expression for $p(m,j)$ assumes that decoder errors do not cancel out errors in the received codeword, an accurate approximation]. Comparing the magnitude of the terms, $p(m,j)$ may be well approximated by:

$$p(m,j) = \begin{cases} p(j) & m = 0, j < 3 \\ p(j) & m = j = 3 \\ p(j) & m = j > 3, j \text{ odd} \\ 0.5\, p(j) & m = j > 3, j \text{ even} \\ 0.5\, p(j) & m - 2 = j > 3, j \text{ even} \\ 0 & \text{otherwise} . \end{cases} \qquad (3)$$

The output BER of the (127,112) code, p, may be calculated from:

$$p = \sum_{m=3}^{127} \sum_{j=0}^{127} \frac{m}{127} p(m,j) \quad . \tag{4}$$

By using equation (4), the output BER of the (127,112) code is evaluated by computer and is shown in Figure 2.
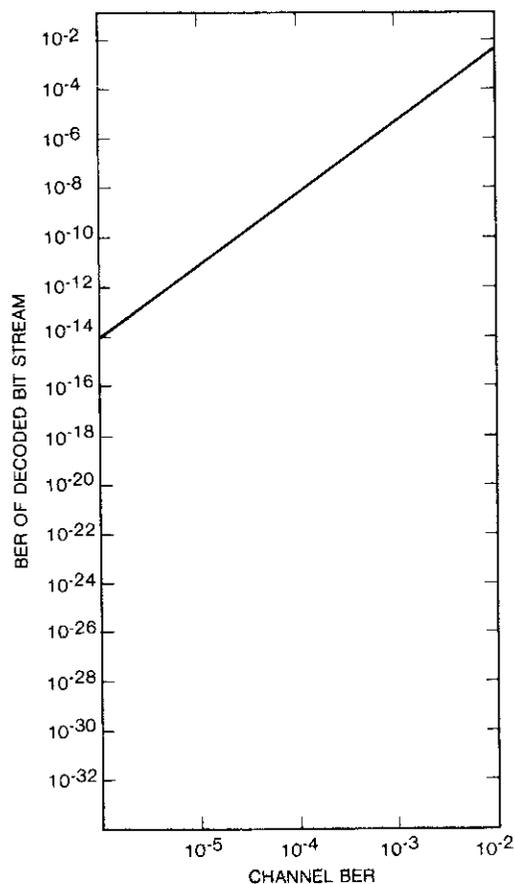
Figure 2. *BER of (127,112) BCH code*

Code location and interface

The BCH code is located in the CTTE (see Figure 1). Since the overall code is (128,112), 112-bit blocks are transferred between the BCH codec and

multiplexer/demultiplexer and 128-bit blocks are transferred between the BCH codec and remainder of the CTTE. The multiplexer/demultiplexer and CTTE operate at the same clock speed with intermittent clocking of the multiplexer/demultiplexer to account for the smaller block size across the multiplexer/demultiplexer-codec interface.

To encode a burst from the multiplexer, the burst is divided into 112-bit blocks, starting at the leading edge of the burst, with the last block possibly less than 112 bits. A block gate signal between the codec and multiplexer is used to transfer successive blocks to the encoder, where 112-bit blocks are encoded into 128 bits. The last block, possibly less than 112 bits, is encoded by assuming that fill zeros precede the block. The succession of 128-bit blocks and the tail block are then transmitted by the CTTE, forming one station burst (see Figure 3). The decoding of a station burst is implicitly defined by the encoding procedure just described.
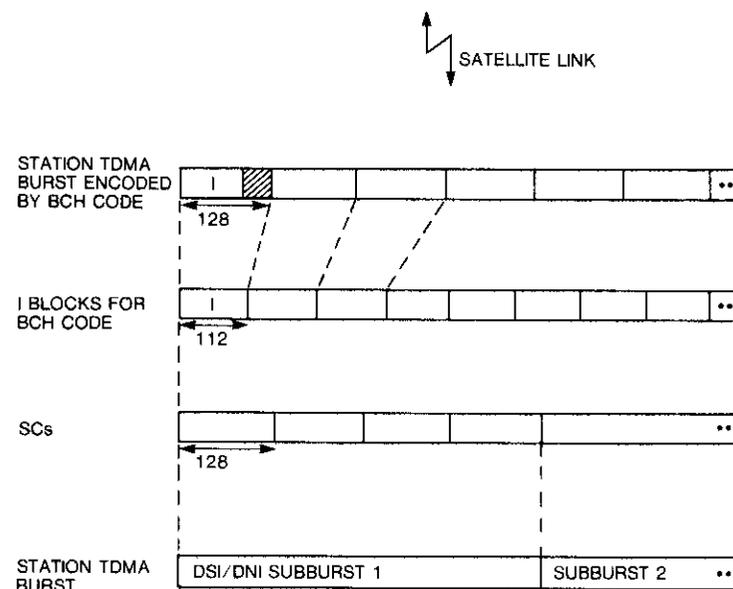
Figure 3. *Transmission Format of TDMA Burst*

Outer codes for the INTELSAT TDMA/DNI channel

The decoded channel

The decoded bit stream consists of a sequence of 112-bit blocks in which errors are independent from block to block; but within a block, errors are

not independent because of the inner code. The distribution of errors within a 112-bit block can be calculated from:

$$p_{112}(k) = \sum_{m=k}^{k+15} \frac{\binom{112}{k}\binom{15}{m-k}}{\binom{127}{m}} \sum_{j=0}^{127} p(m,j) \qquad (5)$$

where $p_{112}(k)$ is the probability of k errors in a 112-bit block. This distribution is plotted in Figure 4, for a raw channel BER of 10^{-3}, and as expected, it shows a clumping about three errors.
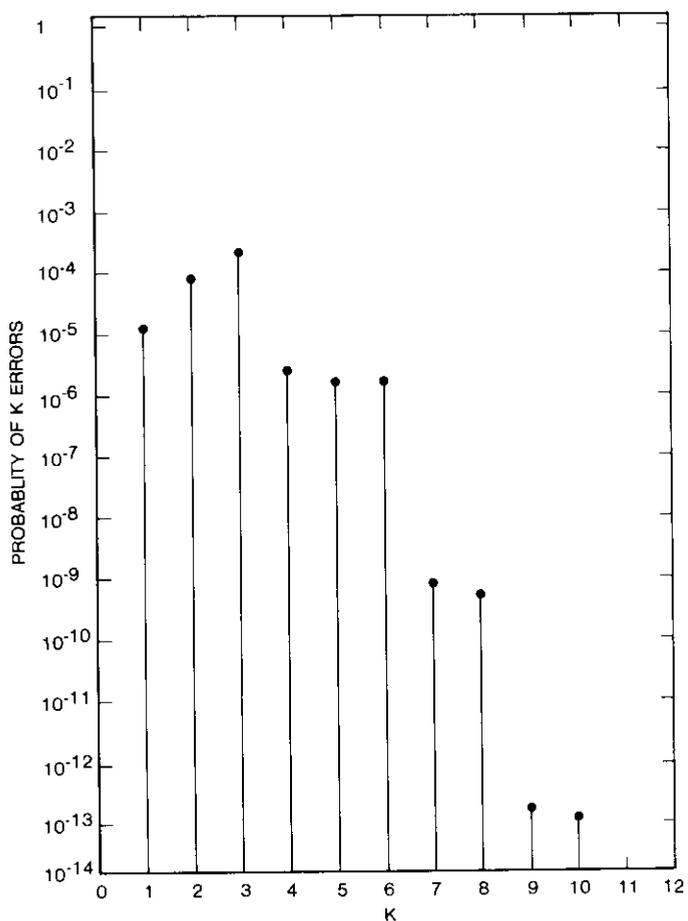
Figure 4. *Distribution of Errors for 112-Bit Decoded Block of (127,112) Inner Code at 10^{-3} Channel BER*

Each 128-bit block SC is composed of portions of two 112-bit blocks. Depending on the location of the SC in the burst, the included portions can be 112:16, 96:32, 80:48, or 64:64. Each different alignment of the 112-bit blocks gives a different distribution of errors in the 128-bit block of the SC. The error distribution can be easily calculated and is shown in Figure 5 for the two extreme alignments, 112:16 and 64:64. As expected, the 112:16 probability distribution shows a relative peak at three errors, and the probability of 4 and 6 errors is approximately the same; the 64:64 distribution is more random.

Figure 5. *Distribution Errors for 112:16 and 64:64 SCs at 10^{-3} Channel BER*

Candidate codes

There are two general types of error-correcting codes: random error-correcting codes and burst error-correcting codes. Since the channel of 112-bit decoded blocks does not correspond to either a random error channel or a burst error channel in which a large number of errors are clumped together, neither a random nor burst error-correcting code is perfectly suitable for use on the decoded channel. For a burst error-correcting code to work properly, the channel must be divided into a sequence of interleaved burst and guard spaces of lengths b and g, respectively; no errors can occur within the guard spaces. Unfortunately, there is no value of b and g that guarantees correction of all error patterns in the decoded channel. The diffuse code, a burst error-correcting code with some random error-correction capability, is not suitable for use on the decoded channel either, since it too requires a guard space around each error burst.

When the channel is not random, interleaving can be used to destroy the statistical dependence between channel bits and produce a memoryless channel, for which a simple random error-correcting code can perform effectively [2],[3]. The BCH decoded channel must be interleaved to depth i, $i \geq 112$, to ensure that the i interleaved bit streams consist of statistically independent bits. In each of the i interleaved bit streams, bit errors are separated on the average by $1/p$ bits. Bit errors in the non-interleaved bit stream, however, are typically clustered into three errors in 112 bits, with each error cluster separated by approximately $3/p$ bits. Therefore, roughly speaking, a simple random error-correcting code used on each of the i interleaved bit streams, which corrects roughly 1 error in $1/p$ bits, can be as effective as a more powerful code used on the non-interleaved bit stream, which must correct at least three errors in 112 bits.

Because of their effectiveness and simplicity of implementation when used with interleaving, t error-correcting $(n - 1, n)$ interleaved self-orthogonal convolutional codes are considered for use on the BCH decoded channel, for small t. The performance of these codes on the BCH decoded channel is examined in a later section. Finally, a block orthogonal convolutional code, which has the same performance as an interleaved self-orthogonal convolutional code but is easier to implement, is discussed.

Outer code placement and interfacing

There are three possible locations for the outer code: position A, between the CTTE and mux/demux; position B, between the mux/demux and individual DSI/DNI units; and position C, at each terrestrial link (see Figure 1). The code location affects the clock speed of the codec hardware, the number of

codecs required, the communications efficiency of the system, system operability, and the performance and suitability of a particular code. The clock speed of the codec hardware affects codec cost and the feasibility of implementing different codes. In position A, the codec hardware operates at 120 Mstep/s; in position B, at the DNI interface rate; and in position C, at the terrestrial data rate. The smallest number of codecs is required in position A. Although position C requires the largest number of codecs, it also allows the codec to be implemented in large-scale integration (LSI) form, reducing the cost of additional coding, even though a large number of codecs are required. Without additional logic to turn the codec on and off, the code position determines the amount of each frame coded, therefore determining the communication efficiency. In position A, an entire burst is encoded; in position B, individual subbursts can be encoded; and in position C, individual data channels can be encoded. The code position affects system operability, since two earth terminals which communicate must have their codec in the same position.

If a convolutional code is used in position A or B and the codec is not flushed at the end of each burst (subburst), an encoder and decoder are required for each transmitted and received burst (subburst), since the state of the encoder and decoder must remain unchanged until the start of the burst (subburst) in the next frame. In position C, it is natural to not flush the convolutional code, since individual data channels are being encoded.

Positions A and B require modification of the current INTELSAT terminal; require coding an entire burst or subburst if complicated control logic is not used; reduce flexibility in system operability; and for convolutional codes, introduce additional complications in code flushing. In this study, therefore, coding in position C is emphasized.

Performance of interleaved self-orthogonal convolutional codes on the TDMA/DNI channel

A t error-correcting $(n - 1, n)$ self-orthogonal convolutional code can be used as the outer code. The parity check matrix for a $t = 1$, and $n = 2$, and 5 self-orthogonal convolutional code is shown below:

$$H = \begin{bmatrix} \cdot \; \cdot \; \cdot \\ 1011 \\ 1011 \\ \cdot \; \cdot \; \cdot \end{bmatrix} \qquad , t = 1, n = 2$$

$$H = \begin{bmatrix} \cdot \; \cdot \; \cdot \\ 000100010001000100011111 \\ 000100010001000100011111 \\ \cdot \; \cdot \; \cdot \end{bmatrix}, t = 1, n = 5 \qquad (6)$$

The $(n - 1, n)$ self-orthogonal convolutional code is interleaved to depth i, $i \geq 112$, using a standard technique [3]. Interleaving requires that n be a divisor of $i - 1$ [3]. Since i must be 112 or larger, $i = 113, 112, 113, 116, 115, 113$, and 113 are used for $n = 2, 3, 4, 5, 6, 7$, and 8, respectively.

Self-orthogonal convolutional codes may be decoded by using threshold decoding. For analytical convenience, this paper considers only definite decoding, in which only uncorrected channel bits are used to calculate the parity check sum. The encoder and decoder for a $t = 1$, $n = 2$, and 5 self-orthogonal convolutional code bit interleaved to depths 113 and 116, respectively, is shown in Figures 6 and 7.

(a) CONVOLUTIONAL ENCODER

(b) THRESHOLD DECODER
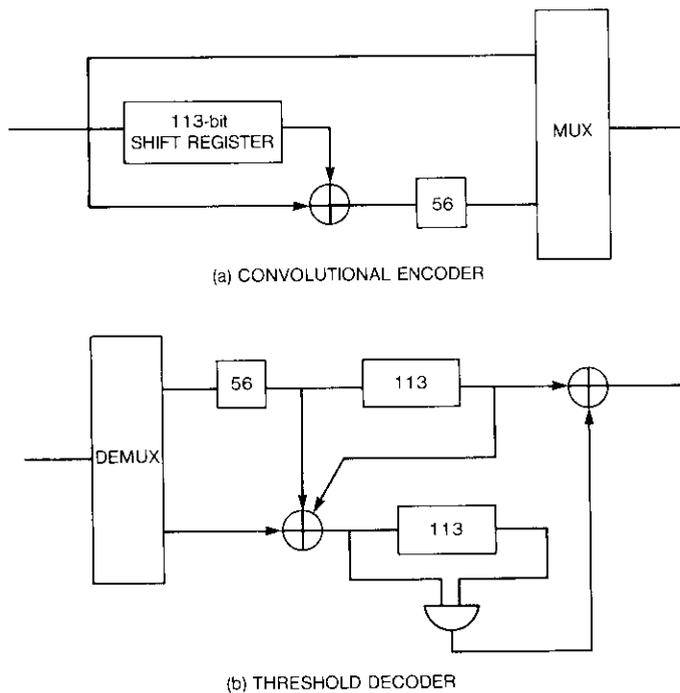
Figure 6. *Convolutional Encoder (a) and Threshold Decoder (b) for* t $= 1$, n $= 2$, *and* i $= 113$

BER Performance

The probability of bit error is calculated for a t error-correcting $(n - 1, n)$ self-orthogonal convolutional code, for definite decoding. Consider an

(a) CONVOLUTIONAL ENCODER

(b) THRESHOLD DECODER
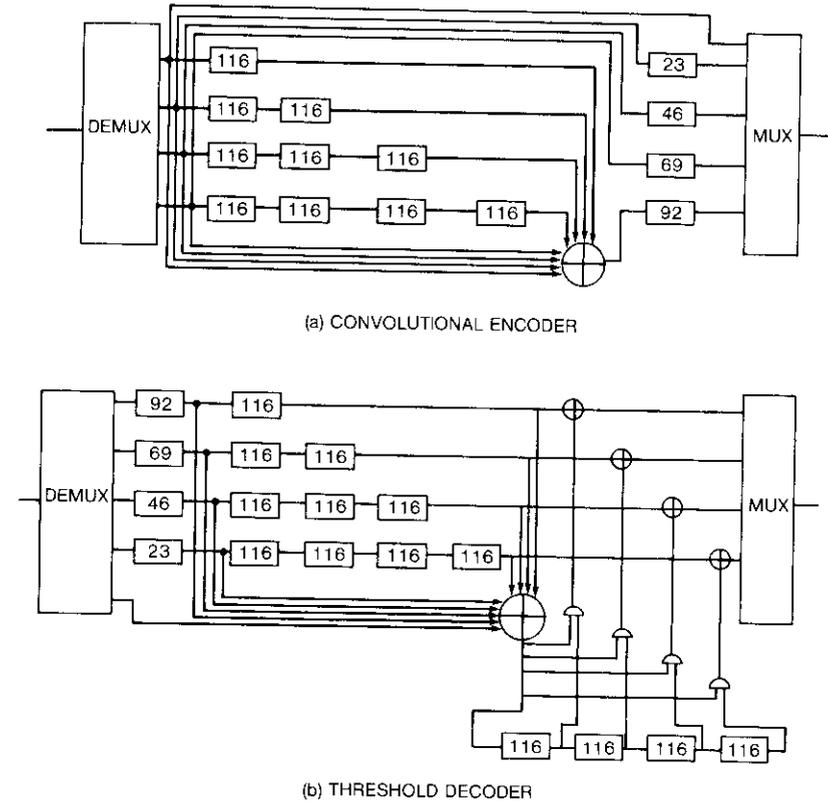
Figure 7. *Convolutional Encoder (a) and Threshold Decoder (b) for* t $= 1$, n $= 5$, *and* i $= 116$

infinite information sequence in which one bit is marked and the remaining bits are unmarked. The probability of bit error of the marked bit, p_e, is calculated. The marked bit appears in $2t$ parity check equations. Each parity check equation has $1 + (n - 1) 2t$ terms, where one term is the marked bit, and $(n - 1) 2t$ terms are unmarked bits. No unmarked bit appears more than once in the set of $2t$ equations. Let the event {miss} be the event, that in one parity check equation, an odd number of unmarked bits are in error. The probability of the event {miss}, p_m, is given by:

$$p_m = \sum_{j=1,3,5,\ldots} \binom{[n-1]2t}{j} p^j (1 - p)^{[n-1]2t-j} \tag{7}$$

where p is the probability of bit error on the BCH decoded channel. The marked bit is decoded incorrectly if:

a. the marked bit is in error and t or more parity check equations give the event {miss}, or

b. the marked bit is correct and more than t parity check equations give the event {miss}.

Then p_e is given by:

$$p_e = p \sum_{k=t}^{2t} \binom{2t}{k} p_m^k (1 - p_m)^{2t-k}$$
$$+ (1 - p) \sum_{k=t+1}^{2t} \binom{2t}{k} p_m^k (1 - p_m)^{2t-k} \quad . \tag{8}$$

Equation (8) is evaluated by computer and shown in Figure 8 for $t = 1$ and $n = 2, 3, 4, 5, 6, 7,$ and 8. Note that the BER curves show only slight degradation as n increases, amounting to small shifts of the BER curve for $n = 2$. This is because the probability of bit error is principally determined by the double error term:

$$p_e \cong 4n(n - 1)p^2 \tag{9}$$

which increases slowly with n, for small n.

It is interesting to compare the BER performance of an interleaved self-orthogonal convolutional code and an outer block code, which does not use block interleaving. Figure 9 shows the BER of a BCH outer block code of block length 127, for $t = 3, 4,$ and 5. The BER is calculated for the 112:16 alignment of inner and outer codes, using the probability distribution shown in Figure 5; the BER for other alignments is almost the same. Comparing Figures 8 and 9 shows that a $t = 5$ (127,92) BCH block code is needed to achieve the same BER performance for this channel as an interleaved $t = 1$ self-orthogonal convolutional code. Since the implementation of the (127,92) BCH code is significantly more complex and the code rate is smaller (0.72), the interleaved self-orthogonal convolutional code is a better choice for the INTELSAT terminal.

Block orthogonal convolutional code

It is desirable to make t and n as large as possible for the t error-correcting $(n - 1, n)$ interleaved self-orthogonal convolutional code, in order to improve BER performance as much as possible and reduce the rate loss of the additional
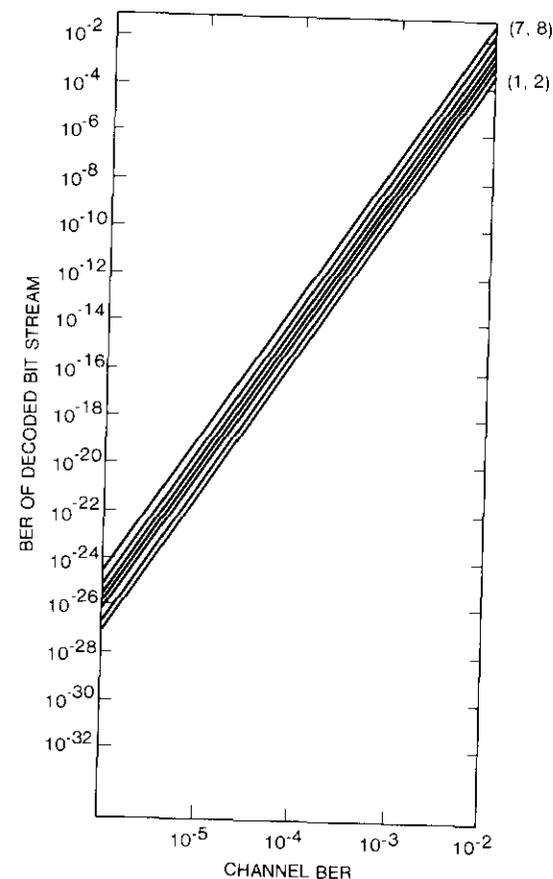
Figure 8. *BER of* t $= 1$, (n $- 1$, n) *Self-Orthogonal Convolutional Codes Concatenated With (127,112) BCH Inner Code*

coding. However, for $t > 1$ and large n, the codec memory required by the interleaved self-orthogonal convolutional code becomes large. In this section, a block orthogonal convolutional code that significantly reduces the codec memory is described. The block orthogonal convolutional code is threshold decodable and achieves the same BER performance as that of the corresponding interleaved self-orthogonal convolutional code.

Code construction and implementation

To understand the code construction, consider the case of a rate-$\frac{4}{5}$ decoder for a systematic convolutional code that decodes a sequence of 112-bit inner
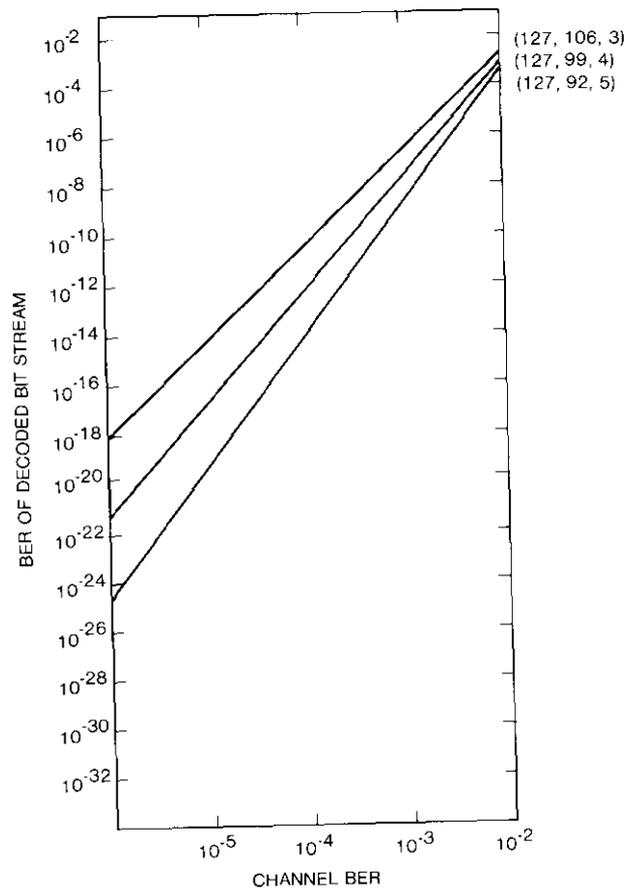
Figure 9. *BER of Outer BCH Code,* t = 3, 4, *and 5, Concatenated With (127,112) BCH Inner Code*

code blocks (see Figure 10). Because of the action of the demultiplexer, each 112-bit block is divided into five portions, with each of the 5-bit streams receiving a 22- or 23-bit portion of each block. Note that if taps are placed in the bit streams with each tap separated by at least 23 bits, within a single row, each tap will correspond to a bit in a distinct 112-bit block; within a column, each tap will correspond to a bit in the same 112-bit block, or to the first bit in the adjacent 112-bit block. To construct the rate-$\frac{4}{5}$ threshold decoder, take the four generator polynomials for a $\frac{4}{5}$ t error-correcting self-
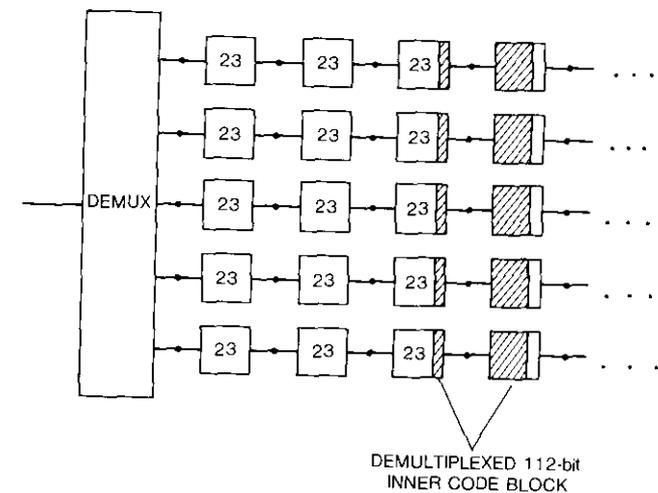
Figure 10. *Demultiplexing of 112-Bit Inner Code Blocks by Rate-$\frac{4}{5}$ Decoder*

orthogonal convolutional code, and arrange the taps of the generator polynomial in the subset of tap locations separated by at least 23 bits, so that no two generator polynomials use the same tap location in any column. Since the generator polynomials for the resulting decoder still have the difference set property, this code is t error-correcting. Moreover, since each tap checks a bit from a different 112-bit block, each block only appears once in each parity check equation. Therefore, the parity check equations are orthogonal on blocks, and the resulting code is t block error-correcting. Of course, an interleaved t error-correcting self-orthogonal convolutional code is also t block error-correcting, but the advantage of the threshold decoder described here is that it requires roughly $1/n$ of the memory of an interleaved ($n - 1$, n) self-orthogonal convolutional code. The decrease in memory can be significant for the high code rates of interest here.

As an example, the construction of a rate-$\frac{4}{5}$ 2 block error-correcting code is given. As given in Reference 3 (p. 419), the generator polynomial taps for a rate-$\frac{4}{5}$ 2 error-correcting self-orthogonal convolutional code are:

$$g_1: 0,16,20,21$$
$$g_2: 0,2,10,25$$
$$g_3: 0,14,17,26$$
$$g_4: 0,11,18,24$$

The generator polynomials can be shifted with respect to one another so that each column has at most one generator polynomial tap, as shown in Figure 11. The resulting encoder and decoder for the 2 block error-correcting code is shown in Figure 12. The encoder can be implemented with 2,472 bits of memory and the decoder with 3,192 bits. An interleaved 2 error-correcting (4,5) self-orthogonal convolutional code requires approximately 5 times as much memory, 11,366 bits for the encoder and 14,382 bits for the decoder.
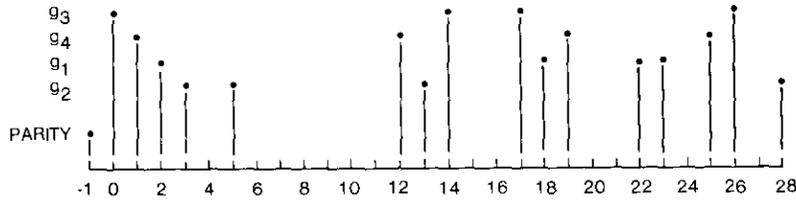
Figure 11. *Generator Polynomials Shifted so That Each Column Has at Most One Tap*

BER performance

As seen from the code construction, the BER performance of the t block error-correcting $(n - 1, n)$ block orthogonal convolutional code is the same as that of the t error-correcting $(n - 1, n)$ self-orthogonal convolutional code. The BER performance of the $(n - 1, n)$ block orthogonal convolutional code is calculated for $t = 2$ and $n = 2, 3, 4, 5, 6, 7,$ and 8, using equation (8), and is shown in Figure 13.

Other codes

The performance of the threshold decoder for an orthogonal convolutional code may be improved by using weight information supplied by the BCH decoder or by using information about the BCH decoded channel. In decoding with weight information, the BCH decoder assigns a weight W, $W = 0, 1, 2,$ or 3, to each decoded bit according to whether 1 or 2 errors have been corrected or 0 or 3 errors detected for that decoded block. The threshold decoder then uses the additional reliability information to form a maximum likelihood estimate of each checked information bit, based on the weight assigned to each bit in the parity check equations.

In decoding with channel information, the collection of 128 syndromes for each BCH-decoded 128-bit block is used to form a block syndrome. Since each BCH-decoded block is either error free or contains three or more errors,
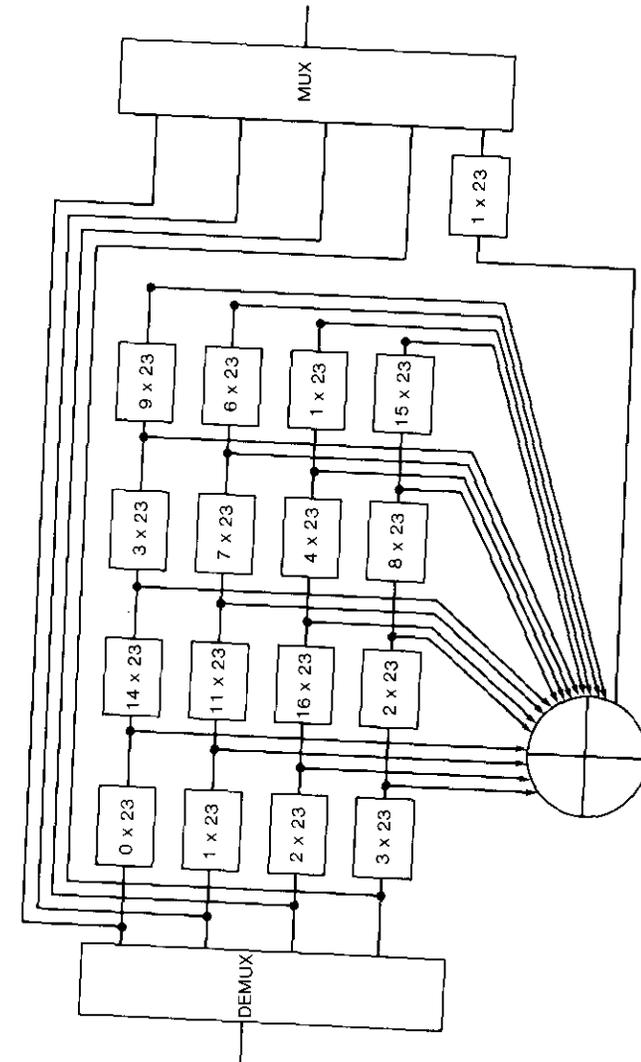
Figure 12a. *Encoder for 2 Block Error-Correcting Code*
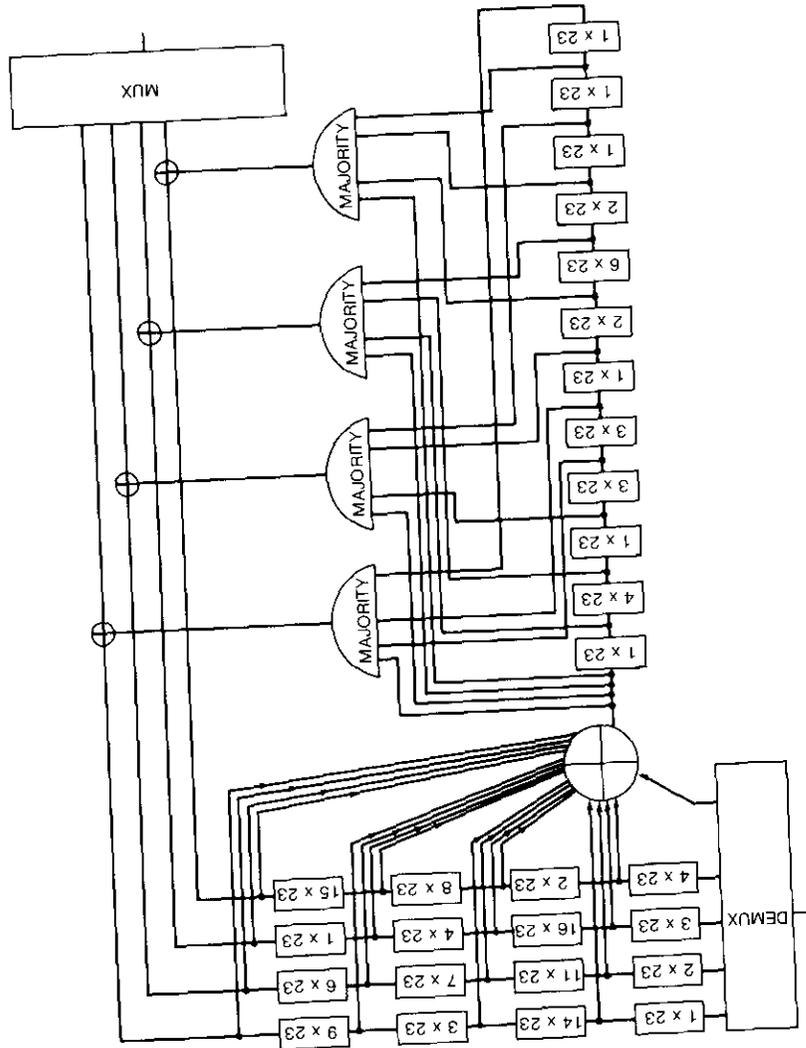
Figure 12b. *Decoder for 2 Block Error-Correcting Code*
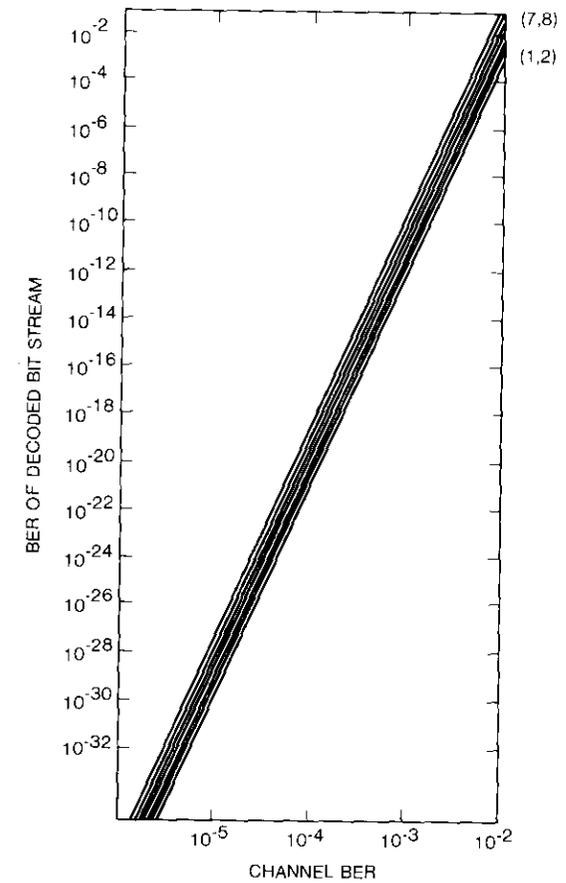
Figure 13. *BER of 2 Block Error-Correcting Convolutional Codes Concatenated With (127,112) BCH Inner Code*

the block syndrome gives information about which decoded blocks are in error, allowing additional bit errors to be corrected.

A second form of threshold decoding with channel information may be obtained by noting that a BCH-decoded block error produces an error pattern in the syndrome register of the threshold decoder, which is related to the register lengths and tap positions in the decoder. By choosing appropriate register lengths and tap positions, and by using information about the number of bit errors in a block error, the number of block errors in the syndrome register can be estimated, allowing correction of additional bit errors.

The performance improvement obtained by using threshold decoding with weight information or channel information is not sufficient to warrant the additional decoder complexity; therefore these schemes are not considered further.

Other outer codes which might be considered are the convolutional code with Viterbi decoding and the Reed-Solomon code. However, implementation of the Viterbi decoder and Reed-Solomon code is significantly more difficult than the block orthogonal convolutional code; for that reason, these codes are not considered further.

Recommended approach

The 2 block error-correcting $(n - 1, n)$ block orthogonal convolutional code is recommended for use in the INTELSAT TDMA terminal, since it gives significant BER improvement for an INTELSAT data channel and is easily implementable, requiring $1/n$ of the codec memory of a $t = 2$ $(n - 1, n)$ interleaved self-orthogonal convolutional code. The major implementation complexity of the codec is the buffer, which can be easily implemented with off-the-shelf components.

The convolutional code is placed in position C, since in this position, coding may be added without hardware or procedural modification to the current terminal; the rate loss resulting from coding is least, since only desired channels are coded; system operability is the most flexible; codec hardware is inexpensive because of the relatively low clock speed and minimal interface to the terminal; and the code need not be flushed. It is possible for the entire codec to be placed on a single LSI chip, thus reducing costs if a large number of codecs are used in position C.

Because of the relatively small variation in BER performance as n increases, n should be as large as possible in order to reduce coding overhead, consistent with the channel data rate to be encoded. Channel data rates are expected to be 1.2, 2.4, 4.8, 9.6, 56, 64, 128, and 144 kbit/s and 1.544 and 2.048 Mbit/s. A 2.048-Mbit/s channel consists of 32 SCS; therefore, a convenient choice of n is 8, so that the coded channel occupies 37 SCS. The throughput efficiency of the concatenated code is then $7/8 \times 7/8 = 77$ percent. For the 1.544-Mbit/s channel, $n = 8$; so the coded channel occupies 28 SCS, and the throughput efficiency is 77 percent. A 144-kbit/s data channel can be coded into 3 SCS with $n = 4$, with a throughput efficiency of $7/8 \times 3/4 = 66$ percent. A 128-kbit/s data channel is coded with $n = 3$, with a throughput efficiency of $7/8 \times 2/3 = 58$ percent. A 64-kbit/s channel is coded with $n = 2$, with a throughput efficiency of $7/8 \times 1/2 = 44$ percent. The 56-kbit/s data channel is coded with $n = 2$, using two SCS for a throughput efficiency of 38 percent.

Data channels of rate 9.6 kbit/s and less can be coded first and then multiplexed into 64-kbit/s data channels, or multiplexed into 64-kbit/s data channels (according to CCITT Recommendation X.50), and then coded. When the low-speed modem channels are individually coded, the interleaving depth need not be as large as for higher data rate channels, since TDMA burst transmission ensures statistical independence of data bits separated by less than 112 bits.

The recommended approach shows that additional channel coding can significantly lower the BER of an INTELSAT data channel. The codec is inexpensive to implement and can be added to the INTELSAT terminal with no hardware or procedural modifications. The revenue loss brought about by additional coding overhead is small because the code rate can be made large, and individual data channels are encoded.

Acknowledgments

References

[1] J. S. Snyder and T. Muratani, "Forward Error Correction for Satellite TDMA in the INTELSAT V Era," AIAA 8th Communications Satellite Systems Conference, Orlando, Florida, April 20–24, 1980, *Proc.*, pp. 674–683.

[2] W. W. Peterson and E. J. Weldon, Jr., *Error Correcting Codes*, Cambridge, Massachusetts: The MIT Press, 1972.

[3] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, New Jersey: Prentice-Hall, 1983.

Ken M. Mackenthun received a B.S.E.E. from George Washington University in 1974, and an M.S.E.E. and Ph.D. from the University of Illinois in 1976 and 1977, respectively. Since 1980, he has been a Member of the Technical Staff at COMSAT Laboratories in the Communications Systems Studies Department. He is interested in signal design for the INTELSAT channel and the mobile-terminal-to-satellite transmission link.

Pseudo-bit-error-rate measurement for 120-Mbit/s TDMA

J. S. SNYDER AND W. J. HERSEY

Abstract

The INTELSAT time-division multiple-access (TDMA) system monitor determines bit error rate (BER) for individual 120-Mbit/s TDMA bursts through on-line application of a shifted-phase-threshold pseudo-error technique which does not interrupt normal data traffic. Performance equations are developed for the ideal quaternary phase-shift-keyed (QPSK) linear channel, while computer simulations extend the analysis to the nonlinear satellite channel with various impairments. Results are compared with the measured performance of a 120-Mbit/s hardware implementation.

The relationship between BER and pseudo-error rate varies little over a wide range of channel conditions. Accuracy of the BER estimate is primarily limited by phase error and by random measurement variation resulting from a finite observation interval. Phase-error effects are investigated through computer simulation, while finite interval effects are readily calculated. Pseudo-error zone width may be selected to trade accuracy for measurement speed so that a pseudo-error measurement may be completed in a much shorter interval than that required for an actual BER measurement of identical accuracy.

Introduction

It is often desirable to monitor the performance of a digital communications link without interrupting normal message flow. Further, an individual measurement should be made in a reasonably short time. Pseudo-error

measuring techniques accomplish these objectives by employing a secondary received path in parallel with the main data path [1]–[6]. The secondary path is intentionally degraded in the sense that countable pseudo-error events in the secondary path occur more frequently than do bit errors in the main path.

The INTELSAT TDMA system monitor (TSM) [7],[8] measures the frequency with which the received QPSK signal falls within predetermined angular pseudo-error regions, as shown in Figure 1a. The result is used to develop an on-line estimate of BER, known as the pseudo-BER (PBER), for individual 120-Mbit/s TDMA bursts. This shifted-phase-threshold detection technique [3] is similar to that reported by Gooding [1] and Leon et al. [2], but requires only a single PBER monitor and does not involve extrapolation. Being immune to amplitude errors, this method is particularly useful in a phase modulation

(a) SHIFTED-PHASE-THRESHOLD

(b) SHIFTED-AMPLITUDE-THRESHOLD
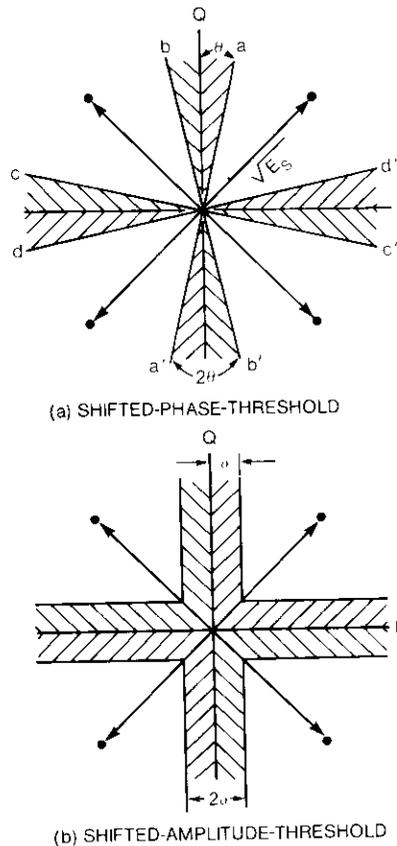
Figure 1. *Shifted-Threshold Detection Pseudo-Error Zones for QPSK*

application such as the INTELSAT TDMA system, where phase is controlled within approximately $\pm 2°$; but demodulated baseband amplitudes vary over a range of two to one.

New and useful performance approximations are developed in this paper, and computer simulation results for typical channel characteristics are reported. Although both bit error probability (P_e) and pseudo-error probability (P_p) versus E_b/N_o vary widely for different kinds of channels, a significant result is that the linear relationship between $\log P_e$ and $\log P_p$ varies little for channels ranging from the ideal linear case to the most nonlinear INTELSAT V case likely to be encountered in practice.

Sources of measurement error in a practical application are discussed, as well as tradeoffs to minimize error. It is shown that the size of the pseudo-error regions (i.e., 2θ) may be selected to trade accuracy for measurement speed. Consequently, a pseudo-error measurement may be completed more quickly than an actual BER measurement of equal accuracy.

A hardware implementation applicable to the 120-Mbit/s INTELSAT TSM is presented, and hardware measurements are compared with computer simulation results.

Pseudo-error measurement techniques

Pseudo-error measuring techniques involve a signal path in parallel with the primary data path in which a parameter is controlled to produce pseudo-errors more frequently than bit errors. The controlled parallel signal path not only permits performance measurement without interrupting message traffic which would be necessary if a test sequence had to be inserted, but also enables the pseudo-error measurement to be completed in less time than a true BER measurement. A BER estimate is obtained from the pseudo-error measurement by means of a known transfer characteristic relating P_e and P_p.

Various methods for generating pseudo-errors are possible [1]–[6],* the most important of which are:

a. shifted-phase-threshold detection,
b. shifted-amplitude-threshold detection,
c. sampling point offset,
d. noise addition, and
e. intersymbol interference.

---

* Reference 4 contains an extensive list of references on pseudo-error and other methods of error rate monitoring.

In the shifted-phase-threshold detection method [1]–[3], the normal decision boundaries between QPSK phase states are shifted by $\pm \theta$, as shown in Figure 1a. Pseudo-errors are counted whenever the received signal falls within the resulting pseudo-error zones of width 2θ. Although influenced by system phase errors, this technique is unaffected by amplitude fluctuations in the demodulator's received IF signal.

The shifted amplitude threshold detection method [4]–[6] applies amplitude shifts to the normal QPSK decision boundaries (Figure 1b). The analog I and Q demodulator signals are individually compared to their respective thresholds to determine whether either falls within a pseudo-error zone. This scheme is extremely sensitive to amplitude variations in the IF signal.

Sampling point offset [4]–[6] involves degrading the pseudo-error signal path by intentionally adjusting the received clock for nonoptimum sampling. Comparing the resulting data stream with the modem output data obtained from correct sampling yields a pseudo-error signal. A disadvantage of this method is its sensitivity to clock phase error.

The noise addition method [3]–[6] processes internal demodulator signals to obtain a demodulated baseband waveform which is more highly degraded by noise than is the normal demodulator output. The degraded output is compared with the normal output to obtain the pseudo-error signal. This method usually requires access to critical internal modem signal points and is not suitable for use outside the modem. In addition, the amount of added noise may not be easily varied.

The intersymbol interference technique [4]–[6] employs a narrower receive filter bandwidth than that required for optimum detection so as to degrade the baseband output through intentional intersymbol interference. Comparing the degraded output to the normal output produces the pseudo-error signal. The amount of degradation critically depends on a special filter, and the degradation introduced is not readily adjusted.

Shifted-phase-threshold pseudo-error theory

The INTELSAT TSM employs the shifted-phase-threshold technique (Figure 1a) for its PBER monitor because of the previously mentioned technical advantages and its suitability for use outside the modem. This pseudo-error method is especially well-suited to use with PSK modulation because phase is usually tightly controlled in a PSK system, whereas amplitude is not. The symmetry of the pseudo-error regions obviates the need for a known test sequence by making pseudo-error occurrence independent of the data pattern.

The performance of this shifted-phase technique and its hardware implementation are examined in the remainder of this paper. An expression for

pseudo-error probability versus E_b/N_o is first presented and compared to computer simulation results. The relationship between P_p and P_e is then developed.

The pseudo-error probability for QPSK transmitted over an ideal linear channel with additive white Gaussian noise (AWGN) is shown in Appendix A to be a function of the distances from each nominal received vector to the pseudo-error zone boundaries and is given by

$$P_p \cong 2\left\{ Q\left[\sqrt{\frac{2E_b}{N_o}}\,(\cos\theta - \sin\theta) \right] - Q\left[\sqrt{\frac{2E_b}{N_o}}\,(\cos\theta + \sin\theta) \right] \right\} \quad (1)$$

where the Q function is defined as

$$Q(\alpha) \triangleq \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-x^2/2}\, dx \quad .$$

Appendix A also shows that equation (1) may be further simplified over the normal range of interest to

$$P_p \cong 2Q\left[\sqrt{\frac{2E_b}{N_o}}\,(\cos\theta - \sin\theta) \right] \quad , \qquad \frac{E_b}{N_o} \geq 7\, dB \quad . \qquad (2)$$

This result is similar to the well-known expression for QPSK error probability over an ideal AWGN channel which is

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_o}} \right) \quad . \qquad (3)$$

Hence, the pseudo-error curve will be similar in shape to the BER curve. However, because the argument of the Q-function in equation (2) is smaller than that of equation (3), and the Q-function value is doubled to determine P_p, the pseudo-error probability is much greater than the bit error probability at a given value of E_b/N_o.

To compare pseudo-error and bit error performance over realistic satellite channels, computer simulations were used. An available satellite channel modeling program [9], capable of including such characteristics as channel filters and power amplifier nonlinearities, was utilized. The capability of specifying a phase offset (i.e., applying a phase rotation to the received IF signal) and computing pseudo-error probability was added to the program.

Simulation of the INTELSAT V channel was accomplished with a 10-dB

input backoff for the earth station's high-power amplifier (HPA) characteristic and a 2-dB input backoff for the traveling wave tube (TWT) spacecraft amplifier. Square-root raised cosine filters with 40-percent rolloff were used for the modulator and demodulator to produce Nyquist signaling. The computer simulation included input and output multiplexer filters in the satellite, up- and down-converter filters, and a down-link group delay equalizer.

The nonideal channel characteristics selected for a given simulation resulted in a scattering of received symbols around the nominal four phase states of Figure 1a before the addition of noise. For each such received symbol, bit error probability and pseudo-error probability were calculated based on the actual position of the perturbed signal vector. These quantities were then averaged over many received symbols to obtain the desired results.

Both P_e vs E_b/N_o and P_p vs E_b/N_o vary greatly for channels ranging from an ideal linear to a worst INTELSAT V nonlinear case, as shown in Figure 2. However, the relationship between P_e and P_p, which is of primary interest, deviates very little over the same wide range of channel conditions (Figure 3). Moreover, since the log-log plot of P_e vs P_p is almost a straight line for
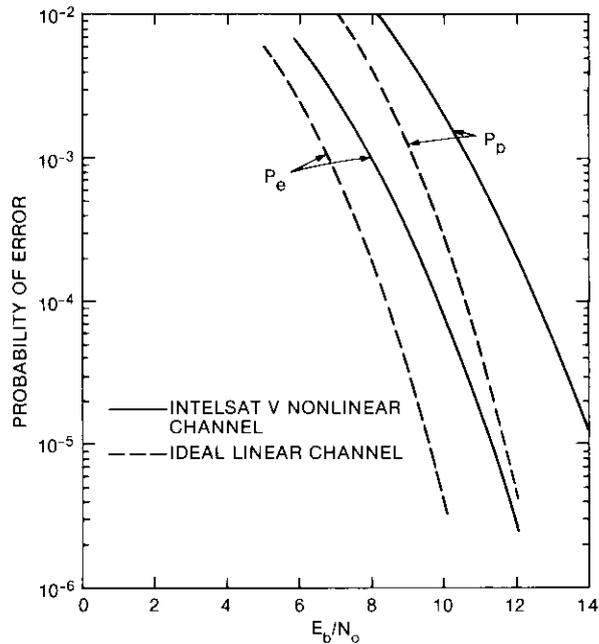
Figure 2. *Simulated Error and Pseudo-Error Probability for* $\theta = 10°$
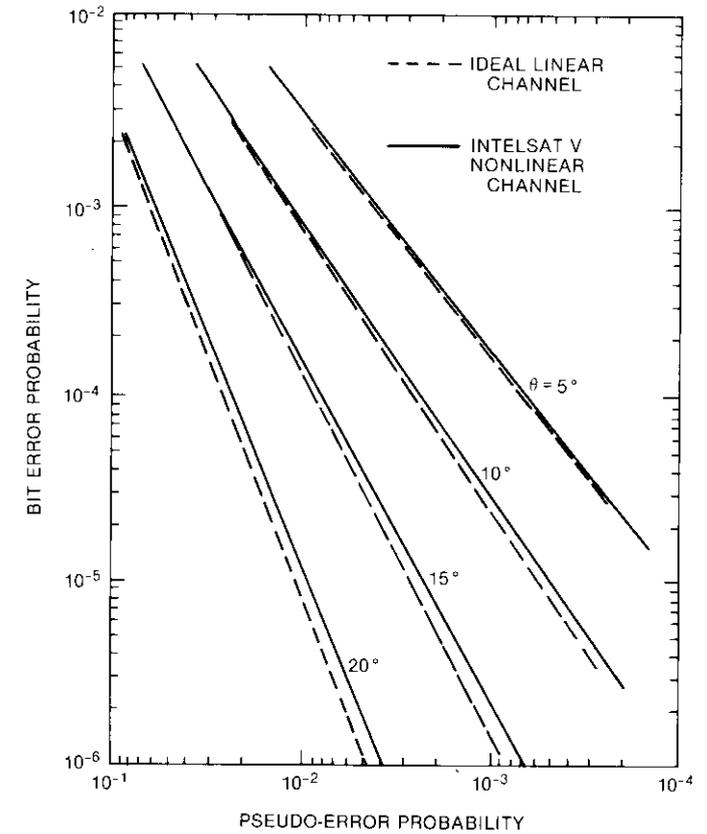
Figure 3. *Simulated Linear and Nonlinear Channel Performance*

the ideal linear AWGN channel, the relationship can be expressed as

$$\log P_e \cong \log C + k \log P_p \qquad (4)$$

or

$$P_e \cong CP_p^k \qquad (5)$$

where C and k are constants which are determined by the choice of the pseudo-error zone angle, θ. [From equations (1) and (3), it may be shown that C is not truly constant, but actually a very weak function of E_b/N_o]. To trade accuracy for measurement speed, C and k may be altered by changing θ.

Measurement error

For a given bit error probability, increasing the value of θ increases the pseudo-error probability (Figure 3), thereby permitting a measurement to be made more quickly. However, the steeper slope which accompanies a larger θ magnifies the effect of pseudo-error measurement inaccuracy on the computed value of BER.

This error magnification may be shown explicitly by noting that for a small deviation, ΔP_p, in pseudo-error probability, the corresponding deviation, ΔP_e, in bit error probability may be determined by differentiating equation (5) with respect to P_p to obtain

$$\frac{\Delta P_e}{\Delta P_p} \cong \frac{kP_e}{P_p}$$

or

$$\frac{\Delta P_e}{P_e} \cong k\left(\frac{\Delta P_p}{P_p}\right) \quad , \tag{6}$$

where k is the slope of the log P_e vs log P_p curve (Figure 3). Thus, the percentage error in determining pseudo-error probability is multiplied by the slope constant k (which is greater than unity) to yield a corresponding percentage error in the determination of P_e.

Sources of measurement error

Errors made in determining pseudo-error probability may be grouped into four types, depending on whether they are the result of:

    a. finite measurement interval,
    b. nonideal channel,
    c. time-varying phase error, or
    d. time-varying amplitude error.

The inaccuracy in a pseudo-error measurement caused by a finite measurement interval comes about because the parameter being observed is a random variable. This randomness results in a measurement variation from one interval to another, which may be reduced by increasing the observation interval.

Nonideal channel characteristics, such as realizable filter parameters and amplifier nonlinearities, offset the linear relationship between log P_e and log P_p and introduce some curvature (Figure 3). In principle, such effects can be calibrated out if they are common to all bursts. This can be accomplished

by storing a unique calibration curve that incorporates actual system conditions. Practical considerations in regard to this calibration will be discussed in a later section.

The effect of time-varying channel parameters cannot be eliminated with calibration. Any real or apparent burst-to-burst variation in the phase positions of the received IF signal will affect the pseudo-error measurement. Phase error shifts a received signal vector toward a pseudo-error zone, increasing the probability that noise will cause a pseudo-error. Phase error also increases the probability of a bit error. Since the argument of the nonlinear Q-function is different for P_p than for P_e [equations (2) and (3)], the two error rate functions do not track perfectly in the presence of phase error.

A time-varying phase error may occur because of differences in phase balance between two modulators. This time-varying parameter will result in different pseudo-error measurements for the two TDMA bursts originating from these modulators, even when they each have the same BER. (Phase jitter caused by noise, however, will have approximately the same effect on all bursts with the same BER, and may be calibrated out.) Further, any degradation that appears to be phase error, such as skew between the demodulated I and Q analog signals or skew between analog signals and clock, will produce measurement inaccuracies.

The amplitude of the IF input signal to the demodulator will vary because of nonideal automatic gain control (AGC). This input variation will result in a burst-to-burst fluctuation as great as two to one in the levels of the analog I and Q signals from the INTELSAT demodulator to the PBER monitor. The shifted-phase-threshold technique (Figure 1a) is insensitive to such variations, unlike the shifted-amplitude-threshold technique (Figure 1b). It will become apparent in the discussion on implementation that IF amplitude fluctuations cause equal variations in the analog I and Q outputs, which cancel in the formation of I/Q ratios in the PBER monitor.

Magnitude of measurement error

PBER measurement errors with the selected shifted-phase technique are primarily a result of the finite measurement interval and time-varying phase error. The effects of these two error sources will now be quantified.

The problem of measuring pseudo-error probability (or bit error probability) is one of estimating the mean value of a population. Since pseudo-errors (or bit errors) occur randomly, an estimate that is based on a finite number of observations does not necessarily yield the true mean, but converges toward it as the number of observations increases. For a large number of observations, the probability is P_c (known as the degree of confidence) that the estimate will differ from the true mean by less than K_c standard errors of the

mean, [10] where P_c is the area under the Gaussian curve between abcissa points $\pm K_c$ σ-units from the mean is given by $P_c = 1 - 2Q(K_c)$. When the number of observations, n, is taken from a much larger population, the standard error of the mean (see Appendix B) for a binary random variable, x, whose probability of being a 1 is p, is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}} \quad .$$

Therefore, with degree of confidence P_c, the error $(\Delta P_p)_{int}$ in a pseudo-error rate measurement made over an interval of n symbols is within

$$(\Delta P_p)_{int} = \pm K_c \sqrt{\frac{P_p(1-P_p)}{n}} \quad .$$

After normalization, the fractional range of uncertainty associated with confidence P_c may be written as

$$\left(\frac{\Delta P_p}{P_p}\right)_{int} = \pm K_c \sqrt{\frac{1-P_p}{nP_p}} \quad .$$

Since P_p is typically much less than unity, the fractional error in the pseudo-error measurement caused by the n-symbol interval is given, with confidence P_c, by

$$\left(\frac{\Delta P_p}{P_p}\right)_{int} \cong \pm \frac{K_c}{\sqrt{nP_p}} \quad . \tag{7}$$

Similarly, the statistical uncertainty in a direct BER measurement made over an n-bit interval is given, with confidence P_c, by

$$\left(\frac{\Delta P_e}{p_e}\right)_{int} \cong \pm \frac{K_c}{\sqrt{nP_e}} \quad . \tag{8}$$

Note that n expresses the measurement interval in symbols in equation (7), but in bits in equation (8).

Thus, the inaccuracy in estimating P_p that results from a finite measurement interval is inversely proportional to the square root of the average number of pseudo-errors in the interval. Further, estimate tolerance widens as the degree

of confidence in it becomes greater, because, as shown in Table 1, K_c increases as confidence level increases.

TABLE 1. DEGREE OF CONFIDENCE, P_c, VS NUMBER OF STANDARD ERRORS OF THE MEAN, K_c

P_c	K_c
0.85	1.44
0.90	1.65
0.95	1.96
0.99	2.58

A time-varying phase error can arise because of differences in phase balance among modulators associated with the various bursts in each TDMA frame. Differences in IF frequency offset will also cause phase error changes in the TSM demodulator. In the INTELSAT V TDMA system, phase errors are expected to be within $\pm 2°$, with 95-percent confidence.

The effect of phase error was investigated in the computer simulations by introducing a fixed phase rotation in the IF signal. A $2°$ phase error is barely perceptible at high values of P_e (Figure 4), but produces a detectable increase in pseudo-error rate at lower values of P_e. Moreover, the effect of phase error worsens with increasing θ.

Parameter tradeoffs

Increasing the value of θ reduces estimation errors caused by a finite measurement interval by increasing the pseudo-error probability corresponding to a given bit error probability [see equation (7)]. However, the inaccuracy caused by phase error increases with increasing θ, as noted above. In addition, the slope, k, of the log P_e vs log P_p transfer curve becomes greater with increasing θ, thereby amplifying the effect of total pseudo-error inaccuracy on the predicted BER value according to equation (6). Experience has shown that total BER inaccuracy is minimized when θ is chosen to make the two error components (finite measurement interval and phase error) approximately equal.

Note that P_e estimation accuracy improves rapidly as the error rate increases. Finite measurement interval error decreases with increasing pseudo-error rate as more pseudo-error events occur within the n-symbol observation window. The effect of phase error on P_e estimation also diminishes at higher error rates (Figure 4). Thus, accuracy is primarily a concern at the minimum error
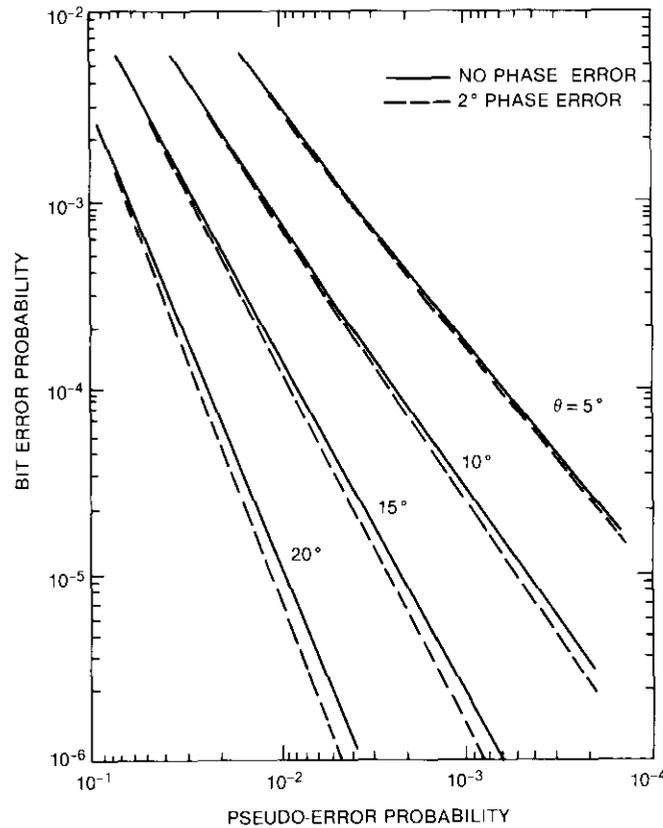
Figure 4. *Phase Error on Simulated INTELSAT V Channel*

rate for which performance is specified, which for the TSM is a BER of 1×10^{-5}. Finally, accuracy may be improved by increasing the number of symbols in the measurement window [equation (7)].

Since finite interval error and phase error are independent random variables, their combined effect on the pseudo-error measurement has a variance given by the sum of the individual variances [11]. Thus, when each component's tolerance is stated with the same degree of confidence P_c, the total pseudo-error tolerance, also with confidence P_c, is given by the square root of the sum of the squares of the individual tolerances, that is,

$$\frac{\Delta P_p}{P_p} = \sqrt{\left(\frac{\Delta P_p}{P_p}\right)^2_{\text{int}} + \left(\frac{\Delta P_p}{P_p}\right)^2_{\phi}} \tag{9}$$

where the subscript ϕ denotes the contribution resulting from phase error.

The TSM objective is to achieve a BER tolerance of around ± 20 percent with 95-percent confidence. Moreover, to minimize measurement time, this accuracy is to be achieved with a measurement interval of only 10^6 symbols. The interval is further subdivided by a 2,000-symbol measurement gate per 2-ms frame. By accumulating data on a given burst over 500 frames, the 10^6-symbol measurement interval is completed in 1 second.

A value of $\theta = 8°$ was selected to minimize total BER tolerance within the objectives of 95-percent confidence and a 10^6-symbol interval based on computer simulations. The nonlinear channel simulation results (Figure 5) may be used in equation (7) to compute the pseudo-error tolerance at a BER of 1×10^{-5} brought about by a measurement interval of 10^6 symbols. With
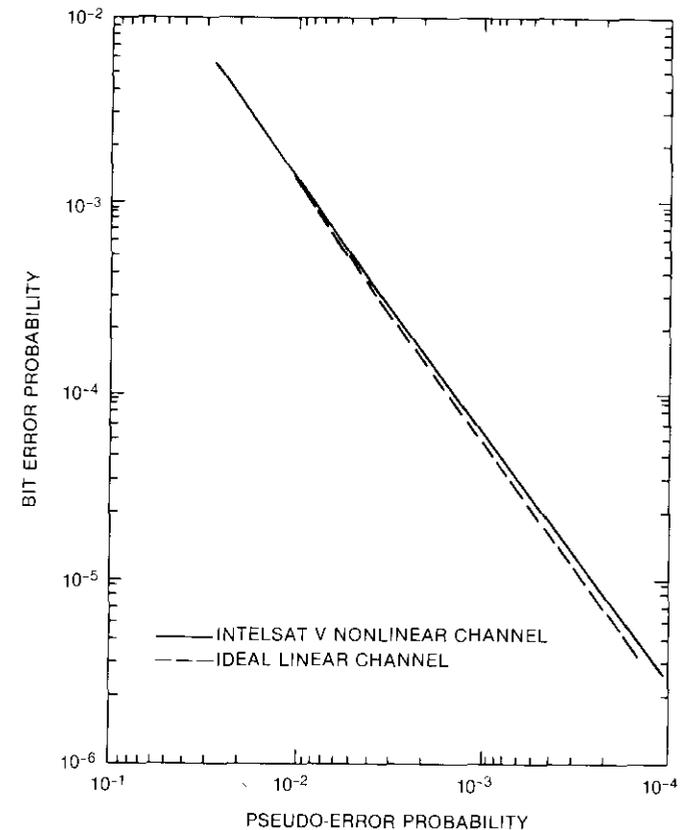
Figure 5. *Simulated Linear and Nonlinear Channels With $\theta = 8°$*

95 percent confidence, the tolerance is 12 percent (see Table 2). The $2°$ phase error simulation results (Figure 6) predict, with 95-percent confidence, a 9.6-percent pseudo-error inaccuracy at a BER of 10^{-5}. From equation (9), the combined pseudo-error tolerance, with 95-percent confidence, is found to be 15.5 percent. From equation (6), using a value of slope constant k = 1.343 (from Figure 5), predicted BER tolerance, also with 95-percent confidence, is found to be ±21 percent.

TABLE 2. PREDICTED MEASUREMENT ERROR AT $P_e = 1 \times 10^{-5}$ FOR $\theta = 8°$ WITH $2°$ PHASE ERROR (95-PERCENT CONFIDENCE)

n 10^6 symbols)	P_e	$\left(\dfrac{\Delta P_p}{P_p}\right)_{int}$	$\left(\dfrac{\Delta P_p}{P_p}\right)_{\phi}$	$\dfrac{\Delta P_p}{P_p}$	$\dfrac{\Delta P_e}{P_e}$
0.7	0.95	0.145	0.096	0.174	0.234
0.7	0.90	0.122	0.081	0.147	0.197
1.0	0.95	0.122	0.096	0.155	0.208
1.0	0.90	0.102	0.081	0.130	0.175
1.4	0.95	0.103	0.096	0.141	0.189
1.4	0.90	0.086	0.081	0.118	0.159
2.0	0.95	0.086	0.096	0.129	0.173
2.0	0.90	0.072	0.081	0.109	0.146

Table 2 illustrates the BER tolerance calculation for several different measurement intervals, as well as for two different confidence levels. The goal of an approximate 20-percent error could be met not only with a 10^6-symbol measurement interval and a 95-percent confidence, but also with a 7 $\times 10^5$-symbol interval if 90-percent confidence were acceptable. Moreover, increasing n beyond 10^6 would improve accuracy by reducing the component that resulted from the measurement interval. Ultimately, such an improvement would be limited by the phase error component.

Although the two error components are treated as random variables, they differ in the nature of their randomness. Finite interval error will vary randomly from one measurement to the next. Phase error will be relatively constant for a given transmitter, but will vary among transmitters. Therefore, combining the two effects to obtain a ±21-percent tolerance with 95-percent confidence means that for a large number of measurements taken over a large number of transmitters, 95 percent would be expected to be within ±21 percent of the actual error rate.

After the sources of measurement error have been properly accounted for,
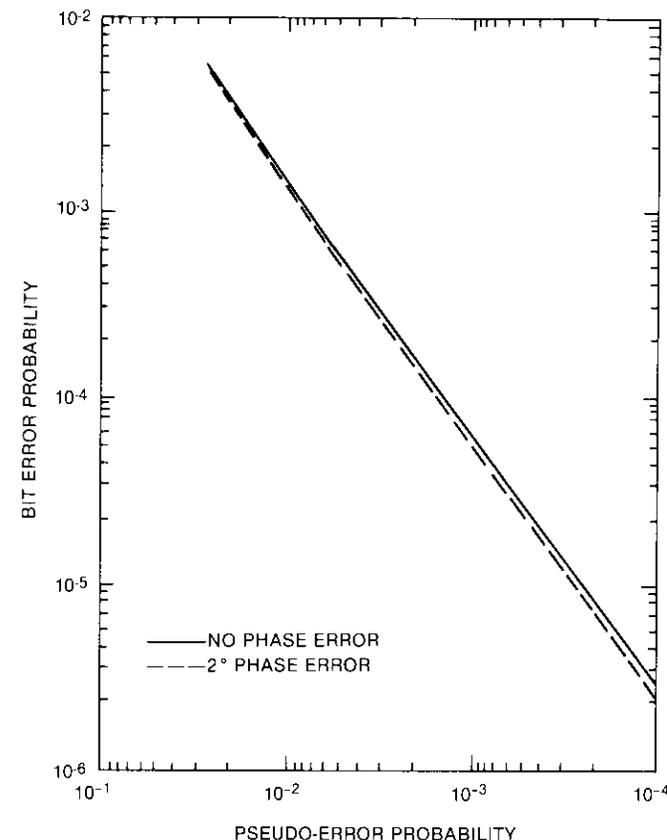
Figure 6. *Phase Error on Simulated INTELSAT V Channel With* $\theta = 8°$

it remains to be demonstrated whether the BER estimate developed by the PBER monitor is actually produced more quickly than a direct BER measurement of comparable accuracy would be. A pseudo-error observation interval of 10^6 symbols results in a confidence of 0.95 that the inaccuracy in the BER estimate will not exceed 21 percent. From equation (8), it is found that to achieve the same 21-percent tolerance with 95-percent confidence for a direct BER measurement would require 8.71 $\times$ 10^6 bits or 4.4 $\times$ 10^6 symbols. Under these circumstances, direct BER measurement would require 4.4 times as many symbols to achieve the same accuracy as that obtained by pseudo-error measurements.

Implementation

Filtered I and Q waveforms are provided to the PBER monitor (Figure 7) by the demodulator in analog form before sampling takes place. The pseudo-error zone comparators then determine where the received signal vector lies in the QPSK signal space (Figure 1a). (Figure 1a may be considered to represent either the amplitude and phase of the received IF signal relative to an unmodulated reference or the amplitudes of the demodulated I and Q baseband signals at the sampling instant, treated as quadrature components.) The number of times this vector falls in one of the pseudo-error zones is tallied in the pseudo-error counter.

Figure 7. *PBER Monitor*

A processor generates control signals necessary to initiate and terminate the measurement within the appropriate TDMA burst. The pseudo-error count is read once per TDMA frame and is accumulated by the processor, which determines BER by interpolating between stored points that define the relationship between pseudo-error rate and bit error rate. This BER estimate is then displayed on a CRT with other TSM measurement data.

Buffer amplifiers distribute the analog I and Q signals among a bank of four comparators (Figure 8). Comparator 1 provides a logical one output when the analog I and Q components define a signal vector which lies below the line c-c' in Figure 1a. Similarly, comparator 2 indicates when the demodulated signal vector lies below line d-d'. Comparator 3 or 4 responds with a logical one when the signal lies to the left of line a-a' or line b-b', respectively. Thus, θ is determined by the ratios of the resistors at the comparator inputs. Moreover, the comparator outputs respond to I/Q ratios rather than to absolute I and Q amplitudes. The demodulator clock is delayed by an adjustable amount to permit the comparator outputs to be clocked at the optimum sampling point.

By choosing comparator polarities as described above, an odd number of binary ones at the comparator outputs indicates that the demodulated signal falls within one of the pseudo-error zones. This condition is determined by exclusively-ORing the sampler outputs. A pseudo-error, which occurs during
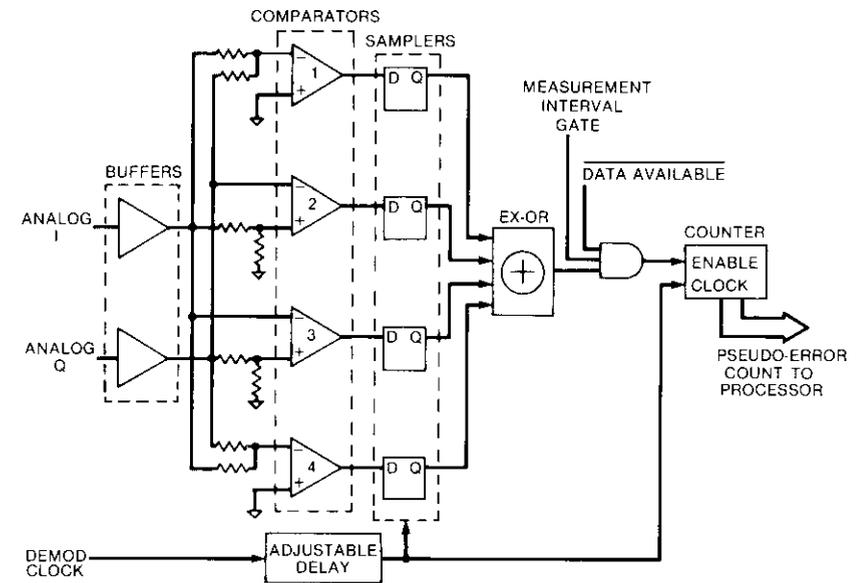
Figure 8. *Pseudo-Error Detection and Counting*

the measurement interval of the desired burst, enables a counter. At the end of the measurement interval, the processor reads the total from the pseudo-error counter.

Each such cycle is typically limited to 2,000 symbols. To achieve the desired accuracy, which requires a 10^6-symbol measurement interval, pseudo-error measurements on a given burst are accumulated over 500 frames by the processor before it causes the operation to step to the next burst. After the accumulation for a given burst has been completed, the pseudo-error rate is transformed into a BER estimate by using stored calibration data.

Alignment and calibration

The use of a 60-MHz symbol rate requires careful attention to circuit design, particularly with respect to timing. At this symbol rate, the effective sampling point for a given clock delay adjustment is uncertain relative to the center of either the I or Q eye pattern, because of propagation delays through buffer amplifiers and comparators and set-up time in the sampling flip-flops. Moreover, since a common clock is used to sample all four comparators, relative uncertainty exists among the four sampling instants, which is aggravated by skew between the analog I and Q signals. For these reasons,

clock timing is best adjusted for minimum pseudo-error rate, rather than on an absolute time basis that is related to maximum eye opening.

A variety of other possible misalignments will cause errors similar to those resulting from phase offset because they result in an incorrect value of the analog I and/or Q component of the signal vector being used, thereby simulating an erroneous phase rotation. Among these misalignments are unequal amplitude balance (*i.e.*, gain) between the analog I and Q data paths, DC offset in the I and Q paths, unequal delays between I and Q, and incorrect demodulator internal clock adjustment. The latter two are particularly critical in high-speed implementations.

Hardware testing showed that the PBER monitor is relatively insensitive to modest amounts of gain imbalance between the I and Q channels (Figure 9).
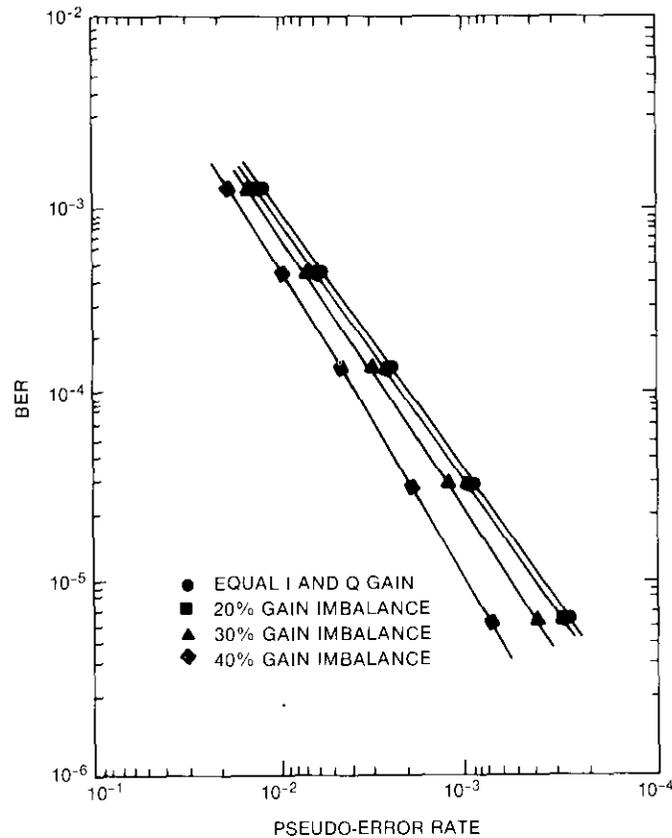
Figure 9. *Gain Imbalance Between I and Q Channels*

Very little effect was observed when the gain of one channel was deliberately adjusted to be 20-percent different from that of the other channel. For gain imbalance exceeding 20 percent, a noticeable degradation of measured performance was noted.

Correct DC balance is even more important. Although only a small effect was noted after adding a DC bias of up to 5 percent of the peak-to-peak analog signal level to one of the two channels (Figure 10), additional offset rapidly degraded performance.
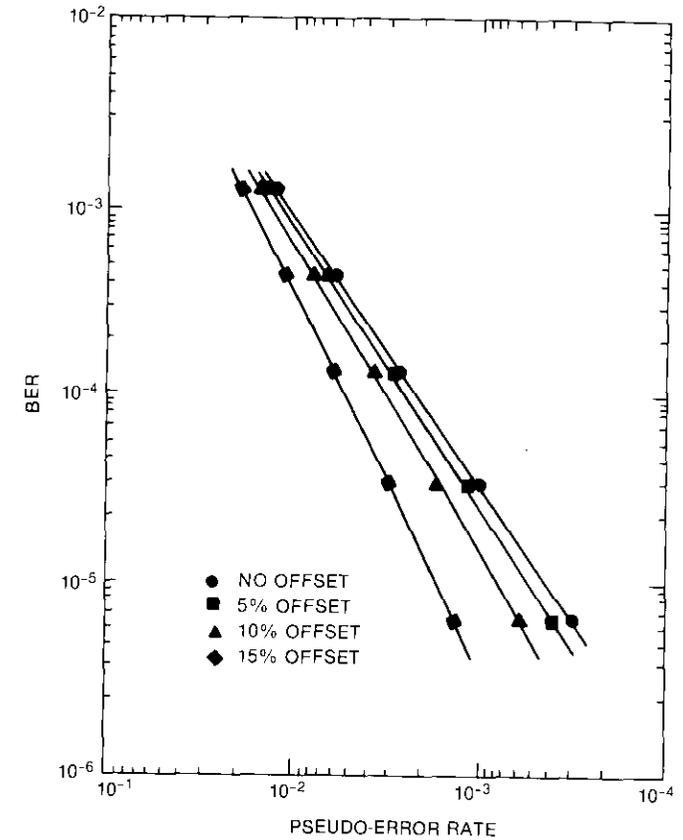
Figure 10. *DC Offset in I or Q Channel*

A unique calibration curve must be made for each PBER monitor and its demodulator because some demodulator variations may affect pseudo-error rate and some may not. Thus, during final calibration, the DC offset and

clock delay adjustments must be made with the actual modem to be used with the PBER monitor. A final calibration table that includes corrections for nonlinear channel operation must then be created.

Since the INTELSAT PBER monitor approximates the $\log P_p$ vs $\log P_e$ transfer curve with two straight-line segments, creation of the calibration table requires that measurements of BER and pseudo-error rate be made for three different BER values, with the modem connected back-to-back and the noise inserted at the 140-MHz IF interface. So that the accuracy of these measurements does not significantly degrade overall PBER accuracy, much greater numbers of bit errors and pseudo-errors must be counted for this calibration than for normal system operation. Measurement of these quantities to no greater than ± 2-percent error with 95-percent confidence requires that at least 10,000 events (errors or pseudo-errors) be observed at each point (equation 7). This requirement is most important at a BER of 10^{-5}, since the accuracy specification is most stringent there.

Measured values of pseudo-error rate must then be corrected because normal system operation will be conducted over the nonlinear INTELSAT V channel, while calibration measurements are taken with the modem looped back at the 140-MHz IF point. These corrections consist of constants to be added to the values of pseudo-error rate in constructing the calibration table, based on the measured performance of representative PBER and modem hardware over a nonlinear channel. The difference in PBER monitor performance between linear and nonlinear channels is small enough to permit fixed offsets in the calibration data in all cases, thereby avoiding the need for nonlinear calibration measurements on each modem.

Results

PBER monitor tests were performed with a prototype 120-Mbit/s modem from Nippon Electric Company (NEC) in a 140-MHz IF loopback configuration, and over an INTELSAT V hardware simulator (Figure 11). Hardware simulator tests were run for HPA/TWTA input backoffs in decibels of 14/14, 10/2, and 0/0. Since little variation was measured for these different channel conditions, for clarity, only the 10/2 case is presented here.

Similarly, data were taken for offsets of up to ± 30 kHz in the 140-MHz demodulator IF input. Although distinct shifts in P_e vs E_b/N_o and P_p vs E_b/N_o were noted, variations in P_p vs P_e were so small over the ± 30-kHz range that they are also omitted here.

The *IF* loopback test shows that the PBER monitor exhibits a response that would be expected if phase offset were present. This shift in the transfer curve, which is constant with time, is attributed to the aggregate of such real
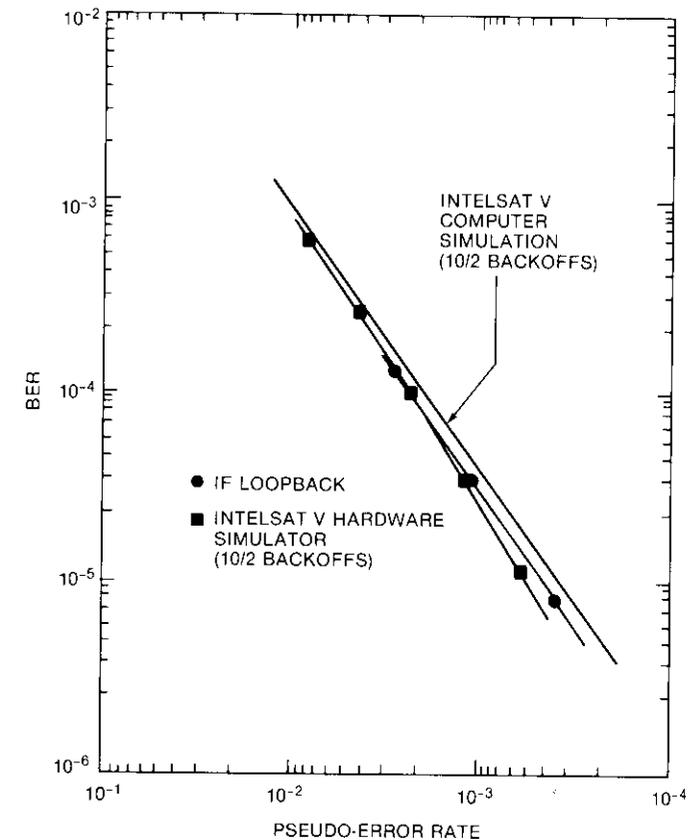
Figure 11. *120-Mbit/s PBER Monitor Measured Performance*

and apparent phase errors as have previously been discussed. Since each PBER monitor is uniquely calibrated with its corresponding demodulator to match the actual P_p vs P_e transfer curve, this effect is of no consequence.

PBER performance measured over the INTELSAT V simulator was not greatly affected by nonlinearities in the channel. The slight clockwise shift in the nonlinear channel transfer curve (Figure 11) is opposite to the slight counterclockwise shift predicted by computer simulations (Figure 3). This difference may be accounted for by noting that the predicted small shift in the P_p vs P_e relationship between ideal linear and INTELSAT nonlinear channels is the difference between the large shifts in P_p vs E_b/N_o and P_e vs E_b/N_o (Figure 2) that occur for the two channels. In a real hardware situation, the resulting small predicted difference may be masked by real or apparent phase errors

and other small degradations which may act in the opposite direction, that is, to increase the pseudo-error rate for a given BER. In any event, the slight difference between linear and nonlinear channel operation is accounted for in the stored calibration table.

Summary and conclusions

Performance of the shifted-phase pseudo-error technique has been analyzed. Equations have been developed for the ideal linear case, while computer simulations have been employed to extend the analysis to nonlinear channels and to include various channel impairments. These results have been compared with the measured performance of a 120-Mbit/s PBER monitor. The relationship between bit error probability and pseudo-error probability,

$$P_e \cong CP_p^k$$

has shown little variation for channels ranging from ideal linear to worst-case INTELSAT V nonlinear.

Sources of measurement error in a practical, high-speed implementation were discussed in detail, and tests confirmed the need for good balance between the analog I and Q channels. The use of the angular pseudo-error zones of the shifted-phase-threshold technique (rather than the rectangular zones of the shifted-amplitude method) was found to provide immunity from amplitude fluctuations in the demodulated signal that result from imperfect IF automatic gain control. Inaccuracies in the estimated BER were caused primarily by statistical variations resulting from a finite measurement interval and by phase error; these inaccuracies are best expressed as measurement tolerances with specified confidence levels. Finite interval errors are readily calculated, while phase error effects were investigated through computer simulation.

The PBER monitor was shown to be capable of accurately determining link performance without requiring that traffic be interrupted to transmit a BER measurement sequence. Further, while pseudo-error zone width, θ, may be varied to trade accuracy for measurement speed, a pseudo-error measurement may be completed in a much shorter interval than that of an actual BER measurement of identical accuracy. At a BER of 1×10^{-5} and an estimation confidence of 0.95, the PBER monitor achieves less than ± 21-percent error in a 10^6-symbol interval with $\theta = 8°$; a direct BER measurement would require more than four times as long for the same accuracy.

References

[1] D. J. Gooding, "Performance Monitor Techniques for Digital Receivers Based on Extrapolation of Error Rate," *IEEE Transactions on Communications*, COM-16, June 1968, pp. 380–387.

[2] B. J. Leon et al., "A Bit Error Rate Monitor for Digital PSK Links," *IEEE Transactions on Communications*, COM-23, May 1975, pp. 518–525.

[3] S. Takenaka et al., "Bit Error Rate Monitor for Four Phase PSK System," *ICC '80 Conference Record*, June 8–12, 1980, pp. 2.5.1–2.5.6.

[4] E. A. Newcombe and S. Pasupathy, "Error Rate Monitoring for Digital Communications," *Proceedings of the IEEE*, Vol. 70, August 1982, pp. 805–828.

[5] M. Keelty and K. Feher, "Pseudo-Error Detector Theory and Applications in QPSK 1.5 Mb/s Data Above FDM Voice Systems," *Proceedings NTC '77*, pp. 43:4-1–43:4-6.

[6] J. M. Keelty and K. Feher, "On-Line Pseudo-Error Monitors for Digital Transmission Systems," *IEEE Transactions on Communications*, COM-26, August 1978, pp. 1275–1282.

[7] D. J. Schaefer, "Techniques for TDMA System Monitoring," *COMSAT Technical Review*, Vol. 9, No. 2A, Fall 1979, pp. 387–412.

[8] J. F. Phiel and C. R. Thorne, "INTELSAT TDMA System Monitor," *Proc. Sixth International Conf. on Digital Satellite Communications*, September 19–23, 1983, pp. IV-14–IV-21.

[9] A. Hamid, S. R. Baker, and W. Cook, "A Computer Program for Communications Channel Modeling and Simulation," *COMSAT Technical Review*, Vol. 13, No. 2, Fall 1983, pp. 355–383.

[10] J. E. Freund, *Modern Elementary Statistics*, Englewood Cliffs, New Jersey: Prentice-Hall, 1979, pp. 228–251.

[11] M. Schwarz, *Information Transmission, Modulation, and Noise*, New York: McGraw-Hill, 1970, p. 351.

Appendix A. Derivation of pseudo-error probability

In this appendix, the probability of a pseudo-error is determined as a function of E_b/N_o and the pseudo-error zone width parameter, θ. The QPSK signal space representation is redrawn in Figure A-1 to illustrate a number of relevant quantities.

The probability of a pseudo-error is the probability of the noisy received signal falling within any of the four pseudo-error zones (I, II, III, or IV), given that any of the four possible phase states (A, B, C, or D) was transmitted. Since each of the possible transmitted phase states is equally likely, the probability of a pseudo-error is given by the probability of receiving a signal in region I, II, III, or IV, given that phase state A was transmitted, or

$$P_p = P[I|A] + P[II|A] + P[III|A] + P[IV|A] \quad . \tag{A-1}$$
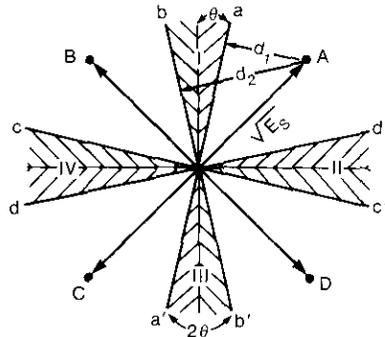
Figure A-1. *Shifted-Phase-Threshold Detection*

Since the probability of being in zone III or IV is much less than the probability of being in zones I or II, given that phase A was transmitted, the last two terms in equation (A-1) can be safely ignored. Furthermore, by the symmetry of the problem

$$P[I|A] = P[II|A] \quad .$$

Therefore, the probability of a pseudo-error is given approximately by

$$P_p \cong 2P[I|A] \quad . \tag{A-2}$$

Although an exact solution to equation (A-2) is difficult to obtain, the probability of the received signal falling within region I may be approximated by the probability of the signal falling to the left of boundary a-a' minus the probability of the signal falling to the left of boundary b-b', or

$$P[I|A] \cong P[(\text{left of a-a'})|A] - P[(\text{left of b-b'})|A] \tag{A-3}$$

(This expression actually yields the probability of being in zone I minus the probability of being in zone III. Since $P[III|A] \ll P[I|A]$, equation (A-3) is a very good approximation.)

Because of the geometry of the problem, equation (A-3) may be rewritten for the AWGN case as

$$P[I|A] \cong Q\left(d_1\sqrt{\frac{2}{N_o}}\right) - Q\left(d_2\sqrt{\frac{2}{N_o}}\right) \quad .$$

Expressing the distances d_1 and d_2 as functions of θ and noting that, for QPSK,

$$E_b = \frac{E_s}{2} \quad ,$$

the pseudo-error probability expressed by equation (A-2) becomes

$$P_p \cong 2\left\{Q\left[2\sqrt{\frac{E_b}{N_o}}\sin(45° - \theta)\right] - Q\left[2\sqrt{\frac{E_b}{N_o}}\sin(45° + \theta)\right]\right\} \quad . \tag{A-4}$$

Expansion of the trigonometric function $\sin(45° \pm \theta)$ yields

$$P_p \cong 2\left\{Q\left[\sqrt{\frac{2E_b}{N_o}}(\cos\theta - \sin\theta)\right] - Q\left[\sqrt{\frac{2E_b}{N_o}}(\cos\theta + \sin\theta)\right]\right\} \quad . \tag{A-5}$$

The Q-function rapidly approaches zero as its argument exceeds 3.5. For normal values of θ, this permits the second term in equation (A-5) to be ignored for values of E_b/N_o greater than or equal to about 7 dB. Thus, a practical approximation, valid for most cases of interest, is

$$P_p \cong 2Q\left[\sqrt{\frac{2E_b}{N_o}}(\cos\theta - \sin\theta)\right] \quad , \quad E_b/N_o \ge 7 \text{ dB} \quad .$$

Appendix B. Mean and variance of binomial random variable

Consider the random variable x, which has only the two values 0 and 1. The probability of each value is

$$P(x = 1) = p$$

and

$$P(x = 0) = 1 - p \quad .$$

The mean or expected value of x in any one binary interval is [B-1, pp. 144–145]:

$$\bar{x} = E\{x\} = p\,(1) + (1 - p)(0) = p \quad .$$

The variance of x is given by

$$\sigma^2 = E\{x^2\} - E^2\{x\} = [p(1)^2 + (1 - p)(0)^2] - p^2$$

or

$$\sigma^2 = p(1 - p) \quad .$$

For random samples of size n taken from an infinite population having the mean $\bar{x}$ and variance σ^2, the resulting sampling distribution of the mean [B-1, pp. 245–246],[B-2] also has the mean value $\bar{x}$ and variance ·

$$\sigma_{\bar{x}}^2 = \frac{\sigma r^2}{n}$$

The expression is also approximately correct for a finite population N, as long as N is much larger than n. The standard deviation $\sigma_{\bar{x}}$ is known as the standard error of the mean. For n-bit samples of the binomial random variable, x, the resulting estimate of $\bar{x}$ has a standard error of the mean given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1 - p)}{n}} \quad .$$

References

[B-1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill, 1965, pp. 144–145; pp. 245–246.

[B-2] J. E. Freund, *Modern Elementary Statistics,* Englewood Cliffs, New Jersey: Prentice-Hall, 1979, p. 228.

*John S. Snyder rceived a B.S.E.E. from Rochester Institute of Technology in 1965 and an M.S.E.E. from the University of Rochester in 1973. He joined COMSAT Laboratories in 1973, and is presently Associate Manager of the Transmission Processing Department, where he is engaged in the research and development of error-correction coding and other digital signal processing techniques applied to satellite communications. Previously, with Taylor Instrument Company and the Department of Defense, he was involved in re-*search and development in the areas of analog and digital signal processing and CRT display systems. Mr. Snyder holds several patents in the areas of error-correction coding and signal processing. He has authored several technical papers for international satellite communications conferences, and is a member of the IEEE.

William J. Hersey received a B.S.E.E. from the University of Maryland in 1981 and an M.S. from the Johns Hopkins University in 1984. He joined COMSAT Laboratories in 1979 while participating in a co-op internship, and became a Member of the Technical Staff upon completion of his undergraduate studies. He has been responsible for the design and development of several error-correcting codecs, and has also been involved in the design of various PSK modems. Mr. Hersey is a member of the IEEE, Tau Beta Pi, and Eta Kappa Nu.

*Equivalence classes of alternant codes**

A. L. BERMAN, H. J. HELGERT†, AND N. G. BERMAN‡

(Manuscript received February 1, 1984)

Abstract

The objective of this paper is to develop a structure for alternant codes and to categorize the resulting detailed code performance. Starting from Helgert's defining construction, a $1:1$ correspondence exists between the elements of a finite extension field, $GF(2^m)$, and the codes that they imply. Partitioning the field into equivalence classes leads to a complete description of the codes that may be generated for the given choice of extension field $GF(2^m)$ and subfield $GF(2^\lambda)$. All the elements within one equivalence class imply a set of equivalent codes, with identical length, information capacity, and code weight spectrum.

Theorems are developed to show that for an equivalence class consisting of the set of elements

$$\left\{ \left(\frac{1}{\alpha_r x_i + y_j} \right)^{2^\lambda} \right\}$$

where

$$\alpha_r \begin{cases} \varepsilon\ GF(2^m) \\[1em] \notin \text{any } GF(2^t) \subset GF(2^\lambda) \end{cases}$$

$$y \, \varepsilon \, GF(2^\lambda)$$
$$x \, \varepsilon \, GF(2^\lambda) - 0$$
$$0 \leq k \leq m - 1$$

the equivalence classes are disjoint, and for a given k, the elements $\{(\alpha_r x + y)^{2^k}\}$ are distinct. Further, the transform $\mathrm{Inv}(\alpha_r) \triangleq (\alpha_r^{2^\lambda} + \alpha_r)^{2^\lambda - 1}$ maps all the elements of a given equivalence class into a unique value $[\varepsilon \; GF(2^m)]$ and its conjugates, and therefore the equivalence class may be characterized either by its invariant, or by the minimum function of its invariant.

The computationally efficient algorithms originally developed to experimentally study the structure of alternant codes are subsequently used to detail code performance. With the equivalence classes limiting the number of cases to be tried, summary performance is given for the best (shortened H-matrix) codes found with an exhaustive search of all cases $m \leq 25$ and selected cases for $m > 25$. Several codes generated from the subfield $GF(2^7)$ were found to be better than any previously known, and cases were found in $GF(2^{25})/GF(2^5)$, $GF(2^{30})/GF(2^5)$, and $GF(2^{60})/GF(2^6)$ that exceeded the minimum distance and information bounds, where none were previously known.

Introduction

As satellite channels are increasingly used for computer-to-computer communications, efficient error correcting codes are of increasing interest for improving the error rates found in typical satellite transmission channels from $\approx 10^{-5}$ to 10^{-10} or better.

Alternant codes were first presented by H. J. Helgert, who described their algebraic construction and established minimum bounds on their information capacity and error correcting capability. Further, he showed that in their most general form, the set of all alternant codes contains codes that satisfy the Varsharmov-Gilbert bound as the code length $\rightarrow \infty$.

Alternant codes as a class are of interest for application to data transmission via satellite links, because these codes are relatively simple to decode (their decoding complexity is of the same order as BCH codes) [2]. In many specific cases, alternant codes are known to be either very close to, equal to, or better than the best known linear block codes [3].

This paper is concerned with a class of binary alternant codes defined by parity check matrices* of the form

$$H = \begin{bmatrix} (x_1 - \alpha)^{-1} & (x_2 - \alpha)^{-1} & \dots & (x_n - \alpha)^{-1} \\ (x_1 - \alpha)^{-2} & (x_2 - \alpha)^{-2} & \dots & (x_n - \alpha)^{-2} \\ \vdots & & & \\ (x_1 - \alpha)^{-2t} & (x_2 - \alpha)^{-2t} & \dots & (x_n - \alpha)^{-2t} \end{bmatrix}$$

where t is a positive integer, x_i are the elements of the Galois field $GF(2^\lambda = n)$, and $\alpha \neq x_i$ (the so-called code generator) is an element of some field $GF(2^m)$ which contains $GF(2^\lambda)$ as a proper subfield.

It is well known [2] that the minimum distance and dimensionality of these (n, k) linear codes are bounded by

$$d_{\min} \geq 2t \left(\frac{\mu}{\lambda} \right) + 1$$

$$k \geq n - \mu t$$

where μ is the smallest integer such that $GF(2^\mu)$ contains both α and $GF(2^\lambda)$, and is in turn a (not necessarily proper) subfield of $GF(2^m)$. For certain subclasses of codes, these bounds are met with equality; whereas for others, $d_{\min}$ and k are herein shown to exceed these values.

For fixed m and λ, the number of distinct (n, k) codes defined by H above is clearly $2^m - 2^\lambda$, corresponding to all permissible values of the code generator α in $GF(2^m)$. It is known that many of these codes have identical weight spectra, and are thus equivalent in this sense. The main objective in this paper is to investigate the relationship between the code generators that give rise to equivalent codes, and to determine the number of distinct weight spectra as a function of the code parameters m and λ. A transformation on α is defined which is then shown to leave the code's weight spectrum unaltered, and thus combines sets of code generators into equivalence classes. Further investigation of these equivalence classes leads to a mapping of the code generators into an element of $GF(2^m)$ that is unique to a particular equivalence class. Upper and lower bounds are then derived for the number of equivalence classes that are asymptotically tight as the code length becomes large. Finally, numerical results are provided for several codes, which are

---

* Defining code constructions for alternant codes are reviewed in the Appendix.

better than any previously published [3]. The weight spectra of codes for selected equivalence classes are also provided.

Equivalence classes of codes

Consider the set of elements of $GF(2^m)$ generated by mapping the code generator α into the element

$$\alpha \rightarrow (\alpha^{2^\ell}x + y)$$

where ℓ is an integer bounded by $0 \leq \ell \leq m - 1$, and x and y range over $GF(2^\lambda)$, with $x \neq 0$. This set, which will be called the equivalence class of α and will be denoted by $E(\alpha)$, clearly contains α, all of its conjugates, and their affine transformations over the field $GF(2^\lambda)$. Since

$$\left(\alpha^{2^{\ell_1}}x_1 + y_1\right)^{2^{\ell_2}}x_2 + y_2 = \alpha^{2^{\ell_1+\ell_2}}\left(x_1^{2^{\ell_2}}x_2\right) + \left(y_1^{2^{\ell_2}}x_2 + y_2\right)$$

the transformation applied to any element of $E(\alpha)$ leads to another element of $E(\alpha)$. The equivalence class is therefore closed. It is easily shown that none of the elements of $E(\alpha)$ belongs to $GF(2^\lambda)$. Thus, the above transformation induces a complete partitioning of the set of code generators into these equivalence classes.

The connection between an equivalence class of code generators and the corresponding code properties is established by

Theorem I

The codes generated by the elements of $E(\alpha)$ have identical weight spectra.

Proof: Without loss of generality, concentrate on the first row of H, denoted by R_1, and consider first the effect of the transformation $\alpha \rightarrow \alpha^{2^\ell}$. Clearly, R_1 is then mapped from the form

$$R_1 = \left[(x_1 - \alpha)^{-1}(x_2 - \alpha)^{-1} \ldots (x_n - \alpha)^{-1}\right]$$

into the form

$$R_2 = \left[\left(x_1 - \alpha^{2^\ell}\right)^{-1}\left(x_2 - \alpha^{2^\ell}\right)^{-1} \ldots \left(x_n - \alpha^{2^\ell}\right)^{-1}\right] \quad .$$

Raising the elements of R_2 to the $2^{m-\ell}$ power leaves the row space of R_2 over $GF(2)$ invariant and results in

$$R_3 = \left[\left(x_1^{2^{m-\ell}} - \alpha\right)^{-1}\left(x_2^{2^{m-\ell}} - \alpha\right)^{-1} \ldots \left(x_n^{2^{m-\ell}} - \alpha\right)^{-1}\right] \quad .$$

However, if $x_i \neq x_j$, then $x_i^{2^{m-\ell}} \neq x_j^{2^{m-\ell}}$. Thus, the set of $x_i^{2^{m-\ell}}$ must be a rearrangement of the set of x_i in R_1. Consequently, R_3 is at most a column permutation of R_1, and its row space over $GF(2)$ has the same weight spectrum as that of R_1 and R_2.

Next, multiply α by x and obtain from R_1:

$$R_4 = [(x_1 - \alpha x)^{-1}(x_2 - \alpha x)^{-1} \ldots (x_n - \alpha x)^{-1}]$$

$$= x^{-1}[(x_1 x^{-1} - \alpha)^{-1}(x_2 x^{-1} - \alpha)^{-1} \ldots (x_n x^{-1} - \alpha)^{-1}]$$

$$= x^{-1}R_5 \quad .$$

Again, if $x_i \neq x_j$, then $x_i x^{-1} \neq x_j x^{-1}$. Thus, R_5 is a column permutation of R_1. Multiplication of the elements of R_5 by x^{-1} produces a linear transformation of the rows of R_5 over $GF(2)$. Therefore, the row space over $GF(2)$ of R_5 is identical to that of R_1.

Finally, consider the effect of adding y to α in R_1. Then R_1 maps into

$$R_6 = [(x_1 - y - \alpha)^{-1}(x_2 - y - \alpha)^{-1} \ldots (x_n - y - \alpha)^{-1}]$$

which is easily seen to be a column permutation of R_1.

Equivalence class invariants

Consider now the function $I(\alpha)$ defined over the equivalence class $E(\alpha)$ and given by

$$I(\alpha) \triangleq \cdot (\alpha^n + \alpha)^{n-1}$$

where $n = 2^\lambda$.

First, note that

$$I(\alpha^{2^\ell}x + y) = [(\alpha^{2^\ell}x + y)^n + (\alpha^{2^\ell}x + y)]^{n-1}$$

$$= (\alpha^{n2^\ell}x + y + \alpha^{2^\ell}x + y)^{n-1}$$

$$= [x(\alpha^{n2^\ell} + \alpha^{2^\ell})]^{n-1}$$

$$= (\alpha^{n2^\ell} + \alpha^{2^\ell})^{n-1}$$

$$= I(\alpha^{2^\ell})$$

$$= [I(\alpha)]^{2^\ell} \quad .$$

Therefore, the transformation is many-to-one and maps the elements of $E(\alpha)$ into the conjugates of the field element $I(\alpha)$ for $0 \le \ell \le m - 1$.

Next, it will be shown that each distinct conjugate is associated with exactly $2^{\lambda}(2^{\lambda} - 1)$ elements of $E(\alpha)$. Toward this end, assume first that

$$[I(\alpha)]^{2^{\ell}} = [I(\alpha)]^{2^{k}}$$

for $\ell \ne k$.

Then

$$\left(\alpha^{n2^{\ell}} + \alpha^{2^{\ell}}\right)^{n-1} = \left(\alpha^{n2^{k}} + \alpha^{2^{k}}\right)^{n-1}$$

and the ratio

$$\frac{\alpha^{n2^{\ell}} + \alpha^{2^{\ell}}}{\alpha^{n2^{k}} + \alpha^{2^{k}}}$$

must be an element of $GF(2^{\lambda})$, say x.

Consequently,

$$\alpha^{2^{k}}x + \alpha^{2^{\ell}} = \alpha^{n2^{k}}x + \alpha^{n2^{\ell}}$$
$$= (\alpha^{2^{k}}x + \alpha^{2^{\ell}})^{n}$$

which implies that $\alpha^{2^{k}}x + \alpha^{2^{\ell}}$ is also an element of $GF(2^{\lambda})$, say y. Clearly then,

$$\alpha^{2^{\ell}} = \alpha^{2^{k}}x + y$$

which shows that the elements of $E(\alpha)$ associated with $[I(\alpha)]^{2^{\ell}}$ are exactly those associated with $[I(\alpha)]^{2^{k}}$. The number of such elements is easily seen to be $2^{\lambda}(2^{\lambda} - 1)$, since the two elements

$$\alpha^{2^{\ell}}x_{1} + y_{1} \text{ and } \alpha^{2^{\ell}}x_{2} + y_{2}$$

are equal if and only if either

$$\alpha^{2^{\ell}} = \frac{x_{1} + x_{2}}{y_{1} + y_{2}} \varepsilon\ GF(2^{\lambda})$$

(a condition ruled out by the definition of α), or else $x_{1} = x_{2}$ and $y_{1} = y_{2}$. Thus, for each distinct pair of (x, y), there exists a distinct element $\alpha^{2^{\ell}}x + y$ which maps into $[I(\alpha)]^{2^{\ell}}$. As an immediate consequence:

Theorem 2

The number of elements in $E(\alpha)$ equals $C \cdot 2^{\lambda}(2^{\lambda} - 1)$, where C is the number of distinct conjugates of $I(\alpha)$.

The conjugates of $I(\alpha)$ may be shown to be unique to $E(\alpha)$ by an argument similar to the one above. Thus, assume that α and β are two code generators and $I(\alpha) = I(\beta)$. Then

$$(\alpha^{n} + \alpha)^{n-1} = (\beta^{n} + \beta)^{n-1}$$

and

$$\left(\frac{\beta^{n} + \beta}{\alpha^{n} + \alpha}\right)^{n-1} = 1 \quad .$$

This implies that

$$\beta^{n} + \beta = x(\alpha^{n} + \alpha)$$

where $x\ \varepsilon\ GF(2^{\lambda})$. Then

$$\alpha x + \beta = \alpha^{n}x + \beta^{n} = (\alpha x + \beta)^{n}$$

and $\alpha x + \beta = y\ \varepsilon\ GF(2^{\lambda})$. Consequently,

$$\beta = \alpha x + y\ \varepsilon\ E(\alpha)$$

which shows that α and β must be in the same equivalence class in order for $I(\alpha)$ to equal $I(\beta)$.

It now follows that the conjugates of $I(\alpha)$, collectively, are a label on $E(\alpha)$, which is unique. A convenient form of this label may be obtained from the minimal polynomial $f_{\alpha}(x)$ of $I(\alpha)$, given by

$$f_{\alpha}(x) = [x - I(\alpha)][x - I(\alpha)^{2}] \ldots [x - I(\alpha)^{2^{c-1}}] \quad .$$

Clearly, $f_{\alpha}(x)$ is an irreducible polynomial over $GF(2)$ whose degree, C, must be a divisor of m. The coefficients of $f_{\alpha}(x)$ are taken as the binary representation of the equivalence class of codes, $E(\alpha)$.

A similar representation may be obtained for the complete set of equivalence classes for fixed m and λ by showing that the product of the minimal polynomials over all equivalence classes is given by

$$F(x) = 1 + \sum_{i=0}^{(m/\lambda)-2} x^{\left(\sum_{j=0}^{i} 2^{j\lambda}\right)} \quad .$$

To establish this, it will first be proved that $I(\alpha)$ is a root of $F(x)$ for every code generator α. Thus,

$$F[I(\alpha)] = 1 + \sum_{i=0}^{(m/\lambda)-2} \left(\alpha^{2^\ell} + \alpha\right)^{(2^\lambda-1)\sum_{j=0}^{i} 2^{j\lambda}}$$

$$= 1 + \sum_{i=0}^{(m/\lambda)-2} \left(\alpha^{2^\ell} + \alpha\right)^{\sum_{j=0}^{i} 2^{(j+1)\lambda} - \sum_{j=0}^{i} 2^{j\lambda}}$$

$$= 1 + \sum_{i=0}^{(m/\lambda)-2} \left(\alpha^{2^\ell} + \alpha\right)^{[2^{(i+1)\lambda} - 1]} \quad .$$

Multiplying both sides by $(\alpha^{2^\ell} + \alpha)$ yields

$$F[I(\alpha)]\left(\alpha^{2^\ell} + \alpha\right) = \sum_{i=-1}^{(m/\lambda)-2} \left(\alpha^{2^\ell} + \alpha\right)^{2^{(i+1)\lambda}}$$

$$= \sum_{i=-1}^{(m/\lambda)-2} \left(\alpha^{2^{(i+2)\lambda}} + \alpha^{(i+1)\lambda}\right)$$

$$= \alpha + \alpha^{2^m}$$

$$= 0 \quad .$$

Next, it is clear that, since the coefficients of $F(x)$ are binary, all the conjugates of $I(\alpha)$ are roots of $F(x)$.

Finally, since each conjugate of $I(\alpha)$ is associated with $2^\lambda(2^\lambda - 1)$ elements in $GF(2^m) \sim GF(2^\lambda)$, the total number of conjugates is exactly

$$\frac{2^m - 2^\lambda}{2^\lambda(2^\lambda - 1)} = \frac{2^{m-\lambda} - 1}{2^\lambda - 1} = 1 + 2^\lambda + 2^{2\lambda} + \ldots + 2^{m-2\lambda}$$

which of course is the degree of $F(x)$. Thus, $F(x)$ contains as roots all conjugates of $I(\alpha)$ for all equivalence classes and no other roots, and is

therefore the product of the minimal polynomials of $I(\alpha)$. As an immediate consequence:

Theorem 3

For fixed m and λ, the number of equivalence classes is equal to the number of irreducible factors of $F(x)$ over $GF(2)$.

Bounds on the number of equivalence classes

With the aid of Theorem 3, enumeration of the equivalence classes for fixed m and λ is in principle a straightforward procedure and quite manageable computationally for values of $2^{m-2\lambda}$ as large as several thousand (Reference 4, Chapter 6). Beyond those values, however, bounding techniques prove to be more productive in determining approximate answers. The following discussion develops upper and lower bounds on the number of equivalence classes which are asymptotically tight, as λ becomes large and reasonably accurate for moderate values of n.

The lower bound is based on the observation that the degree of the irreducible factors of $F(x)$ can be no larger than m. Consequently, since $F(x)$ is of degree $(2^{m-\lambda} - 1)/(2^\lambda - 1)$, the number of equivalence classes, N, must be at least $(2^{m-\lambda} - 1)/[m(2^\lambda - 1)]$. For large values of λ, this bound approaches the simple form

$$N \geq \frac{2^{m-2\lambda}}{m} \quad .$$

To obtain an upper bound on N, note that the number of irreducible factors of $F(x)$ cannot exceed the number of factors of the polynomial which is the product of all irreducible polynomials of degree $1, 2, \ldots$, and so on, taken in order until the degree of the product is at least that of $F(x)$. Thus, if I_k is the total number of irreducible polynomials of degree k over $GF(2)$, the number of factors of $F(x)$, and therefore N, is upper bounded by

$$N \leq \sum_{k=1}^{p} I_k + r$$

where p is the largest integer such that

$$(p + 1)r + \sum_{k=1}^{p} kI_k = \frac{2^{m-\lambda} - 1}{2^\lambda - 1}$$

and $0 \leq r \leq I_{p+1}$.

Using some well-known bounds on I_k [4], the left side of the last inequality can be shown to satisfy

$$\sum_{k=1}^{p} kI_k \geq 2^{p+1} - 2^{(p+3)/2} \quad .$$

Thus, for large $\lambda < m/2$, p may be set equal to $m - 2\lambda - 1$, resulting in

$$r \leq \frac{2^{m-2\lambda} - (2^{m-2\lambda} - 2^{(m-2\lambda+2)/2})}{m - 2\lambda}$$

$$= \frac{2^{(m-2\lambda+2)/2}}{m - 2\lambda} \leq I_{m-2\lambda}$$

and

$$N \leq \sum_{k=1}^{m-2\lambda-1} I_k + \frac{2^{(m-2\lambda+2)/2}}{m - 2\lambda} \quad .$$

Therefore, the following has been established:

Theorem 4

For sufficiently large $\lambda < m/2$, the number of equivalence classes is bounded by

$$\frac{2^{m-2\lambda}}{m} \leq N \leq \sum_{k=1}^{m-2\lambda-1} I_k + \frac{2^{(m-2\lambda+2)/2}}{m - 2\lambda} \quad .$$

The upper bound of Theorem 4 can in most cases be improved by recognizing that the degree of the irreducible factors of $F(x)$ must be a divisor of m. Then,

$$N \leq \sum_{\substack{k=1 \\ k|m}}^{p} I_k + r$$

where p is the largest divisor of m such that

$$(Sr) + \sum_{\substack{k=1 \\ k|m}}^{p} kI_k = \frac{2^{m-\lambda} - 1}{2^{\lambda} - 1} \quad ,$$

S is the next largest divisor of m following p, and $0 \leq r \leq I_s$.

As an example, let $m = 20$, $\lambda = 5$. The divisors of 20 are 1, 2, 4, 5, 10, and 20, and the corresponding I_k are given by $I_1 = 2$, $I_2 = 1$, $I_4 = 3$, $I_5 = 6$, $I_{10} = 99$, and $I_{20} = 52{,}377$. Consequently, since $(2^{m-\lambda} - 1)/(2^{\lambda} - 1) = 1{,}057$, then $p = 10$, $r = 2$, and $N \leq 113$ are obtained. Theorem 4 would, of course, yield $N \leq 134$.

Consider now a number of special cases. If $m = 2\lambda$, then $F(x) = 1 + x$, which is irreducible. Consequently, from Theorem 3, $N = 1$. For $m = 3\lambda$, the sum of the degrees of all irreducible polynomials of degree less than $m/3$ and a divisor of m is at most

$$\sum_{\substack{k=1 \\ k|m}}^{m/4} kI_k \leq \sum_{k=1}^{m/4} kI_k \leq 2^{(m/4)+1} \quad .$$

On the other hand, the degree of $F(x)$ equals $2^{\lambda} + 1$. The fraction of the degree of $F(x)$ contributed by polynomials of degree less than $m/3$ is therefore bounded by

$$\frac{2^{(3\lambda/4)+1}}{2^{\lambda} + 1}$$

which approaches zero for large λ.

Consequently, for sufficiently large λ, the factors of $F(x)$ must be of degree at least $m/3$, and N is asymptotically bounded by

$$\frac{2^{\lambda}}{3\lambda} \leq N \leq \frac{2^{\lambda}}{\lambda} \quad .$$

By a similar reasoning, it may be shown that, for $m = 4\lambda$, N is asymptotically bounded by

$$\frac{2^{2\lambda}}{4\lambda} \leq N \leq \frac{2^{2\lambda}}{2\lambda} \quad .$$

Finally, for $m > 4\lambda$, the lower and upper bounds of Theorem 4 converge to the value

$$N = \frac{2^{m-2\lambda}}{m} \quad .$$

These conclusions are summarized in the following:

Theorem 5

For large λ, N is asymptotically bounded by

$$\frac{2^{m-2\lambda}}{m} \leq N \leq k\frac{2^{m-2\lambda}}{m}$$

where the value of k is as follows:

m	k
2λ	m
3λ	3
4λ	2
$>4\lambda$	1

Numerical results

Partitioning the Galois field $GF(2^m)$ into the set of equivalence classes implied by its selected subfield $GF(2^\lambda)$ produces a compact description of all possible alternant codes that may be generated for the given choices of m and λ. In many cases, the limited number of equivalence classes permits a complete characterization of the codes that are generated. Several of these cases are presented below.*

For the case of $m/\lambda = 2$, where there is only one equivalence class for each value of m, the nonzero values of the G matrix and H matrix weight spectra are given in Table 1 for $6 \leq m \leq 12$. For the cases of $m/\lambda = 3, 4, 5$, where there are many equivalence classes, representative sample spectra are given in Tables 2, 3, and 4, respectively.

It should be noted that, since any row of the H matrix may be replaced by a linear combination of that row with other rows, the minimum weight codeword given by the H matrix spectrum may be used to determine the best possible shortened code.†

Table 5 summarizes a computer search to determine whether any of the resulting codes were equal to or better than those previously published, as well as whether the resulting code performance exceeded either the minimum distance or minimum information bounds of Reference 2. For those codes that either exceeded the minimum construction bounds or were better than any previously known code, construction parameters are given in Tables 6 through 8.

---

* A complete summary of numerical results is presented in Reference 1.

† Codes are shortened by deleting the minimum number of columns required to allow the deletion of one row of the H matrix [3], [5].

TABLE 1. WEIGHT SPECTRA OF ALTERNANT CODES, $m/\lambda = 2$

Weight of Code Vector	Number of Code Vectors Having That Weight							
	Prime Code				Dual Code			
	$m = 6$	$m = 8$	$m = 10$	$m = 12$	$m = 6$	$m = 8$	$m = 10$	$m = 12$
0	1	1	1	1	1	1	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	7	0	0	0
3	0	0	0	0	18	0	0	0
4	0	0	0	0	15	0	0	0
5	2	24	232	2016	12	24	0	0
6	1	44	1044	19824	9	44	0	0
7	0	40	3200	150816	2	40	0	0
8	0	45	10000	1074564	0	45	0	0
9		40	27410	6724292		40	22	0
10		28	63043	36983606		28	105	0
11		24	126348	181562148		24	100	0
12		10	221109	801899487		10	95	0
13		0	338450	3207100848		0	180	0
14		0	459325	11683010232		0	85	0
15		0	553104	38945583600		0	80	0
16		0	587673	119270849775			150	0
17			552708	336759114600			70	0
18			460590	879315465900			65	0
19			338120	2128877967720			60	0
20			219778	4789975427370			11	0
21			127228	10036132584720			0	0
22			63614	19616077324680			0	0
23			26640	35820653775120			0	120
24			9990	61193616865830			0	234
25			3562	97909825691292			0	304
26			959	146864738536938			0	259
27			140	206698450488572			0	324
28			25	273137238145613			0	420
29			10	339067000563552			0	204
30			1	395578167324144			0	396
31			0	433859858044896				384
32			0	447417978608799				186
33				433859884789296				360
34				395578130249064				261
35				339066984516912				196
36				273137293083068				216
37				206698451825792				156
38				146864689455168				75
39				97909835209152				0
40				61193647005720				0
41				35820640796220				0
42				19616065197930				0
43				10036143286620				0
44				4789977477705				0
45				2128871501040				0
46				879316489560				0
47				336762120240				0
48				119266917585				0
49				38944498920				0
50				11683349676				0
51				3207401448				0
52				801850362				0
53				181500048				0
54				36972232				0
55				6733328				0
56				1082142				0
57				149988				0
58				18102				0
59				2052				0
60				171				0
61				0				0
62				0				
63				0				
64				0				

TABLE 2. WEIGHT SPECTRA OF ALTERNANT CODES, $m/\lambda = 3$

Weight of Code Vector	Number of Code Vectors Having That Weight					
	Prime Code			Dual Code		
	$m = 12$	$m = 15$	$m = 18$	$m = 12$	$m = 15$	$m = 18$
0	1	1	1	1	1	1
1	0	0	0	0	0	0
2	0	0	0	5	0	0
3	0	0	0	31	0	0
4	0	0	0	131	0	0
5	0	0	0	276	0	0
6	0	0	0	467	0	0
7	4	128	2664	731	11	0
8	5	400	18828	815	75	0
9	3	800	103748	704	265	0
10	2	1903	572702	515	475	0
11	0	4072	2840094	261	1010	0
12	0	6876	12534666	109	1896	0
13	1	10360	50108013	44	2430	0
14	0	14420	182540826	5	3360	0
15	0	17448	608511057	1	4496	0
16	0	18381	1863665889	0	4575	0
17		17336	5261937489		4440	0
18		14330	13739087990		3830	0
19		10360	33263548033		2430	73
20		6860	74843684910		1560	270
21		4136	156814663947		1042	549
22		2068	306501312966		520	1281
23		760	559698222303		245	2295
24		250	956148972660		85	3699
25		136	1529839442787		15	6591
26		47	2294763930504		7	8352
27		0	3229665525677		0	10888
28		0	4267767537668		0	16149
29		0	5297920408962		0	18171
30		0	6180907976772		0	21849
31		0	6779059946346		0	27303
32		0	6990909720435		0	26577
33			6779062335594			25602
34			6180903844428			24792
35			5297919971994			19125
36			4267772629372			14472
37			3229664614902			12135
38			2294761057164			8019
39			1529839605318			5313
40			956148983772			4107
41			559699687242			2151
42			306502738062			1224
43			156812478552			786
44			74842344594			261
45			33265266753			81
46			13739798658			19
47			5261055669			9
48			1863417963			0
49			608819061			0
50			182604510			0
51			50037909			0
52			12516918			0
53			2847951			0
54			579534			0
55			104451			0
56			16740			0
57			2271			0
58			300			0
59			45			0
60			0			0
61			0			0
62			0			0
63			0			0
64			0			0

TABLE 3. WEIGHT SPECTRA OF ALTERNANT CODES, $m/\lambda = 4$

Weight of Code Vector	Number of Code Vectors Having That Weight			
	Prime Code, $m = 20$		Dual Code, $m = 20$	
	$\alpha = \xi^{21}$	$\alpha = \xi^{39}$	$\alpha = \xi^{21}$	$\alpha = \xi^{39}$
0	1	1	1	1
1	0	0	0	0
2	0	0	0	1
3	0	0	0	0
4	0	0	0	0
5	0	0	25	35
6	0	0	310	254
7	0	0	595	698
8	0	0	3260	2920
9	10	38	5705	6629
10	116	88	17280	15903
11	82	135	29250	31082
12	260	207	56987	54577
13	320	331	84690	86286
14	475	464	113540	114292
15	460	519	141915	138788
16	619	560	141705	146473
17	554	514	141690	136922
18	360	400	113080	116207
19	410	341	85060	84308
20	176	245	57398	55802
21	172	140	28917	31327
22	40	72	17142	15310
23	40	29	5825	7202
24	0	11	3290	2366
25	0	0	605	945
26	0	0	280	177
27	0	0	10	66
28	0	0	15	5
29	0	1	0	0
30	1	0	0	0
31	0	0	1	0
32	0	0	0	0

TABLE 4. WEIGHT SPECTRA OF ALTERNANT CODES, $m/\lambda = 5$

WEIGHT OF CODE VECTOR	NUMBER OF CODE VECTORS HAVING THAT WEIGHT PRIME CODE $m = 25$
0	1
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	15
13	0
14	25
15	0
16	45
17	0
18	28
19	0
20	9
21	0
22	3
23	0
24	2
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0

TABLE 5. SUMMARY OF CODE PERFORMANCE

m	Nominal Prime Code (n,k,d)	Best First-Order Shortened Prime Code (n,k,d)	Best Second-Order Shortened Prime Code (n,k)	Best Dual Code
		$m = 2\lambda$		
6	8,2,5	6,1,5	5,1,5	8,6,2
8	16,8,5	11,4,5	8,2,5	16,8,5
10	32,22,5	21,12,5	14,6,5	32,10,11
12	64,52,5	39,28,5	26,16,5	64,12,25
14	128,114,5	74,61,5	46,34,5	128,14,54
16	256,240,5	143,128,5	86,72,5	256,16,95
18	512,494,5	275,258,5	154,138,5	512,18,201
20	1024,1004,5	542,523,5	197,279,5	
		$m = 3\lambda$		
12	16,4,7	14,3,7	12,2,7	16,12,2
15	32,17,7	25,11,7	20,7,7	32,15,7
18	64,46,7	45,28,7	34,18,7	64,18,19
21	128,107,7	85,65,7	57,38,7	128,21,43
24	256,232,7	156,133,7	99,77,7	
		$m = 4\lambda$		
16	16,0,9[a]	16,1,9	9,1,9	16,5,1 to 9,8,2
20	32,12,9	30,11,9	26,8,9	32,20,5
24	64,40,9	51,28,9	41,19,9	6
28	128,100,9	93,66,9		
		$m = 5\lambda$		
25	32,7,11[b]	30,6,11		
35	128,93,11	98,64,11		
		$m = 6\lambda$		
30	32,2,13	32,4,13[c]		
42	128,86,13	102,61,13		
		$m = 7\lambda$		
49	128,79,15	106,58,17		
		$m = 10\lambda$		
60	64,4,21	64,5,21[d]		

a. $k = k_{min} + 1$ (16,1,9).
b. $d = (2m/\lambda + 1) + 1 = d_{min} + 1$ (32,7,12).
c. Dimensionality of row space of $H = m - 2$.
d. Dimensionality of row space of $H = m - 1$.

TABLE 6. CODES EXCEEDING THE MINIMUM DISTANCE BOUND
$(2n/m + 1)$, EXHAUSTIVE SEARCH $n \leq 25$

m	λ	EQUIVALENCE CLASS LEADERS (exponent of the primitive element ξ)
25	5	445, 2267, 2867, 7075, 7683

TABLE 7. CODES EXCEEDING THE MINIMUM INFORMATION BOUND
$(2^m - n)$, EXHAUSTIVE SEARCH $n \leq 25$, LIMITED
SAMPLING $n \leq 84$, $m \leq 7$

m	λ	EQUIVALENCE CLASS LEADERS (exponent of the primitive element ξ)	AMOUNT BY WHICH INFORMATION IS EXCEEDED, Δk
		Search Range: Odd Exponents up to 305	
30	5	61, 99, 165, 171	2
		15, 25, 49, 55, 59, 85, 89, 101, 119, 121 169, 183, 187, 191, 193, 207, 233, 247, 251, 263, 273, 283, 301, 303	1
		Search Range: Odd Exponents up to 51	
60	6	7, 17, 33	1

TABLE 8. CODES WHOSE PERFORMANCE IS EQUAL TO OR BETTER
THAN THOSE PREVIOUSLY PUBLISHED

m	λ	EXPONENTIAL OF PRIMITIVE ELEMENT $(\alpha = \xi^i)$	NOMINAL G MATRIX CODE (N, K, d)	FIRST-ORDER SHORTENED G MATRIX CODE (N, K, d)
14	7	1	128,114,5	74,61,5
21	7	7	128,107,7	85,65,7
28	7	1	128,100,9	93,66,9
35	7	27	128,93,11	98,64,11
42	7	5	128,86,13	102,61,13
49	7	13	128,79,15	106,58,17

Conclusions

This paper has developed a structure for alternant codes and categorized the resulting detailed code performance. Starting from the defining construction of Reference 2, a 1:1 correspondence exists between the elements of a finite extension field $GF(2^m)$ and the codes that they imply. Partitioning the field into equivalence classes leads to a complete description of the codes that may be generated for the given choice of extension field $GF(2^m)$ and subfield $GF(2^\lambda)$. All the elements within one equivalence class imply a set of equivalent codes, with identical length $(2^\lambda = n)$, information capacity $(2\lambda -$ the dimension of the row space of $H \geq 2^\lambda - m)$, and code weight spectrum. For those equivalence classes whose elements $\alpha_r \in GF(2^m)$ are not contained in any subfield $GF(2^t) \subset GF(2^\lambda)$, the implied code will satisfy the minimum distance bound $(d_{min} \geq 2m/\lambda + 1)$.

Theorems have been presented which show that, for an equivalence class consisting of the set of elements

$$\left\{ \left(\frac{1}{\alpha_r x_i + y_j} \right)^{2^k} \right\}$$

where

$$\alpha_r \begin{cases} \in GF(2^m) \\ \notin \text{ any } GF(2^t) \subset GF(2^\lambda) \end{cases}$$

$$y_j \in GF(2^\lambda)$$

$$x_i \in GF(2^\lambda) - 0$$

$$0 \leq k \leq m - 1 \quad.$$

The equivalence classes are disjoint, and for a given k, the elements $\{(\alpha_r x + y)^{2^k}\}$ are distinct. Further, the transform $\text{Inv}(\alpha_r) \triangleq (\alpha_r^{2^\lambda} + \alpha_r)^{2^\lambda - 1}$ maps all the elements of a given equivalence class into a unique value $[\in GF(2^m)]$ and its conjugates; therefore, the equivalence class may be characterized either by its invariant, or by the minimum function of its invariant.

The number of repetitions of each element in a particular equivalence class is the same for every element in the class, and further is 1:1 with the number of repetitions in the conjugates of the invariants. Following from this, upper and lower bounds have been developed for the number of equivalence classes for a particular combination of $GF(2^m)$ and $GF(2^\lambda)$. The number of equivalence classes is of exponential order $[\mathcal{O}(2^{m-2\lambda})]$, and, in most cases of interest,

tends toward the lower bound $(\simeq 2^{m-2\lambda-\log m})$. In particular, for $m/\lambda = 2, 3$, it can be shown [1] that

$$\underline{\lambda = 2} \qquad\qquad \underline{\lambda = 3}$$

$T(\alpha) \; \varepsilon \; GF(2^m) \qquad\qquad T(\alpha_r) \; \varepsilon \;$ equivalence class of α_r

$Inv(\alpha) = 1$

where

$$T(\alpha) \overset{\triangle}{=} (\alpha^n + \alpha)$$

$$Inv(\alpha) \overset{\triangle}{=} (\alpha^n + \alpha)^{n-1} \qquad .$$

With the equivalence classes limiting the number of cases to be tried, summary performance (length, number of information bits, and minimum distance) has been given for the best (shortened H matrix) codes found with an exhaustive search of all cases $n \leq 25$.

The computational routines were also used to sample the performance of codes generated from the subfield $GF(2^7)$, which were found to be better than any previously known. In addition, cases were found in $GF(2^{25})/GF(2^5)$, $GF(2^{30})/GF(2^5)$, and $GF(2^{60})/GF(2^6)$ that exceeded the minimum distance and information bounds.

References

[1] A. Berman, "Structural Properties of Alternant Codes," Doctoral Dissertation, George Washington University, Department of Electrical Engineering and Computer Science, January 1982.

[2] H. J. Helgert, "Alternant Codes," *Information and Control,* Vol. 26, No. 4, December 1974, pp. 369–380.

[3] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes,* Amsterdam: North Holland Publishing Co., 1978, p. 592.

[4] E. R. Berlekamp, *Algebraic Coding Theory,* New York: McGraw-Hill, 1968.

[5] H. J. Helgert and R. D. Stinaff, "Minimum-Distance Codes for Binary Linear Codes," *IEEE Transactions on Information Theory,* Vol. IT-19, No. 3, 1973, pp. 344–356.

Appendix. Review of defining code construction for alternant codes

To provide a context for the discussion of the structure of alternant codes, this Appendix reviews the defining constructions of alternant codes, as described in References A-1 and A-2. In these papers, Helgert gives a generalized formulation for the H matrix of a code in terms of elements and the functions of elements from a Galois field $GF(q^n)$. The form of the matrix leads to a Vandermonde determinant and, therefore, a bound on the minimum distance of the code. Successive specializations of the initial formulation produce redundancies in the rows of H, and elimination of these redundant rows leads to the final form of the alternant code used for this study.

General formulation and derivation of minimum distance

With the elements

$\{y_i|i = 1, 2, \ldots, n\}$ elements of $GF(q^n) - 0$ (not necessarily distinct)

$\{x_i|i = 1, 2, \ldots, n\}$ distinct elements of $GF(q^n)$

and the polynomials $\{g_j(x)|j = 1, 2, \ldots, t\}$ of the form

$$g_j(x) = c_{0,j} + c_{1,j}x + c_{2,j}x^2 + \cdots + c_{t-1,j}x^{t-1}$$

with coefficients from $GF(q^n)$, an (N, K) linear error-correcting code is represented by its parity check matrix, H, having elements from $GF(q^n)$ in the form:

$$H = \begin{bmatrix} y_1 g_1(x_1) & y_2 g_1(x_2) & \cdots & y_N g_1(x_N) \\ y_1 g_2(x_1) & y_2 g_2(x_2) & \cdots & y_N g_2(x_N) \\ \vdots & \vdots & & \vdots \\ y_1 g_t(x_1) & y_2 g_t(x_2) & \cdots & y_N g_t(x_N) \end{bmatrix}_{t \times N} \quad \begin{array}{l} N, n, t, K, \\ \text{integers} > 0 \\ nt \geq N - K \end{array}$$

The conventional H matrix form over $GF(q)$ is obtained from this matrix by expanding each of the elements from $GF(q^n)$ as a column vector (n-tuple) with elements from $GF(q)$. The row space V of this matrix is defined to be all linear combinations of the rows of H over $GF(q)$. This H matrix may be

represented as the product of three matrices

$$H = CXY$$

where

$$C = \begin{bmatrix} C_{01} & C_{11} & \cdots & C_{t-1,1} \\ C_{02} & C_{12} & \cdots & C_{t-1,2} \\ \vdots & \vdots & & \vdots \\ C_{0t} & C_{1t} & \cdots & C_{t-1,t} \end{bmatrix}_{t \times t}$$

$$X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_N \\ x_1^2 & x_2^2 & & x_N^2 \\ \vdots & \vdots & & \vdots \\ x_1^{t-1} & x_2^{t-1} & \cdots & x_N^{t-1} \end{bmatrix}_{t \times N}$$

$$y = \begin{bmatrix} y_1 & 0 & \cdots & 0 \\ 0 & y_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & y_N \end{bmatrix}_{N \times N} .$$

With Y a diagonal matrix and X having a Vandermonde form, if C is nonsingular, then any $t \times t$ determinant selected from H will be nonzero, thus establishing the minimum distance of the corresponding N, K code as

$$d \geq t + 1 .$$

Further, the requirement $(nt \geq N - K)$ ensures that the code will have no more than mt parity check bits over $GF(q)$.

First specialization

With

$t = rs$, where r, s integers > 0
$z_i \in GF(q^n) - 0$
$\{w_j | j = 1, 2, \ldots , s\}$ distinct elements of $GF(q^n)$
different from all the x_i

then let

$$y_i = \frac{z_i}{\displaystyle\prod_{j=1}^{s} (x_i - w_j)^r}$$

$$g_{(\ell-1)r+k}(x_i) = \frac{z_i}{y_i(x_i - w_\ell)^k}$$

where

$$i = 1, 2, \ldots , N$$
$$\ell = 1, 2, \ldots , s$$
$$k = 1, 2, \ldots , r .$$

This specialization transforms the H matrix to the form

$$H_{t \times N} = \begin{bmatrix} H_1 \\ \hline H_2 \\ \hline \vdots \\ \hline H_s \end{bmatrix} \begin{matrix} \updownarrow r \\ \updownarrow r \\ \\ \updownarrow r \end{matrix} \quad rs = t$$

where

$$H_\ell = \begin{bmatrix} \dfrac{z_1}{(x_1 - w_\ell)^1} & \dfrac{z_2}{(x_2 - w_\ell)^1} & \cdots & \dfrac{z_N}{(x_N - w_\ell)^1} \\ \dfrac{z_1}{(x_1 - w_\ell)^2} & \dfrac{z_2}{(x_2 - w_\ell)^2} & \cdots & \dfrac{z_N}{(x_N - w_\ell)^2} \\ \vdots & \vdots & & \vdots \\ \dfrac{z_1}{(x_1 - w_\ell)^r} & \dfrac{z_2}{(x_2 - w_\ell)^r} & \cdots & \dfrac{z_N}{(x_N - w_\ell)^r} \end{bmatrix} .$$

From References A-1 and A-2, C is nonsingular, and every $t \times t$ determinant of $H \neq 0$, thus maintaining the distance of the NjK code to be

$$d \geq rs + 1 .$$

Again, the number of parity check symbols is at most nrs over $GF(q)$. Also, since the s $\{w\}$'s are distinct and different from the distinct set of $\{x\}$'s, and there are at most q^n elements in the field, there are at most $(q^n - s)\{x\}$'s and $N \le (q^n - s)$.

Second specialization

Let $n = \lambda m$, where λ, m integers > 1. Then

$$GF(q^m) \subset GF(q^{\lambda m}) = GF(q^n) .$$

Further, set $s = \lambda$

$$N \le q^m$$

$$\{x_i | i = 1, 2, \ldots , N\} \, \varepsilon \, GF(q^m)$$

$$z_i \, \varepsilon \, GF(q^m) - 0 \qquad (i = 1, 2, \ldots , N)$$

$$w_j \stackrel{\triangle}{=} w^{q^{m(j-1)}} \qquad (j = 1, 2, \ldots , \lambda)$$

$$w \begin{cases} \varepsilon \, GF(q^n) \\ \not\varepsilon \text{ any } GF(q^t) \subset GF(q^n) \end{cases}$$

then the element in the ℓth row and i^{th} column of submatrix H_1 equals the $q^{m(j-1)}$ power of the element in the ℓth row and i^{th} column of submatrix H_j for

$$\ell = 1, 2, \ldots , r$$

$$j = 2, 3, \ldots , \lambda .$$

That is, for the ℓth row and i^{th} column of

$$H_1 = \frac{z_i}{(x_i - w_i)^\ell} = \frac{z_i}{(x_i - w)^\ell}$$

raising to the $q^{m(j-1)\text{th}}$ power gives

$$\left\{ \frac{z_i}{(x_i - w)^\ell} \right\}^{q^{m(j-1)}} = z_i \left\{ \frac{1}{(x_i - w)^{q^{m(j-1)}}} \right\}^\ell$$

$$= \frac{z_i}{(x_i - w_j)^\ell} = \ell^{\text{th}} \text{ row and } i^{\text{th}} \text{ column of } H_j$$

since a row of H over $GF(q^n)$ and the $(q^k)^{\text{th}}$ power of that row over $GF(q^n)$ contributes the same set of rows to the row space of H over $GF(q)$. These redundant rows may be deleted, giving over $GF(q^n)$

$$H = \begin{bmatrix} H_1 \\ \hline H_2 \\ \hline \vdots \\ \hline H_s \end{bmatrix} \Bigg\} rs \qquad \qquad \overset{\leftarrow N \rightarrow}{[H_1]} \quad \updownarrow r$$

Third specialization

Let

$$N = q^m$$

$$z_i = z \qquad \qquad i = 1, 2, \ldots , N$$

Further, set

$$r = pq \qquad \qquad (p = \text{integer} > 0)$$

For this case, the ℓq^{th} row of H is the qth power of the ℓth row of H, and is therefore redundant for $\ell = 1, 2, \ldots , p$. The resulting minimum distance is

$$d = rs + 1 = \lambda pq + 1 .$$

The final specialization of $p = 1$, $q = 2$ results in the form

$$H = \left[\frac{1}{w + x_1} \frac{1}{w + x_2} \cdots \frac{1}{w + x_N} \right]$$

having a length $N = 2^m$ and a minimum distance

$$d = 2\lambda + 1 .$$

The number of parity check symbols over $GF(2)$ is at most n. It may be verified by substitution that the elements of H are roots of the equation

$$x^{2^m} + \sigma x^{2^m - 1} + \sigma = 0$$

where

$$\sigma = \left(\frac{1}{w^{2^m} + w}\right) \quad .$$

References

[A-1] H. J. Helgert, "Alternant Codes," *Information and Control*, Vol. 26, No. 4, December 1974, pp. 369–380.

[A-2] H. J. Helgert, "Noncyclic Generalization on BCH and Srivastava Codes," *Information and Control*, Vol. 21, No. 3, October 1972, pp. 280–290.

Hermann J. Helgert received the B.S. degree in electrical engineering from the University of Buffalo, Buffalo, N.Y., in 1962, and the M.S. degree in electrical engineering and the Ph.D. degree in communications theory, both from the State University of New York, Buffalo, in 1964 and 1966, respectively. From 1963 to 1964 he was with Union Carbide Research Laboratory, Tonawanda, N.Y., working in superconductivity. From 1964 to 1966 he was a Research Fellow at the State University of New York at Buffalo. From 1966 to 1969 he was with Cornell Aeronautical Laboratories, Buffalo, working on the statistical analysis of weapons systems. Currently he is Professor of Electrical Engineering and Computer Science at the George Washington University. Dr. Helgert is a member of the IEEE.

Arnold L. Berman received a B.S.E.E. from the Massachusetts Institute of Technology in 1957; and an M.S.E.E. and D.Sc. from George Washington University in 1977 and 1982, respectively. Since joining COMSAT in 1964, he has been responsible for the design and development of microwave communications circuits; the analytic modeling of these circuits to predict the linear and nonlinear distortion effects upon communications signals; the measurements of in-orbit communications performance; and the supervision of research and development in the areas of satellite-borne antennas and transponders, earth station receiving and transmitting equipment, microwave integrated circuit design, and monolithic circuit development. He is currently Assistant Director of COMSAT Laboratories.

Nancy G. Berman received a Bachelor of Science degree in mathematics from the Massachusetts Institute of Technology, and a Ph.D. in mathematical statistics from American University. She is currently employed as a Project Manager in the Statistical Services Division of Group Operations, Incorporated, a Washington, D.C.-based consulting firm.

Transponder supply/demand analysis for the geostationary orbit

W. R. SCHNICKE, J. B. BINCKES, AND D. H. LEWIS

(Manuscript received May 23, 1984)

Abstract

Geostationary orbit transponder supply and demand for the U.S. domestic and international arcs are analyzed using an orbital capacity model. Capacities are computed on the basis of single-carrier access of a 36-MHz transponder, including link budget calculations, degradation by external carrier-to-interference ratio (c/i) resulting from neighboring satellites (uniformly spaced), and modulation performance [FM, companded FM (CFM), companded SSB (CSSB), or QPSK]. The resulting transponder capacity estimates are then used with a recent projection of demand (telephone, TV, and data) to determine the supply of transponders needed to meet the demand in each category.

The transponder supply is estimated from an orbit use model based on reasonable assumptions for the number of frequency reuses, the available visible orbital arc, and the satellite spacing.

The analysis reveals that, if available technology is employed to make efficient use of satellite systems, the supply of satellite transponders will be adequate to meet the demand through the year 2000 with the use of C-band and K-band (14/11-12 GHz) satellites only. These conclusions are based on conservative improvements in satellite technology with respect to the efficiencies gained by using digital modulation techniques and improved low-sidelobe antennas. Advanced digital modulation, low-rate encoding, and on-board processing techniques currently under development can potentially increase the capacity of satellites by a factor of 5 to 10 over that used in this study.

Introduction

The central issue to be considered at the 1985 World Administrative Radio Conference on the Use of the Geostationary Satellite Orbit (WARC-85) is the means for ensuring access to the geostationary orbit by all countries. Of particular concern to developing countries is the possibility that adequate orbital and spectrum resources may not be available in the future because of rapid expansion of satellite systems by developed countries.

This paper reviews the expected demand for communications satellite services through the year 2000. An orbit capacity model which takes into account the impact of new technology is used to assess the ability of the geostationary orbit to support this demand.

Under NASA sponsorship, several other studies have been made of U.S. domestic telecommunications demand for satellite services through the year 2000 and the likely levels of transponder capacity that would be available to meet that demand. The most recent of these studies, performed by the Western Union Telegraph Company [1], was an update of their 1979 study [2] which discussed the need for capacity in the 18/30-GHz band. These studies reflect the significant changes that have occurred in the satellite communications industry in the intervening years. A separate study of demand for 18/30-GHz capacity was conducted by ITT in 1981 [3].

The latest NASA study estimates the demand requirement for transponders by converting traffic forecasts from voice expressed in channels, data expressed in megabits per second, and video expressed in channels into equivalent 36-MHz transponders. These estimates are based on several assumptions regarding the percent of voice traffic carried by analog versus higher efficiency digital modulations. High and low estimates of demand are then compared to the transponder capacities expected to be available. Briefly, the estimates of transponder supply presented in this paper are comparable to those obtained by NASA; however, the NASA study continues to forecast demand requirements significantly greater than those assumed here.

Service demand

The first step in transponder supply and demand analysis is to define the service demand to be met; i.e., the expected requirement for telephone channels, TV channels, teleconferencing channels, and megabits of data capacity to be met by the Fixed Satellite Service (FSS). Table 1 presents the demand projections used in the analysis for the following:

 a. The U.S. domestic market, based on literature surveys, consolidation of projections from prior studies, and user surveys [4].

 b. The U.S. Atlantic Ocean Region (AOR) international market, estimated from the INTELSAT traffic data base [5] which is compiled annually at a meeting of U.S. International Service Carriers and overseas correspondents. With respect to the data base, 10 percent is assumed to be data traffic and an additional 10 percent is added to account for single-channel-per-carrier (SCPC) and companded FM small-capacity carriers.

TABLE 1. DEMAND PROJECTIONS

SERVICE	1985	1990	1995	2000
U.S. DOMESTIC TRAFFIC				
Voice (Channels)	197,000	490,000	790,000	1,270,000
Video (Channels)	183	368	492	628
Teleconferencing (Channels)	264	1,068	2,652	4,272
Data (Mbit/s)	2,664	8,100	14,256	19,080
U.S. AOR INTERNATIONAL TRAFFIC				
Voice (Channels)	60,000	114,000	209,000	360,000
Video (Channels)	15	25	36	65
Data (Mbit/s)	380	730	1,340	2,300

The table reflects a healthy growth in demand for satellite telecommunications services through the year 2000, although the growth rates decrease over time and do not approach the 45-percent per annum FSS growth enjoyed from 1975 to 1980, or the 25-percent growth rate from 1980 to 1982. The 14-percent annual growth rate for international voice traffic, the bulk of the total demand, is the rate INTELSAT has used in its system planning studies. There is some evidence that this rate will be lower in the future, perhaps closer to 10 percent.

Demand for teleconferencing in the U.S. domestic market was forecast assuming the availability of a TV codec operating at 3 Mbit/s. International demand was not projected due to a lack of data, although analysts speculate that occasional-use TV incorporates some teleconferencing applications. Further, in the international demand category, requirements for the newly initiated INTELSAT Business Services (IBS) have not been included. Although IBS are expected to grow rapidly, there is currently no basis for accurately projecting demand. However, the inclusion of preliminary estimates of future demand for IBSS would not have impacted the conclusions of this study.

The demand projections presented in Table 1 in terms of channels and megabits cannot by themselves be used to project consequent demand for

space segment capacity in terms of 36-MHz transponders. Rather, the "raw" channel and megabit demand estimates must be considered along with estimates of transponder capacity (channels or megabits per transponder) to express demand in units of transponders. Estimation of transponder capacity is based on numerous system parameters, including satellite and earth station e.i.r.p. and G/T, modulation type, and interference from other systems. Further, the estimation of transponder supply needed to project the ability of the geostationary orbit to meet transponder demand, which is a function of the number of satellites that can be packed into a given orbital arc, is based on these same system parameters.

New technology, such as improved earth station antennas, new modulation types, and launcher improvements that allow the use of more powerful satellites, will have a substantial effect on both the demand for transponders and the supply. Greater efficiency in transponder use lowers the number of transponders needed to meet a given demand, while improved earth station sidelobe patterns permit tighter packing of satellites, increasing transponder supply. Thus, the use of improved technologies has a dual impact on the supply/demand equation.

Orbit capacity model

The major system parameters which affect transponder supply/demand estimation, and which are responsive to the introduction of new technology, have been linked in a computerized orbit capacity model. Table 2 summarizes the influence of several major variables on orbit capacity. Although inhomogeneities between satellites and satellite antenna patterns are included in this table, they have not yet been included in the model.

General description

Numerous computer models of considerable complexity have been developed to design satellite network configurations which maximize the capacity of the geosynchronous orbit, including specific transponder frequency plans and satellite antenna coverages [6]–[8]. This paper does not address any specific system configuration, but rather the first-order impact of various technological advances on orbit capacity. The model which was developed is straightforward in design and is small enough to be easily run on a personal or business computer.

Figure 1 shows the flow of computations in the orbit capacity model. The segments of the model can be described as follows:

TABLE 2. SUMMARY OF FACTORS AFFECTING ORBITAL UTILIZATION

ITEM	PARAMETERS AFFECTED, RELATIONSHIP	IMPACT ON ORBITAL RESOURCES	
		SUPPLY OF TRANSPONDERS	CONSUMPTION OF TRANSPONDERS
Ground Antenna Size	The larger the size and higher the gain, the better the isolation or external C/I.	Larger antenna permits closer satellite spacing, thus more transponders per unit of orbital arc.	Larger antennas permit greater capacity per transponder; traffic forecast satisfied with fewer number of transponders.
Ground Antenna Sidelobe Pattern	Lower off-axis gain improves isolation between satellites.	Better isolation permits closer spacing of satellites, thus more transponders per arc.	For a given spacing, lower sidelobes allow higher C/I, higher capacity per transponder.
Modulation Bandwidth Efficiency (channels per MHz) or Circuit Multiplication	For a given C/N, higher efficiency modulation provides higher channel density.	Supply side unaffected.	Rate of consumption of transponders is inversely proportional to the realizable channel density.
Number of Frequency Reuses per Satellite	Affects reuse (internal) C/I, more reuses yield lower C/I.	Multiplies number of transponders available per slot.	The more reuses the lower the capacity per transponder; this slowly increases the rate of consumption.
Inhomogeneities Between Satellites: Power, Antenna Gains, etc.	Directly (dB for dB) reduces external C/I for a given spacing.	For a given C/I objective, larger spacing, thus fewer slots.	Rate of consumption unaffected for a given C/I.
Satellite Antenna Patterns	Rejection by spatially isolated beams or cross-polarization.	Better isolation permits closer spacing, increases number of slots.	Rate of consumption unaffected for a given C/I.
C/I Objective of Satellite Operator	Affects modulation performance; tolerable C/I level depends on type of transmission system.	The lower the C/I, the closer the satellites can be spaced, the greater the number of orbital slots.	Capacity per transponder drops as C/I is lowered; this slowly increases the rate of consumption.
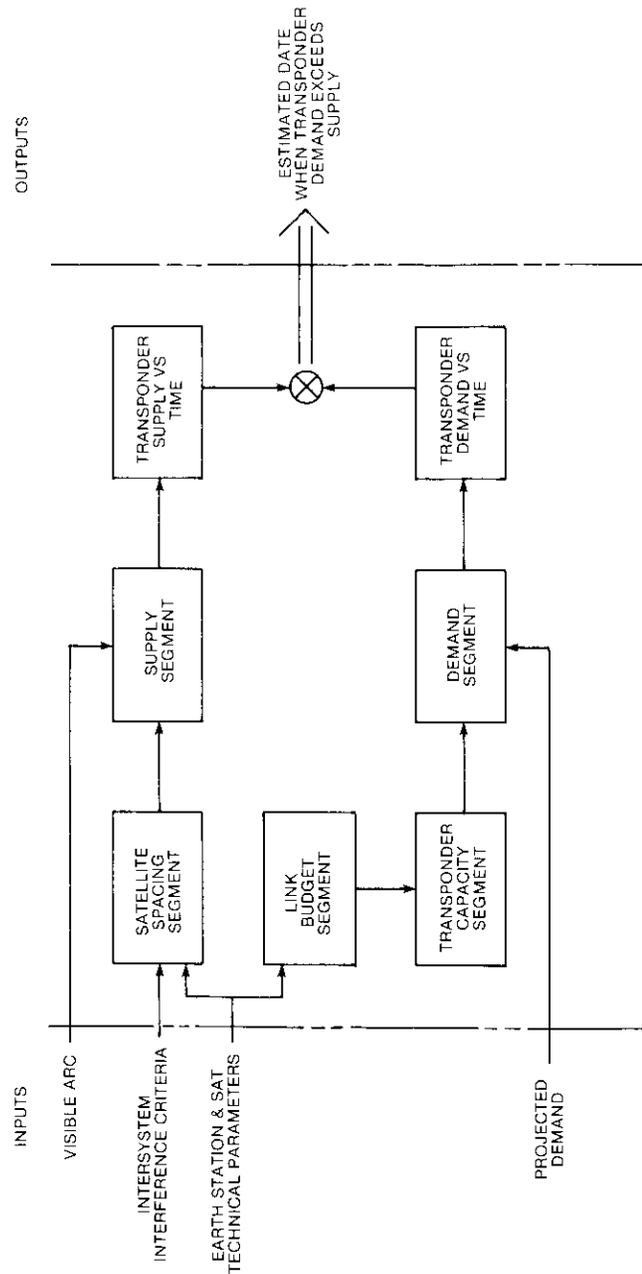
Figure 1. *Orbit Capacity Model*

a. Satellite spacing segment. The spacing between satellites in the system is computed by using inputs describing the earth stations and satellites in the homogeneous system (frequencies, antenna diameters, and sidelobe patterns), together with the C/I which can be accepted by the interfered-with satellite. The C/I arising from RF interference caused by one or more other satellites is termed external C/I.

b. Supply segment. The estimated satellite spacing and inputs describing the visible arc and the number of transponders per satellite are used to compute the total transponder supply. Repeating this computation for successive points in time gives the transponder supply time history.

c. Link budget segment. Input data on earth station and satellite power, receive sensitivity, transponder bandwidth, and satellite self-interference (frequency reuse) are used to compute the carrier-to-noise ratios (C/Ns) for the up- and down-links and for the overall transmission path, including self- and external interference. The C/I arising from RF interference on the same satellite from multiple reuse of the same frequency band is termed reuse C/I.

d. Transponder capacity segment. The basic link budget information obtained in item (c) is used, along with inputs describing the modulation techniques (minimum channel quality, companding gains, energy-per-bit for digital systems), to compute the number of channels that can be obtained from each transponder as a function of modulation type.

e. Demand segment. Dividing input demand projections for telephony, TV, and data channels by the estimated capacity in channels per transponder gives the number of transponders that will be needed to satisfy demand in each category at a particular point in time. Again, repeating this computation for successive points gives the transponder demand time history. The point in time at which supply and demand intersect gives the system saturation date; beyond this date, demand will exceed supply.

A single run of the model produces the outputs listed in Table 3. Most notably, transponder supply and demand correspond to input values of "raw" demand for each of the four service categories for a particular point in time. A time history (transponder supply and demand *vs* time) is developed by making a series of runs, each using an input demand component, for a succession of points in time. Use of the model in this manner does not capture the effects of the phased introduction of new technology; however, it does show how homogeneous systems such as all-FM or all-digital systems using either current or improved antennas would handle increasing demand over the entire study period. Studies are under way to examine the time-

phased introduction of earth station antennas with improved sidelobe patterns (as mandated by the FCC) by January 1, 1987, as well as the gradual increase in the use of digital modulation.

TABLE 3. ORBIT CAPACITY MODEL: OUTPUT

MODEL SEGMENT	OUTPUT
Satellite Spacing	Earth station transmit and receive gains
	Satellite spacing
Link Budget	C/N up and down
	C/N total, including C/I external and C/I reuse
Transponder Capacity	FM capacity
	CFM capacity
	CSSB capacity
	QPSK power- and bandwidth-limited capacities
Supply/Demand	Number of orbital slots
	Total transponders available (supply)
	Total transponders required to meet demand

Table 4 describes input parameters which are required to operate the model. As mentioned, once the service area and up- and down-link frequencies have been selected, only the demand estimates change from run to run.

TABLE 4. ORBIT CAPACITY MODEL: INPUT

MODEL SEGMENT	INPUT
Satellite Spacing	External C/I
	Earth station transmit frequency
	Earth station receive frequency
	Earth station antenna diameter
	Earth station antenna sidelobe pattern
Link Budget	Earth station e.i.r.p. and G/T
	Satellite e.i.r.p. and G/T
	Transponder bandwidth
	Reuse C/I
	Fade margins
Transponder Capacity	Companding gain, CSSB, and CFM
	Energy per bit per noise density, E_b/N_o, QPSK
	Bandwidth-symbol duration product, BT, QPSK
Supply/Demand	Visible arc
	Demand, voice channels
	Demand, video channels
	Demand, teleconferencing channels
	Demand, data (Mbit/s)

Technical assumptions

SATELLITE SPACING

Satellite spacing is calculated in the program as a function of the C/I objective, based on a homogeneous intersatellite interference model [9] in which isolation is calculated from earth station transmit/receive gains and the sidelobe roll-off pattern. For these calculations, it is assumed that the adjacent satellite systems and their earth station networks have identical performance characteristics and are spaced at uniform longitudinal intervals in the geostationary orbital arc. It is also assumed that the service coverage area is the same for each satellite. Thus, no isolation of radiated power by virtue of satellite antenna discrimination is present.

TRANSPONDER CAPACITIES

Transponder capacities are computed on the basis of single-carrier access of a 36-MHz transponder. The calculations are performed in several steps. First, the model completes a set of link budget calculations to determine the up- and down-link C/Ns and the C/N total, based on the system characteristics and free space losses (rain margin can be added) for the frequency band of interest. This total is then degraded by the C/I resulting from N neighboring satellites at a spacing of θ degrees, which is obtained from the intersatellite spacing calculation previously determined in the satellite spacing segment. Finally, the C/(N + I) is further degraded by a frequency reuse C/I, which is usually specified as 26 dB per reuse. In general, internal intermodulation noise is not considered because single-carrier-per-transponder access is assumed. However, in the case of CSSB, intermodulation noise was taken into account due to the nonconstant amplitude envelope that is characteristic of this transmission mode [10].

This final C/(N + I) can then be appropriately applied to particular modulation performance equations contained in the model for the following modulation types: FM, CFM, CSSB, and QPSK. For QPSK, a voice coding rate of 64 kbit/s was assumed, plus a voice-encoding [digital speech interpolation (DSI)] gain of 2. The resulting voice transponder capacity estimates are then divided into the raw demand projection specified for a particular year in order to determine the number of "configured" transponders needed to meet the demand for voice channels. For TV transmission, two FM channels per transponder were assumed, while for digital modulation a TV codec operating at 20 Mbit/s was assumed. This provided three channels per transponder for digital systems operating at the bandwidth limit of 60 Mbit/s per 36-MHz transponder. Analog transmission of teleconferencing video was at the rate of two channels per transponder. Digital transmission of teleconferencing

assumed a highly compressed transmission of 1.5 Mbit/s per video signal, yielding 20 channels per transponder when operating at the bandwidth limit.

TRANSPONDER SUPPLY

Satellite link parameters and spacing are used together with input data on satellite bandwidth, transponder bandwidth, and visible orbital arc to compute the transponder supply, which is the product of the number of slots and transponders per orbital slot. The orbital arc divided by the satellite spacing gives the number of slots; the number of transponders per slot is simply the number of transponders per satellite. A 10-percent reduction is then taken to account for incompatibilities between certain signal types (e.g., TV interfering with SCPC/PSK) to obtain the net supply estimate used in supply/demand analysis. Further, no allowance is made for in-orbit sparing, which is consistent with U.S. domestic practice but is not typical of international systems.

The values used for the model input parameters are shown in Table 5 for the U.S. domestic system, and in Table 6 for the international system.* Common values were used for the domestic and international cases whenever possible to facilitate comparison of the two systems.

Results for U.S. domestic service

U.S. domestic transponder demand

For a given system frequency and thermal C/N, transponder voice channel capacity, and consequently transponder demand, is primarily a function of the modulation type and external C/I selected by the system operator, as shown in Figure 2. Conventional FM peaks at about 1,600 channels per transponder and is relatively insensitive to external C/I over the range of values investigated. Companded FM is only slightly more sensitive to external C/I, and peaks at a capacity of 2,800 voice channels per transponder, considerably higher than FM.

Digital modulation operates in the bandwidth-limited mode due to the assumed QPSK bandwidth-symbol duration product, BT, of 1.2 (typical of INTELSAT modems). This gives 60 Mbit/s in a 36-MHz transponder, or 1,875 voice channels per transponder with a DSI gain of 2. Use of 32-kbit/s voice encoding doubles this to 3,756 channels per transponder, about the same as

---

* Tables 5 and 6 assume 8- and 12-m antennas, while Kratochivil *et al.* [2] assume a preponderance of smaller diameter antennas (on the order of 5 m). The study presented here assumes a more efficient utilization of the orbital arc at the cost of larger antennas.

TABLE 5. SYSTEM PARAMETERS USED IN ORBIT UTILIZATION MODEL: U.S. DOMESTIC SATELLITES

PARAMETER	C-BAND	K-BAND
Visible Arc (deg)	65	56
C/I, External (dB)	24	24
Reuses	2	2
C/I, Reuse (dB)	26	26
Transmit Frequency (GHz)	5.925–6.425	14.0–14.5
Receive Frequency (GHz)	3.700–4.200	11.7–12.2
Earth Station		
Antenna Diameter (m)	12	8
G/T (dB/K)	32.4	34
e.i.r.p. (dBW)	81	73
Satellite		
Transponder Bandwidth (MHz)	36	36
G/T (dB/K)	−4	1
e.i.r.p. (dBW)	36	41
Companding Gain (net)		
CFM	9	9
CSSB	9	9
Average Speaker Level (dBm0)	−21	−21
E_b/N_o (dB)		
BER = 10⁻⁶	10.5	10.5
Implementation	0.8	0.8
	11.3	11.3
BT (cycles/symbol)	1.2	1.2

CSSB for an external C/I of 24 dB and a reuse C/I of 26 dB (typical of domestic systems).

Based on the data in Figure 2, use of CSSB at an external C/I of 28 dB (or greater) appears to be the most attractive option. It shows better than a 30-percent improvement over digital modulation with low-rate voice encoding (LRE), and a factor of 2 improvement over CFM. However, as will be seen, transponder supply for CSSB and the other modulation types drops off sharply as external C/I approaches 28 dB because of the greater intersatellite spacing required to maintain that degree of isolation. At an external C/I of 24 dB (the value typical of domestic systems and the one selected for use in these studies), CSSB has a transponder channel capacity which is about the same as that for digital modulation with voice encoding.

TABLE 6. SYSTEM PARAMETERS USED IN ORBIT UTILIZATION MODEL: INTERNATIONAL SATELLITES

PARAMETER	C-BAND		K-BAND	
Visible Arc (deg)	16		16	
C/I, External (dB)	24		24	
Reuses	2–4		2	
C/I, Reuse (dB)	27–22		27	
Transmit Frequency (GHz)	5.925– 6.425		14.0 –14.5	
Receive Frequency (GHz)	3.700– 4.200		10.95–11.2	
			11.45–11.7	
Earth Station				
Antenna Diameter (m)	12	40	8	17
G/T (dB/K)	32.4	40.7	34	39
e.i.r.p. (dBW)	81	90	73	84
Satellite				
Transponder Bandwidth (MHz)	36		36	
G/T (dB/K)	–10		3	
e.i.r.p. (dBW)	28		41	
Fade Margins = 99%				
Up (dB)	0		3.6	3.6
Down (dB)	0		6	11.3
Companding Gain (net)				
CFM	9		9	
CSSB	9		9	
Average Speaker Level (dBm0)	–15		–15	
E_b/N_o (dB)				
BER = 10^{-6}	10.5		10.5	
Implementation	0.8		0.8	
	11.3		11.3	
BT (cycles/symbol)	1.2		1.2	

As with external C/I, decreasing reuse C/I causes a sharp drop in CSSB transponder capacity. While this is not a factor in the typical twofold-reuse domestic system (C/I reuse of 26 dB), it will have a significant impact on four-to-sixfold-reuse international systems (C/I reuse of 20–22 dB; see Figure 2). For an international system operating at an external C/I of 24 dB, the transponder capacity for CSSB will be less than that for companded FM, and about the same as that for digital modulation with no LRE.
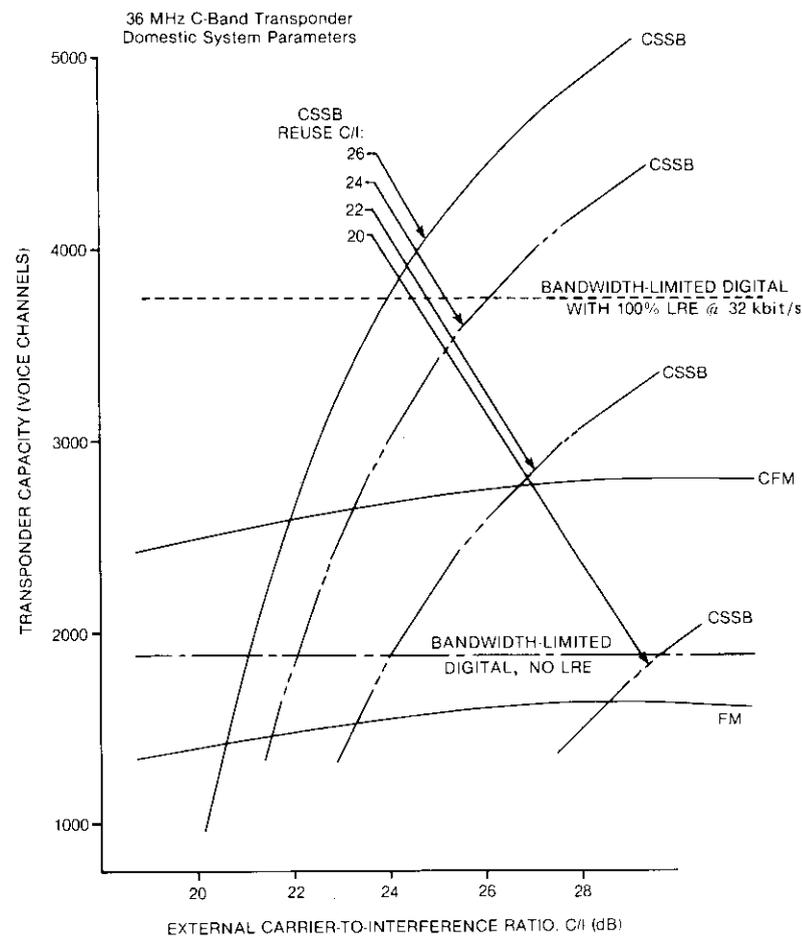
Figure 2. *Effect of Modulation Type and External C/I on Transponder Voice Channel Capacity*

Given the data of Figure 2 and the preference for a common set of assumptions for domestic and international case studies, FM and all-digital modulation and an external C/I of 24 dB were used in all subsequent analyses.

Figure 3 shows the transponder demand projected for the period 1985–2000, assuming the raw channel demand projected in Table 1 and the use of FM for voice, TV, and video conferencing and QPSK for data. Teleconferencing
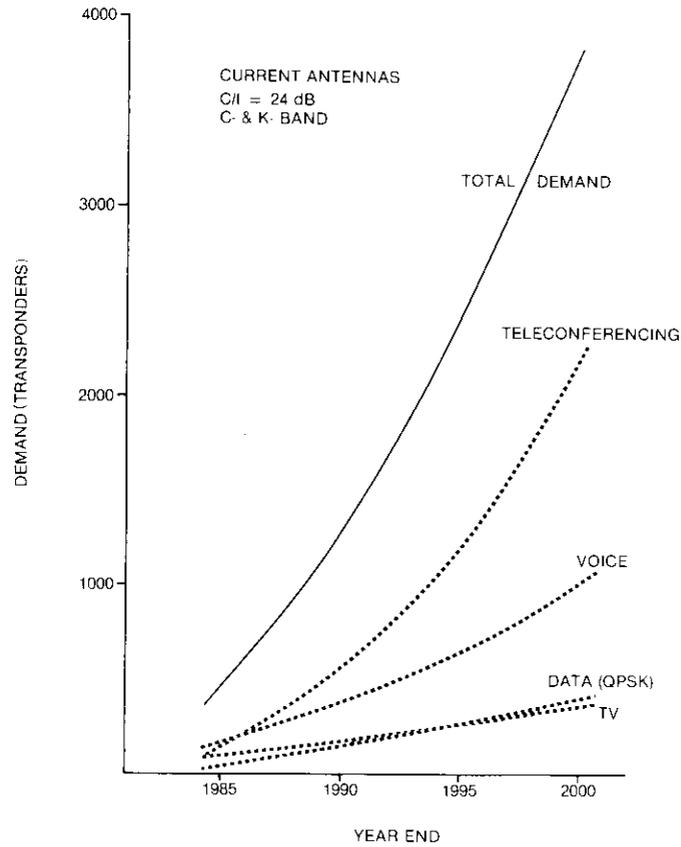
Figure 3. *U.S. Domestic Demand With Current Frequency Modulation*

accounts for about one half of the total demand throughout the period, with voice contributing about one quarter of the total.

In Figure 4, the same raw voice demand is assumed to be met by using digital modulation, including DSI for voice circuits (no LRE). The total transponder demand in year 2000 is less than half what it was with FM modulation, with the bulk of the reduction due to greatly improved teleconferencing transponder efficiency (twenty 1.5-Mbit/s channels as compared to only two FM channels per transponder). The reduced demand which would result if 100-percent 32-kbit/s voice encoding using DSI and LRE were employed is also shown. (If analog CSSB were used for voice, the demand would be equivalent to this latter case.)
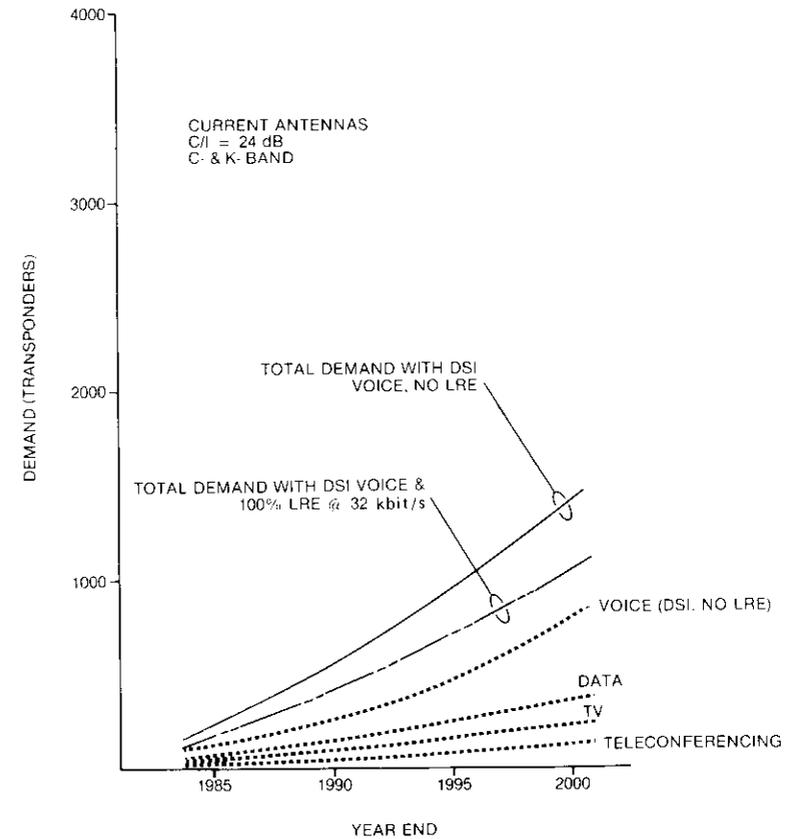
Figure 4. *U.S. Domestic Demand With All-Digital Modulation*

U.S. domestic transponder supply

The calculations required to estimate transponder supply are much more straightforward than those for transponder demand. The supply estimates from the orbital use model are time invariant and are determined solely by the number of transponders per (500-MHz) frequency band (6/4 GHz or 14/12 GHz), the number of frequency reuses, the available visible orbital arc, and the satellite spacing, which is in turn determined by the assumed C/I and antenna sidelobe performance. Simply stated, the number of orbital slots, determined from the visible arc and the satellite spacing, is multiplied by the number of transponders per satellite.

Assessment of transponder supply is facilitated by referring to the tabulation of orbit model results presented in Table 7. It can be seen that the total number of available orbit slots, and hence the total supply of transponders, is driven by the C/I objective and the sidelobe performance of the earth station antennas. This is equally true for both the C- and K-bands. For a given type of antenna, either "current" ($32 - 25 \log \theta$) or "improved" ($29 - 25 \log \theta$), a 4-dB lower C/I objective results in over 50-percent net increase in the supply of transponders because the lower C/I permits closer spacing of the adjacent satellite systems. For any given C/I objective, the 3-dB reduction in antenna sidelobes results in approximately a 40-percent increase in the pool of transponders available.

For purposes of this analysis, it was decided that three satellite slots (approximately 10 percent) would be kept vacant as a "guard band" orbit to allow for possible adjustments in some orbit locations due to inhomogeneities between satellite systems or types of satellite services.

Comparison of supply and demand for U.S. domestic transponders

The orbit capacity model provides estimates of the available orbital resources in terms of the number of orbit slots and the total number of transponders, as well as estimates of the "configured demand" transponder requirements which result from the application of selected satellite/earth station technologies to the raw demand forecasts. A 15° portion of the arc suitable for U.S. domestic service was set aside for Canadian requirements at C-band. Likewise, for K-band, a 10° subarc was reserved for Canadian systems. The remainder of the arc visible from CONUS—namely 65° of arc at C-band (5° minimum elevation) for U.S. domestic and 56° of arc at K-band (10° minimum elevation)—was assumed to be available for U.S. domestic services.

The impact of demand by countries south of the U.S. on U.S. domestic capacity will not be addressed here. Note that where geographical separation allows 30 dB of isolation (C/I) using spacecraft antennas, the same orbital arc can be used for both territories. Further, since the bulk of the South American land mass is significantly east of North America, an additional segment of the geostationary arc east of that suitable for CONUS service is available to meet South American demand.

Figure 5 and Table 8 show the results of application of the orbit utilization model to U.S. domestic satellite demand. The figure contains information on both transponder supply and demand. The horizontal straight lines show the transponder supply limitations corresponding to two reuses of the spectrum for each orbital slot. The curves depict the configured transponder demand associated with two different modulation types: current FM and all-digital. Current FM modulation represents FM modulation applied to the demand

TABLE 7. RANGE OF SUPPLY OF RESOURCES IN THE U.S. DOMESTIC GEOSTATIONARY ORBIT

C/I Objective (dB)	Total Visible Arc (deg)	Average Spacing (deg)	Available Slots	Usable Slots	Number of Reuses	Transponders Per Slot	Total Number of Transponders C-Band	Total Number of Transponders K-Band
CURRENT ANTENNAS								
C-Band								
24	65	2.52	25	22	2	24	528	
26	65	3.03	21	18	2	24	432	
28	65	3.65	17	14	2	24	336	
K-Band								
24	56	1.91	29	26	2	24		624
26	56	2.29	24	21	2	24		504
28	56	2.76	20	17	2	24		408
IMPROVED ANTENNAS								
C-Band								
24	65	1.91	33	30	2	24	720	
26	65	2.30	28	25	2	24	600	
28	65	2.77	23	20	2	24	480	
K-Band								
24	56	1.45	38	35	2	24		840
26	56	1.74	32	29	2	24		696
28	56	2.10	26	23	2	24		552
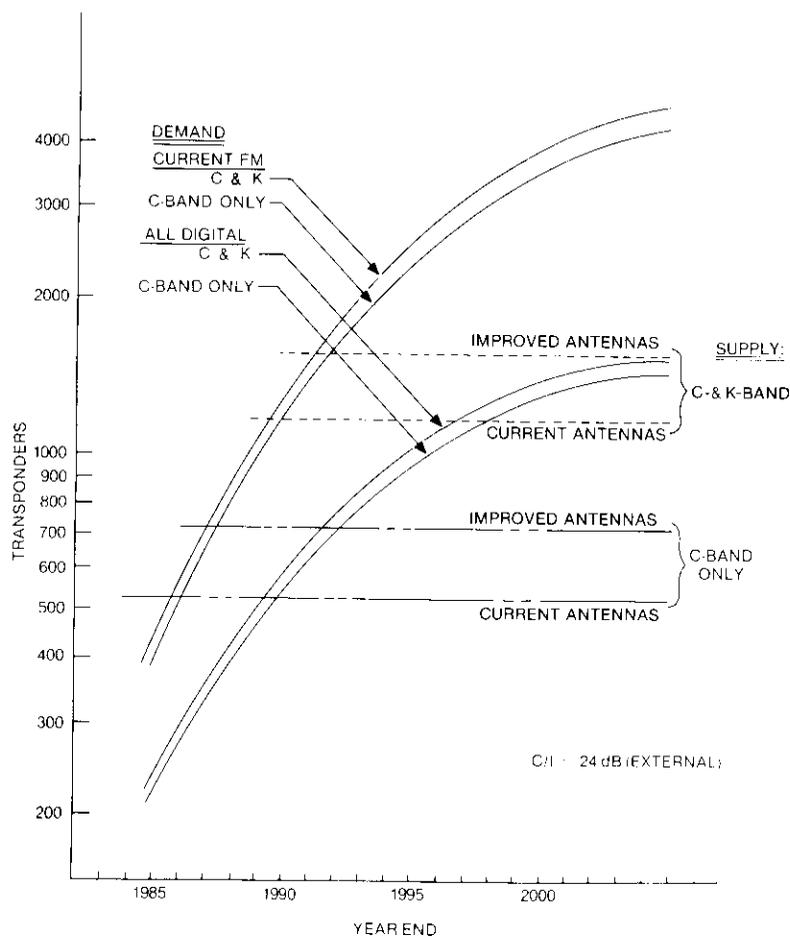
Figure 5. *U.S. Orbital Arc: Transponder Supply/Demand*

projections for voice and TV services (both TV distribution and video conferencing), with QPSK reserved for data services only. All-digital modulation refers to the use of QPSK modulation for all demand categories.

Similarly, transponder supply limits have been developed for two types of earth station antennas: current and improved sidelobes. Supply is shown for C-band only, as well as for C-band and K-band together. Likewise, configured demand is indicated for C-band only and for both bands together.

TABLE 8. CONFIGURED U.S. DOMESTIC SATELLITE DEMAND (36-MHz TRANSPONDERS)

YEAR	C-BAND TRANSPONDERS		K-BAND TRANSPONDERS		COMBINATION OF C- AND K-BAND TRANSPONDERS[a]	
	CURRENT FM[b]	ALL-DIGITAL	CURRENT FM[b]	ALL-DIGITAL	CURRENT FM[b]	ALL-DIGITAL
1985	394	216	469	253	435	236
1990	1,171	544	1,358	639	1,273	596
1995	2,321	888	2,613	1,013	2,480	956
2000	3,592	1,310	4,210	1,490	3,929	1,408

[a] Weighted average (45 percent C-band, 55 percent K-band) resulting from a greater supply of K-band transponders due to additional orbit slots with available closer satellite spacing at K-band.

[b] Current modulation is defined as FM applied to voice and TV services; QPSK for data. The table depicts the case for an external C/I objective of 24 dB.

Conclusions for U.S. domestic supply/demand

Several significant conclusions result from analysis of the data in Table 8 and Figure 5, assuming a C/I objective of 24 dB. With current antennas and current FM modulation techniques, configured demand would equal the capacity potentially available using C- and K-band transponders by 1989. However, with improved antennas (mandated by the FCC by January 1, 1987) and continued use of current modulation techniques, demand would not exceed available capacity until after 1991. The use of all-digital modulation techniques would allow the supply of transponders to keep pace with demand until 1997, even with continued use of current antennas. Furthermore, use of all-digital modulation techniques coupled with improved antennas would provide a supply of transponders that exceeds demand until beyond the year 2000.

Relationship of results to earlier studies

Earlier studies have addressed the question of supply versus demand for the U.S. domestic satellite arc. The most recent Western Union study indicates that the maximum potential supply of C- plus K-band transponders would vary from about 850 to 1,450 transponders in the year 2000, whereas the required demand would exceed 2,000 transponders and could even reach as high as 3,000 transponders.

In the present study, the supply limits for transponders were calculated to vary from about 1,000 to 1,600, depending on sidelobe performance of the earth stations which determines satellite spacing. This finding is consistent with the most recent Western Union study [1]. However, the present study found that the demand forecast for the year 2000 could be accommodated within these limits, provided that all-digital modulation was used.

While it is unclear why the demand forecasts differ, it is believed that differences are largely attributable to assumptions regarding transponder use efficiency. The more optimistic conclusion regarding the adequacy of transponder supply follows from the assumed use of advanced digital modulation techniques, which result in a lower effective demand for transponders.

The results regarding demand are similar to those of the Western Union study if the less efficient analog modulation (FM) is assumed. In this case, the present study also projects that demand will exceed supply, and that a net deficit of 1,000 to 2,000 transponders will result.

While it is too early to project with certainty, it is believed that economics will dictate a wholesale conversion to digital transmission by all satellite operators. This will translate into a supply of transponders adequate to meet

demand through the better part of this century. The only factor that could upset this conclusion would be the widespread use of earth station antennas with diameters significantly less than 8 to 12 m. Such use currently seems appropriate for special-purpose applications at C- and K-bands, whereas the 8- to 12-m antennas appear to meet the needs of more general FSS applications, including telephony.

Results for U.S. international service

U.S. Atlantic international transponder demand

Figure 6 shows the configured international transponder demand for the AOR, assuming FM for voice and TV, and QPSK for data. Voice traffic clearly dominates the total demand, in contrast to the domestic case where teleconferencing was the most important element. Moving to digital modulation for voice reduces the total demand by 40 percent. Use of 100-percent 32-kbit/s voice employing LRE and DSI results in a further 24-percent reduction (relative to the FM demand level), as shown in Figure 7. In this international case, CSSB is less efficient than in the domestic case because there are more band reuses on international satellites. Thus, use of CSSB in lieu of FM for voice produces results comparable to digital modulation without voice encoding, rather than digital modulation with voice encoding as was seen in the domestic case.

Interpretation of these configured international demand curves is further developed in Table 9. Figures 6 and 7 showed that configured demand for a given year is highest for the least efficient current FM modulation, and lowest for the more efficient all-digital modulation. However, in the international case, two dimensions of the configured demand side of the equation which are not as significant in the domestic arc have been added, namely, the effects of earth station antenna size and of various levels of frequency reuse. The use of larger antennas reduces the net demand requirements for transponders by about 20 percent, all other factors being equal.

Use of all-digital QPSK modulation further reduces the net/configured demand for transponders to 18 to 30 percent below that for current modulation, depending on the reference antenna size. With the satellite design parameters assumed here, both 12- and 30-m antennas are operating in a bandwidth-limited mode, as opposed to a power-limited mode. Thus, antenna size does not impact the transponder throughput bit rate, and consequently the digitally configured transponder demand is not sensitive to antenna size. It should be recalled, however, that antenna size does control the available supply of transponders.
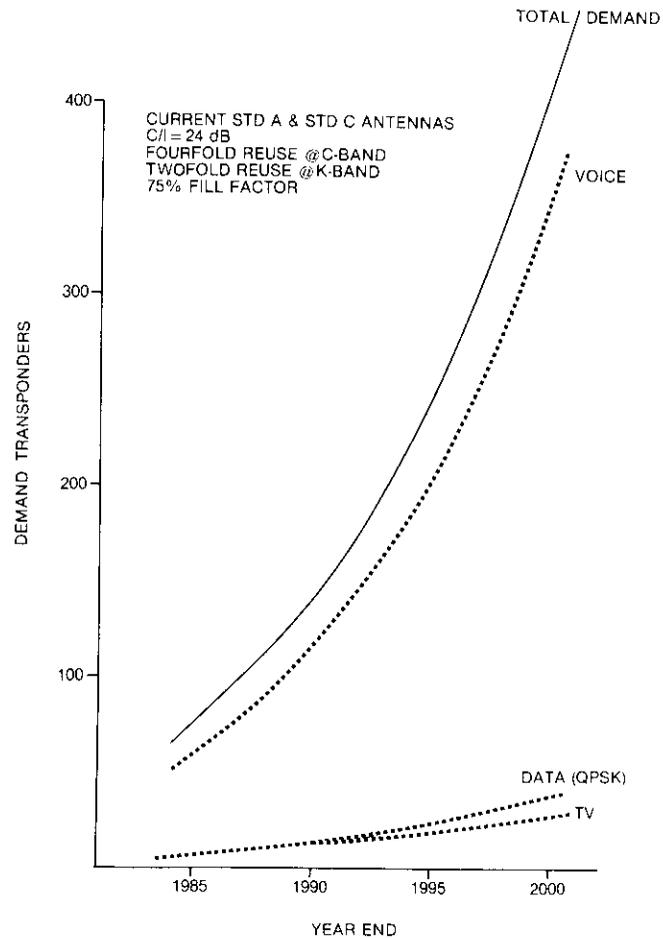
Figure 6. *Atlantic International Demand (INTELSAT System—AOR) With Current FM Modulation*

U.S. Atlantic international transponder supply

The international C-band transponder supply for the AOR is given in Table 10 to show the effects of antenna size and number of reuses. (The effects on K-band supply would be smaller.) As was the case for the domestic orbital arc (Table 7), it is apparent that the total number of available international orbital slots, and hence the total supply of transponders, is driven by antenna sidelobe performance. For this tabulation, two earth station antenna diameters
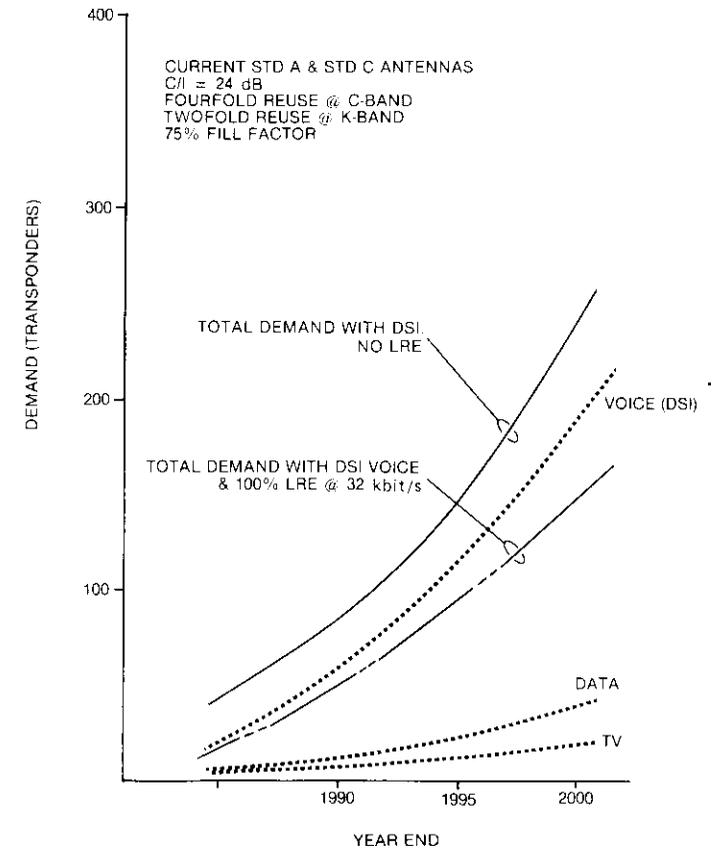
Figure 7. *Atlantic International Demand (INTELSAT System—AOR) With All-Digital Modulation*

are assumed (12 and 30 m), with each size having either current sidelobe levels $(32 - 25 \log \theta)$ or improved sidelobes $(29 - 25 \log \theta)$.

As was postulated for the domestic case, 10 percent of the satellite slots have been set aside as a guard band orbit to allow for differences between satellites or services. This accounts for the "usable" slots *versus* the "available" slots in the table.

Another refinement indicated in the table that was not applied to the domestic case is the use of a satellite fill factor (configuration efficiency) of 75 percent to account for the difference between the total available transponders and usable transponders. This was done in the international case because a

TABLE 9. CONFIGURED INTERNATIONAL DEMAND
(36-MHz TRANSPONDERS)*

| YEAR | ANTENNA SIZE (m)/REUSES | | | |
| | CURRENT FM | | | ALL-DIGITAL |
	12/2	30/2	30/4	12/2 or 30/4
1985	90	69	75	44
1990	168	129	130	83
1995	305	230	248	147
2000	514	389	423	249

* Analysis is presented for C-band only. For the case of all-digital modulation, the achievable bit rates projected by the orbit model appear insensitive to earth station antenna size or number of frequency reuses, because of the bandwidth-limited operation with such links.

certain percentage of transponders on a given satellite will always be underutilized due to the nonsymmetric geographic distribution of traffic, which never exactly matches the available capacity.

The effective range of supply of available C-band transponders for a 16° orbital arc spanning 325°E to 341°E, which is commensurate with the locations of international gateway-type stations, is summarized in Table 11. The variation shown results mainly from varying the number of reuses of the band between two and six times. Thus, depending on the ground station equipment, the assumed tolerable level of interference (isolation = 24 dB), and the number of reuses the international satellite can effectively utilize, a supply range of almost 10-to-1 is indicated for usable transponders in the international arc.

Comparison of U.S. supply and demand for international Atlantic services

The K-band frequency allocation is an orbital resource which should be included to provide a complete picture of international transponder supply and demand. To allow assessment of C- plus K-band transponder supply and demand while minimizing the number of variables, several representative pairs of earth station antenna sizes were assumed, each pair with one size at C-band and one size at K-band, along with associated reuse levels. It was also assumed that C-band is reused four times and K-band two times, levels which represent current INTELSAT plans for the 1990s. INTELSAT VI has six C-band reuses, but only two-thirds of the bandwidth is reused because of the need to reserve bandwidth for global coverage. Three earth station pairs were selected: large gateways, Standard A (C-band) plus Standard C (K-band); small gateways, Standard B (C-band) plus 8-m K-band; and urban collectors,

TABLE 10. RANGE OF SUPPLY OF RESOURCES[a] IN THE INTERNATIONAL (MID-ATLANTIC) ORBIT

ANTENNA SIZE (m)	TOTAL VISIBLE ARC (deg)	AVERAGE SPACING (deg)	AVAILABLE SLOTS	USABLE SLOTS	NUMBER OF REUSES	TRANSPONDERS PER REUSE[b]	AVAILABLE TRANSPONDERS	TOTAL NUMBER OF USABLE[c] TRANSPONDERS
CURRENT ANTENNAS								
12	16	2.52	6	5	2	12	120	90
12	16	2.52	6	5	3	12	180	135
12	16	2.52	6	5	4	12	240	180
12	16	2.52	6	5	5	12	300	225
12	16	2.52	6	5	6	12	360	270
30	16	1.21	13	11	2	12	264	198
30	16	1.21	13	11	3	12	396	297
30	16	1.21	13	11	4	12	528	396
30	16	1.21	13	11	5	12	660	495
30	16	1.21	13	11	6	12	792	594
IMPROVED ANTENNAS								
12	16	1.91	8	7	2	12	168	126
12	16	1.91	8	7	3	12	252	189
12	16	1.91	8	7	4	12	336	252
12	16	1.91	8	7	5	12	420	315
12	16	1.91	8	7	6	12	504	378
30	16	0.92	17	15	2	12	360	270
30	16	0.92	17	15	3	12	540	405
30	16	0.92	17	15	4	12	720	540
30	16	0.92	17	15	5	12	900	675
30	16	0.92	17	15	6	12	1,080	810

[a] C-Band transponders with c/t objective set at 24 dB.
[b] Assumes twelve 40-MHz transponders in 500-MHz band.
[c] Assumes 75-percent maximum utilization of transponders under INTELSAT spacecraft configuration.

TABLE 11. RANGE OF SUPPLY OF INTERNATIONAL C-BAND (36-MHz) TRANSPONDERS

ANTENNAS	CURRENT ANTENNA SIDELOBES	IMPROVED ANTENNA SIDELOBES
12-m	90–270	126–378
30-m	198–594	270–810

8-m C-band plus 5-m K-band. A 75-percent fill factor was used, but no allowance was made for providing a sparing capability in orbit.

Conclusions for international Atlantic services

Figure 8 shows the transponder supply and demand data for combined C- and K-band usage. With the large gateway stations (Standard A/Standard C), transponder supply is adequate to meet demand beyond the year 2000, even with current antenna sidelobes and FM modulation. With small gateway stations (Standard B/8-m K-band), the system will saturate in 1994 if current FM and current antenna sidelobes are used. Improved antenna sidelobes will delay saturation until 1997, while all-digital modulation will extend the system beyond 2000 with either antenna. The smallest antennas, the urban collectors (8-m C-band/5-m K-band), show both substantial increases in demand because of decreased transponder use efficiency, and decreases in supply due to the increased satellite spacing necessary to maintain the desired external C/I. The combination of these two effects is such that a system using these antennas with current FM and sidelobes would saturate in 1987; improving the antennas would extend the system to 1991. Unlike the larger antennas, the 8-m C-band/5-m K-band pair is power-limited at a digital throughput below the bandwidth limit. This results in an all-digital demand well above that for the other station pairs, as reflected in Figure 8. Consequently, all-digital transponder modulation does not extend system saturation for these small earth stations nearly as much as it did for the larger ones. Even with improved antennas, saturation occurs in 1994.

Summary of overall results

If efficient satellite communications systems employing digital techniques are utilized on a widespread basis, the supply of satellite transponders will meet demand through the year 2000 with the use of C- and K-band satellites. This is true even with the conservative technological improvements assumed in this analysis. Advanced digital modulation, LRE, and on-board processing techniques currently under development can potentially increase the capacity of satellites by a factor of 5 to 10 over the capacity assumed in this study.
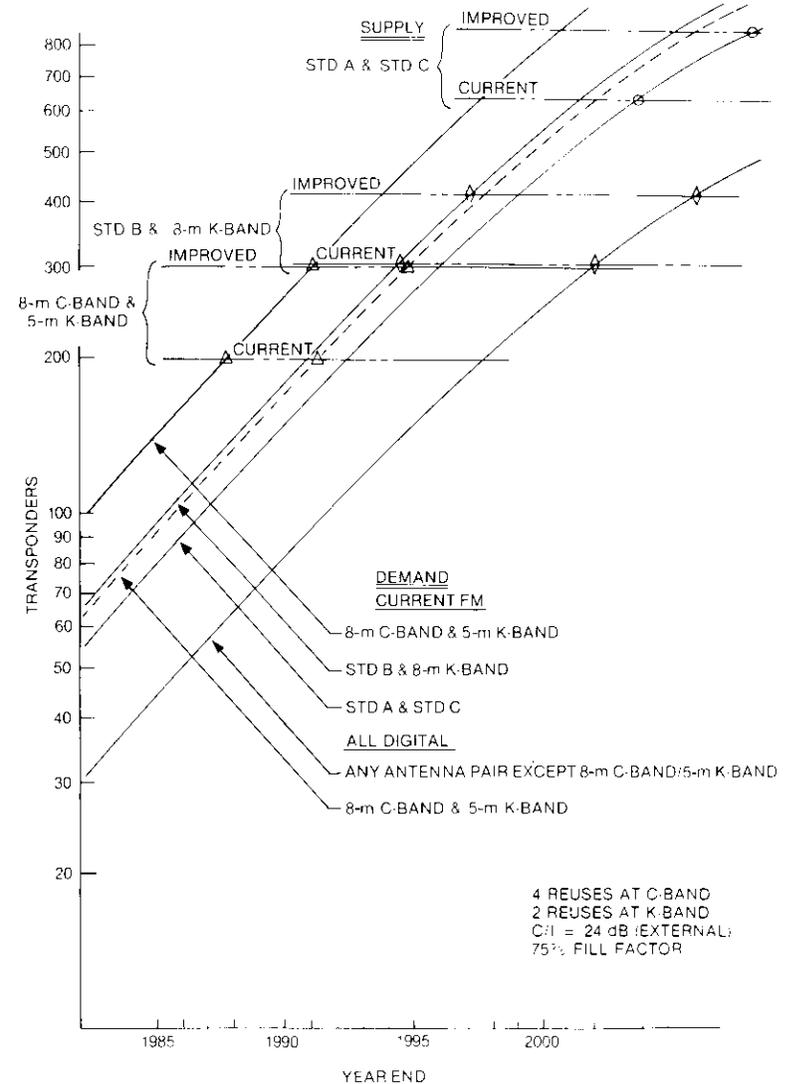
Figure 8. *Atlantic International Service Transponder Supply/Demand: Combined C- and K-Band*

With the use of advanced voice modulation techniques in domestic systems, either LRE (32 kbit/s) or CSSB could reduce effective demand for transponders by nearly 33 percent, with CFM providing a reduction of 20 percent, both

relative to continued use of techniques with efficiencies comparable to those achieved today. For the international service, increased self-interference caused by fourfold frequency reuse results in modest demand reductions for the analog techniques only when used with large gateway (Standard A) earth stations. For smaller antennas, CFM is slightly less efficient and CSSB is substantially less efficient than the baseline 64-kbit/s voice encoding. However, substantial gains can be realized through use of LRE in the international service. These findings are summarized in Table 12.

TABLE 12. EFFECT OF ADVANCED VOICE MODULATION TECHNIQUES ON TRANSPONDER DEMAND[a]

SERVICE	BASELINE 64-kbit/s[b]	LRE 32-kbit/s[c]	CSSB	CFM
Domestic (2 reuses)	1,311	942 (−28%)[d]	914 (−30%)	1,046 (−20%)
International (4 reuses)				
Standard A	249	153 (−39%)	213 (−14%)	237 (−5%)
Standard B	249	153 (−39%)	638 (+156%)	279 (+12%)
8 m	378	255 (−33%)	1,223 (+224%)	438 (+16%)

[a] Year 2000 demand assumed met using C-band only; transponder supply not affected.

[b] Voice transmission using modulation indicated; video, teleconferencing, and data remain QPSK (20 Mbit/s per video channel, 1.5 Mbit/s per teleconferencing channel).

[c] Includes DSI gain of 2.

[d] Percentage change in demand relative to baseline.

Partially counterbalancing this projection of capacity improvement are additional factors that may make it difficult to realize improvements. These factors include the high cost of incorporating technological improvements, particularly when existing equipment and facilities have not been amortized; the tendency of new services to impose greater restrictions on intersatellite interference, as is the case with small antenna applications; the apparent lack of homogeneity in the development of new satellite systems; and the reluctance of some developing nations to use improved technology.

References

[1] D. Kratochivil et al., "Satellite-Provided Fixed Communications Services: A Forecast of Potential Domestic Demand Through the Year 2000," Vol. 2, for NASA Contract NAS3-22894, NASA Lewis Research Center, NTIS CR168146, August 1983.

[2] T. Gabriszeski et al., "18/30-GHz Fixed Communications System Service Demand Assessment," Western Union, for NASA Contract NAS3-21359, NASA Lewis Research Center, NTIS CR159837, July 1979.

[3] R. B. Gamble, "Market Capture by 20/30-GHz Satellite Systems," ITT, for NASA Contract NAS3-21266, NASA Lewis Research Center, NTIS CR165231, April 1981.

[4] FCC Space WARC Advisory Committee, "First Advisory Committee Report for ITU WARC ORB 85," Docket File 80-741, December 1983.

[5] INTELSAT Board of Governors, "Traffic Data Base," Document BG-56-10, September 1983.

[6] T. Mizuno, Y. Ito, and T. Muratani, "Efficient Use of Geostationary Satellite Orbit Through Optimization of Satellite Locations," IEEE International Communications Conference 1979, Conference Record, Vol. 3, 1979, pp. 40.7.1–40.7.6.

[7] Y. Ito, T. Mizuno, and T. Muratani, "Effective Utilization of Geostationary Orbit Through Optimization," IEEE Transactions on Communications, Vol. COM-27, No. 10, October 1979, pp. 1551–1558.

[8] P. Sawitz and N. Shusterman, "Spectrum/Orbit Utilization Program User's Manual," Operations Research Inc., NASA Contract NAS5-21647, NASA Goddard Space Flight Center, May 16, 1984.

[9] CCIR Green Book, Volume IV, Part 1, Report 455, Geneva, 1982.

[10] K. Jonnalogadda, "Single Sideband, Amplitude Modulated, Satellite Voice Communication System Having 6000 Channels per Transponder," RCA Review, Vol. 43, September 1982, pp. 464–488.

William R. Schnicke received an S.B. from the Massachusetts Institute of Technology in 1966 and an M.S. from the University of Pennsylvania in 1969, both in electrical engineering. He joined COMSAT in 1974 as Senior Staff Scientist at Headquarters, where his principal activities involved INTELSAT system planning under the Management Services Contract. In 1978, he joined the U.S. Signatory organization, where he is currently Senior Director of INTELSAT Systems Planning and serves as U.S. delegate to the Advisory Committee on Planning and as Alternative Governor to the INTELSAT Board.
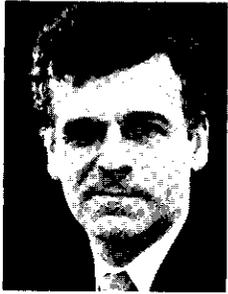
Jeffrey B. Binckes received a B.Sc. in physics from Miami University (Ohio) in 1963. He joined COMSAT in 1979, and is currently Manager of the Spacecraft Studies and Technology Department in the International Systems Planning Division. He has led studies which culminated in the development of several new spacecraft concepts for use by INTELSAT in the 1990s, such as an INTELSAT VI follow-on medium-size Z-satellite for carrying high-density trunk traffic. Previously, he participated in the INTELSAT VI U.S. Signatory proposal eval-uation team, and prior to that led a shuttle optimization study that recom-mended specification modifications for growth capabilities and a 10-year mission life on the INTELSAT VI program.

Mr. Binckes has been active in the CCIR Fixed-Satellite Service Group (SG-4) for many years, and has been a working group chairman of the FCC Space WARC Advisory Committee (SWAC) since 1981. At present, he is chairman of SWAC ORB-85 Proposal Evaluation Group.

Donald H. Lewis received B.S.E. degrees in Mathematics and Aeronautical Engineering from the University of Michigan, and an M.S. in Aero-nautical Engineering from The Massachusetts In-stitute of Technology. As an Engineer and Project Manager at TRW, he participated in and directed systems engineering and guidance and control tasks on the Lunar Orbiter and Apollo programs, includ-ing design and development of the control center space guidance sotfware used on the Apollo mis-sion. He joined COMSAT in 1981 after holding various positions with the U.S. government, where he was responsible for operations research studies of environmental and energy policy alternatives. He is currently Assistant Director of Systems Engineering for COMSAT Maritime Services.
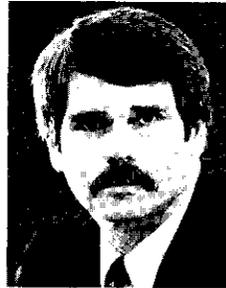
Measurement of data activity on satellite carrier VF channels

G. G. SZARVAS, S. AKERMAN, AND M. R. SHIU

(Manuscript received June 29, 1984)

Abstract

The amount of in-band data traffic that may be carried by a digital speech interpolation (DSI) system can influence the quality of the transmitted speech. Very little is known about the amount of in-band data carried in channels designated for speech. To collect such data, a traffic analyzer (TA) was developed to simultaneously measure the signal activity on 60 analog VF channels. Data signals of interest were differentiated from speech signals on the basis of signal continuity.

Measurements were made at the Andover (Maine) Earth Station on carriers containing a mixture of voice and data traffic from the United States to seven European destinations. The measurements showed that with the present traffic loading, for supergroups with 45 or more of the 60 channels designated as voice channels, the data load averaged over a 1-min. interval was equal to or less than 4.1 percent. A small digital speech interpolation system (60 input channels to 30 output channels) can operate with this data load and a 37-percent speech activity factor without loss of speech quality. For nonvoice traffic, the measured data load averaged over 1-min. intervals was 60 to 100 percent, demonstrating that this traffic should be excluded from DSI systems.

Introduction

Over the next several years, the INTELSAT FDM/FM analog transmission mode will gradually be augmented through the use of time-division multiple-access/digital speech interpolation (TDMA/DSI) transmission. Because of the greater activity of in-band data signals, the available DSI gain, g_{eff}, which is defined as the ratio of the number of terrestrial VF channels (served channels) to the number of satellite channels (serving channels), depends on the proportion of data loading in the VF channels. INTELSAT signatories have

been encouraged to measure data and speech statistics on channels designated for voice traffic to permit a proper accounting of the data traffic load when selecting the operating DSI gain.

This paper reports the results of data-loading measurements made at the Andover Earth Station in August and September 1983, on carriers containing a mixture of voice and data traffic from the United States to the United Kingdom, Italy, Spain, France, Germany, Switzerland, and Scandinavia.

Measuring technique

Traffic analyzer description and features

The TA is a flexible, unattended, remotely controlled data collection system specifically designed to detect nonspeech traffic that is carried in voice frequency channels. Figure 1 shows the functional block diagram of the TA (A), its remote control and information retrieval unit TA (B), and the ancillary equipment consisting of a baseband amplifier, baseband-to-supergroup de-
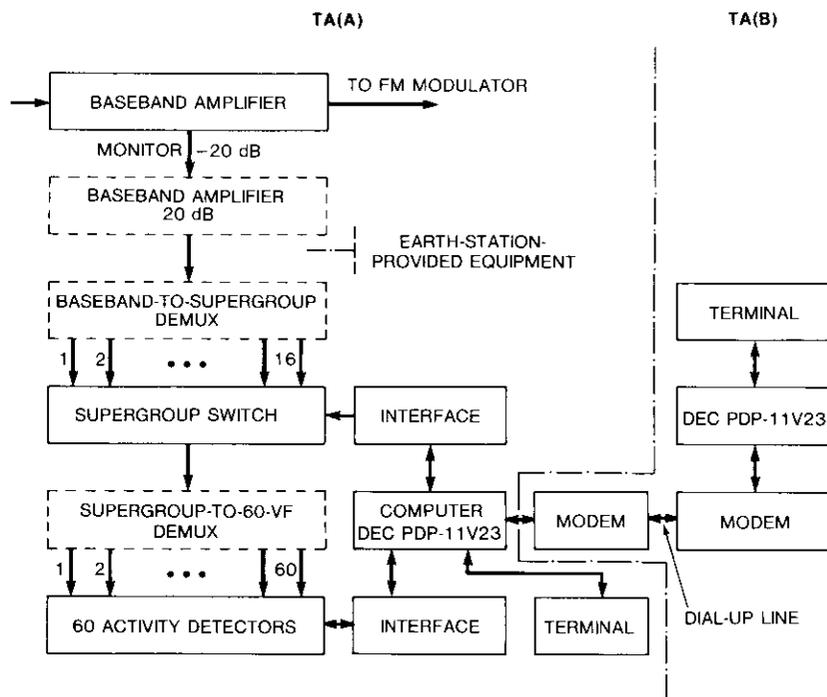
Figure 1. *Traffic Analyzer Functional Block Diagram*

multiplexer, and supergroup-to-VF channel demultiplexer. The TA (A) computer selects one of the supergroups in the baseband via the supergroup switch, which can connect a maximum of 16 supergroups. Activity in each of the 60 VF channels of the selected and demultiplexed supergroup is simultaneously detected by 60 threshold detectors. The detector outputs are sampled and their ON/OFF states are transferred to the computer of the TA (A) for further analysis.

The TA has two operating modes: a high-resolution mode in which a single supergroup containing 60 VF channels is monitored by 60 activity detectors, and their outputs are sampled 10 times per second; and a low-resolution mode, in which the supergroup switch cyclically selects four defined supergroups every second. In the low-resolution mode, the 60 activity detectors are time shared, and 240 VF channels are sampled once per second.

All input VF channels were assigned to one of three categories. Channels that were spare or otherwise out of service were omitted from the measurements and assigned to the "0" pool. Channels designated for voice service (usually DSI input channels) were assigned to the "*I*" pool. The remaining channels, carrying various classes of nonvoice traffic such as alternate voice data (AVD), voice frequency telegraphy (VFTG), simultaneous voice/data (data-over-voice), and facsimile are usually carried in digital noninterpolated (DNI) channels and were assigned to the "*N*" pool.

After 16 supergroups had been connected to the supergroup switch, the operation of the TA was remotely controlled through a dial-up telephone line. A test plan that defined operational parameters (supergroups to be monitored, test start time, and duration) was transmitted to the remote location via a 1,200-bit/s modem. After a test run was completed, the partially reduced data were transmitted to the remote control site for further data reduction, analysis, plotting and storage.

Activity detection

The primary purpose of the TA was to measure the incidence of in-band data traffic. Because the in-band data signals of interest are continuous and at a relatively high level (-13 dBm0), a simple threshold detector was used, consisting of a full-wave rectifier followed by an averaging circuit, a DC amplifier, and a threshold comparator. The threshold was set at -35 dBm0, well above the noise level. The activity measured with this simple threshold detector is signal-level dependent. The level dependency is negligible for the high-level, continuously present data signals, but not for speech signals that change dynamically.

Figure 2 shows the threshold detector organization and the timing diagram in the low-resolution mode. According to the multiplexer/demultiplexer

scheme, the 60 detectors are organized into five groups with 12 detectors in each group. The computer addresses the selected group through the 3 → 8 decoder. The ON/OFF status of the channels within a selected group of 12 VF channels is transferred through 12 bus lines to the computer for processing. The activity detectors are reset only in the low-resolution mode because they are time-division multiplexed over four supergroups. The timing diagram in
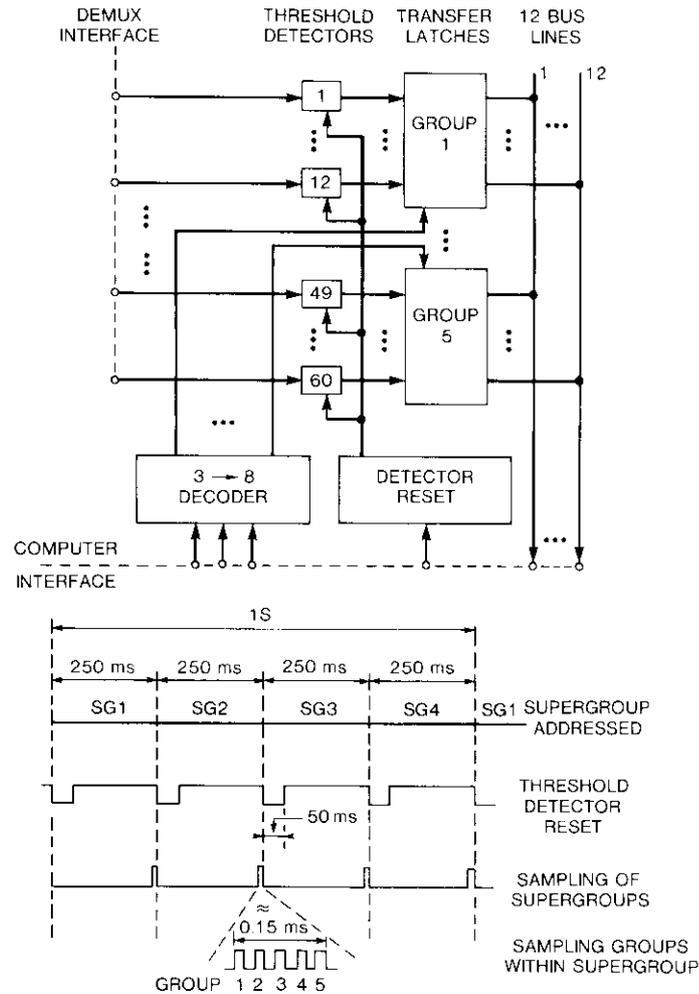
Figure 2. *Threshold Detector Organization and Timing Diagram in Low-Resolution Mode*

Figure 2 shows that in the low-resolution mode, activity detection time per channel is limited to 200 ms, with a sample rate of 1 sample per second per channel.

After the TA (A) installation, the FDM demodulation channel bank was aligned to a channel level accuracy of ±0.5 dB at the TA (A) input. The fixed threshold of the activity detectors was also calibrated to within ±0.5-dB accuracy. The calibration, which was rechecked during and after the measurement test period, was found to be stable within ±0.5-dB.

Discrimination of data and speech signals

VF data signals are continuously present, as opposed to the on-off structure of speech signals. Signal continuity for at least 10 s (100 consecutive "ones" in the high-resolution mode or 10 in the low-resolution mode) was selected as the criterion to declare the signal as data. Other signals were declared as speech.

During extensive laboratory testing, no speech signals were classified as data in the high-resolution mode. When the low-resolution mode was used with an abnormally high active speech level (exceeding −6 dBm0), a small number of speech passages were interpreted as data with call durations of less than 20 s. The low probability of occurrence of such a high speech level combined with the low number of false detections and their short duration produces a negligible measurement error.

CCITT No. 5 signaling sequences are generally interpreted as speech. There are exceptions, such as the busy-flash signal, if it is assumed that the calling party listens to the busy tone long enough.

Data reduction

Ensemble activity data for the I and N pools for each time interval ΔT of test runs were reduced to five numbers. They are shown for the I-pool in equations (1) through (5).

Average ensemble speech activity:

$$a_{Is} = \frac{\sum_{k=1}^{m} n_{Isk}}{mn_I} = \frac{S_{Is}}{mn_I} \tag{1}$$

Average ensemble data activity:

$$a_{Id} = \frac{\sum_{k=1}^{m} n_{Idk}}{mn_I} = \frac{S_{Id}}{mn_I} \tag{2}$$

Maximum ensemble speech activity:

$$\hat{a}_{Is} = \frac{\hat{n}_{Is}}{n_I} \qquad (3)$$

Maximum ensemble data activity:

$$\hat{a}_{Id} = \frac{\hat{n}_{Id}}{n_I} \qquad (4)$$

Maximum combined ensemble activity:

$$\hat{a}_{I(s+d)} = \frac{\hat{n}_{I(s+d)}}{n_I} \qquad (5)$$

where n_I is the number of channels in the I pool; n_{Isk} and n_{Idk} are the number of active channels with speech and data, respectively, at sample number k; maxima of n_{Isk} and n_{Idk} during ΔT are $\hat{n}_{Is}$ and $\hat{n}_{Id}$; $\hat{n}_{I(s+d)}$ is the maximum number of simultaneously active channels (speech plus data) during ΔT; $1 \le k \le m$ with $m = \Delta T x$ ($x =$ sampling rate). The time interval, ΔT, used for averaging was generally 1 min. Therefore, m, the number of samples for averaging, was 600 or 60 for the high- and low-resolution modes, respectively. The time interval, ΔT, was increased to 2 min. for overnight measurement runs and to 4 min. for weekend measurement runs.

By changing the subscript from I to N in equations (1) through (5), the set of five numbers describing the N-pool activity during ΔT is obtained. On-line processing by the TA (A) computer produces the numerators of equations (1) through (5). An example of this on-line processing is shown in Table 1, in which an I pool of only five channels is used. Channel activity is symbolized with vertical line segments, and the averaging is limited to include only 11 samples ($m = 11$).

Besides detecting the number of data calls, the traffic analyzer also monitors the time duration of each data call. From this information a data-call length distribution is obtained. Figure 3a is a composite plot of the observed data-call length distribution for over 2,000 data calls, and Figure 3b is a normalized data-call second distribution for each interval in Figure 3a. The normalized data-call second loading for each interval is given by

$$DCS_i = \frac{N_i Tm_i}{\sum_{j=1}^{16} N_j Tm_j} \times 100$$

TABLE 1. ON-LINE PROCESSING EXAMPLE

SAMPLE NUMBER m	ACTIVITY DETECTOR OUTPUT (CHANNELS 1–5) n_d 1 2 3 4 5	SAMPLED VALUE OF OUTPUT (CHANNELS 1–5) 1 2 3 4 5	ACTIVE CHANNELS $n_{I(s+d)k}$	ACTIVE SPEECH CHANNELS n_{Isk}	ACTIVE DATA CHANNELS n_{Idk}
11	— —┤ — —	1 0 0 1 0	2	1	1
10	— —┼ — —	1 0 1 0 0	2	1	1
9	┆- ┼ ┆—┼—	1 1 1 1 1	5	4	1
8	— — — —	1 0 0 0 0	1	0	1
7	—┆— + —	1 1 0 1 0	3	2	1
6	— — —	1 0 0 0 0	1	0	1
5	—┆— — —┆—	1 1 0 0 0	2	1	1
4	— + —	1 0 0 1 0	2	1	1
3	—┐ — — ┼	1 1 0 0 0	2	1	1
2	—└ —┤┼	1 1 0 1 0	3	2	1
1	— — — —	1 0 0 0 0	1	0	1

where i = interval number, 1, . . ., 16
Tm_i = midpoint of the interval duration (for $i = 16$, a value of $Tm_{16} = 3{,}000$ s was used)
n_i = number of data calls terminating within the interval i

Not unexpectedly, the sets of data for Figure 3a and 3b can reasonably be shown to be samples from simple exponential distributions given by the general expression

$$f(d) = ke^{-kd}$$

with $k = 0.06$ for the data in Figure 3a and $k = 0.01$ for the data in Figure 3b. Figures 3a and b show that a large number of short-duration data calls were observed, but these calls represent only a small portion of the total data load. One consequence of this is that the maximum observed ensemble data activity may be considered an extreme peak because several of the observed data calls are likely to be in the short-duration category, and thus present a lighter load to the DSI system.

(a) DATA-CALL DURATION

(b) NORMALIZED DATA-CALL SECOND
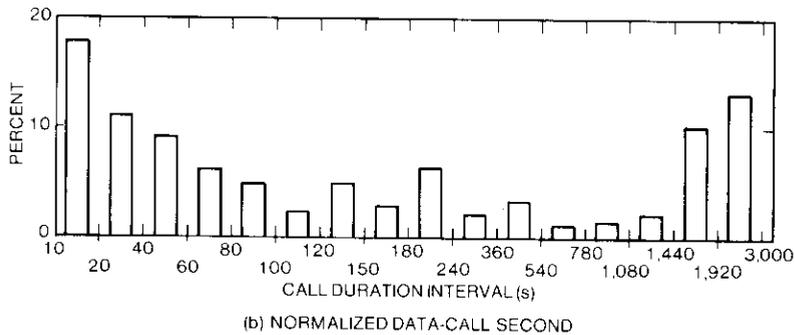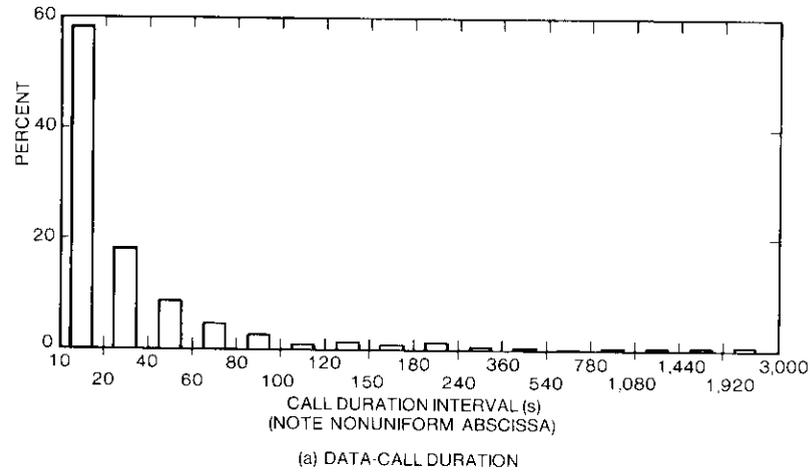
Figure 3. *Composite Data-Call Distributions*

Activity measurement

Test organization

The test setup was identical to the block diagram of Figure 1. Tests were performed on the VF channels of seven carriers destined for Italy, Spain, the United Kingdom, France, Germany Switzerland, and Scandinavia. Table 2 shows the carrier number, the number of channels, and their distribution by service categories. The 12 channels of group A, allocated in the 12- to 60-kHz baseband frequency range, were not monitored.

TABLE 2. MONITORED CARRIERS AND THEIR VF CHANNEL DISTRIBUTION

CARRIER NUMBER	DESTINATION	NUMBER OF CHANNELS[a]	SERVICE CATEGORIES[b] 1	2	3	4	5	6	7	8	POOLS I	N	O
4	UK	960	78.4	9.7	8.1		1.4	0.3		2.1	78.4	11.9	9.7
3	UK/Italy	600	73.3	21.3	2.8		0.5			2.0	73.3	5.4	21.3
5	UK/Spain	300	91.0	2.0	4.7		1.0			1.3	91.0	7.0	2.0
2	France	420	72.9	15.5	4.8		0.5			1.4	72.9	11.6	15.5
11	Germany/ Israel Abadan	600	69.0	22.7	5.7		1.2	0.2		1.3	69.0	8.3	22.7
1	Switzerland	240	80.4	14.2	1.7		2.1	0.4		1.3	80.4	5.4	14.2
12	Scandinavia	420	69.0	29.0	0.5		1.0			0.5	69.0	2.0	29.0
Total Channels		3,540	2,690	584	169		37	5		55	2,690	266	584
Percent		100	76.0	16.5	4.8		1.0	0.1		1.6	76.0	7.5	16.5

[a] A-group channels excluded.

[b] Service category codes:

1 = Voice	6 = Simultaneous voice/ data
2 = Spare	
3 = Alternate voice/data	7 = Facsimile
4 = Reserved	8 = 4-kHz data
5 = Voice frequency telegraphy	

Service category information was obtained from the operations data base and was entered in the TA computer in the form of service tables. A service table identifies each channel of a given carrier with its service code and place in the multiplexing hierarchy. The service tables were translated into classification tables by grouping each channel according to its service into the *I*, *N*, or *O* pool corresponding to the DSI and DNI pools of the future TDMA/DSI operation, or into the pool to be omitted (*O*). All voice channels were assigned to the *I* pool; the spares were omitted; and channels of the remaining service categories were lumped into the *N* pool (see Table 2).

Table 3 shows how data were collected over a representative set of carriers and supergroups. The high-resolution mode test runs were usually of 30-min. duration.

The ensemble average data and speech activities were determined over intervals of 1 min., and were also averaged over a 30-min. test run. Figure 4 is a plot of a typical speech test run for a supergroup with 50 of its 60 channels designated as voice channels. The average of the 1-min. ensemble speech activities for this 30-min. run was 36.6 percent, while the 1-min. ensemble activities ranged from 31 to 39 percent. Variation in the 1-min.

TABLE 3. TEST RUNS

CARRIER NUMBER	NUMBER OF SUPERGROUPS	NUMBER OF TEST RUNS	
		HIGH RESOLUTION	LOW RESOLUTION
4[a]	15	32	8
3	10	16	1
5	5	18	4
2	7	73[b]	6
11	10	21	12
1	4	10	3
12	7	7	3
Total	58	177	37

[a] One spare supergroup is deleted.

[b] The large number of test runs is due to the fact that runs performed in August, which is a vacation month in France, were repeated in September. No significant differences were observed between the two test run sets.
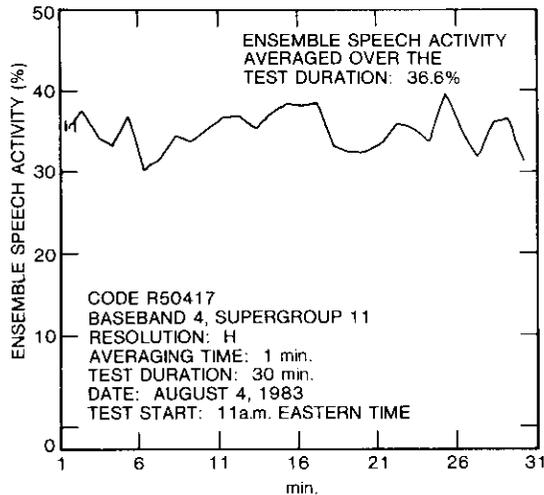
Figure 4. *Variation of the Normalized and Per-Minute Averaged Ensemble Speech Activity of 50 Voice Channels*

ensemble activities was less for data than for speech, and less for larger ensembles with high activity than for smaller ensembles with low activity. Figure 5 shows the activity plot of a typical overnight run.

A total of 34 high-resolution measurement runs of 30-min. duration were conducted, using 1-min. averaging intervals, for supergroups with at least
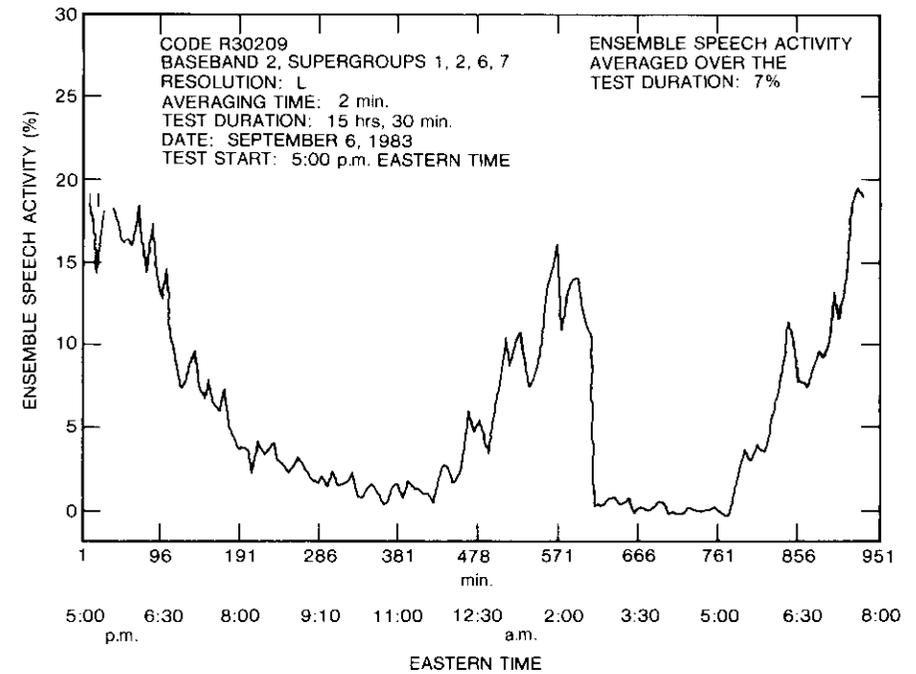
Figure 5. *Variation of the Normalized and Per Two-Minute Averaged Ensemble Speech Activity of 221 Voice Channels*

40 of the 60 channels designated for voice traffic. The largest observed 1-min. average ensemble data signal activity for each test run was selected; the results are plotted in Figure 6a. The maximum percent of channels carrying data for this set of observations was 4 percent, indicating that during any 1-min. interval, only 2 or 3 of the 60 channels in the supergroups continuously carried in-band data. The small table in Figure 6a shows the same data for measurements made on larger ensembles of channels. A maximum data signal loading of 4.1 percent was observed on a 126-channel ensemble, but only 5 or 6 channels continuously carried in-band data signals during the most heavily loaded 1-min. interval. The same measurement runs also provided similar information on the ensemble speech activity, which is shown in Figure 6b.

Measurements were also made on the data loading in channels designated for nonspeech service. The number of channels in the measurement ensemble were reduced because of the limited number of channels in these service categories. Figure 7a is a histogram of the largest observed 1-min. average

FOR LARGER CARRIERS	
NUMBER OF CHANNELS	MAXIMUM DATA ACTIVITY
227	1.2
193	1.3
224	2.6
126	4.1
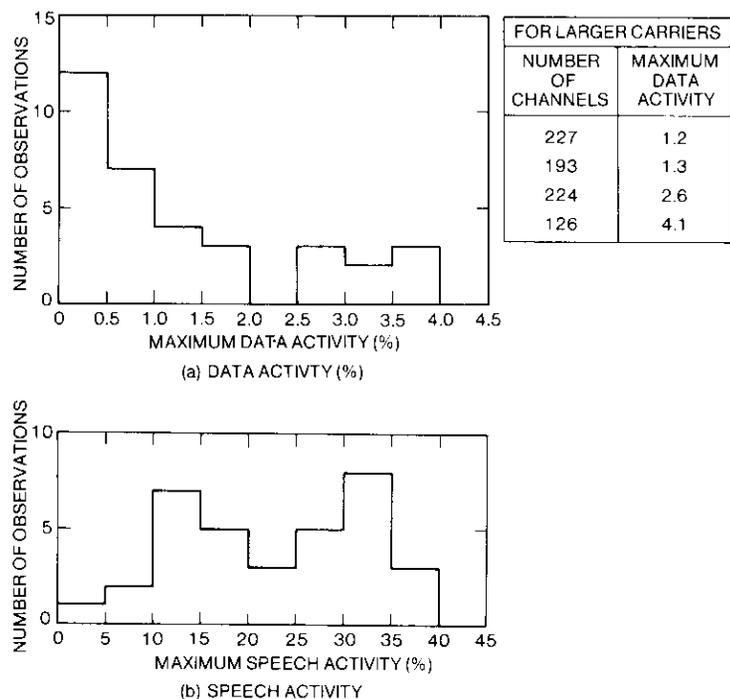
Figure 6. *Maximum Average Data and Speech Activity on Voice Channels Over Intervals* ΔT

ensemble data signal activity for supergroups in which at least 20 of the 60 channels were designated for non-voice service categories.

Similar information on ensemble data signal activity for supergroups carrying from 5 to 20 nonspeech designated channels is shown in Figure 7b. The results confirm the expectation that as the number of channels designated for nonspeech traffic decreases, the probability of all such channels being active increases.

During the initial phase of the measurement program, speech level measurements and high- and low-resolution mode activity measurements were also performed. Speech level measurements performed with a speech voltmeter [1] on 57 randomly selected VF channels of carrier 4 resulted in the average active speech level of -19 dBm0 with a 5-dB standard deviation. This is at least 8.3 dB lower than the level recommended by the CCITT for system-loading calculations. Although it was not part of the original measurement program, the lower speech power level on carrier 4, which was

(a) MORE THAN 20 NONVOICE CHANNELS

(b) 5—20 NONVOICE CHANNELS
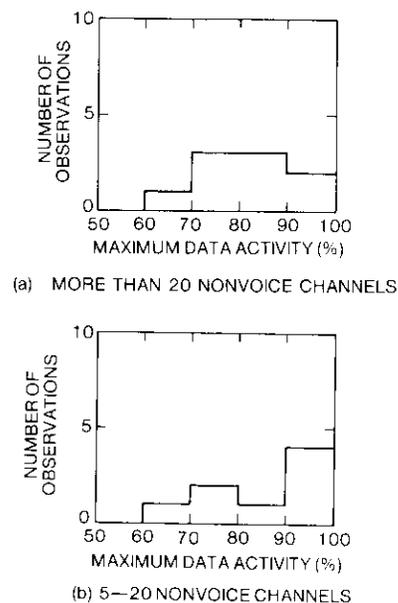
Figure 7. *Maximum Data Activity on Nonvoice Channels*

observed with a speech voltmeter, suggests that further measurements should be undertaken.

Consecutive test runs of 15-min. duration were repeatedly performed by monitoring the 60 VF channels of the same supergroup. These test runs, which were conducted first in low- and then in high-resolution modes, resulted in almost identical normalized ensemble speech activity for both modes when the mode calibration curves shown in Appendix A were used.

Discussion

The 266 channels in the nonvoice service category (see Table 2) constitute 9 percent of the total (2,690 + 266) channels in active service on the seven carriers that were monitored. The maximum data activity of these channels is between 60 and 100 percent with a weighted average of 82 percent, as Figure 7 shows. To avoid excessive clipping of the speech channels, it is necessary to route data channels to DNI slots when data loads exceed 8 percent.

By using the observed values of 37-percent speech activity, a worst-case DSI gain estimate for a DSI pool consisting of 240 and 120 voice channels is

as follows. Figure 8 [2] shows that available DSI gain as a function of the served channels, assuming speech activity only. The curves are calculated based on the creation of additional overload channels with bit reduction and on the permissible average quantizing noise increase of 1 dB, as suggested by Campanella [2]. For 37-percent speech activity, Figure 8 shows a gain, g, of 2.40, 2.26 and 2.13 for 240-, 120- and 60-channel systems, respectively.

In-band data has a 100-percent activity factor. If data is to be carried in a DSI system at the required level of DSI performance (*i.e.*, a DSI gain of 2.4 for a 240-channel system with 37-percent speech activity and no data channels), then additional satellite channels are needed to account for the higher data channel activity. This causes the effective DSI gain to be reduced. The effective DSI gain, g_{eff}, is given by

$$g_{eff} = \frac{g}{1 - a_d(g - 1)} \tag{6}$$

where g = DSI gain for speech and no data
  a_d = ratio of the number of data channels to the number of input terrestrial channels

By using this relationship and allowing for an 8-percent data load while maintaining the same performance as that of a 240-channel system with a DSI gain of 2.4 and a 37-percent speech activity factor, the effective DSI gain is calculated to be 2.16. Similarly, a 120-channel system with a DSI gain of 2.26 without data reduces to an effective DSI gain of 2.07 with an 8-percent data load. A 60-channel system with a DSI gain of 2.13 without data reduces to an effective DSI gain of 1.95 with an 8-percent data load, and to 2.05 with a 4-percent data load.
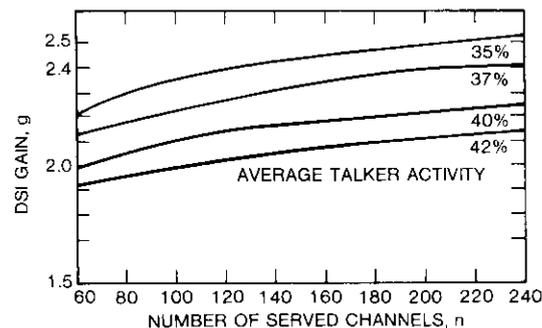
Figure 8. *DSI Gain, g, vs Number of Served Channels, n (Bit reduction with 1-dB permitted average quantizing noise increase)*

Summary and conclusions

The objective of the measurement program was to determine to what extent in-band data signals were being transmitted over channels that were designated for voice traffic. Data signal activity measurements were conducted on the voice frequency channels of seven carriers with a mixture of voice and in-band data traffic between the Andover Earth Station and several European countries.

The results of the measurements were as follows:

● For supergroups with 45 or more of the 60 channels designated as voice channels, the maximum observed data load, averaged over a 1-min. interval, was 4.0 percent.

● For the same channels in 76 percent of the test runs, the maximum observed data load, averaged over a 1-min. interval, was 2.0 percent or less.

● The maximum observed data load in designated voice channels, averaged over a 1-min. interval, was 4.1 percent; this load occurred on a 126-channel carrier.

● For larger carriers (193 to 227 designated voice channels), the maximum observed data load, averaged over a 1-min. interval, ranged from 1.2 to 2.6 percent.

● For channels designated as voice channels, a total of 2,316 calls were classified as data. Of these, 58 percent were between 10 and 20 s in duration, and 86 percent were less than 1 min. in duration. The data-call second loading attributed to the 10- to 20-s duration calls was 17.9 percent of the total observed data-call seconds; for calls of up to 1-min. duration, it was 38.3 percent.

● For supergroups with 20 or more of the 60 channels designated as nonspeech channels, in approximately 88 percent of the test runs the maximum observed data load, averaged over a 1-min. interval, was 70 percent or more.

● For these same channels, the maximum observed data load, averaged over a 1-min. interval, was 100 percent.

These measurement results permit the following conclusions:

● A 240-channel DSI system operating with a DSI gain of 2.0 has sufficient margin to accommodate the maximum observed data loading of 4.1 percent without loss of speech quality.

● A 60-channel DSI system, operating with a DSI gain of 2.0 and a 37-percent speech activity factor, can accommodate the maximum observed data loading of 4.1 percent without loss of speech quality.

● The high percentage of short-duration calls indicates that the maximum observed data signal loading, averaged over a 1-min. interval, may be viewed as an extreme loading indicator, since the observations were heavily weighted toward calls of short duration.

● Voice frequency channels that are designated for nonspeech traffic should be excluded from DSI systems.

Currently, the effect of the observed in-band data loading in designated voice channels on the speech quality of the DSI system is negligible. However, this is an important parameter that should be periodically checked to develop a history from which future in-band data-loading projections may be made.

Acknowledgments

The authors wish to thank F. Giorgio, J. Kolsrud, W. Bogaert, and M. Benjamin for their encouragement and interest; H. Suyderhoud for an extensive and positive critique and A. Gatfield for valuable discussions; M. Onufry for discussions and extensive help in clarifying the text; and D. Fifield and his staff for their help and excellent collaboration at the earth station.

The monitoring program was developed by K. Bittle and W. Ferng. The transmission program developed by C. T. Wu and R. Curtis was instrumental in the hardware implementation, in its partial design, and in the measurements.

References

[1] R. Carson, "A Digital Speech Voltmeter—The SV6," *British Telecommunications Engineering,* Vol. 3, Part 1, April 1984, pp. 23–30.
[2] S. J. Campanella, "Digital Speech Interpolation Techniques," National Telecommunications Conference, Birmingham, Alabama, December 1978, *Proceedings,* pp. 14.1.1–14.1.5.

Appendix A.   Speech activity considerations

The objective of the measurement program was to collect in-band data-loading statistics for channels designated for voice traffic only. A substantial body of speech activity data was also collected, but raw data contain measurement errors resulting from the fixed threshold activity detector and the detector output sampling frequency.

To correct the data for these factors, an SV6 speech level meter [A-1] was used to make a series of activity measurements as a function of speech level for several talkers by using prerecorded speech. The traffic analyzer was then used to measure the identical speech passages in both the high- and low-resolution modes. From these data, factors m_H and m_L can be obtained for the high- and low-resolution modes, respectively. When multiplied by the traffic analyzer (TA) activity measurement a_{ts}, these factors yield the activity, as measured by the SV6 speech level meter. Curves of m_H and m_L vs active speech level are given in Figures A-1 and A-2.

To illustrate the use of the calibration curves, assume that a high-resolution mode
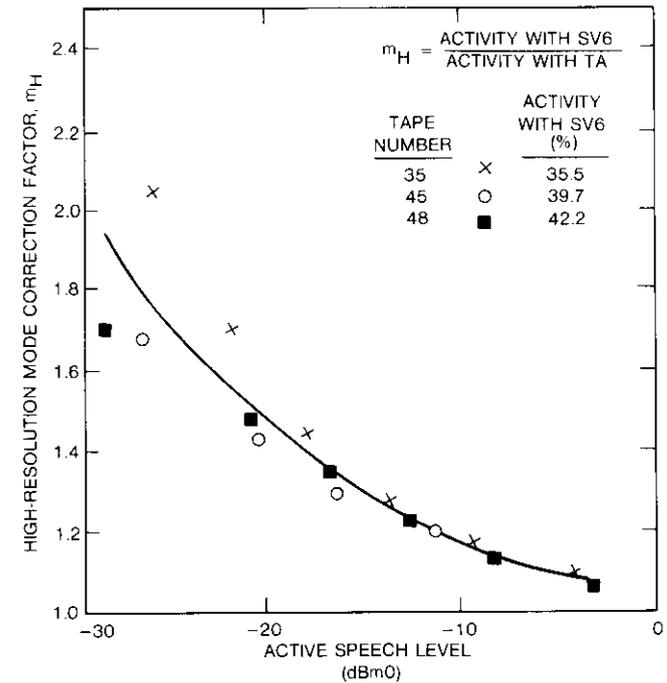
Figure A-1. *Multiplication Factor for Activity in High-Resolution Mode, m_H vs Speech Level*

measurement was made at talker levels of −20 and −10 dBm0. Figure A-1 shows that at −20 and −10 dBm0, the activities measured with the TA must be multiplied with the m_H factors of 1.45 and 1.15, respectively, to obtain the measured activity. The activity measured with the TA at −10 dBm0 exceeds that at −20 dBm0 by a factor of 1.26 (1.45/1.15).

In another application, the SV6 was used to measure the speech level of several talkers in a supergroup; the average was found to be −19 dBm0. From Figures A-1 and A-2, the corresponding multiplication factors of $m_H = 1.4$ and $m_L = 1.5$ were found. When the same supergroup was measured with the TA for two consecutive 15-min. intervals, once in the high-resolution mode and once in the low-resolution mode, the normalized activity for each mode was almost identical.

Maximum-to-average activity relationship

When the activity over the interval ΔT is averaged, the single activity values (activity distribution) are lost. The maximum activity during ΔT, however, is recorded and can be used as a check. For the simple case of only speech loading and with a ΔT value of 1 min., Figure A-3 shows the theoretical relation of maximum activity
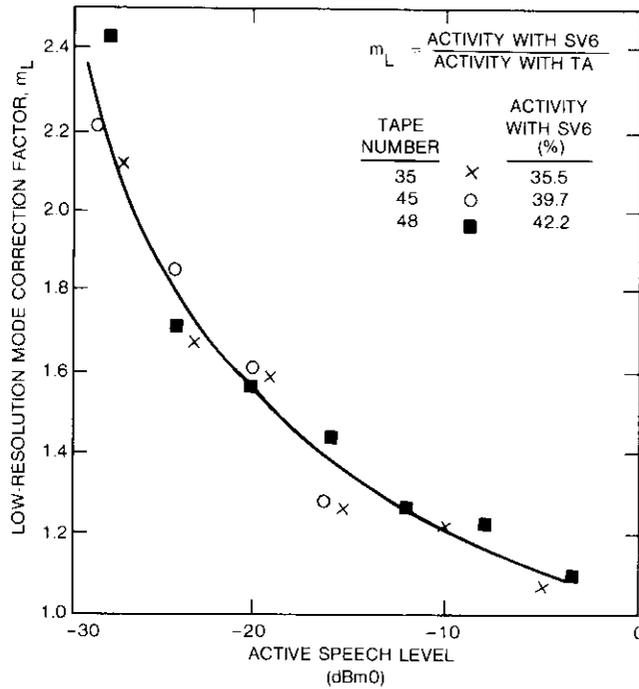
Figure A-2. *Multiplication Factor for Activity in Low-Resolution Mode,* m_L
vs Speech Level

to the average activity, $\eta = \hat{a}_{ts}/a_{ts}$, vs average activity with channel number n and sampling rate (resolution mode) as parameters. The curves show the η values of maximum probability. Measured η values are expected to be below the theoretical values because of the simplifying assumptions used in the following theoretical derivation.

A pool of n channels containing only speech is assumed. The probability, p, that a channel is active is assumed to be the same for each channel. The number of active channels, n_s, is the statistical variable. The channels are sampled m times during a ΔT time interval. It is assumed that n_s at a given sampling time is independent from the n_s value of any other sampling time. The probability to be found is that a given $\hat{n}_s$ is the maximum of the n_{sk} values, $P(n_{sk} = \hat{n}_s)_m$, for $1 \leq k \leq m$.

Assigning f_s for the probability that a given sample, n_{sk}, has the value of $\hat{n}_s$

$$P(n_{sk} = \hat{n}_s) = f_s \quad . \tag{A-1}$$

approximating the discontinuous n_{sk} with a continuous variable and assuming that the cumulative distribution F_s exists such that

$$\frac{dF_s}{dn_s} = f_s \tag{A-2}$$

$$P(n_{sk} = \hat{n}_s)_m = \binom{m}{1} f_s F_s^{(m-1)} = \frac{dF_s^m}{dn_s} \quad . \tag{A-3}$$

In equation (A-3), f_s is the probability that a given sample value is $\hat{n}_s$ which is multiplied with the probability that all other $(m-1)$ samples are equal to or less than $\hat{n}_s (F_s^{(m-1)})$. Factor $\binom{m}{1}$ expresses that it is immaterial which of the m samples takes the prescribed value $\hat{n}_s$.

By approximating the F_s, the cumulative binomial distribution by the normalized Gaussian cumulative distribution function F_N with mean and standard deviation of np and $\sqrt{np(1-p)}$, equation (A-3) can be expressed as

$$P(\hat{a} = \eta a)_m = \binom{m}{1} f_N(t) F_N^{(m-1)}(t) \tag{A-4}$$

where the maximum activity $\hat{a}$ is $\hat{n}_s/n$; the average activity a is $\left[\left(\sum_{k=1}^{m} n_{sk}\right)\Big/ mn\right] \approx p$; and the normalized variable t is given by

$$t = (\eta - 1) \sqrt{\frac{na}{1-a}} \quad . \tag{A-5}$$

The η values of maximum probability, that is, the maximum of the frequency distribution of equation (A-4) vs average activity a are shown in Figure A-3 for the given parameters of resolution mode and number of channels n. It can be shown that in the range of interest, the cumulative binomial distribution exceeds a normal distribution with an additive value R, $R \rightarrow 0$, only if $n \rightarrow \infty$. The η values obtained from the approximation are larger than those that would be obtained by using the cumulative binomial distribution. Also, if the assumed statistical independence of n_s does not exist, the maximum activity is a function of the average activity, which further decreases the value of η. These two reasons may explain at least qualitatively that as Figure A-4 shows, the measured η values are systematically below the calculated values.

Reference

[A-1] R. Carson, "A Digital Speech Voltmeter—The SV6," *British Telecommunications Engineering,* Vol. 3, Part 1, April 1984, pp. 23–30.
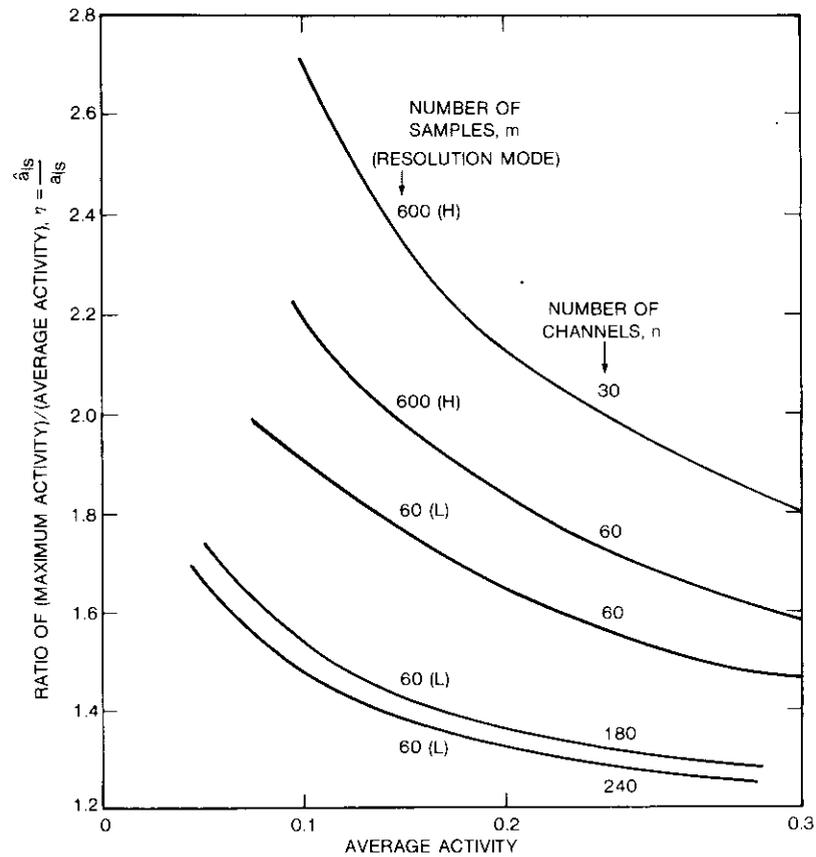
Figure A-3. *Maximum-to-Average Speech Activity, η, vs Average Speech Activity With Number of Samples, m, and Number of Channels, n, as Parameters*
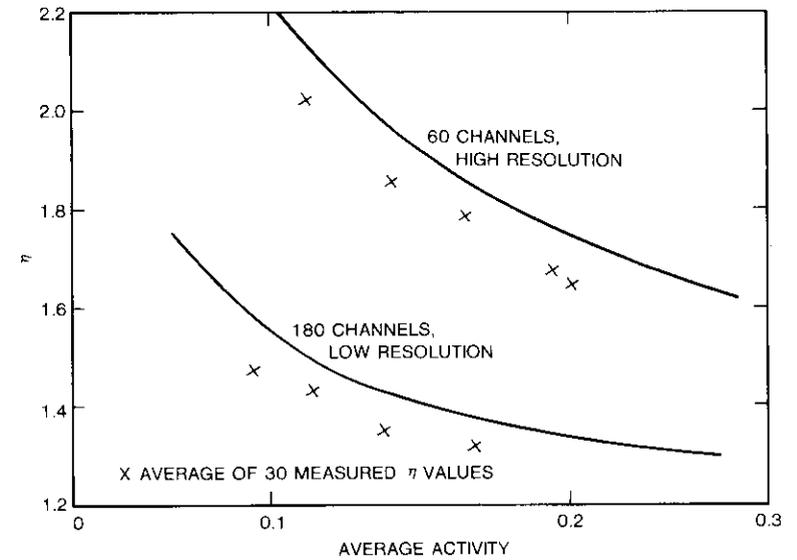
Figure A-4. *Measured Maximum-to-Average Speech Activity, Valves and Theoretical Curves*

George G. Szarvas received a Dipl. Eng. Degree in electrical and mechanical engineering from the Technical University in Budapest, Hungary. He completed postgraduate courses at the Royal Institute for Technology in Stockholm, Sweden. Before joining COMSAT in 1969, Mr. Szarvas worked for the Telecommunications Research Institute in Hungary, for L. M. Ericsson and Facit Electronics in Sweden, and for the Swedish Teleadministration. In December 1983, he retired from a position as a Member of the Technical Staff in the Voice Band Processing Department of the Communications Techniques Division at COMSAT Laboratories, where he worked in the areas of speech detection and companded systems. He is a member of the IEE.

Steve Akerman received a B.S. degree in electrical engineering from the American Institute of Technology, Chicago, and has pursued graduate work in electrical engineering at Syracuse University. Before joining COMSAT in 1981, he was with the Department of Defense for 22 years, 13 of those with the Defense Communication Engineering Center in satellite communications systems engineering control and test. From 1981 to 1984, he worked for COMSAT World Systems Division, Engineering and Development, as Transmission Systems Engineer until his retirement in June 1984.

Menfai Robert Shiu received a B.S. in mathematics from Princeton University in 1961 and an M.S. in mathematics from Georgetown University in 1967. Prior to joining COMSAT in 1980, he was a Research Associate and System Manager with the Biophysics Computer Center at Georgetown University. Since joining COMSAT as a Member of the Technical Staff, he has been involved in software development for the Programmable Interface Processor Communications (PIPCOM) project. Currently, he is responsible for systems analysis and software development in the Software Development Services Department. He has been a member of Pi Mu Epsilon, and is presently a member of IEEE and an Associate Editor of The Journal of the Pattern Recognition Society.

A computer-controlled satellite communications monitoring system for time division multiple access

J. R. POTUKUCHI, F. T. ASSAL, AND R. C. MOTT

(Manuscript received July 3, 1984)

Abstract

This paper discusses the operational features of automated satellite communications system monitors (CSMs), which are necessary for the maintenance of high-level system performance of fully variable, demand-assigned, time-division multiple-access (FV/DA/TDMA) satellite communications networks. Earth station and spacecraft parameters that are critical to the operation of such communications systems are also identified, and specific monitoring techniques are developed. In particular, the overall system design, as well as the implementation of a fully automated CSM designed and developed for Satellite Business Systems' FV/DA/TDMA network, are described.

Introduction

The satisfactory operation of a communications satellite network requires that the spacecraft and all of the earth stations maintain critical operational parameters within the system specifications. Therefore, accurate monitoring is needed to periodically assess and characterize the long-term and/or gradual performance degradation trends of the satellite and earth station hardware, and to minimize down times or outages by identifying catastrophic failures that require switchover to redundant equipment.

The selection of a specific centralized monitoring system architecture, the choice of parameters to be monitored, and the monitoring system implementation depend on the specific characteristics of the spacecraft and the network to be monitored. Until recently, commercial satellite communications systems have employed frequency division multiple access (FDMA) for the majority of communications links. In this mode, earth stations continuously transmit signals at different preassigned frequencies. To maintain efficient system operation, the down-link transmissions are routinely monitored [1] by filtering the desired signal from the multiplexed group and routing it to appropriate measurement equipment.

With the introduction of time-division multiple-access (TDMA) techniques to satellite communications, a new set of monitoring functions is now required [2]. In TDMA, all earth stations within one transponder employ the same transmission frequency, but cyclically transmit their data in short bursts. In an FV/DA/TDMA network, the duration of burst transmissions can be dynamically varied in accordance with traffic demands. To verify these transmissions, a performance monitor must be capable of synchronizing to the FV/DA/TDMA network and providing selective signal gating that is synchronous with each burst.

In this paper, spacecraft and earth station transmission parameters that are critical for monitoring the operation of such a network are identified, and specific monitoring techniques are presented. In addition, the overall system design and implementation of a fully automated CSM designed and developed for Satellite Business Systems (SBS) are also described. Modified monitoring techniques were developed to accommodate the different systems, such as the frame structure and the channel bandwidth occupancy of the INTELSAT V TDMA system. These differences and the corresponding measurement techniques are also described.

TDMA network architectures

In its baseline configuration, SBS provides a fully variable demand assignment algorithm to access the satellite [3]. The FV-TDMA frame architecture is designed to dynamically allocate transmission capacity among earth stations of the network, based on the total network activity.

Figure 1 shows the structure of the 15-ms cyclic frame. Each frame starts with a control field, followed by a traffic field. The control field, which is 10½ channels (5,376 bits) in duration, consists of the frame reference burst (FRB) and five time slots for the transmit request burst (XRB). Each earth station in the network transmits an XRB once every 20 frames (or 1 superframe). Therefore, a maximum of 100 stations can be accommodated per transponder.
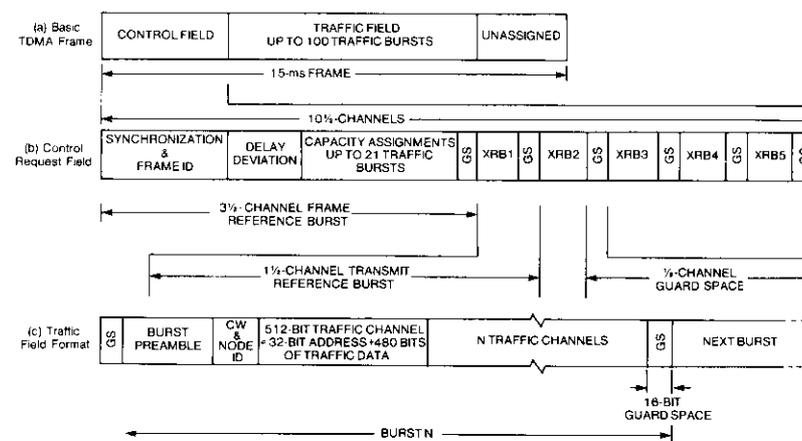
Figure 1. *SBS TDMA Frame Structure*

Nominally, the six guard times in the control field are one-half channel or 256 bits in duration. These guard times are used in the design of the CSM to establish the power drive conditions on each satellite transmitter.

Based on the capacity requested by the different XRBs, the reference station transmits the FRB, which contains the traffic burst time boundaries and may be updated every superframe of 300 ms. To monitor the performance characteristics of each earth station, it is necessary to decode the FRB and to monitor burst positions, durations, and guard times.

The traffic field, consisting of assigned and unassigned regions, contains the individual traffic bursts from each operating earth station that uses the transponder. Each traffic burst is separated from the previous burst by 16 bits of guard space. At the beginning of each burst, a preamble is included for carrier and bit timing acquisition and synchronization. This is followed by a common codeword [or unique word (UW)], unique station identification code (node ID), and data channels. Finally, each 512-bit traffic channel consists of a 32-bit address followed by 480 bits of traffic data.

In contrast with the 15-ms frame duration, INTELSAT architecture is based on a 2-ms frame [4], as shown in Figure 2. For the case shown, the frame begins with a primary reference burst (RB1), followed by a "redundant secondary" reference burst (RB2) and a sequence of traffic bursts. The traffic burst positions and durations are preassigned and change infrequently (only when the overall network is rearranged). Hence, the reference burst durations are 288 symbols, as compared to 896 symbols for SBS. It should also be noted that the SBS system operates on 48-Mbit/s digital carriers in 43-MHz
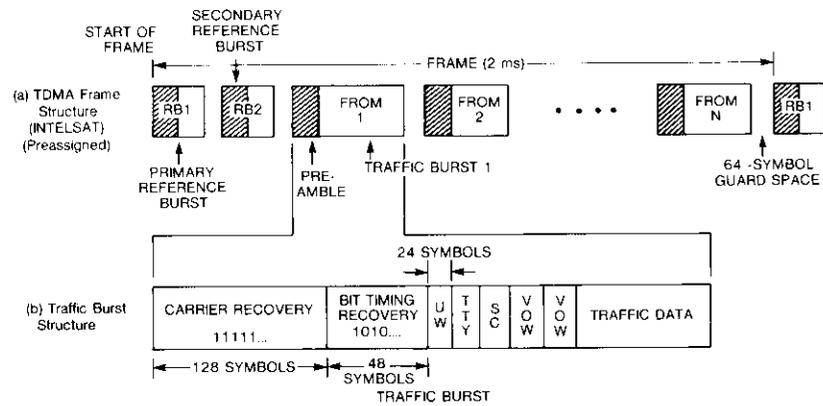
Figure 2. *INTELSAT TDMA Frame Structure*

transponders, whereas in the INTELSAT system, the transmission rate and transponder bandwidth are 120 Mbit/s and 72 MHz, respectively. As a result, the INTELSAT reference burst duration is only 4.8 μs, and the measurement of its carrier frequency therefore requires a more sophisticated technique than that used for the SBS system. In addition, the maximum guard space between bursts in INTELSAT is significantly shorter than that of SBS [48 symbols (0.8 μs) vs 128 symbols (5.3 μs)]. Hence, INTELSAT has provided a relatively large guard time of 288 symbols between the last traffic burst and RB1 to monitor the power drive levels of each burst.

Critical operational TDMA system parameters

The INTELSAT V and SBS satellites [5] are known as nearly "transparent" satellites because their primary functions are to receive up-link signals, translate them to a down-link frequency, power amplify, and retransmit without intentionally changing the incoming signal characteristics (no on-board signal processing). Figure 3, which shows the SBS satellite transponder block diagram and frequency plan, is an example of such a transparent satellite. Performance degradations occur because of receiver noise, the temperature of the earth, imperfect channel filtering, adjacent and co-channel interference, and spacecraft transmitter nonlinearity.

Figures of merit have been developed to quantify and characterize the performance of such satellites [6],[7]. These include: saturation flux density, G/T, translation frequency, e.i.r.p., transmitter transfer characteristics, channel frequency response and group delay, and antenna patterns. After a satellite
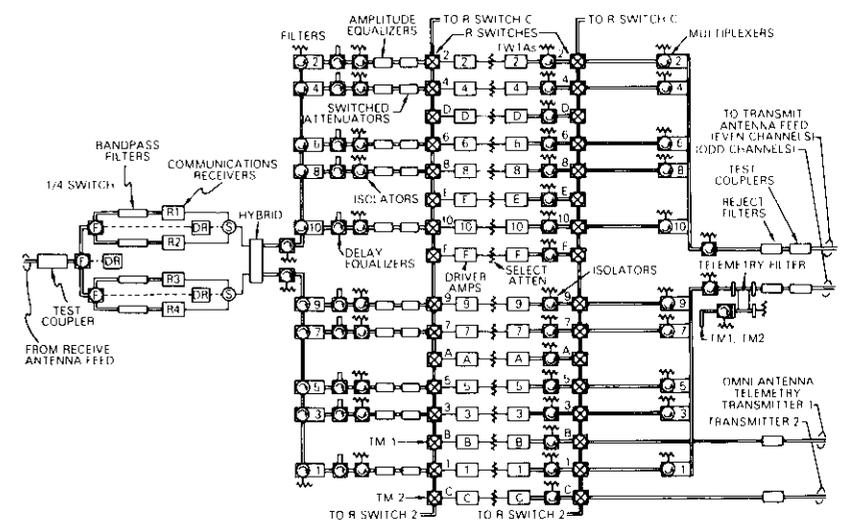
(a) SBS Transponder Block Diagram
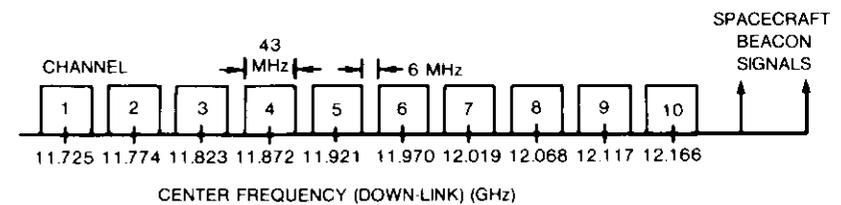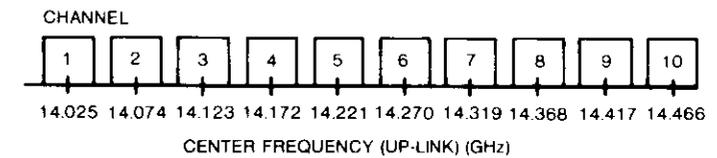
(b) Frequency Plan

Figure 3. *Satellite Communications Subsystem Block Diagram and Frequency Plan*

launch, in-orbit tests are performed to ensure that the satellite has survived, and to establish its baseline performance in terms of these figures of merit. The tests are performed by using continuous signals at a time when the satellite is dedicated to in-orbit performance testing.

As an alternative, CSMs can be used to monitor the health of the satellite communications payload and the quality of relayed signals. These CSMs are designed to measure and/or derive the satellite's figures of merit and the system parameters that are critical to satisfactory operation of the network. If a critical satellite parameter or a traffic burst parameter changes beyond system specifications, its impact on the overall TDMA network depends on the specific parameter.

Performance variation in a satellite communications subsystem can affect the traffic within a specific transponder, all traffic, or traffic emanating from specific geographic regions. For example, a reduction in the saturated output power of a specific transponder transmitter (and hence its e.i.r.p.) proportionally reduces the received burst powers of all traffic within the transponder in question. Conversely, a reduction in satellite receiver gain [and hence the gain-to-noise-temperature ratio (G/T)] produces proportionate changes in the input backoffs (IBOs) of all traffic bursts in all transponders. Similarly, degradation in G/T due to an increase in receiver noise temperature results in a corresponding degradation in link carrier-to-noise ratio (C/N) of all traffic.

In TDMA/QPSK, the satellite transmitters are operated near saturation. Hence, changes in received burst power and C/N are relatively insensitive to changes in IBO. Performance anomalies associated with transmissions from specific geographic regions result from changes in antenna coverage due to factors such as antenna pointing and reflector distortions. These anomalies will be detected as associated changes in IBO of traffic bursts emanating from the affected region.

Performance anomalies associated with earth stations other than the network reference station generally affect a specific link or a specific pair of bursts. For example, if a specific traffic burst's up-link power or frequency change is abnormal, the station that receives this transmission is affected. Also, transmit timing errors and burst transmissions that are significantly longer than those assigned affect the adjacent bursts. However, if these errors affect the reference station, the entire network will be affected.

Based on the above observations, communications performance monitoring of the satellite and the TDMA network requires the accurate measurement of the following traffic burst parameters:

    a. IBO at the transponder transmitters,
    b. link C/N,
    c. power,
    d. frequency,
    e. position relative to the reference burst,
    f. duration, and
    g. guard space following burst.

The following satellite parameters must also be monitored:

    a. satellite local oscillator frequency,
    b. satellite beacon level, and
    c. transponder linear gain variations.

Among the satellite's figures of merit discussed earlier, changes in local oscillator frequency are monitored by a direct measurement. Any changes in e.i.r.p., G/T, and antenna patterns can be deduced from the appropriate CSM measurements and processing of the measured data. For example, the transmitter's saturated power output (and from it, the e.i.r.p.) can be derived from the IBO, burst power, and beacon measurements of the traffic bursts. By comparing the IBOs of all transmitting stations, changes in antenna patterns can be identified. The IBO, C/N measurements, and transponder linear gain measurements provide a means of identifying changes in G/T. This approach is described in greater detail in the subsections that follow.

Input backoff

Monitoring the IBO of all traffic bursts at the satellite is useful in adjusting the burst power transmitted to the satellite, identifying problems associated with each earth station, monitoring gain changes in the satellite subsystems, and deriving transponder saturated power and e.i.r.p.

In TDMA systems, spacecraft power amplifiers can be operated near saturation to maximize DC-to-RF power conversion efficiency and minimize the effect of up-link power variations. However, a large overdrive condition tends to further spread the spectrum of the band-limited signals, resulting in adjacent channel interference. The overdrive condition may also degrade bit error performance because of the bit stream crosstalk and intersymbol interference caused by filters and by the nonlinearity of the power transmitter. Conversely, a significant underdrive condition decreases the up-link C/N and may degrade the overall bit error rate performance. Hence, IBO is monitored to ensure proper up-link drive conditions and also to identify problems associated with earth station transmitters.

The traffic burst IBO data provide an indirect method of monitoring variations in the satellite receive antenna coverage pattern. For example, an east-west or north-south antenna pointing error is identifiable by comparing the IBOs of all traffic burst transmissions from these regions. Changes in the antenna patterns due to reflector distortions can be identified by comparing the IBOs of all traffic bursts across the entire coverage area. If a sufficient number of stations are distributed across the coverage area, the actual antenna pattern can be deduced from the above measurements.

Monitoring of IBO can also identify performance variations in some of the satellite communications subsystems. For example, deviation of all burst IBOs by the same value is caused by receive gain variations; whereas, a common IBO degradation of bursts in only one channel indicates performance changes in the channel subsystems.

IBO MONITORING PRINCIPLE

The IBO monitoring technique discussed in this paper is based on the principle that, in the presence of a carrier operating in the nonlinear region of a power amplifier, a small CW carrier and Gaussian noise are suppressed by the amplifier in a mathematically predictable manner. This suppression is a function of the IBO and the specific characteristics of the TWTA. This principle is illustrated in Figure 4, which shows the small CW signal (pilot) levels at the output of a 12-GHz TWTA for four discrete IBO conditions of a continuous modulated QPSK carrier. For the cases shown, the pilot frequency was offset from the channel center frequency by 20 MHz, and its level was 20 dB below the level required for TWTA saturation. To clearly show the pilot suppression, the output spectrum (in the absence of the QPSK carrier) is also shown, with a pilot frequency offset of 18 MHz.

The photographs of Figure 4 were taken at the IF output of a receiver which down-converted the 12-GHz TWTA output to 70 MHz. It can be observed that the pilot suppression increases from 3 to 17 dB as the QPSK signal input drive is varied from −5-dB IBO to 4-dB overdrive. Typical TWTA transfer and pilot suppression characteristics are shown in Figure 5.

Implementation of this scheme requires transmission to the satellite of a pilot signal that is located within the TDMA/QPSK signal bandwidth, and measurement of the received pilot level in the absence and presence of the TDMA signal. From these measurements, the pilot signal suppression in the presence of the TDMA signal and the traffic burst IBO can be computed. The pilot frequency and its power level must be selected to permit accurate suppressed pilot level measurement and minimal interference with TDMA signal performance.

If a digitally modulated constant envelope carrier were transmitted continuously to the satellite, the suppressed pilot could be recovered and its level measured accurately by narrowing the effective filter bandwidth (*i.e.*, by improving the suppressed pilot-to-QPSK signal interference ratio). In this case, the measurements are average-power or steady-state measurements.

To accurately monitor the transmitter IBO, suppressed pilot measurements must be performed well within the duration of the burst. This minimizes or avoids the effects of transients caused by the pilot selection filter. These
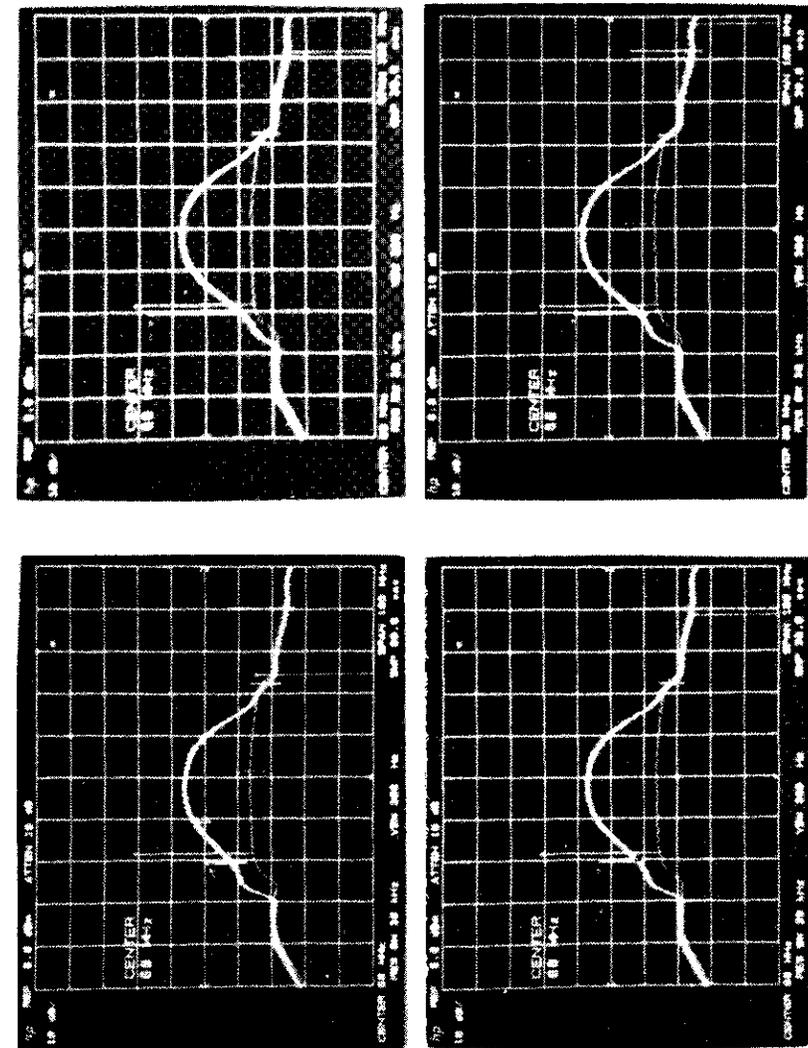
Figure 4. Power Spectra of the CW Pilot With and Without the 48-Mbit/s QPSK Signal
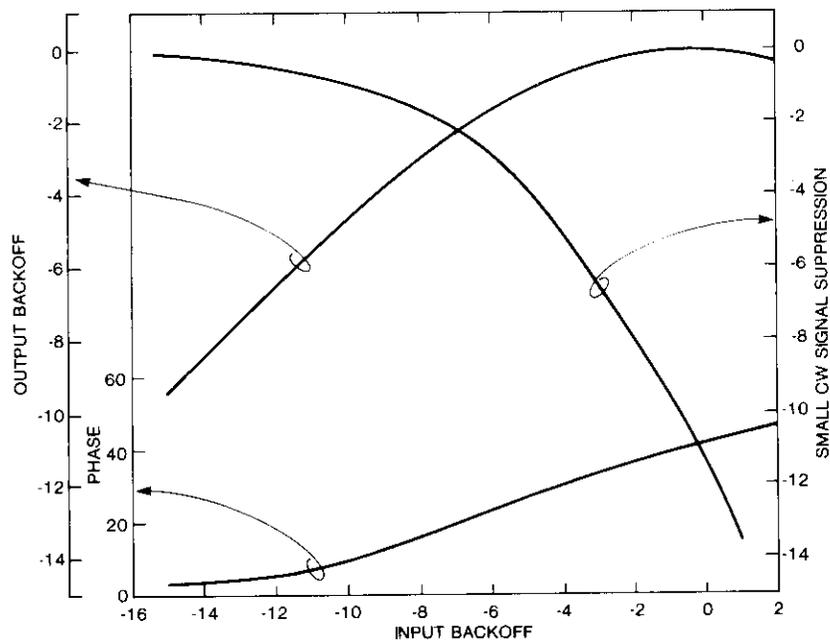
Figure 5. *TWTA Transfer and Small CW Signal Suppression Characteristics*

measurements are possible only if, after selecting the appropriate filter bandwidth, the resulting pilot power-to-QPSK signal interference ratio is sufficiently large to achieve the desired measurement accuracy. This technique was chosen for the SBS system because the ratio of the transponder bandwidth to the transmission rate of the TDMA/QPSK signal is 1.8 (24 Msymbol/s in a 43-MHz channel), and, with the pilot located 20 MHz below the channel center frequency, the QPSK signal spectral power density is approximately 15 dB below the peak, as shown in Figure 4.

The suppressed pilot measurement uncertainty can be minimized by simultaneously measuring the interfering QPSK signal power of two filters of the same noise bandwidth as that of the suppressed pilot filter, and performing the appropriate corrections. The center frequencies of the interference measurement filters are symmetrically offset by 1 MHz from the pilot frequency (Figure 6).

The unsuppressed pilot signal is measured during the fixed half-channel guard spaces within the XRB segment of each frame (Figure 1). These guard times are long enough to allow a transient-free measurement.

(a) Filter for Pilot and QPSK Interface Measurement

(b) Simplified Hardware Implementation
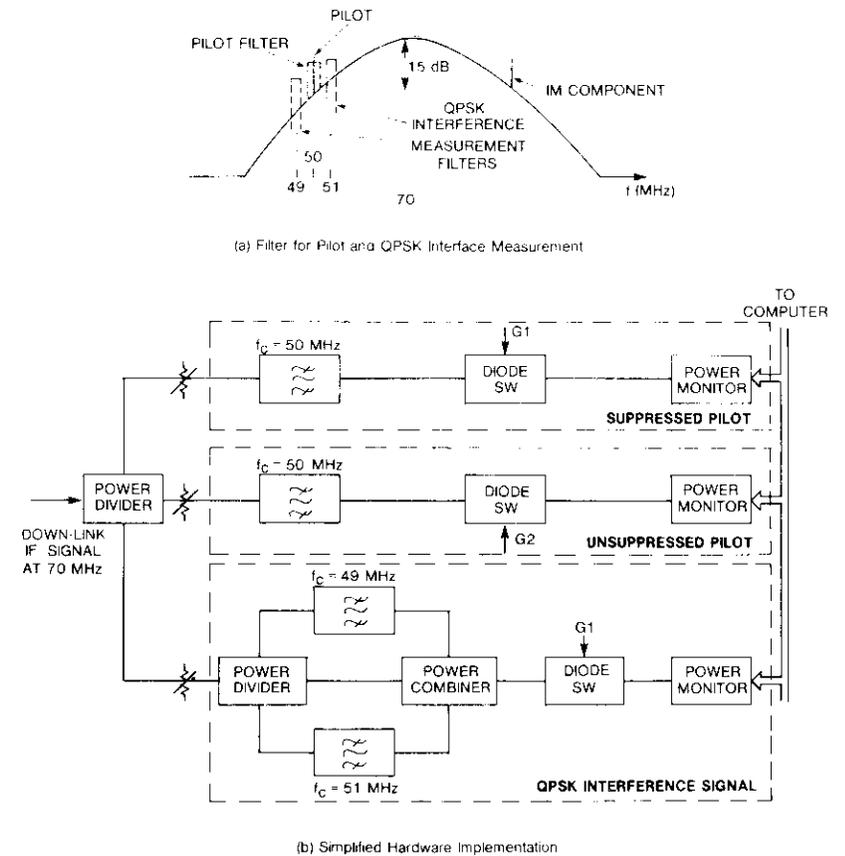
Figure 6. *Input Backoff Measurement (SBS)*

SBS IBO MONITORING TECHNIQUE

A simplified IBO monitor implementation in the SBS CSM is shown in Figure 6. To perform the measurement, two control signals are required: one is synchronous with the start of the selected traffic bursts transmission (G1), and the other is synchronous with the preassigned half-channel XRB guard spaces. These signals are generated by the CSM TDMA processor, which synchronizes to the selected network, decodes the FRB, and from it generates the network assignment map.

Since IBO computation is based solely on the difference between suppressed and unsuppressed pilot measurements, both of which are performed simul-

taneously, measurement accuracy is unaffected by path loss and equipment variations at the monitoring stations.

INTELSAT IBO MONITORING TECHNIQUE

In the INTELSAT system, the allocated transponder bandwidth-to-transmission rate is 1.2 (60 Msymbol/s in a 72-MHz channel), and the QPSK signal spectral power density within the transponder bandwidth is no lower than 6 dB (Figure 7). If the suppressed pilot filter bandwidth is reduced to compensate for this 9-dB-higher QPSK interference level (compared to that of the SBS system), the time required to measure this pilot power becomes excessive. The problem is worsened because the minimum length of burst in the INTELSAT system is only 4.8 µs. This situation led to the development of a preamble-based measurement technique for the INTELSAT CSM [8],[9], as described below.
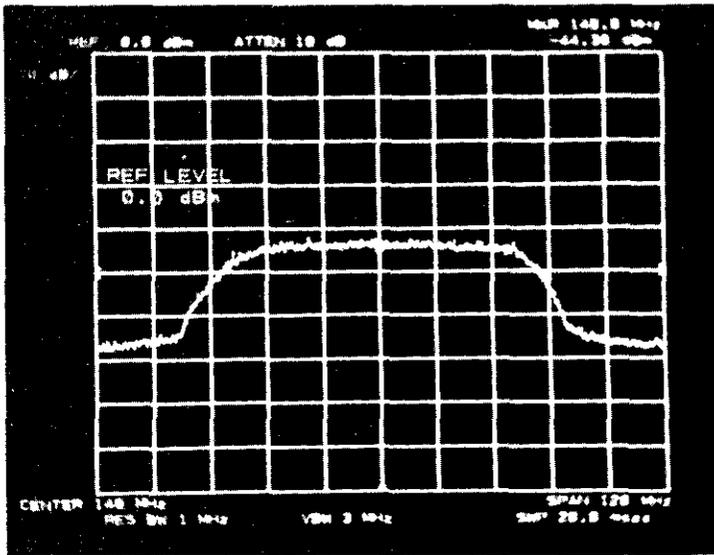
Figure 7. *QPSK Spectrum with Pilot (INTELSAT)*

INTELSAT PREAMBLE-BASED MEASUREMENT TECHNIQUE

All burst transmissions in a TDMA network begin with an identical preamble consisting of carrier and bit timing recovery. In the INTELSAT system, the carrier recovery segment consists of a 128-bit unchanging bit sequence (either all 1s or all 0s) which produces an unmodulated carrier. The bit timing

recovery segment consists of a 48-bit alternating 1,0 sequence which produces 0° and 180° phase alterations of the carrier frequency at the symbol rate.

Figure 8 shows the carrier spectra for each segment of the traffic burst and
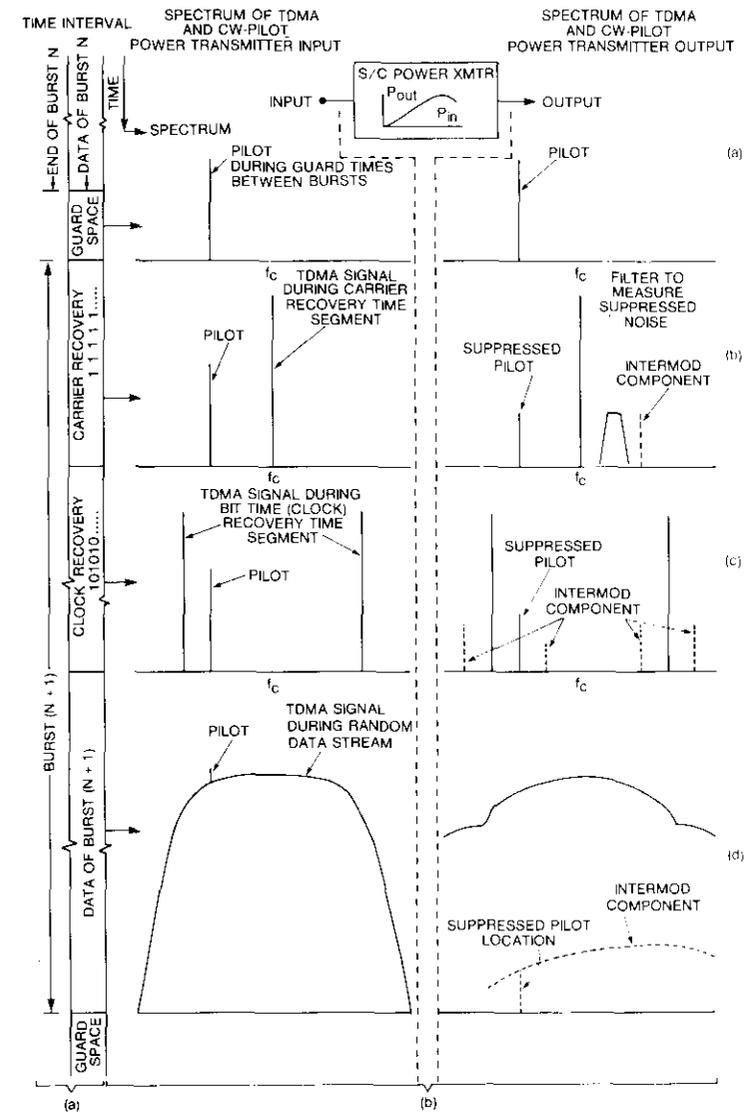
Figure 8. *INTELSAT Traffic-Burst-Dependent Carrier Spectra*

the pilot at the transmitter input and output. As can be expected, during the guard times between bursts, only the unsuppressed pilot is observed at the input and output (Figure 8a). In the carrier recovery segment, the output spectrum consists of an unmodulated carrier at f_c, the suppressed pilot at f_p, and the intermodulation product at $2f_c - f_p$ (Figure 8b). During the bit timing recovery segment, the output spectrum consists of two equal carrier components at $(f_c - f_{clock})$ and $(f_c + f_{clock})$ where f_{clock} is the symbol rate, the suppressed pilot at f_p, and their intermodulation products (Figure 8c). Over the data portion of the burst, the output spectrum (Figure 8d) contains the traffic burst's QPSK spectrum, the suppressed pilot at f_p, and a broadband low-level spectrum that results from the convolution of the pilot and the QPSK spectrum.

Since the TDMA carrier spectrum during the preambles consists only of CW components, interference contribution to the suppressed pilot measurement is limited to the thermal noise of the system. Hence, the suppressed pilot can be measured accurately during the carrier or bit timing recovery segment of each traffic burst preamble. Uncertainty can be reduced by measuring the suppressed noise with a filter of the same noise bandwidth as that of the suppressed pilot, and correcting the pilot-plus-noise measurement. The noise filter characteristics should be selected to exclude any intermodulation products within the filter's measurement bandwidth. The unsuppressed-pilot level can be measured in a preassigned periodic guard space within the frame.

Carrier-to-noise ratio

The C/N of all traffic bursts is monitored to assess their transmission quality. This measurement is also useful in identifying performance anomalies associated with satellite noise temperature.

In QPSK/TDMA systems, C/N measurement in an operating system must account for the fact that during the data transmission, noise is suppressed by a nonlinear transmitter, and the amount of suppression depends on traffic burst IBO. With the preamble-based measurement scheme discussed above, the CW carrier power and the suppressed noise can be measured accurately during the carrier (or bit timing recovery) segment of the burst. The carrier level can be measured with a filter centered at f_c (or $f_c \pm f_{clock}$), and noise can be measured with a filter centered away from known signals and their intermodulation products, as shown in Figure 6.

In addition to its simplicity of implementation and its ability to accurately monitor the overall performance of all traffic bursts, this measurement scheme provides a means to assess performance anomalies associated with the satellite G/T. For example, a common degradation of all traffic burst C/Ns in a satellite can be caused by degradation in the receiver and/or the receive antenna gain,

or the receiver noise temperature. As discussed earlier, comparison of all IBO measurements can account for changes in the receive subsystems. Hence, any additional degradations are attributable to an increase in receive noise temperature.

Burst power

Burst power levels are monitored to derive the transponder saturated power output (e.i.r.p.), and also to ensure that the received burst levels and burst-to-burst level variations are within the dynamic range and automatic gain control capabilities of the receive modem.

If the satellite has a stable on-board beacon located close to or within the communications band and transmits through the communications antenna (as is the case with SBS), the measured burst power levels can be related directly to the transmitter output power. Since the IBO for each traffic burst is measured, the saturated power output can be computed by using the known transfer characteristics of the transmitter. The above measurements are performed for each burst in the transponder; therefore, averaging of this data provides significant improvement in measurement accuracy.

For analog transmission, the carrier power level can be obtained by comparing a continuous down-link power measurement with the beacon signal power to remove any fluctuations in the transmission path. In the SBS FV/DA/TDMA systems, traffic burst transmission durations and frame occupancy vary with time. In a preassigned TDMA, while the burst durations are fixed by the network burst time plan (BTP), individual transmission durations generally differ. In addition, the BTP is subject to occasional change when a network reconfiguration occurs. Hence, a power measurement scheme based on measurement of fixed-duration samples of the traffic burst is preferable. The selected measurement duration should be less than the minimum expected duration of the burst. Uncertainties in burst power measurements can be minimized by correcting for low-noise amplifier (LNA) and communications receiver (consisting of D/C and IF amplifiers) gain variations, and down-link path loss variations.

The LNA and receiver gain calibration is performed by injecting a known level of a small CW signal at the LNA input and measuring its level at an appropriate IF output of the receiver. To avoid interference, this injected signal should be located outside the burst power measurement filter bandwidth.

Burst power is measured in SBS CSM as follows. The IF signal from the communications receiver is filtered and applied to a solid-state switch. A fixed-duration burst-synchronous measurement gate (generated by the TDMA processor) is applied to the control element of this switch, which outputs a fixed-duration sample of the traffic burst taken once every frame. An averaging

(thermistor-type) power monitor measures this power. Since the frame and the measurement duration are fixed, peak power can be computed. These data are corrected as described above. Figure 9 shows the LNA and communications receiver gain calibration scheme.
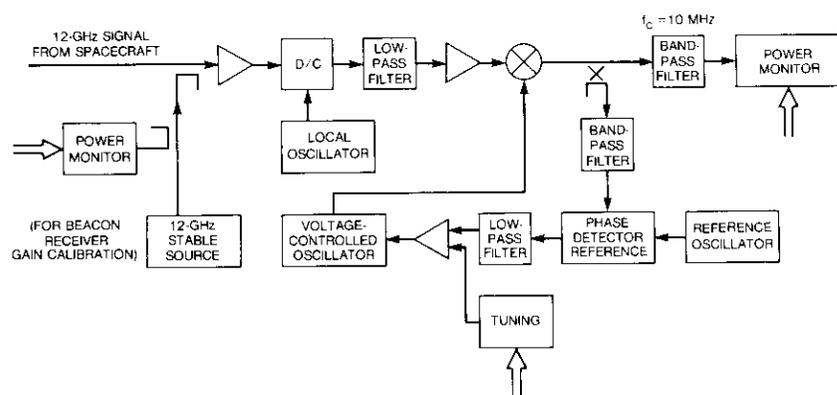
Figure 9. *Beacon Receiver*

Saturated output power

The saturated output power of the transmitter is a derived parameter which is useful in assessing the health of the transmitters and identifying and characterizing any long-term performance degradation trends. SBS satellites and many other relatively high-power satellites use traveling wave tube amplifiers (TWTAs) as transmitters. However, historically TWTAs are among the satellite's most life-limiting devices.

The primary life-limiting element of the TWTA is the cathode. With aging, the space-charge-limited cathode current at the beginning of life (BOL) progresses toward a temperature-limited condition and then decreases noticeably. Since TWTA linear gain is directly related to the cathode current, it also decreases over time, followed eventually by a noticeable reduction in saturated power output. The specific degradation rates of these parameters depend on factors such as cathode type, operating temperature, and the initial setting of the space-charge-limited current. Typical TWTA gain characteristics at BOL and at the end of 7 years [10] (assuming a 1-percent cathode current drop per year) are shown in Figure 10. For the case shown, the linear gain and the saturated output power decrease by 3 and 0.7 dB, respectively, at the end of life (EOL). In addition, EOL operation at saturation requires an increase in drive level of 4 dB compared to that at BOL.
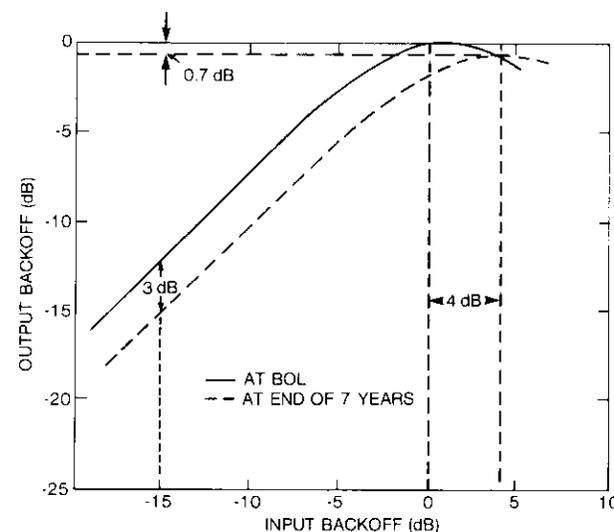
Figure 10. *BOL/EOL TWTA Characteristics*

To optimize link performance and minimize the effects of satellite receiver and TWTA gain variations, commandable gain steps are included within the transponder. In addition, redundant transmitters are normally included in the satellite so that if the transmitter experiences an unacceptable drop in saturated power output, or a total failure, it can be replaced by using an appropriate ground command. To use these resources efficiently and to maintain continuity of service, variations in the transponder linear gain and the saturated power output of the satellite transmitter must be monitored.

SBS SATURATED OUTPUT POWER MONITORING TECHNIQUE

The transmitter saturated output power in the SBS CSM is derived from traffic burst IBO, received burst power, and the received level of the on-board beacon(s) [11]. Of these, the backoff and burst power measurements were discussed earlier. The following is a brief description of the beacon-level measurement scheme.

Each of the first-generation SBS satellites has two on-board beacons located at the upper edge of the 12-GHz communications band. To provide the ability to monitor either of the beacons of any one of the satellites, a programmable phase-locked loop (PLL) type beacon receiver configuration, shown in Figure 11, was chosen.
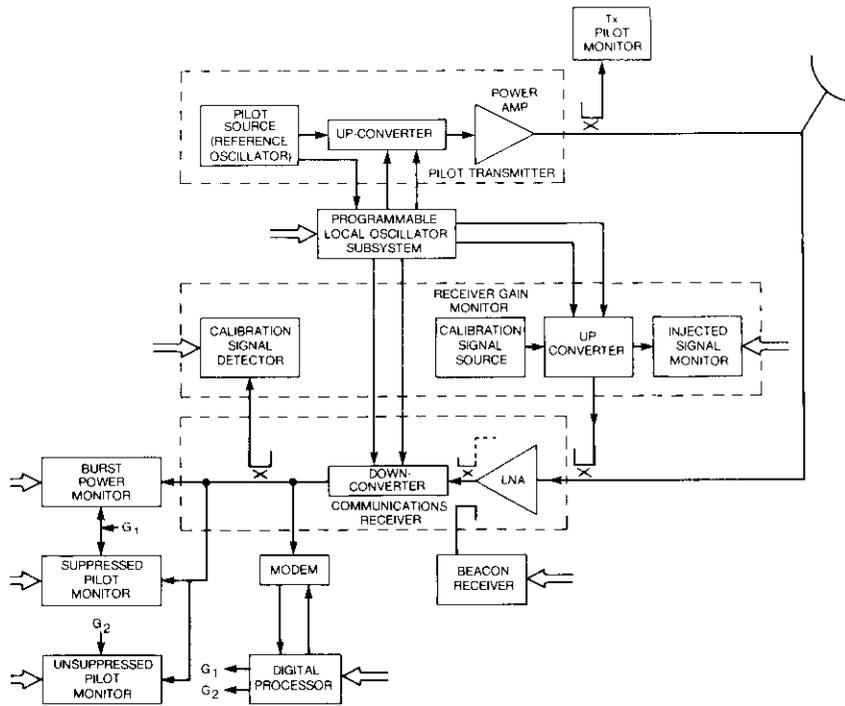
Figure 11. *TWTA Saturated Power Output Monitor Block Diagram*

For the case shown, the received down-link signals are filtered by a narrowband bandpass filter centered at 12.19 GHz, amplified by a 12-GHz amplifier, and frequency-translated to 70 MHz. After additional filtering and amplification, this signal is applied to another mixer. A programmable voltage-controlled oscillator acts as the local oscillator, and its frequency is chosen such that the selected spacecraft beacon frequency is translated to 10 MHz. The PLL consists of a 10-MHz bandpass filter phase detector, a low-pass filter, a voltage-controlled oscillator, and a mixer. The PLL output is always locked to the 10-MHz highly stable reference source. This output is filtered with a very narrow bandwidth bandpass filter centered at 10 MHz to improve the beacon signal-to-noise ratio, and applied to the power meter.

MEASUREMENT UNCERTAINTY REDUCTION

Uncertainties in the spacecraft beacon-level measurement caused by CSM LNA and beacon receiver gain variations are minimized by periodically measuring the overall gain and performing necessary corrections. Since the

LNA is common to both the communications and beacon receivers, this correction can be performed in three steps, as follows:

a. The LNA communication receiver cascade is calibrated by injecting a known level of an out-of-communications-band CW signal into the LNA input and comparing it with the generated output level at an appropriate filtered IF output.

b. The communications receiver gain is calibrated by injecting the same CW signal at the input of the receiver down-converter and comparing it with the generated output level at the same IF port as in step (*a*).

c. The beacon receiver gain is calibrated by injecting another low-level CW signal into the input of the beacon receiver and comparing it with its filtered output at the same output port where the spacecraft beacon is measured. From steps (*a*) and (*b*), the LNA gain is calculated and added to the beacon receiver gain to evaluate the gain of the LNA beacon receiver cascade.

TWTA SATURATED POWER COMPUTATION

The TWTA saturated power is computed from the traffic burst power and its IBO, the beacon level, the LNA and communications receiver gain, and the LNA and beacon receiver gain, combined with the above gain calibration data. The computational procedure is as follows:

a. The measured beacon and burst power at IF are referred to the LNA input by using the gains of the respective chains.

b. From the burst IBO and the known tube characteristics, the expected burst power of the saturated TWTA is computed.

c. The difference between the burst power of a saturated TWTA and the beacon level referred to the CMS LNA input is the same at the satellite transmit antenna flange. By comparing this result with that measured during in-orbit testing (or at prelaunch), TWTA saturated output power can be computed. These measurements are performed for all bursts within the transponder, and the data are averaged to minimize the effects of the individual measurement uncertainties. It should be noted that overall measurement accuracy depends primarily on beacon power stability.

Transponder linear gain

As discussed earlier, the TWTA linear (small signal) gain generally degrades more rapidly, with aging, compared to the saturated output power degradation.

Also, the drive level required to saturate the tube increases at approximately the same rate as the reduction in the linear signal gain. Hence, a degrading tube can be identified earlier with a high degree of reliability from an observed long-term linear gain degradation than from an observed drop in its saturated power output.

The IBO measurement is performed by transmitting a small CW signal (pilot) from the CSM and measuring its received level in the presence and absence of the traffic burst. A relative measure of transponder linear gain can be obtained by measuring the transmitted pilot level and comparing it with the unsuppressed level. Observed variations in a long-term profile of transponder linear gain can be attributed to satellite receive and transmit antenna gain changes and pointing errors; satellite receiver, driver, and transmitter gain changes; and propagation effects. Of these, changes due to transmit antenna gain, antenna pointing, and propagation effects can be identified by comparing the computed linear gain with the received beacon level. Changes in receive antenna pointing and in the receiver affect the linear gain of all transponders and hence can be identified easily. The remaining changes in linear gain will then be due to the driver and the transmitter of the transponder.

Satellite local oscillator frequency

The spacecraft local oscillator frequency is computed from a measurement of the transmitted and received pilot frequencies. This is the same signal used in the IBO monitor.

Since the received pilot (from the spacecraft) is typically 30 dB below that of the TDMA carrier, filtering it for a frequency measurement directly at the down-link frequency is impractical. Therefore, in the CSM, this frequency is measured after a dual down-conversion and appropriate narrowband filtering to improve its C/N. The transmitted pilot is derived from a highly stable source and up-converted to the 14-GHz band by a dual-frequency up-converter. Because all of the local oscillators are phase-locked to a common reference source, they do not contribute any additional errors.

Burst frequency

Monitoring of the transmit carrier frequencies of traffic bursts ensures that these frequencies and the burst-to-burst frequency variations are within the AFC pulling range and burst-to-burst carrier recovery capability of the demodulator. For continuous QPSK transmission, the transmit carrier frequency can be computed by measuring the recovered carrier frequency of the QPSK demodulator with a direct-reading frequency counter and correcting it for satellite frequency translation errors.

In TDMA systems, this measurement must be completed within the duration of the shortest burst in the frame. Hence, excessive uncertainty may be associated with a single measurement because of truncation errors in the direct frequency counter. For this reason, the recovered carrier frequency monitor employs a gated reciprocal frequency counter. The external gate to the counter is synchronous with the traffic burst, and of a duration slightly less than the minimum expected burst duration. Measurement uncertainty is reduced by averaging measurements over as many frames as necessary to achieve the desired accuracy. Uncertainty can also be reduced by performing measurements at a lower frequency (i.e., by down-converting the modem's recovered carrier).

The transmit carrier frequency can be computed from the recovered carrier frequency measurements by correcting these measurements for frequency translation errors caused by the satellite and CSM local oscillator. The CSM local oscillator frequency error can be computed by periodically injecting a known stable CW signal into the LNA input, as was done for down-link chain gain (calibration), and measuring the down-converted frequency of this injected signal. The satellite local oscillator frequency measurement error was described previously.

Burst duration and guard spaces between bursts

In TDMA systems, the earth terminals must precisely time (i.e., synchronize) their transmissions to ensure that the bursts are properly interleaved at the satellite. Guard spaces are also provided between consecutive bursts to minimize the risk of burst overlaps.

NEED FOR MONITORING

Significant errors in transmit burst positions and/or transmit durations can create burst overlaps and cause a specific link outage at one extreme. A network outage can occur if the reference burst is affected. For example, errors in transmit burst position can cause the burst code word to fall outside the aperture set for its detection. The burst modem cannot detect the code word, and a link outage results. Conversely, if a burst transmits at an expected time but the transmission is terminated prematurely, the TDMA processor/modem cannot identify this occurrence. By monitoring burst transmission durations and the guard spaces between bursts, such anomalies can be identified.

For example, if a traffic burst of proper duration is transmitted earlier than expected, guard space monitoring will indicate an abnormally long guard space between this and the following burst. Similarly, delayed traffic burst transmission can be identified by the presence of a large guard space before

the burst. Traffic burst transmission of shorter-than-expected duration can be identified by the burst duration monitor and comparison with the expected duration.

MONITORING DIFFICULTY

Conventional burst duration/guard space monitoring techniques involve envelope detection of the TDMA transmissions and measurement of envelope on-off times. However, envelope detection of a band-limited QPSK signal can contain apparent signal-off times produced by the 180° phase transitions, as shown in Figure 12. The scheme implemented in CSM (Figure 13), as described below, is immune to these short-duration nulls [12],[13].
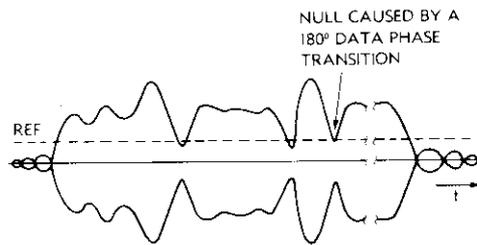
Figure 12. *Typical Received TDMA Burst Waveform*

CSM BURST DURATION AND GUARD SPACE MONITOR REALIZATION

For the case shown, the modulated signal is amplified by an automatic leveling amplifier, envelope detected, filtered, and applied to a comparator with a manually settable threshold. The comparator output contains not only a logic "low" output for each of the guard spaces, but also short-duration logic "low" outputs caused by the 180° phase transitions of the modulated signal. Without further processing, these short-duration nulls can cause measurement errors in the form of a premature end to the burst. Thermal noise voltage spikes during the guard spaces can also cause errors in the guard-space measurements.

To overcome these problems, the comparator output is sampled only during a no-null period. These periods are predictable since the nulls, which naturally occur between two consecutive symbols, happen at the symbol rate and are in-phase with the recovered PSK clock signal from the TDMA demodulator. As shown in Figure 13, the recovered clock is frequency-doubled, phase-shifted, and fed to a master/slave flip-flop to produce a bit-rate sampled and held output which is relatively null-free (Figure 14c). Further error correction
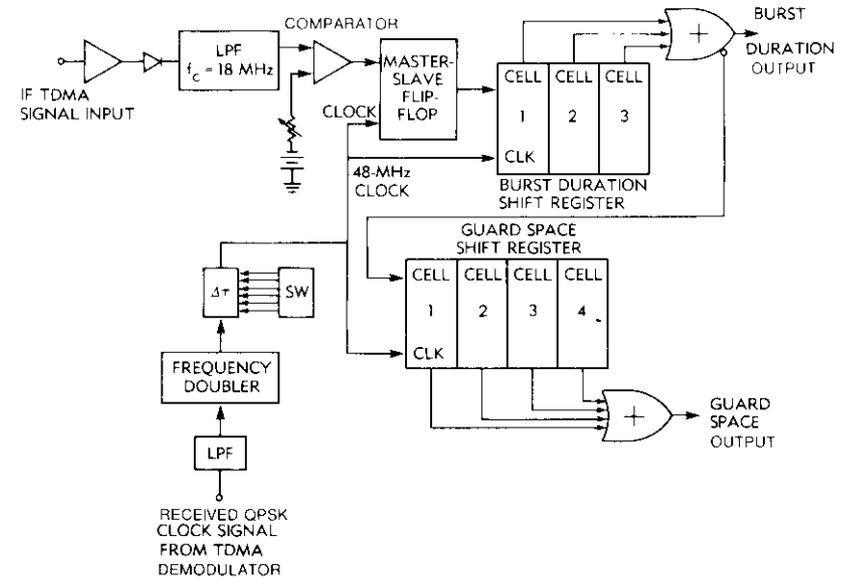
Figure 13. *Schematic of Burst Duration and Guard Space Monitor*

is accomplished by serially shifting the flip-flop output through a 3-cell burst duration shift register. The unequally delayed parallel cell outputs are then combined using an "or" logic gate to produce a null-free burst duration signal. A pattern of three successive "low" samples is considered to indicate the end of the burst, and hence the beginning of a guard space.

Waveforms (a) through (g) of Figure 14 illustrate the operation of this correction mechanism. The dots, numbered 1 through 23, on the detected envelope (Figure 14b) indicate sampling times which coincide with the leading edge of the 48-MHz clock. Figure 14d shows the contents of the 3-bit shift registers, and Figure 14f shows the output of the logic gate (burst duration output) at each sampling instant. Since the burst duration transition can only occur at the end of three successive "zeros" of the shift register contents, the burst duration output (Figure 14e) is not affected by the detected envelope nulls caused by the 180° phase transitions of the waveform shown in Figure 14b. However, the digitized burst duration envelope is longer than the actual duration by exactly 3 bits.

The guard space measurement is initiated by the digitized burst duration transition and terminated when the 4-bit guard space shift register contains all zeros. As a direct consequence of this guard space definition, the digitized guard space waveform (Figure 14g) is both reduced in length by 1 bit times

(a) 48-MHz Clock
(b) Detected Envelope
(c) Flip-Flop Output
(d) Burst Duration SR Content
(e) Burst Duration
(f) Guard Space SR Content
(g) Guard Space
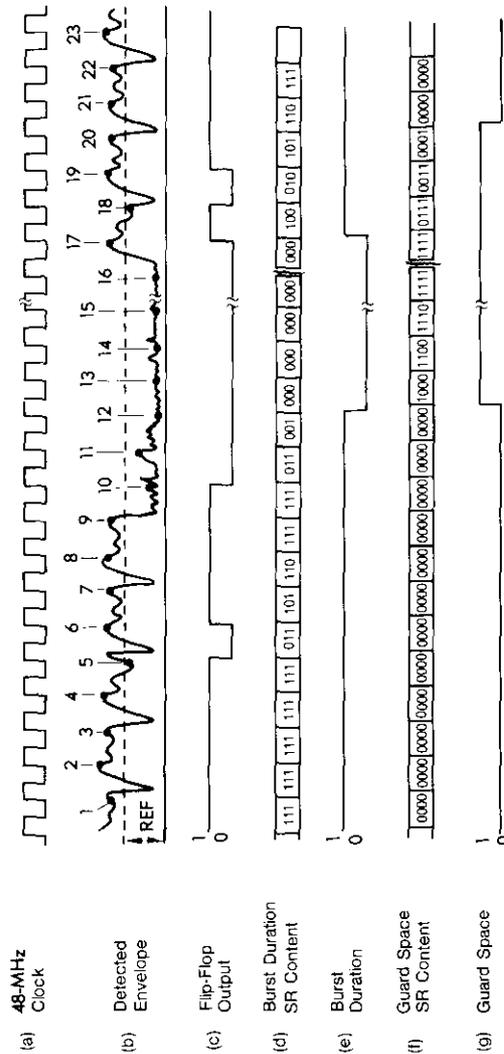
Figure 14. *Worst-Case Detected Envelope Signal of Burst Duration and Guard Space Monitor*

the number of burst duration shift register cells, and increased by 1 bit times the number of guard space shift register cells. Conveniently, the two effects tend to be self-correcting.

To measure the burst duration and the guard space, the outputs shown in Figures 14e and 14g are applied to two separate computer-controlled counters, and the fixed offsets discussed above are software-corrected.

CSM hardware configuration

Figure 15 is a photograph and Figure 16 is a functional block diagram of the SBS CSM. As shown in Figure 16, the signals received from the satellite are amplified by an LNA and applied to a dual-frequency conversion communications receiver and a PLL beacon receiver. The output of the communications receiver is applied to a burst modem and to the individual communications parameter monitors described previously.

The modem interfaces with the TDMA processor, which acts as a TDMA receive-only terminal. The processor synchronizes to the network, decodes the FRB and derives the network assignment map, and sets up apertures for the modem's codeword detector. It then processes the detected codeword data, compares the data with the network assignment map, and generates the real-time control signal necessary for the individual performance monitors.

The transmit pilot signal is derived from a common highly stable crystal oscillator, up-converted by a dual-frequency up-converter, amplified by a power amplifier, and transmitted by the CSM antenna. The transmit and receive subsystem local oscillators are phase-locked to the common reference oscillator.

The CSM computer interfaces with and controls all of the measurement subsystems, the local oscillator frequencies, the beacon receiver, and the TDMA processor. The computer also performs all instrument control and data processing functions. Computer peripherals include a hard-disk memory to store all raw and processed parameter data, a fine printer and plotter, and a real-time clock.

CSM operating modes

The SBS CSM is designed to operate in three monitoring modes: sequential, fast-scan, and dwell. The operational features of these modes are discussed in the subsections that follow.

Sequential monitoring mode

The sequential monitoring mode (SMM) is the normal CSM operating mode. The simplified flowchart in Figure 17 illustrates the measurement sequence.
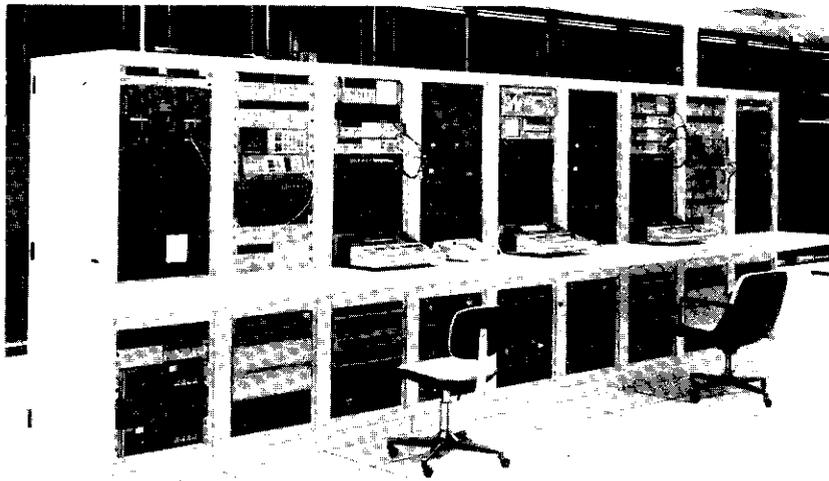
Figure 15. *Photograph of the CSM*
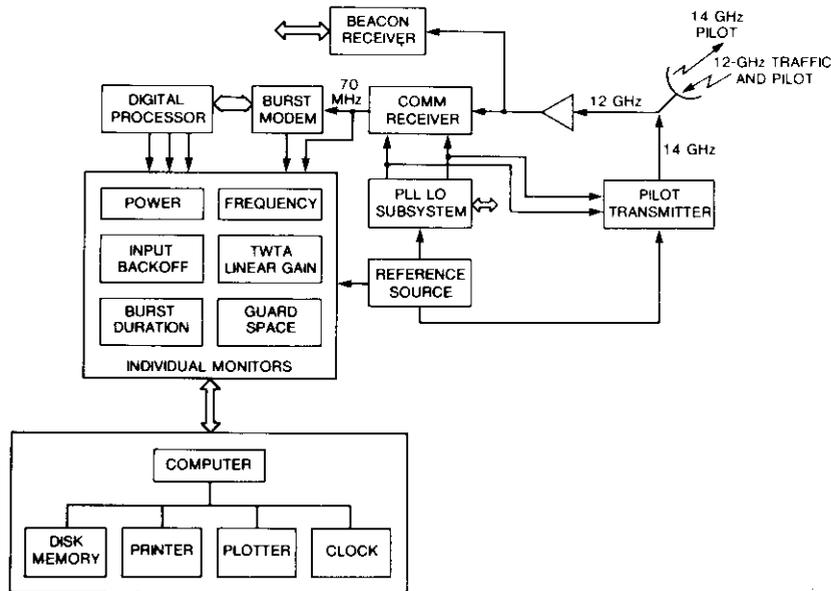
Figure 16. *Simplified Functional Block Diagram of the CSM*

Monitoring in the SMM begins with the first on the list of selected transponders. Initially, the CSM local oscillator frequencies are set by the
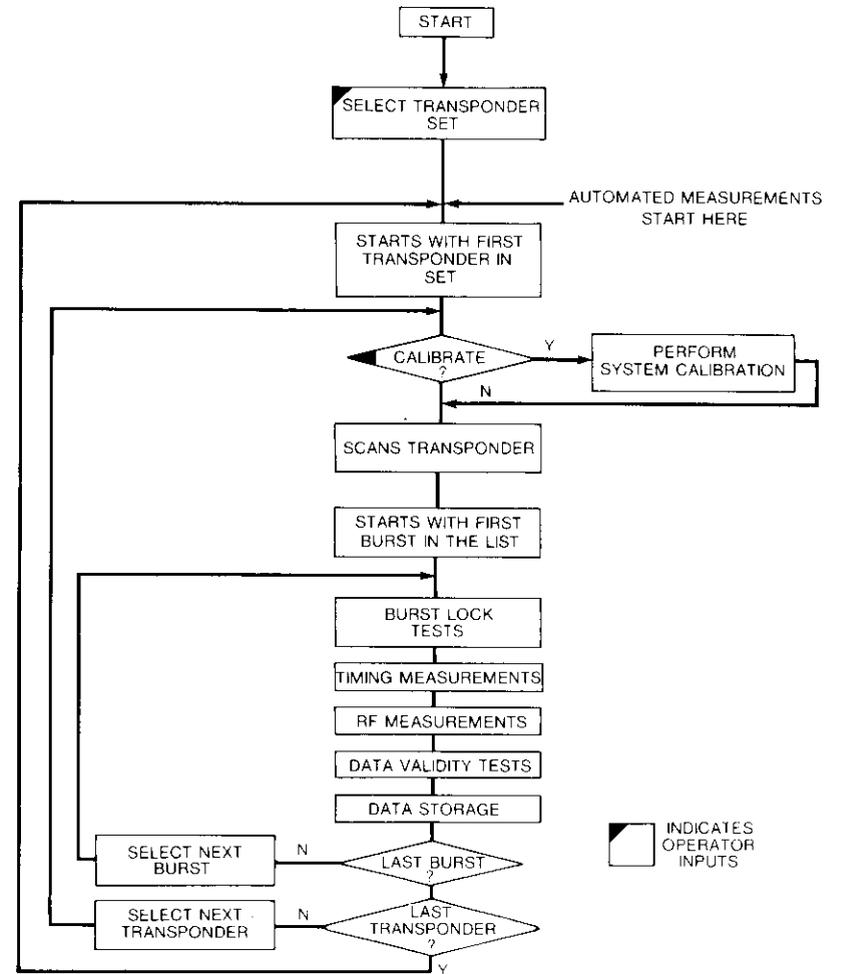
Figure 17. *Simplified Autoscan and Monitor Flow Chart*

computer in order to select the TDMA signals from the desired satellite channel, transmit the CSM pilot signal at a frequency 20 MHz below the selected transponder up-link center frequency, and set the locally injected calibrating signal frequency 21 MHz higher than the down-link channel center frequency.

Next, a complete system calibration is performed, during which the gains

of the CSM communications and beacon receivers are measured. In addition, the frequencies of transmitted and received pilot signals are measured, and the satellite local oscillator frequency is computed based on these measurements.

Simultaneously, the TDMA processor synchronizes to the selected network and generates the network assignment map. In the fast scan mode described in the next subsection, the processor also generates a map of all active traffic bursts, compares them with the expected assignment map, and identifies the bursts whose code words could not be detected.

After system calibration is completed, the computer interrogates the processor and collects the above data. The burst monitoring process begins with the first burst in the network assignment map. The computer transmits the assignment number and identification code of the burst to the processor, and the processor attempts to "lock" and track the burst. If the process is successful, the computer then measures the position of the burst relative to FRB, as well as the total number of active traffic bursts in the network, and stores the information. The processor also generates all of the control signals necessary to perform the RF measurements. If the unique word of the traffic burst is not detectable, the processor sets up a unique bit pattern in the status register of the processor-computer interface which the computer checks before initiating measurements.

After ensuring that the processor has tracked the burst, the computer interrogates all of the CSM monitoring subsystem instruments sequentially. At the end of this process, the processor interface status register is checked again to determine if the burst has lost lock or if any assignment changes have occurred during the measurement process. No-lock conditions are identified by unique bit patterns in the status register.

If neither of the above conditions has occurred (*i.e.*, the processor stays locked to the burst), then all of the measurement data are considered valid and are stored. The above measurement sequence repeats sequentially for all bursts in the network.

If the status register indicates a loss-of-lock condition while the measurements are in progress, another attempt is made to lock onto this burst. If the burst lock does not occur, the burst is considered to be inactive. An alarm is generated, and the monitoring continues with the next burst in the list.

If the burst assignments change during the monitoring process, the data are considered invalid, and the digital processor is commanded to track the burst again.

If the traffic burst loses lock during a measurement sequence, the dwell mode monitoring described below can be invoked to identify the specific performance anomaly that caused the condition.

Fast scan mode

All of the TDMA processor functions described above are performed by digital hardware, and thus are accomplished in near-real time. However, some of the burst mode measurements, such as frequency, require multiple measurements and subsequent averaging to improve measurement accuracy. In other words, measurement accuracy of burst mode signals is achieved at the expense of the time it takes to perform a measurement. Thus, near-real-time TDMA network monitoring can be performed if the RF measurement segment of the CSM is skipped. The monitoring is restricted to network synchronization and network activity (related to the presence and absence of traffic bursts) based only on the modem's ability to detect burst code words. This is referred to as the fast scan (monitoring) mode (FSM).

Dwell monitoring mode

In an operating system, situations may arise making it necessary to dwell on a specific burst which is experiencing trouble, for an extended period. For example, this step may be required when an anomaly observed during SMM or FSM requires further exploration to identify the specific cause; or a station entering into the network may need CSM assistance in antenna pointing and/or setting up-link power levels; or a station which shows consistently abnormal IBO needs CSM assistance for up-link-level adjustment.

The dwell monitoring mode (DMM) can be selected by using the CSM computer to interrupt the normal SMM. When in the DMM, the transponder and the specific traffic burst on which the CSM is to dwell can be selected via the computer keyboard. Either after a preset time interval or after the completion of a specific number of measurements, the CSM will revert automatically to the SMM and resume its normal operations from where the SMM was initially interrupted. It is also possible to revert the CSM to SMM at any time from the keyboard by using a designated special function key. Figure 18 is a simplified dwell mode flow chart.

TDMA processor

The following sections outline the architecture and operation of the TDMA processor which was designed and developed to synchronize with the SBS's DA/TDMA network, track any burst within the frame, and provide the real-time control signals required by the individual CSM monitors. The processor performs these functions by using a variety of interface connections, as shown in Figure 19. These interfaces can be classified in three general categories: modem control signals, CSM central processor connections, and real-time gating signals to the monitor subsystems.
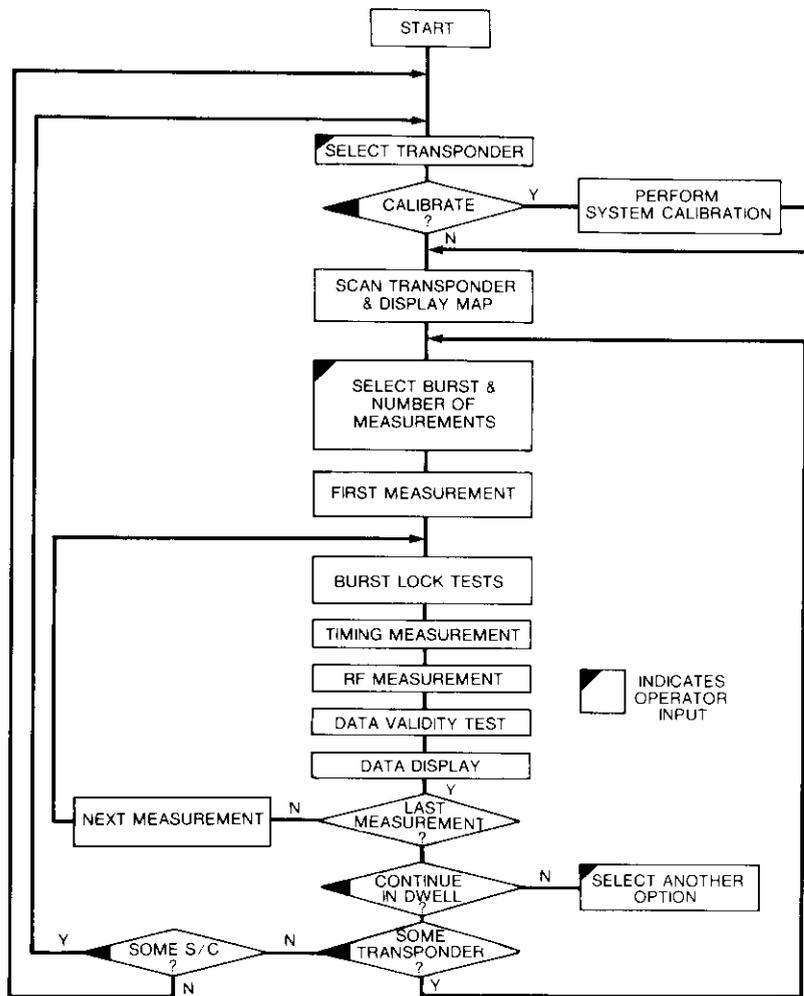
Figure 18. *Simplified Dwell Mode Flow Chart*

The TDMA burst modem must·have a control signal that enables the modem's receive codeword hardware. This real-time gate signal, referred to as the codeword aperture, must be generated for every TDMA burst which is to be received by the modem. These apertures are generated by the TDMA processor as a part of its modem interface hardware. Via this interface, the
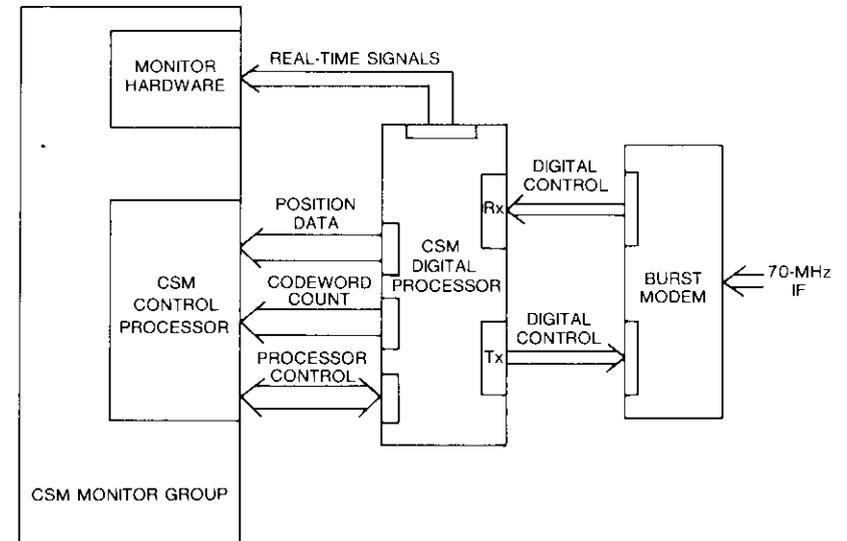
Figure 19. *CSM Digital Processor Interface*

codeword detector provides a detection pulse to the digital processor each time a burst codeword is detected within the modem receive aperture.

The TDMA processor interface and the CSM central processor interface consist of three separate 16-bit parallel data lines: a bi-directional processor control port and two read-only ports. All control signaling between the two processors is performed via the bi-directional 16-bit parallel interface. A 16-bit control word containing the identification information for the selected traffic burst is transmitted to the TDMA processor via this bus. The TDMA processor initiates a burst acquisition process upon receipt of this control word, and returns a response to this measurement request to the central processor via the same interface. If the acquisition is successful, this response includes the selected traffic bursts, the expected position of the next traffic burst relative to FRB, and the number of traffic burst assignments in the FRB. In case of an anomaly, appropriate identification messages are sent to the central processor.

In addition to the routine TDMA synchronization function, the TDMA processor also measures the detected burst position relative to FRB, and the number of burst codeword hits in a frame. These results are transferred to the central processor via the two 16-bit parallel read-only interfaces.

Three real-time control signals (G_1, G_2, and G_3) are generated by the digital processor to synchronize individual CSM monitors to the selected

traffic burst. Traffic burst gate G_1 begins at the expected location of the selected traffic burst and has a nominal duration of 1,280 TDMA symbol intervals (5 channels). The XRB guard space gate, G_2, is a series of six pulses per TDMA frame which are synchronous with the guard time intervals between the FRB and XRB bursts. Gate G_3, which is initiated by the next receive aperture following the selected TDMA traffic burst, is a control signal that is used to reset the burst duration monitor if the guard space following this burst is missing due to an anomaly.

As indicated in Figure 20, the CSM TDMA processor consists of six major functional blocks: a modem interface, a forward error correction (FEC) processor, an assignment processor, a synchronizer, and two gate generators.
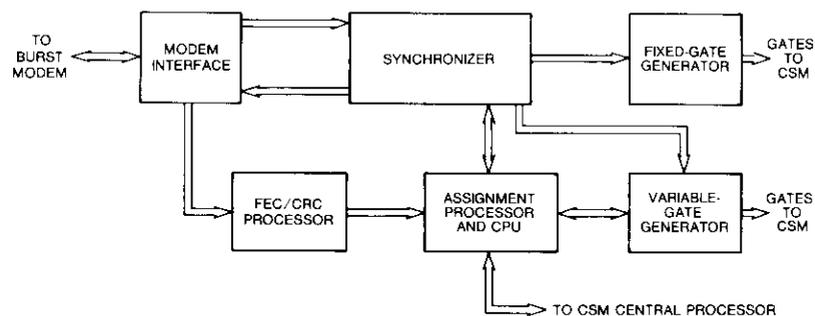
Figure 20. *CSM Digital Processor Block Diagram*

The modem interface has been described previously. The purpose of each of the other functional blocks is as follows:

a. FEC/CRC Processor. The data fields in the FRB and TDMA traffic burst control fields are protected by a quadratic residue code capable of correcting up to two errors per 16-bit block transmitted. The FEC processor performs the necessary operations to recover the original corrected data from the FEC encoded bit stream.

In addition to FEC protection, the FRB assignment data also contain a cyclic redundancy check (CRC) capable of detecting but not correcting data errors. The CRC circuitry contained in the FEC/CRC processor detects these data errors and inhibits the use of incorrect assignment data.

b. Assignment Processor. The assignment processor is a microprocessor-controlled system which processes the decoded FRB assignment information to construct TDMA traffic burst assignment maps. This

processor also controls the synchronization operation of the TDMA processor, while providing a control interface to the CSM central processor.

c. Synchronizer. The TDMA processor synchronizer circuitry contains the hardware necessary to synchronize with the TDMA frame. All frame timing, modem aperture generation, and gating originate in this section.

d. Gate Generators. Two types of control gates are generated by the processor: the fixed gate and the variable gate. The fixed-gate generator provides the CSM monitor equipment with all the real-time gate-control signals that do not change on a frame-to-frame basis. The variable-gate generator provides the CSM measurement equipment with all the real-time gate signals that depend on traffic burst assignments.

CSM data processing

The CSM computer performs all of the monitoring subsystem control, data collection, processing, and storage functions. As a monitoring controller, it performs such control functions as transponder selection and monitoring-instrument interrogation. The raw data accumulated by instrument interrogation are then processed with calibration information. These corrected data are further processed to extract the traffic burst and spacecraft system parameters. For example, to compute the IBO of a traffic burst, the suppressed pilot power is corrected by subtracting the interfering QPSK signal power; small-signal suppression is then computed by using this and the unsuppressed pilot measurement; and finally the IBO is computed by using the algorithm that relates IBO to small-signal suppression.

The computer performs the following data processing functions: computation of daily averages and computation of standard deviations of the monitored spacecraft and earth station parameters. The computer also generates short-term (diurnal) and long-term (monthly and yearly) data profiles, either on demand or when a specific parameter's average and/or standard deviation exceed the preset thresholds.

Data are stored in three categories: satellite-related (spacecraft beacon level and local oscillator frequency), transponder-related (TWTA saturated power output and transponder linear gain), and burst-related (all traffic burst parameters). All of the data are stored as one record for each monitoring cycle of traffic in one satellite. All records associated with the same satellite, transponder, and burst are linked. Burst-related information is retrieved by using this stored linkage information. Data pertinent to other satellites are stored in separate files.

Data storage structure

Each day's data measurements and the statistical analysis results of the previous day (before 00.00 GMT) are stored in six random-access files. Of these, two files store burst-related measurements, two store calibration measurements, and two store the averages and standard deviations of the previous day's data. Of each file pair, one serves as an index file and the other as a data file. The thresholds (for alarm condition generation) and the standard deviation norms (for a hard-copy profile of the previous day's data) are stored in a separate file. Figure 21 illustrates the burst index and data file structure, and Table 1 presents related layouts for the burst data file.
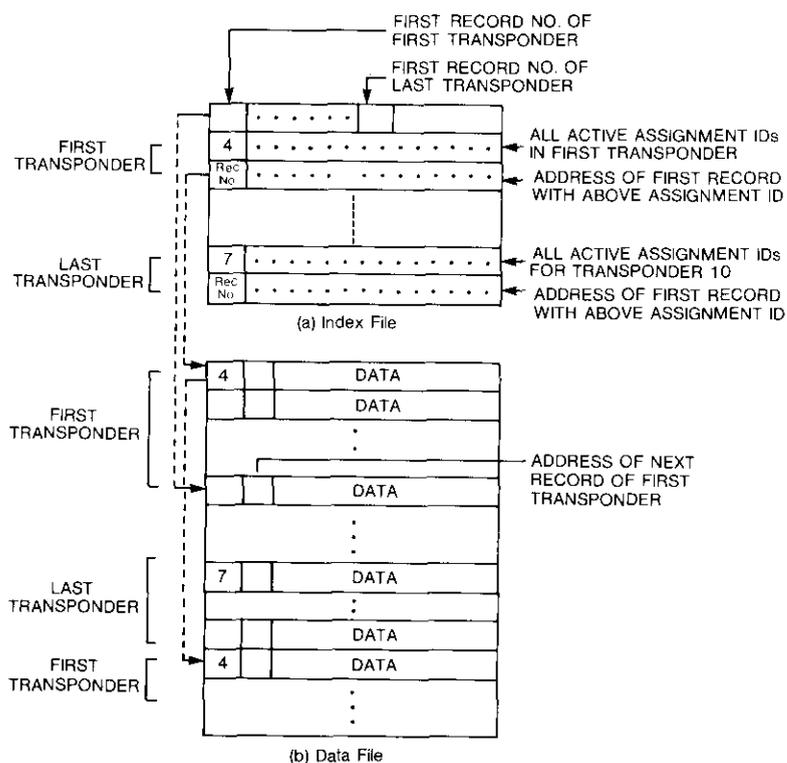
Figure 21. *Burst Index and Data File*

The measurement software can perform data analysis in either a standalone or an interleaved (overlapped) mode. In the interleaved mode, data before

TABLE 1. BURST DATA FILE LAYOUT

FIELD NO.	FILE LAYOUT
1	Assignment number of the burst
2	Address of next record of same transponder
3	Number of expected bursts in frame
4	Number of active bursts in frame
5	Number of inactive bursts
6	Guard space between this and the following burst
7	Corrected burst frequency
8	Corrected beacon level
9	Corrected burst power
10	Burst input backoff
11	Burst output backoff
12	TWTA output power
13	TWTA saturated output power
14	Time of day (hour)
15	Time of day (minute)
16	Time of day (second)

and after 00.00 GMT are considered as "previous day" and "current day" data, respectively. The analysis of previous-day data includes the computation of averages, standard deviations, maxima, and minima for each parameter. These values are stored in "average" files, from which diurnal diagnostic profiles are generated when needed. The decision to generate plots is made by comparing current-day averages and standard deviations with those of the previous day. If the deviations for any parameter exceed a specific preset threshold, a diurnal plot of that parameter is generated.

Some examples of typical data outputs are presented in Figures 22 through 25. Figure 22 shows the profile of a traffic burst. A typical diurnal variation of traffic burst IBO at the satellite transmitter is illustrated in Figure 23. The cyclical variations in IBO are typical of a nontracking transmit antenna. A high peak-to-peak variation would indicate a need for proper transmit antenna pointing.

Figure 24 is a diurnal plot of the spacecraft beacon as measured at CSM, and includes the propagation effects of rain and/or snow from 1400 to 1800 GMT. The gaps between data clusters indicate that CSM was monitoring traffic in another satellite during these periods.

The usefulness of long-term performance profiles in identifying degradation trends is illustrated in Figure 25, which shows a 40-day plot of a traffic burst transmit frequency. For the case shown, the burst transmit frequency drifted by approximately 1 kHz during the 40-day period, and thus required a frequency correction.

```
SATELLITE _ TRANSPONDER _ BURST _ NODE _

TIME OF MEASUREMENT _                                    EXPECTED   MEASURED   DEVIATION


ACTIVE NODES IN TRANSPONDER                              19         18         1

POSITION, FROM FRB (SYMBOLS)                             7,424      7,424      0

DURATION (SYMBOLS)                                       15,992     15,993     1

GUARD SPACE (SYMBOLS)                                               7

RECEIVED CARRIER FREQUENCY (MHz) - DEVIATION (Hz)                   70.003     322.200

SMALL SIGNAL SUPPRESSION (dB)                                      13.38

SPACECRAFT TWTA INPUT BACKOFF (dB)                                 1.45

SPACECRAFT TWTA OUTPUT BACKOFF (dB)                                -0.11

CORRECTED BURST POWER (dBm)                                       -40.56

CORRECTED BEACON LEVEL (dBm)                                      -62.95

TWTA POWER OUTPUT (dBm)                                            41.30

TWTA SATELLITE POWER OUTPUT (dBm)                                  42.85
```

Figure 22. *Typical Burst Profile*

Summary

This paper has discussed the operational features of a performance monitor which are required in order to establish and maintain the high level of overall system performance necessary in a demand-assigned TDMA satellite communications system. Critical earth station and spacecraft parameters have been identified, and techniques have been presented for monitoring their performance. Overall system design and implementation for a fully automated communications system monitor developed to monitor the SBS fully variable demand-assigned TDMA network were also presented. The CSM has been in continuous operation for over 3 years, and has been instrumental in maintaining successful operation of the SBS TDMA network.

Acknowledgments

INPUT BACKOFF SATELLITE TRANSPONDER BURST FOR MONTH __ DAY __


```
AVERAGE MEASUREMENT:        -1.2
NUMBER OF POINTS PLOTTED:    18
MAXIMAL MEASUREMENT:        -0.76
MINIMAL MEASUREMENT:        -1.65
STANDARD DEVIATION:          0.28
PREVIOUS DAY'S AVERAGE:     -0.48
```
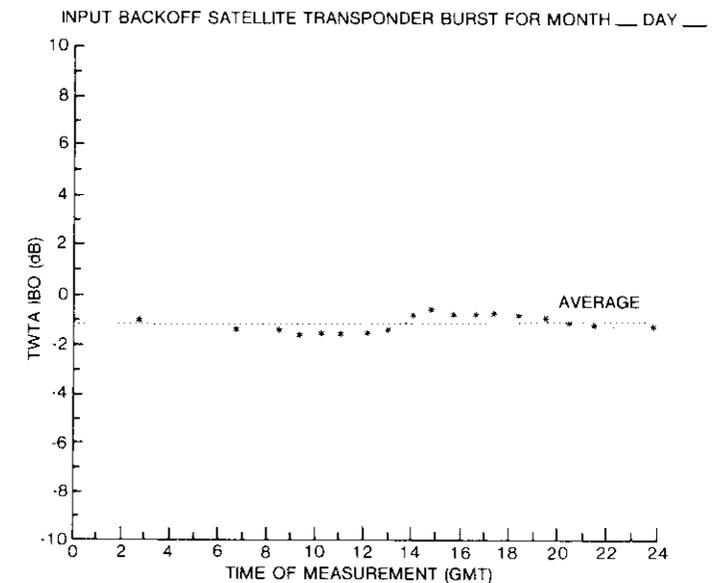
Figure 23. *Diurnal Plot of Traffic Burst Input Backoff*

Figure 24. *Diurnal Beacon Plot*

AVERAGE MEASUREMENT:	-62.2
NUMBER OF POINTS PLOTTED:	1444
MAXIMAL MEASUREMENT:	-61.7
MINIMAL MEASUREMENT:	-65
STANDARD DEVIATION:	0.4

References

[1] E. E. Steinbrecher and L. F. Gray, "A Computer-Controlled Satellite Signal Monitoring System," *COMSAT Technical Review*, Vol. 1, No. 1, Fall 1971, pp. 79–116.

[2] D. J. Schaefer, "Techniques for TDMA System Monitoring," *COMSAT Technical Review*, Vol. 9, No. 2A, Fall 1979, pp. 387–412.

[3] W. H. Curry Jr., "SBS System Evolution," *COMSAT Technical Review*, Vol. 11, No. 2, Fall 1981, pp. 267–291.

[4] P. Bargellini and S. J. Campanella, "An Overview of Satellite Transmission Techniques," *Innovations in Telecommunications*, Part B, New York: Academic Press, 1982.

[5] G. G. Churan and W. E. Leavitt, "Summary of the SBS Satellite Communications Performance Specifications," *COMSAT Technical Review*, Vol. 11, No. 2, Fall 1981, pp. 421–431.
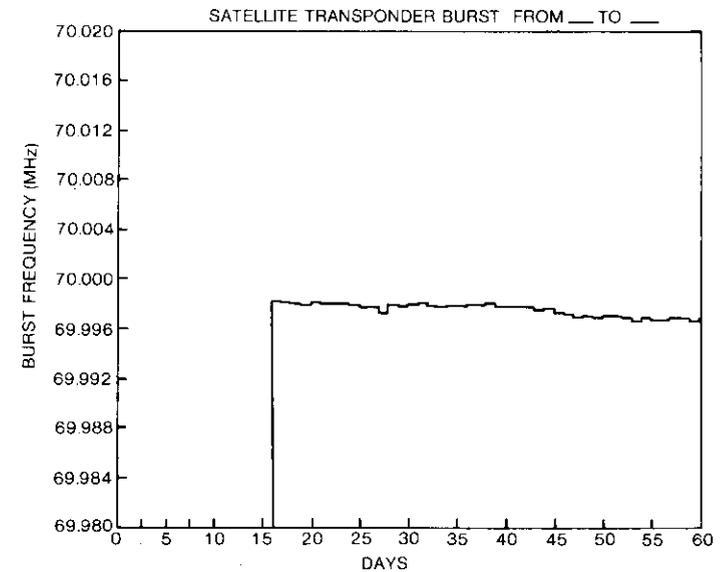
Figure 25. *Long-Term Profile of Traffic Burst Frequency*

[6] I. Dostis, C. Mahle, V. Riginos, and I. Atohoun, "In-Orbit Testing of Communications Satellites," *COMSAT Technical Review*, Vol. 7, No. 1, Spring 1977, pp. 197–226.

[7] C. Emmert, V. Riginos, and J. Potukuchi, "In-Orbit Measurement of the SBS Satellite," Paper 82-0465, AIAA 9th Communications Satellite Systems Conference, San Diego, March 1982.

[8] F. Assal, A. Berman, and J. Potukuchi, "Monitoring of Input Backoff and Carrier-To-Noise Ratio of All Transmissions in a Satellite TDMA Network." Patent Pending.

[9] J. Potukuchi, F. Assal, and R. Mott, "TDMA Performance Monitoring of Communications Satellite Transmitters," National Telecommunications Conference, San Francisco, California, November 1983, *Proceedings*, pp. 410–414.

[10] R. Strauss and J. Owens, "Past and Present INTELSAT TWTA Life Performance," AIAA 8th Communications Satellite System Conference, Orlando, Florida, April 1980, *Proceedings*, pp. 98–106.

[11] J. Potukuchi, F. Assal, and R. Mott, "Communications Satellite System Monitoring for Time-Division Multiple Access," Global Telecommunications Conference (Globecom), San Diego, California, November 1983, *Proceedings*, pp. 1338–1346.

[12] R. Mott, F. Assal, and R. Kroll, "Guard Space and Burst Duration Monitor for TDMA," Satellite Communications Conference, Ottawa, Canada, June 14–17, 1983.

[13] R. Mott, F. Assal, and R. Kroll, "Satellite Receiver TDMA Burst Duration and Guard Space Monitor," Patent 4,470,148.

Janaki R. Potukuchi received his baccalaureate degree in mathematics from Andhra University, Waltair, India, and his B.S., M.S., and Ph.D. degrees in electrical engineering from the Madras Institute of Technology, the Indian Institute of Technology, and North Carolina State University in 1963, 1965, and 1974. From 1966 to 1969, he was a lecturer in the Department of Electronics and Communications Engineering at Osmania University, Hyderbad, India. During 1976, he was a Principal Engineer at Fairchild Industries in Germantown, Maryland, where he was responsible for the design and development of low-noise GaAs FET amplifiers.

Dr. Potukuchi joined COMSAT Laboratories in 1976 as a Member of the Technical Staff. He is presently Manager of the Microwave Systems Department, and is responsible for supervising research and development in the areas of satellite payloads, earth stations, and state-of-the-art monolithic and hybrid microwave integrated circuits.

Francois T. Assal received B.E.E. and M.E.E. degrees from the City University of New York. Presently, he is Associate Director of the Microwave Technology Division at COMSAT Laboratories, where he is responsible for directing research and development of satellite payloads, earth stations, state-of-the-art microwave circuits, and monolithic microwave integrated circuits. Previously, his responsibilities included the development of a 3.5- to 6.5-GHz, 8 × 8 microwave switch matrix, the first TDMA communications system monitor for the SBS operating system, and a 14/11-GHz earth station. He holds patents in communications satellite configurations and hardware.

Richard C. Mott received B.S.E.E. and M.S.E.E. degrees from George Washington University in 1967 and 1973, respectively. After joining COMSAT Laboratories in 1966, he developed microwave integrated circuitry for the DOMSAT simulator, and several different satellite-borne waveguide circuits for the COMSTAR Centimeter Wave Beacon transmitters. He is presently a Staff Scientist in the Microwave Systems Department, where he designs and develops microwave satellite and earth station components and systems. He has managed projects for development of satellite redundancy switches at 4 and 12 GHz, the SBS fixed-transponder communications system monitor of a QPSK/TDMA system, and a 14/11-GHz earth station using 600W brazed-helix TWTAs.

10-W solid-state power amplifier for C-band TWTA replacement

S. M. CHOU, I. G. ATOHOUN, P. B. ROSS,
M. J. URCUYO, AND R. R. JOHNSON

(Manuscript received June 19, 1984)

Abstract

Design and performance data are presented for a solid-state power amplifier (SSPA) that has been developed as a potential replacement for the traveling wave tube amplifier (TWTA). The SSPA weighs 208 g and delivers 10 W of RF power, with gain greater than 65 dB and 35-percent power-added efficiency over the 3.7- to 4.2-GHz satellite communications band. The amplifier consists of a driver, as well as medium- and high-power amplifier sections.

Introduction

For communications via satellite, power amplification is traditionally performed by TWTAs with high-voltage power supplies. Power output ranges from a few to several hundred watts. TWTAs, however, consume most of the DC power available to the communications payload, and are among the heaviest communications components.

With advances in gallium arsenide field-effect transistor (GaAs FET) technology [1]–[4], SSPAs have been developed which offer the following advantages over TWTAs:

a. low voltage requirement,
b. greater reliability and life expectancy,

c. small phase shift and better behaved nonlinearities of both phase-shift and gain transfer characteristics, and

d. significantly lower weight.

The need to reduce overall spacecraft weight while maintaining reliable operating conditions and increasing communications channel capacity make SSPAs strong candidates to replace TWTAs in communications and broadcast satellites.

A C-band SSPA was designed to provide an RF output power of 10 W, with associated gain greater than 60 dB and phase shift of less than 12° over the satellite communications bandwidth of 500 MHz (3.7 to 4.2 GHz). A weight goal of 225 g was also imposed on the design. The SSPA was designed, fabricated, integrated, and tested under both single- and multicarrier operating conditions. It consists of a driver, as well as medium- and high-power amplification sections with a total of six amplification stages. The driver provides an RF gain of 24.5 dB over a 1-GHz bandwidth. The medium- and high-power sections are designed to provide a saturated gain of 40 dB over the entire 3.7- to 4.2-GHz bandwidth.

The overall SSPA weighs 208 g and delivers 10 W of RF output power with a gain greater than 64 dB and 35-percent typical power-added efficiency over the entire satellite communications bandwidth. Under this output power and at 4 GHz, the phase shift is 1°, while the third-order carrier-to-intermodulation ratio (C/I) for a carrier separation of 1 MHz is -16 dBc. At 5 W (which corresponds to a -30 dBm input drive level), the phase shift and C/I are 2° and -28 dBc, respectively. In addition, at a drive level of -28 dBm, the phase shift is 2.4° and decreases monotonically to 0° for the -35 dBm input drive level. The SSPA survives 10-dB (relative to saturation) overdrive operating conditions, under which the maximum phase shift is 3° while the C/I is -11 dBc. Figure 1 is a photograph of the SSPA.

Amplifier design

The SSPA consists of three amplification sections: driver, medium power, and high power. Figure 2 is a block diagram of the SSPA, showing these sections and the devices used in each amplification stage. The figure also indicates the gain value and power level at each stage while the amplifier delivers 10-W RF power. The three amplification sections provide six amplifier stages. To obtain 10-W output RF power, the last stage combines the power from four high-power GaAs FETs.

The input and output matching networks are realized with microstrip transmission lines on 0.635-mm-thick alumina substrates. The substrates are
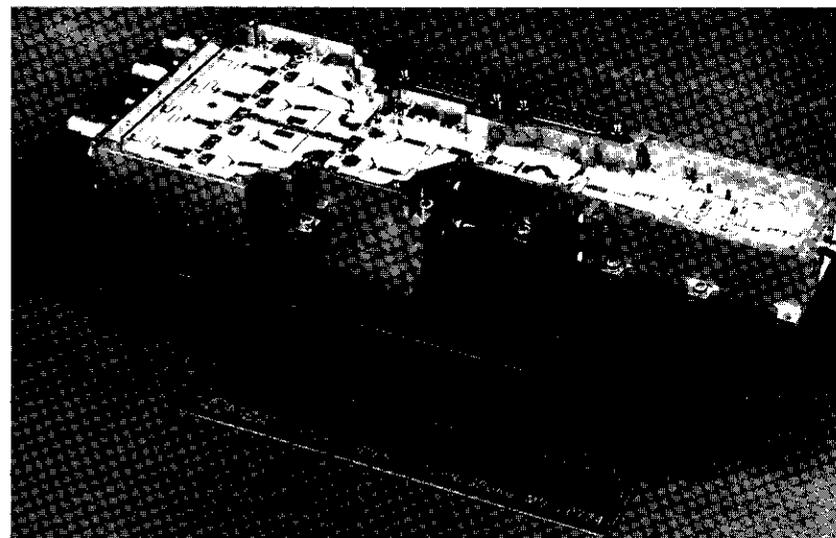
Figure 1. *Photograph of the SSPA*

AMPLIFICATION STAGE	1ST	2ND	3RD	4TH	5TH	6TH
GAIN (dB)	13	12	16.6	9	7.5	6.5
DEVICE	HP1101	HP1101	FSX52WF	FLC15	FLC30	FLC30
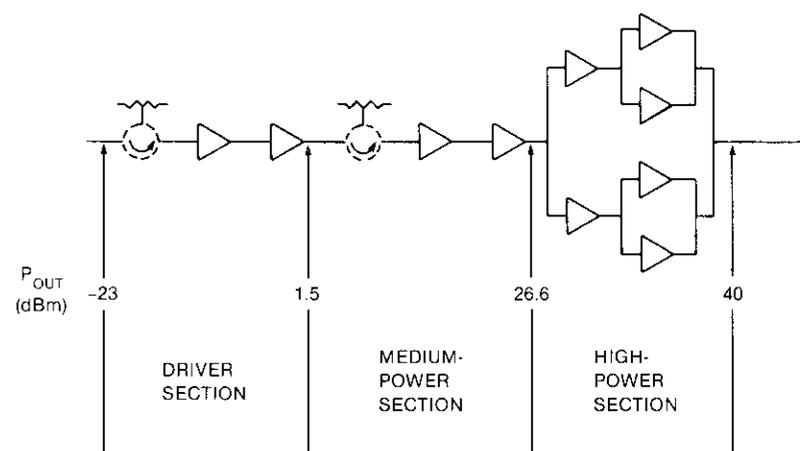
Figure 2. *Block Diagram of the SSPA Amplification Stages*

soldered on Kovar carriers and then secured by several screws on an aluminum housing. The FETS are mounted on the metal ridge of the housing to ensure an adequate heat sink. The amplifier sections were built and tested separately before the sections were integrated. Good impedance matching between sections was ensured by adding isolators or providing quadrature hybrids in the circuit.

Driver

Hewlett-Packard GaAs FETs fabricated with a 1-μm gate (HP 1101) were chosen for the design of a 2-stage amplifier with a minimum RF gain of 22 dB. Since linear amplification was required, the input, interstage, and output networks were designed by using measured S-parameters for linear bias conditions [5] where $V_{DS} = 4.0$ V and $I_{DS} =$ 50-percent I_{DSS} ($I_{DS} \sim 43$ mA). The networks were realized on a 0.635-mm-thick alumina substrate using distributed transmission lines.

These networks had to be designed so that RF gain variations were minimal and uniform over the prescribed frequency range. Since the design required the use of input and output isolators, the input and output return losses had to be reasonable; however, overall circuit stability also had to be addressed. To satisfy these conditions, gain, return losses, and stability factors were computed over small frequency intervals. The computations were performed with SUPER-COMPACT, a computer-aided design analysis program. The length and width of each transmission line were determined by using SUPER-COMPACT's computational capabilities and graphical display. When necessary, the width and/or length of the transmission lines were varied to optimize the frequency responses of the circuit. The designed amplifier RF gain is 26.3 dB over the 3.7- to 4.2-GHz frequency range, and the concomitant variations are less than 0.1 dB.

Medium-power section

The amplifiers in the medium-power section are intended to be operated linearly in order to minimize the overall phase shift and intermodulation while maintaining high efficiency. The output stage of this section uses a Fujitsu FLC 015 medium-power FET which delivers 32.5 dBm of RF power with 8-dB associated gain. The FET selected for the input stage was a Fujitsu FSK 052 that delivers 23.5 dBm maximum RF power with a 15-dB gain at 4 GHz. The 2-stage amplifier utilizes interstage matching techniques to minimize the circuit size. The overall power-added efficiency of this section is approximately 32 percent at a 30-dBm output power level.

High-power section

The high-power amplifier section, which is composed of hybrid-combined stages, uses a total of six FLC 30 power GaAs FETs. The output of the amplifier combines the power of two balanced amplifiers, each driven by a single-ended amplifier. Quadrature Lange couplers are used for power splitting and combining. The insertion loss of the coupler and the 25.4-mm-long 50Ω line (which is necessary for the measurement) is typically 0.15 dB per path. The return losses of the input and output ports of the coupler are greater than 21 dB within a 500-MHz frequency band. The power-added efficiency of this section is 36 percent with a 13-dB gain. The design and performance of the single-ended and balanced amplifier that is used as the basic building block of this section have been reported earlier [6], [7].

Before fabrication, all of the FETs were tested using a standard test jig, and then paired. The RF power, efficiency, and linearity were optimized in two steps: first, the inputs of the FETs were matched for maximum flat small-signal gain; then, the outputs of the FETs were matched for large-signal linearity and maximum power. Two of the high-power sections have been built, and their performances are essentially identical. The FETs can be replaced and interchanged easily without altering the RF performance of the amplifier.

Measurements

Transfer characteristic and DC power

Figure 3 depicts the input RF power vs output RF power relationship. This relationship was measured at three different frequencies, as shown in the figure. During the measurement, bias currents and voltages were monitored at each amplification stage to determine the overall DC power consumption. The DC power consumed by the SSPA is also illustrated. A maximum DC power consumption of about 35 W occurs at the "quasisaturation" point and decreases monotonically under overdrive operating conditions. The DC power consumed in the linear region of the transfer characteristic is almost constant at about 32.5 W.

Gain vs frequency

Figure 4 illustrates the swept gain responses of the SSPA for six different input drive levels. As indicated, the quasisaturated gain (which results from the input RF power of -25 dBm) is greater than 64 dB. The figure also shows that the overdrive conditions affect the RF gain of the SSPA, resulting in a decrease in overall gain.
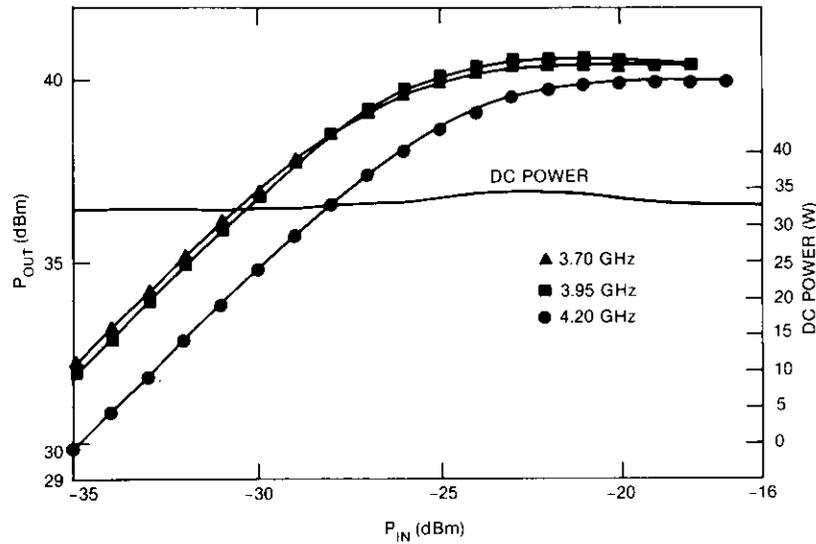
Figure 3. *Transfer Characteristics and Consumed DC Power*
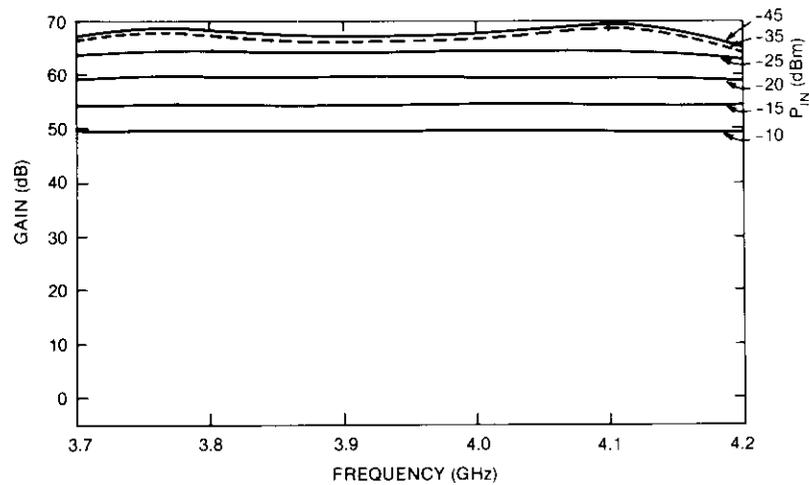
Figure 4. *Swept RF Gain Responses*

Phase shift vs RF drive level

The single-carrier phase shift of the SSPA was measured under CW operating conditions at 4 GHz. As shown in Figure 5, the maximum phase shift of the
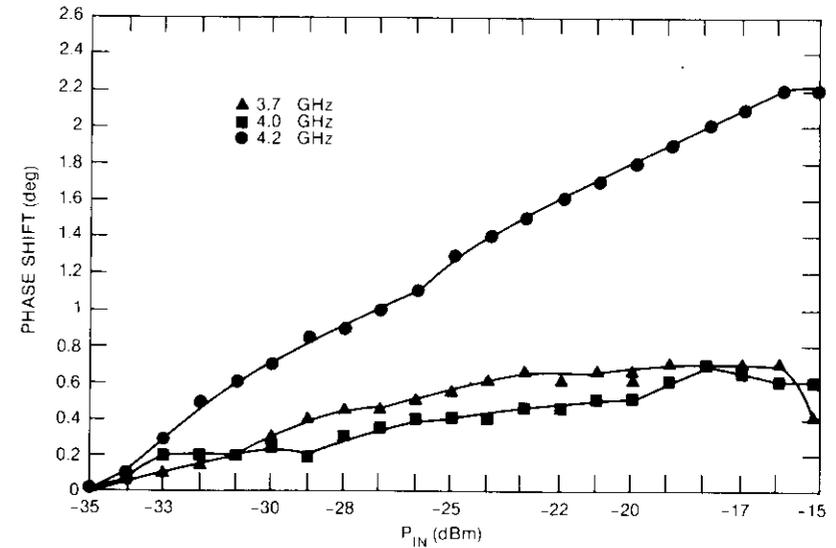
Figure 5. *Phase Shift of the Driver Section* vs *Frequency*

driver is 2.2° for RF drive-level variations of −35 to −15 dBm, and it occurs at 4.2 GHz; while at 4 GHz, the maximum phase shift is 0.7° over the same RF drive level. Figure 6 gives the maximum phase shift of the combined medium/high-power amplifier section, about 3° at 4 GHz.
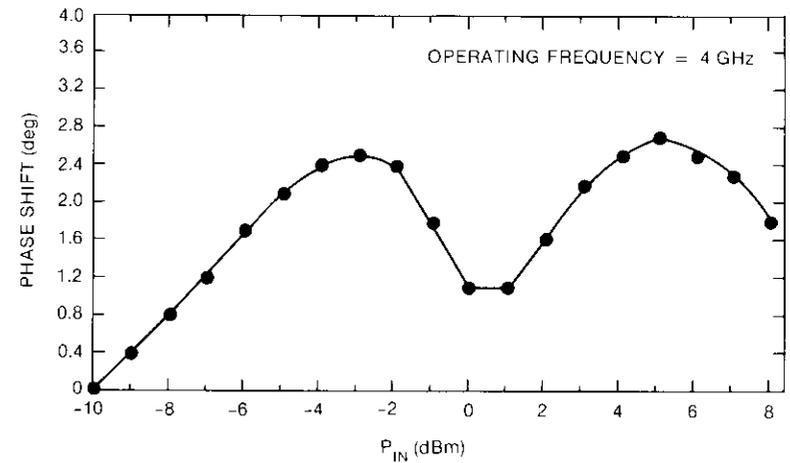
Figure 6. *Phase Shift of the Medium/High-Power Stages Combination* vs *Frequency*

The phase shift of the entire SSPA is plotted in Figure 7. The maximum phase shift contributed by the SSPA is less than 6° under overdrive conditions. The results of phase shift measurements suggest that the contribution of the driver to the SSPA phase shift is negligible when the driver phase shift is greater than zero and less than 0.5°. Otherwise, this contribution is additive; that is, the SSPA phase shift is the sum of the driver phase shift and the medium/high-power section phase shift. When the driver phase shift is less than 0.5°, the medium/high-power section alone defines the SSPA phase shift.
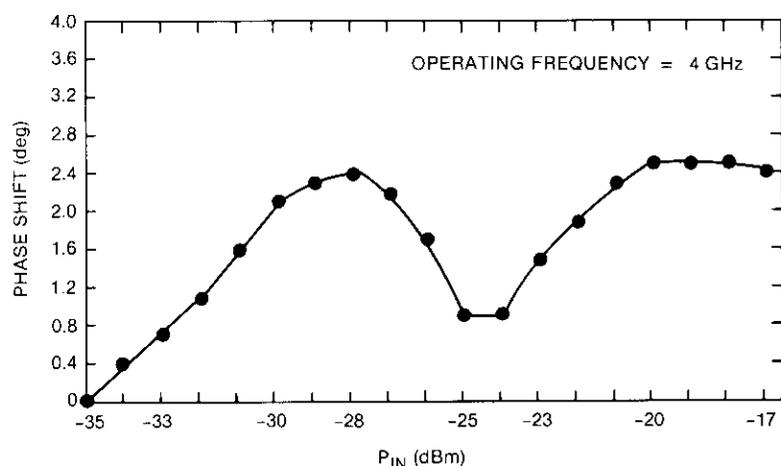
Figure 7. *Overall SSPA Phase Shift* vs *Frequency*

Carrier-to-intermodulation ratio

The C/I was measured for carrier separations of 1 and 5 MHz, respectively. For those measurements, the RF signal generated by two equal carriers set either 1 or 5 MHz apart was applied (after proper system calibration) to the input of the SSPA. The resulting third-order and fifth-order intermodulation products were measured at the SSPA output by setting the spectrum analyzer for a linear display with a scan width that permitted simultaneous viewing of the intermodulation products and an injected reference signal. At the RF drive level of interest, the power level of the injected signal was first set with a precision attenuator to equal that of the carrier, and was then adjusted to equal the power level of the intermodulation product. The difference between the attenuator readings corresponds to the C/I value at the given RF drive level. Figures 8 and 9 plot typical values of C/I measured at 4 GHz.
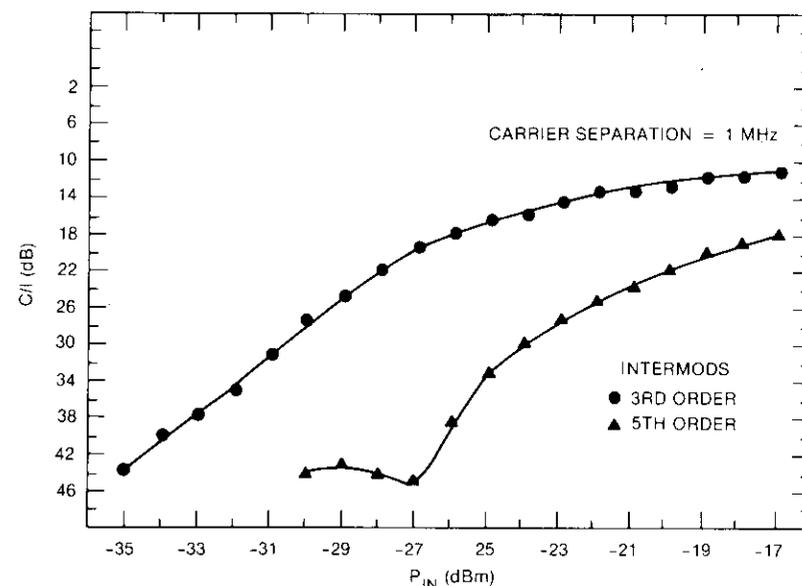
Figure 8. *C/I* vs *RF Drive Level for 1-MHz Carrier Separation*
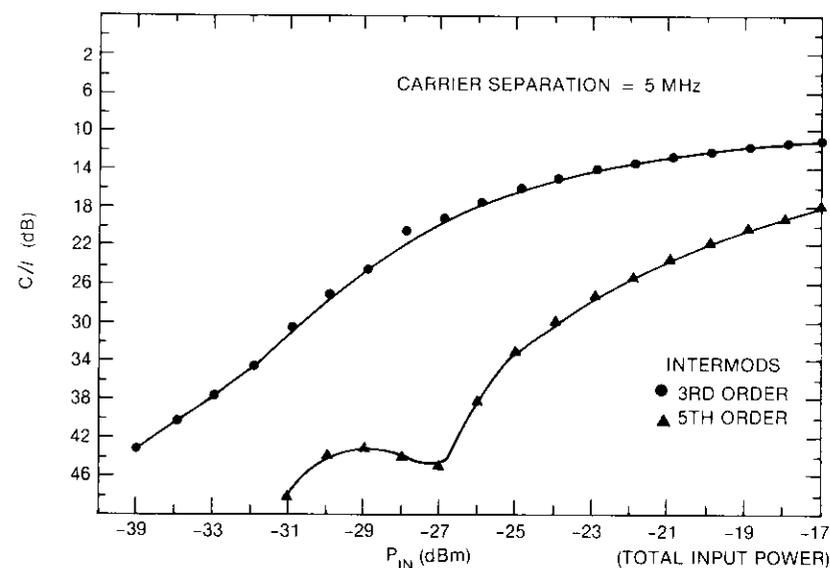
Figure 9. *C/I* vs *RF Drive Level for 5-MHz Carrier Separation*

Performance under single- and multicarrier conditions

The behavior of the SSPA under multicarrier conditions was investigated by measuring its output power at 4 GHz under a 15-MHz injected white noise. The noise was generated by two cascade bipolar amplifiers, with the first amplifier input terminated into a 50-Ω load. The average output power was measured, and power-added efficiencies were computed. Figure 10 depicts the performance difference between single- and multicarrier conditions. For the multicarrier case, the output level decreases by 1.5 dB at the 10-W output power point for CW operating conditions. Reduction in efficiency is also shown.
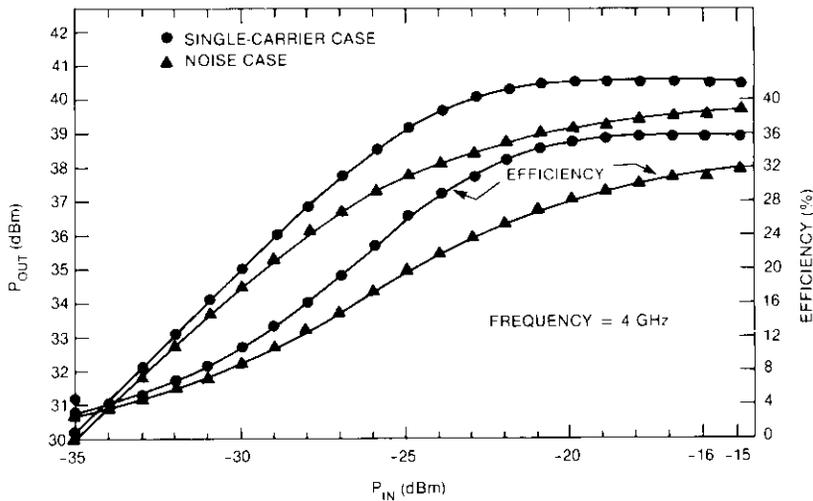
Figure 10. *Transfer Characteristics and Efficiencies of the SSPA Under Single-Carrier and Injected Noise*

Linearity vs efficiency

For communications applications, the SSPA will be required to amplify multiple RF signals in a given bandwidth. To relate the intermodulation distortion effects to efficiency, a curve representing 2-tone C/I *vs* power-added efficiency was plotted for the SSPA and compared with a single-collector TWTA. Figure 11 illustrates this relationship.
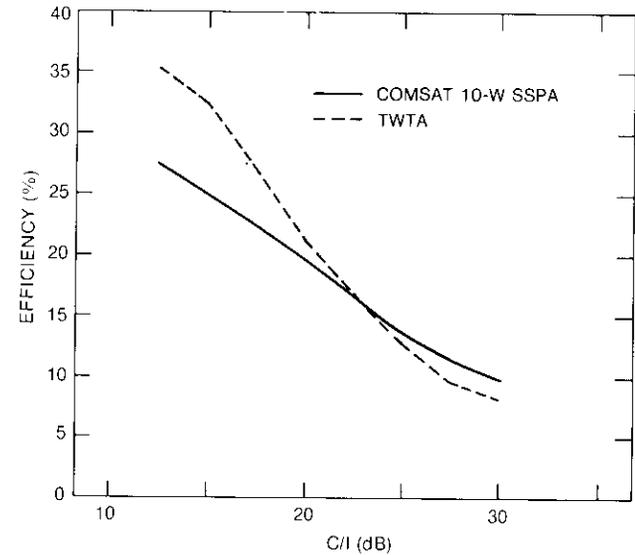
Figure 11. *2-Tone Power-Added Efficiency vs Linearity*

Conclusion

A high-efficiency, high-power SSPA has been demonstrated. This amplifier delivers 10 W of RF output power, 64-dB gain, and 35-percent power-added efficiency. The third-order intermodulation distortion at the SSPA at 10 W output power level is −16 dBc with two equal-amplitude carriers 1 MHz apart. Table 1 compares the general performance of the SSPA with that of an equal-power-level, state-of-the-art, 3-collector TWTA. The table also compares the COMSAT SSPA with another C-band SSPA. With all its advantages in linearity, weight, size, reliability, and overdrive limiting characteristics, the COMSAT C-band SSPA is a strong candidate to replace the TWTA.

TABLE 1. COMPARISON OF POWER AMPLIFICATION

PARAMETERS	HUGHES 246H 3-COLLECTOR TWTA	COMSAT SSPA	C-BAND SSPA [8]
Frequency Band (GHz)	3.7-4.2	3.7–4.2	3.7–4.2
Maximum Output Power (W)	11.5	11.5	8.5
Small Signal Gain (dB)	65.2	65	60
Phase Shift (deg)	32.5 (at saturation)	<3 (overall)	22

TABLE 1. (Continued) COMPARISON OF POWER AMPLIFICATION

PARAMETERS	HUGHES 246H 3-COLLECTOR TWTA	COMSAT SSPA	C-BAND SSPA [8]
Linearity (C/I$_3$)[b]			
0 dB	12	14	13
−3 dB	15.6	17	
−6 dB	18.4	22	
Mass (Without EPC) (g)	820	208	368.5
Size (cm)	35.56 × 5.08 × 5.08	22.86 × 7.62 × 2.54	
Power-Added Efficiency (%)	44	35	27 (with DC supply)

[a] SSPA saturation = output power. Increases by 0.25 dB when input power increases by 1 dB.
[b] Input power of the two carriers relative to single-carrier power required for maximum output power.

References

[1] L. S. Napoli et al., "High Power GaAs FET Amplifier—A Multiple Structure," International Solid-State Circuit Conference, 1973, *Digest of Technical Papers*, pp. 82–83.

[2] F. Sechi, H. Huang, and V. Riginos, "High Efficiency MESFET Linear Amplifier Operations at 4 GHz," International Solid-State Circuit Conference, Philadelphia, 1977, *Digest of Technical Papers*, pp. 164–165.

[3] P. T. Ho, C. M. Pham, and R. L. Mencik, "A 10-Watt C-Band FET Amplifier for TWTA Replacement," IEEE MTT-S International Microwave Symposium, Orlando, Florida, April 1979, *Symposium Digest*, pp. 128–130.

[4] Y. Arai, S. Murai, and T. Sakane, "High Power GaAs FET Amplifier for TWT Replacement," *Fujitsu Scientific and Technical Journal*, Vol. 15, No. 3, September 1979, pp. 63–82.

[5] Hewlett-Packard Co., "Diode and Transistor Designer's Catalog, 1982–83," p. 4.

[6] S. Chou, C. Chang, and F. Assal, "High Efficiency Broadband FET Power Amplifier for C-Band TWTA Replacement," IEEE International Conference on Communications, Philadelphia, 1982, *Conference Record*, pp. 1E.2.1–1E.2.4.

[7] S. Chou and C. Chang, "4-GHz High-Efficiency Broadband FET Power Amplifiers," *COMSAT Technical Review*, Vol. 12, No. 2, Fall 1982, pp. 399–411.

[8] D. Aubert et al., "Power Amps Replace TWTAs in Satellite Transponders," *Microwave System News*, April 1983, pp. 62–76.

Su Min Chou received a B.S.E.E. from Chinese Naval College of Technology in 1954, an M.S. from Chiao Tung University in 1961, and a Ph.D. from University of Utah in 1967. He joined COMSAT Laboratories in 1968 and is presently a Senior Staff Scientist in the Microwave Technology Division. He has been involved in the design of a low-noise parametric amplifier, a high-power IMPATT amplifier, a high-power bipolar transistor amplifier, and a high-power FET amplifier at various frequency bands for both earth station and satellite applications.

Ignace G. Atohoun received a B.S.E.E. from Colorado State University; an M.S.E.E. from Université Laval, Québec, Canada; and a Ph.D. in electrophysics from Stevens Institute of Technology, Hoboken, New Jersey. From January 1968 to September 1969 he was a Research and Teaching Assistant at the Université Laval and worked on semiconductor integrated antennas. He joined the Stevens Institute of Technology in September 1970 as an Assistant Instructor.

As Member of the Technical Staff of the Transponders Department of the Microwave Technology Division at COMSAT Laboratories since August 1974, Dr. Atohoun has been engaged in the modeling of Gunn-effect devices and the communications characterization of Gunn-effect amplifiers and multicollector traveling wave tube amplifiers operating in the 10- to 15-GHz band. In addition, he has participated in the in-orbit testing of various communications satellites and has developed a third-order phase-locked loop for in-orbit test equipment and satellite television receivers, as well as 12- to 13-GHz low-level phase detectors and 11-GHz hybrid microwave integrated amplifier circuits on GaAs. He is currently engaged in the design and development of an 11-GHz amplifier using monolithic microwave integrated circuits, with emphasis on active matching techniques. He is a member of Sigma Xi, IEEE, and AAAS.

Patricia B. Ross received a B.S.E.E. from Purdue University in 1950. She joined COMSAT in 1975 working for the Maintenance and Supply Center in the positions of Technical Coordinator, Digital Engineer, and Staff Engineer. Her principal area of responsibility was SCPC and SPADE equipment. In 1979, she became a Member of the Technical Staff of the Microwave Systems Department, and is currently an Associate Engineer in the Microwave Circuits Department. Projects she has contributed to are the RFSM beacon level monitor, MMIC element development, power amplifier design and testing, and MMIC mask design. She is a member of IEEE.

Marco J. Urcuyo is a Senior Technician in the Microwave Technology Division at COMSAT Laboratories. Since joining COMSAT in July 1975, he has worked on the development of the microwave switch matrix, as well as the communications characterization of various microwave components. Currently, he is involved in the development of an L-band solid-state power amplifier, a C-band solid-state 10-W amplifier, and a 12-GHz medium-power amplifier. He is enrolled at the Northern Virginia Community College, where he is pursuing an Associate Degree in Electronics Technology.

Ronald R. Johnson is an Assistant Staff Member in the Transponders Department of the Microwave Technology Division at COMSAT Laboratories. After serving 4 years in the Air Force as a Microwave Technician, he worked in the Special Projects Lab at Radiation Systems, Inc., where he was engaged in research and development of antennas and feed systems for 5 years. Since joining COMSAT in 1974, he has been involved with waveguide filter design and in-orbit test systems installation and integration. He is currently working in MIC and MMIC design and testing.

Reducing the effects of structural modes on the performance of attitude control systems

E. ELATI*

(Manuscript received November 30, 1983)

Abstract

The trend toward high-performance satellites with complex mechanical structures has led to structures with natural frequencies close to the frequency band of the attitude control loops. This closeness can cause instability of the control loop. This paper analyzes the use of digital notch filters to cancel the effect of structural modes on attitude control system performance. Tradeoffs in the design of such a cancellation system, and the influence of various parameters of notch filter implementation on control system performance, are discussed. The case in which the structural frequencies are not accurately known or vary during the satellite mission is also examined. Finally, experimental results are presented for a control system that includes a structural mode whose negative effect on the stability of the control loop is cancelled by a digital notch filter.

Introduction

Figure 1 is a simplified block diagram of a single-input/single-output control loop. The transfer function $P(s)$ can be approximated by

$$P(s) = \frac{1}{Is^2} \sum_{i=1}^{N} \frac{(s^2 + 2\xi_i\omega_i s + \omega_i^2)}{(k_i s^2 + 2\xi_i\omega_i s + \omega_i^2)} \tag{1}$$

where I = moment of inertia

ω_i = frequency of the i-th structural mode

k_i = coefficient whose value (larger or smaller than 1) depends on the structure of the flexible body and the location of the loop sensor relative to its actuator

ξ_i = damping ratio of the i-th structural mode

N = total number of structural modes considered.
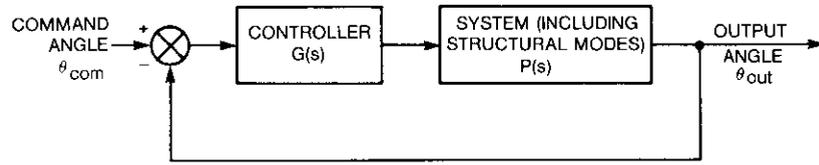
Figure 1. *Simplified Block Diagram of a Control Loop*

In the case where $k_i > 1$ (pole before zero), the control loop could be unstable. This effect is illustrated in the Bode diagrams of the open-loop transfer function, $P(j\omega)$ (Figure 2), which show that, at a pole ω_n, the amplitude of the open-loop transfer function is greater than 1, while the phase lag is more than $180°$. The system is therefore unstable. Cancellation of this effect is more difficult when the structural frequencies are not accurately known or vary with time.

This paper describes two major aspects of the proposed cancellation system: a learning/adapting system for identification of the structural modes, and a scheme for cancelling the negative effect of structural modes. This scheme is a chain of digital notch filters.

Finally, results of experiments performed on the single-axis flexible spacecraft simulator are presented. These results support theoretical conclusions that were reached prior to the tests.

Methods for eliminating the influence of a structural mode

Three different approaches were considered in selecting the cancellation system: a variable structure system, a phase-locked loop, and a chain of adaptive notch filters.

Variable structure system

In its simplest application, this technique is based on switching the control input to the system between two distinct values when a switching plane is
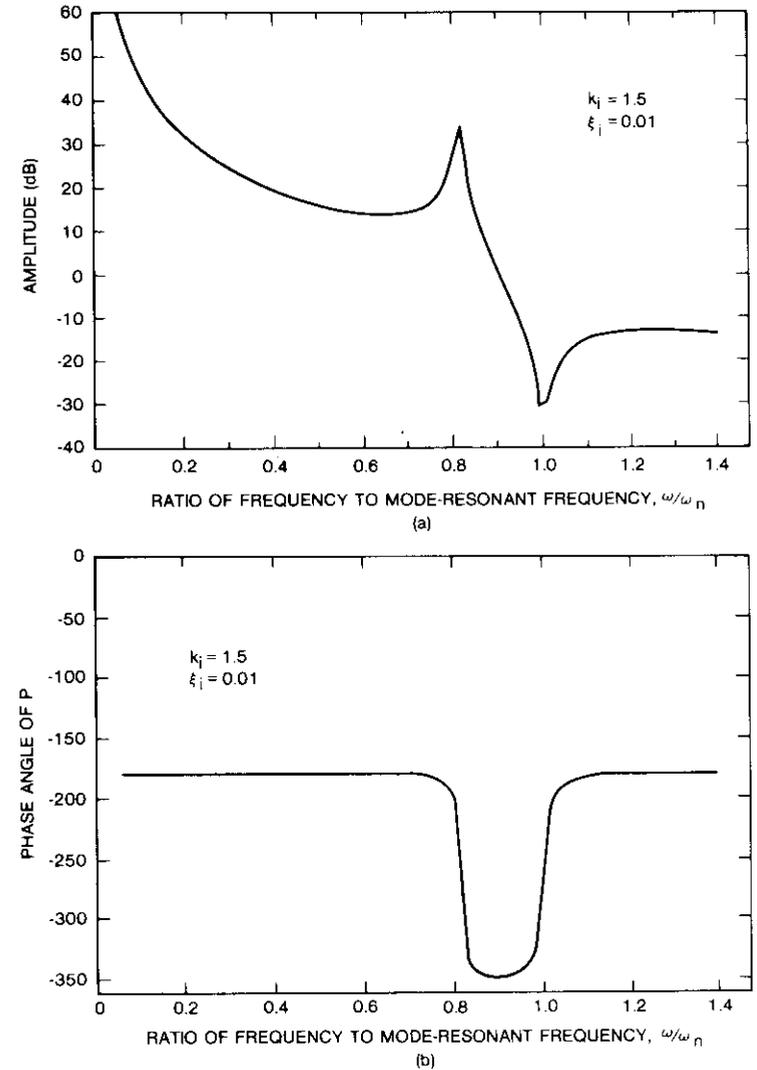
Figure 2. *Open Loop Transfer Function of a System Containing a Structural Mode*

crossed, and then converging to zero error along the plane [1]. This method can handle some uncertainty, but would be unacceptable for spacecraft control because of the large number of states in the system. Identification of these states would require many differentiations, which would result in an extremely noisy system.

Phase-locked loop

The elimination of structural mode influence has been given much consideration in recent years [2],[3]. Most of the literature deals with the case where the modal frequencies are known and a controller is to be designed. A phase-locked loop has been suggested to overcome the problem of varying frequencies; however, this solution seems inadequate because of the very low frequencies to be detected.

Chain of adaptive notch filters

The transfer function of a notch filter has the form

$$NF(s) = \frac{1 + \dfrac{2\xi_z s}{\omega_n} + \dfrac{s^2}{\omega_n^2}}{1 + \dfrac{2\xi_p s}{\omega_n} + \dfrac{s^2}{\omega_n^2}} \tag{2}$$

where ξ_z = zero damping ratio
ξ_p = damping ratio of the pole
ω_n = frequency of the notch filter.

Figure 3 shows the frequency response of the notch filter. As illustrated, the notch filter can attenuate the amplitude-frequency response at a distinct frequency without affecting the rest of the spectrum. Notch filters at the structural frequencies would thus be useful in eliminating the effect of these frequencies. Detection of the unknown frequencies will be performed separately. This is the method that was chosen for analysis.

General description of the cancellation system

Figure 4 is a block diagram of the cancellation system. This system operates in two distinct phases, the learning phase and the operating phase.

Learning phase

During this phase, the structural frequencies of the system are identified. All the switches are set to 1, and the learning proceeds as follows:

a. Gain K_ℓ is set so that the output signal is maximum but does not saturate the learning network.
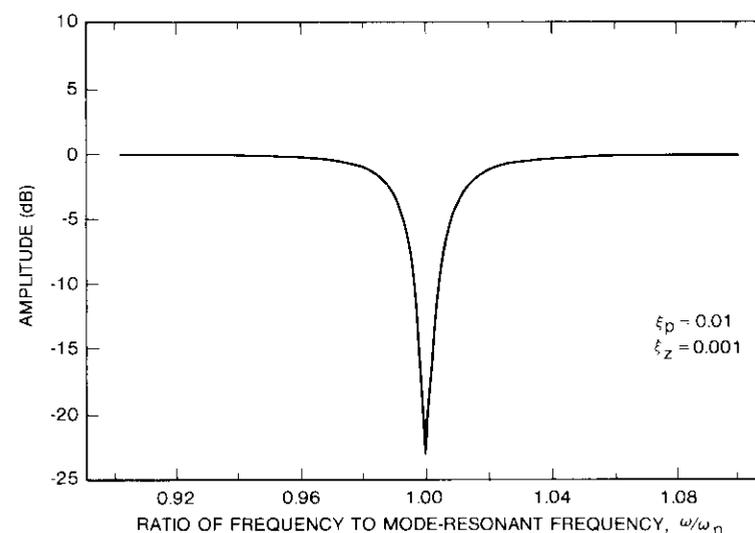
Figure 3. *Frequency Response of a Notch Filter*

b. The system is excited by white noise produced by the processor, and the outputs of the sensors are monitored at a sampling period T_1.

c. According to the learning method chosen, the learning network detects the modal frequencies. The processor then switches to 2.
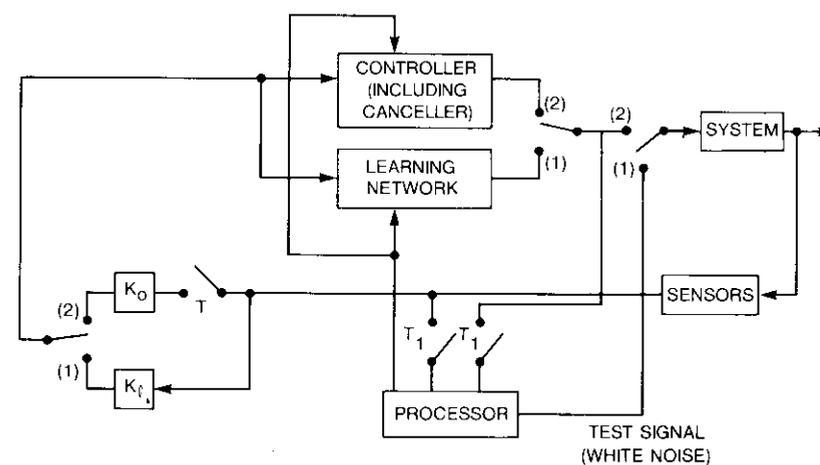
Figure 4. *System for Cancelling the Effects of Structural Modes*

Operating phase

In this phase, the control loop is closed with a controller designed to achieve the required performance, including the canceller (a chain of digital notch filters). The sensor outputs are sampled at a period T. Also during this phase, the structural frequencies may change their values. If this occurs, an adjusting procedure should be initiated, as discussed later.

The notch filter

The transfer function of the notch filter was presented in equation (2). The primary goal in designing the notch filter is to have sufficient attenuation at the frequencies of the structural modes to cancel the amplification caused by the modes without affecting the overall performance of the control system, whose bandwidth is only slightly lower than that of the structural frequencies.

The performance of the filter depends on the damping ratios of its zeros (ξ_z) and poles (ξ_p), assuming that the notch frequency, f_n, equals (or is very close to) the structural frequency. These parameters affect the characteristics of the notch filter, as follows:

a. Gain at notch frequency

$$\left| NF(j\omega) \right|_{\omega = \omega_0} = \frac{\xi_z}{\xi_p} \quad . \tag{3}$$

b. Phase lag (for $\xi_z \ll \xi_p$, which is generally the case)

$$\phi \cong -2\xi_p \frac{f}{f_n} [1 - |f/f_n|^2]^{-1} \quad . \tag{4}$$

c. Bandwidth

$$BW \cong 2\xi_z f_n \quad . \tag{5}$$

The term "bandwidth" is used here to mean the difference between the frequencies at which the gain is 3 dB greater than the gain at the notch frequency. Equation (5) is valid for $2\xi_z^2 \ll \xi_p^2$ and $\xi_z^2 \ll 1$. These conditions are generally satisfied in the present context.

As can be seen, for a fixed attenuation of the notch filter, there is a conflict between the need for low phase lag (small ξ_p) to protect control system stability, and high bandwidth (higher ξ_z, and therefore higher ξ_p). Later, it will be shown that ξ_z and ξ_p also affect parameters in the digital system.

Digitization of the notch filter

Since it is necessary to vary the notch frequencies during the learning or adjusting phases, the notch filters must be implemented digitally on the control system computer. The desired digital transfer function is achieved by using the bilinear transformation with prewarping [4]

$$s \rightarrow \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \tag{6}$$

$$\omega_n \rightarrow \frac{2}{T} \tan \frac{\omega_n T}{2} \tag{7}$$

where z^{-1} is the unit delay operator, and T is the delay time.

The digital transfer function is

$$NF(z) = \frac{a + bz^{-1} + cz^{-2}}{d + bz^{-1} + ez^{-2}} \tag{8}$$

where

$$a = 0.5F^2 + \xi_z F + 0.5$$

$$b = F^2 - 1$$

$$c = 0.5F^2 - \xi_z F + 0.5$$

$$d = 0.5F^2 + \xi_p F + 0.5$$

$$e = 0.5F^2 - \xi_p F + 0.5$$

$$F = \tan \frac{\omega_n T}{2}$$

For $f_n \cdot T < 0.25$ and an infinite word length, there is good agreement between the transfer functions (phase and amplitude response) of the analog and digital notch filters.

Notch filter implementation

The canceller is composed of a cascade of notch filters, each corresponding to a structural frequency with ξ_z and ξ_p, as required. This method was selected over methods such as direct or parallel implementations [5] which might be more accurate, because the values of the structural frequencies are not constant, and therefore the coefficients of the canceller must be updated with time. The coefficients of the direct or parallel form are the results of a large

number of finite-length word calculations, and therefore accumulate a large error; while the six coefficients of the notch filter cascade form result from the straightforward calculations given in equation (8).

A single notch filter can be implemented in either the direct or canonical form. The direct form is represented by the difference equation

$$Y(n) = \frac{a}{d}X(n) + \frac{b}{d}X(n-1) + \frac{c}{d}X(n-2)$$

$$- \frac{b}{d}Y(n-1) - \frac{e}{d}Y(n-2) \ . \tag{9}$$

where $Y(n)$ is the output signal, and $X(n)$ is the input. This implementation uses five multiplications and four memory locations $[X(n-1), X(n-2), Y(n-1),$ and $Y(n-2)]$.

The canonical implementation is represented by the following difference equations:

$$V(n) = \frac{a}{d}X(n) - \frac{b}{d}V(n-1) - \frac{e}{d}V(n-2) \tag{10}$$

$$Y(n) = V(n) + \frac{b}{a}V(n-1) + \frac{c}{a}V(n-2) \ . \tag{11}$$

Here also, five multiplications are needed, but only two computer memory locations $[V(n-1)$ and $V(n-2)]$ are required for each notch filter. This method uses an additional variable, $V(n)$.

The transfer function between V and X is

$$\frac{V(z)}{X(z)} = \frac{a}{d + bz^{-1} + ez^{-2}} \tag{12}$$

$$a \approx 0.5 \ .$$

The delay operator z^{-1} can be represented as

$$z^{-1} = e^{-j\omega T} = \cos \omega T - j \sin \omega T \ . \tag{13}$$

Using the expressions for d, b, and e, together with the trigonometric relations

$$\tan \frac{U}{2} = \frac{1 - \cos U}{\sin U} = \frac{\sin U}{1 + \cos U} \tag{14}$$

$$\cos 2U = 2 \cos^2 U - 1 = 1 - 2 \sin^2 U \tag{15}$$

$$\sin 2U = 2 \sin U \cos U \tag{16}$$

yields

$$\left. \left| \frac{V}{X} \right| \right|_{\omega = \omega_n} = \frac{a}{2\xi_p (1 - \cos \omega_n T)} \tag{17}$$

$$\left. \left| \frac{V}{X} \right| \right|_{\omega = 0} = \frac{a}{2 \tan^2 (\omega_n T/2)} \ . \tag{18}$$

Thus, low ξ_p and $\omega_n T \ll 1$ require a high amplification of X to obtain V. This forces attenuation of the signal before it is input to the notch filter, which is compensated for by amplifying the output of the notch filter. This problem, which does not occur with the direct method, might be an important factor in choosing the implementation method for the notch filter. It is also clear that the higher the value of ξ_p and of the product $\omega_n T$, the less attenuation and compensation are needed.

Gain sensitivity to design parameters

The most important characteristic of the notch filter is its gain, not only at the notch frequency, but also across its 3-dB bandwidth. Adequate attenuation over a finite interval centered at the notch frequency is required to account for errors in determining the structural frequencies.

The expression for z^{-1} from equation (13) is used to obtain the gain function

$$\left| \frac{Y}{X} \right|^2 = \left| \frac{a + b \cos \omega T + c \cos 2\omega T + j(b \sin \omega T + c \sin 2\omega T)}{d + b \cos \omega T + e \cos 2\omega T + j(b \sin \omega T + e \sin 2\omega T)} \right|^2 \ . \tag{19}$$

The gain $\left| \frac{Y}{X} \right|$ is a function of a, b, c, d and e, which in turn are functions of the design parameters ξ_p, ξ_z, ω_n, and T [equation (8)]. The sensitivity of gain about a nominal set of design parameters can be measured by the total error, which is given by the expression [6]

$$\sigma^2[f(z)] = \sigma^2 + \sum_i \left| \frac{\partial f(z)}{\partial z_i} \right|^2 \sigma_{z_i}^2 \tag{20}$$

where $f = \left| \frac{Y}{X} \right|$

$\sigma^2 =$ variance for quantization error due to finite wordlength [4], [5]

$\sigma_{z_i}^2 =$ variance of the parameter z_i being varied to determine sensitivity

and

$$\sigma^2 = \frac{(LSB)^2}{12} \tag{21}$$

where LSB is the quantization step.

The influence of the following parameters on gain sensitivity was measured: sampling rate, tangent memory size (i.e., resolution for tan $\omega_n T/2$), word length, and effect of damping ratios. The effect of these parameters was evaluated by calculating the sensitivity for various notch filters, and the following observations were made:

a. Sensitivity is smaller for higher values of the product $f_n \cdot T$ (optimal value $0.1 < f_n \cdot T < 0.2$).

b. Accurate prewarping (good resolution for the tangent function of $\omega_n T/2$) is essential.

c. Word length has a direct influence on sensitivity, which increases for shorter words (by a factor of 2 per bit).

d. The zero damping ratio, ξ_z, of the notch filter directly influences gain sensitivity, while the pole damping ratio has little effect. The smaller the value of ξ_z is, the higher the gain sensitivity.

Based on these observations, it is recommended that the digital notch filter parameters be selected by choosing a sampling period in the range of $0.1 < f_n T < 0.2$, using the largest possible word size, and using the largest possible notch filter bandwidth (large ξ_z and therefore ξ_p).

Effects of quantization noise

This subsection analyzes the effect of the structure and parameters of the notch filter on the noise that results from the finite length of the digital word. Figure 5 shows the noise, θ_{out}, produced by a digital notch filter as an input to the control system.
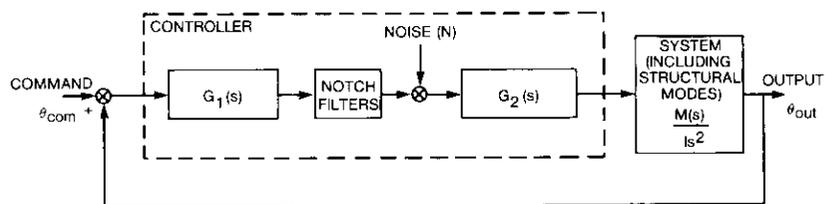
Figure 5. *Quantization Noise in the Control Loop*

Frequencies beyond the loop bandwidth (expected noise spectrum) yield

$$\frac{\theta_{out}}{N} = G_2(s) \frac{M(s)}{Is^2} \tag{22}$$

where $G_2(s)$ is the gain after the digital filter.

The first recommendation would therefore be to place the highest gain possible (given saturation and overflow considerations) prior to the notch filters. The overall loop gain can be preserved by adjusting the gain after the notch filters.

DIRECT IMPLEMENTATION

Figure 6 shows the noise model for direct implementation. The average output noise is [5]

$$\sigma_{0d}^2 = 5 \frac{\sigma^2}{2\pi j} \oint \frac{1}{zB^2(z)} dz \tag{23}$$
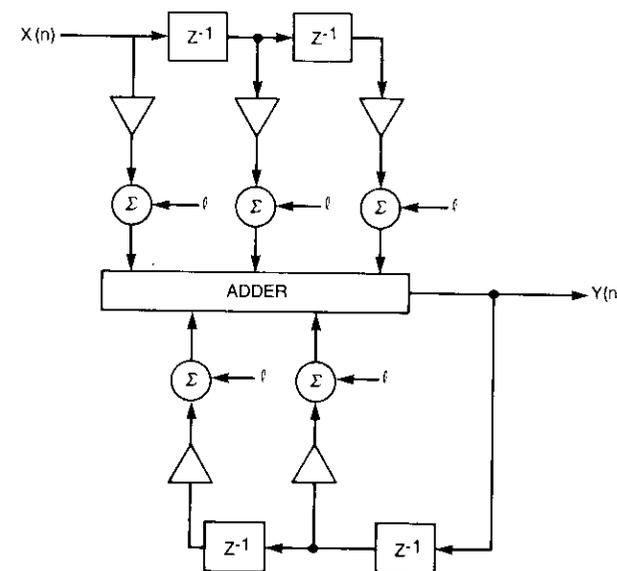
Figure 6. *Noise Model for Direct Implementation*

where σ^2 is presented in equation (21), and $B(z)$ is the denominator of the notch filter transfer function. The power spectral density of the output noise is

$$S_{0d} = S_{ii}(z) \frac{1}{B(z)} \left[\frac{1}{B(z)} \right]^* \tag{24}$$

where S_{ii} is the power spectral density of the input noise and $*$ denotes the complex conjugate. Thus, the noise is passed through a lightly damped complex pole whose behavior equals that of the poles in equation (12). There is some amplification at DC, and substantial amplification at the notch frequency, which is further increased by the transfer function of the structural modes.

CANONICAL IMPLEMENTATION

The noise model for this implementation method is presented in Figure 7. The average output noise is

$$\sigma_{0_c}^2 = 2\sigma^2 + \frac{3\sigma^2}{2\pi j} \oint H(z) |H(z)|^* \frac{dz}{z} \quad . \tag{25}$$
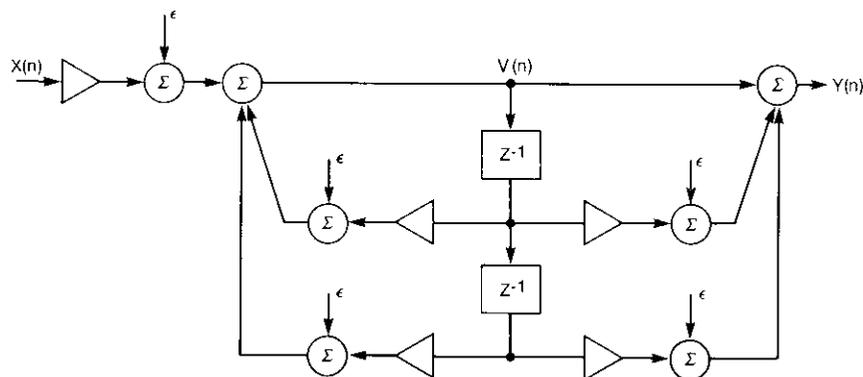
Figure 7. *Noise Model for Canonical Implementation*

This form of implementation does not amplify specific frequencies, as the direct implementation does. However, the noise is amplified because of the gain adjustment made before and after each filter, as discussed earlier. In both cases, careful sequencing of the notch filter cascade for the various structural modes is necessary.

SUMMARY OF RECOMMENDATIONS

Selection of the implementation method for each control system should be made separately. The general recommendations for notch filter implementation, in addition to a long digital word, are:

a. $0.1 < f_n \cdot T < 0.2$;

b. the largest possible ξ_z and ξ_p to achieve the desired gain while preventing a large phase lag; and

c. the highest gain possible prior to the digital controller, including the notch filters.

The learning phase

Since the modal frequencies are not accurately known or may vary with time, the structural frequencies must be detected in order to calculate the notch filter parameters. This learning is based on excitation of the system by white noise [6] or by an impulse, and performance of a spectrum analysis on the system's response. This phase must be executed prior to the operational phase of the satellite.

Spectral analysis

There are two possible approaches to performing the spectral analysis. In the first approach, the output of the detecting filter is continuously calculated, and the process is repeated at each sampling point. Very little memory is used (two or four memory locations for previous outputs and some locations for intermediate results). For low frequencies with narrow filters (high resolution), this approach is time-consuming.

The second approach is based on gathering a certain number of data samples from the sensor output into memory, and then performing the spectral analysis on those points. The speed of this analysis depends only on the processor's capability. This method is preferred for the detection of low frequencies, as in the case under discussion.

Learning methods

The learning phase can be accomplished by using the scanning bandpass filter, the fast Fourier transform (FFT), or the narrowing bandpass filter, as follows:

a. *Scanning Bandpass Filter*. This method is based on varying the frequency of a bandpass filter and measuring its output to determine the frequencies of maximum response.

b. *Fast Fourier Transform*. The FFT is a well known technique [7] for rapidly performing the discrete Fourier transform while avoiding repetitive multiplications. This technique reduces the number of multiplications necessary to perform a spectral analysis for N frequency points from N^2 multiplications to $N \log N$ multiplications.

c. *Narrowing Bandpass Filter*. Like the scanning filter method, this method passes the signal through a bandpass filter and uses the output value as a criterion for spectral analysis. The learning process is as follows:

(1) Gather the necessary data.

(2) Choose the bands and final resolution desired.

(3) Begin calculating the outputs of large-bandwidth bandpass filters in each band.

(4) Subdivide the frequency increments for which the output is largest into narrow bands.

(5) Repeat the calculation until the desired resolution is reached and the structural frequency is detected.

Comparison of learning methods

Analyses and computer simulations were performed to compare the three methods described. The important characteristics of these methods can be summarized as follows:

a. Accuracy. Accuracy was good for all the learning methods. Errors that occur in the narrowing bandpass filter method arise from either close structural modes or low amplitude modes. However, an inaccurate notch filter is adequate in either case.

b. Learning Time. The FFT takes much longer than the two other methods to calculate the low frequencies of the satellite modes.

c. Memory Length. A large amount of memory is required for the FFT.

d. Software. Software is more complex for the narrowing bandpass filter method because of the many decision points in the program.

Thus, the scanning bandpass filter appears to be the best method available for low-frequency spectral analysis.

Adjustment of notch filter frequencies

The frequencies of the notch filters should be adjusted under the following circumstances:

a. Periodically.

b. After an alert (a gradual increase in output not due to command or disturbance).

c. Following a critical event, such as deployment or maneuvers.

One way to accomplish this adjustment is to open the loop and perform a short search using less information. A second method is to move the notch filter frequencies by small increments about the initial frequencies (one at a time) and observe any decrease in output.

A third approach, for cases where stability decreases rapidly (such as after an alert), is to use wider notch filters to cover variations in structural frequencies, even if damping decreases momentarily. The notch frequencies are then adjusted, and when a minimum is reached, the width of the notch filter is slowly reduced.

For the latter two approaches, the dominant frequency at the output is first detected, and optimization is performed around this frequency. Then, the remaining filters are adjusted one at a time. All these steps should be performed while the control loop is closed.

Experimental results

To demonstrate the effect of some of the parameters discussed on a flexible spacecraft, a modified single-axis flexible spacecraft (SAFS) simulator was used. Figure 8 is a schematic diagram and Figure 9 is a photograph of the SAFS simulator, which includes:

a. a system with two circular disks of different moments of inertia floating on a pair of air bearings (with negligible friction), connected by an aluminum rod;

b. a momentum wheel to rotate the upper disk;

c. position sensors to measure the motion of a laser beam across a sensor; and

d. a minicomputer, including the controller.
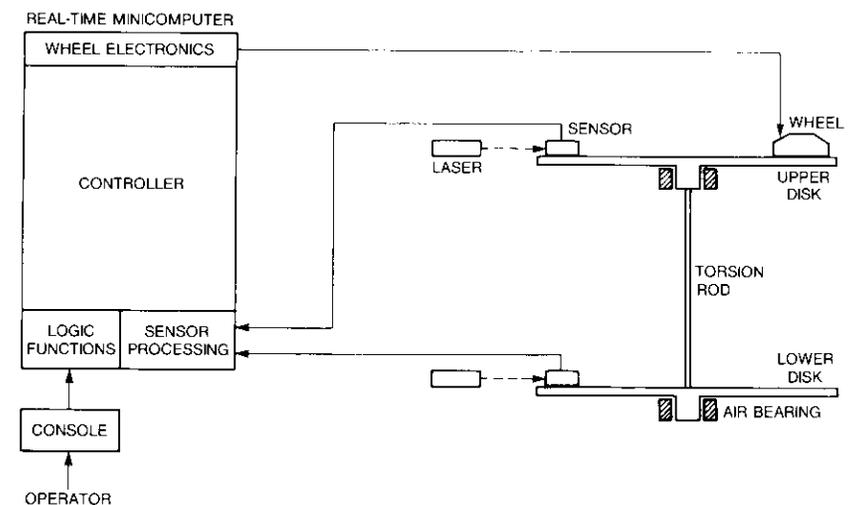
Figure 8. *Schematic Drawing of SAFS*
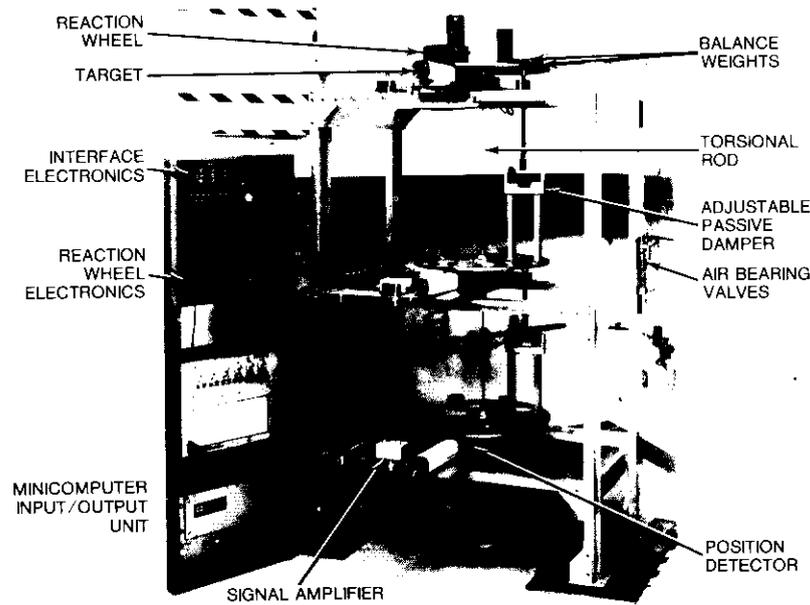
Figure 9. *Photograph of SAFS Simulator*

To create instability in the control loop, the output of the lower disk sensor was used as a feedback signal to the controller which feeds the actuator (momentum wheel) that moves the upper disk, thus creating an unstable structural mode (pole before zero).

Since the control loop performance (bandwidth, stiffness) was not of primary interest here, no change was made to the existing controller

$$G(s) = 1.943 + \frac{0.544}{s} .$$ (26)

All the experimental results are in good agreement with the results of the theoretical analysis presented in the previous sections.

The learning phase

With the control loop open, the structural modes were excited by applying an external disturbance to one of the disks. The output was sampled, and the data were collected in the computer memory. Later, the scanning filter method was applied to the collected data, and a structural frequency of

0.308 Hz was detected. A real-time FFT analyzer detected a frequency of 0.3075 Hz, thus demonstrating the accuracy of the scanning bandpass filter method. After this search, the loop was closed with a notch filter included.

The notch filter

To determine the attenuation to be provided by the notch filter, the output of the control system was examined during tests with different values of attenuation in the control loop. The control system remained unstable for attenuations of less than 3.

To eliminate the effect of the structural mode, a notch filter with an attenuation of 5 ($\xi_z = 0.03$, $\xi_p = 0.15$) was chosen. The notch filter transfer function is then

$$NF(s) = \frac{1 + \dfrac{2 \cdot 0.03s}{2\pi0.308} + \dfrac{s^2}{(2\pi0.308)^2}}{1 + \dfrac{2 \cdot 0.15s}{2\pi0.308} + \dfrac{s^2}{(2\pi0.308)^2}} .$$ (27)

This filter is implemented on the computer by using the bilinear transformation described in equations (6) through (8). Use of the notch filter stabilized the control loop.

Effect of various parameters on system performance

Several parameters of the notch filter design can affect the stability of the system which was measured at the system output (lower disk sensor output).

INACCURACY OF THE NOTCH FILTER FREQUENCY

Figure 10a through d shows the output of the sensor for four notch frequencies. As can be seen, the output is well damped for frequencies close to the structural frequency (0.308 Hz). However, for frequencies below 0.3 Hz or above 0.32 Hz, the system tends to instability. These results show one way to adjust a notch filter frequency—by checking outputs of notch filters of various frequencies and choosing the one which causes the greatest damping of the output.

NOTCH FILTER DAMPING RATIO

Figure 11 illustrates the effect of the damping ratios of the notch filter for several ratios. In Figures 11a and b, the filter gain at the notch frequency is 0.2; while in Figures 11c and d, it is 0.1. For low damping ratios (Figure 11c and d), the system becomes oscillatory, or even unstable. This is because

(a) $f_n = 0.32$ Hz

(b) $f_n = 0.3$ Hz

(c) $f_n = 0.315$ Hz

(d) $f_n = 0.305$ Hz
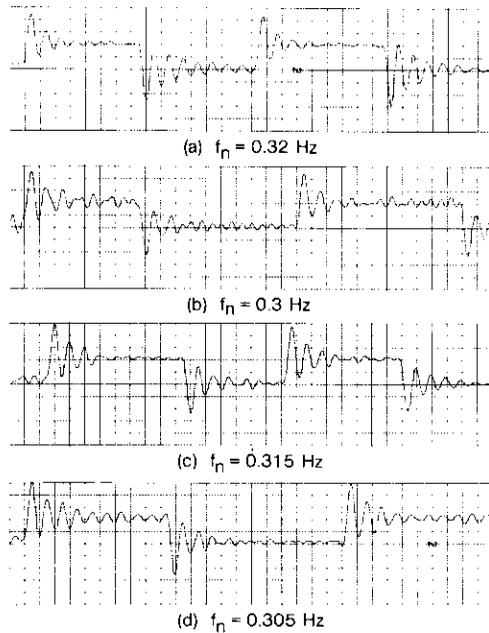
Figure 10. *Effect of Frequency Shift of the Notch Filter*

the narrow bandwidth of the notch filter response does not cover all of the band required by the structural mode.

WORD LENGTH

Figure 12 shows the output of the lower disk sensor for four word lengths. As the word length becomes shorter, the output becomes noisier, and a 12-bit word length yields to instability.

GAIN PARTITIONING

The effect of inserting high or low gain prior to the notch filter was examined, keeping the overall gain constant by compensating after the notch filter. Figure 13 shows the results for three cases. As expected, lower attenuation before the notch filter (higher gain) results in less noise at the output.

IMPLEMENTATION METHOD

The notch filter was implemented by using both direct and canonical methods (Figure 14). The direct method was found to be noisier, as predicted.

(a) $\xi_p = 0.25$, $\xi_z = 0.05$

(b) $\xi_p = 0.15$, $\xi_z = 0.03$

(c) $\xi_p = 0.1$, $\xi_z = 0.01$

(d) $\xi_p = 0.05$, $\xi_z = 0.005$
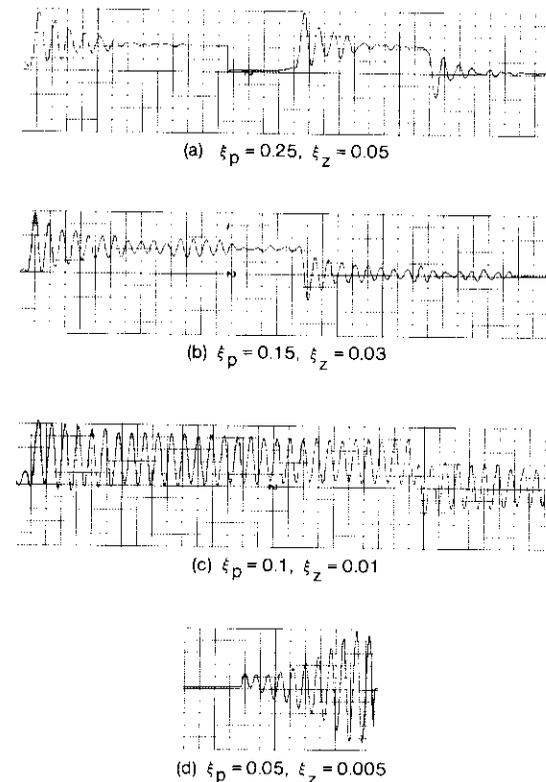
Figure 11. *Effect of Notch Filter Damping Ratios on Control Loop*

However, this result is not instructive since the gain was not adjusted prior to the canonical notch filter [equations (17) and (18)]. The use of gain adjustment would have increased the noise at the output of the canonical notch filter.

SAMPLING RATE

Figure 15 shows the output of a notch filter for sampling periods of 0.118 and 0.236 s. As expected, the noise level is higher for shorter sampling periods. Examination of the expression $1/(1 - \cos \omega_n T)$ for a notch frequency of 0.31 Hz and sampling periods of 0.236 and 0.118 s reveals a ratio of 1:4, which is approximately the ratio of the noise levels for the two sampling rates.

(a) 16 bits

(b) 14 bits

(c) 12 bits

(d) 10 bits
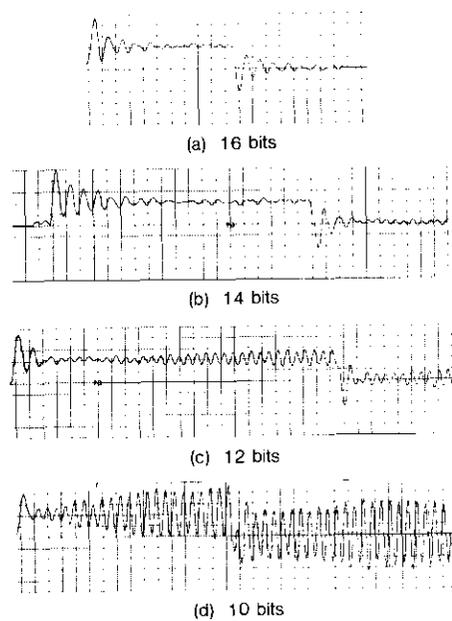
Figure 12. *Effect of Word Length on System Output*

(a) Attenuation = 28.6

(b) Attenuation = 5.73

(c) Attenuation = 2.29
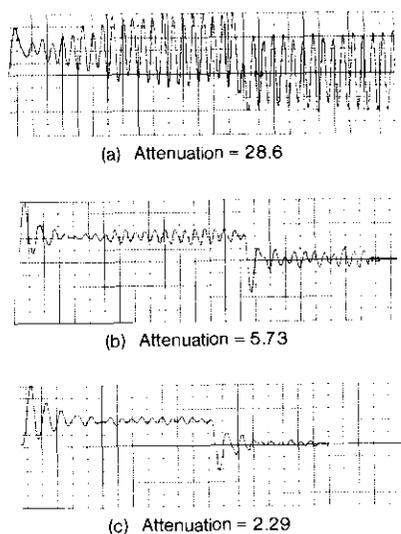
Figure 13. *Effect of Gain Partitioning on System Output*

(a) Direct Form

(b) Canonical Form
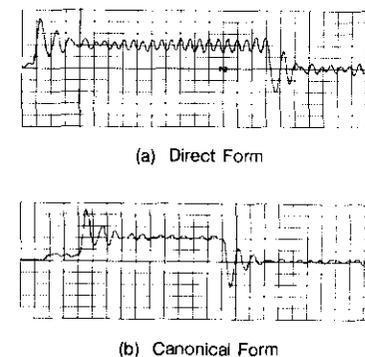
Figure 14. *Effect of Notch Filter Implementation Method on System Output*

(a) Sampling Period = 0.118 s

(b) Sampling Period = 0.236 s
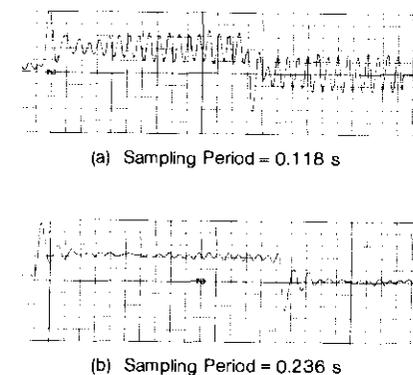
Figure 15. *Effect of Sampling Rate on System Noise*

Conclusions

A method has been presented for counteracting the destabilizing effects arising from the structural modes of complex satellites. This method is based on detecting structural frequencies (learning) and then closing a control loop which includes notch filters at the structural frequencies. Use of the scanning notch filter is recommended for detecting structural frequencies.

To obtain the optimum design for the notch filters, the effect of various parameters on control loop performance was analyzed. Theoretical results were substantiated by experiments performed on the SAFS simulator. The

principal results of these experiments emphasized the need to use the highest gain possible prior to the notch filter, the widest notch filter possible, the longest digital word possible, and a sampling rate of $5f_n$ to $10f_n$, where f_n is the notch frequency.

Acknowledgments

References

[1] V. Itkis, *Control Systems of Variable Structure*, Jerusalem: Keter Publishing House, 1976.

[2] Mark J. Balas, "Active Control of Flexible Systems," Symposium on the Dynamics and Control of Large Flexible Spacecraft, Blacksburg, Virginia, 1977, *Proceedings*, pp. 217–236.

[3] N. K. Gupta, "Frequency Shaped Cost Functionals: Extension of Linear Quadratic Gaussian Design Methods," *Journal of Guidance and Control*, Vol. 3, No. 6, November–December 1980, pp. 529–535.

[4] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

[5] R. E. Bogner and A. G. Constantinides, *Introduction to Digital Filtering*, New York: John Wiley and Sons, 1975.

[6] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Applied Mathematics Series 55, National Bureau of Standards, November 1955.

[7] E. O. Brigham, *The Fast Fourier Transform*, Englewood Cliffs, New Jersey: Prentice-Hall, 1974.

Eliezer Elati received B.Sc. and M.Sc. degrees in electrical engineering from the Technicon, Israel Institute of Technology, Haifa, in 1973 and 1978, respectively. Since 1974, he has been with the Armament Development Authority, Haifa, Israel, working on the development of control systems. During 1981 and 1982, while on sabbatical leave from his job in Israel, he was a Member of the Technical Staff in the Stabilization, Telemetry, and Command Department of the Spacecraft Laboratory at COMSAT Laboratories, where he worked on the stabilization of large flexible spacecraft, and on INTELSAT V support.

CTR Notes

A K-band distributed microstrip termination

A. EZZEDDINE AND H-L. A. HUNG

(Manuscript received June 29, 1984)

Introduction

Because of their small size, active or passive millimeter-wave components are widely used in microstrip line configurations. However, most communications subsystem interconnections remain in waveguide format to reduce unnecessary radio frequency (RF) losses. The waveguide-to-microstrip transition thus becomes a necessary subcomponent of such interconnections. Accurate evaluation of waveguide-to-microstrip transition performance (such as the return-loss parameter) requires a matched 50-Ω microstrip termination that performs satisfactorily at millimeter wave frequencies.

The time-domain reflectometer (TDR) method, which is used to study network discontinuities at lower microwave frequencies, does not provide useful information at K-band because the frequency bandwidth required for such a measurement is at least an order of magnitude higher. Other methods include transforming the frequency domain data from a wideband measurement (0 to 40 GHz) into the time domain by using a fast Fourier transform (FFT) algorithm. By identifying the parasitic response, the discontinuities caused by the waveguide-to-microstrip transition can be calculated. However, such an indirect method does not provide sufficiently accurate results.

Microstrip terminations can also be used to accurately characterize 2-port devices such as metal semiconductor field-effect transistors (MESFETs) or bipolar transistors, as well as hybrid and monolithic circuits at microwave frequencies. In S-parameter

Amin K. Ezzeddine is a Member of the Technical Staff in the Microelectronics Division at COMSAT Laboratories.

Hing-Loi A. Hung is Manager of the Monolithic Microwave Techniques Department in the Microelectronics Division at COMSAT Laboratories.

measurements, a series of calibrations with an open, shorted microstrip reference line and, most importantly, a precision broadband matched 50-Ω microstrip termination, is required to accurately include the mismatch effect of the launcher in either the coaxial-to-microstrip or waveguide-to-microstrip configuration.

This paper describes a new approach in the design of such a microstrip termination for operating frequencies in the DC to 25-GHz band. Results of both theoretical analysis and experimentation are presented.

Theoretical analysis

A PERIODICALLY LOADED MICROSTRIP LINE

In a microwave integrated circuit (MIC) with characteristic impedance Z_0 (for example, 50 Ω), a lumped 50-Ω thin-film resistor on a substrate usually provides a good match at frequencies below X-band. However, at high frequencies $[f > 1/(2\pi RC)]$, the parasitic capacitance, C, of a resistor of value R can cause significant reflection that renders the 50-Ω resistor ineffective as a matched termination. Also, since the resistor must be connected to ground, the inductive effect of the wraparound becomes significant at higher frequencies. The parasitic capacitance could be reduced by decreasing the dimensions of the resistor. However, the dimensions cannot be decreased below certain limits which are determined by the inhomogeneity of the thin-film layer, the dimensional tolerances during fabrication, and the parasitic effect caused by a small resistor width coupled to the wide 50-Ω line.

In the present design, the capacitive effect is minimized by periodically loading a 50-Ω microstrip line with small resistors that gradually absorb the microwave power without significant reflection. Figure 1 is a circuit schematic of such a distributed termination.

A 50-Ω microstrip line of width W is periodically loaded at intervals ℓ_2 with a series resistor of length ℓ_1 and a shunt resistor of width ℓ_3 and length S, as shown in Figure 2. The dimensions of these resistors should be small enough so that the components remain resistive over the entire frequency band of interest. This condition can be expressed as:

$$C \approx \frac{\varepsilon_o \varepsilon_r A}{h} << \frac{1}{2\pi R f_{max}}$$

where f_{max} = highest frequency of interest
  R = resistor value
  C = resistor parasitic capacitance
  A = resistor area
  h = substrate thickness
  ε_o = permittivity of free space
  ε_r = relative dielectric constant of the substrate material.
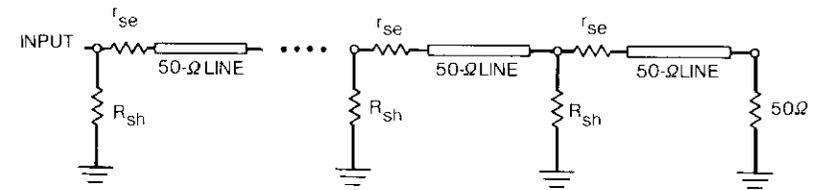
Figure 1. *Circuit Schematic of a Multisection Microstrip Distributed Termination*
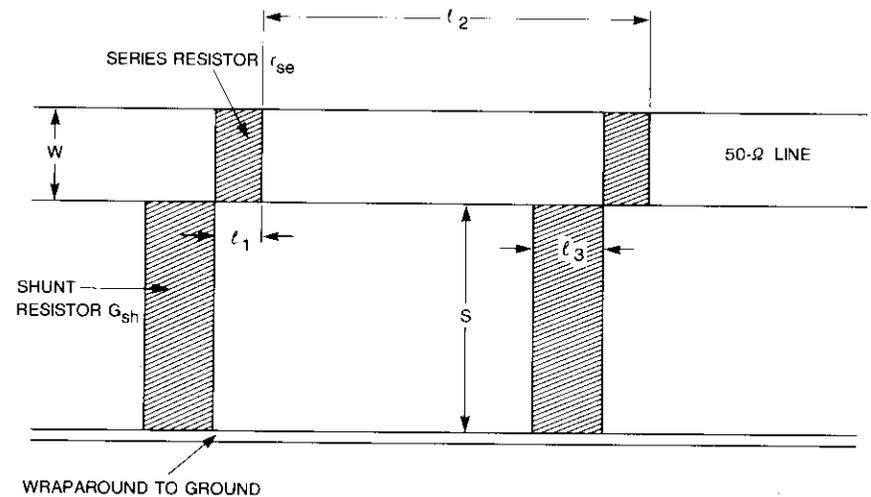
Figure 2. *Section of the Periodically Loaded Microstrip Line*

In the following analysis, ℓ_2 is assumed to be much smaller than λ, the microstrip wavelength of the highest frequency of interest.* For each unit section, let the series resistance be r_{se} and the shunt conductance G_{sh}; then, in the distributive limit (when $\ell_2 << \lambda$), the characteristic impedance, Z_c, of the microstrip line and the propagation constant, γ, are given by

$$Z_c = \sqrt{\frac{r_{se}/\ell_2 + j\omega L}{G_{sh}/\ell_2 + j\omega C}} \tag{1}$$

$$\gamma = \sqrt{(r_{se}/\ell_2 + j\omega L)(G_{sh}/\ell_2 + j\omega C)} \tag{2}$$

---

* In practice, this condition is always met; however, the design can be implemented with $\ell_2 \geq \lambda$.

where, for the quasistatic mode on the microstrip,

L = unit length inductance
C = unit length capacitance
$\omega = 2\pi f$.

Note that the characteristic impedance of the loaded line is frequency dependent. By properly adjusting r_{se} and G_{sh}, the characteristic impedance can be made frequency independent, as follows. If

$$\frac{r_{se}}{G_{sh}} = \frac{L}{C} \tag{3}$$

then

$$Z_c = \sqrt{\frac{L}{C}} \tag{4}$$

$$\gamma = (r_{se}/\ell_2 + j\omega L)\sqrt{\frac{C}{L}} \quad . \tag{5}$$

Equation (5) shows that attenuation, the real part of γ, is also frequency invariant.

The expressions in equations (3), (4), and (5) can now be expressed in terms of the physical parameters of the microstrip line termination. Assuming R_s is the sheet resistance of all resistors, then

$$r_{se} = \frac{R_s \ell_1}{W} \tag{6}$$

$$G_{sh} = \frac{\ell_3}{R_s S} \quad . \tag{7}$$

Equation (3) gives

$$\frac{R_s^2 S \ell_1}{\ell_3 W} = \frac{L}{C}$$

or

$$\frac{\ell_1}{W}\frac{S}{\ell_3} = \frac{L}{CR_s^2} \quad . \tag{8}$$

Since $\sqrt{L/C} = Z_0$ is the characteristic impedance of the microstrip line without loading, and $\lambda = 2\pi/(\omega\sqrt{LC})$ is the wavelength on the line, then

$$\frac{\ell_1}{W} \cdot \frac{S}{\ell_3} = \frac{Z_0^2}{R_s^2} \quad . \tag{9}$$

From equation (5),

$$\gamma = j\omega\sqrt{LC}\left(1 - j\frac{r_{se}}{\omega L \ell_2}\right)$$

or

$$\gamma = \frac{R_s \ell_1}{Z_0 W \ell_2} + j\frac{2\pi}{\lambda} \tag{10}$$

SENSITIVITY TO SHEET RESISTANCE

The ability to control R_s is the limiting factor on the performance of a periodically loaded termination. Therefore, the variation of Z_c with small variations in R_s, must be calculated as

$$Z_c = Z_0\sqrt{\frac{1 - j(R_s\ell_1/W\ell_2\omega L)}{1 - j(\ell_3/\ell_2 R_s S\omega C)}} \tag{11}$$

hence

$$\frac{\partial Z_c}{\partial R_s} = -jZ_0\frac{\ell_1/W\ell_2\omega L}{1 - j(R_s\ell_1/W\ell_2\omega L)} \quad . \tag{12}$$

DESIGN OF A 50-Ω LINE

The distributed resistor also serves as an attenuator, significantly reducing any reflections at the other end of the resistor.

A 50-Ω periodic line is designed on 10-mil fused silica substrate with tantalum-nitride thin-film resistors. The sheet resistance of the thin film is 50 Ω per square. The resistor dimensions are chosen to be small enough to simulate pure resistors, but large enough to overcome the error introduced in the photolithography process without affecting the return loss. Since the dimensional error is nearly 0.2 mil, an 8-mil-long series resistor should adequately satisfy the above two conditions. The 50-Ω microstrip conductor is 3-μm gold-plated over a thin layer of chrome-gold; it is 21.8 mil wide; hence ℓ_1 is 8 mil, W is 21.8 mil, S is 21.8 mil, and ℓ_3 is 8 mil.

The amount of spacing, ℓ_2, selected between successive series-shunt resistors is 50 mil. This selection is based on the requirement that ℓ_2 be sufficiently short to minimize the overall length of the termination, but large enough to reduce the variation of the characteristic impedance with respect to R_s. Also, from equation (12), with the magnitudes of R_s and Z_0 equal to 50 the expression can be simplified to

$$\frac{\partial Z_c}{\partial R_s} = \frac{1}{1 + j\omega L (W\ell_2/R_s\ell_1)}$$

Therefore, at 20 GHz for example,

$$\left|\frac{\partial Z_c}{\partial R_s}\right| = 0.366$$

indicating that Z_c is relatively insensitive to small variation in R_s.

A 10-section periodic line with the above specification was analyzed and optimized by using the SUPER COMPACT computer program. Figure 3 is a listing of the program file for analysis of the distributed termination. All parasitics, such as the resistor capacitance and T-junctions, were modeled accurately, and the line was terminated with a 50-Ω resistor shorted to ground. The length of the shunt resistor, S, and the length of the 50-Ω resistor were optimized to obtain the best return loss from the circuit. A return loss greater than 40 dB was predicted by SUPER COMPACT from DC to 25 GHz.

```
list
     * VARIABLES
     WMS:21.29MIL
     BLK
     TRL  1 2 W=WMS P=34MIL SUB1
     TEE  2 3 4 W1=WMS W2=WMS W3=8MIL SUB1
     TFR  4 6 W=8MIL P=29.18MIL RS=51.60H SUB1
     RES  6 R=0.0000010H
     TFR  3 5 W=WMS P=8MIL RS=51.60H SUB1
     CELL: 2POR 1 5
     END
     LAD
     CELL  1 2
     CELL  2 3
     CELL  3 4
     CELL  4 5
     CELL  5 6
     CELL  6 7
     CELL  7 8
     CELL  8 9
     CELL  9 10
     CELL  10 11
     TFR  11 12 W=WMS P=20.844MIL RS=51.60H SUB1
     RES  12 0 R=0.0000010H
     MATCH1: 1POR 1
     END
     LAD
     TRL  1 2 W=?9.7641MIL? P=1000MIL SUB
     MATCH: 2POR 1 2
     END
     DATA
     SUB: MS H=10MIL ER=9.7 MET1=AU 3.5UM
     SUB1: MS H=10MIL ER=3.8 MET1=AU 3.5UM
     END
     FREQ
     STEP 1GHZ 25GHZ 0.5GHZ
     END
     OPT
     MATCH F=1GHZ 20GHZ MS11=0 W=100
     END
     OUT
     PRI MATCH S
     END
```

Figure 3. *Program Listing for Distributed Termination Analysis*

Experimental results

Figure 4 is a photograph of two fused-silica substrates mounted on two different carriers to facilitate evaluation of the circuits in both coaxial and waveguide measuring systems. Each Invar carrier has two ridges of height equal to twice the substrate thickness. This arrangement further improves RF ground at those locations for the shunt resistors and the terminating resistors that are already grounded through wrap-around gold ribbon connections. Thus, parasitic inductance can be minimized.
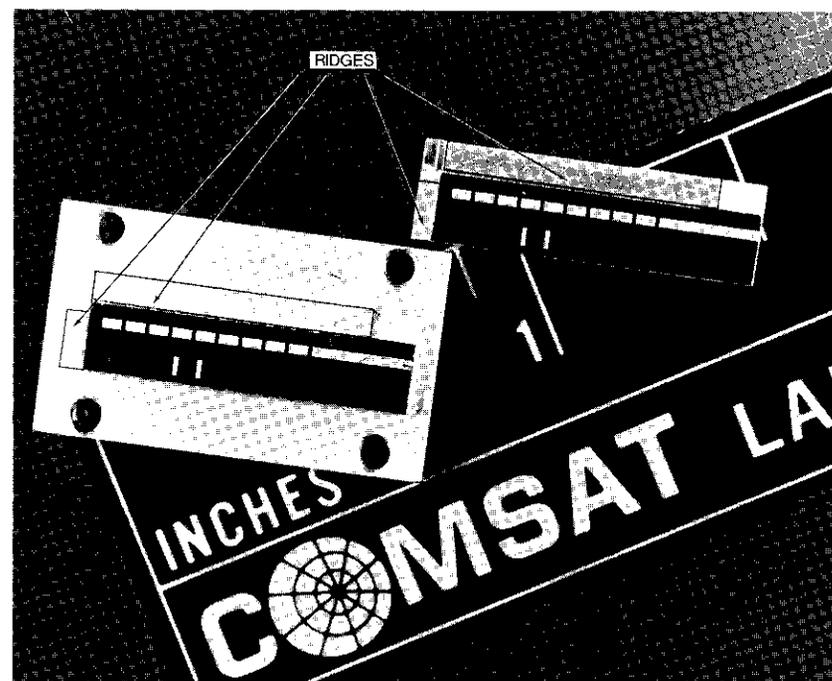
Figure 4. *50-Ω Microstrip Terminations on Two Different Carriers*

DC MEASUREMENTS

The DC resistance measured at the input to ground of the fabricated substrate samples varied from 52 to 53 Ω. The deviation from 50 Ω is caused by small dimensional variation of the resistor components and by sheet resistance that is slightly greater than the designed 50-Ω/square. Sheet resistance, measured on a 50 $\times$ 50-mil test resistor included on each substrate, was found to be 51.6 Ω. Figure 5 shows the calculated return loss of the distributed termination resulting from the slightly higher sheet resistance. A predicted value of greater than 30 dB from DC to 25 GHz was
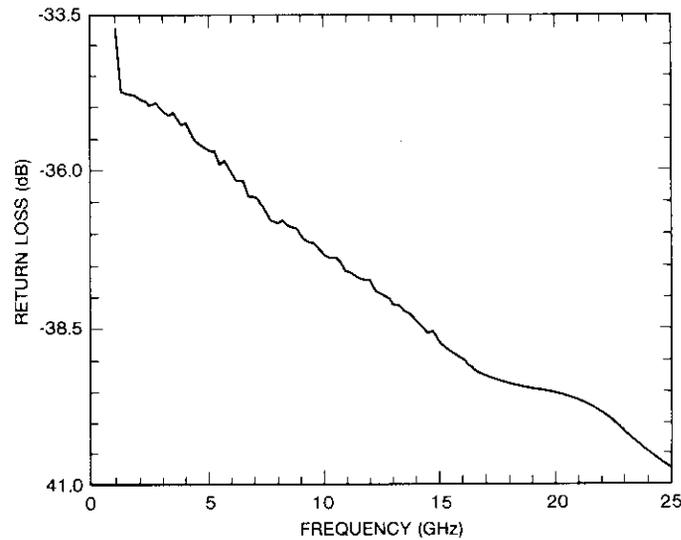
Figure 5. *Computed Return Loss of the Load for a Sheet Resistance* $R_s = 52\ \Omega$

obtained, indicating that substantial margin in performance can be attained with 5-percent variation in the total DC resistance value.

RF MEASUREMENTS WITH COAXIAL-TO-MICROSTRIP LAUNCHER

Two sample substrates were first evaluated in a coaxial measurement setup. The return losses of these two resistors were measured by using an APC7 coaxial-to-microstrip transition. Figure 6 shows the input return loss of the two sample circuits, as measured on an automatic network analyzer (ANA). Two sets of measurements were taken for each resistor by using two launchers.

To understand how the launchers could disturb the measurements, the two launchers were placed back-to-back and connected by a 50-Ω line on 10-mil fused silica. One of the launchers was attached to the ANA; the other was connected to an APC7 precision coaxial load. The return loss of this arrangement, and that of the 50-Ω coaxial load alone, are shown in Figure 7. Each launcher provides appreciable reflection, which makes interpretation of the return loss measurements difficult.

Certain conclusions can be drawn from the measured results illustrated in Figure 6. The coaxial-to-microstrip launcher/termination provided a return loss greater than 20 dB from DC to 18 GHz, and greater than 25 dB from DC to 15 GHz. For frequencies above 18 GHz, the APC launchers caused excessive reflection. Thus, evaluation of the distributed resistor is degraded starting at 15 GHz, and is limited for frequencies above 18 GHz by the performance of the coaxial-to-microstrip line transition.
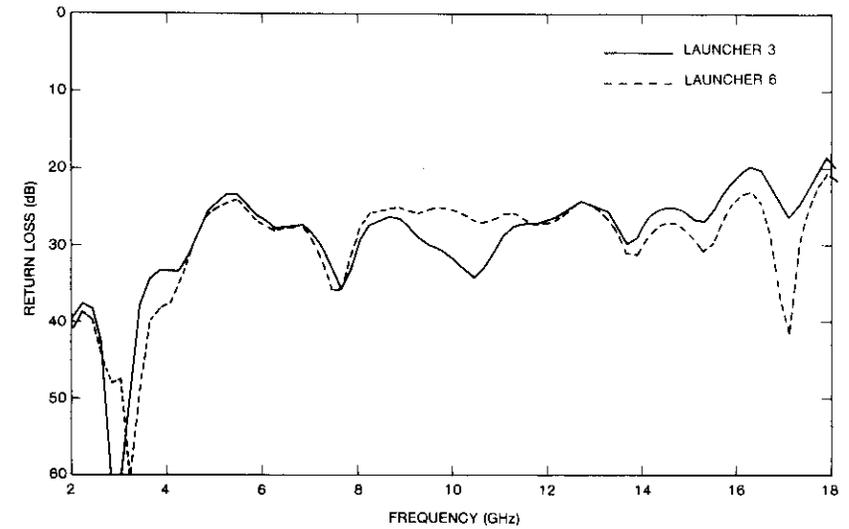
(a) Termination 1
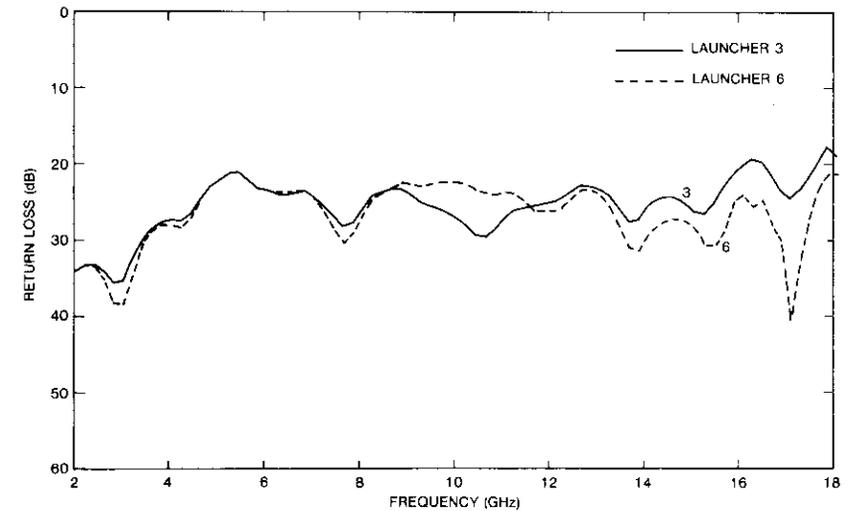
(b) Termination 2

Figure 6. *Measured Return Loss of Two Microstrip Terminations With APC7 Launchers*

Attempts were made to evaluate the microstrip termination at a higher frequency by using SMA-type launchers. A cable-wave SMA connector with a 20-mil center conductor diameter provided better performance than an APC7 connector at higher
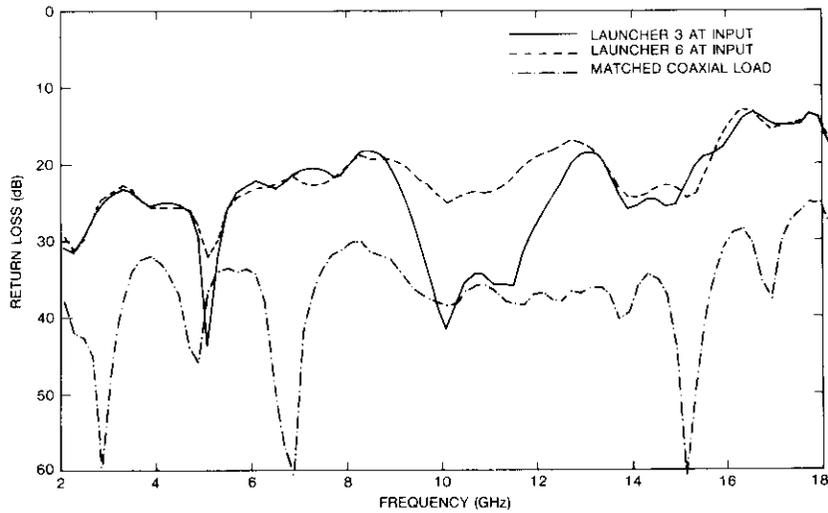
Figure 7. *Comparison of Discontinuity Effect of APC7 Launchers on Microstrip Line and Matched Coaxial Load*

frequencies; however, reflection from the transition was still excessive, as indicated in Figure 8. A Weinschel Engineering (3.5-mm, WPM-3) sliding-pin connector was
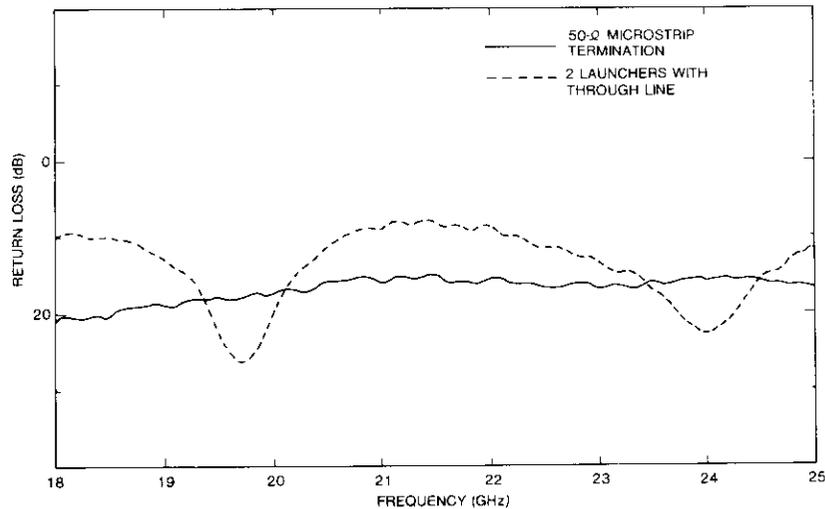
Figure 8. *Return Loss of Microstrip Termination With an SMA-to-Microstrip Connector*

also tested with the resistor in a special housing, as shown in Figure 9. The return loss was then measured from 18 to 25 GHz, and, because of the relatively better performance of the connector, a return loss greater than 20 dB was obtained up to 23 GHz, as illustrated in Figure 10.
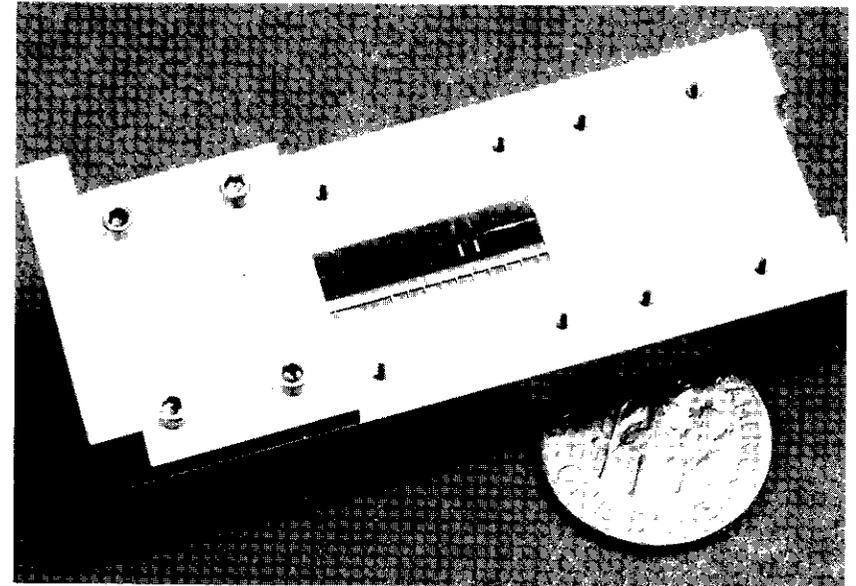
Figure 9. *Test Fixture for the 50-Ω Microstrip Termination With a Weinschel (35-mm, WPM-3) Connector*

RF MEASUREMENTS WITH WAVEGUIDE-TO-MICROSTRIP TRANSITION

Finally, a sample of the microstrip termination was mounted on a special carrier (the lower component in Figure 4) to be inserted in a test housing with a waveguide-to-microstrip transition, as shown in Figure 11. The transition return loss was around 25 dB in the 17.7- to 22.5-GHz band. This approximate return loss value was derived from the results of measurements (Figure 12) of two such transitions that are connected by a 50-Ω microstrip line. The worst return loss value at the peaks in the frequency range, measured with one of the ports terminated with a matched load, is 20 dB. At the peaks, the phases of incident and reflected wave were added at the input transition; thus, the actual return loss for one waveguide-to-microstrip transition is 6 dB lower. Figure 13 shows that the return loss of this resistor was lower than 25 dB from 18 to 22.5 GHz. Beyond that frequency range, the measurement of distributed resistor performance is mainly limited by the degradation in the return loss of the waveguide-to-microstrip transition.
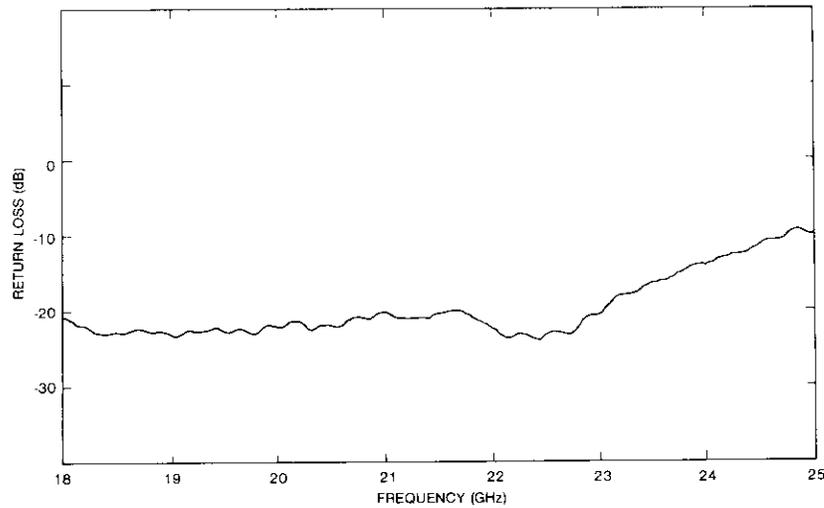
Figure 10. *Return Loss of the 50-Ω Microstrip Termination With the Weinschel Connector*
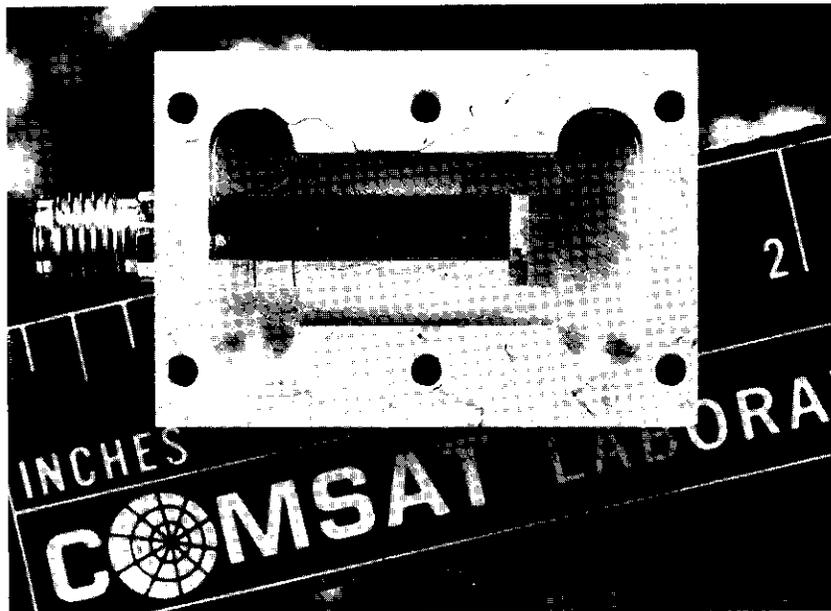
Figure 11. *Housing of an 18- to 23-GHz Waveguide-to-Microstrip Transition*
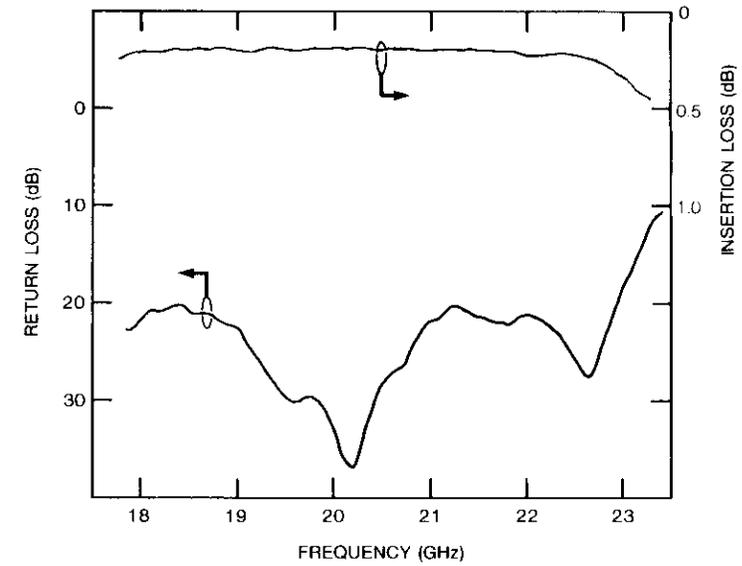
Figure 12. *Return Loss and Insertion Loss of Two Waveguide-to-Microstrip Transitions Connected by a 0.2-in., 50-Ω Microstrip Line*
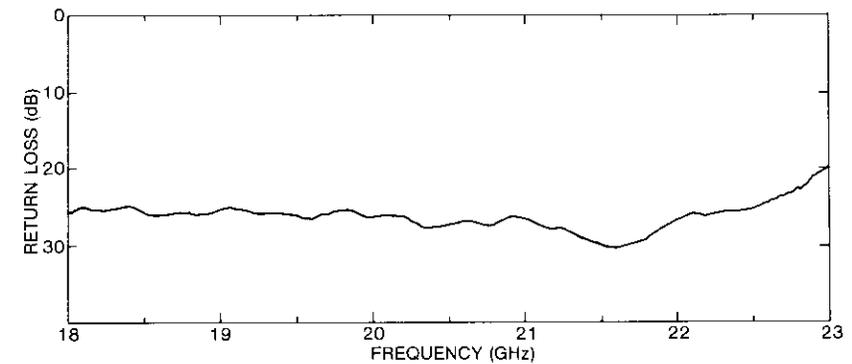
Figure 13. *Return Loss of the 50-Ω Resistor With a Waveguide-to-Microstrip Transition*

Conclusions

A new design has been presented for a K-band microstrip-matched termination that gradually attenuates microwave power with minimal reflection, and with a predicted

return loss of approximately 30 dB from 0 to 25 GHz. The lower values of return loss obtained can be attributed to limitations of the launchers and available measurement techniques. This design approach can be readily extended to a higher frequency band and to substrate materials other than fused silica, such as alumina or sapphire. The new microstrip termination is useful in microwave devices and circuit characterizations which use an error-correction scheme that requires precision 50-Ω loads on microstrip lines, with either a waveguide-to-microstrip transition at millimeter waves or a coaxial-to-microstrip transition at lower microwave frequencies. The distributed resistor network can also be used as an attenuator in microwave integrated circuits.

Acknowledgments

Translations of Abstracts

Estimateur-detecteur combiné pour transmissions de données en MPD-4

J. J. POKLEMBA

Sommaire

La structure d'un récepteur à état stable destiné à la transmission de données en MDP-4 est mise au point, et les performances en sont évaluées. Cette structure repose sur les principes de la théorie de détection et d'estimation. Puisque la théorie de la détection optimale en bande limitée appliquée à la technique MDP-4 est bien connue, cet effort s'oriente surtout vers l'utilisation d'un estimateur—dont le fonctionnement est optimal du point de vue des erreurs quadratiques moyennes—et d'un détecteur combinés. Les trois paramètres à estimer sont l'amplitude et la phase de la porteuse du signal reçu, et la base de temps des symboles des données qui y sont modulées. Ces estimations de base de temps et de phase conduisent à deux boucles verrouillées en phase (PLL) interdépendantes assistées de données. L'estimation d'amplitude est également fonction des signaux de sortie des autres estimateurs et du détecteur. Etant donné que chaque boucle assiste les autres de façon cohérente, le meilleur rapport de signal-à-bruit à la sortie est obtenu lorsque le bruit de voie et les processus à estimer sont tous gaussiens.

Les mesures effectuées peuvent se diviser en trois grandes catégories: la gigue de phase de la porteuse et celle de la boucle de poursuite de base de temps des symboles, le saut de cycles, et le taux d'erreur sur les bits (TEB). D'après les résultats, lorsque le taux de symboles est égal à 32 kbit/s, et que les largeurs de bande sont étroites ou moyennes (à peu près égales ou inférieures à 320 Hz), la gigue de phase sur la porteuse de référence récupérée se situe à 0,4° (valeur quadratique moyenne) de la valeur mesurée dans une boucle à verrouillage de phase fonctionnant avec une porteuse non modulée. En outre, les mesures de gigue de phase applicables à la base de temps de référence récupérée des symboles, avec une séquence de données de 1010 . . . concordent directement avec les prévisions de boucle à verrouillage de phase sur toute la gamme de largeurs de bande faisant l'objet de l'essai, c'est-à-dire, de 32 Hz à 1 kHz.