

COMSAT

Technical Review

Volume 19 Number 2, Fall 1989

Advisory Board

Joel R. Alper
Joseph V. Charyk
John V. Evans
John S. Hannon, Jr.
Willard R. Nichols

Editorial Board

Richard A. Arndt, *Chairman*
James F. Allison
Ali E. Atia
S. Joseph Campanella
Dattakumar M. Chitre
William L. Cook
Russell J. Fang
Ramesh K. Gupta
Ivor N. Knight
Larry C. Palmer
David V. Rogers
Hans J. Weiss
Daniel R. Wells
Amir I. Zaghloul

Past Editors

Pier L. Bargellini 1971–1983
Geoffrey Hyde 1984–1988

Editorial Staff

MANAGING EDITOR
Margaret B. Jacocks
TECHNICAL EDITOR
Barbara J. Wassell
PRODUCTION
Barbara J. Wassell
Virginia M. Ingram
CIRCULATION
Merilee J. Worsey

COMSAT TECHNICAL REVIEW is published twice a year by the Communications Satellite Corporation (COMSAT). Subscriptions, which include the two issues published within a calendar year, are: one year, \$15 U.S.; two years, \$25; three years, \$35; single copies, \$10; article reprints, \$2.50. Overseas air mail delivery is available at an additional cost of \$18 per year. Make checks payable to COMSAT and address to Records Department, Communications Satellite Corporation, 22300 Comsat Drive, Clarksburg, MD 20871-9475, U.S.A.

ISSN 0095-9669

© COMMUNICATIONS SATELLITE CORPORATION 1989
COMSAT IS A TRADE MARK AND SERVICE MARK
OF THE COMMUNICATIONS SATELLITE CORPORATION

COMSAT TECHNICAL REVIEW

Volume 19 Number 2, Fall 1989

- 163** SIMULATING SPREAD SPECTRUM SYSTEMS
L.C. Palmer AND P. Y. Chang
- 195** DIGITAL SIGNAL PROCESSOR-BASED PROGRAMMABLE BPSK/QPSK/
OFFSET-QPSK MODEMS **L-N. Lee, A. Shenoy AND M. K. Eng**
- 235** QUALITY-OF-SERVICE MEASUREMENT OF SATELLITE-BASED
INTERNATIONAL PACKET-SWITCHED NETWORKS **D. S. Gokhale,
R. Kohli AND M. Kacker**
- 259** GENERATION AND DISPLAY OF SATELLITE ANTENNA PATTERNS
W. L. Cook, D. D. Opiekun AND L. R. Karl
- 299** DEMONSTRATION OF A SIMPLIFIED TRACKING MECHANISM FOR USE
WITH INCLINED-ORBIT SATELLITES **E. A. Faine, D. K. Garlow AND
R. W. Gruner**
- 311** A SUN INTERFERENCE PREDICTION PROGRAM
K. T. Lin AND L. J. Yang
- 333** TRANSLATIONS OF ABSTRACTS
FRENCH 333 SPANISH 336

Simulating spread spectrum systems

L. C. PALMER AND P. Y. CHANG

(Manuscript received May 17, 1989)

Abstract

Spread spectrum systems expand the bandwidth of communications signals in an apparently random way, thereby achieving covertness and/or interference immunity. Simulation of these systems requires representation of the very large bandwidth-time products ($W_s T_n$) that are characteristic of such signals. This paper describes a computer program that simulates direct-sequence pseudonoise and frequency hopping (FH) systems. Test results presented include the application of importance sampling to reduce the length of Monte Carlo simulations of FH systems.

Introduction

Software simulation is a well-established adjunct to communications channel modeling and link design. Existing software programs are used to predict performance and to permit tradeoff and sensitivity studies prior to extensive investment in hardware. Some of these software packages have evolved over a number of years [1],[2].

In a typical channel simulation, random information bits are generated and used to modulate a sampled complex-envelope [3] representation of the signal. The modulated signal is oversampled using at least eight samples per symbol to maintain adequate accuracy. The sampled signal is then operated on in turn by software elements that simulate filtering, channel nonlinearities, and additive noise and interference. A simulated receiver operation is used to synchronize to and demodulate the signal, evaluating link performance by

either counting bit errors (the direct Monte Carlo approach) or estimating bit error rate (BER) using hybrid simulation/analysis techniques.

In principle, these simulation techniques can be applied directly to spread spectrum signals [4]–[8]. However, in practice, the scale of the simulation expands by several orders of magnitude when dealing with such signals. This expansion is best illustrated by considering a binary phase shift keyed signal that is converted into a direct-sequence pseudonoise (DSPN) signal by modulo-2 adding a sequence of pseudorandom binary elements to the original data bits.

In typical DSPN applications, the rate of the pseudorandom binary elements might be two or three orders of magnitude higher than the original data bit rate, so that each original data bit is subdivided into a large number of pseudorandom elements, or chips. The ratio of bit duration, T_b (equals $1/R_b$, the bit rate) to chip duration, T_c (equals $1/W_s$), $\gamma = T_b/T_c = W_s/R_b = W_s T_b$, is referred to variously as the signal bandwidth-time product, the processing gain (of the correlation receiver), or the bandwidth expansion (ratio of spread bandwidth to data bandwidth). To simulate these DSPN systems using existing channel modeling and simulation programs, samples of the individual chips must be created within the program. This increases the simulation length by the bandwidth expansion factor γ , which can be 1,000 or more.

In contrast to DSPN systems that expand bandwidth by using a high-chip-rate (R_c) signal to modulate the carrier, thereby creating a signal with a wide power density spectrum, frequency-hopping (FH) signals have narrow instantaneous bandwidth but hop around in frequency in a pseudorandom manner to cover a very wide bandwidth. After the intended receiver is synchronized to the pseudorandom code used at the transmitter, the receiver can follow and de-hop the incoming signal to restore it to the original, relatively narrow, baseband where demodulation can take place.

Implementation of frequency hopping requires a frequency synthesizer that can output orthogonal frequencies over a wide bandwidth (W_s) with spacing R_h . The ratio W_s/R_h can be very large and is analogous to processing gain in the DSPN system. Frequency hopping can be either slow or fast relative to the information bit rate, R_b . With slow frequency hopping, the transmitter changes frequency slowly relative to the bit rate, so that many bits are sent before the synthesizer makes a large change in frequency. With fast frequency hopping, the synthesizer hops to a new frequency several times during each data bit. Demodulation then requires combining the metrics from the diversity hops that make up each data bit. Because hopping is almost always incoherent, FH systems typically use a noncoherent modulation technique such as orthogonal M -ary frequency shift keying (M -ary FSK).

This paper describes a simulation program called SPREAD† that models and simulates both DSPN and FH systems. The simulation structure repetitively generates and processes large blocks of N simulated samples representing one or more spread spectrum signals. Certain elements, such as signal generators, are unique to the SPREAD program; other elements such as filters, nonlinearities, noise generators, and certain demodulator elements have been adapted from the CHAMP [9],[10] program.

After the program structure was defined, the program was validated by testing using classical channel modeling configurations [11] with known results. For example, DSPN signals with different processing gains were simulated through linear and hard-limiting channels. Monte Carlo, or error-counting, simulations were also made for FH systems using noncoherent M -ary FSK.

The program was developed on an IBM 3083G computer and then converted for use on an Apollo Model DM3000 workstation. Typical simulations run at a rate of 0.1 to 1 bit/s. DSPN simulations can use hybrid simulation/analysis techniques that determine an average BER from the scatter diagrams collected at the output of the correlation receivers. Because of the highly nonlinear nature of M -ary FSK demodulators, which are implemented as a bank of M matched filters followed by envelope detectors, Monte Carlo (error-counting) simulations must be conducted.

Since lengthy Monte Carlo simulations of FH systems become impractical for large numbers of data bits (10,000 or more), importance sampling [12],[13] has been investigated as an alternative means of predicting low bit error probabilities. The application of importance sampling to FH simulations has given reasonably accurate results for some known cases. More experimentation is needed to gain confidence with this technique.

Simulation approach

The SPREAD program is a time-domain simulation that models and simulates spread spectrum systems to evaluate BER performance. The program contains signal generation, signal analysis, and signal reception functions. These elements provide a flexible capability to generate different spread spectrum and jamming signals and to evaluate the end-to-end performance of spread spectrum links. This section describes the overall structure and function of the program.

† The SPREAD program is being developed by COMSAT Laboratories under corporate sponsorship.

Overall program structure

The SPREAD program can simulate one or more DSPN or FH signals. For example, multiple DSPN signals can be simulated to evaluate code-division multiple access (CDMA) systems. The program is divided into four major parts, as shown in Figure 1: initialization, signal generation, channel analysis, and receivers. Signal samples are processed one block at a time. In each block, a signal generation section creates samples of DSPN or FH signals. These signals have a common sampling rate, f_s , and are added together to create a sum signal that occupies the total bandwidth, W_s . The blocks correspond to either one data bit of the lowest-rate DSPN signal or one hop of the FH signal with the longest dwell time. A large number of signals can be generated, added together, and then demodulated individually.

Spectrum spreading of the signals is controlled by selectable PN generators. Currently, linear shift register generators with a maximum of 31 stages are used. For each signal, identical generators are included at the transmitter and the receiver. Each code generator can be initialized independently or synchronized to be exactly in phase.

After generation and summing, the composite signal can be operated on by conventional channel elements such as filters and nonlinearities before noise and interfering or jamming signals are added. After these elements are combined, the signals can be processed using routines that perform operations such as spectral analysis.

The final section of the program contains receivers that demodulate each signal. These receivers operate sequentially on each block of data to demodulate the one signal to which they are synchronized. Individual PN generators are used which, when synchronized, correlate the composite signal with local references to de-spread the DSPN signals or de-hop the individual FH signals. Following these operations, the individual baseband signals are demodulated using conventional simulation modules. In particular, standard hybrid simulation/analysis techniques [10],[13] are used to determine BER for DSPN signals, and error-counting routines are used to determine BER for FH signals.

Signal generation

DSPN signals are generated in sampled complex-envelope form as

$$S_k = \sqrt{P} D_I^t P_N^t \exp(j 2\pi f_o k/f_s + j\theta) + j\sqrt{P} D_Q^t P_N^t \exp(j 2\pi f_o k/f_s + j\theta) \quad (1)$$

where S_k = k th complex signal sample
 P = power in the signal

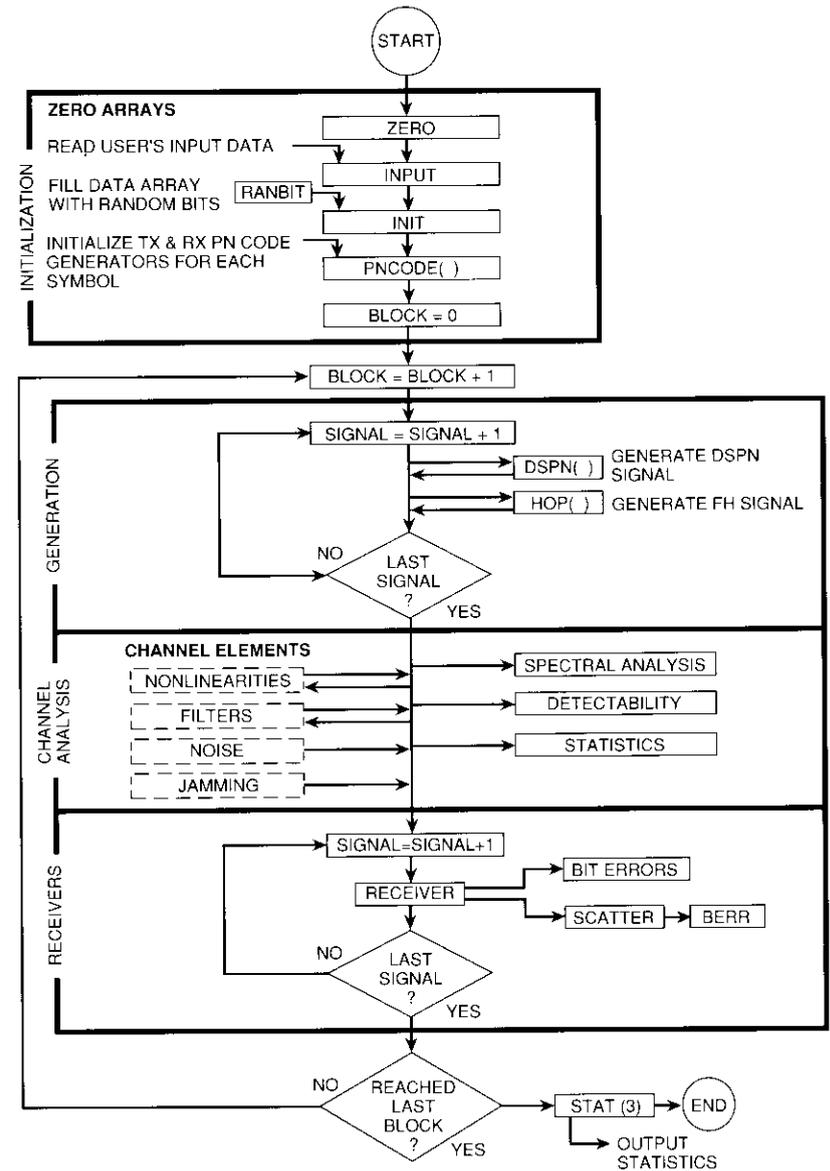


Figure 1. Elements of the Spread-Spectrum Simulation Program

- $D_i^f, D_Q^f = \ell$ th in-phase and quadrature data bits (± 1), respectively. [For binary phase shift keyed (BPSK) data, $D_i = D_Q$.]
- $PN_i^j, PN_Q^j = j$ th in-phase and quadrature chips (± 1), respectively. (Typically, j advances 1, 2, . . . , N , where N denotes the integer number of chips per bit, and then the data bit changes.)
- $f_o =$ signal frequency (which can equal 0)
- $f_s =$ sampling rate
- $\theta =$ initial phase angle of this signal. For coherent DSPN signals this remains constant throughout the simulation run.
- $k =$ continuous index of the samples.

FH signals are generated using the inverse discrete Fourier transform (DFT) as

$$S_k = \sum_{r=0}^{N-1} A_r \exp\left(j \frac{2\pi kr}{N} + j\theta\right) = \sqrt{P_a} \exp\left(j \frac{2\pi ka}{N} + j\theta\right) \quad (2)$$

- where $S_k = k$ th complex signal sample
- $a =$ index [$0 \leq a \leq (N - 1)$] of the single complex DFT coefficient that is *nonzero*. This index is selected as a random B -bit integer, $B = \log_2(N/4)$. The least significant m bits [$m = \log_2(M)$] are supplied by the random data source for M -ary FSK. The most significant $(B - m)$ bits are obtained from the PN generator and control the coarse hopping.
- $N =$ block size of the DFT
- $A_r =$ complex DFT coefficient at discrete spectral line r . For a single hopping signal, all $A_r = 0$ except the line with index $r = a$.

The power in the hopping signal at discrete spectral line a is selected as $A_a A_a^* = P_a$, where $*$ denotes the complex conjugate. The phase angle $\theta = \tan^{-1}\{\text{Im}(A_a)/\text{Re}(A_a)\}$ is selected randomly on the interval $(0 - 2\pi)$ from hop to hop. Hopping is restricted to the interval $\pm 1/4 f_s$ to allow for subsequent filtering and to be compatible with the DSPN generation, which must oversample the PN chips using at least eight samples per chip.

The frequency resolution of the DFT is f_s/N , where N is the block size. For a typical 4-ary FSK FH simulation with a DFT block size of 8,192 complex samples, groups of four randomly selected tones can be hopped over $\pm 1/4 f_s$, or over $4,096/4 = 1,024$ groups of four frequency bins. This subdivision of the discrete frequency domain, and its relationship to the discrete spectrum of a high-rate DSPN signal, are shown in Figure 2.

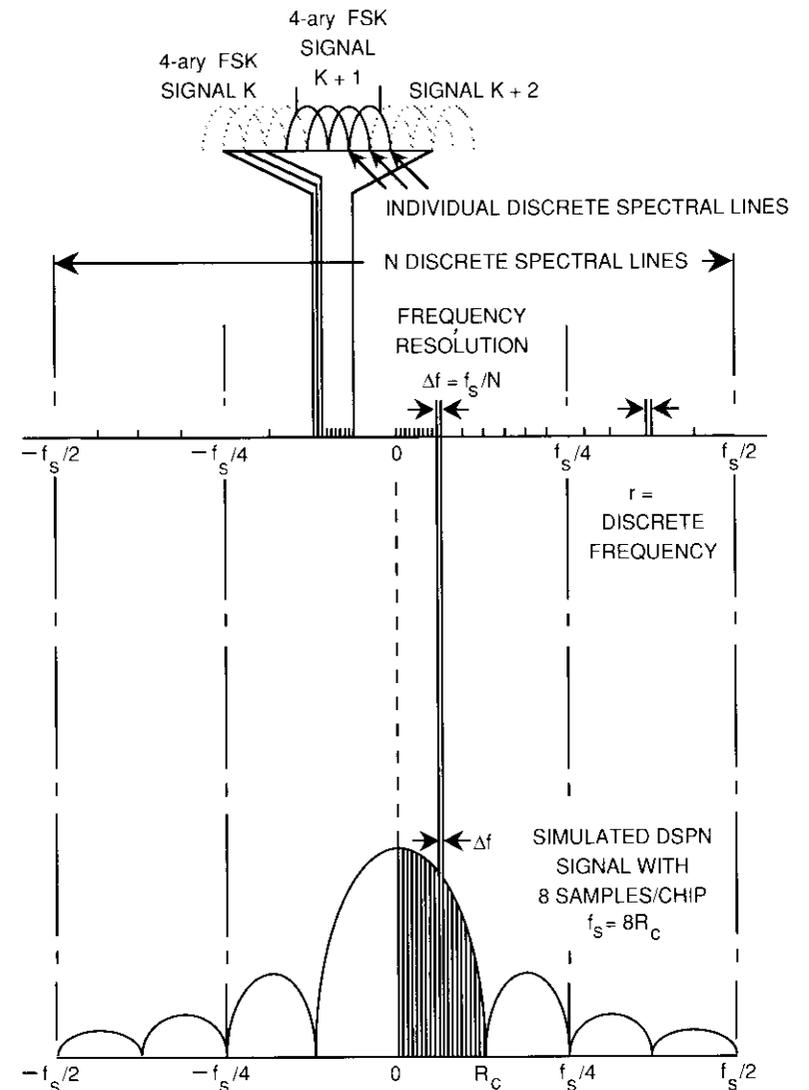


Figure 2. Representation of DSPN and FH Signals in the Discrete Frequency Domain

Correlation receivers

The spreading operation performed at the transmitter must be undone at the receiver in order to collapse the bandwidth back to that of the original data. This is accomplished in the simulation program by implementing correlation receivers for each signal to be demodulated. For DSPN signals, these correlators multiply the incoming waveform, chip by chip, by a locally derived PN sequence. The individual products are accumulated by an integrate-and-dump operation over each data symbol. With perfect carrier-phase and chip-timing synchronization, this operation reverses the spreading that was done at the transmitter. The operation is shown conceptually in Figure 3.

The receiver for FH signals also uses a local PN generator to de-hop the incoming signal, leaving a baseband containing the M -ary FSK tones of that particular signal. Several options then exist in the receiver implementation to normalize the signal levels prior to accumulating the matched-filter metrics, making the final selection of the filter with the largest output, and mapping to the corresponding data bits. These alternatives are shown conceptually in Figure 4.

Summary of program features

The simulation program is implemented to allow different PN generators to spread BPSK or quadrature phase shift keyed (QPSK) signals or to control the hopping of M -ary FSK signals with $M = 2, 4, 8,$ or 16 . Noise or tone jamming can be added to the signals, in addition to like-signal interference (for CDMA simulations). Block size is currently limited to 2^{14} (16,384); therefore, frequency hopping over $\pm f_c/4$ allows 8,192 frequency bins to be used. With 4-ary FSK, 2,048 hopping positions are available. Signals are processed one block at a time, thus providing unlimited simulation length. Conventional graphic routines for spectra and scatter plots are taken from the CHAMP program.

The SPREAD program is written in Fortran 77 and requires ≤ 3 Mbyte of core memory. It is currently being run on an Apollo DM3000 workstation.

Simulation results

Tests were conducted to demonstrate program capabilities and to validate the simulation of DSPN and FH systems. These tests included channel configurations where performance trends can be predicted from known results.

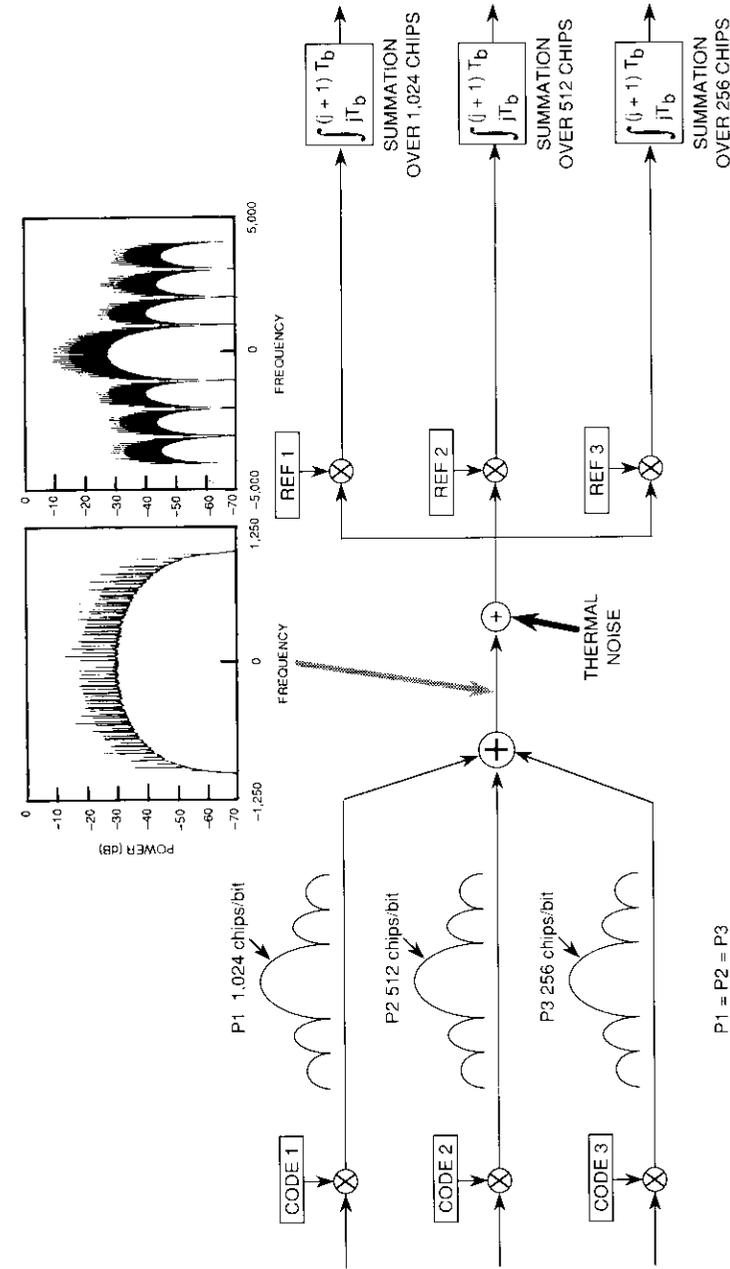


Figure 3. Model for Three DSPN BPSK Signals

P1 = P2 = P3

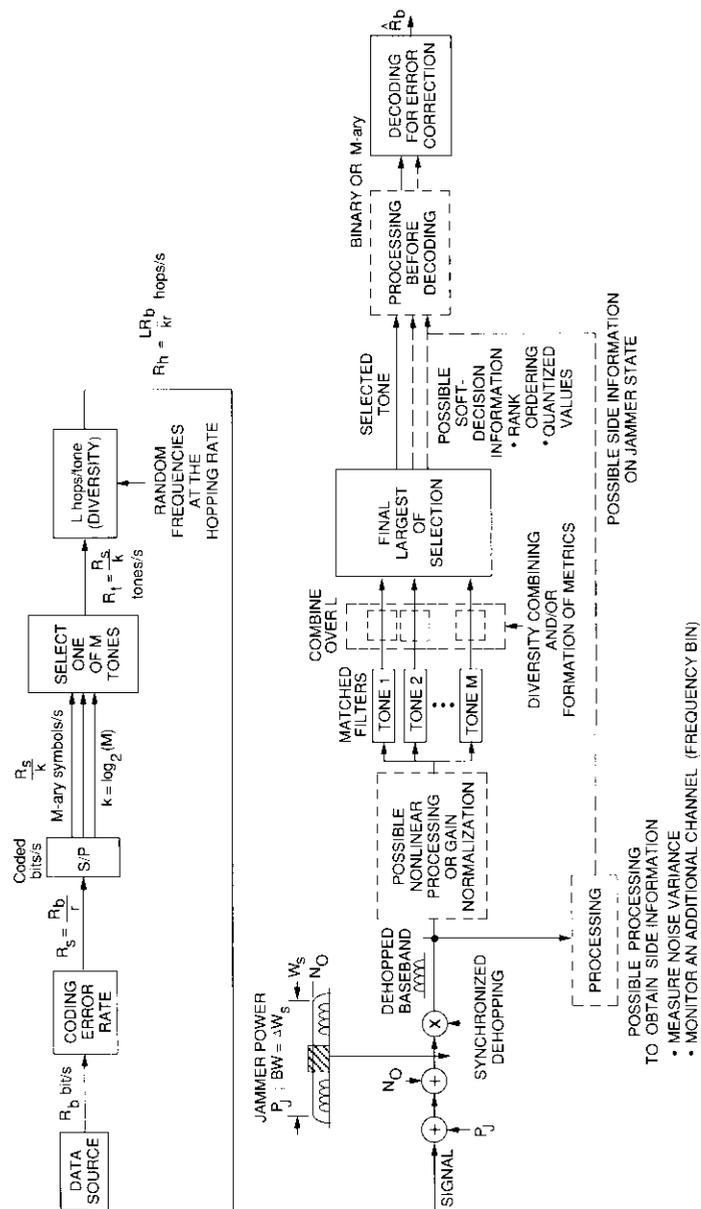


Figure 4. Frequency-Hopping Transmitter/Receiver

Simulation results for DSPN systems

THREE DSPN SIGNALS

Simulations were run to test the DSPN correlation receivers and calibrate the BER estimation process. Figure 3 illustrated a test case with three BPSK DSPN signals in a CDMA configuration, using different PN codes but the same chip rate. These signals have equal power but a different number of chips per bit (1,024, 512, or 256), as indicated.

The codes for the signals are generated by linear shift-register configurations with 4, 5, and 13 stages, respectively. The short codes repeat several times in a bit duration, resulting in discrete line spectra which are evident in the composite spectrum, as depicted in the insert of Figure 3. Figure 5 shows scatter diagrams in the decision region, as well as estimated average BER results. The performance degradation due to mutual interference increases as the number of chips per bit (processing gain) decreases.

INTERFERING SIGNALS

A second series of simulation runs was made with a single DSPN signal and the following three interference conditions:

- One sine wave jamming signal much larger than the communications signal
- Ten sine wave jamming signals with combined power much greater than the communications signal
- White noise over the transmission bandwidth.

These runs were conducted for jamming-power to signal-power ratios (J/S) of 13, 17, 20, and 23 dB. The tests were then repeated with an ideal hard-limiter in the channel, as shown in Figure 6. Figure 7 depicts simulated spectra for these cases.

BER results are plotted in Figure 8 for a single sine wave jammer vs J/S . The jammer power, J , appears in the despread baseband channel as noise in the in-phase and quadrature channels. These additional noise components have total power, or variance, $\sigma_j^2 \cong [(J/S)/PG]^2$, where PG is the processing gain or number of chips per bit. One-half of this power falls into each of the in-phase and quadrature channels and adds to the thermal noise that is already present on these channels with relative variance of $\sigma_n^2 = N_o/2E_b$. As an example, for the signal with a processing gain of 512, the jammer variance out of the correlator would be $(100/512) = 0.195$ when $J/S = 20$ dB. This value adds to thermal noise. For example, when $E_b/N_o = 10$ dB, one-half of the jammer noise adds to each channel, effectively doubling the noise or halving the E_b/N_o .

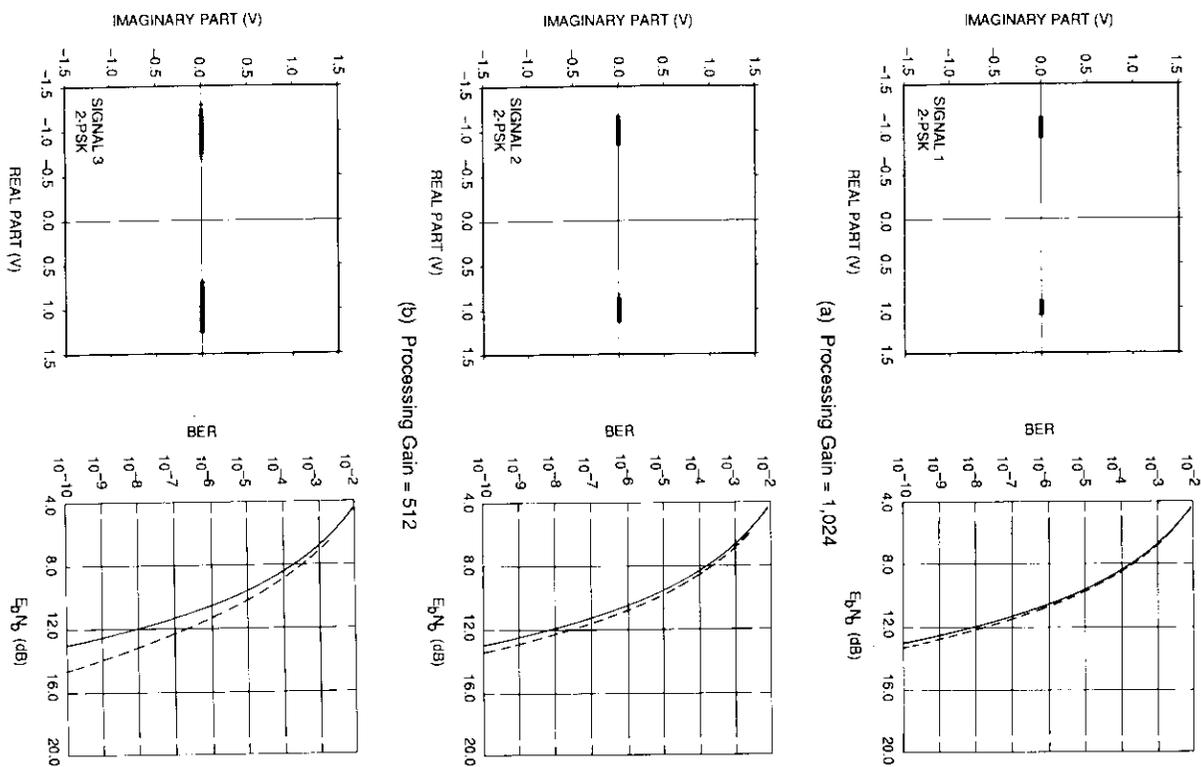


Figure 5. BER Results for Three DSPN Signals

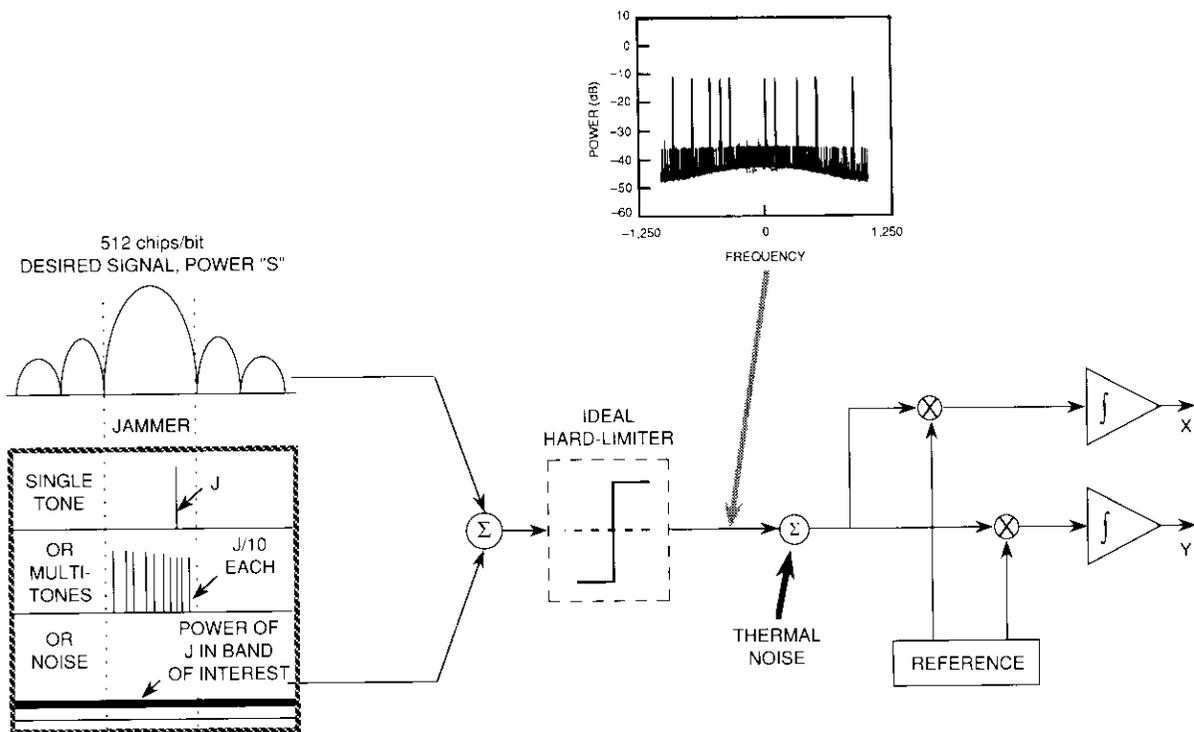


Figure 6. Model for One DSPN QPSK Signal With Various Jammers

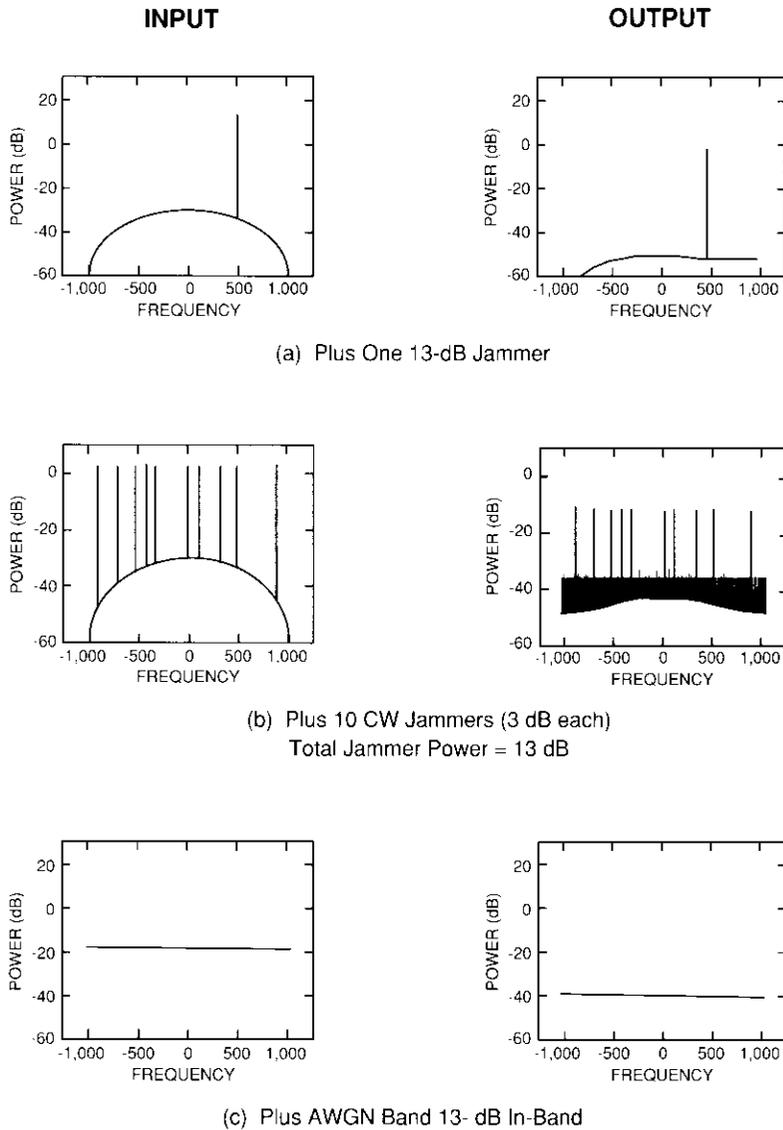


Figure 7. Simulated Spectra Into and Out of Hard-Limiter

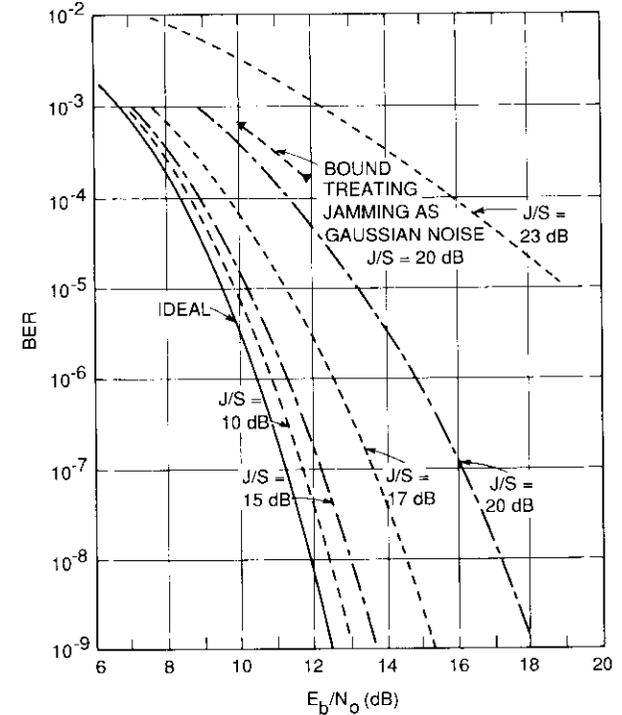


Figure 8. Simulation Results: DSPN With Liner Channel and Single Sine-Wave Jammer

Figure 9 summarizes results from a final series of six DSPN tests. BER is plotted for three different types of jamming signals, with J/S kept constant at 13 dB, and with and without the hard-limiter in the channel. Without the limiter, the three types of jammers degrade the channel by approximately 1 dB at a BER of 10^{-6} . Ten sine wave jammers are almost equivalent to Gaussian noise. When the limiter is included in the channel, a small additional degradation is noted for the multiple sine wave and noise jammers. With the single sine wave jammer, an additional loss is noted because the desired carrier is suppressed in the limiter.

The trend of these results agrees with analytical findings [14] and confirms the general behavior expected. For a hard-limiting channel, approximately 1 dB of additional degradation is expected for a noise jammer, as compared with the linear channel. The sine wave jammer will produce additional performance degradation due to signal suppression in the limiter. For

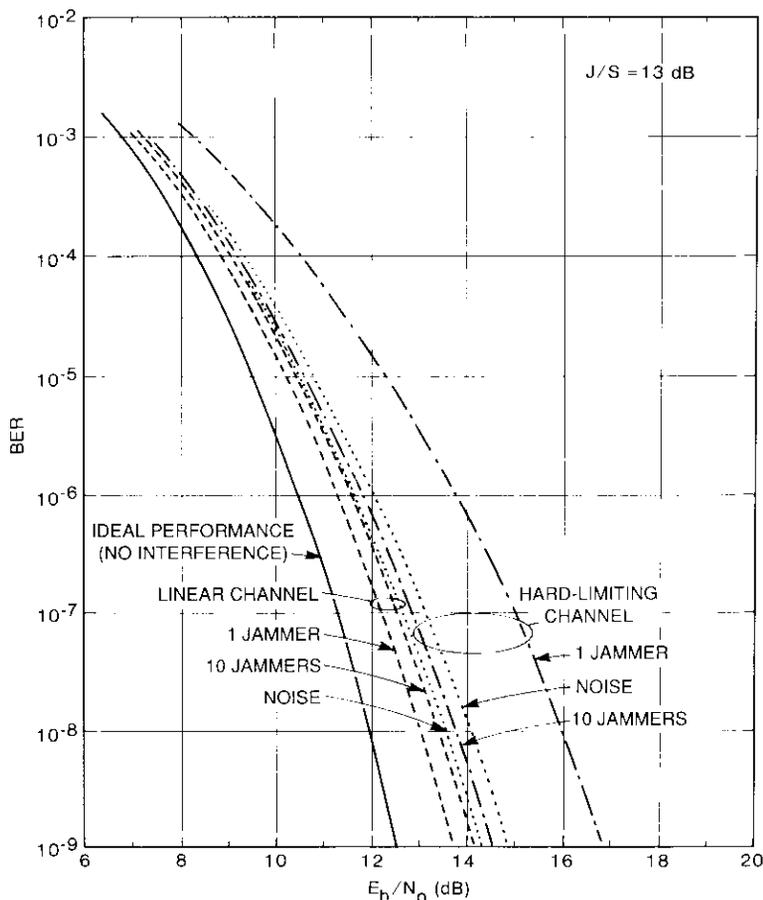


Figure 9. Jamming Results for Single BPSK DSPN Signal (512 chips/bit) With Quaternary PN

quaternary PN and BPSK data, this additional loss would approach 6 dB as J/S became very large [14].

Simulation results for FH systems

As illustrated in Figure 4, FH systems were simulated using a block size, N , of 8,192. Noncoherent M -ary FSK is simulated with $M = 2, 4$, and 8. Tones are generated by selecting a random B -bit word [$B = \log_2(8,192) = 13$], with the $m = \log_2(M)$ least significant bits determined by the random data bits, and the $(B - m)$ most significant bits determined by the PN

generator at the transmitter. For example, with 4-ary FSK, hopping would take place over 1,024 sets of four contiguous tones covering $\pm 1/4 f_c$.

The phase angle of the tone is random on $(0 - 2\pi)$ from hop to hop. After frequency-to-time conversion using the inverse DFT (IDFT), Gaussian random numbers are added to the in-phase and quadrature channels. After passing through the channel, the DFT (time-to-frequency) precedes the de-hopping step, where a synchronized PN generator at the receiver selects the M discrete spectral lines corresponding to the M possible tones of the M -ary FSK signal. The discrete spectral lines are then converted to envelope as $P_r = A_r A_r^*$.

The above operation implements a bank of M matched filters, followed by envelope detectors. If one hop corresponds to one symbol, then the M envelope values are compared to find the largest ("largest-of-selection"), and this largest line is mapped to recover the demodulated information bits. These bits are then compared to the transmitted bits, and errors are counted. With fast frequency hopping, transmitted symbols are divided into L diversity chips, each of which is hopped over a wide bandwidth. The matched filter outputs on each of these diversity hops must be saved and combined before a final decision can be made on the symbol. Figure 4 showed options for this combining, as well as for the use of side information such as information on jammer state.

To evaluate the performance of FH systems, Monte Carlo simulations must be conducted because of the highly nonlinear nature of the demodulation process. Evaluation of BER by error counting requires extremely long simulation runs, with the time increasing as lower BERs (higher E_b/N_o) are evaluated. It was necessary to investigate some means of reducing the CPU time required for these simulations. The importance sampling technique [12] offers promise in this area, and results obtained with importance sampling are given later in this paper.

Table 1 summarizes the results of Monte Carlo simulations used to determine the BER of FH systems with one hop per M -ary symbol. Figure 10 compares these results with theoretical results for $M = 2, 4$, and 8. For an M -ary orthogonal FSK system with $m = \log_2(M)$, the theoretical bit error probability is [15]

$$P_b = \frac{1}{2(M-1)} \sum_{n=2}^M (-1)^n \binom{M}{n} \exp \left[-m \frac{E_b(n-1)}{N_o n} \right] \quad (3)$$

which, for $M = 4$, gives

$$P_b = e^{-E_b/N_o} - 2/3 e^{-4/3 E_b/N_o} + 1/6 e^{-3/2 E_b/N_o}$$

TABLE 1. MONTE CARLO SIMULATION RESULTS FOR FREQUENCY HOPPING USING M-ARY FSK IN WHITE NOISE

M	E_b/N_o (dB)	NO. BITS SIMULATED	NO. ERRORS	BER	1 σ (67%) CONFIDENCE INTERVALS	
					LOWER	UPPER
2	5	2,048	209	0.102	0.095	0.108
	7	3,000	130	0.043	0.040	0.047
	8	2,048	48	0.023	0.020	0.027
	10	7,000	20	0.00286	0.0022	0.0035
4	0	2,048	502	0.245	0.235	0.254
	2	4,096	562	0.137	0.132	0.142
	4	4,096	250	0.061	0.057	0.065
	6	17,384	256	0.0156	0.0146	0.0166
	7	32,768	221	0.0067	0.0063	0.0072
	8	18,000	35	0.0019	0.0016	0.0023
8	3	1,536	103	0.067	0.061	0.073
	4	6,144	215	0.035	0.032	0.0373
	5	12,288	187	0.0152	0.0141	0.0163
	6	18,000	83	0.0046	0.0041	0.0051
	7	22,000	23	0.0010	0.0008	0.0013

Good agreement is noted between the results shown in Figure 10 and the theoretical results.

Additional Monte Carlo simulations were conducted for $E_b/N_o = 1, 2,$ and 3 dB, with results as shown in Table 2. Note that the simulation runs for E_b/N_o of 1 and 2 dB are quite accurate. A fractional accuracy of $(Np)^{-1/2}$ is expected at the 1 σ confidence level, which would be 3 percent at $E_b/N_o = 1$ dB and 4 percent at $E_b/N_o = 2$ dB. At $E_b/N_o = 3$ dB, a 1 σ accuracy of about 3.5 percent is expected, and the observed error is twice this value.

For these three simulation runs, convergence of the simulated BER results was checked vs run length. The results are shown in Figure 11, where simulation run lengths range from 4,096 to 8,192 bits. Runs of this length are practical, but still require about 1 hour of CPU time. Runs that are 10 times this length (in an attempt to estimate lower BER) quickly become impractical.

Even for coded operation, there is usually greater interest in error probabilities lower than 10^{-1} , and it is generally desirable to obtain estimates in the region of $10^{-3}, 10^{-4}$, or even lower. Because such low estimates

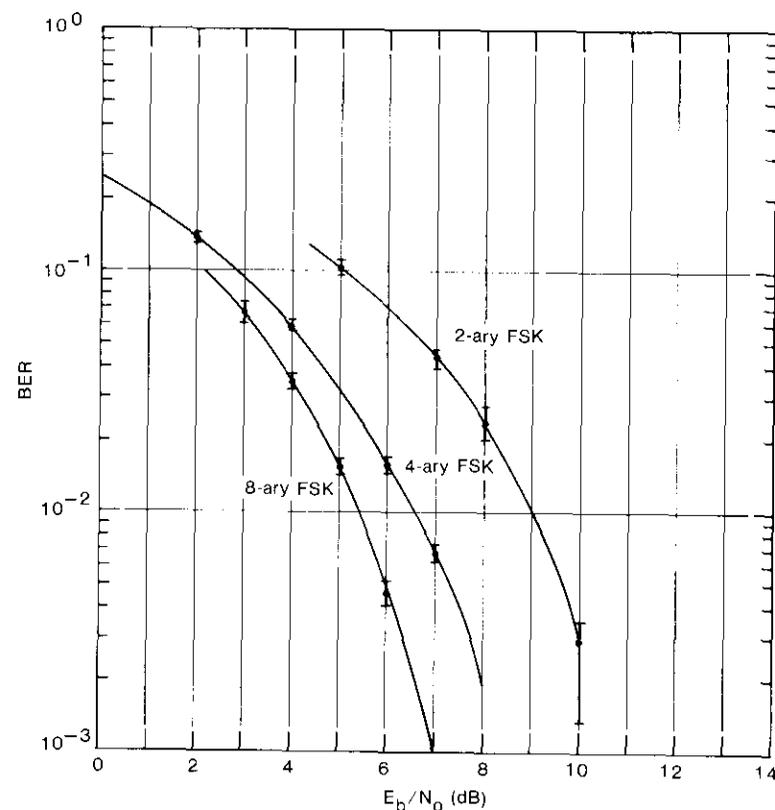


Figure 10. BER Results for FH: White Gaussian Noise Channel

TABLE 2. EXPERIMENTAL RESULTS USING MONTE CARLO SIMULATION

E_b/N_o (dB)	NO. BITS SIMULATED	NO. ERRORS	BER		ERROR (%)	1 σ CONFIDENCE INTERVALS	
			EXPERIMENTAL	THEORETICAL		LOWER	UPPER
1	4,096	733	0.1790	0.1848	-3	0.1781	0.1915
2	4,096	562	0.1372	0.1399	-2	0.1341	0.1457
3	8,192	741	0.09045	0.09772	-7.4	0.0943	0.1012

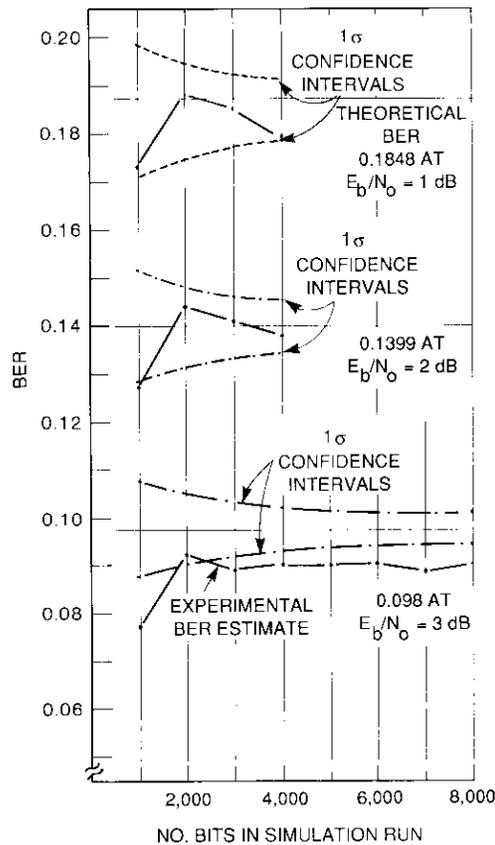


Figure 11. Experimental BER Estimate vs Run Length

cannot be obtained using Monte Carlo methods, alternative techniques must be considered. With the importance sampling technique outlined in the next section, estimates of BER can be extrapolated to much lower values of P_b .

Application of importance sampling

For normal Monte Carlo error-counting simulations, BER is estimated as $\hat{p} = E/N$, where N bits are simulated and E errors are experienced and counted. If p is the true error probability and \hat{p} is the estimate of that probability, then an experiment with N trials should give an expected value of E of $\bar{E} = Np$, and the standard deviation of E would be $\sigma_E = \sqrt{Np(1-p)}$.

If a Monte Carlo simulation experiment is designed to estimate p with a fractional accuracy $\pm a$, and $k\sigma$ confidence is desired in the answer, then

$$k\sqrt{Np(1-p)}/N \leq ap \tag{4}$$

giving the required simulation length

$$N \geq \left(\frac{k}{a}\right)^2 \frac{(1-p)}{p} \tag{5}$$

For example, for ± 10 -percent accuracy ($a = 0.1$) with 67-percent ($k = 1$, 1σ) confidence,[†] the number of bits simulated should be $100(1-p)/p \approx 100/p$.

The $k\sigma$ confidence intervals for the simulation experiment that runs N bits and observes E errors can be obtained as

$$\text{Lower confidence interval} = \frac{E}{N} - \frac{k\sqrt{E}}{N}$$

$$\text{Estimate of } p \hat{=} \hat{p} = E/N$$

$$\text{Upper confidence interval} = \frac{E}{N} + \frac{k\sqrt{E}}{N}$$

If $p = 10^{-3}$ and 100 errors are observed in a simulation run of 10^5 bits, then the estimate, \hat{p} , is $100/10^5 = 10^{-3}$, and the 1σ confidence intervals are 0.9×10^{-3} to 1.1×10^{-3} , which is ± 10 percent of the true BER.

The SPREAD program models an FH system by generating a 4-ary FSK signal in the frequency domain and then displacing this signal randomly over a large frequency range to simulate hopping. One symbol duration of simulated signal is then transformed to the time domain, where noise is added to the I and Q samples. This block of data typically contains 8,192 complex samples.

After the addition of noise, the FH signal is transformed back to the frequency domain, where the hopping is removed. The remaining 4-ary FSK signal is demodulated to produce the data bits. Because of the large number of samples that must be generated for each 4-ary symbol (which conveys 2 bits), simulation of the FH system takes a large amount of CPU time. To experience 100 errors at a BER of 10^{-3} , for example, 10^5 bits or 50,000

[†] $k = 1$ corresponds to 1σ or 63-percent confidence (*i.e.*, two out of three random simulations should be within the required accuracy range); $k = 2$ corresponds to 95-percent confidence (19 out of 20 random simulations should be within the required range); etc.

symbols would have to be simulated. Typically, a day of CPU time is needed on a large computer, and almost 10 days on a workstation, to accurately simulate a BER of 10^{-3} . Thus, it is not practical to obtain direct Monte Carlo simulation results for BERs much below 10^{-2} .

Importance sampling

Importance sampling [12],[13] has been investigated as a means of reducing the length of Monte Carlo simulations for FH systems. With importance sampling, a Monte Carlo simulation run to confirm an error rate of p , which ordinarily would require that N bits be simulated to observe E errors (yielding an estimate $\hat{p} = E/N$), can be run for a much smaller number of bits. To implement importance sampling, the noise input into the simulated demodulator is biased, or increased, so that more errors are produced and counted than normally would be experienced. The bias is computed as the simulation progresses, and this computed weight can then be used to unbiased the error count to correct the experimentally measured error rate.

Importance sampling is implemented using a biased probability density function (pdf), $f^*(n)$, of the noise into the simulated demodulator, rather than the unbiased pdf, $f(n)$. A weight, $w(k)$, is produced for each bit as $w(k) = f^*(n_k)/f(n_k)$, where n_k denotes the random noise numbers that are added to the k th bit. Appropriate processing is performed to obtain the weight if more than one noise number is involved.

In FH simulations, thousands of noise numbers are generated on each simulated symbol. The error count is then "unbiased" by the weight to produce an unbiased BER estimate, as

$$\text{BER estimate} = \frac{1}{mK} \sum_{k=1}^K w(k) E(k) \quad (6)$$

where $E(k)$ is either 1, 2, . . . , or m (the possible number of bit errors on the k th symbol) if symbol k is in error, and 0 otherwise. Each symbol conveys m bits, and K symbols are simulated. This technique was implemented, and the test results are given below.

Simulation results using importance sampling

The simulation of FH systems is a natural application for importance sampling [2] because of the extremely large bandwidth-time products that must be represented and modeled. In the current SPREAD program, each bit is represented by 8,192 samples, so the program takes about 1,000 times as long to run as comparable CHAMP simulations. Also, because the underlying M -ary FSK modulation requires a nonlinear demodulator, it is not possible to

use the hybrid simulation-analysis approach that yields such large time savings in CHAMP.

For example, if a simulation run is desired that will estimate a BER of 10^{-5} , then 10^7 bits must normally be run to experience and count 100 errors. If the input noise is biased to cause errors to occur much more frequently (say if 100 errors are experienced in only 10^4 bits), and if this bias can be measured in the simulation program, then the error count can be corrected to give an estimate of the BER at $p = 10^{-5}$. In this example, if comparable accuracy is achieved, a 1,000-to-1 saving is possible because performance of the 10^{-5} BER level is estimated with a Monte Carlo simulation of length 10^4 bits instead of 10^7 bits.

APPLICATION OF IMPORTANCE SAMPLING

A fairly straightforward way to bias the input noise numbers is to simply increase the noise by using variances σ_2^2 for the noise probability density function $f^*(n_k)$, rather than σ_1^2 , where $\sigma_2^2 > \sigma_1^2$. The simulation is thus made at a relatively low value of E_b/N_o , the bias is computed, errors are counted, and the error count is then corrected to estimate performance. In effect, this extrapolates the results obtained at a relatively high BER (relatively low E_b/N_o) to lower BERs (higher E_b/N_o).

The simulation process is shown in Figure 12. After frequency domain generation, the FH signals are transformed to the time domain, where Gaussian random numbers are added to the block of signal samples. These random numbers have zero mean and a standard deviation of σ_2 . Referring to the complex noise sequence as z ,

$$z \triangleq (x_1, y_1; x_2, y_2; x_3, y_3; \dots x_N, y_N) \quad (7)$$

The joint pdf for N pairs of noise samples is

$$f(z) = \prod_{i=1}^N \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{1}{2} \frac{x_i^2}{\sigma_2^2}\right) \exp\left(-\frac{1}{2} \frac{y_i^2}{\sigma_2^2}\right) \quad (8)$$

Given that noise numbers are drawn from the population with a biased standard deviation, σ_2 , rather than the smaller standard deviation, σ_1 , then a weight can be determined as

$$\begin{aligned} w_i &= \frac{f_{\sigma_1}(x_i, y_i)}{f_{\sigma_2}(x_i, y_i)} = \frac{1/\sigma_2^2}{1/\sigma_1^2} \exp \left[\frac{-(1/2) 1/\sigma_2^2 (x_i^2 + y_i^2)}{-(1/2) 1/\sigma_1^2 (x_i^2 + y_i^2)} \right] \\ &= \frac{\sigma_1^2}{\sigma_2^2} \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) (x_i^2 + y_i^2) \right] \end{aligned} \quad (9)$$

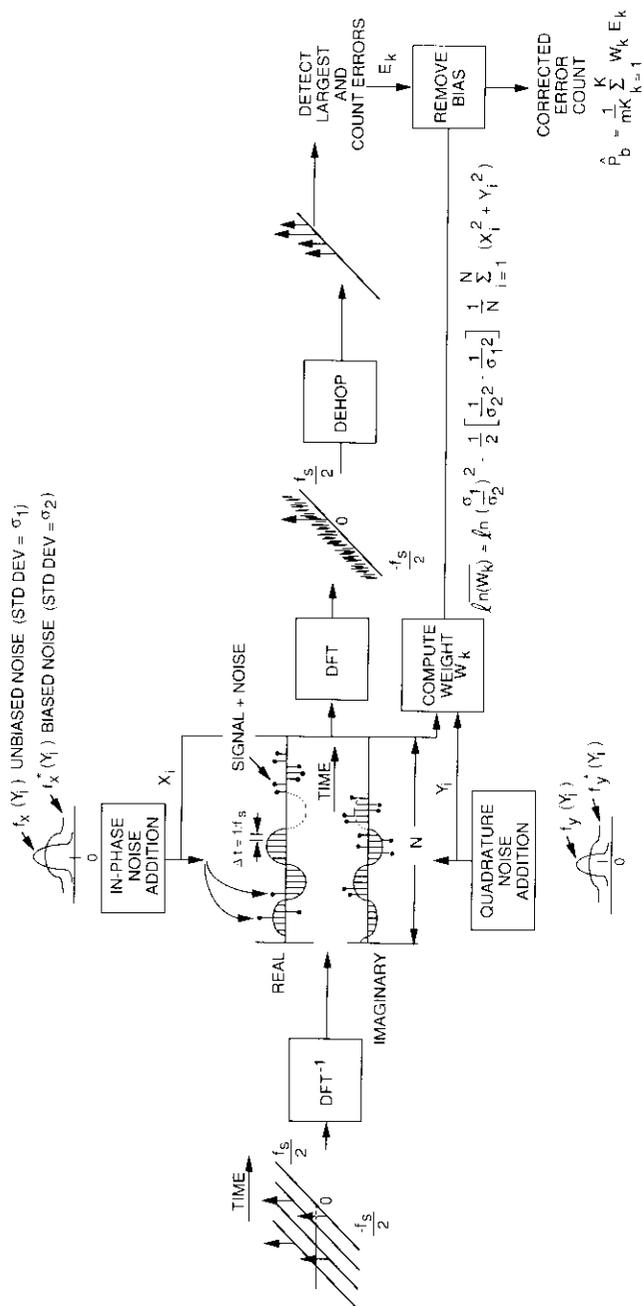


Figure 12. Generation of Successive Blocks of Frequency-Hopping Signal

The weight for the N pairs of noise numbers applied to the samples of the k th symbol is

$$W_k = \prod_{i=1}^N \frac{\sigma_1^2}{\sigma_2^2} \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) (x_i^2 + y_i^2) \right] \quad (10)$$

Taking the natural logarithm gives

$$\ell n(W_k) = 2N \ell n(\sigma_1/\sigma_2) - \frac{1}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^N (x_i^2 + y_i^2) \quad (11)$$

The average weight is obtained by dividing by N to give

$$\overline{\ell n(W_k)} = \ell n(\sigma_1/\sigma_2)^2 - \frac{1}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \frac{1}{N} \sum_{i=1}^N (x_i^2 + y_i^2) \quad (12)$$

from which the weight is determined by taking the inverse of the logarithm.

This weight is computed on each of the k symbols and is used to correct for the bias in the error count. If K symbols are simulated and an error count, E_k , is produced ($E_k = 0$ if no error occurs; $E_k = 1, 2, \dots, m$ if one or more bit errors occurs), then an estimate of BER is

$$\hat{P}_b = \frac{1}{mK} \sum_{k=1}^K W_k E_k \quad (13)$$

where m is the number of bits per symbol. Typically, a direct error count will be made with $W_k = 1$, which corresponds to a conventional Monte Carlo simulation at a low baseline value of E_b/N_o . At the same time, BER estimates can be made for a series of higher values of E_b/N_o by computing the corresponding weights, W_k , on each symbol. These error counts are then unbiased, using the weights, to estimate BER at each of the higher values of E_b/N_o .

SIMULATION RESULTS

Simulation results for 4-ary FSK with one hop per symbol are summarized in Table 3 and plotted in Figure 13 using importance sampling to estimate BER at much lower values of p than are experienced directly in the simulation. The results are somewhat mixed. The first simulation run (run 1, baseline $E_b/N_o = 1$ dB) corresponds to a very noisy case in which 733 error events were recorded. Because of the large number of errors, good accuracy is obtained when compared to the theoretical (solid) curve in Figure 13. This accuracy degrades somewhat for the extrapolated points, but then gives a

TABLE 3. SIMULATION RESULTS USING IMPORTANCE SAMPLING

E_b/N_o (dB)	RUN 1 ^a				RUN 2 ^b				RUN 3 ^c			
	THEORETICAL BER	EXPERIMENTAL BER	ERROR (%)	AVERAGE WEIGHT	EXPERIMENTAL BER	ERROR (%)	AVERAGE WEIGHT	EXPERIMENTAL BER	ERROR (%)	AVERAGE WEIGHT	EXPERIMENTAL BER	ERROR (%)
1	0.1848	0.1790	-3.1	1.0	0.1372	-2	1.0	—	—	9.045 × 10 ⁻²	-7.5	1.0
2	0.1399	—	—	—	—	—	—	—	—	5.547 × 10 ⁻²	-9.8	0.613
3	0.09772	—	—	—	—	—	—	—	—	—	—	—
4	0.06156	0.0333	-46	0.186	4.82 × 10 ⁻²	-21	0.35	—	—	—	—	—
5	0.03395	—	—	—	—	—	—	—	—	—	—	—
6	0.01579	6.54 × 10 ⁻³	-58	0.0365	1.21 × 10 ⁻²	-23	0.088	1.681 × 10 ⁻²	+6.4	1.86 × 10 ⁻¹	+6.4	1.86 × 10 ⁻¹
7	5.914 × 10 ⁻³	2.30 × 10 ⁻³	-61	0.0128	5.04 × 10 ⁻³	-15	0.037	7.96 × 10 ⁻³	+34.6	8.8 × 10 ⁻²	+34.6	8.8 × 10 ⁻²
8	1.684 × 10 ⁻³	6.51 × 10 ⁻⁴	-61	0.0036	1.76 × 10 ⁻³	+4.5	0.013	3.32 × 10 ⁻³	+97	3.67 × 10 ⁻²	+97	3.67 × 10 ⁻²
9	3.394 × 10 ⁻⁴	1.42 × 10 ⁻⁴	-58	7.9 × 10 ⁻⁴	5.0 × 10 ⁻⁴	+47	0.004	1.16 × 10 ⁻³	+241	1.28 × 10 ⁻²	+241	1.28 × 10 ⁻²
10	4.437 × 10 ⁻⁵	2.20 × 10 ⁻⁵	-50	1.23 × 10 ⁻⁴	1.09 × 10 ⁻⁴	+145	7.9 × 10 ⁻⁴	3.27 × 10 ⁻⁴	+637	3.6 × 10 ⁻³	+637	3.6 × 10 ⁻³
11	3.375 × 10 ⁻⁶	2.24 × 10 ⁻⁶	-34	1.25 × 10 ⁻⁵	1.68 × 10 ⁻⁵	+397	1.2 × 10 ⁻⁴	7.14 × 10 ⁻⁵	+2,016	7.9 × 10 ⁻⁴	+2,016	7.9 × 10 ⁻⁴
12	1.305 × 10 ⁻⁷	1.34 × 10 ⁻⁷	-3	7.4 × 10 ⁻⁷	—	—	—	1.10 × 10 ⁻⁵	+8,329	1.2 × 10 ⁻⁴	+8,329	1.2 × 10 ⁻⁴

^a $E_b/N_o = 1$ dB, 4,096 bits, 733 error events.
^b $E_b/N_o = 2$ dB, 4,096 bits, 562 error events.
^c $E_b/N_o = 3$ dB, 8,192 bits, 741 error events.

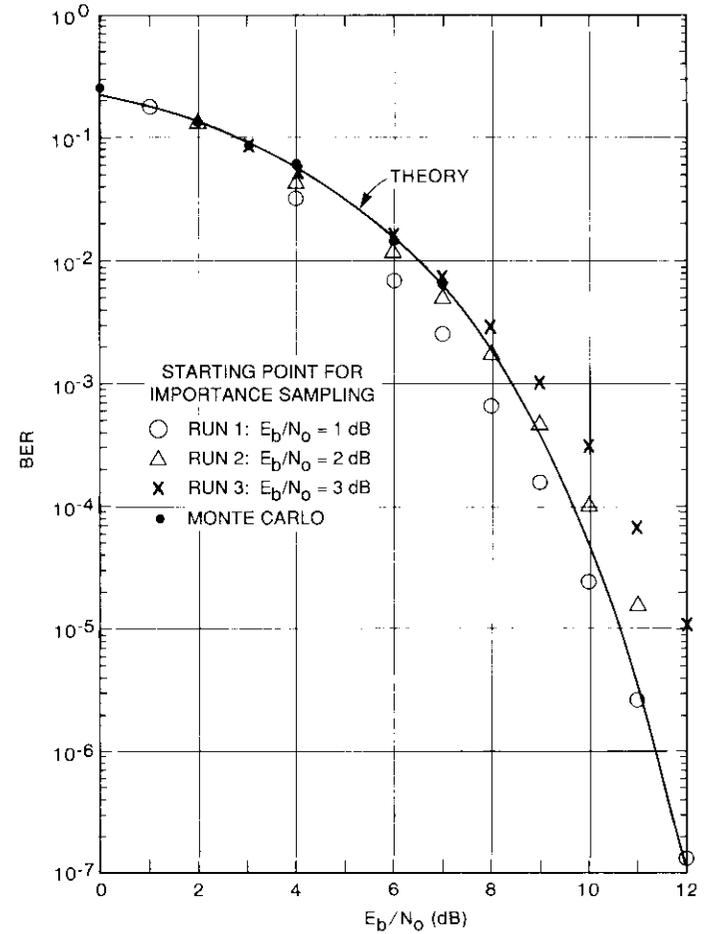


Figure 13. Simulated BER Results vs E_b/N_o .

surprisingly accurate estimate at $p \approx 10^{-7}$. To achieve 3-percent accuracy at a BER of 10^{-7} , about 700 errors (requiring 7×10^9 bits) would have to be recorded. At 1 bit/s, such a simulation would take *several hundred years*.

Run 2 in Table 3 also gives reasonable results starting at a baseline E_b/N_o of 2 dB; however, accuracy becomes poor for the extrapolated points at BER lower than 10^{-4} . Similarly, the accuracy of run 3, which has a baseline E_b/N_o of 3 dB, degrades rapidly for P_b lower than 10^{-3} . For this particular set of experiments, no single baseline E_b/N_o value gives uniformly low errors in estimating the lower BER value at higher E_b/N_o .

A general conclusion based on these simulation experiments with importance sampling is that ± 1 -dB accuracy is achieved down to a BER of 10^{-3} to 10^{-4} . For lower error rates, results tend to be very erratic. The technique was applied here to a channel for which the answer can be computed analytically (*i.e.*, the answer is known). Much more experience is needed with this technique before it can be applied with confidence to more general channels for which answers are not known.

FAST-HOPPING SIMULATION RESULTS

The simulation model has been modified to implement L chip diversity for 4-ary and 8-ary FSK. Simulations were run for $L = 4$ to test the program. The accumulated outputs of the matched filters, Z_i ($i = 1, 2, \dots, m$) over the L chips, where the individual in-phase and quadrature outputs are denoted by X_{it} , Y_{it} , are formed as

$$Z_i = \sum_{t=1}^L (X_{it}^2 + Y_{it}^2) \quad (14)$$

The matched filter with the largest output is selected, and the $m = \log_2(M)$ information bits that correspond to that filter are provided as output. Although this strategy is not necessarily the best approach with partial band jamming, it was used to validate the simulation model for a case where theoretical results can be calculated.

Figure 14 shows simulation results for 4-ary FSK with four hops per symbol. Results for 8-ary FSK with four hops per symbol, using importance sampling, are given in Figure 15. For both of these tests, importance sampling gives reasonably accurate results when BER is extrapolated to a value that is one or two orders of magnitude lower than that experienced directly in the Monte Carlo simulation at a high baseline E_b/N_o .

Conclusions

A special simulation program (SPREAD) has been developed and tested to model and simulate spread spectrum systems. Both DSPN and FH systems can be simulated. The program has been successfully transferred to a workstation environment, where it runs at about one-tenth the speed as on COMSAT's mainframe computer. For a DFT block size of 8,192 (equivalent to a $W_s T_b$ product of 1,024), the simulation runs at about 0.1 bit/s on the workstation. Increasing the block size by a factor of 2 or even 4 might be feasible (with a corresponding increase in simulation time).

The model can be used to investigate BER performance for DSPN systems by employing existing hybrid simulation/analysis techniques that process a

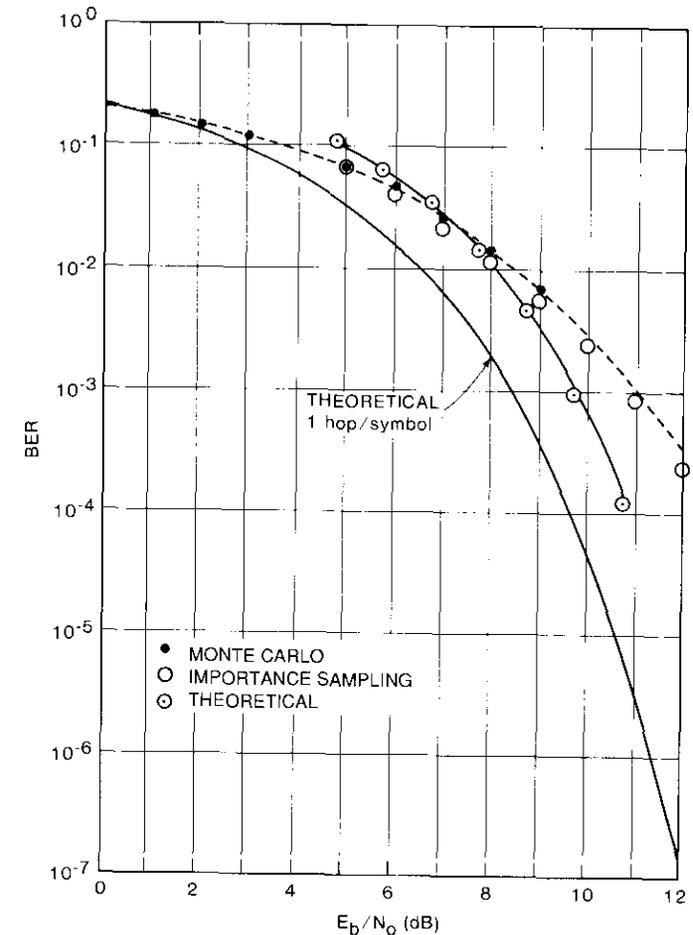


Figure 14. Simulated P_e vs E_b/N_o Results for 4-ary FSK With 4 hops/symbol

scatter diagram. For FH systems, with a highly nonlinear demodulation process, superposition does not hold and Monte Carlo (error-counting) simulations must be used. Here, importance sampling appears to offer the potential to obtain reasonably accurate BER estimates with reasonable simulation lengths. More experimentation is needed to gain confidence in this technique.

It is anticipated that SPREAD will have applications other than for end-to-end performance estimates. For example, by adding signal and spectral

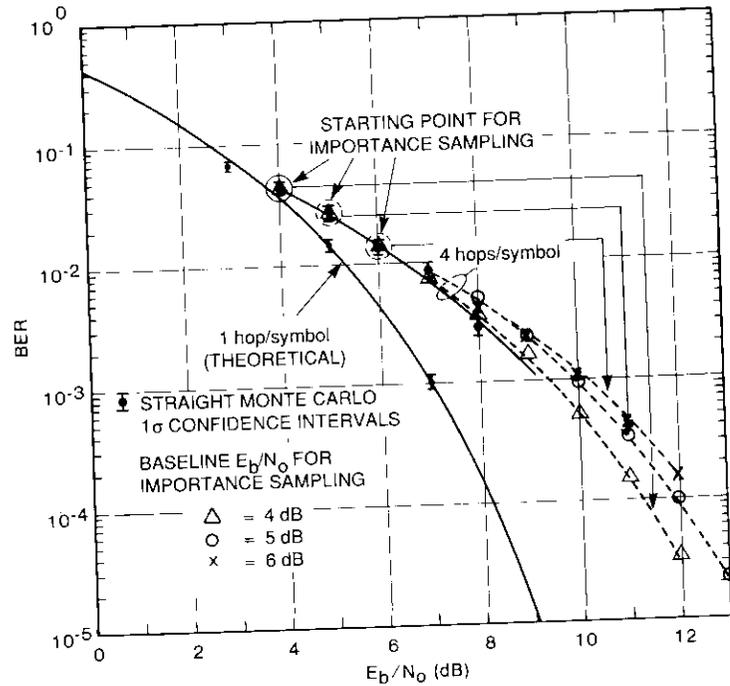


Figure 15. Simulated BER vs E_b/N_o for 8-ary FSK With 4 hops/symbol

analysis elements, the program can be used to evaluate signal detectability. By implementing explicit PN code tracking algorithms, it could also be used to study particular receiver implementations.

Acknowledgments

The authors wish to acknowledge several helpful discussions with K. Mackenthun on the implementation of importance sampling.

References

- [1] *IEEE Journal on Selected Areas in Communications*, Special Issue on Computer-Aided Modeling, Analysis, and Design of Communication Systems, Vol. SAC-2, No. 1, January 1984.
- [2] *IEEE Journal on Selected Areas in Communications*, Special Issue on Computer-Aided Modeling, Analysis, and Design of Communication Systems II, Vol. SAC-6, No. 1, January 1988.

- [3] J. Dugundji, "Envelopes and Pre-Envelopes of Real Waveforms," *IRE Transactions on Information Theory*, Vol. IT-4, No. 1, March 1958, p. 53.
- [4] K. R. Matis and J. W. Modestino, "Interactive Communication System Simulation on a High-Speed PC Workstation," *IEEE Journal on Selected Areas in Communications*, Special Issue on Computer-Aided Modeling, Analysis, and Design of Communication Systems II, Vol. SAC-6, No. 1, January 1988, pp. 24-33.
- [5] M. M. Drossman *et al.*, "Digital Computer Simulation of Spread Spectrum Communications Systems," International Conference on Communications, Denver, Colorado, June 1981, *Conf. Rec.*, Vol. 3, Paper 45.4.
- [6] M. E. Mortara and P. K. Leong, "ICSSM—A Computer-Aided Modeling and Simulation Tool for Communication Systems," IEEE Global Telecommunications Conference, San Diego, California, November/December 1983, *Conf. Rec.*, Vol. 3, Paper 46.2.
- [7] K. S. Shanmugan and V. S. Frost, "Simulation of Spread-Spectrum Systems Using ICSSM," IEEE Global Telecommunications Conference, San Diego, California, November/December 1983, *Conf. Rec.*, Vol. 3, Paper 46.3.
- [8] G. Liu and J. Rowe, "A Computer Simulation of a Jammed Direct Sequence Spread Spectrum Communications Link," National Telecommunications Conference, Washington, D.C., November 1979, *Proc.*, Vol. 1, Paper 15.6.
- [9] W. L. Cook, "Interactive Computer Simulation of Satellite Transmission Systems," 5th Annual Pittsburgh Conference on Modeling and Simulation, Pittsburgh, Pennsylvania, April 1974, *Proc.*, Vol. 5, pp. 867-872.
- [10] L. Palmer, "Computer Modeling and Simulation of Communications Satellite Channels," *IEEE Journal on Selected Areas in Communications*, Special Issue on Computer-Aided Modeling, Analysis, and Design of Communication Systems, Vol. SAC-2, No. 1, January 1984, pp. 89-102.
- [11] *IEEE Transactions on Communications*, Special Issue on Spread Spectrum Communications, Vol. COM-25, No. 8, August 1977.
- [12] K. S. Shanmugan and P. Balaban, "A Modified Monte-Carlo Simulation Technique for the Evaluation of Error Rate in Digital Communication Systems," *IEEE Transactions on Communications*, Vol. COM-28, No. 11, November 1980, pp. 1916-1924.
- [13] M. C. Jeruchim, "Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems," *IEEE Journal on Selected Areas in Communications*, Special Issue on Computer-Aided Modeling, Analysis, and Design of Communication Systems, Vol. SAC-2, No. 1, January 1984, pp. 153-170.
- [14] J. J. Jones, "Hard-Limiting of Two Signals in Random Noise," *IEEE Transactions on Information Theory*, Vol. IT-9, No. 1, January 1963, pp. 34-42.
- [15] S. W. Houston, "Modulation Techniques for Communication, Part 1: Tone and Noise Jamming Performance of Spread Spectrum M-ary FSK and 2, 4-ary DPSK Waveforms," IEEE National Aerospace and Electronics Conference, Dayton, Ohio, June 1975, *Proc.*, pp. 51-58.



Larry C. Palmer received a B.S. from Washington and Lee University and a B.E.E. from Rensselaer Polytechnic Institute, both in 1955, and an M.S. and Ph.D. in electrical engineering from the University of Maryland in 1963 and 1970, respectively. From 1955 to 1957, he served in the U.S. Army Signal Corps as an Electronics Officer. Following this, he spent 6 years with the Radcom-Emertron Division of Litton Systems, where he was engaged in the design and development of airborne electronics systems. In 1963, he joined IIT INTELCOM and worked on the military communications

satellite system. After the acquisition of INTELCOM by Computer Sciences Corporation in 1965, he was involved in systems analysis, modeling, and simulation related to communications and navigation satellite systems.

Dr. Palmer joined COMSAT Laboratories in 1974 and is currently Principal Scientist in the Communications Technology Division. He served as Program Manager for COMSAT's participation in the DARPA SATNET program, and was Systems Engineering Project Manager for the NASA Advanced Communications Technology Satellite (ACTS) program under subcontract to RCA. He is a Senior Member of IEEE, and a member of the IEEE Communications Society, the IEEE Information Theory Group, Tau Beta Pi, and Eta Kappa Nu.

Peter Y. Chang received a B.S.E.E. in 1966 from Cheng Kung University, Taiwan, China, and an M.S.E.E., M.Ph., and Ph.D. from Columbia University in 1969, 1972, and 1975, respectively. He joined COMSAT Laboratories in 1976 and is currently a Staff Scientist in the Systems Modeling and Analysis Department of the Communications Technology Division. His work at COMSAT has included TDMA and FDMA modulation, digital signal transmission, intermodulation analysis involving the nonlinear effect of transponders, direct-broadcast (TV, video, and audio) satellite transmission systems, intersatellite links, simulation of satellite marine and mobile communications systems with fading channels, and simulation of digital signal performance in satellite channels involving various modulation and coding schemes. Recently, he was involved in the NASA ACTS program, including low-bit-rate system study analysis and simulation, and the analysis, modeling, and simulation of rain fade effects. He has also worked on the modeling and simulation of a TDMA random-access-with-notification network for the VSAT system. He is currently involved in the study, simulation, and analysis of spread spectrum techniques.



Prior to joining COMSAT, Dr. Chang was with the Satellite Communications Division of Computer Sciences Corporation. He is a member of IEEE, Eta Kappa Nu, and Sigma Xi.

Index: digital transmission; mobile communication systems; phase-locked systems; signal processing; modulation, demodulation, modems

Digital signal processor-based programmable BPSK/QPSK/offset-QPSK modems

L-N. LEE, A. SHENOY, AND M. K. ENG

(Manuscript received October 4, 1989)

Abstract

The generalized hardware and software algorithm designs of a class of digital signal processor (DSP)-based programmable modems are described. These low-cost modems can operate at bit rates from a few hundred bits per second to several hundred kilobits per second in wide-ranging applications such as small earth stations for thin routes, and single-channel-per-carrier land, maritime, and aeronautical mobile terminals. Modulation formats of BPSK/QPSK/offset-QPSK, in burst and continuous modes, were implemented in the prototype modems. To accommodate the wide range of bit rates, a multiprocessor architecture that can be adapted to various application-specific configurations was devised. Actual hardware implementations of this architecture for candidate applications are discussed, and the software algorithms for fast burst- and continuous-mode acquisition are described. Typical bit error rate performance test results on an additive white Gaussian noise channel are given for the candidate applications. These results demonstrate very good correspondence with the theoretical predictions.

Introduction

In a power-limited satellite communications channel, phase-shift-keyed (PSK) modulated signals with coherent demodulation, particularly when combined with channel coding and soft-decision decoding, have proved to

be extremely power-efficient for the reliable transmission of digital information [1]. Each of the various PSK-based modulation techniques, including binary PSK (BPSK), quaternary PSK (QPSK), minimum shift keying (MSK), and offset-QPSK (O-QPSK), has unique transmission properties that are best suited for a specific channel environment. For example, BPSK is more robust in the presence of severe phase noise, and is therefore most appropriate for low-bit-rate applications such as in the INMARSAT Standard-C service at 1,200 bit/s [2]. QPSK, which can be efficiently filtered to achieve 1.4- to 1.5-bit/s/Hz without degradation from optimal performance, is the most bandwidth-efficient and thus finds wide application in almost all types of digital communications up to several hundred megabits per second.

MSK, which can be considered as a class of O-QPSK with constant envelope, is especially attractive when the channel is highly nonlinear.

O-QPSK, which has less amplitude fluctuation than QPSK and yet is generally more bandwidth-efficient than MSK (with proper filtering), is a good compromise between QPSK and MSK and finds application in the INMARSAT Standard-B and high-rate aeronautical voice services, where the ability to achieve both moderate bandwidth efficiency and power efficiency in the presence of channel nonlinearities is important [3].

In the past, due to mismatch of the transmit and receive filters implemented with analog techniques, filtered PSK transmissions typically exhibited an implementation loss of 0.5 to 2 dB (due to component tolerances and other variations), even if the channel was perfectly linear. Additional throughput inefficiencies may also be experienced in time-division multiple access (TDMA) operations due to the need to send long preambles for burst acquisition.

Through the use of digital signal processing technology, power and bandwidth efficiencies can be substantially improved in terms of both the filter mismatch loss and the overhead needed for burst-mode acquisition. For low and medium data rates up to a few hundred kilobits per second, digital implementations have also been more cost-competitive in recent years because of the availability of inexpensive digital signal processors (DSPs) such as those in the Texas Instruments (TI) 320 family. In addition to improvements in efficiency, digital implementation also allows for modulation and spectral shaping programmability such that a general hardware design can be used to implement almost all PSK modulation types with a family of filtering characteristics for bit rates ranging from several hundred bits per second up to several hundred kilobits per second.

Based on the above concept, a number of single, dual, and multiprocessor-based PSK modems of this type were developed. This paper describes the generalized architecture of these modems and briefly discusses specific implementations that achieve high speed and fast burst-mode acquisition.

Typical bit error rate (BER) test results show that these digitally implemented modems perform very well in an additive white Gaussian noise (AWGN) channel, with negligible implementation loss.

Generalized DSP-based BPSK/QPSK/O-QPSK modem architecture

Figure 1 is a block diagram of a generalized DSP-based modem. At the modulator side, serial digital data are converted into modulated samples through the digital modulator. The modulated samples are then converted into analog waveforms by a digital-to-analog (D/A) converter at a convenient, fixed, low intermediate frequency (IF) or at baseband, depending on the application. To interface with standard IF equipment, and to facilitate channel selection, this low-IF or baseband signal is converted to a standard IF via an analog up-converter and channel selector.

At the demodulator side, the analog signal is first down-converted to a fixed, low IF or to baseband via an analog down-converter and channel selector. The resultant signal is then sampled by an analog-to-digital (A/D) converter. Certain applications require complex in-phase and quadrature (I&Q)

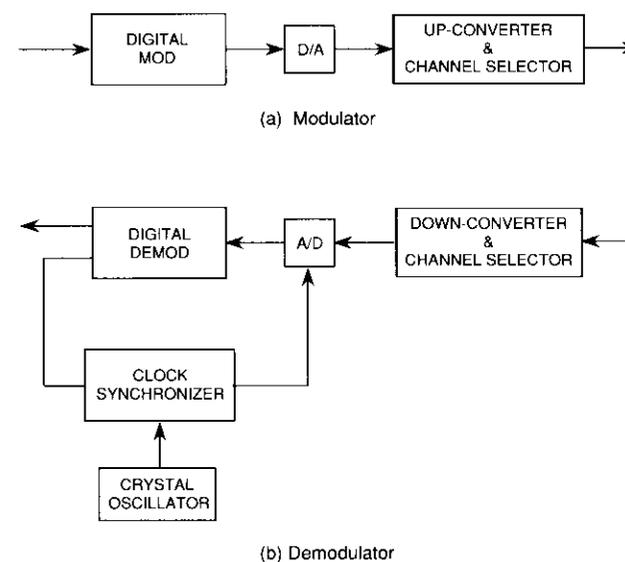


Figure 1. Block Diagram of a Generalized DSP-Based Modem

demodulator inputs at baseband, in which case two A/D converters are used. The quantized samples out of the A/D converter are processed by the DSP chips for carrier recovery, automatic gain control (AGC), and other demodulation functions, whose final outputs are then the demodulated digital data. A digital clock synchronizer driven by the clock recovery algorithm of the digital demodulator is used to synchronize the sampling clock.

This general modem structure is designed to keep the sampling frequency of the digital modulator and demodulator to only a small multiple of the modulation symbol rate (typically 4 to 16 samples/symbol), thus minimizing the speed of the critical digital hardware, including the DSPs and D/A and A/D converters. The optional low-IF frequency is generally an integer multiple of the sampling frequency. Considerations for the fixed low-IF selection are somewhat different for the modulator and the demodulator.

In general, the modulator algorithm is relatively straightforward and can be implemented with digital logic and a lookup table for bit rates up to several hundred kilobits per second. Thus, the digital modulator hardware can be operated at these speeds without substantial hardware complexity. A very low IF at the transmit side, however, causes image rejection difficulties in the up-converter design. Therefore, practical low IFs tend to be in the range of hundreds of kilohertz to several megahertz.

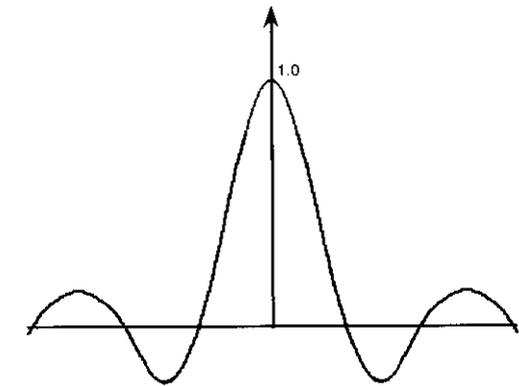
At the receive side, image rejection is not a problem in the down-converter design because the conversion local oscillator is in the center of the IF band. Hence it is desirable to keep the IF as low as possible, sometimes even at baseband, which may lead to a design with different IF frequencies at the transmit and receive sides. Since the modem interfaces with external RF equipment at a standard IF at the modulator output and demodulator input (typically 70 MHz), the use of different internal transmit and receive low-IF frequencies is not a problem.

Sampling considerations

A theoretical minimum of two samples per carrier cycle is needed to faithfully represent the modulated signal. At the modulator side, four samples per carrier cycle are chosen because at this frequency a sinusoidal wave can easily be represented by the samples $+1, 0, -1, 0$. Modulating the carrier (sampled using these values) by the baseband signal is straightforward with medium-scale integration digital hardware, thus eliminating the need for multiplication and table lookup. Further, successive samples are exactly 90° apart, so that the quadrature signal has non-zero samples only when the in-phase signal has zero values. Thus, adding two quadrature signals is equivalent to time-multiplexing these two signals sample-by-sample.

Modulator structure

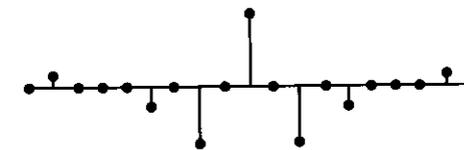
Once the sampling rate is determined, the modulator processing can be represented as shown in Figure 2. One bit of digital data, which is represented as a signal impulse, is first passed through a waveform-shaping filter to produce an impulse response (Figure 2a), and then modulated by a sinusoidal waveform (Figure 2b), resulting in a modulated response (Figure 2c). Since the impulse response has significant non-zero values over a finite number of symbols, the modulated signal during a particular symbol interval is affected by the impulse response of a finite number of neighboring symbols. Typically, for a waveform-shaping filter with a reasonable rolloff factor (40 to 60



(a) Impulse Response Envelope of the Spectral Shaping Filter



(b) Modulating Carrier



(c) Samples After Modulation and Filtering

Figure 2. Time-Series Representation of the Modulation and Filtering Process

percent), the impulse response has a practical duration of no more than 7 to 9 symbols. Therefore, all possible waveform shapes representing a symbol over a symbol duration can be stored as a function of the 7 to 9 adjacent data symbols that affect the outcome.

Figure 3 depicts a generalized implementation of the modulator using this approach, with QPSK as the representative modulation. The serial data are first serial-to-parallel converted to I&Q streams. Each data stream is then fed into a shift register. The contents of the shift register address a lookup table, which outputs the modulated samples. The I&Q components from the lookup tables are then time-multiplexed to produce the final output. For O-QPSK, the timing of the two shift registers is offset by half a symbol. For BPSK modulation, each serial data bit is routed to both the I and Q channels. Alternatively, the quadrature component can be disabled. However, zeroing the quadrature component reduces the output level by 3 dB compared to the QPSK or O-QPSK case.

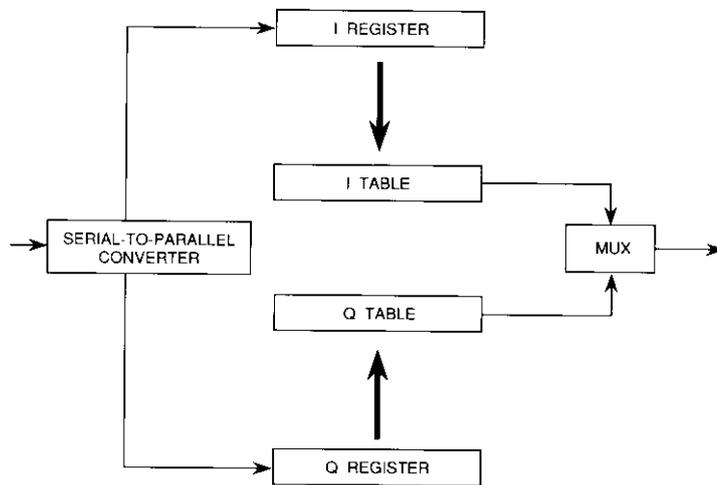


Figure 3. General Structure of the Digital Modulator

Demodulator structure

The selection of four samples per carrier cycle does not necessarily provide similar computational advantages on the demodulator side, since the demodulator must track the slightest frequency and phase differences between the transmitted signal and receive-side local references. However, it is advan-

tageous to choose the sampling rate to be an even multiple of the symbol rate of the modulated signal, because the demodulator requires a minimum of two samples per symbol—one at the maximum eye opening for data detection, and the other at the data transition for clock recovery. (For O-QPSK, data transitions in the in-phase channel occur at the same point as the maximum eye openings in the quadrature channel, and vice versa.) Since the modulated signal has a bandwidth about 1.5 times greater than the symbol rate, and the noise bandwidth of the signal input (for square-root Nyquist pulse-shaping) is equal to the symbol rate (R_s), a sampling rate of four samples per symbol is selected to avoid aliasing.

The demodulator assumes one of two general structures, depending on the symbol rate. When the modulation symbol rate is low, a standard configuration with baseband matched filters, as shown in Figure 4a, is used. QPSK is again used as the representative modulation. A numerically controlled oscillator (NCO) generates samples of the carrier phase reference, ϕ . The sine and cosine values of ϕ are multiplied by the incoming signal, and the resultant samples are filtered by digital low-pass filters (LPFs) whose impulse responses are matched to the transmit pulse-shaping filter. The output samples corresponding to maximum eye openings are used for data detection, while those corresponding to data transitions are used for symbol clock tracking. For O-QPSK demodulation, the baseband in-phase component is delayed by half a symbol to compensate for the half-symbol delay imposed on the quadrature component at the transmit side.

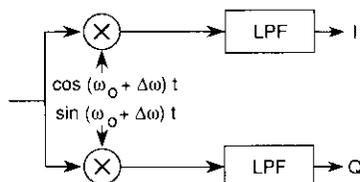
Since the computational requirement increases in proportion to the sample rate (and hence the modulation symbol rate, since the sample and symbol rates are directly related), a demodulator structure that requires significantly reduced computation is desirable for higher bit rates. The structure shown in Figure 4b employs a fixed bandpass digital filter (BPF) for each of the matched filters for the I&Q components [4]. When the sampling frequency equals four times the center frequency, half of the finite impulse response (FIR) filter coefficients are zero, which reduces the computation for the filtering by a factor of 2. The matched filter operation is computationally the most demanding in the demodulator, and the BPF structure reduces these computational requirements substantially.

This latter approach achieves computational savings only if the actual received carrier frequency is very close to the center frequency of the bandpass filter (*i.e.*, one-fourth of the sampling frequency), to avoid spectral truncation. Since the sampling frequency is four times the symbol rate, the carrier frequency must be very close to the symbol rate (typically ± 10 percent). For higher bit rates, the frequency error derived from the RF subsystem is usually only a few percent of the symbol rate, and thus the bandpass

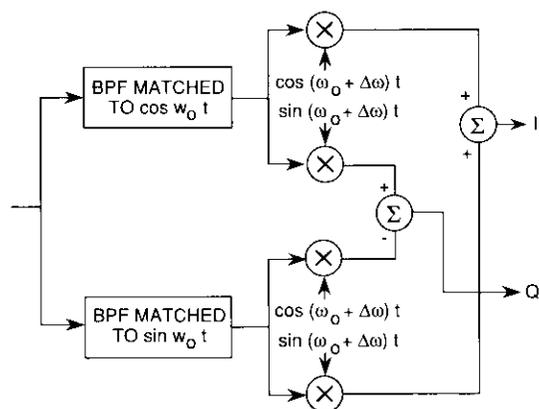
implementation is feasible. This approach is not suitable for low symbol rate applications where the frequency offsets are comparable to the symbol rate, which leads to significant spectral truncation in the bandpass matched filters.

With the BPF approach, the residual frequency error is corrected by an NCO with a mixing configuration that effectively performs a single-sideband demodulation. The mixing configuration needs to be performed only twice per symbol, compared to four times per symbol in the baseband approach of Figure 4a. Therefore, there is no increase in the total number of multiplications per symbol, although four mixers instead of two are required.

There is usually substantial overhead associated with the service routines for handling processor interrupts, which are the usual form for receiving incoming samples. [Later-generation DSP devices now offer the option of direct memory access (DMA) data transfers.] It is generally desirable to



(a) Mixing Section of a Conventional Quadrature Carrier Demodulator Structure



(b) Quadrature Carrier Front-End Requiring Less Computation Using Bandpass Matched Filters

Figure 4. Alternative Demodulator Structures

include a small first-in first-out (FIFO) buffer between the A/D converter and the DSP(s) to reduce the need for processor interrupts to once per symbol. This transfer rate was chosen because the symbol rate is a natural period for all demodulator operations. Additionally, an unnecessarily long delay may affect the performance of the clock and carrier recovery loops.

When the symbol rate is higher than can be accommodated by the processing power of a single DSP, a number of processors operating in parallel are used to increase the speed. Figure 5 illustrates a general architecture for a multiprocessor demodulator based on the TI TMS32020 DSP device. Each DSP contains its own erasable programmable read-only memory (EPROM), random-access memory (RAM), and RAM storage area. High-speed (25-ns) static data and program RAMs are used for the DSP data and program memories so that the external memory is addressed without "wait" states. On power-up, a boot program is executed from the EPROM that downloads the running program from the EPROM into the faster program RAMs. For time-critical routines that contain special instructions which require using the internal memory, sections of the running program are read into the internal RAM storage of the DSP, where the instructions execute at full speed. Extensive use is made of programmable logic devices in order to derive a compact design for the DSP peripheral logic, such as the address decoders, memory control signals, port decoders, and I/O interfacing.

Interprocessor communications, primarily consisting of block data transfers and the exchange of control/status information between two or more DSPs and a host processor (if any), take place through a high-speed (15-ns) global RAM. Software-driven I/O flags are used by the individual processors to signal when new data are available in the global memory. The global memory is mapped into the identical data memory address space in each processor. Address bus buffers and bidirectional data bus transceivers isolate the global memory bus from the processor buses.

A bus arbitration circuit monitors global memory accesses to prevent bus contention. If a current instruction addresses a data operand in the global memory, the DSP asserts a "Bus Request" to request control of the global memory bus. The bus arbitrator grants access based on the priority of the requesting processor. The individual processor priorities are determined in the system design according to the need for, and frequency of, data interchanges. The "Ready" line is used by the bus arbitrator to signal when the global memory bus may be accessed. The number of computations for each task in the demodulation process is quite predictable, once the demodulation algorithms are established. Thus, it is generally possible to schedule the workload of each processor so that bus contentions for global memory access can be avoided. In the case of simultaneous global memory access

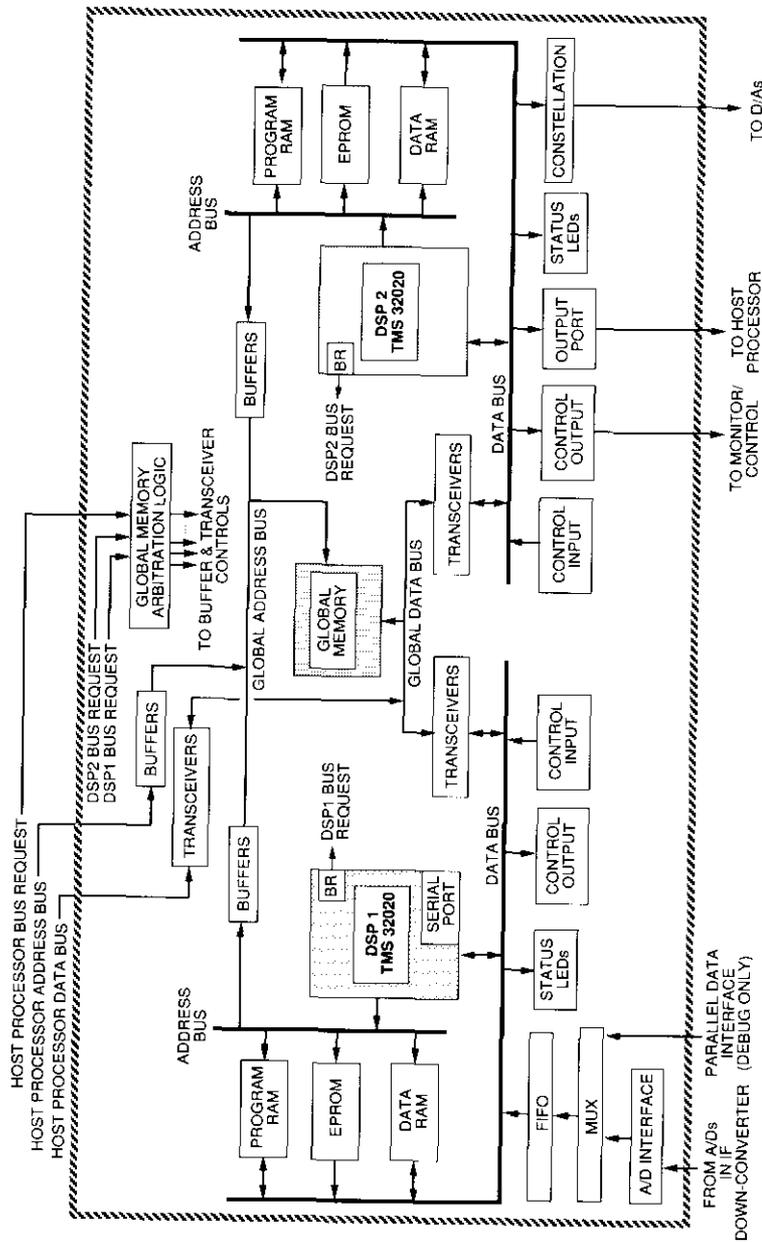


Figure 5. Block Diagram of a Dual-Processor-Based DSP Demodulator

requests, the "Ready" line is de-asserted to cause lower-priority processors to wait.

Multimode operation of the demodulator is achieved by storing the running programs for each mode (BPSK, QPSK, and O-QPSK) as three program blocks in one EPROM. A change in the operating mode, originating from either a control interface or a front panel switch, interrupts the processors. The interrupt routine reads the bit pattern from a control port, which indicates the new operational mode. The processors then re-initialize, download the appropriate EPROM block into the program RAM, and begin execution from the RAM.

Digital models for the demodulator functions

The key processing functions for continuous operation of the demodulator include the matched filters, the carrier tracking loop, and the clock tracking loop. Digital models and implementation techniques for each of these processes are described below. The AGC function is generally implemented at the IF stages where the AGC signal is derived from the average noise floor. For digital processing of the received signal, a digital form of gain control or scaling may be needed. The scaling function is described later in this paper.

Nyquist matched filters

Square-root Nyquist filters were selected for the matched filters because, when cascaded, these filters have no intersymbol interference at the detection sampling instant in a linear channel. A closed form of the impulse response of the square-root Nyquist filter can be obtained by evaluating the inverse Fourier transform of its frequency response. A convenient form of the frequency response is given as [5]

$$P_r(f) = \begin{cases} T_s & |f| \leq \frac{R_s}{2} - \beta \\ T_s \cos^2 \left[\frac{\pi}{4\beta} \left(|f| - \frac{R_s}{2} + \beta \right) \right] & \frac{R_s}{2} - \beta < |f| \leq \frac{R_s}{2} + \beta \\ 0 & |f| > \frac{R_s}{2} + \beta \end{cases} \quad (1)$$

where R_s is the symbol rate, $T_s = 1/R_s$ is the symbol duration, and β is the rolloff factor expressed with respect to $R_s/2$. Let $\alpha = (R_s/2) - \beta$ and $\delta = (R_s/2) + \beta$. Then the time domain response is obtained by taking the inverse Fourier transform, as

$$\hat{P}_s(t) = \frac{4\beta}{(4\beta + \pi\alpha)(1 - 64\beta^2 t^2)} [\cos(2\pi\delta t) + 8\beta t \sin(2\pi\alpha t)] + \frac{\pi\alpha}{(4\beta + \pi\alpha)} \text{sinc}(2\pi\alpha t) \quad (2)$$

The tap coefficients for the matched filters can now be obtained by replacing t with $\pm nT_{\text{samp}}$, where n is the number of samples on either side of the peak (0, 1, 2, ...) that are required to encompass the desired response duration (7 to 9 symbol periods), and T_{samp} is the sampling interval. For the DSP implementation, these coefficients are conveniently quantized to 16 bits. This high resolution allows better suppression of the undesirable sidelobes in the matched filter implementation.

The impulse response can be calculated and truncated when the amplitude of ringing becomes negligible (usually on the order of 1 percent). At four samples per symbol, the baseband matched filters are implemented as 29-tap FIR filters in the DSP. As mentioned previously, about half of the coefficients in the bandpass matched filter implementation are zero, which effectively reduces the computational length of the FIR filters by about one half (*i.e.*, 15 taps for the filter matched to the cosine phase, and 14 taps for the filter matched to the sine phase).

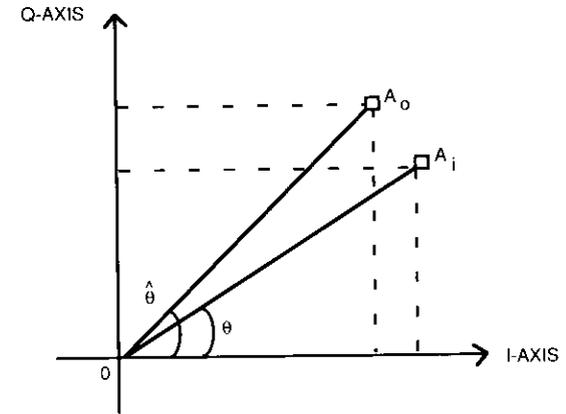
Carrier tracking loop

Carrier recovery is implemented using a decision feedback loop which is updated once per symbol. \hat{I} and \hat{Q} samples at the assumed midsymbol points are used to generate the phase error estimates. With \hat{I} and \hat{Q} being the reference coordinates of the ideal signal point in the quadrant indicated by the received I and Q , the quantity $\hat{I}Q - \hat{Q}I$ is used as the phase detection error signal input to the loop filter. From trigonometric relations, one can write

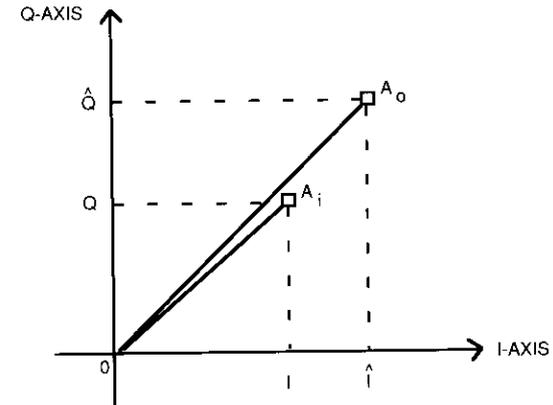
$$\sin(\theta - \hat{\theta}) = \sin\theta \cos\hat{\theta} - \cos\theta \sin\hat{\theta} \quad (3)$$

With reference to Figure 6, equation (3) can be written as

$$\sin(\theta - \hat{\theta}) = \frac{\hat{Q}I - I\hat{Q}}{\sqrt{I^2 + Q^2} \sqrt{\hat{I}^2 + \hat{Q}^2}} \quad (4)$$



(a)



(b)

Figure 6. Vector Diagrams for Calculating K_d

For a small phase error $\Delta\theta = \theta - \hat{\theta} = \sin(\theta - \hat{\theta})$, or

$$\Delta\theta = \frac{Q\hat{I} - I\hat{Q}}{\sqrt{I^2 + Q^2}\sqrt{\hat{I}^2 + \hat{Q}^2}} \quad (5)$$

or

$$\Delta\theta = \frac{Q\hat{I} - I\hat{Q}}{A_i A_o} \quad (6)$$

where A_i and A_o are the received and reference signal amplitudes, respectively. The phase error voltage that is input to the loop filter is

$$V_d = Q\hat{I} - I\hat{Q} \quad (7)$$

Using equation (7) in equation (6), gives

$$V_d = A_i A_o \Delta\theta \quad (8)$$

Since A_o is a constant, the phase detector gain is proportional to the input signal level. The phase detector gain in the conventional sense is

$$K_d = \sqrt{2E_s} \quad (9)$$

where A_o is assumed to be composed of $\hat{I}, \hat{Q} = \pm 1$, and $A_i^2 = E_s$.

The NCO is implemented with a lookup table that has L entries, where L is typically a power of 2. For $L = 256$, the table stores the value of the cosine function in 1.4° steps. A pointer is used to address the table two times per symbol (at the symbol sampling and symbol transition instants for the carrier and clock loops) to obtain the values of the cosine function for phase error corrections. Because the sine function lags the cosine function by 90° , another pointer that is offset by $L/4$ entries is used to obtain the sine values. Alternatively, the sine and cosine functions may be stored as pairs in the table, so that the pointer can access both values within a single increment operation.

If the cosine pointer is incremented by 1 twice per symbol, $L/2$ symbols will be required in order to complete a cycle. Therefore, for a unit step, the NCO creates a frequency shift of $2R_c/L$. The NCO gain in the classical sense is therefore

$$K_o = \frac{4\pi R_c}{L} \quad (10)$$

The loop filter is configured as shown in Figure 7. This is a step-invariant implementation of the analog filter

$$F(s) = a + \frac{b}{s}$$

The impulse-invariant implementation is generally applicable only for filter functions having sharp cutoff characteristics, due to the frequency aliasing errors inherent in this approach [6]. The step-invariant method provides an additional 6-dB/octave high-frequency attenuation, so for a given filter function, the aliasing errors are usually smaller.

From Reference 6, the step-invariant form $H(z)$ is obtained as

$$H(z) = \frac{z-1}{z} Z \left\{ L^{-1} \left[\frac{F(s)}{s} \right] \right\} \quad (11)$$

Substituting for $F(s)$ and evaluating equation (11) yields

$$H(z) = a + \frac{bT_s z^{-1}}{(1-z^{-1})} \quad (12)$$

where the symbol time, T_s , is the rate at which the delay element in the loop filter is updated. In this case, $T_s = 1/R_s$.

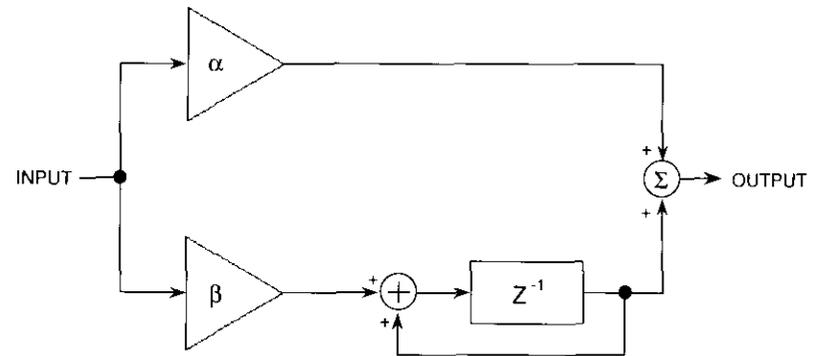


Figure 7. Realization of the Step-Invariant Loop Filter

Comparing equation (12) with the realization of Figure 7 gives

$$\alpha = a \quad (13a)$$

$$\beta = bT_s \quad (13b)$$

Following Gardner [7], and using equation (13), gives

$$\tau_1 = 1/\beta R_s$$

and

$$\tau_2 = \alpha\tau_1 = \alpha/\beta R_s \quad (14)$$

The natural frequency of the second-order loop is

$$\omega_n = \frac{(K_o K_d)^{1/2}}{\tau_1} = R_s \sqrt{\frac{4\pi\beta A_o A_i}{L}} \quad (15)$$

and the damping factor is

$$\xi = \frac{\tau_2 \omega_n}{2} = \frac{\alpha \omega_n}{2\beta R_s} \quad (16)$$

The loop bandwidth is

$$B_L = 0.5 \omega_n \left(\zeta + \frac{1}{4\zeta} \right) \quad (17)$$

For a small phase error and a fixed incoming signal amplitude, the loop bandwidth and damping factor can be adjusted by changing α and β . For a given signal energy, the damping factor is proportional to α and inversely proportional to $\sqrt{\beta}$.

Symbol timing loop

Symbol timing recovery is implemented with a transition-tracking decision feedback loop [8]. If a negative-to-positive change in the I channel between the assumed midsymbol samples of the previous symbol and the current symbol is detected, the I channel sample at the assumed transition period is added to an accumulator. Likewise, if a positive-to-negative change is detected, the sample at the transition point is subtracted from the accumulator. This can be expressed as

$$\tau_e = I_m \operatorname{sgn}(\Delta \hat{I}) + Q_m \operatorname{sgn}(\Delta \hat{Q}) \quad (18)$$

where τ_e is the timing error estimate; I_m , Q_m are the midsymbol (*i.e.*, transition point) demodulated values; and $\Delta \hat{I}$, $\Delta \hat{Q}$ are the differences between the current and previous \hat{I} and \hat{Q} values.

If the value of the accumulator exceeds a positive threshold, the accumulator is reset to zero, while an "advance" signal is sent to the external clock synchronizer hardware which causes the symbol clock to advance by a small step whose size is determined by the clock synchronizer dividers, as shown in Figure 8. Similarly, if the value of the accumulator becomes less than a negative threshold, a "retard" signal is sent to the clock synchronizer, causing the clock to retard by one step while the accumulator is reset to zero. By adjusting the thresholds up and down, the noise bandwidth of the loop can be changed. The behavior of this loop is equivalent to a first-order phase-locked loop (PLL), with the accumulator acting as the NCO.

The receive clock synchronizer is implemented digitally under control of the DSP software. The clock synchronizer includes a master oscillator at a frequency which is N times the symbol rate (typically $10 < N < 100$). The symbol-rate clock is divided down from the master clock frequency. Under DSP software control, the edges of these clocks can be advanced or retarded by one clock interval or $1/N$ of a symbol. This is accomplished by changing the divider ratio to $N - 1$ or $N + 1$, as illustrated in the figure.

Demodulator acquisition algorithms

The programmable modem can operate in either continuous mode or burst mode. The continuous mode is characterized by a continuous data stream without an acquisition preamble, thus necessitating the acquisition of carrier and clock from the modulated data stream itself. The burst mode is characterized by a relatively short data packet preceded by an acquisition preamble consisting of an unmodulated carrier segment for carrier frequency and phase acquisition, a modulated carrier segment for symbol clock acquisition, and finally, a predetermined unique-word segment to resolve the carrier phase ambiguity and mark the beginning of the data segment. Consequently, the two operation modes require radically different acquisition algorithms, major elements of which are briefly described below.

Continuous-mode acquisition

The acquisition procedures for the continuous mode are relatively straightforward. The demodulator depends on an external signal derived from a frame synchronizer or the BER detector in a Viterbi decoder to determine whether it is in synchronization. If frame synchronization is lost or if the BER is excessively high for a predetermined amount of time, say 5 seconds,

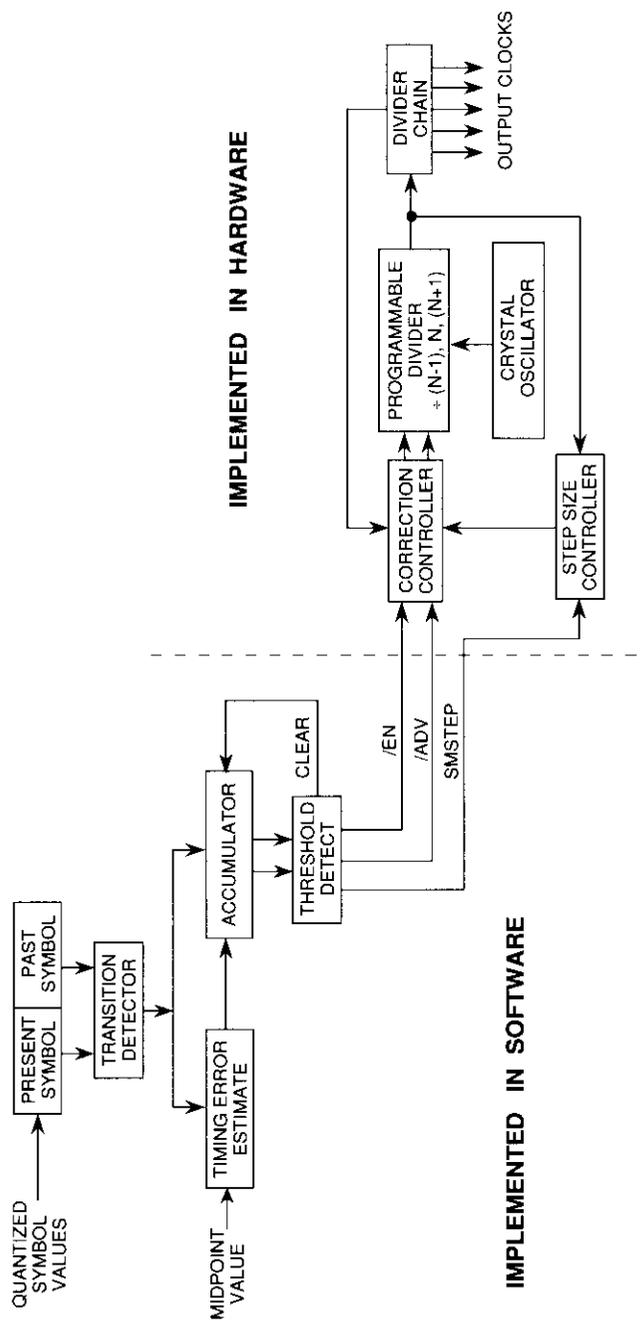


Figure 8. Block Diagram of the Symbol Timing Tracking Algorithm

the demodulator assumes that synchronization is lost and begins the acquisition sequence.

Acquisition involves frequency scanning and carrier phase and clock acquisition by the respective recovery loops. The frequency scanning procedure is needed because the loop bandwidths of the recovery loops, optimized for low energy-per-bit to noise-power density ratio (E_b/N_o) operation, are generally narrower than the frequency uncertainty range. Taking advantage of the fact that the frequency offset between the transmit and receive side is most likely small (typically less than $R_s/10$) and that the digital NCO can be set to any frequency within the uncertainty range instantly in the DSP software, a scanning pattern (Figure 9) is used. Frequency acquisition can therefore be achieved very quickly.

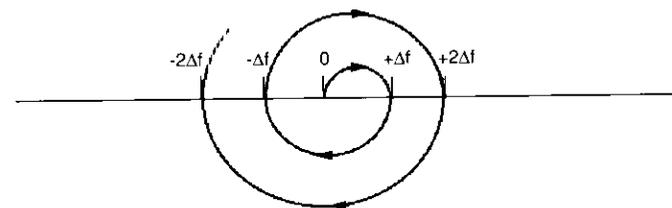


Figure 9. Pattern for Frequency Scanning

Burst-mode acquisition

The procedure for burst-mode acquisition involves the following four steps:

- Incoherent energy detection with hysteresis
- Frequency estimation
- Carrier phase acquisition
- Clock acquisition.

These are accomplished by a demodulator structure that uses multiple FIR filters for each preamble segment, as shown in Figure 10. A narrowband LPF is used during the unmodulated carrier portion of the preamble, followed by a narrowband BPF for the clock acquisition sequence. Finally, a matched square-root Nyquist filter is used for unique word and data demodulation. Since the multiple filters are used sequentially, each with the same number

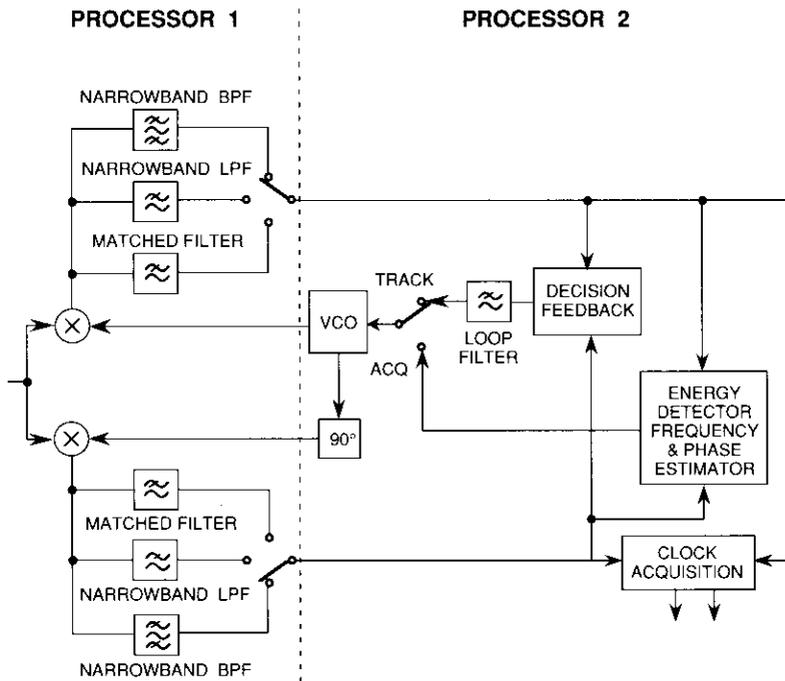


Figure 10. Block Diagram of the Burst Demodulator

of taps, and since the computational demands of the individual filters are the same, the additional computational complexity of this approach is minimal. The burst-mode acquisition steps are described below.

ENERGY DETECTION

Initially, both the local carrier reference frequency and the sampling clock are free-running. Assuming that the initial portion of the preamble contains an unmodulated carrier, the narrowband LPFs are switched in and their outputs are monitored by the energy detector. If the average filter output level over a sufficiently large number of symbols (typically >25) exceeds a preset threshold, the demodulator assumes that a burst is present. A second threshold, which is lower than the first, is then used. As long as the average filter output level is above the second threshold, the demodulator assumes that the burst continues. This threshold approach with hysteresis minimizes the probability of false energy detection and the premature termination of a burst.

FREQUENCY ESTIMATION

A receiver equivalent to the Costas loop for coherent demodulation of a phase-modulated digital signal (Figure 11) is assumed. The reference oscillator is implemented either numerically, as in the case of a software implementation, or in hardware by a voltage-controlled oscillator (VCO) whose controlling voltage is generated digitally via a D/A converter.

As in the case of energy detection, the preamble is assumed to contain an unmodulated carrier segment. Before acquisition, the reference oscillator in the receiver is left free-running, and the filter outputs are sampled once per symbol time. The frequency of sampling will be a function of the maximum frequency uncertainty to be resolved. Typically, if the worst-case frequency uncertainty is less than plus or minus one-half of the symbol rate (R_s), one sample per symbol is sufficient. The exact range depends on the bandwidth of the LPF used before the signal is sampled.

Both outputs from the I&Q phase-matched filters are sampled and quantized. At time t , these samples are quantized and represented by I_t and Q_t . The phase angle corresponding to the samples, θ_t , is obtained through a lookup table that implements a rectangular-to-polar conversion. The accuracy of the phase angle estimate is a function of the number of quantization levels used. High accuracy requires finer quantization, but also requires a larger table.

In a noiseless channel, the frequency error between the unmodulated carrier and the reference is obtained by taking the difference between the phase angle measurement between two adjacent symbol samples, that is,

$$f_i = (\theta_{t+1} - \theta_t) \text{ mod } 2\pi \cdot \frac{R_s}{2\pi} \tag{19}$$

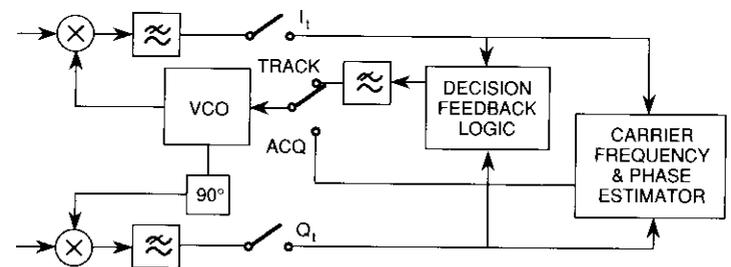


Figure 11. A Typical Costas Loop Demodulator

Here the operation mod 2π is used to ensure that the angle difference is in the range $(-\pi, \pi)$. For a noisy channel, the accuracy of the measurement is affected by the noise. To reduce the effect of channel noise, the measurement must be made over a number of symbols. If the measurement is taken over $(n + 1)$ symbols, there are n noisy estimates of the frequency error, as

$$f_i = (\theta_{i+1} - \theta_i) \bmod 2\pi \cdot \frac{R_s}{2\pi}, \quad i = 1, 2, \dots, n \quad (20)$$

Taking an average of these n estimates results in

$$\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i \quad (21)$$

A flowchart of the process is shown in Figure 12.

This basic frequency estimation technique can be made more robust by applying two enhancements. First, sample-by-sample scaling can be applied to all inputs to the frequency estimator. This scaling minimizes the quantization

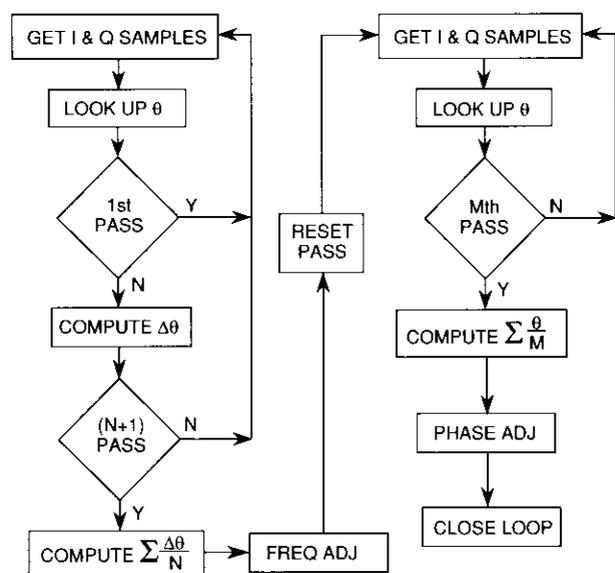


Figure 12. A New Frequency and Phase Acquisition Technique

error caused by the rectangular-to-polar coordinate conversion table used in the frequency estimation routine. As shown in Figure 13, the accuracy of the table is poor when both the I and Q samples are small. The scaler senses this condition and shifts the samples upward until at least one of them has an amplitude greater than 25 percent of the full range.

Second, a check for abnormal phase changes can be made for all angle estimates. Given the fact that the maximum frequency uncertainty present is known beforehand from the RFFIF system designs, a sample representing a large phase jump which corresponds to an instantaneous frequency change much greater than the maximum expected frequency certainty is unreasonable and is ignored. This large phase jump will be less than $\pm\pi$ because the frequency estimator operates at the symbol rate, and the maximum expected frequency uncertainty is usually on the order of $R_s/10$. Therefore, a modulo- 2π adjustment in the phase change will not occur.

CARRIER PHASE ACQUISITION

Two techniques can be used for phase acquisition. The first is simpler to implement, but requires the transmission of k additional preamble symbols of unmodulated carrier.

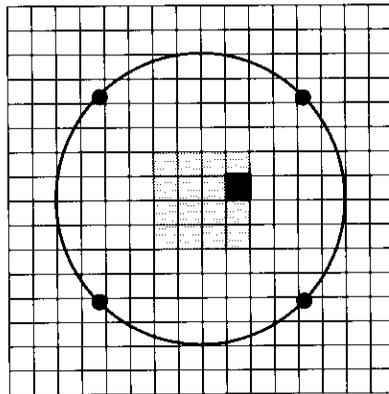
As soon as the frequency error is determined, proper correction is introduced into the VCO so that a correct reference frequency can be generated. The receiver continues to sample the filter outputs I_t and Q_t once per symbol, and determines the phase angles each time. Since the reference frequency is now very close to the transmitted frequency, the phase angle measurement should remain constant each time, except for the uncertainty caused by noise. The noise variance can be reduced by averaging over k symbols. The estimated phase

$$\bar{\theta} = \frac{1}{k} \sum_{t=1}^k \theta_t \quad (22)$$

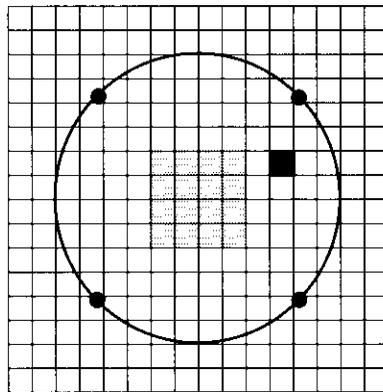
is obtained, and proper phase correction is introduced. After phase correction, the proper switches are closed to form a decision-feedback carrier recovery loop. Figure 14 is an example of this operation, showing phase error as a function of time.

An improved phase estimation technique exists which requires no additional transmission overhead, but more processing. After the frequency offset is estimated, the set of angle measurements $\theta_1, \theta_2, \dots, \theta_n$, can be adjusted by

$$\theta'_i = (\theta_i + \bar{\theta}_i) \bmod 2\pi \quad (23)$$



(a) Without Scaling



(b) With Scaling

 AREA OF POOR ACCURACY
 RECEIVED SIGNAL

Figure 13. Use of Scaling to Reduce Quantization Error in the Rectangular-to-Polar Conversion

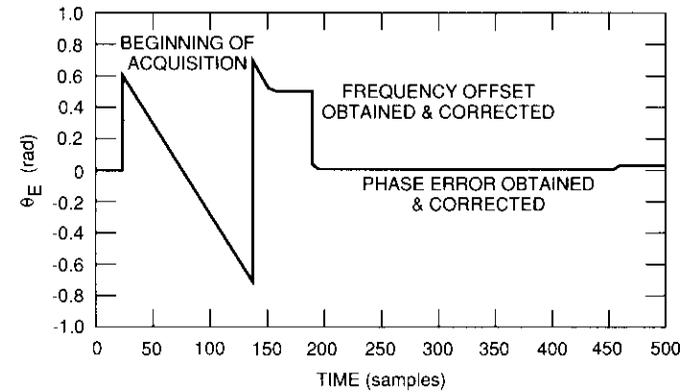


Figure 14. Example of the Carrier Frequency and Phase Acquisition Operation

where $\tilde{\theta}_i$ is obtained by a backwards accumulation of the phase angle attributed by the frequency offset, that is,

$$\tilde{\theta}_n = 0 \tag{24a}$$

$$\tilde{\theta}_{i-1} = \tilde{\theta}_i + 2\pi \cdot \bar{f}T_s \tag{24b}$$

The phase angle can then be estimated by averaging θ'_i over as many samples as desired, up to $(n + 1)$. Because the same set of samples is used for the frequency error estimate, any residual frequency error does not contribute to the error in the phase estimate.

A slight problem exists with both techniques described above because the phase angle assumes a finite value, either between $-\pi$ and π or between 0 and 2π . When the range $(-\pi, \pi)$ is assumed, there is a discontinuity at π . When the true phase angle is close to the discontinuity, an incorrect estimate may result.

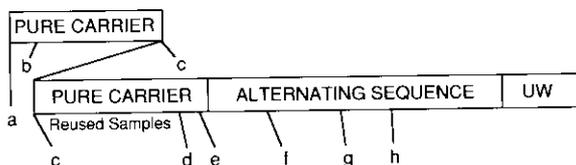
This problem is circumvented by checking the difference between the current angle and the initial angle. If the difference is within the interval $(-\pi, \pi)$, no corrective action is needed. If the difference is less than $-\pi$, the running total should be adjusted by adding 2π ; if the difference is greater than π , it should be adjusted by subtracting 2π . The adjusted sum is divided by k . If the result of the division is within the range desired, it is accepted as the average phase angle. If the result is out of the range desired, it should again be adjusted by adding or subtracting 2π .

This phase acquisition technique is slightly modified for the multiprocessor demodulator implementation. Due to the filtering delay and the time needed for interprocessor communications, some delay exists in the loop. This necessitates the use of a very narrowband loop filter (bandwidth $< R_s/100$) to maintain phase lock. However, use of such narrow loop bandwidth creates difficulties for initial acquisition. By experimentation, it was determined that a three-step filter switching procedure works well, as shown in Figure 15.

Additional delay is introduced by the global memory, which is used as a ring buffer to store the input data samples that have been used for frequency estimation, in order to reuse them for acquisition of the carrier phase. The ring buffer concept is illustrated in Figure 16. The first DSP takes the data from the FIFO and writes them into the global memory at an address specified



(a) Conventional Acquisition Sequence



(b) New Acquisition Sequence With Stored Samples

- | | |
|--|------------------|
| a-b. Energy detection | } Narrowband LPF |
| b-c. Open-loop frequency acquisition | |
| c-d. Wideband phase locking (200 Hz) | |
| d-e. Midband phase locking (150 Hz) | } BPF |
| e-f. Coarse clock acquisition
carrier loop uses BPSK phase detector | |
| f-g. Narrowband clock acquisition
Narrowband carrier tracking | |
| g-h. Narrowband clock acquisition
carrier coasting | } Root Nyquist |
| h. Narrowband carrier tracking
acquisition complete | |

Figure 15. Burst Acquisition Sequence With Stored Samples

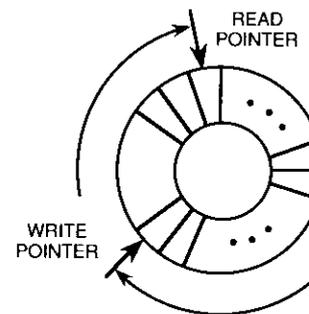


Figure 16. Concept of the Ring Buffer

by the write pointer, which is automatically incremented. Initially, this DSP reads back the data immediately after it finishes writing. After the frequency estimate is obtained, the read pointer is set back by $n \times 4$ samples, or n symbols. Thus, these samples are reused by both DSPs when the carrier recovery loop is closed. Samples prior to the incoherent energy detection are not reused because they may not be part of the burst. The net effect of this technique is to introduce an n -symbol delay into the clock tracking loop.

Figure 15 illustrates the entire burst acquisition sequence, which indicates that the net result of this technique is to increase the preamble length by n symbols. The ring buffer method for reusing samples is useful for cases where the unmodulated carrier segment of the preamble is insufficient for both the frequency and phase acquisitions.

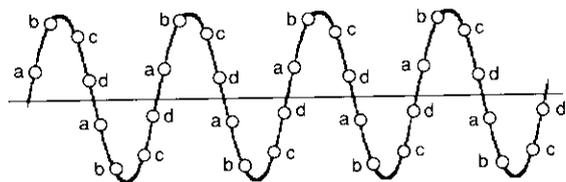
SYMBOL TIMING ACQUISITION

A problem with the burst symbol timing acquisition for O-QPSK modulation concerns the preamble design. For QPSK modulation, the most desirable preamble pattern for symbol timing acquisition is an alternating 0101 . . . sequence on both the I and the Q channels. However, this pattern corresponds to an unmodulated tone for O-QPSK, with a frequency offset of half the symbol rate. Thus, it creates a problem for conventional carrier recovery loops. Even though the decision-directed carrier recovery loop described previously can normally operate with such a signal, initial experiments indicate that a false-lock condition does exist [9]. A compromise design is therefore chosen for the symbol timing acquisition sequence. This sequence has a fixed 000 . . . pattern in the I channel, and an alternating 0101 . . . pattern in the Q channel. The sequence has only half the energy in the clock component than the previous sequence, but does not create false lock.

Symbol timing acquisition includes three phases. Acquisition is initiated by detection of a signal level drop at the output of the narrowband LPF in the Q channel that was used for the carrier acquisition. This is because the alternating 0101 . . . sequence corresponds to a frequency of half the symbol rate, which is outside the passband of the LPF. A narrowband BPF is then switched in which provides noise rejection but allows the half-symbol-rate component to pass through. In doing so, the signal in the 000 . . . modulated channel is totally rejected. A phase detector suitable for BPSK signaling is switched in to maintain carrier phase tracking. In the meantime, a coarse symbol timing acquisition algorithm is initiated while the sampling clock is left free-running. The coarse clock acquisition algorithm groups samples on the Q channel into four groups according to their relative positions within a symbol, as shown in Figure 17. After an accumulation over a number of symbols (e.g., 16 symbols), the group with the smallest magnitude is considered to contain the zero-crossing samples. The samples which are two samples away from the zero-crossings are then considered as data detection samples. This completes the coarse acquisition phase. This procedure reduces the symbol timing error to less than one-eighth of the symbol duration.

The second phase allows the symbol timing to settle after the coarse timing adjustment loop. The narrowband BPFs are still used. The symbol timing loop is the same as that used previously, except that only the minimum step size is used.

The third phase begins when the square-root Nyquist filter is switched in, in preparation for data demodulation. To avoid a sudden drop in the signal-to-noise ratio in the carrier recovery loop, the loop is made to coast for 16 symbols after the square-root Nyquist filter is switched in. The acquisition sequence is completed when the carrier recovery loop resumes normal operation.



Samples in four groups: a, b, c, and d

Figure 17. Grouping Samples for Coarse Clock Acquisition

Examples of DSP-based modems

Three DSP-based modems were implemented, each with some unique design features and functions tailored for specific applications. These modems are described briefly to illustrate the flexibility of the general architecture and implementation discussed above. The modems are as follows:

- 24- and 21-kbit/s BPSK/QPSK/O-QPSK burst/continuous modem for the INMARSAT Standard-B and aeronautical test bed.
- 144-kbit/s continuous QPSK modem.
- 1,200- and 600-bit/s INMARSAT Standard-C half-duplex, continuous-mode ship earth station (SES) BPSK demodulator.

The 24- and 21-kbit/s BPSK/QPSK/O-QPSK modem was designed for, and integrated into, the INMARSAT Standard-B and aeronautical test bed to help select the modulation and optimize the parameters of these systems. The modem is capable of operating at 24 or 21 kbit/s for both QPSK and O-QPSK operation, and at 12 or 10.5 kbit/s for BPSK operation. The demodulator is based on two TI TMS32020 DSPs in a dual-processor configuration, and is capable of both continuous- and burst-mode demodulation. All the functions described previously were implemented. The first processor performs all the mixing and filtering functions, while the second performs the synchronization and data detection functions. The matched filters are implemented at baseband. This versatile modem is capable of operating in six modes at four different speeds, and is ideally suited for the intended application. Figure 18 is a photograph of the modem drawer.

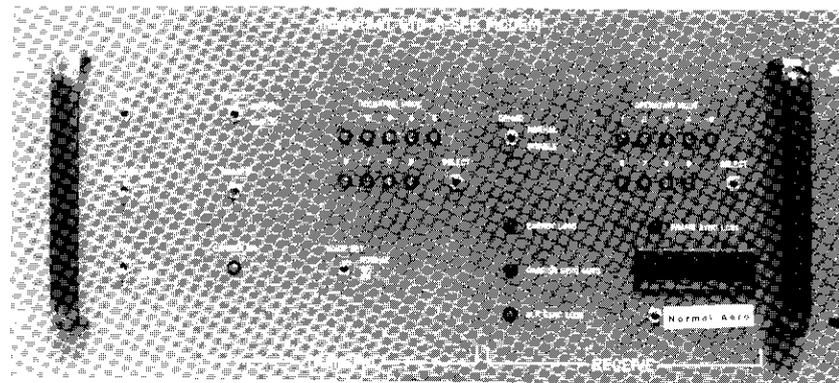


Figure 18. 21- and 24-kbit/s Multimode Standard-B/Aeronautical Modem Drawer

The 144-kbit QPSK continuous-mode demodulator was implemented to demonstrate the potential of multi-DSP-based modems with parallel processing for moderately high-speed operations. To minimize computations, the matched filters are implemented at low IF, since the frequency uncertainty is less than 10 percent of the symbol rate. The demodulator is based on four TI TMS32020 processors, each with a 200-ns instruction cycle.

The functional task partitioning of the processors is shown in Figure 19. The first processor reads the data from a FIFO. It performs the FIR filtering for the in-phase matched filter with the odd-numbered samples and forwards the even-numbered samples to the second processor via the global memory. The second processor uses the even-numbered samples to perform the FIR filtering for the quadrature-phase matched filter. The third processor performs the recovered carrier phase lookup, carrier mixing, and data demodulation. The fourth processor performs data detection, phase error computation, carrier loop filtering, symbol timing error calculations, and control of the external clock synchronizer. Figure 20 is a photograph of the four-processor demodulator board.

This modem is also capable of operating at 128, 64, and 32 kbit/s. Instead of choosing different IFs for different operating speeds, the sampling rate is increased from a nominal 4 samples per symbol to 8 and 16 samples per symbol for the lower data rates. The length of the FIR matched filters is also increased by the same factor. The computational requirements for the first and second processors remain about the same, since the amount of time available for lower-speed operation is also increased by the same factor.

The INMARSAT Standard-C SES demodulator is a BPSK demodulator operating at either 1,200 or 600 bit/s. The modulator is a simple balanced mixer which modulates the carrier without filtering, as specified by the Standard-C system definition. Because of the extremely low symbol rate, the IF for the demodulator input was chosen to be 9.6 kHz, which is 8 or 16 times the symbol rate. The matched filter is an integrate-and-dump section, which is a FIR filter with a duration of only one symbol. With this exception, the demodulator is implemented in a single TI TMS32010 processor, based on the same general architecture and algorithm described above.

The SES is operated in a half-duplex mode, which requires that the received signal be switched off when the SES is transmitting. To ensure rapid acquisition of the received signal after completion of its transmission, the SES demodulator stores the frequency last present in the NCO before the received signal was switched off, and then uses this information to reacquire the phase with the carrier recovery loop when the receiver is switched back on. Figure 21 is a photograph of the prototype SES unit.

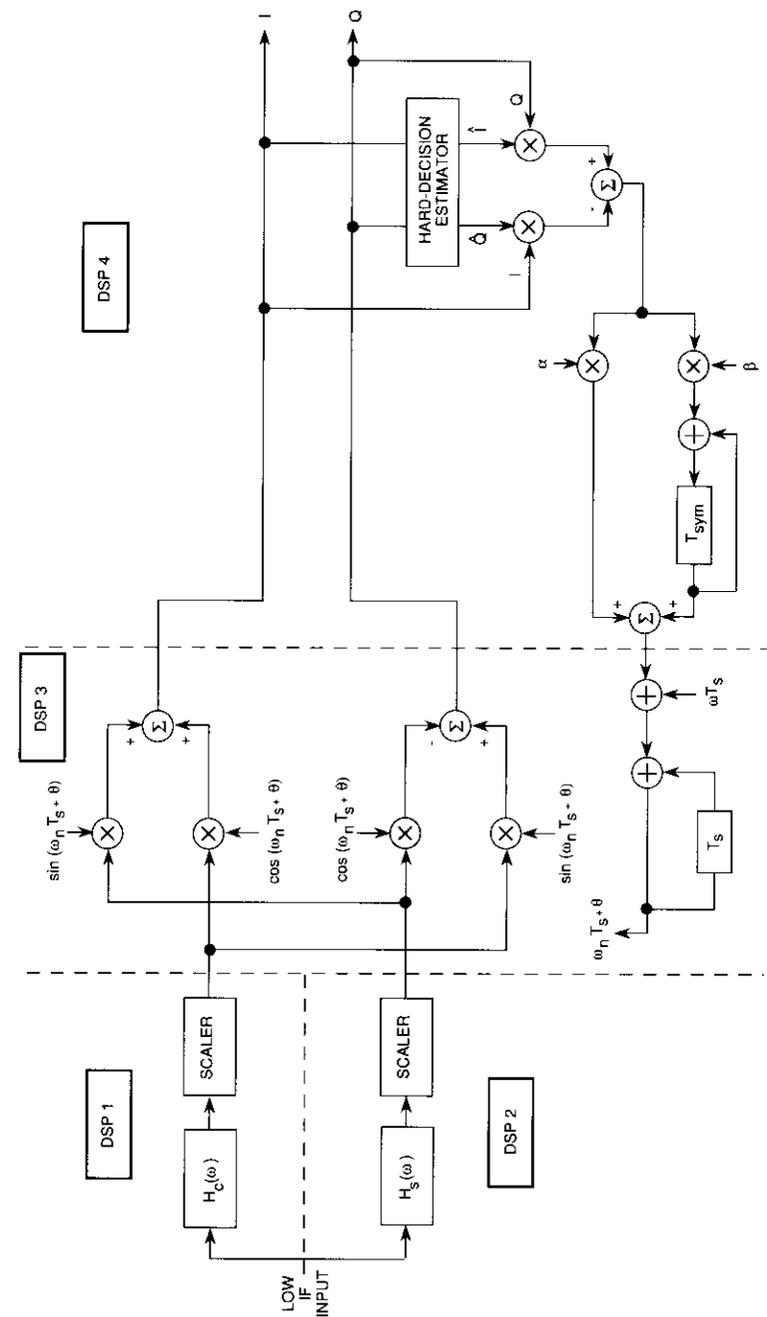


Figure 19. Functional Block Diagram of the Digitally Implemented Carrier Recovery Loop

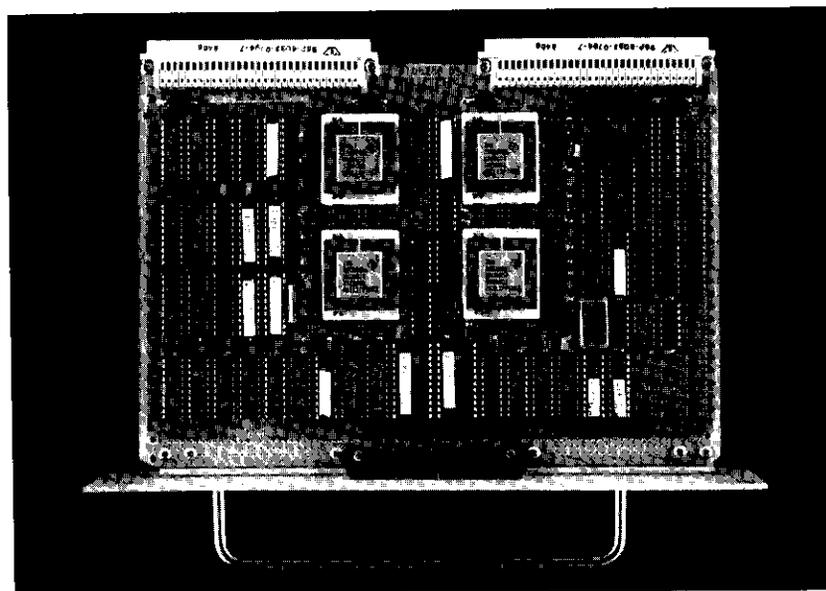


Figure 20. 16- to 144-kbit/s Four-Processor Continuous-Mode QPSK Demodulator Board

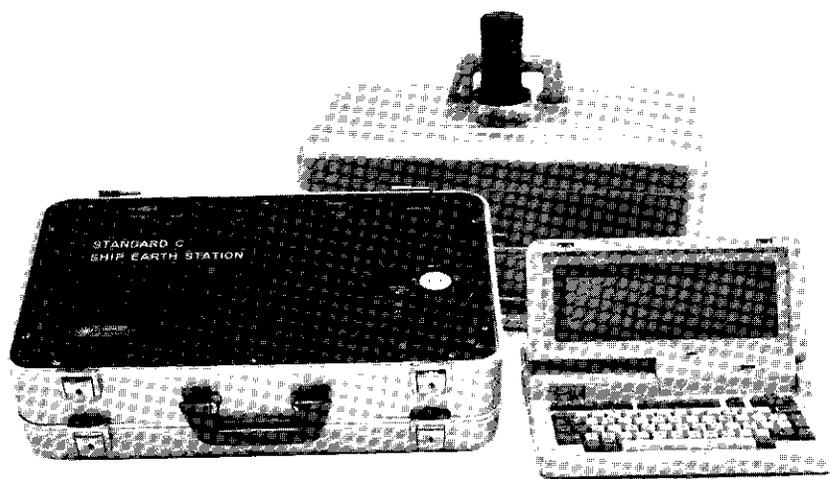


Figure 21. Standard-C SES Functional Prototype Unit

Representative test results

The BER performance of the Standard-B and aeronautical DSP-based demodulators operating in the continuous mode has been measured in an IF loopback configuration with AWGN introduced at 21.4 MHz. For both QPSK and O-QPSK, the performance is indistinguishable from theory. The BER performance for 24-kbit/s O-QPSK is shown in Figure 22 (where E_{mb} is the energy per modulation bit). When a maximum frequency offset of 1 kHz is introduced between the transmit and receive units, a degradation of no more than 0.2 dB is observed.

The probabilities of burst acquisition in the AWGN channel for frequency offsets of 0, 300, 600, and 1,000 Hz were also measured. A burst is considered acquired if the unique word is detected. The test results, summarized in Table 1, demonstrate that the acquisition probabilities are slightly higher if a longer preamble is used. With a frequency offset of 1,000 Hz, slightly greater degradation is observed due to the spectral truncation of the narrowband carrier recovery LPF at the input.

The measured BER performance for the four-processor, continuous-mode, 144-bit/s QPSK modem in 70-MHz IF loopback under AWGN conditions is shown in Figure 23. A representative plot of the frequency spectrum for $R_s = 72$ ksymbol/s is presented in Figure 24. From Figure 23, the four-processor QPSK demodulator exhibits an implementation loss of 0.4 to 0.8 dB at a BER

TABLE 1. NUMBER OF PACKETS RECEIVED AS A FUNCTION OF E_{mb}/N_o AND FREQUENCY OFFSET*

E_{mb}/N_o (dB)	0 (Hz)	300 (Hz)	600 (Hz)	1,000 (Hz)
QPSK Modulation (preamble = 508 bits)				
1	96	92	90	85
2	98	98	97	90
3	99	99	98	97
QPSK Modulation (preamble = 432 bits)				
1	88	85	85	80
2	96	96	94	90
3	98	97	97	95
O-QPSK Modulation (preamble = 432 bits)				
1	88	89	88	80
2	96	96	95	88
3	98	98	98	94

*Data based on 100 transmitted bursts.

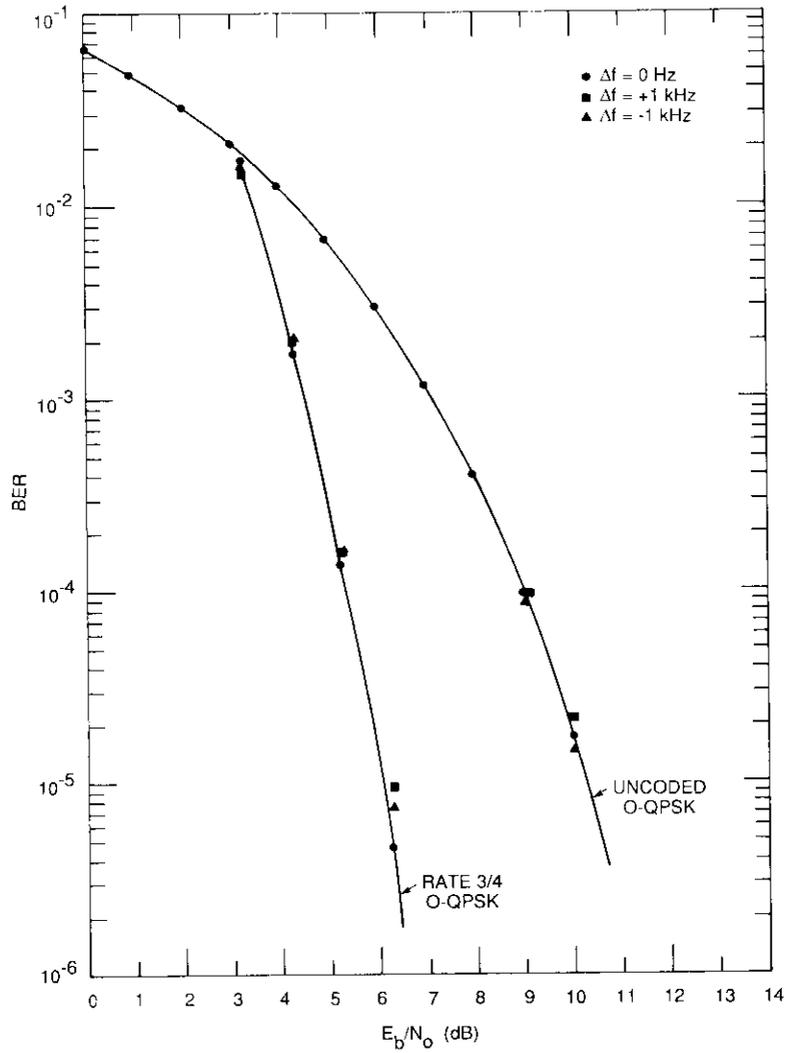


Figure 22. Measured BER Performance for the Dual-Processor 24-kbit/s Continuous-Mode O-QPSK Demodulator With 21.4-MHz IF Loopback Under AWGN

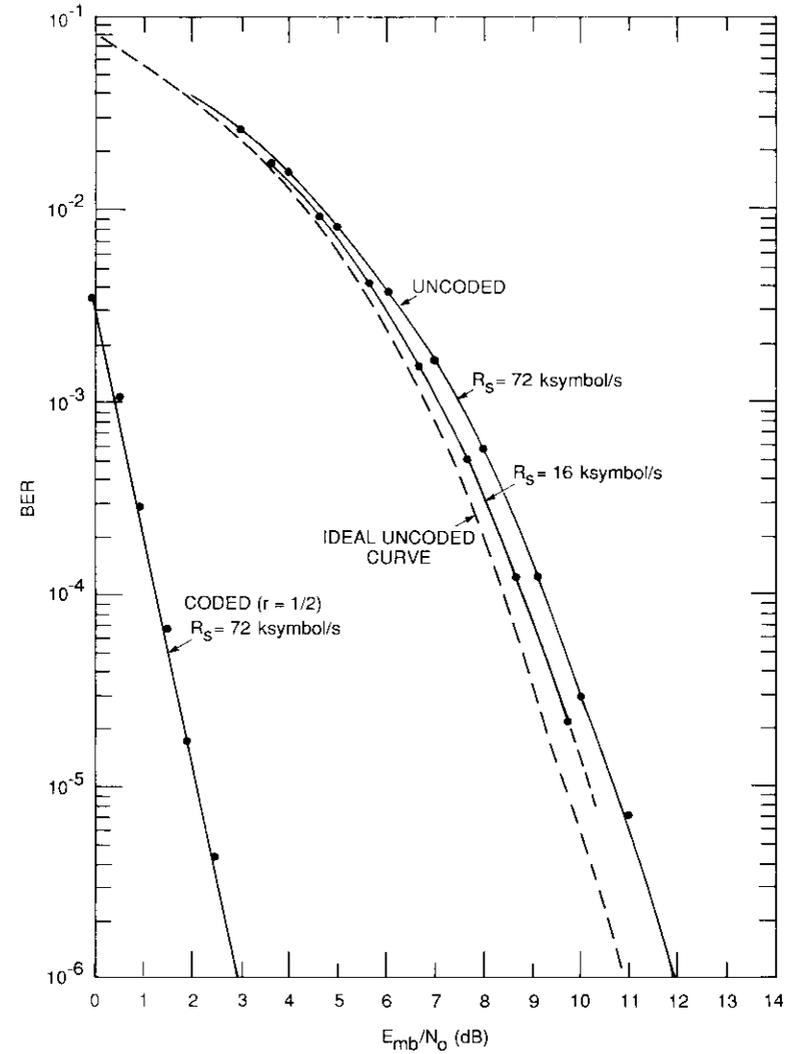


Figure 23. Measured Coded and Uncoded BER Performance for the Four-Processor 144-kbit/s Continuous-Mode QPSK Demodulator With 70-MHz IF Loopback Under AWGN

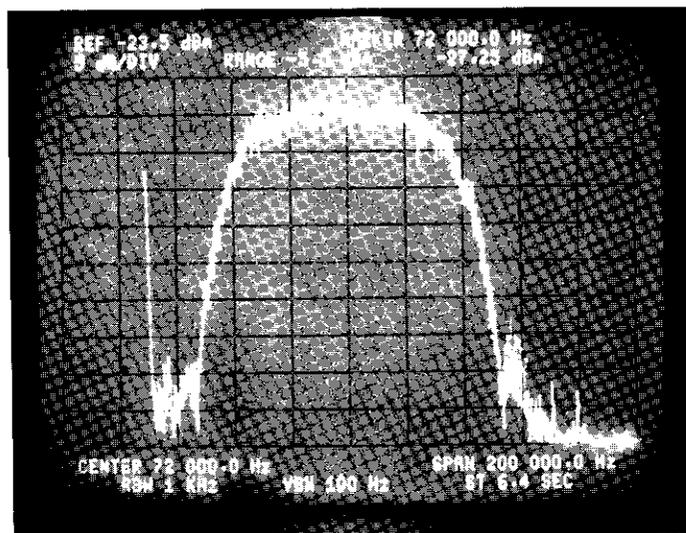


Figure 24. Received Signal Spectrum at the Down-Converter Output for 144-kbit/s Continuous QPSK Modulation With No Noise

of 10^{-4} , depending on the bit rate. The coding gain is 4.7 dB at $BER = 10^{-4}$, after allowing for a 3-dB difference due to the rate 1/2 code. A major component of the implementation loss results from the analog components and phase noise limitations of the IF converters. This can be deduced from Figure 25, which shows the BER performance of the four-processor QPSK modem back-to-back (no IF). The DSP modem has an implementation loss of 0.2 dB for all $BER < 10^{-3}$. This is indicative of the excellent back-to-back performance of the digitally implemented QPSK modem.

After a small conversion loss in the IF subsystem, these low implementation losses (0.2 to 0.6 dB, depending on the symbol rate) are a substantial improvement over the implementation losses previously reported (≈ 1 dB at BER of 10^{-3}) with earlier bit-slice microprocessor devices [10].

Conclusions

A programmable multiprocessor DSP-based modem architecture has been discussed, and time-domain algorithms for burst- and continuous-mode acquisition have been presented. Implementations of this architecture using single, dual-, and multiple-DSP devices have shown excellent BER correspondence with theory, with little performance degradation.

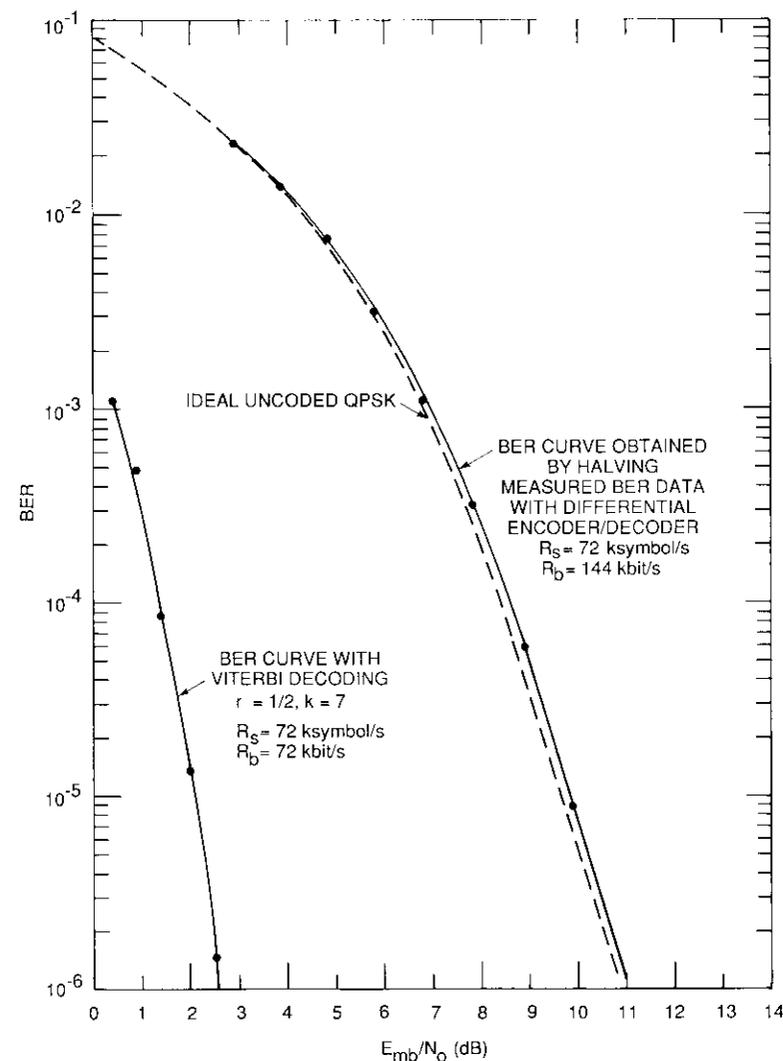


Figure 25. Measured Coded and Uncoded BER Performance for the Four-Processor 144-kbit/s Continuous-Mode QPSK Demodulator With Local Loopback Under AWGN

Since the input to these DSP-based modems is at baseband or low IF, and the output is soft-decision demodulated data that have been processed in software, an increase in the operating data rate will readily be achieved with a next-generation, higher-speed DSP. With the current generation of DSP devices, and particularly with the third generation of floating-point DSPs that are now available, the same architecture clearly can be used to obtain much higher operating data rates (*e.g.*, on the order of 500 kbit/s) than reported here. Higher operating rates also require changes to the DSP peripheral logic to support shorter access time constraints. This does not pose any significant problem, as the speeds of static RAM and programmable logic devices have increased substantially since these DSP-based modems were designed. Conversely, multiple- or dual-processor schemes utilizing the older-generation processors can be reduced to a more compact single-processor design by using current-generation devices, with no loss in operating speed.

The general multiprocessor-based modem architecture described in this paper continues to evolve. Prototype single- and dual-DSP-based modems with sophisticated frequency-domain-based burst- and continuous-mode acquisition algorithms are currently under development to combat severe multipath fading, Doppler shifts, and phase noise effects for low-bit-rate mobile satellite applications.

Acknowledgments

The authors would like to express their appreciation to N. Becker, whose initial design of a digitally implemented modem based on a bit-slice microprocessor architecture had significant influence on the work reported here, and to T. Lin, whose multiprocessor DSP hardware design provided the basic hardware platform for the generalized multiprocessor modem architecture.

References

- [1] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, New York: McGraw-Hill, 1979.
- [2] COMSAT Laboratories, "Digital Modulation and Coding Techniques Applicable to INMARSAT Standard-C Ship Earth Stations," Final Report (2 vols.) to INMARSAT under Contract INM/83-063, October 1984. COMSAT Data Catalog Nos. 84DC138 (Vol. I) and 84DC139 (Vol. II).
- [3] "INMARSAT Standard-B System Definition Manual," Issue No. 1.9, International Maritime Satellite Organization, London, England, July 1987.
- [4] N. Becker and V. Dhadesgoor, "A Microprocessor Based All-Digital Mobile Earth Terminal Simulation Facility," COMSAT Laboratories Technical Memorandum, October 1983. COMSAT Data Catalog No. 84DC016.

- [5] K. Shanmugan, *Analog and Digital Communication*, New York: John Wiley & Sons, 1979, p. 195.
- [6] W. D. Stanley, *Digital Signal Processing*, Reston Publishing Company (dist. by Prentice-Hall, Englewood Cliffs, New Jersey), 1975.
- [7] F. Gardner, *Phase Lock Techniques*, 2nd ed., New York: John Wiley & Sons, Ch. 2, 1979.
- [8] W. J. Hurd and T. O. Anderson, "Digital Transition Tracking Symbol Synchronizer for Low SNR Coded Systems," *IEEE Transactions on Communication Technology*, Vol. COM-18, No. 2, April 1970, pp. 141-147.
- [9] G. L. Hedin, J. K. Holmes, W. C. Lindsay, and K. T. Woo, "Theory of False Lock in Costas Loops," *IEEE Transactions on Communications*, Vol. COM-26, No. 1, January 1978, pp. 1-12.
- [10] C. Hegard, J. A. Heller, and A. J. Viterbi, "A Microprocessor-Based PSK Modem for Packet Transmission Over Satellite Channels," *IEEE Transactions on Communications*, Vol. COM-26, No. 5, May 1978, pp. 552-564.

Lin-Nan Lee received a B.S.E.E. from National Taiwan University in 1970, and an M.S. and Ph.D. in electrical engineering from the University of Notre Dame in 1973 and 1976, respectively. He is a Principal Scientist in the Communications Technology Division at COMSAT Laboratories. His research interests include digital signal processor (DSP)-based modems, forward error correction (FEC) coding techniques, cryptography, video transmission, and high-definition television (HDTV). Before joining COMSAT in 1977, he was a communications officer in the Chinese Air Force, a Research Assistant at the University of Notre Dame, and a Senior Scientist at Linkabit Corporation. He is a member of IEEE and Sigma Xi.





Ajit Shenoy received a B.Tech. in electrical engineering (electronics) in the First Class, with Distinction, from the Indian Institute of Technology, Madras, India, in 1979, and an M.Eng. (E.E.) from the Stevens Institute of Technology in 1981. He completed an additional year of graduate study in telecommunications and digital systems before joining COMSAT Laboratories in June 1982. He is currently Manager of the Digital Techniques Department, where his responsibilities include the design and implementation of advanced microprocessor-based communications hardware using digital signal processing techniques.

Prior to joining COMSAT, he worked on electronic motor speed control at Philips India, Ltd., and was a consultant for Synfax Manufacturing, Inc., Belleville, N.J., where he designed instrumentation systems for quality control. Mr. Shenoy is a recipient of several awards and scholarships from the Government of India. He has also coauthored a number of papers in the areas of digital modem design, computer simulation modeling, and digital communications systems design.

Michael K. Eng received a B.S.E.E. from the University of Maryland, College Park, in 1976, and an M.S.E.E. from George Washington University in 1981. He is a Senior Member of the Technical Staff in the Communications Technology Division at COMSAT Laboratories. His interests include the analysis, design, and implementation of communications and digital signal processing systems. Prior to joining COMSAT in 1985, he was employed by M/A-COM Linkabit Corporation as a Senior Engineer.



Index: communication satellites, networks, packet switching, performance tests

Quality-of-service measurement of satellite-based international packet-switched networks

D. S. GORHALE, R. KOHLI, AND M. KACKER

(Manuscript received December 15, 1988)

Abstract

Satellites are increasingly being used to provide international packet-switching services. CCITT Recommendation X.135 specifies the minimum quality requirements that must be met when providing such services. To verify the correctness and feasibility of these requirements, quality-of-service measurements were performed for two experimental configurations. The first involved one satellite link, one transatlantic cable link, and one transit network. The second used two satellite links and one transit network. The parameters measured included throughput capacity, connection establishment delay, connection release delay, and data packet transfer delay. The results demonstrate that the performance of packet-switched services over international satellite links is well within the specified requirements. It was also shown that components other than satellite propagation delay (such as nodal processing delay and transmission delay) contribute significantly to the overall delay.

Introduction

Performance issues are critical in the design of national and international communications networks. These networks use internationally standardized protocols to enable the successful interoperation of heterogeneous equipment from multiple vendors. Because the equipment is connected using different types of media, the protocols must operate in various environments. Therefore

the performance criteria specified for protocol operation must consider the diversity of environments in different networks. Among other factors, these criteria are influenced by the characteristics of the transmission media. This paper focuses on measuring of the performance of international data communications networks that include satellite links.

The measurements described here are part of a continuing program at COMSAT Laboratories which entails conducting experiments and submitting the results to the appropriate data communications standards organizations. These organizations include the International Telegraph and Telephone Consultative Committee (CCITT) and the International Standards Organization (ISO) internationally, and various American National Standards Institute (ANSI) committees in the U.S.A. The primary role of COMSAT in these organizations has been to ensure that satellite circuit characteristics are taken into consideration when data communications standards are established. Experiments conducted by COMSAT include joint experiments with the National Institute of Standards and Technology (NIST) [formerly the National Bureau of Standards (NBS)] to measure the performance of different ISO protocols over satellite links, and a joint experiment with TYMNET and the National Telecommunications and Information Administration (NTIA) to measure the quality of service of national packet networks. Related work in this area is described in References 1, 2, and 3.

Data communications networks typically employ the internationally standardized protocols specified in CCITT Recommendations X.25 and X.75. These protocols are used to provide packet-switched data services and are introduced briefly in the next section. CCITT Rec. X.135 specifies the minimum quality-of-service requirements that must be met when providing such packet-switched services. An experiment was conducted by COMSAT in cooperation with American Telephone and Telegraph (AT&T), Deutsche Bundespost, and INTELSAT to verify the correctness and feasibility of these requirements. The quality-of-service parameters measured were throughput capacity, connection establishment delay, data transfer delay, and connection release delay. A secondary objective was to experimentally verify protocol parameter values appropriate for use over satellite circuits.

Background

The communications environment has changed dramatically over the last few years. Whereas voice has hitherto been the dominant traffic carried by networks, the increasing requirements of data generation and data processing applications are now causing a shift in emphasis from voice to data. These data applications include electronic mail, file transfer, remote database access, and remote computing. In addition, the proliferation of personal computers,

workstations, and local area networks is increasing the demand for data communications networks.

Packet switching is one of the most commonly used techniques for data communications. This technique groups data in blocks, commonly referred to as packets. Each packet contains addressing information that allows intermediate switches to route the packet to its final destination. A typical packet-switched connection consists of three phases: a connection is established (call setup phase), data are transferred (data transfer phase), and the connection is released (call clearing phase). In the context of a packet-switched call, call setup involves making logical associations and allocating resources at nodes in the connection path. This is in contrast to a circuit-switched connection where a physical circuit is dedicated to the call. The call clearing phase allows the nodes to release allocated resources such as buffers and routing table entries. Since no physical circuit is exclusively allocated to the packet-switched call, other packet-switched calls can be statistically multiplexed over the same circuit. Due to the bursty and variable traffic characteristics of typical data applications, packet switching results in a significant increase in capacity utilization.

Most packet-switched networks use a seven-layer model (standardized by the ISO) for network architecture. The Appendix provides a brief outline of this model, as well as a concise description of the various layers and their relationship to Rec. X.25.

To access packet-switching services, data terminal equipment (DTE) in the customer's end system communicates with data circuit-terminating equipment (DCE) in the network. Rec. X.25 [4] specifies the physical, link, and network layer protocols for this type of communication. For packet-switched connections that traverse more than one network, Rec. X.75 [5] defines the protocol for communications between any two networks. The node used for communication with a different network is referred to as a half-gateway. The combination of Rec. X.25 and X.75 allows DTE on different networks to communicate. Rec. X.25 allows a packet-switched service user to set up multiple calls to the same (or different) locations over the same physical circuit simultaneously. These connections are referred to as virtual circuits or logical channels.

The parameters associated with the packet-switching protocols specified in Rec. X.25 and X.75 include packet size, window size, and retransmission timer value. These parameters (defined in the Appendix) are primarily dependent on the physical characteristics of the link, which include transmission rate, propagation delay, and bit error rate. The selection of appropriate values for the protocol parameters is extremely important in obtaining high performance.

Satellites are increasingly being used for packet-switched network links. The significant features offered by satellite circuits include distance insensitivity, high-capacity links, and broadcast capability. The ability to provide high-capacity links that are location-insensitive makes satellite circuits well suited for wide-area (long-haul) data networks. Satellite-based networks are even better suited for environments where configuration and communications needs change dynamically. Due to its inherent broadcast capability, a satellite-based network can be reconfigured almost instantaneously, without the complex routing algorithms required by terrestrial networks.

Any two end systems can be connected directly via a satellite link; therefore, the number of switching nodes in such a connection may be lower than in a terrestrial connection. The increased nodal processing and queueing delays associated with the increased number of switches in an all-terrestrial connection could lead to delays comparable to the additional satellite propagation delay. Moreover, selection of appropriate protocol and operational parameter values results in extremely high throughputs over satellite links. One such critical parameter is the window size. It specifies the maximum number of packets that a transmitting entity is allowed to send before obtaining an acknowledgment that the packets have been received at the destination. For full channel utilization on a 9,600-bit/s satellite circuit with a propagation delay of 270 ms, nodal processing delay of 30 ms, and packet size of 64 bytes, the minimum window size is the smallest integer exceeding $9,600 \times 2 \times (0.270 + 0.030) / (64 \times 8)$, that is, 12 packets. The round-trip propagation and processing delay is $2 \times (0.270 + 0.030)$ seconds, and the transmission time for a 64-byte packet is approximately $64 \times 8 / 9,600 = 0.053$ s. Thus, if a window size of less than 12 packets is chosen, the channel will be underutilized even when there are no errors, because the transmit side will have to wait for acknowledgments from the receive side. Further, link layer protocols that provide selective retransmission capability offer much better performance than do protocols that use "go back N " [6] techniques. Other key protocol parameters that affect performance over the satellite link include maximum packet size and retransmit timer value.

Quality-of-service standards

Some of the key parameters that specify the quality of service for packet switching are call setup delay, data transfer delay, call clearing delay, and throughput capacity. These are collectively called speed-of-service parameters. Other parameters that define the quality of service are accuracy, blocking, and availability. Only speed-of-service parameters were measured during the experiments reported here.

A single X.25 connection may traverse multiple national and/or international networks. The communications path is logically partitioned into several sections for specifying performance measurement procedures and objectives. Packet-level reference events are defined in order to identify the events during which time measurements are to be taken. An example of such an event is the receipt of an X.25 call request packet at a given node. The boundary separating two adjacent sections (referred to as the apportionment boundary) serves as the reference location at which the circuit is monitored and time measurements are taken. CCITT Rec. X.134 [7] specifies the apportionment boundaries and the X.25 and X.75 packet-level reference events to be used in performance measurement. Figure 1 shows the apportionment boundaries.

Rec. X.135 [8] specifies the requirements for speed-of-service parameters when international packet-switched services are provided. It also specifies the performance of the national and international portions of such connections. These portions are further classified as Type A (must be all-terrestrial circuits) and Type B (may involve satellite circuits). Table 1 gives detailed definitions of Type A and Type B. International links that include satellite hops are classified as International B portions. For the packet-switching experiment, the performance of the International B portion of the connection (between boundaries B_5 and B_{n-2} in Figure 1) was measured. The speed-of-service requirements specified in Rec. X.135 are shown in the last two columns of Tables 2 and 3, which appear in the Data Analysis and Results section of this paper. These target specifications are independent of the data rate or bit error rate on the circuits used.

TABLE 1. VIRTUAL CONNECTION PORTION TYPES FOR WHICH PERFORMANCE OBJECTIVES ARE SPECIFIED (ADAPTED FROM CCITT REC. X.134)

PORTION TYPE	TYPICAL CHARACTERISTICS
National A	Terrestrial connection via an access network section.
National B	Connection via an access network section with one satellite circuit; or via an access network section and one or more transit network sections.
International A	Connection via a direct terrestrial internetwork circuit section.
International B	Connection via two satellite circuits and one transit network section; or via one satellite circuit and two or more transit network sections.

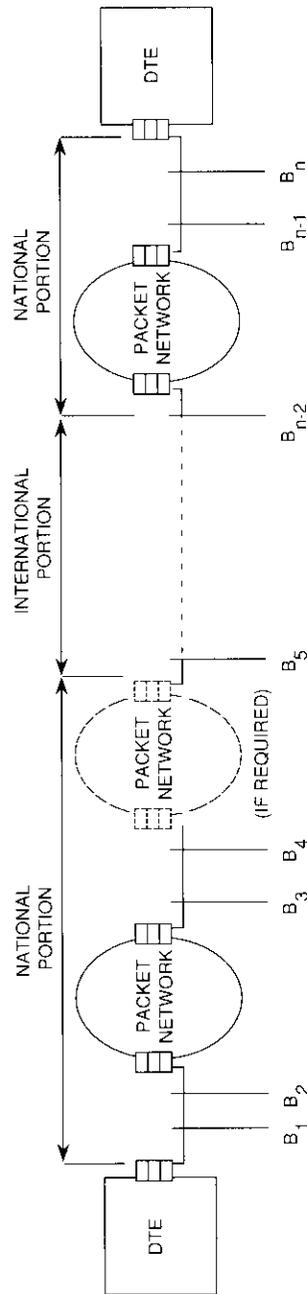


Figure 1. Apportionment Boundaries for Evaluating the Performance of International Packet-Switched Services (Adapted from CCITT Rec. X.134)

Experimental setup

This section describes the configurations, participants, equipment, and circuits used in this study.

Configurations

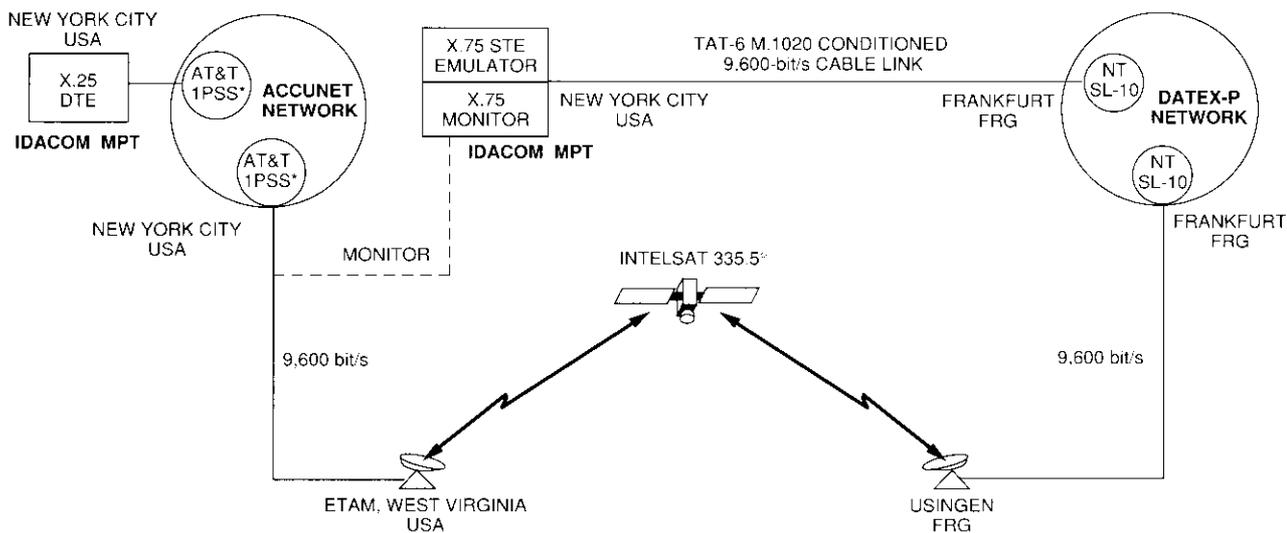
For these experiments, two configurations representative of the International B portion were selected. Two public data networks were used: AT&T's ACCUNET in the U.S.A. and Deutsche Bundespost's DATEX-P in the Federal Republic of Germany (F.R.G.).

Configuration 1 consisted of one satellite hop and one transit network (Figure 2), which typifies the most common environment. ACCUNET and DATEX-P were used as the source and transit networks, respectively, connected by an INTELSAT satellite link. The destination network was an X.75 protocol emulator located in New York City. Therefore the International B portion in this configuration consisted of a satellite circuit from the U.S.A. to the F.R.G., the DATEX-P network, and a transatlantic cable link from the F.R.G. back to the U.S.A.

Configuration 2 consisted of two satellite hops and a single transit network (Figure 3), which represents a less frequently used case. The source and destination networks were X.75 protocol emulators located at Etam, West Virginia, U.S.A., while DATEX-P was the transit network. The networks were connected by satellite links. Thus the International B portion in this configuration consisted of a satellite link from the U.S.A. to the F.R.G., the DATEX-P network, and a second satellite link from the F.R.G. back to the U.S.A.

Participants

The experiment was conducted jointly by AT&T, Deutsche Bundespost, INTELSAT, and COMSAT. AT&T provided the services of the ACCUNET network, the North American termination of the international transatlantic cable circuit to the F.R.G., and a Standard 2C earth station at Etam. Deutsche Bundespost provided the services of the DATEX-P network, the West German termination of the transatlantic cable circuit, and a Standard 1C earth station at Usingen. INTELSAT provided the space segment, two 9,600-bit/s voice-grade analog circuits in a Ku-band transponder (with up/down-link frequencies of 14/11 GHz) on the INTELSAT V satellite at 335.5°E. COMSAT designed, coordinated, and conducted the experiment; provided the protocol emulators; and developed the software for conducting the experiment and processing the data.



*IN THE EXPERIMENT, THE SAME 1PSS SWITCH WAS USED AT THE TWO LOCATIONS SHOWN.

Figure 2. Configuration 1 Equipment and Circuits

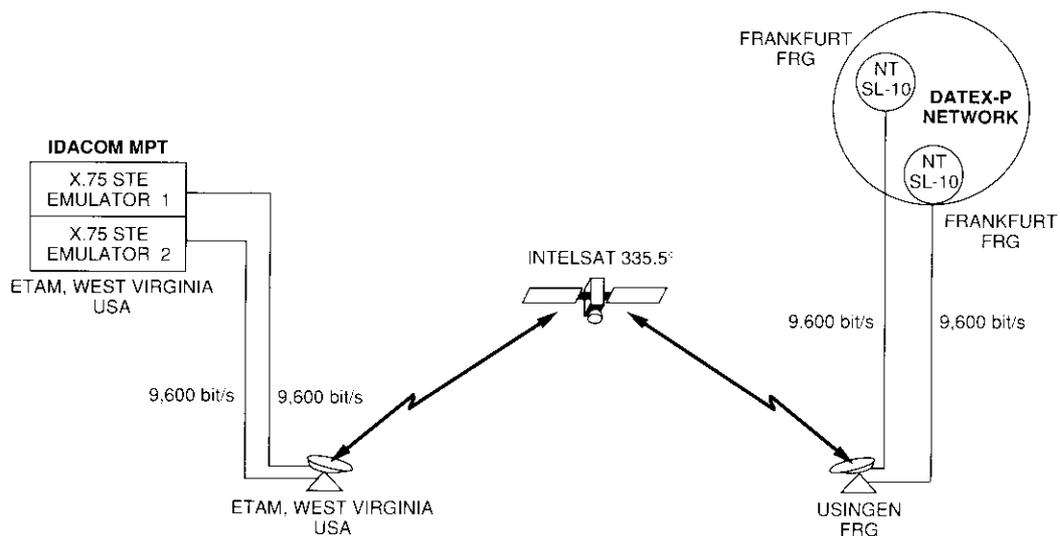


Figure 3. Configuration 2 Equipment and Circuits

Measurement equipment and circuits

PACKET-SWITCHING EQUIPMENT

The ACCUNET X.25 and X.75 ports were on an AT&T IPSS switch. The X.75 ports of the DATEX-P network were on two Northern Telecom SL-10 switches. All of these were high-performance switches typical of those used at the node sites of commercial public data networks.

PROTOCOL ANALYZERS

Protocol analyzers are useful developmental and diagnostic tools that generally serve two functions: protocol monitoring and protocol emulation. In the monitoring mode, these analyzers can identify and record predetermined events on a physical circuit (typically, the arrival of a type of packet that conforms to the encoding rules of a protocol on the circuit). In the emulator mode, the equipment can generate and respond to packets as required by the protocol.

In this experiment, two IDACOM Multi-Port Tester (MPT) Model 368.2 protocol analyzers were used. Each MPT has two ports that can be used independently. In configuration 1, one port of an MPT was used as X.25 DTE. The two ports of the second MPT were used as an X.75 monitor at B_5 and as an X.75 protocol emulator at B_{n-2} . In configuration 2, only one MPT was required, with both ports configured as X.75 protocol emulators (at B_5 and B_{n-2}).

The calculation of data transfer and clear indication delays requires that the real-time clocks at the B_5 and B_{n-2} boundaries be synchronized. Errors in delay measurements can arise for two reasons: inconsistent initialization of the real-time clocks, and differing clock drifts at the two boundaries. Since a single protocol analyzer unit was used for both X.75 ports in the experiment, the clocks at B_5 and B_{n-2} were generated by the same unit and so remained synchronized.

SATELLITE CIRCUITS

Analog voice-grade satellite circuits were used to provide two 9,600-bit/s links between Etam and Usingen. Only one circuit was used in configuration 1, while both were used in configuration 2. These circuits were terminated by V.29 modems* at New York City and Frankfurt in configuration 1, and at Etam and Frankfurt in configuration 2.

* 9,600-bit/s modems standardized for use on point-to-point, four-wire leased telephone-type circuits.

TRANSATLANTIC CABLE CIRCUIT

One leased 9,600-bit/s transatlantic cable circuit was used between the ACCUNET X.75 gateway in New York City and the DATEX-P X.75 gateway in Frankfurt. This circuit was over the TAT-6 cable and was conditioned according to the requirements of CCITT Rec. M.1020. The circuit was terminated by V.29 modems on either end.

OTHER CIRCUITS

A leased terrestrial C-2 conditioned analog 9,600-bit/s circuit was used between the earth station at Etam and the ACCUNET gateway at New York City. Similar circuits were used between the earth station at Usingen and the DATEX-P gateway at Frankfurt.

Experimental procedure

The experiment was conducted by running specially developed test software on the protocol analyzers. The tests were divided into two categories: delay measurement and throughput capacity measurement.

Delay measurement

Delay measurement experiments were performed by repeatedly setting up a call, sending three data packets in sequence, and then clearing the connection. Packets were time-stamped and stored on a protocol analyzer as they passed the B_5 and B_{n-2} boundaries. The time measurements at these boundaries were synchronous because both ports of the protocol analyzers used the same clock. The time stamps made it possible to determine the call setup, data transfer, and clear indication delays.

Data transfer delay was obtained for packets of three different lengths: 32, 64, and 128 octets (a CCITT term for a byte, *i.e.*, 8 bits). A large number of samples were obtained by repeating the three-phase procedure outlined below:

Step 1: Call Setup Phase. Compose and send a call request packet.

Step 2: Data Transfer Phase.

- (a) Compose and send a data packet with 32 octets of user data.
- (b) Compose and send a data packet with 64 octets of user data.
- (c) Compose and send a data packet with 128 octets of user data.

Step 3: Call Clearing Phase. Compose and send a clear indication packet.

In configuration 1, the packets were generated by a protocol analyzer emulating X.25 DTE (*i.e.*, the emulator functioned as a terminal connected to the ACCUNET network). In configuration 2, the packets were generated by a protocol analyzer emulating an X.75 gateway. In this case, the analyzer emulated an X.25 network in the U.S.A. generating traffic over an X.75 link.

Throughput capacity measurement

Throughput capacity measurements were obtained for the cases of four and eight logical channels. The experiments were conducted by setting up connections for either four or eight logical channels, transmitting 2,000 data packets, and releasing all connections. Each packet contained 128 octets of user data. Data packets were transmitted until the packet-level window closed. As receive-ready (acknowledgment) packets were received, the packet-level window opened and more data packets were transmitted. The data packets were stored as they passed the B_{n-2} boundary. A large number of samples were obtained by repeating the three-phase procedure outlined below:

Step 1: Call Setup Phase. Compose and send a call request packet on four or eight logical channels.

Step 2: Data Transfer Phase. Send 2,000 data packets.

Step 3: Call Clearing Phase. Compose and send clear indication packets on all logical channels.

Data analysis and results

The experimental data were transferred from the protocol analyzer to a Digital Equipment Corporation (DEC) VAX 11/750 minicomputer. Specially designed software was written to analyze the data and calculate delay and throughput. For delay, the cumulative distribution function is used to represent the percentage of packets that fall below a certain value. Throughput capacity is represented by using the complementary cumulative distribution function, because the number of samples that exceed a certain value is of interest.

Table 2 summarizes the measured values for call setup delay, data packet transfer delay, and clear indication delay, while throughput capacity is summarized in Table 3. The mean and 95th percentile requirements specified in Rec. X.135 are shown in the last two columns of the tables for comparison with the measured results.

In Table 3, the throughput has been evaluated with and without overhead. Throughput values are a composite of all the logical channels and represent

TABLE 2. MEASURED CALL SETUP, DATA TRANSFER, AND CALL CLEARING DELAYS (ms)

DELAY	CONFIGURATION 1				CONFIGURATION 2				X.135 REQUIREMENTS	
	MIN	MEAN	95%ILE	MAX	MIN	MEAN	95%ILE	MAX	MEAN	95%ILE
Call Setup	1,111	1,168	1,235	1,450	1,558	1,593	1,620	1,702	2,000	3,000
Data Transfer										
32 octets	408	447	460	580	654	661	675	697	—	—
64 octets	467	508	530	684	714	721	730	753	—	—
128 octets	586	624	640	746	832	839	845	917	900	1,200
Call Clearing	400	421	440	568	622	630	635	668	900	1,200

a user throughput based on a 128-octet packet size, excluding network and data link layer overheads. If overhead is included, the throughput is higher. Overhead can be broken down into the following components:

- Network layer header (3 octets)
- Cyclic redundancy check (2 octets)
- Data link layer header (3 octets)
- Flag (1 octet)

Call setup delay

Call setup delay is the time that elapses between the transmission of a call request packet and the receipt of the corresponding call connected packet. If t_1 is the time at which a call request packet is sent from the source X.75 gateway, and t_2 is the time at which a call connected packet is received at the source X.75 gateway, then the International B portion call setup delay at B_5 is $t_2 - t_1$. Measurements t_1 and t_2 were obtained by reading time stamps associated with the call request and call connected packets.

Figure 4 shows the cumulative distribution functions for the call setup delay in both configurations. In configuration 1, 90 percent of the packets had a call setup delay of less than 1,200 ms, while in configuration 2, 90 percent of the packets had a delay of less than 1,600 ms.

Delay can be analyzed by examining three components: transmission delay, propagation delay, and network delay. Transmission delay is dependent on packet size and link rate. For example, a 12-octet (96-bit) packet on a 9,600-bit/s link would have a transmission delay of 96/9,600 seconds, or 10 ms. Propagation delay is a function of the distance between nodes and the medium used to interconnect them. For example, the propagation delay

TABLE 3. MEASURED THROUGHPUT (bit/s)

NUMBER OF LOGICAL CHANNELS	CONFIGURATION 1				CONFIGURATION 2				X.135 REQUIREMENTS		
	MIN	MEAN	95%ILE	MAX	MIN	MEAN	95%ILE	MAX	MEAN	95%ILE	
	User Data Transfer Throughput										
4	7.163	7.678	7.225	8.429	8.082	8.244	8.225	8.368	1.500	1.200	
8	8.504	8.788	8.540	8.849	8.208	8.297	8.214	8.504	1.500	1.200	
	Data Transfer Throughput, Including Overhead										
4	7.667	8.218	7.733	9.022	8.650	8.824	8.803	8.956	—	—	
8	9.102	9.406	9.140	9.471	8.785	8.880	8.792	9.102	—	—	

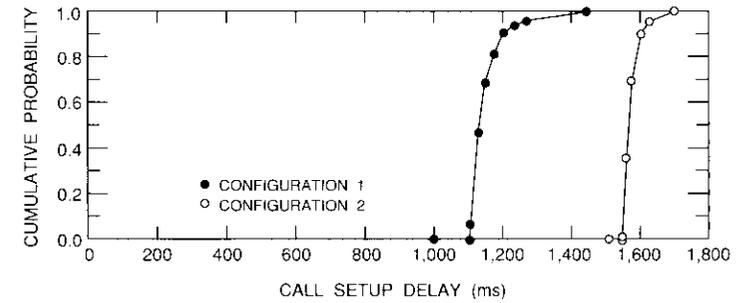


Figure 4. Call Setup Delay Cumulative Distribution Function

is approximately 270 ms from one earth station to another via satellite. Network delay is the composite of various switching, transmission, and propagation delays within a network.

From loopback measurements, the propagation delay from New York to Frankfurt via the satellite link was determined to be 280 ms. This included the terrestrial propagation delays from New York to the Etam earth station, and from the Usingen earth station to Frankfurt. The propagation delay from New York to Frankfurt via the transatlantic cable circuit was 60 ms.

The transmission delay for a call request packet (29 octets) is $29 \times 8/9,600 = 24$ ms. If the DATEX-P network delay is excluded, in configuration 1 the time interval between the transmission of the call request packet at the B_5 boundary and its receipt at the B_{n-2} boundary is $24 + 280 + 24 + 60 = 388$ ms.

The call connect packet is 15 octets long; therefore, its transmission delay over a 9,600-bit/s link is 13 ms. Using reasoning similar to that used in the call request packet case, the time interval between the transmission of the call connected packet at the B_{n-2} boundary and its receipt at the B_5 boundary is $13 + 60 + 13 + 280 = 366$ ms. Hence, excluding the DATEX-P network delay, the call setup delay is $388 + 366 = 754$ ms.

From Table 2, the mean call setup delay in configuration 1 is 1,168 ms. Therefore, the mean delay introduced by the DATEX-P network is $1,168 - 754 = 414$ ms. The round-trip satellite propagation delay of 560 ms for the call setup sequence contributes less than half ($560/1,168$) the overall delay. Thus, it is evident that delay components other than satellite propagation delay contribute considerably to the overall call setup delay. In a practical situation, for calls originating at a considerable distance from the cable head, the satellite propagation delay will contribute an even smaller percentage of the overall delay. A comparison of measured values and Rec. X.135

requirements in Table 2 reveals that the mean and 95th percentile delay values for call setup are well within the Rec. X.135 specifications for both configurations.

Data packet transfer delay

Data packet transfer delay is the time elapsed between the transmission of a data packet at the B_5 boundary and its receipt at the B_{n-2} boundary. If t_3 is the time at which a data packet is sent from the source X.75 gateway, and t_4 is the time at which a data packet is received at the destination X.75 gateway, then the International B portion data packet transfer delay at B_5 is $t_4 - t_3$. Measurements t_3 and t_4 were obtained by reading time stamps associated with a data packet.

Figure 5 shows the cumulative distribution functions for the data packet transfer delay in both configurations for the 128-octet data packets. In contrast to the call setup delay, the data transfer delay is a one-way delay. The total number of octets in a packet with 128 octets of user data is 136 or 137, depending on whether normal or extended numbering is used. This packet size results in a transmission delay of 114 ms over a 9,600-bit/s link. If the DATEX-P network delay is excluded, in configuration 1 the time interval between the transmission of the data packet at the B_5 boundary and its receipt at the B_{n-2} boundary is $114 + 280 + 114 + 60 = 568$ ms. From Table 2, the measured mean data transfer delay for a 128-octet user data packet is 624 ms. Therefore, the mean delay introduced by the DATEX-P network is $624 - 568 = 56$ ms.

If a transatlantic cable circuit were used instead of a satellite circuit, the total delay would be reduced by only one third ($220/624$) because of the

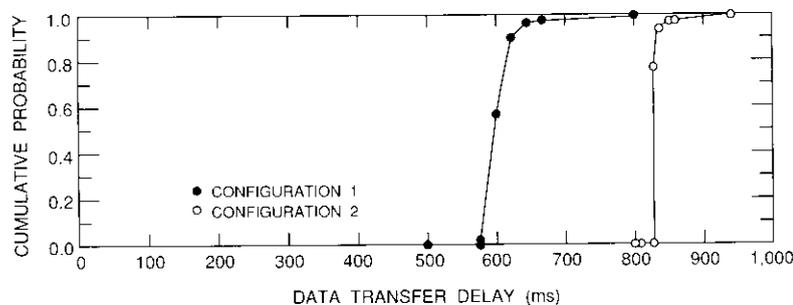


Figure 5. Data Transfer Delay Cumulative Distribution Function for Packets With 128 octets of User Data

additional delay introduced by terrestrial circuits and intermediate nodes from the source and/or destination to the cable head. A comparison of the measured values and Rec. X.135 requirements in Table 2 reveals that the mean and 95th percentile delay values for data packets are well within the specifications of Rec. X.135 for both configurations.

Clear indication delay

The clear indication delay is the time elapsed between the transmission of a clear request packet at the B_5 boundary and its receipt at the B_{n-2} boundary. If t_5 is the time at which a clear request packet is sent from the source X.75 gateway, and t_6 is the time at which a clear request packet is received at the destination X.75 gateway, then the International B portion of clear indication delay is $t_6 - t_5$. Measurements t_5 and t_6 were obtained by reading time stamps associated with a clear request packet. Figure 6 plots the cumulative distribution functions for the clear indication delay for both configurations.

As with the data packet transfer delay, the clear indication delay is one-way. The transmission delay for a 15-octet clear request packet is $15 \times 8/9,600 = 12.5$ ms. From Table 2, the mean clear indication delay for configuration 1 is 421 ms. Given the measured satellite and cable propagation delays, the mean clear indication delay introduced by the DATEX-P network is $421 - 12.5 - 280 - 12.5 - 60 = 56$ ms. Thus it is evident that delay components other than satellite propagation delay contribute significantly to the overall clear indication delay. By comparing the measured values and Rec. X.135 requirements in Table 2, it can be seen that the mean and 95th percentile values for clear indication delay are well within the specifications in Rec. X.135 for both configurations.

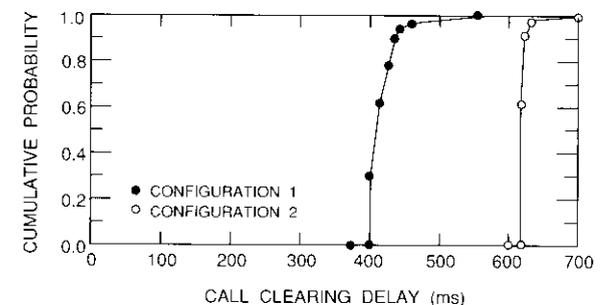


Figure 6. Call Clearing Delay Cumulative Distribution Function

Throughput capacity

Let T_{n-2} denote the time required to successfully transmit a group of 200 packets of 128 user octets each across the B_{n-2} boundary. Then, the throughput capacity, η , for the International B portion is given by

$$\eta = \frac{200 \times 128 \times 8}{T_{n-2}} \text{ bit/s}$$

Measurements were obtained by transmitting 2,000 packets in each run of the experiment. The first and last 100 packets were discarded for steady-state measurements. The remaining 1,800 packets were then divided into nine groups of 200 packets. Throughput capacity was obtained for both the four and eight logical channel cases.

The complementary cumulative distribution function for the throughput of the eight-logical-channel case is plotted in Figure 7. By comparing the measured values and Rec. X.135 specifications given in Table 3, it can be seen that the throughput was 400- to 500-percent higher than that required by Rec. X.135 for both the mean and 95th percentile values for both configurations.

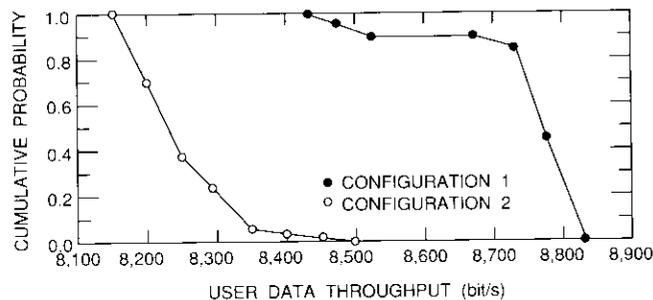


Figure 7. User Data Throughput Complementary Cumulative Distribution Function for Packets With 128 octets of User Data (8 logical channels)

Conclusions

Experimental results have been presented which demonstrate that the mean and 95th percentile values of call setup delay, data packet transfer delay, call clearing delay, and throughput capacity for configurations using both

single- and double-hop transmission are well within the requirements specified in CCITT Rec. X.135. On the 9,600-bit/s satellite link, a peak throughput as high as 9,471 bit/s was obtained, and it was found that satellite propagation delay accounted for less than half the total delay. Other delay components such as transmission and nodal processing delays contribute significantly to the overall delay. Furthermore, since satellite circuits may bypass some terrestrial circuits and switching nodes, the end-to-end delay in satellite-based configurations is often comparable to that in terrestrial/transoceanic cable configurations.

The difference between the measured mean data transfer delays in the two configurations examined was approximately 200 ms, thus demonstrating that an additional satellite hop can easily be accommodated in data circuits, with only marginal impact on delay. It should be noted that the choice of appropriate operational parameters is critical for obtaining high performance across satellite circuits.

Acknowledgments

The authors wish to thank all those involved with the development, implementation, and execution of these experiments. Special thanks are due to D. Chitre and I. Knight at COMSAT, E. Wetmore at AT&T, K. Peter-Steinruck at Deutsche Bundespost, and T. Oishi at INTELSAT.

References

- [1] J. S. McCoskey, W. Redman, W. Morgart, B. Hung, and A. Agarwal, "An International High-Speed Packet-Switching Experiment," *COMSAT Technical Review*, Vol. 14, No. 1, Spring 1984, pp. 1-24.
- [2] L. Palmer, J. Kaiser, S. Rothschild, and D. Mills, "SATNET Packet Data Transmission," *COMSAT Technical Review*, Vol. 12, No. 1, Spring 1982, pp. 181-212.
- [3] S. H. Richman, "A Comparison of Field Performance Measurements of Multi-National Packet Switching Networks," IEEE Global Telecommunications Conference, Tokyo, 1987, *Proc.*, Vol 3, pp. 1627-1634.
- [4] CCITT, "Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Terminals Operating in the Packet Mode and Connected to Public Data Networks by Dedicated Circuit," Recommendation X.25, VIIIth Plenary Assembly, Malaga-Torremolinos, October 1984, *Red Book*, Vol. VIII, Pt. 3, pp. 108-242.
- [5] CCITT, "Terminal and Transit Call Control Procedures and Data Transfer System on International Circuits Between Packet-Switched Data Services," Recommendation X.75, VIIIth Plenary Assembly, Malaga-Torremolinos, October 1984, *Red Book*, Vol. VIII, Pt. 4, pp. 152-235.

- [6] A. Tannenbaum, *Computer Networks*, Englewood Cliffs, New Jersey: Prentice-Hall, 1981.
- [7] CCITT, "Apportionment Boundaries and Packet Level Reference Events for Defining Packet-Switched Performance Parameters," Recommendation X.134, IXth Plenary Assembly, Melbourne, November 1988, *Blue Book*, Vol. VIII, Pt. 3, pp. 358-368.
- [8] CCITT, "Speed of Service (Delay and Throughput) Objectives for Data Networks When Providing International Packet-Switched Service," Recommendation X.135, IXth Plenary Assembly, Melbourne, November 1988, *Blue Book*, Vol. VIII, Pt. 3, pp. 368-391.
- [9] CCITT, "Reference Model of Open Systems Interconnection for CCITT Applications," Recommendation X.200, IXth Plenary Assembly, Melbourne, November 1988, *Blue Book*, Vol. VIII, Pt. 4, pp. 3-56.

Appendix. Review of the Open Systems Interconnection model and CCITT Recommendation X.25

The ISO and CCITT have defined a layered architecture to facilitate communications between data equipment made by different manufacturers. This architecture is referred to as the Open Systems Interconnection (OSI) model [9]. The term "open" signifies that, by conforming to these standards, a system will be capable of interacting with systems conforming to the same standards worldwide. This architecture enables each entity in a network of open systems to be viewed as a stack of seven layers, each layer providing a different function (see Figure A-1). Each layer adds to the services provided by the lower layers, so that the full seven-layer stack provides the set of services needed for distributed applications. A brief description of the functions performed by each layer of the OSI architecture is given below.

1. **Physical Layer:** Transmits bits over the physical link between two adjacent nodes (nodes may be terminals, computers, switching equipment, etc.). It provides mechanical, electrical, functional, and procedural characteristics to establish, maintain, and release physical connections between link layer entities. The architecture of this layer permits the use of a variety of media for interconnection (e.g., cables, satellites, fiber optics). It is the lowest layer of the architecture.
2. **Link Layer:** Corrects for possible transmission errors in the physical layer and transforms the link so that it appears to be free of transmission errors. This is accomplished by first segmenting data into units called

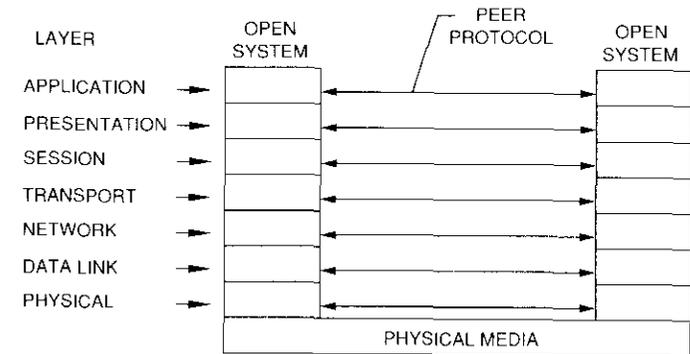


Figure A-1. Seven-Layer OSI Architecture

frames and then retransmitting any frames that may have experienced errors in transmission. This layer makes the "wire" provided by the physical layer appear to be very reliable. It also provides mechanisms that control data flow to prevent a faster transmitter from overloading a slower receiver.

3. **Network Layer:** Establishes, maintains, and terminates connections between communicating end-system transport layer entities. This also involves the routing of data between intermediate entities. In addition, the network layer performs such functions as congestion control, flow control, addressing, and the segmentation and reassembly of long packets. If a connection traverses networks that employ dissimilar protocols, this layer may be responsible for the appropriate protocol conversion.
4. **Transport Layer:** Controls the transport of data from the source-end system to the destination-end system. This layer relieves higher layer entities of concern regarding the transport of data and makes the end-to-end transfer reliable. It also handles the problem of end-to-end flow control.
5. **Session Layer:** Provides enhanced services to end-to-end transport connections. One such service is *dialog control* for managing the full/half-duplex nature of the connection. Another session service is *synchronization*, which is used to mark reference points from which data transfer can be resumed (or synchronized) in a situation such as a system crash in the midst of a large file transfer.

- 6. Presentation Layer:** Concerned with the syntax and semantics of the information transmitted. The encoding of data is standardized to allow the exchange of information between systems that use different information codes (e.g., ASCII, EBCDIC). This layer can also provide data compression and cryptography.
- 7. Application Layer:** Defines applications for the end users. Some common examples are file transfer, electronic mail, virtual terminals, and transaction processing.

At the transport layer level, end-to-end transfer capability can be provided by reserving a full path and bandwidth from the source to the destination system for the duration of the connection. However, the intermittent nature of most end-user traffic provides opportunities to economize on line costs by sharing links between adjacent nodes for data traffic between multiple end-user sites. One technique for achieving this is *packet switching*. Whenever the source user has data to be transmitted to a remote destination, the data are inserted in a packet and delivered to the network. Each packet has addressing information that allows the network to route the packet to its final destination. Packet switching is provided by a number of commercial public data networks.

In Rec. X.25, CCITT standardized the physical, link, and network protocols for accessing this service. These protocols define the interface between the source equipment (DTE) and network equipment (DCE).

The physical layer in Rec. X.25 is specified by referencing CCITT Rec. X.21, which defines the use of a duplex point-to-point synchronous circuit. However, by bilateral agreement, other arrangements are allowed for the physical layer.

The link layer in Rec. X.25 provides for the reliable transmission of information between two adjacent nodes. Each unit of information at this layer is called a frame, and three types of frames are used. Information frames are used to send data; supervisory frames are used for acknowledgment (or negative acknowledgment of data) or to slow the sender (for flow control); and unnumbered frames are for control purposes (and to acknowledge other control frames). These control frames are used to announce such conditions as invalid frames (protocol violations), a node going down, a node coming up, and so forth.

The frames described above have fields called control (for sequence numbers, acknowledgments, etc.), data (for information carried), and checksum (for detecting transmission errors). Sequence numbers are used to order frames, to request retransmission if an out-of-sequence frame is received, or

to acknowledge receipt for in-sequence frames. This layer has parameters that are dependent on the characteristics and quality of the link. For efficient operation, appropriate parameters should be set by network managers. One such parameter is the time the sending side waits for an acknowledgment before considering a frame lost. This parameter is called retransmission time, since the sending side will retransmit a frame that is perceived as lost.

The network layer of Rec. X.25 manages connections between two pieces of DTE. Two forms of connection are provided: switched virtual circuits and permanent virtual circuits. Switched virtual circuits are similar to ordinary telephone calls. A connection is established, information is exchanged, and the connection is released. In contrast, a permanent virtual circuit is similar to a leased line. It is always present, and the DTE can exchange data at any time without a call setup sequence.

End-to-end DTE connections (virtual calls) are established as follows. The DTE builds a call request packet and sends it to the DCE. The network delivers the packet to the destination DCE, which then gives it to the destination DTE. At this point, the call request packet becomes the incoming call packet. If the destination DTE wishes to accept the call, it builds and sends back a call accepted packet. When the originating DTE receives the call accepted packet (now called the call connected packet), the virtual call is established.

During this procedure, some parameter values may be negotiated. One such parameter is the maximum number of bytes in a data packet (*packet size*). Another is the number of sequential packets that can be transmitted without waiting for an outstanding acknowledgment (*window size*). The values of these parameters are influenced by throughput and flow control requirements. Such negotiable parameters are referred to as *facilities* in CCITT nomenclature.

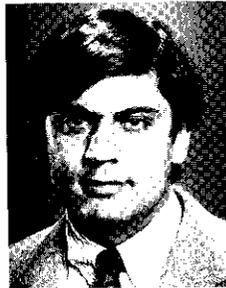
After the call setup phase, the DTE uses the full-duplex channel to exchange data packets. When either side wants to terminate the connection, it sends a clear request packet to the other side, which then sends a clear confirmation packet back as an acknowledgment to the originating DTE.

A virtual circuit can also be set up between DTE connected to different networks. A common approach to linking two networks is to have a relay node (called a gateway) which is linked to both networks. However, since the two networks may be in different countries and run by different organizations, the relay node is conceptually split into what can be called *half-gateways*, and the two halves are connected by a link. The protocols use these half-gateways to communicate with each other, as defined in Rec. X.75 (which is almost identical to Rec. X.25). In CCITT nomenclature, a half-gateway is called a *signaling terminal* (STE).



Dilip S. Gokhale received a B. Tech. degree in electronics and communications from the Regional Engineering College, Surathkal, India, in 1981, and an M.S. degree in electrical engineering from the New Jersey Institute of Technology, Newark, in 1984. He joined COMSAT in 1984 and is presently an Associate Department Manager in the Network Technology Division. He is actively involved in research and development in the areas of data communications, satellite communications, network management systems, and multiprocessor architectures.

Ranjit Kohli received his Bachelor of Technology degree in electrical engineering from the Indian Institute of Technology, Delhi, in 1984, and an M.S. from Rensselaer Polytechnic Institute, Troy, New York, in 1986. He joined COMSAT Laboratories in 1986 as a Member of the Technical Staff in the Network Technology Division. Mr. Kohli has been actively involved with ISDN and OSI standards and has represented COMSAT at meetings of T1S1, T1Q1, T1E1, X3S3.3, X3S3, and CCITT Study Group XVIII. In 1988, he was leader of the U.S. delegation to the meeting of ISO/IEC/JTC1/SC6/WG4 in Tokyo. In 1989, he served as Secretary to Technical Subcommittee T1S1.1, which deals with ISDN services and architecture. He was also actively involved in enhancing the OSI Transport Protocol Class 4 (ISO 8073) for obtaining better performance over long-haul connections such as satellites. Mr. Kohli is currently pursuing an MBA at the Wharton School of Management.



Mukesh Kacker received his Bachelor of Engineering (B.E.) degree in electronics and communications from the University of Roorkee, Roorkee, India, in 1982; a Master of Technology (M. Tech.) in computer technology from the Indian Institute of Technology, New Delhi, India, in 1984; and an M.S. in computer science from the University of Wisconsin, Madison, in 1986. He joined COMSAT Laboratories in 1987 as Member of the Technical Staff in the Network Technology Division. Mr. Kacker has been involved in performance measurements of X.25- and X.75-based packet-switched networks with satellite links, and in the implementation and performance evaluation of the upper layers of emerging OSI standards, especially the X.400 Message Handling System, OSI presentation layer, and Common Management Information Protocol (CMIP). He has represented COMSAT at ANSI X3S3.3 and X3T5.5 standards meetings.

Index: antennas, communication, satellites, software

Generation and display of satellite antenna patterns

W. L. COOK, D. D. OPIEKUN, AND L. R. KARL

(Manuscript received October 3, 1989)

Abstract

The problems involved in generating and displaying satellite antenna patterns on the surface of the earth are addressed, and a computer program developed for this purpose is described. The various coordinate systems used in formulating the problem are described, and the equations for transforming points between coordinate systems are derived. This is followed by a discussion of the techniques used to generate and display geographic features, antenna gain contours, and flux density contours for various types of beams. Related topics addressed include the calculation of great circle paths, earth station azimuth and elevation angles, and the effect of satellite pointing errors on gain patterns. Plots generated by the program are used to illustrate the types of analyses that can be performed.

Introduction

Communications satellites use directional antennas to focus radiation patterns on the surface of the earth. These patterns vary from a simple circular beam covering the entire earth to complex patterns conforming to continental boundaries. The effectiveness of a particular antenna pattern can be determined by calculating the gain or flux density patterns projected on the surface of the earth. It is often desirable to view the earth using one of a number of alternative projections, and to supplement the contour data with geographic boundaries, grids, and elevation angle contours.

Two computer programs have been developed to perform these functions. The Antenna Coverage Program (ACP) executes input commands stored in a text file. It runs on an IBM mainframe and on an IBM PC/AT. The Interactive Antenna Coverage Program (IACP) is a menu-driven program that also runs on an IBM mainframe and IBM PC/AT, as well as on a VAX workstation. Both programs are written in FORTRAN and use the Graphical Kernel System for generating graphical displays.

These programs allow a user to create satellite antenna patterns on a world map. The user may define a satellite location by specifying the satellite's longitude, latitude (for satellites in inclined orbit), and distance from the earth. Next, the user can display an orthographic, perspective, or equi-rectangular projection of the world which includes lines of latitude and longitude, country boundaries, and state boundaries. The user may also define the map's center, height, and width. After the map parameters are established, the user can then define the location and shape of an antenna pattern: circular, elliptical, shaped, user-specified contours, or measured gain values in antenna coordinates. Both ACP and IACP can generate and display several antenna patterns in the form of equal gain contours or equal flux density contours on a single map.

ACP and IACP can also generate contours of constant elevation angle for a set of user-specified elevation angles. Several other features are available in both programs which allow a user to add a title, add a footnote, and label points on the map. Along with the plots that ACP and IACP generate, both programs create a table of values for a set of user-defined points which includes the elevation angle, azimuth, gain, and flux density at each point.

Coordinate system definitions

Three principal coordinate systems are required when generating antenna contours on the earth's surface: an earth-centered system, a satellite-centered system (also the antenna coordinate system), and a viewing system. The locations and relative orientations of these systems are described in this section.

Earth-centered system

The fundamental coordinate system used in antenna coverage analysis is the earth-centered system. This system is a Cartesian coordinate system (x_e, y_e, z_e) whose origin is at the center of the earth, whose positive z_e -axis passes through the North Pole, and whose positive x_e -axis passes through the Greenwich meridian. Hence, the x_e - y_e plane lies in the plane of the Equator. The orientation of the earth-centered coordinate system is shown in Figure 1.

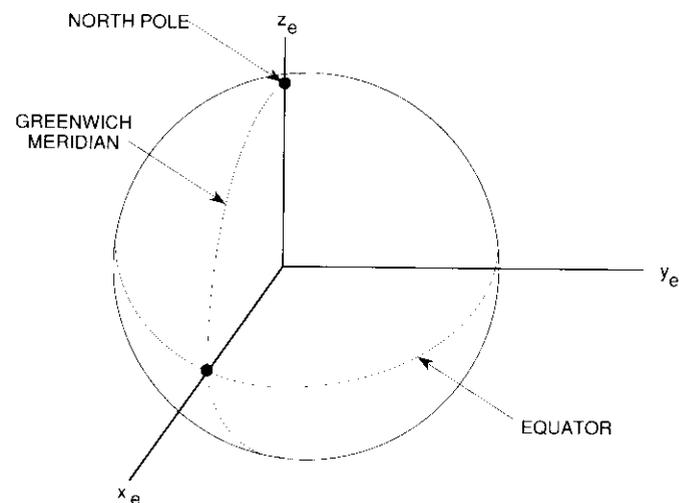


Figure 1. Earth-Centered Coordinate System

All other coordinate systems are defined in terms of the earth-centered system.

It is convenient to specify a point on the earth's surface in terms of its latitude and longitude (see Figure 2). Latitude is indicated by the symbol ρ_e and longitude by the symbol μ_e . If the radius of the earth is given by R , then the Cartesian coordinates are related to latitude and longitude by

$$\begin{aligned}x_e &= R \cos \rho_e \cos \mu_e \\y_e &= R \cos \rho_e \sin \mu_e \\z_e &= R \sin \rho_e\end{aligned}$$

The latitude and longitude are related to the Cartesian coordinates by

$$\begin{aligned}\rho_e &= \sin^{-1}(z_e/R) \\\mu_e &= \tan^{-1}(y_e/x_e)\end{aligned}$$

Satellite-centered system

It is useful to define a satellite-centered coordinate system (x_s, y_s, z_s) whose origin is at the satellite and whose positive z_s -axis passes through the subsatellite point and through the center of the earth. The positive x_s -axis of this system points North, parallel to the axis of the earth (*i.e.*, the z_e -axis). The y_s -axis lies in the plane of the Equator. Figure 3 shows the orientation of the satellite-centered system. As will be seen below, this system is useful for defining satellite antenna beam patterns and satellite pointing angles.

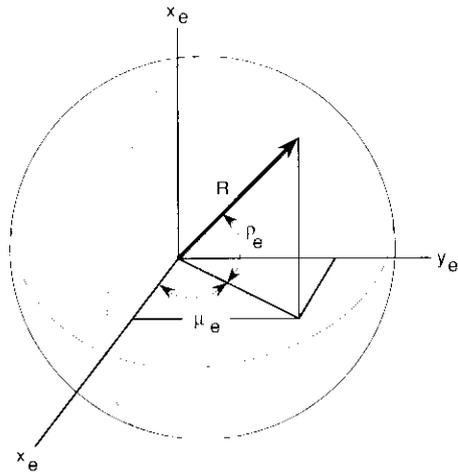


Figure 2. Latitude and Longitude Coordinates

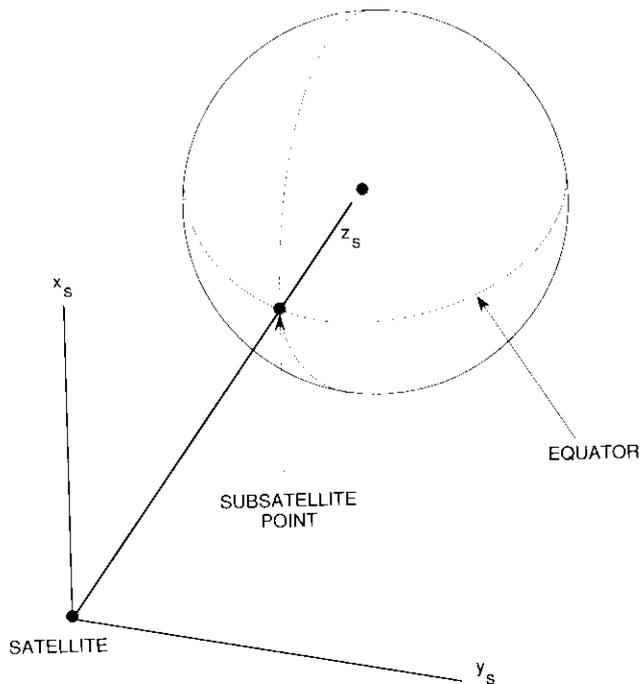


Figure 3. Satellite-Centered Coordinate System

Coincident with the satellite-centered Cartesian system is a spherical system having coordinates T_s , θ_s , and ϕ_s (see Figure 4). The angle θ_s is the angle T_s makes with the z_s -axis, while ϕ_s is the angle the projection of T_s on the x_s - y_s plane makes with the x_s -axis. The transformation between Cartesian and spherical coordinates is given by

$$\begin{aligned} x_s &= T_s \sin \theta_s \cos \phi_s \\ y_s &= T_s \sin \theta_s \sin \phi_s \\ z_s &= T_s \cos \theta_s \end{aligned}$$

Antenna patterns are typically described in a so-called "tilde" spherical system $(\tilde{\theta}, \tilde{\phi})$ in which $\tilde{\theta}$ is the angular distance between T_s and the x -axis, and $\tilde{\phi}$ is the angular distance between T_s and the y -axis. As will be shown below, the calculations are performed in angular coordinates α and β , which are the north-south and east-west pointing angles, respectively, from the satellite. Angles measured in the tilde system can be converted to pointing angles by using the following equations:

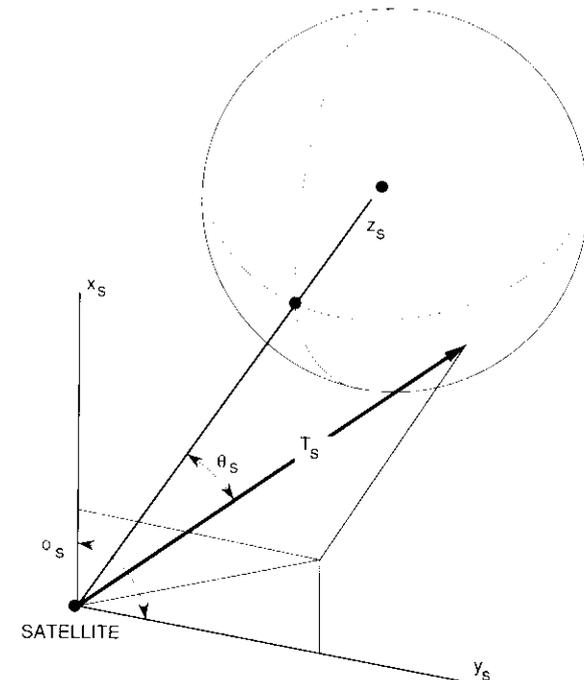


Figure 4. Satellite-Centered Spherical System

$$\alpha = \pi/2 - \bar{\theta}$$

$$\beta = \pi/2 - \bar{\phi}$$

Viewing coordinate system

The graphical representation is actually displayed in the viewing coordinate system. The viewing system is defined below for each type of projection considered: orthographic, perspective, and equirectangular [1].

ORTHOGRAPHIC PROJECTION

In an orthographic projection, the earth's surface is projected in a straight line onto a plane which is perpendicular to the defined line of sight. Figure 5 shows a typical orthographic view of the earth.



Figure 5. Orthographic View of the Earth

In the viewing coordinate system (x_v, y_v, z_v) for an orthographic projection shown in Figure 6, the z_v -axis defines the line of sight of the viewer and the x_v - y_v plane is the projection plane. The only restriction on the viewing system is that the positive z_v -axis must pass through the center of the earth, that is, the viewer must be looking directly (not obliquely) down on the earth. The point where the line of sight intersects the earth's surface is defined as the view center. The distance, D , between the viewer and the view center is arbitrary for this projection since it does not affect the displayed image. The x_v -axis lies in the plane defined by the meridian passing through the view center. Note that the x_v -axis is parallel to the axis of the earth (the z_e -axis) only in the special case where the view center lies on the Equator (known as an equatorial view).

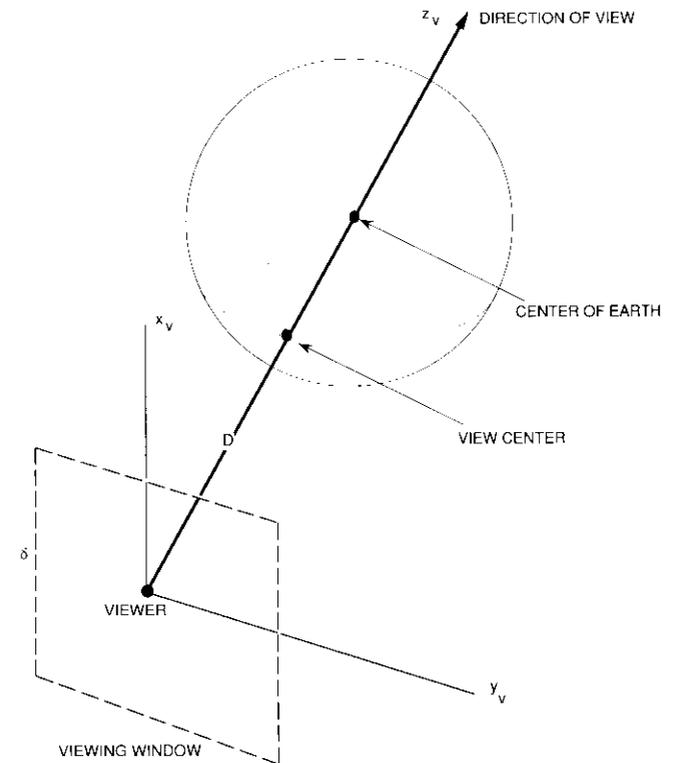


Figure 6. Viewing System: Orthographic Projection

A viewing window must also be specified to define the extent of the view. For example, a square viewing window δ units on a side will cover the region defined by

$$\begin{aligned} -\delta/2 &\leq x_v \leq \delta/2 \\ -\delta/2 &\leq y_v \leq \delta/2 \end{aligned}$$

Any point falling outside this region will not be displayed, nor will any point hidden by the earth (*i.e.*, a point for which $z_v > D + R$).

The coordinates in the projection plane (x_p, y_p) that are actually displayed on the screen are the same as the viewing system coordinates, that is,

$$\begin{aligned} x_p &= y_v \\ y_p &= x_v \end{aligned}$$

PERSPECTIVE PROJECTION

A perspective projection is an attempt to represent the earth as seen in a perspective view from a specific location a finite distance, D , from the earth's surface. In a perspective projection, portions of the earth closer to the viewer will appear larger than portions further away. An example of a perspective projection is given in Figure 7, which shows the earth as seen from a distance of 1,000 miles (*i.e.*, $D = 1,000$).

In the special case where D is equal to the satellite range, the earth will be shown as it appears from the satellite. The orientation of the viewing system is defined as in the case of the orthographic projection; however, the coordinates displayed in this case are the angular coordinates α_v and β_v (as shown in Figure 8) rather than the linear coordinates x_v and y_v . The angular coordinates are related to the linear coordinates by

$$\begin{aligned} \alpha_v &= \tan^{-1}(x_v/z_v) \\ \beta_v &= \tan^{-1}(y_v/z_v) \end{aligned}$$

Spherical coordinates may then be determined from

$$\begin{aligned} \tan^2 \theta_s &= \tan^2 \alpha_v + \tan^2 \beta_v \\ \tan \phi_s &= \tan \beta_v / \tan \alpha_v \end{aligned}$$

Finally, the coordinates in the projection plane (x_p, y_p) may be obtained from the spherical coordinates and normalized to one earth radius by

$$\begin{aligned} x_p &= K\theta \sin \phi \\ y_p &= K\theta \cos \phi \end{aligned}$$



Figure 7. Perspective View of the Earth

where K is the normalization factor given by

$$K = R/\phi = R/\sin^{-1}[R/(D + R)]$$

The viewing window is a square of size δ , such that

$$\begin{aligned} -\delta/2 &\leq x_p \leq \delta/2 \\ -\delta/2 &\leq y_p \leq \delta/2 \end{aligned}$$

Any point falling outside this region will not be displayed, nor will any point hidden by the earth [*i.e.*, a point for which $z_v > D + R - R^2/(D + R)$].

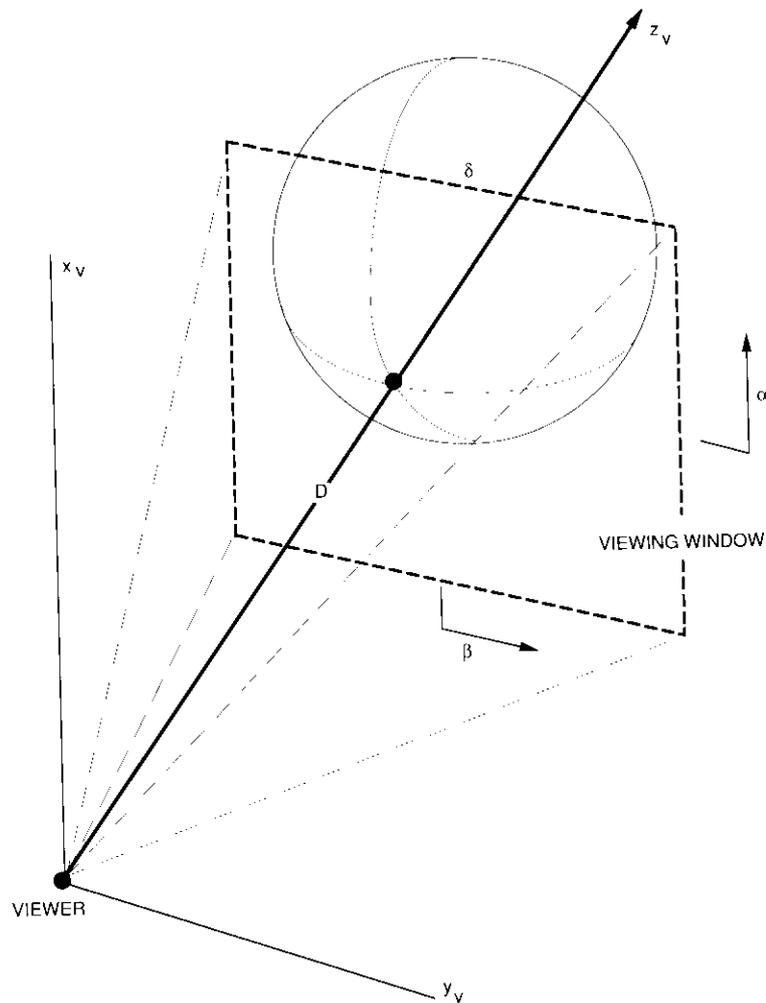


Figure 8. Viewing System: Perspective Projection

EQUIRECTANGULAR PROJECTION

In an equirectangular projection, the Earth's surface is displayed with lines of latitude and longitude forming a square grid. Figure 9 is an example of an equirectangular projection.

The viewing system for an equirectangular projection is shown in Figure 10. In this case, the x_v - y_v viewing plane is parallel to the plane of the flat projection, with the positive x_v -axis pointing north. The units of x_v are in degrees latitude, and the units of y_v are in degrees longitude. The range of the coordinates for a square viewing window of aperture δ and a view center (ρ_{view} , μ_{view}) are

$$\begin{aligned} \rho_{\text{view}} - \delta/2 &\leq x_v \leq \rho_{\text{view}} + \delta/2 \\ \mu_{\text{view}} - \delta/2 &\leq y_v \leq \mu_{\text{view}} + \delta/2 \end{aligned}$$

Any value falling outside this range will not be displayed.

Coordinate transformations

In the process of determining satellite antenna coverages, coordinate values must often be transformed from one coordinate system to another. This section presents a general solution to the problem of transforming coordinates from one Cartesian system to another. The results are then particularized to relate the satellite-centered system and the viewing system to the earth-centered system. Finally, the relationships between Cartesian and angular (e.g., latitude and longitude) coordinates in each system are defined.

General Cartesian transformations

Consider the local Cartesian coordinate system having coordinate axes x_s , y_s and z_s , which is shown in Figure 11. (For the moment, this system may represent either the satellite-centered system or the viewing system.) The orientation of this system with respect to the earth-centered system (x_e , y_e , z_e) may be defined by specifying the components of three vectors (r , s , and t) in the earth-centered system. Vector r defines the origin of the local system, vector s defines the direction of the positive z_s -axis, and vector t lies anywhere in the z_s - x_s plane. The cross product of s and t defines the positive y_s -axis, so that t may be chosen to lie along the positive x_s -axis. Denoting the vector components by

$$\begin{aligned} r &= r_1 i + r_2 j + r_3 k \\ s &= s_1 i + s_2 j + s_3 k \\ t &= t_1 i + t_2 j + t_3 k \end{aligned}$$

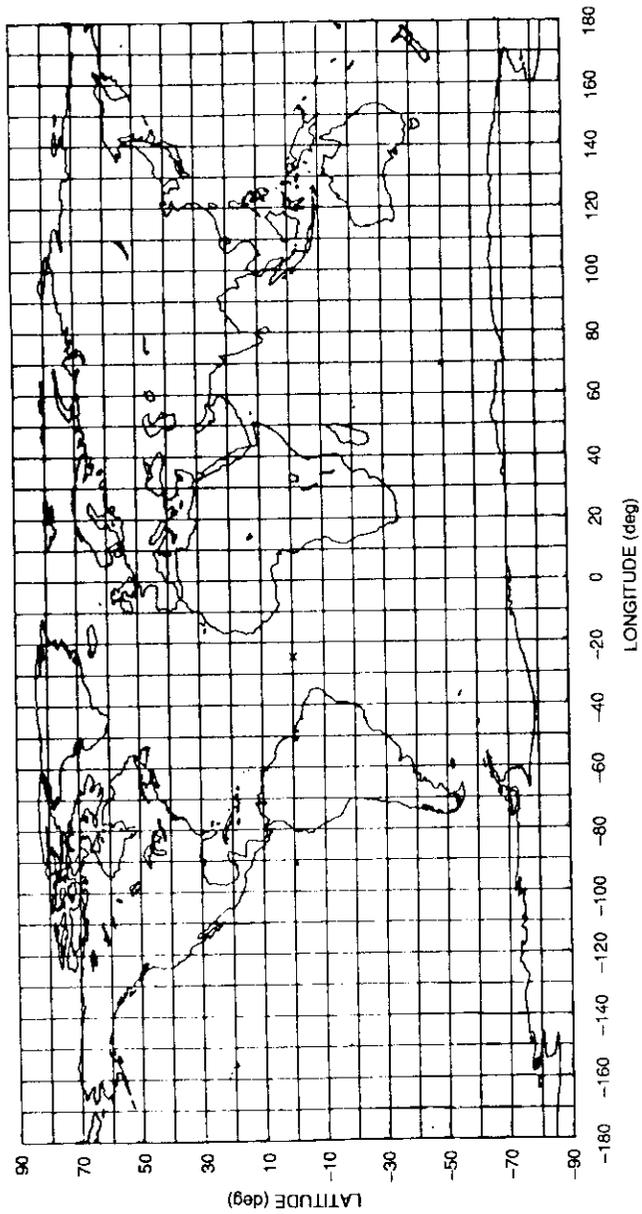


Figure 9. Equirectangular Plot of the Earth

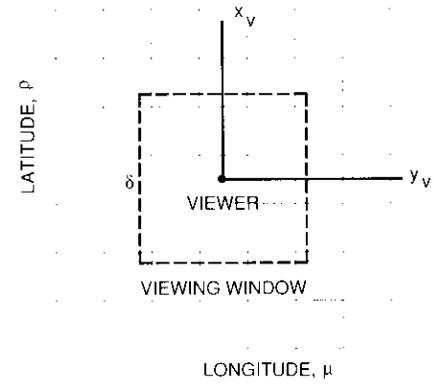


Figure 10. Viewing System: Equirectangular Projection

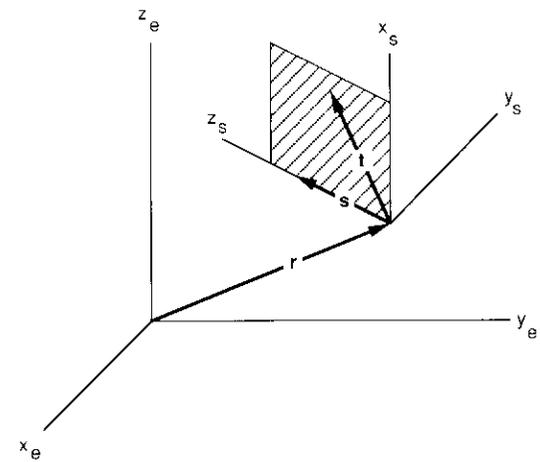


Figure 11. Local System Orientation

where i , j , and k are the unit vectors in the earth-centered system, the unit vectors l , m , and n in the local system may be written as

$$\begin{aligned} l &= a_1 i + a_2 j + a_3 k \\ m &= a_4 i + a_5 j + a_6 k \\ n &= a_7 i + a_8 j + a_9 k \end{aligned}$$

where

$$\begin{aligned} a_1 &= [(s_3 t_1 - s_1 t_3) s_3 - (s_1 t_2 - s_2 t_1) s_2] / M_1 \\ a_2 &= [(s_1 t_2 - s_2 t_1) s_1 - (s_2 t_3 - s_3 t_2) s_3] / M_1 \\ a_3 &= [(s_2 t_3 - s_3 t_2) s_2 - (s_3 t_1 - s_1 t_3) s_1] / M_1 \\ a_4 &= (s_2 t_3 - s_3 t_2) / M_2 \\ a_5 &= (s_3 t_1 - s_1 t_3) / M_2 \\ a_6 &= (s_1 t_2 - s_2 t_1) / M_2 \\ a_7 &= s_1 / M_3 \\ a_8 &= s_2 / M_3 \\ a_9 &= s_3 / M_3 \end{aligned}$$

and where M_1 , M_2 , and M_3 are the magnitudes of $s \times t \times s$, $s \times t$, and s , respectively. (Note that the magnitude of the cross product of two vectors, a and b , is equal to $|a| |b| \sin \theta$, where $|a|$ and $|b|$ are the magnitudes of the vectors and θ is the angle between them.)

Any point (x_s, y_s, z_s) in the local system may be transformed to a point (x_e, y_e, z_e) in the earth-centered system by

$$\begin{Bmatrix} x_e \\ y_e \\ z_e \end{Bmatrix} = \begin{bmatrix} a_1 & a_4 & a_7 \\ a_2 & a_5 & a_8 \\ a_3 & a_6 & a_9 \end{bmatrix} \begin{Bmatrix} x_s \\ y_s \\ z_s \end{Bmatrix} + \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \end{Bmatrix}$$

The inverse transformation, which transforms coordinates in the earth-centered system to the local system, is

$$\begin{Bmatrix} x_s \\ y_s \\ z_s \end{Bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{Bmatrix} x_e \\ y_e \\ z_e \end{Bmatrix} - \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \end{Bmatrix}$$

Satellite-centered system to earth-centered system

The results of the preceding section may be applied to determine the transformation between the satellite-centered system and the earth-centered system. Assume that the radius of the earth is given by R and the range of the satellite (*i.e.*, its distance from the subsatellite point) is given by S .

Further assume that the subsatellite point latitude and longitude are given by ρ_{sat} and μ_{sat} , respectively. Then the components of r , s , and t may be specified as follows:

$$\begin{aligned} r_1 &= (R + S) \cos \rho_{\text{sat}} \cos \mu_{\text{sat}} \\ r_2 &= (R + S) \cos \rho_{\text{sat}} \sin \mu_{\text{sat}} \\ r_3 &= (R + S) \sin \rho_{\text{sat}} \\ s_1 &= -\cos \rho_{\text{sat}} \cos \mu_{\text{sat}} \\ s_2 &= -\cos \rho_{\text{sat}} \sin \mu_{\text{sat}} \\ s_3 &= -\sin \rho_{\text{sat}} \\ t_1 &= 0 \\ t_2 &= 0 \\ t_3 &= 1 \end{aligned}$$

Note that $M_3 = 1$, since s is a unit vector. M_1 and M_2 are equal to 1 only in the case where the viewer is located in the plane of the Equator. The coefficients of the transformation matrix are then

$$\begin{aligned} a_1 &= -\sin \rho_{\text{sat}} \cos \rho_{\text{sat}} \cos \mu_{\text{sat}} / M_1 \\ a_2 &= -\sin \rho_{\text{sat}} \cos \rho_{\text{sat}} \sin \mu_{\text{sat}} / M_1 \\ a_3 &= \cos^2 \rho_{\text{sat}} / M_1 \\ a_4 &= -\cos \rho_{\text{sat}} \sin \mu_{\text{sat}} / M_2 \\ a_5 &= \cos \rho_{\text{sat}} \cos \mu_{\text{sat}} / M_2 \\ a_6 &= 0 \\ a_7 &= -\cos \rho_{\text{sat}} \cos \mu_{\text{sat}} \\ a_8 &= -\cos \rho_{\text{sat}} \sin \mu_{\text{sat}} \\ a_9 &= -\sin \rho_{\text{sat}} \end{aligned}$$

The above coefficients are found by using the steps explained in the previous section. The inverse transformation is used to convert earth-centered coordinates to satellite-centered coordinates.

Viewing system to earth-centered system

In a similar manner, the transformation between the viewing system and the earth-centered system may be derived. Assume that the radius of the earth is given by R and the distance of the viewer from the surface of the earth is given by D . Further assume that the viewer is looking directly down on the surface of the earth to a point whose latitude and longitude are given by ρ_{view} and μ_{view} , respectively. Then the components of r , s , and t may be specified as follows:

$$\begin{aligned}
 r_1 &= (R + D) \cos \rho_{\text{view}} \cos \mu_{\text{view}} \\
 r_2 &= (R + D) \cos \rho_{\text{view}} \sin \mu_{\text{view}} \\
 r_3 &= (R + D) \sin \rho_{\text{view}} \\
 s_1 &= -\cos \rho_{\text{view}} \cos \mu_{\text{view}} \\
 s_2 &= -\cos \rho_{\text{view}} \sin \mu_{\text{view}} \\
 s_3 &= -\sin \rho_{\text{view}} \\
 t_1 &= 0 \\
 t_2 &= 0 \\
 t_3 &= 1
 \end{aligned}$$

Note that $M_3 = 1$, since s is a unit vector. M_1 and M_2 are equal to 1 only in the case where the viewer is located in the plane of the Equator. The coefficients of the transformation matrix are then

$$\begin{aligned}
 a_1 &= -\sin \rho_{\text{view}} \cos \rho_{\text{view}} \cos \mu_{\text{view}} / M_1 \\
 a_2 &= -\sin \rho_{\text{view}} \cos \rho_{\text{view}} \sin \mu_{\text{view}} / M_1 \\
 a_3 &= \cos^2 \rho_{\text{view}} / M_1 \\
 a_4 &= -\cos \rho_{\text{view}} \sin \mu_{\text{view}} / M_2 \\
 a_5 &= \cos \rho_{\text{view}} \cos \mu_{\text{view}} / M_2 \\
 a_6 &= 0 \\
 a_7 &= -\cos \rho_{\text{view}} \cos \mu_{\text{view}} \\
 a_8 &= -\cos \rho_{\text{view}} \sin \mu_{\text{view}} \\
 a_9 &= -\sin \rho_{\text{view}}
 \end{aligned}$$

Analysis algorithms

This section presents the algorithms used to create the world map, to plot great circle paths, to calculate azimuth and elevation angles of specified locations and plot elevation angle contours, to calculate and plot antenna gain and flux density patterns, and to plot the effects of satellite pointing errors.

Geographic features

Geographic features may be plotted on an orthographic, perspective, or equirectangular map to serve as a background for displaying antenna pattern data. These geographic features include earth station or city locations, and continent, state, or country boundaries.

Each earth station or city location is defined by its coordinates in latitude and longitude (ρ_e, μ_e). Each pair of coordinates must first be transformed to

Cartesian earth-centered coordinates and then to the viewing system for display.

Continent, country, or state boundaries are typically represented by one or more groups of points, each point being defined by its Cartesian coordinates (x_e, y_e, z_e) in the earth-centered system. Consecutive points within each group are connected by straight lines in the display, following their transformation to the viewing system.

Figure 12 is a plot of the United States including state boundaries and the locations of certain selected cities.

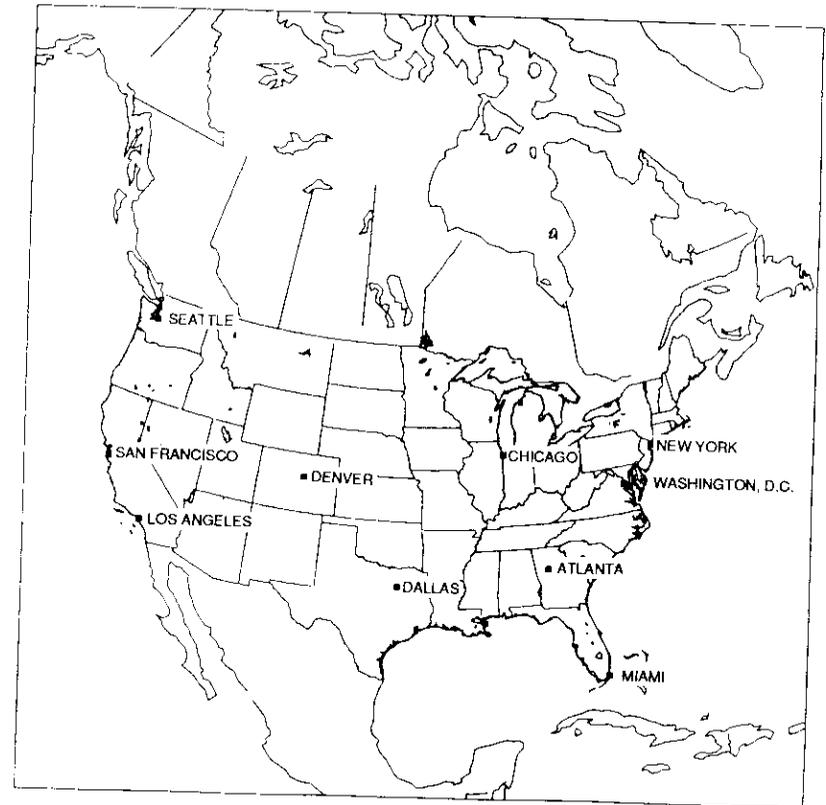


Figure 12. Plot Showing State Boundaries and Cities

Great circle path

A great circle path represents the shortest distance between any two specified points on the earth's surface. A great circle is formed by the intersection of the sphere representing the surface of the earth and a plane passing through the center of the earth and the two specified points (see Figure 13).

The algorithm for generating a great circle path is as follows:

1. Convert the latitude and longitude of the end points (ρ_a, μ_a) and (ρ_b, μ_b) to Cartesian coordinates in the earth-centered system by means of the equations

$$\begin{aligned} x_a &= R \cos \rho_a \cos \mu_a \\ y_a &= R \cos \rho_a \sin \mu_a \\ z_a &= R \sin \rho_a \end{aligned}$$

and

$$\begin{aligned} x_b &= R \cos \rho_b \cos \mu_b \\ y_b &= R \cos \rho_b \sin \mu_b \\ z_b &= R \sin \rho_b \end{aligned}$$

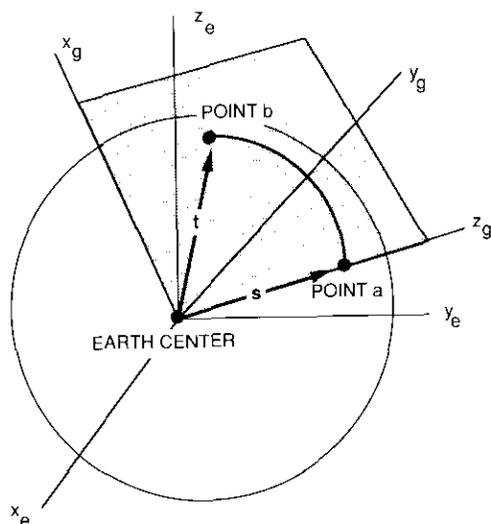


Figure 13. Great Circle Paths

2. Define a local coordinate system (x_g, y_g, z_g) such that the y_g - z_g plane contains the great circle path connecting points a and b . Specifically, the origin of the local system lies at the center of the earth, the positive z_g -axis intersects point a , and the x_g - z_g plane contains point b . Hence, the components of the r, s , and t vectors defining this system in terms of the earth-centered system are

$$\begin{aligned} r_1 &= 0 \\ r_2 &= 0 \\ r_3 &= 0 \\ s_1 &= x_a \\ s_2 &= y_a \\ s_3 &= z_a \\ t_1 &= x_b \\ t_2 &= y_b \\ t_3 &= z_b \end{aligned}$$

3. Convert the coordinates of point b from the earth-centered system to the local system.

The angular separation between points a and b is found by

$$\Delta\theta = \cos^{-1} [s \cdot t / (|s| |t|)]$$

where $s \cdot t$ is the dot product of vector s and vector t . Next calculate n points along the circular arc between a and b and connect consecutive points with straight line segments. The coordinates of each point in the local system are given by

$$\begin{aligned} x_i &= R \sin (i \Delta\theta / n) & 1 \leq i \leq n \\ y_i &= 0 \\ z_i &= R \cos (i \Delta\theta / n) \end{aligned}$$

Each point is first converted to the earth-centered system and then to the viewing system prior to plotting.

Azimuth and elevation angles

The azimuth and elevation angles define the direction of the satellite as seen from a specific earth station location. The azimuth is the angle between the satellite's angular position (as seen from a point on the earth) and true North, while the elevation is the angle between the satellite's angular position and the (sea-level) horizon. Elevation angles below some nominal value

(typically 5° to 10°) may result in significant signal degradation due to atmospheric disturbances and multipath effects. Hence, elevation angle contours are often used in conjunction with satellite antenna pattern contours to determine the effective coverage area of a satellite antenna beam.

The following procedure is used to determine the azimuth and elevation for an earth station whose latitude and longitude are given by ρ_e and μ_e , respectively, and a satellite at a distance S from the earth, whose latitude and longitude are given by ρ_s and μ_s , respectively. First, the earth station and satellite locations must be identified in the basic Cartesian coordinate system. The earth station location is

$$\begin{aligned} x_e &= R \cos \rho_e \cos \mu_e \\ y_e &= R \cos \rho_e \sin \mu_e \\ z_e &= R \sin \rho_e \end{aligned}$$

and the satellite location is

$$\begin{aligned} x_s &= (R + S) \cos \rho_s \cos \mu_s \\ y_s &= (R + S) \cos \rho_s \sin \mu_s \\ z_s &= (R + S) \sin \rho_s \end{aligned}$$

Next, a new local coordinate system $(x, y, z)_L$ is created such that the origin is at the earth station location (x_e, y_e, z_e) , the z -axis is normal to the surface of the earth, and the x -axis is pointing due north. The components of the r , s , and t vectors defining the local system in terms of the basic system are

$$\begin{aligned} r_1 &= x_e \\ r_2 &= y_e \\ r_3 &= z_e \\ s_1 &= x_e \\ s_2 &= y_e \\ s_3 &= z_e \\ t_1 &= 0 \\ t_2 &= 0 \\ t_3 &= 1 \end{aligned}$$

The satellite location is then transformed from the basic system (x, y, z) to the local system $(x_s, y_s, z_s)_L$, and the azimuth (α) and elevation (ϵ) angles, as shown in Figure 14, are obtained from the following equations:

$$\begin{aligned} \alpha &= -\tan^{-1}(y_s/x_s) \\ \epsilon &= \sin^{-1}\left(\frac{z_s}{\sqrt{x_s^2 + y_s^2 + z_s^2}}\right) \end{aligned}$$

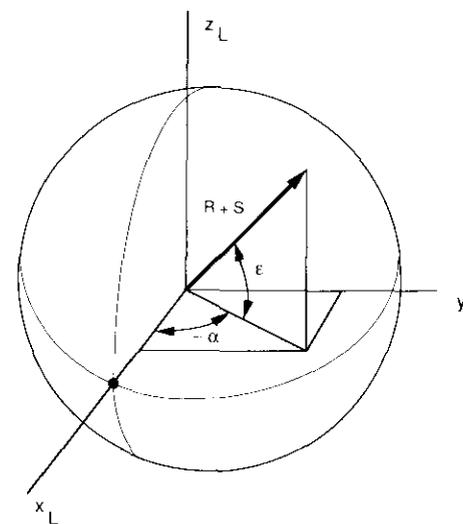


Figure 14. Azimuth and Elevation Angles

with the additional restrictions that

$$\begin{aligned} \alpha &= \pi && \text{when } \rho_e = \pi/2 \\ &= 0 && \text{when } \rho_e = -\pi/2 \\ &= 0 && \text{when } \rho_e = \rho_s \text{ and } \mu_e = \mu_s \end{aligned}$$

Contours of constant elevation angle may be generated and plotted for one or more values of elevation angle, ϵ . Figure 15 shows a slice through the center of the earth in a plane containing the satellite and a point on the earth whose elevation angle is to be calculated. The point on the earth is defined in the satellite-centered system by the coordinates (T_s, θ_s, ϕ_s) , and the elevation angle is denoted by ϵ .

The process begins by finding a value for θ_s which corresponds to an elevation angle of ϵ . The relationship between θ_s and ϵ may be derived by equating two separate expressions for the line segment from the earth center and perpendicular to the earth station line of sight, that is,

$$R \cos \epsilon = (S + R) \sin \theta_s$$

It immediately follows that

$$\theta_s = \sin^{-1}[(R \cos \epsilon)/(S + R)]$$

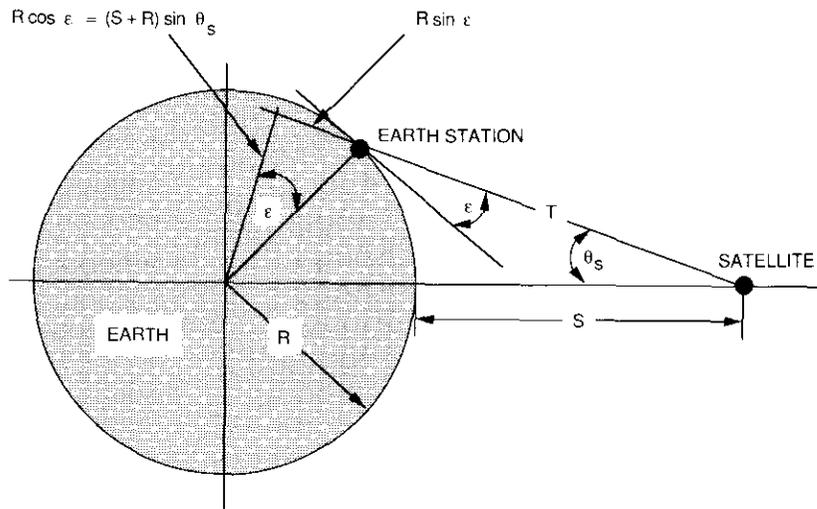


Figure 15. Earth/Satellite Geometry

Knowing θ_s , the distance T_s may be determined from

$$T_s = (S + R) \cos \theta_s - R \sin \epsilon$$

A contour circle is drawn by looping over a set of values for ϕ_s , spanning the region 0 to 2π radians. For each value of ϕ_s , the coordinate set (T_s, θ_s, ϕ_s) is converted first to the Cartesian satellite-centered system, then to the Cartesian earth-centered system, and finally to the viewing system for plotting.

An example of a plot of contours of constant elevation angle is shown in Figure 16. Contours for elevation angles of 0° , 5° , and 10° are shown on an equirectangular projection of the earth.

Antenna gain patterns

A variety of options are provided for the specification of antenna gain patterns. In cases where a simple representation of the pattern will suffice, the pattern may be approximated by a circular beam, an elliptical beam, or a shaped beam. The circular beam is defined by its boresight direction and 3-dB beamwidth; the elliptical beam is defined by its boresight direction, 3-dB beamwidths along the major and minor axes, and the orientation of the major axis from the horizon. In either of these cases, the shape of the radiation pattern depends on the assumed distribution of power over the aperture. Shaped beam patterns are formed by illuminating the antenna

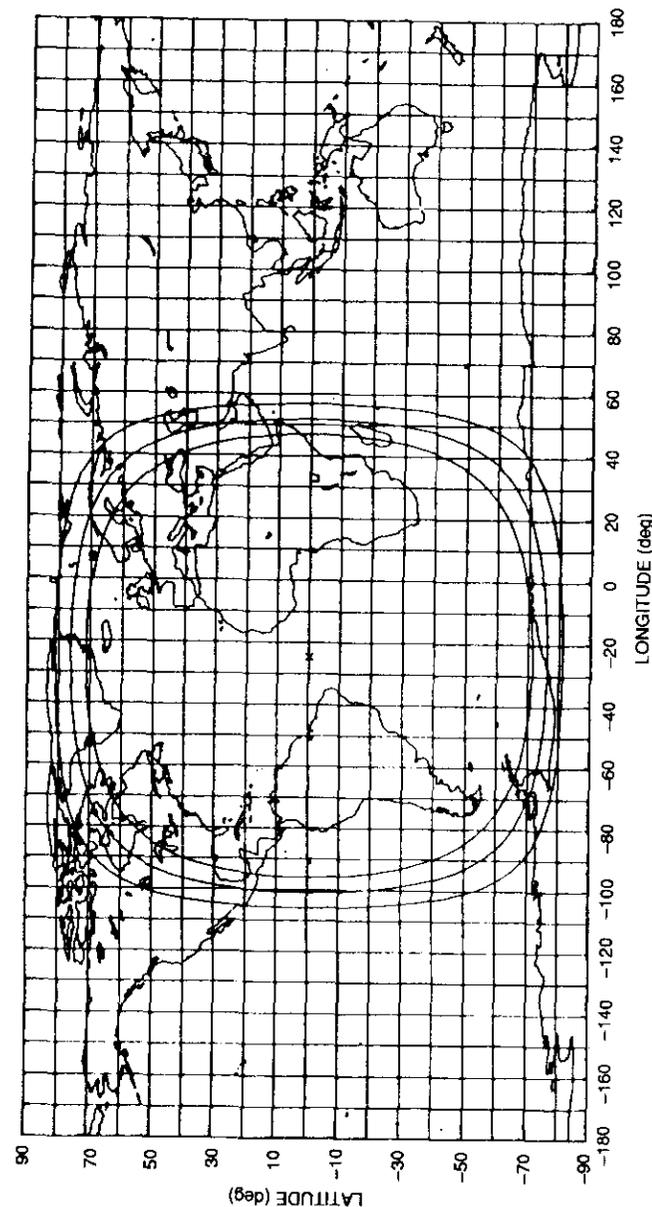


Figure 16. Elevation Angle Contours

reflector with a number of individual feeds, each oriented in a specific direction and having a specified amplitude and phase. When greater precision is needed, the pattern can be represented by specifying the gain at equally spaced values of (θ and ϕ) in the antenna coordinate system.

CIRCULAR BEAM

A model commonly used to approximate the aperture distribution for a circular antenna is

$$f(r) = (1 - r^2)^p$$

When the exponent p is zero, the aperture distribution is uniform. As p increases, the distribution becomes more highly tapered toward the edge of the aperture [2]. The far-field radiation pattern, $E(\theta)$, generated by this distribution is proportional to

$$E(\theta) \sim J_{p+1}(u) / u^{p+1}$$

and the gain distribution is proportional to

$$G(\theta) \sim J_{p+1}^2(u) / u^{2(p+1)}$$

where $u = (\pi D / \lambda) \sin \theta$, and where $J_p(x)$ is the Bessel function of the first kind of order p and argument x .

The 3-dB (or half-power) beamwidth, B , has the following values (in degrees), depending on the value of p :

$$\begin{aligned} B &= 58.91 / D & p &= 0 \\ B &= 72.71 / D & p &= 1 \\ B &= 84.31 / D & p &= 2 \end{aligned}$$

Hence,

$$\begin{aligned} u &= (58.9 \pi / B) \sin \theta & p &= 0 \\ u &= (72.7 \pi / B) \sin \theta & p &= 1 \\ u &= (84.3 \pi / B) \sin \theta & p &= 2 \end{aligned}$$

An n -by- n antenna pattern grid is created which represents equally spaced points in α and β . The distance between the circular beam center, c , on the grid and each grid point, p , is then substituted for θ in the gain equation. This distance may be determined from

$$\tan^2 \theta = \tan^2 (\alpha_p - \alpha_c) + \tan^2 (\beta_p - \beta_c)$$

The gain $G(\alpha, \beta)$ is computed for each grid point and accumulated in a corresponding n -by- n array representing the equally spaced points in α and β . The gain is normalized by dividing all gain values by the maximum gain, and then converted to dB. The gain contours are generated and plotted using the contour generation algorithm described later in this paper.

ELLIPTICAL BEAM

The same model for the far-field radiation pattern is used for elliptical beams as for circular beams. In the case of an elliptical beam, the 3-dB beamwidth, B , for a cut through the pattern at ξ degrees is a function of the semi-major axis (e), the semi-minor axis (f), and the orientation angle of the major axis (η) (see Figure 17). The equation for the elliptical beamwidth is

$$B = [\cos(\eta - \xi)^2 / e^2 + \sin(\eta - \xi)^2 / f^2]^{-1/2}$$

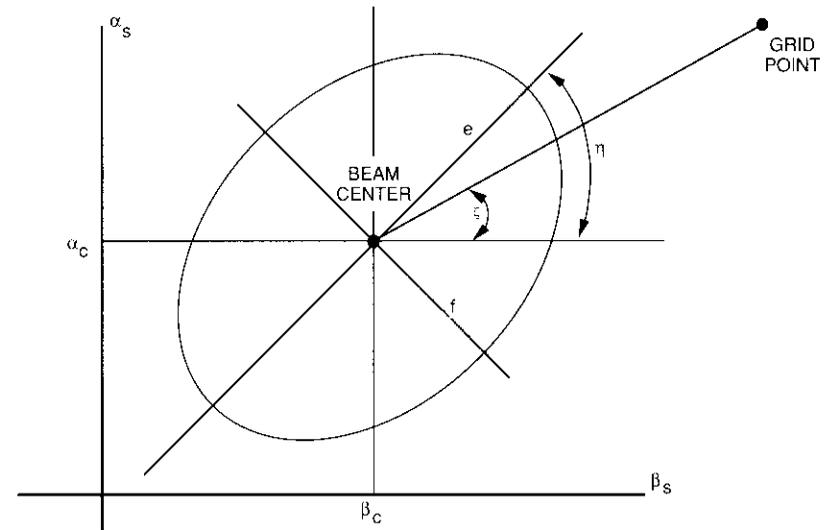


Figure 17. Ellipse Parameters

SHAPED BEAM

One approach to designing beam coverage patterns that have complex shapes is to use a cluster of antenna feeds. The orientation of each feed is adjusted so that a desired composite pattern is projected onto the surface of the earth (see Figure 18). The far-field radiation pattern generated by each

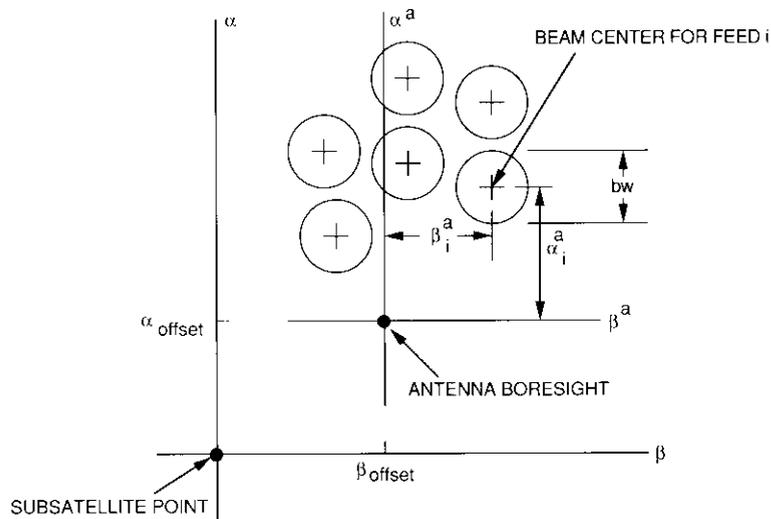


Figure 18. Multiple Antenna Feeds

feed is calculated from the characteristics of the radiation field over the antenna aperture, and from the relative amplitude and phase weighting associated with each feed. In this approach, the following information must be provided for each feed, i :

- Beam pointing angles (α_i^a and β_i^a) in the antenna boresight coordinate system
- Relative amplitude (A_i)
- Relative phase (ψ_i).

Parameters which are assumed to be constant for all feeds are the antenna diameter (d), the aperture edge taper (t), the taper index (n), and the wavelength of the radiation (λ). The boresight coordinate system is presumed to be offset from the subsatellite point by the angular coordinates α_{offset} and β_{offset} .

The aperture fields associated with each feed are approximated by analytical expressions which are integrable in closed form. First a complex amplitude, c_i , is defined for each feed, which is a function of the relative feed amplitude, A_i , and phase, ψ_i , as

$$c_i = A_i \exp(j\psi_i) = A_i (\cos \psi_i + j \sin \psi_i)$$

The beam center direction of each feed in the satellite-centered system is defined by the pointing angles α_i and β_i , where

$$\alpha_i = \alpha_i^a + \alpha_{\text{offset}}$$

$$\beta_i = \beta_i^a + \beta_{\text{offset}}$$

Equivalently, these directions may also be expressed in spherical coordinates, θ_i and ϕ_i , as

$$\theta_i = \tan^{-1} \left(\sqrt{\tan^2 \alpha_i^2 + \tan^2 \beta_i^2} \right)$$

$$\phi_i = \tan^{-1} (\tan \beta_i / \tan \alpha_i)$$

The complex amplitude $C_i(\alpha, \beta)$ at any arbitrary point in the far field due to an individual feed patterns, i , is given by

$$C_i(\alpha, \beta) = c_i (2\pi r / \lambda) [2J_1(\sigma_i) / \sigma_i + \tau 2^{n+1} n! J_{n+1}(\sigma_i) / \sigma_i^{n+1}] \cdot \sqrt{1 + [2\tau / (n + 1)] + [\tau^2 / (2n + 1)]}$$

where $\tau = t - 1$

$$\sigma_i = (2\pi r / \lambda) \tan^{-1} \sqrt{(\tan \alpha - \tan \alpha_i)^2 + (\tan \beta - \tan \beta_i)^2}$$

The real and imaginary components of the total radiation field amplitude, $C(\alpha, \beta)$, are found by summing the contributions of each individual feed, that is,

$$\text{Re}\{C(\alpha, \beta)\} = \sum_{i=1}^N \text{Re}\{C_i(\alpha, \beta)\}$$

$$\text{Im}\{C(\alpha, \beta)\} = \sum_{i=1}^N \text{Im}\{C_i(\alpha, \beta)\}$$

where N is the number of feeds. The contributions of each individual feed to the far-field pattern are accumulated in a complex n -by- n array representing points equally spaced in α and β . At the completion of this process, the real and imaginary amplitude values are squared and summed, and the resulting gain values are normalized to a maximum value of unity and converted to dB. The gain contours are generated and plotted according to the contour generation algorithm described below.

Examples of shaped beam patterns are shown in Figures 19 and 20, which are perspective views of the earth as seen from the satellite. The 3-dB contours in Figure 19 are the circular beams generated by each of 10 feeds used to generate the shaped beam. Figure 20 shows the 1-, 2-, and 3-dB gain contours of the composite field created by summing the patterns of all 10 feeds, assuming a uniform amplitude and phase weighting for each.



Figure 19. Beams Generated by Each Individual Feed

MEASURED BEAM PATTERN

Antenna patterns that cannot be represented by any of the relatively simplistic models presented above can still be accommodated by the program provided that gain values $G(\bar{\theta}, \bar{\phi})$ are known as a set of equally spaced values of $\bar{\theta}$ and $\bar{\phi}$ in the antenna coordinate system. The corresponding values of the pointing angles in the satellite-centered coordinate system are

$$\alpha = \pi/2 - \bar{\theta}$$

$$\beta = \pi/2 - \bar{\phi}$$

In this case, the gain values $G(\alpha, \beta) = G(\bar{\theta}, \bar{\phi})$ are already in the proper form required for processing by the contour generation algorithm.

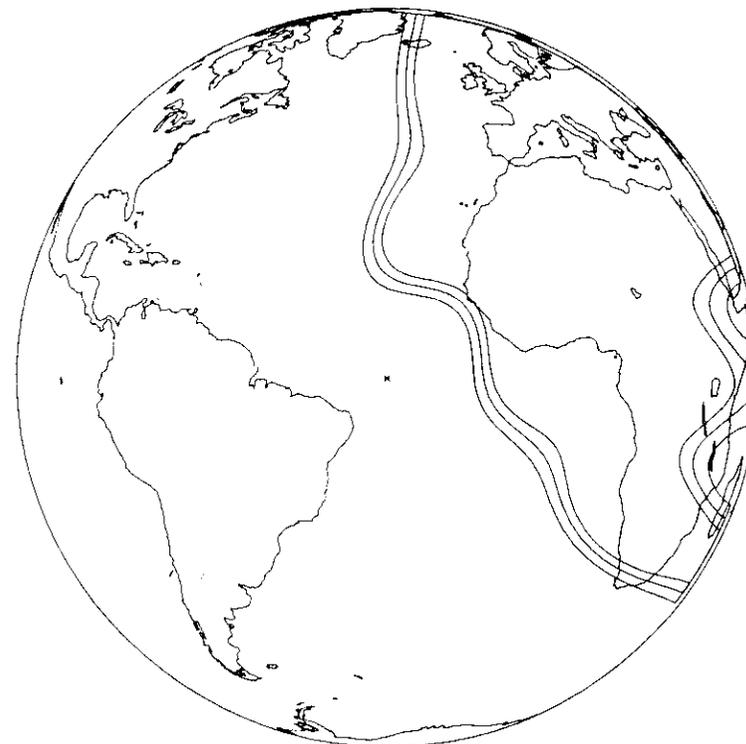


Figure 20. Composite Field From All Feeds

USER-SPECIFIED GAIN CONTOUR

An alternative means of specifying an antenna pattern is to list a number of discrete points that define the outline of the coverage region (*i.e.*, a contour of equal gain values). The angular coordinates of each point, α and β , are specified in the satellite-centered coordinate system.

The user-specified gain contour is plotted by the following procedure. Each point in angular coordinates is first converted to spherical satellite-centered coordinates. If the point is visible from the viewer's position, it is converted to Cartesian satellite-centered coordinates and then to earth-centered coordinates and added to an array of points to be plotted. If the point is not visible (*i.e.*, it falls off the edge of the earth as seen from the viewer), a point on the edge of the earth closest to the original point is generated.

converted to earth-centered coordinates, and added to the array of points to be plotted. Next, each point in the array of earth-centered points is converted to the appropriate viewing system and then to plotting coordinates, and a closed line is drawn connecting all the points on the plot.

Flux density patterns

Flux density patterns show contours of constant normalized flux density on the earth's surface. Flux density is proportional to the antenna gain and inversely proportional to the path loss. The path loss is in turn proportional to the square of the distance between the satellite and the point on the earth's surface.

Figure 21 shows a slice through the center of the earth in a plane containing the satellite and a point on the earth, *p*, for which the flux density is to be calculated. The coordinates of point *p* in a satellite-centered system are (*T*, θ_s , ϕ_s). The path loss factor is a function of the distance, *S* + *R*, between the satellite and the earth's center, and the distance, *T*, between the satellite and point *p*. The equation for the path loss factor, *L*, is

$$L = (S/T)^2$$

where the distance *T* is found to be

$$T = (S + R) \cos \theta_s - \sqrt{R^2 - (S + R)^2 \sin^2 \theta_s}$$

After the path loss factor is computed, it is converted to dB using the equation

$$L_{(dB)} = 10 \log_{10}(L)$$

and subtracted from the normalized gain, *G*(θ , ϕ), to produce a normalized flux density, *F*(θ , ϕ), as

$$F(\theta, \phi) = G(\theta, \phi) - L_{(dB)}(\theta, \phi)$$

for a shaped beam, and

$$F(\theta) = G(\theta) - L_{(dB)}(\theta)$$

for a circular or elliptical beam.

Satellite pointing errors

Errors in the satellite pointing direction due to motion of the satellite can significantly affect the satellite antenna gain at specific points on the earth's

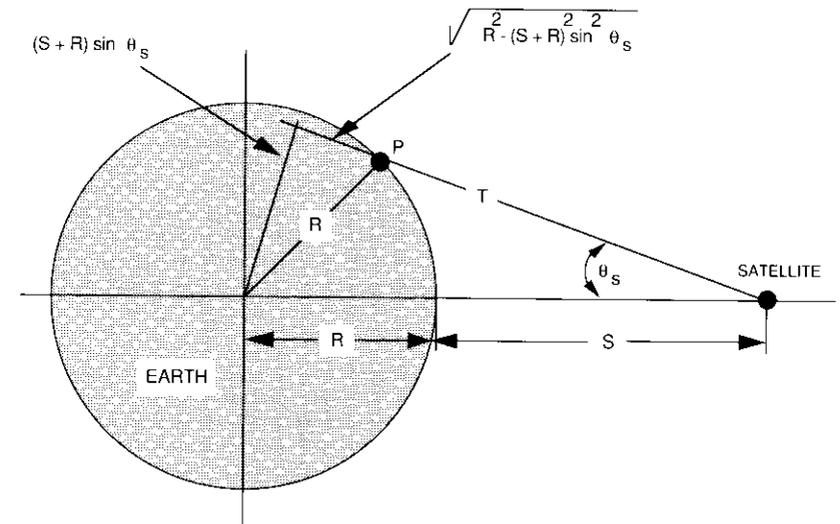


Figure 21. Path Loss Calculations

surface, particularly in the case of narrow spot beams. For this reason it is frequently of interest to graphically represent the impact of satellite pointing errors on the antenna gain or flux density patterns.

Satellite pointing errors are typically described in terms of the maximum values of inclination (*i.e.*, angular coordinate α in the north-south direction) and attitude (*i.e.*, angular coordinate β in the east-west direction). These maximum angular errors are defined by $\pm \alpha_e$ and $\pm \beta_e$, respectively. The effect of these errors may be displayed by means of either "pointing error boxes" or a worst-case antenna gain pattern plot.

POINTING ERROR BOXES

Pointing error boxes indicate the range of earth station locations relative to satellite antenna gain or flux density pattern. Each box is generated by plotting the outline of a region defined by

$$\alpha_0 - \alpha_e \leq \alpha \leq \alpha_0 + \alpha_e$$

$$\beta_0 - \beta_e \leq \beta \leq \beta_0 + \beta_e$$

where (α_0 , β_0) are the angular coordinates of the earth station of interest in the satellite-centered coordinate system. Each side of the box is divided into

a set of n rays defined by the angular coordinates (α_i, β_i) , $1 \leq i \leq n$. For example, the bottom side is represented by

$$\alpha_i = \alpha_0 - \alpha_e \quad 1 \leq i \leq n$$

$$\beta_i = \beta_0 - \beta_e + 2\beta_e(i - 1)/(n - 1)$$

where n is the number of rays in each coordinate direction. For each ray (α_i, β_i) the spherical coordinates (θ_i, ϕ_i) are determined from

$$\theta_i = \tan^{-1}(\sqrt{\tan^2 \alpha_i + \tan^2 \beta_i})$$

$$\phi_i = \tan^{-1}(\tan \beta_i / \tan \alpha_i)$$

The next step is to determine the distance, T_i , from the satellite to the point where the ray intersects the earth's surface. From Figure 21, the following relationship can be derived:

$$T_i = (S + R) \cos \theta_i - R \sin \{\cos^{-1}[(S + R) \sin \theta_i / R]\}$$

The coordinates (T_i, θ_i, ϕ_i) of the intersection points are first converted to Cartesian coordinates in the satellite-centered system, then to the earth-centered system, and finally to the viewing system. Consecutive corners of the pointing error box are then connected by straight line segments.

An example of the display of pointing error boxes is shown in the equirectangular projection of Figure 22. By comparing the boxes surrounding each earth station location with the gain contours, it is possible to determine the variation in gain caused by instabilities in the orientation of the satellite.

WORST-CASE ANTENNA PATTERN

An alternative means of displaying the effects of satellite pointing errors is by a modified antenna coverage pattern which represents a worst-case situation. When generating antenna gain values on the n -by- n antenna pattern grid, the following four beam center locations are used in place of (α_j, β_j) for beam j :

$$(\alpha_j - \alpha_e, \beta_j - \beta_e)$$

$$(\alpha_j + \alpha_e, \beta_j - \beta_e)$$

$$(\alpha_j - \alpha_e, \beta_j + \beta_e)$$

$$(\alpha_j + \alpha_e, \beta_j + \beta_e)$$

Each point in the antenna pattern grid is assigned a value equal to the worst (lowest) of the four gain values calculated. The resulting pattern contours (either gain or flux density) are plotted in the usual manner (refer to discussion of the contour generation algorithm).

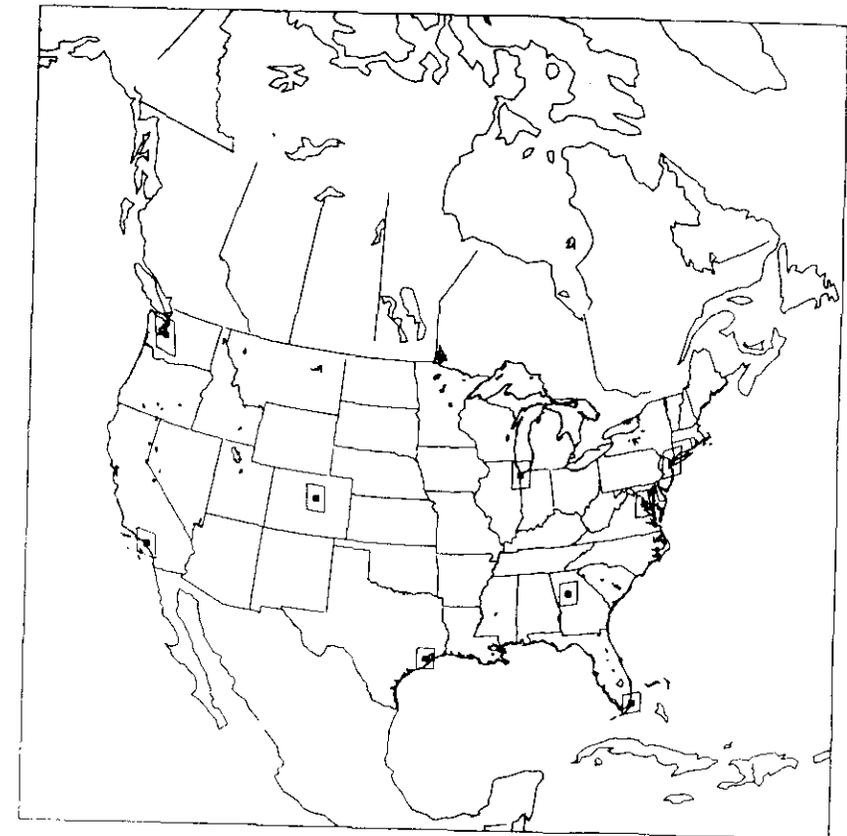


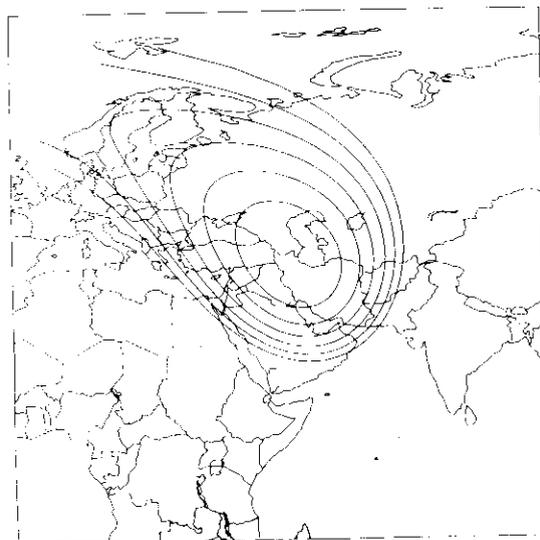
Figure 22. Pointing Error Boxes

Figure 23 is example of the degradation in a gain pattern due to pointing errors. Figure 23a shows a circular beam pattern in the absence of pointing errors, while Figure 23b shows the same pattern when satellite pointing errors in the north-south direction are $\pm 0.1^\circ$.

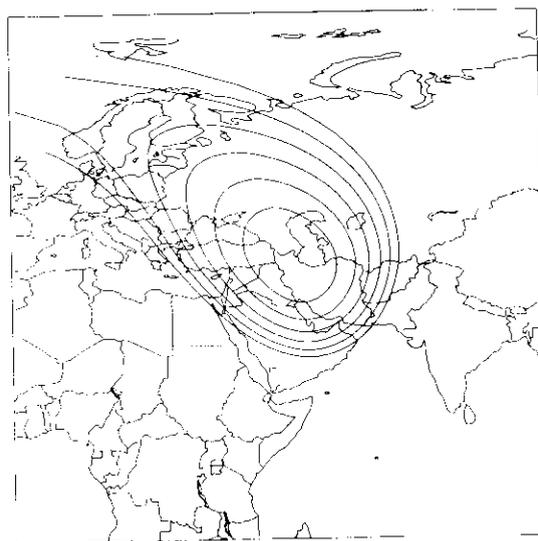
Contour generation algorithm

Antenna gain contours are generated from an n -by- n antenna pattern grid representing equally spaced points in α and β . Each grid point is assigned a gain level by one of the beam models. The gain level is the computed gain for a north-south angle, α , and an east-west angle, β (see Figure 24).

The antenna pattern grid is searched square by square from left to right and bottom to top for a specific gain contour level, (e.g., a 3-dB gain contour



(a) Without Pointing Errors



(b) With Pointing Errors

Figure 23. Example of Degradation in Gain Pattern due to Pointing Errors

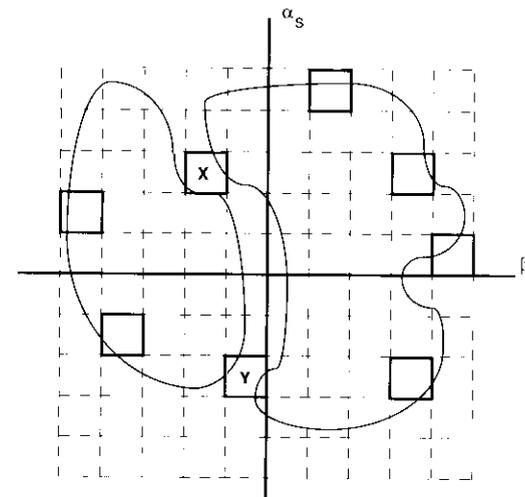


Figure 24. Gain Contours

level). Each side of a square is examined to determine if the gain contour and the side intersect. If the gain contour intersects the side, the intersection point is found in angular coordinates (α_i, β_i) . If one intersection point is found, then at least one other side of the square contains an intersection point.

A gain contour (at a particular level) and a grid square may intersect two times, four times, or not at all. If two intersection points are found, then the gain contour may intersect a square six different ways, as shown in Figure 25. The second intersection point is converted to angular coordinates, both intersection points are converted to earth-centered coordinates, and a line is drawn between the two points. This procedure is repeated for each grid square.

Four intersection points may be found on the gain contour and a grid square, as shown by the squares in Figure 26. Each intersection point is converted to angular coordinates (α_i, β_i) . Once the four intersection points are converted, it is necessary to determine how to connect the four points.

The points may be connected in two different ways to form either line segments AD and BC or line segments AB and CD . The correct choice can be determined by computing the distances of the line segments that can be drawn and summing them, as

$$D_1 = |AD| + |BC|$$

$$D_2 = |AB| + |CD|$$

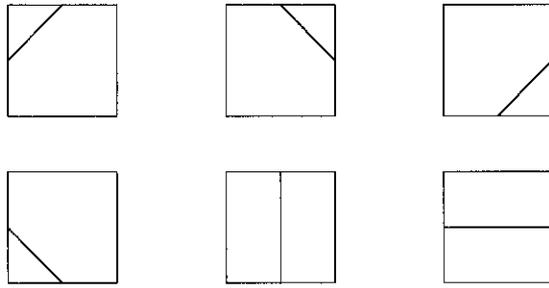


Figure 25. Six Different Intersection Pairs

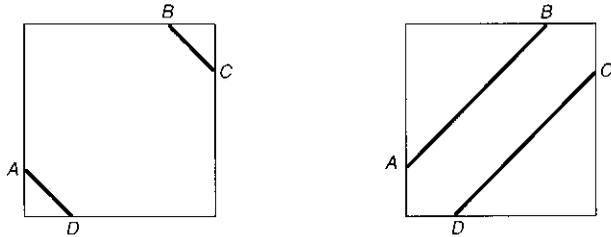


Figure 26. Four Intersection Points

The segments with the shortest combined total distance, AD and BC , are the segments that will be drawn.

After determining which way to connect the points, all four intersection points are converted to earth-centered coordinates and the two line segments are drawn. The procedure described above is repeated for each grid square. A series of connected line segments will generate the visible gain contour.

Antenna gain at a point

It is sometimes necessary to compute the antenna gain, G_p , at a point, P , which lies within an antenna pattern grid square (see Figure 27). A second-order polynomial expansion is used to approximate the antenna gain, G , over the square element, as

$$G = a_0 + a_1x + a_2y + a_3xy$$

where $a_0, a_1, a_2,$ and a_3 are constant coefficients and x and y are the east-west and north-south pointing angles relative to one corner of the square

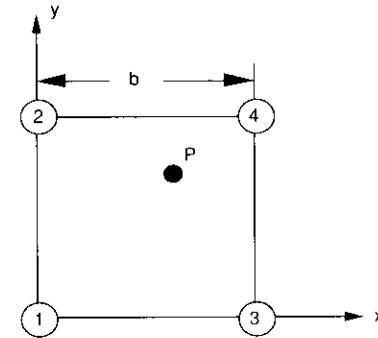


Figure 27. Gain at a Point

element. The values of the coefficients are derived from the antenna gain values at each corner of the grid square. The antenna gains at each of the four grid points are

$$\begin{aligned} G_1 &= a_0 \\ G_2 &= a_0 + a_1b \\ G_3 &= a_0 + a_2b \\ G_4 &= a_0 + (a_1 + a_2)b + a_3b^2 \end{aligned}$$

respectively, where b is the width of the square. The coefficients are then simply

$$\begin{aligned} a_0 &= G_1 \\ a_1 &= (G_2 - G_1) / b \\ a_2 &= (G_3 - G_1) / b \\ a_3 &= (G_4 - G_3 - G_2 + G_1) / b^2 \end{aligned}$$

The gain at point P , whose coordinates are (x_p, y_p) , is then given by

$$\begin{aligned} G_p &= G_1 + (1/b)(G_2 - G_1)x + (1/b)(G_3 - G_1)y \\ &\quad + (1/b^2)(G_4 - G_3 - G_2 + G_1)xy \end{aligned}$$

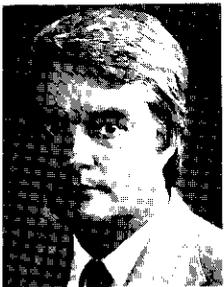
Conclusions

A computer program has been developed to generate and display satellite antenna patterns on the surface of the earth. This paper has described the techniques used by the program to generate and display geographic features, antenna gain contours, and flux density contours for various types of beams,

as well as great circle paths, earth station azimuth and elevation angles, and the effects of satellite pointing errors on the gain patterns. Plots generated by the program were used to illustrate the types of analyses that can be performed. Work is proceeding on extensions to the program to generate and display contours of constant availability for transmission in the presence of up-link rain attenuation and down-link rain degradation.

References

- [1] *Cartographic Automatic Mapping Program Documentation*, Central Intelligence Agency Document GC 77-10126, June 1977.
- [2] R. Johnson and H. Jasik, *Antenna Engineering Handbook*, New York: McGraw-Hill, 1961.



William L. Cook received a B.S. in engineering mechanics from Lehigh University in 1964, an M.S. in engineering sciences from Purdue University in 1966, and a D.Sc. in computer sciences from the George Washington University in 1973. He is currently Associate Director of the System Development Division at COMSAT Laboratories, where his responsibilities include systems analysis software development, as well as architectural design and software engineering for complex computer-based distributed systems. While at COMSAT, Dr. Cook has been responsible for the developing

analysis techniques and computer software used in the evaluation and optimization of communications satellite systems and subsystems. He was also Project Manager for experimental programs in high-speed computer communications via satellite and for the development of a monitoring and control facility for INTELSAT's TDMA network. Prior to joining COMSAT, Dr. Cook was a structural engineer at NASA's Goddard Space Flight Center and a member of the group responsible for developing the NASA Structural Analysis Program (NASTRAN).

Deborah D. Opiekun received a B.S. in computer science with a minor in mathematics from Virginia Polytechnic Institute in 1985. She is currently completing work toward an M.S. in systems engineering from the University of Maryland. She is a Senior Member of the Technical Staff in the System Development Division at COMSAT Laboratories, where her current work includes the research and development of a user interface management system and research on distributed database management systems. She has also been responsible for the design and development of the INTELSAT Satellite Services Data Base Management Facility, the Strategic Planning Data Base Management Facility, the Interactive Antenna Coverage Program, and the Antenna Coverage Program. In addition, her work has included the research and evaluation of project management and computer-aided software engineering tools.



Lewis R. Karl received a B.S. in mechanical engineering in 1985, and a B.S. in computer science in 1986, both from the University of Pittsburgh. He is currently a graduate student in computer science at the University of Maryland. As a Member of the Technical Staff in the System Development Division at COMSAT Laboratories, his work includes the development and maintenance of antenna modeling and analysis software. He is currently implementing a new version of the COMSAT General Antenna Program, which is a tool for analyzing the radiation patterns generated by reflecting antenna systems. Mr. Karl recently completed the implementation of an availability analysis relating to antenna coverages, using measured rainfall data in the Antenna Coverage Program.

Demonstration of a simplified tracking mechanism for use with inclined-orbit satellites

E. A. FAINE, D. K. GARLOW, AND R. W. GRUNER

(Manuscript received May 8, 1989)

Abstract

A simplified satellite tracking antenna system that allows small-aperture earth terminal antennas to operate with inclined-orbit satellites has been developed. The tracking ability of the system was successfully tested and demonstrated. The system operates as a single-axis, open-loop program tracker. By using a special mount design, satellite tracking is limited to one axis of movement. The actuator and controller equipment used are low-cost components that can be purchased through television receive-only (TVRO) distributors.

Introduction

Satellites in geosynchronous orbit are acted upon by many disruptive forces, including the gravitational pull of the nonspherical earth, the moon, and other celestial objects. This results in a tilted orbital plane. To maintain the satellite's position, thrusters are frequently fired to overcome these disruptive forces—a process called stationkeeping. INTELSAT's conventional stationkeeping system keeps the satellite at its nominal orbital position to within $\pm 0.01^\circ$ latitude and longitude. The amount of fuel required to maintain the satellite's position is a primary factor limiting the operational lifetime of the spacecraft.

In 1988, COMSAT Corporation developed a new stationkeeping system called "The COMSAT Maneuver"* which extends the operational lifetime of geosynchronous satellites. The COMSAT Maneuver corrects the east-west position of the satellite, but does not make corrections in the north-south direction. When the north-south stationkeeping thrusters are no longer used, the satellite takes on an inclined orbit. In inclined-orbit operation (Figure 1), the satellite appears (to an observer on the earth) to move sinusoidally orthogonal to the equatorial arc, completing a cycle in one sidereal day. The earth station antenna must track this movement.

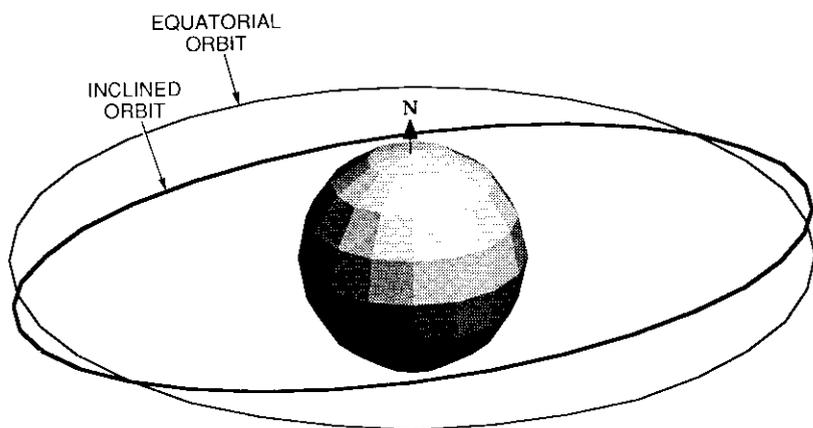


Figure 1. Inclined-Orbit Operation

Antennas with relatively expensive and complex tracking systems are widely used in larger-sized (greater than 7 m) earth stations. The larger antennas track the satellite in both azimuth and elevation and can follow its true periodic motion. The smaller, less expensive, very small aperture terminal (VSAT)-type antennas do not usually have tracking systems. A major objective of this project was to develop a simple tracking system for VSAT antennas to allow them to access inclined-orbit satellites. Such a system was developed and was first used in a test and demonstration** for the PEACESAT organization.

* L. H. Westerlund, "Method of Orienting a Synchronous Satellite," U.S. Patent No. 4,776,540, issued October 11, 1988.

** Jointly sponsored by COMSAT and INTELSAT, and conducted by COMSAT during the week of July 25, 1987, at the Paumalu earth station in Hawaii.

The tracking system was installed in 1.8- and 2.9-m-diameter C-band VSAT antennas. Figure 2 shows the geostationary arc and the path of the accessed INTELSAT IV-A F3 satellite, as observed from Paumalu. The enlarged area shows the periodic angular motion of the satellite and the 1-dB receive band beamwidths of the VSAT antennas.

This paper describes an inclined-orbit satellite tracking system for small earth station antennas, including the mount and timer control circuit. A procedure for aligning the system to track an inclined-orbit satellite is also presented. The measured results obtained during the PEACESAT test and demonstration verified the tracking ability of the system.

Tracking antenna system

The tracking antenna system shown in Figures 3 and 4 consists of an antenna reflector, tracking mount, tripod, actuator, controller, and timer control circuit. The collapsible tripod selected for the test facilitated transportability and installation. In a permanent installation, alternative rigid support structures such as a pole mount would be more likely to be used and would be less expensive.

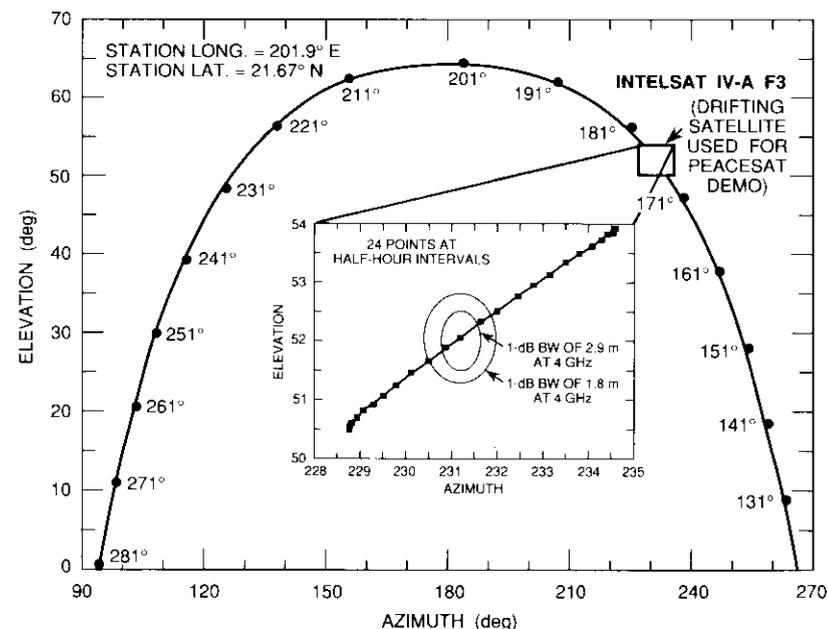


Figure 2. Geostationary Arc From Paumalu

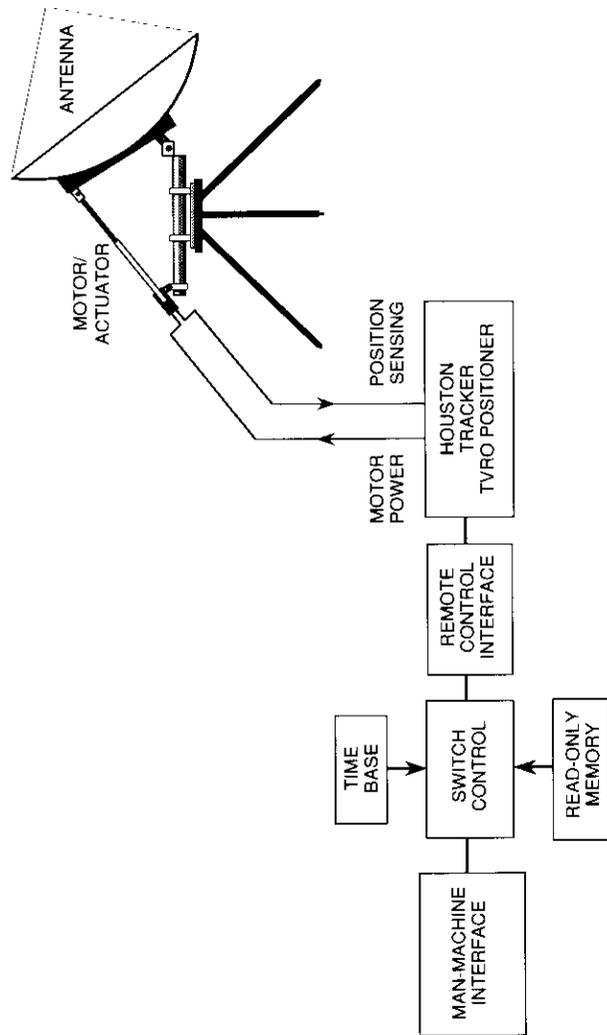


Figure 3. Antenna Tracking System Block Diagram

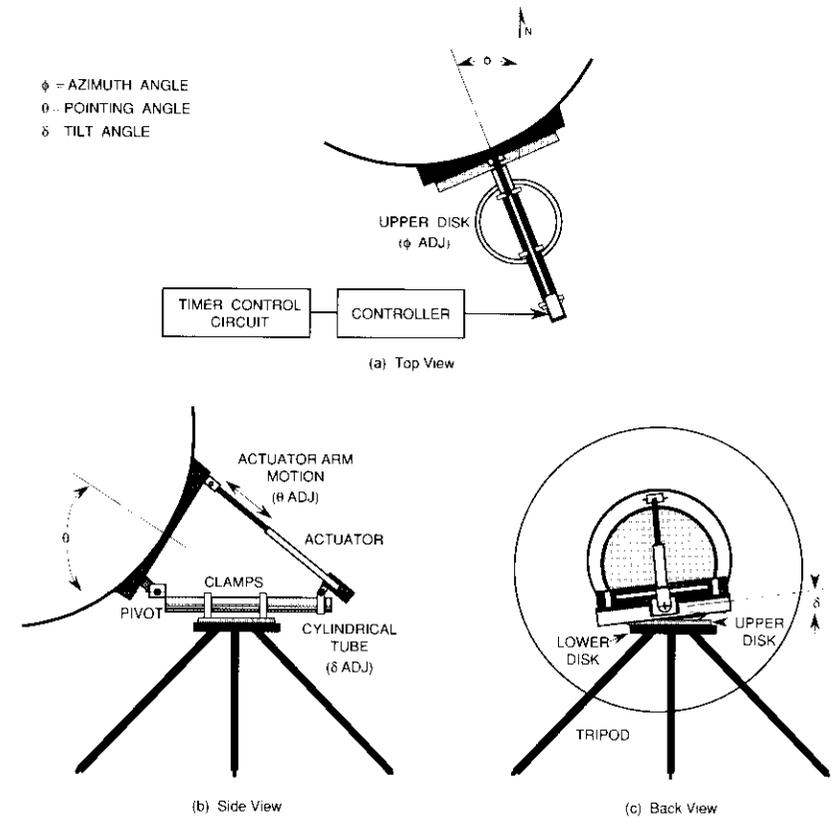


Figure 4. Functional Sketch of the Antenna Tracking System

Two aluminum disks separated by a thin plastic sheet bearing facilitate azimuthal rotation of the upper disk relative to the tripod. Once the azimuthal position is determined, the azimuth table is fastened to the tripod by a clamp system located at the outer circumference of the aluminum disks. Axis tilt is provided by a cylindrical tube mounted above the azimuth table. The tube is rotated to the desired angular position and secured by a pair of clamps. A pivot assembly attached perpendicular to the cylindrical tube serves as the rotational axis for the tracking motion.

The automatic positioning system moves the antenna reflector to track the inclined-orbit satellite. The system consists of an actuator (a motor-driven

jackscrew device), a controller, and a timer control circuit. The actuator is attached between the top of the antenna reflector and the back of the cylindrical tube. It is commanded by the controller, which receives signals from the timer control circuit.

A Houston Tracker Systems television receive-only (TVRO) antenna actuator and controller with sufficient memory to store 24 different satellite positions was used. The controller, which is normally operated from switches located on the front panel, was now commanded by the specially designed timer control circuit to switch the controller automatically at precisely controlled time intervals. The following is a description of the timer control circuit.

Time base

A crystal oscillator was built to provide precise timing control. The crystal has an oscillator frequency 1.0027 times the standard watch crystal frequency of 32,768 kHz. Timing is provided by a 14-stage oscillator/divider chip. The output signal taken off the 14th stage of the divider is a 0.5-sidereal-second square wave. This signal serves as the clock for portions of the controller, such as the analog switch driver, the setup switch network, and the word time control circuitry. Further division of the time base is accomplished using divide-by- N counters to increase the period to a sidereal half-hour.

Switch control

In the normal mode of operation for the Houston Tracker controller, the pointing of the antenna is changed to a pre-stored position by pressing a combination of two switches on the controller. The first switch on the front panel of the controller is labeled *SATELLITE TYPE*, and the second is labeled *SATELLITE NUMBER*. Valid positions consist of a combination of the two numbers pre-stored in the tracker memory. For this application, switch combinations for 24 positions were used, with *SATELLITE TYPE* being thrown 1, 2, or 3 times, and *SATELLITE NUMBER* being thrown 1 to 8 times. Each of the 24 positions is used twice during a sidereal day. The position is incremented every half hour in ascending order (as the satellite moves northward), and then in descending order (as the satellite moves southward).

Switching data are stored in an 8K x 8-bit programmable read-only memory (PROM). The 8-bit data contained in the PROM address consist of two 4-bit switch control words. The lower 4 bits determine the number of times the *SATELLITE NUMBER* switch on the tracker is to be pulsed, while the upper 4 bits determine the number of times the *SATELLITE TYPE* switch is to be pulsed. To access the PROM, a modulo-48 counter is incremented at the time of each pointing to provide the lower address bits for the data lookup.

Interfacing to the controller

The Houston Tracker controller contains a connector to allow installation of a remote control circuit board. A signal applied to a switch line throws the appropriate switch in the controller. Interfacing to this connector is accomplished through an analog switch on the control circuit board.

The man-machine interface

The timer control circuit can be set manually by using two switches on the front panel labeled *RUN/SET* and *START*. The *RUN/SET* toggle switch is used to set the starting address or initial position of the antenna. In the *SET* mode, a 1-to-48 counter increments continuously, and the trigger pulse is temporarily disconnected from the rest of the circuit. This counter displays the PROM address and the position number (1 to 48). A four-digit liquid crystal display (LCD) on the front panel displays the PROM address and the data for a visual reference (*i.e.*, the number of times each of the tracker switches are to be triggered at that particular position).

Placing the switch in the *RUN* position stops the counter and allows normal operation. Pressing the *START* switch triggers the controller and resets the half-hour timer to zero without changing to the next PROM address. Successive triggering of the controller by the timer circuitry follows the normal operating mode.

Alignment to track an inclined-orbit satellite

Aligning the single-axis tracking antenna system at the installation site requires knowledge of the azimuth, tilt, and pointing angles (ϕ , δ , and θ), as shown in Figure 4. These angles are calculated based on satellite position and inclination information determined from the INTELSAT 11-parameter ephemeris and epoch data, and on the earth station latitude and longitude.

First, the tilt angle (δ) is established. For a fixed tilt angle, movement of the actuator arm causes the antenna pointing vector to trace a path in both the azimuth and elevation coordinates. A tilt angle is chosen that best fits the pointing of the earth station antenna to the predicted satellite movement. Using these data, the pointing angle (θ) is determined for a specific time.

After the tilt and pointing angles are set, the antenna is scanned in the azimuth angle (ϕ) until the satellite signal is observed. The pointing angle and azimuth angle are then adjusted for maximum signal strength. The time of day and the pointing angle are noted. The calculated pointing angles are then adjusted to correspond with the measured pointing angle. The tracker controller can now be programmed to automatically update the antenna pointing angle each sidereal half hour.

PEACESAT test and demonstration

The PEACESAT test and demonstration of the tracking antenna system was conducted in two primary phases. The first consisted of a presentation of INTELSAT system-related information to members of the PEACESAT organization, including a demonstration of the quality of the satellite communications link. This was followed by technical verification of the performance of the inclined-orbit tracking system.

Test configuration

The two VSATs were operated in an INTELSAT Vista Star Network mode, with the Paumalu 1 Standard-A earth station serving as the hub station. All three earth stations simultaneously tracked the inclined-orbit INTELSAT IV-A F3 satellite located at 177°E. The satellite orbit inclination was $\pm 2.9^\circ$. Figure 2 shows the elevation and azimuth angular movements required of the earth station antenna.

The C-band VSATs were equipped with a single voice/data circuit capability, each using a 64-kbit/s binary phase shift keying (BPSK) modulator/demodulator (modem). The Paumalu hub station was equipped with two such modems to perform the turnaround function.

Four single-channel-per-carrier (SCPC)/BPSK carriers were transmitted to establish the communications link between the 1.8- and 2.9-m VSATs. Two carriers were required for the "go" link from the 1.8-m VSAT to the hub, and from the hub to the 2.9-m VSAT. Two additional carriers were required in order to establish the return link.

Results

The accuracy of the VSAT tracking system shown in Figure 5 was determined by monitoring the signal at the Paumalu 1 hub station. The Paumalu 1 antenna was placed in autotrack and tracked the inclined-orbit satellite during the measurement period. The strip-chart recording depicted in Figure 6 shows the up-link signal level of the 2.9-m antenna in the direction of the satellite. As the data demonstrate, the maximum signal loss due to tracking error was 1.5 dB, which corresponds to an angular pointing error of approximately 0.4° . The 2.9-m VSAT antenna transmit signal variation exhibited the expected sawtooth amplitude pattern, corresponding to the anticipated falloff in signal between satellite positioner corrections that were automatically programmed to occur every half hour. The relatively constant signal level at the beginning and end of the strip-chart recording occurs at the extremes of the figure-eight pattern, where the satellite angular motion (as observed from the earth) is at a minimum. The amount of signal loss depends on the length of the update

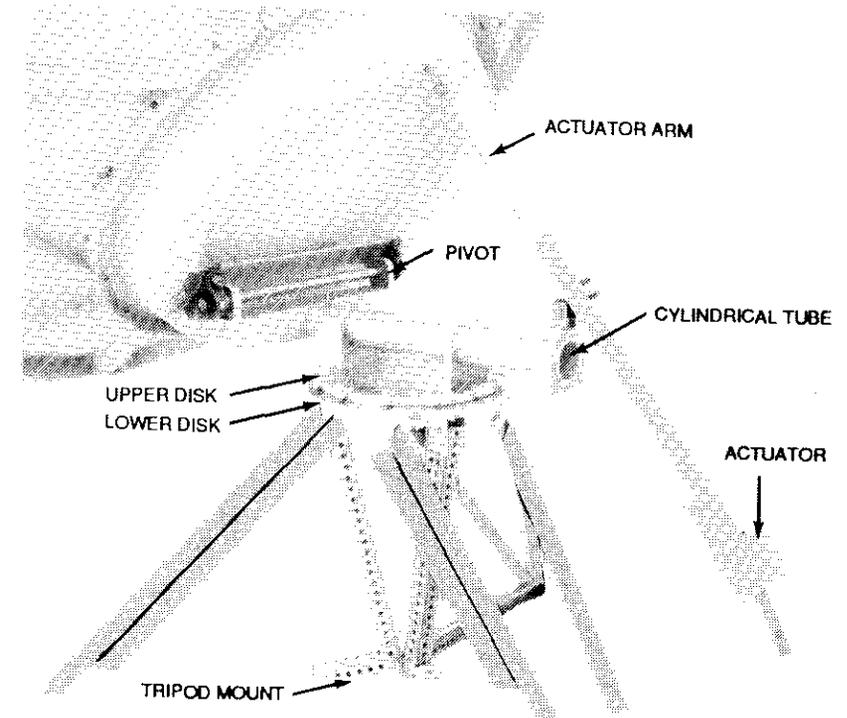


Figure 5. VSAT Tracking System

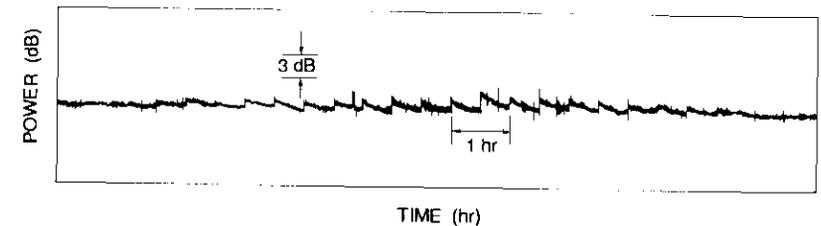


Figure 6. Signal Level of the 2.9-m VSAT

period, the half-power beamwidth of the VSAT antenna, and the inclination angle of the satellite orbit. The signal received using the 1.8-m antenna was not monitored; however, the signal variation introduced by the tracking mechanism would be less because of the antenna's broader half-power beamwidth.

Conclusions

The successful PEACESAT test and demonstration of the simplified satellite-tracking antenna mechanism indicates that the concept is suitable for use in VSAT networks that access inclined-orbit satellites. The demonstration also showed the feasibility of operating a voice and data communications network on an inclined-orbit INTELSAT satellite.

Acknowledgments

The PEACESAT test and demonstration was the result of a collaboration between INTELSAT and three divisions within COMSAT Corporation. INTELSAT provided the space segment. COMSAT World Systems Division's INTELSAT Satellite Services (U.S. Signatory to INTELSAT) served as overall project coordinator and provided the facilities at the Paumalu site, including the Paumalu Standard-A earth station. COMSAT General (COMGEN), a subsidiary of COMSAT Systems Division, provided the VSATs. COMSAT Laboratories designed, developed, and retrofitted the COMGEN VSATs with simplified tracking mechanisms, and provided technical support at the Paumalu site during the test and demonstration. G. Bish, a co-op student with the Earth Terminal Antenna Department and the University of Cincinnati designed and built the timer control circuits. The authors also wish to acknowledge significant contributions by E. Carpenter to the calculation of the satellite pointing angles and other aspects of the project.

Edward A. Faine received a B.S.E.E. from Ohio State University in 1963 and an M.S.E.E. from the State University of New York at Buffalo in 1969. He has been with COMSAT since 1973, and for the past 10 years has held several positions in the International Systems Planning Department of INTELSAT Satellite Services, World Systems Division. Mr. Faine has been a member of the Advisory Committee on Planning to the INTELSAT Board of Governors for the past 5 years.



Douglas K. Garlow joined COMSAT Laboratories as a student trainee in 1980, and is currently a Senior Technician in the Earth Terminal Antenna Department of the Microwave Laboratory. He has participated in the design and development of high-performance earth station components, formulating test procedures for these components and integrating them into operational earth terminals. In addition, he has been involved in the design and installation of numerous prototype transportable earth terminals for advanced communications satellite systems.

Robert W. Gruner received a B.S.E.E. from Northeastern University in 1962. He joined COMSAT Laboratories in 1969 and is currently Manager of the Earth Terminal Antenna Department of the Microwave Laboratory. He has participated in antenna and microwave circuit research applicable to the design of both satellites and earth stations, designing numerous microwave components such as broadband polarizers, orthogonal mode junctions, corrugated feed horns, and multiplexers. Mr. Gruner has authored more than 12 published papers on satellite earth station antenna and feed designs. He is a member of Eta Kappa Nu and Tau Beta Pi.



A sun interference prediction program

K. T. LIN AND L. J. YANG

(Manuscript received May 17, 1989)

Abstract

A computer program to estimate sun interference for a satellite-to-earth link has been developed which runs on an IBM 3083 processor under VM/CMS. A second version of the program runs on an IBM PC/AT under DOS, without graphic outputs. The program is designed as a system impairment analysis tool and is sufficiently general to be used for a geostationary satellite at any longitude, an earth station anywhere on earth, and for any given time between the years 1950 and 2050. The predictions include time of peak sun interference, sun interference duration, noise temperature increase due to solar noise during the affected periods, and resulting (degraded) antenna G/T during these periods.

Introduction

The sun is a wideband radio emitter. Its emitted radio energy is random, and is therefore in the form of noise. At microwave frequencies, noise power is usually measured in terms of its equivalent blackbody radiation (or brightness) temperature. The sun's equivalent brightness temperature varies from 12,000 to 40,000 K at 12 GHz, depending on solar activity [1]. An earth station antenna is usually directional and acts as a collector of radio waves, including both wanted signals and noise.

When the sun passes through the main beam of an earth station antenna, there is a dramatic increase in the received noise. The noise power increase during a sun transit depends on the relative position of the sun with respect to the antenna gain pattern; the size of the solar disk as observed from the

earth station; and the effective noise-temperature distribution across the solar disk. The increased noise degrades the earth station receive performance, and may cause an outage. This effect is called sun interference.

The sun may also transit the main beam of a global-coverage satellite antenna during orbital sunrise and sunset each day; however, the effect is relatively small, since only a small portion of the main beam of the antenna is filled by the sun. The computer program described here addresses sun interference on the down-link path only.

For geostationary satellites and fixed earth stations, sun interference occurs for several minutes daily during several days near the equinoxes. It occurs before the spring equinox and after the fall equinox if the earth station is located in the Northern Hemisphere, and occurs after the spring equinox and before the fall equinox if the station is in the Southern Hemisphere [2]. The degradation during these periods is measured in terms of reduction in the figure of merit of the receive system (G/T), which is defined as the ratio of the antenna gain (G) to the total effective noise temperature (T), including the system noise referred to the antenna terminal.

For a large receiving antenna, the noise temperature increase is almost equal to the average solar disk noise temperature during peak interference (assuming that the noise temperature across the solar disk is uniform), because the entire main beam is filled by the solar disk. On the other hand, for a small antenna with a half-power beamwidth larger than the solar disk, the noise temperature increase is less than that experienced by large antennas during peak interference. Therefore, the degradation in G/T is typically less for a small antenna than for a large one.

Vuong and Forsey [3] recently presented a prediction of sun transit. They noted that their predictions for high-gain antennas could be inaccurate because the emitted noise distribution over the solar disk is not uniform, as assumed in their simplified analysis. For greater accuracy, the noise temperature distribution over the solar disk should be used; however, it must be measured only a couple of days in advance of the prediction period in order to be extrapolated accurately, since it does not remain the same for longer periods of time. Because of this difficulty, the simplifying assumption of an average solar disk noise temperature is also used in the prediction given here. It should be noted that this assumption can give rise to errors approaching 3 dB.

Where sun outages cannot be tolerated, strategies can be devised to avoid them. Some satellite time-division multiple access (TDMA) system operations avoid sun outages by using two reference earth stations that have adequate separation. To analyze this problem for stations in a relatively small coverage area, precise predictions of the time and duration of the sun outages are

necessary [4]. For some systems that employ large antennas, system unavailability due to additional disruption of service by sun interference may reach 0.02 percent. However, as could be expected for systems using small antennas, a recent calculation showed that sun transit effects in many Ku-band very small aperture terminal (VSAT) systems are far milder than in conventional satellite communications systems [5]. Thus, a means of sun interference prediction is important to both system designers and system operators, who should be aware of sun outage occurrences and their system effects. However, the program presented here should be used with caution by system operators, due to the simplifying assumptions used.

Constraints on the SUNOUT program model

A computer program called SUNOUT was developed to predict the approximate time, duration, and degree of degradation of sun interference for a down-link path. One version of the program runs on an IBM 3083 processor under VM/CMS; another version runs on an IBM PC/AT under DOS. SUNOUT is intended to be used as a system impairment analysis tool, and should also have useful applications in operational environments, although as noted below it cannot be used to predict precise sun interference times.

In order to simplify the computation, the following reasonable assumptions were made:

- a. A spherical earth with earth stations at zero altitude above sea level is assumed.
- b. An approximate mathematical model is used to predict the sun's position [6]. The resulting error is within 0.01° or, for a given position, within 6 seconds.
- c. Only antennas with diameters greater than 10 wavelengths are considered.
- d. The satellite is assumed to be true geostationary, with zero declination.
- e. Variations in the earth's rotation are not considered.
- f. Parallax correction is neglected in obtaining the coordinates of the sun.
- g. Noise from the sun is assumed to be uniformly distributed over its disk.
- h. The size of the solar disk is assumed to be fixed as viewed by any earth station.

The simplification in orbital parameters arising from the assumption in item (d) may have a significant impact on the sun outage time prediction.

For a satellite orbit with an inclination of 1° , the error in the predicted time of peak interference may be 3 to 5 minutes, while for satellites with larger inclinations the error may be substantially greater. However, these errors are acceptable in a systems analysis tool. The relevant formulas used in the program will be derived in the following sections.

SUNOUT program model

Coordinate systems and astronomical triangle

Two coordinate systems were used in the computation. One is the horizon coordinate system (HCS), and the other is the equatorial coordinate system (ECS) [7]. The HCS is defined by an observer on earth with a local reference horizontal plane and a north-pointing vector. The direction to an object is given by the azimuth and elevation angles. The azimuth is the horizontal angle measured clockwise from north, and the elevation is the angle between the object and its projection onto the horizontal reference plane. Although the coordinates of a celestial object in the HCS change continuously because of the earth's rotation, this system is convenient for specifying an antenna pointing to a geostationary satellite. The observer is at the origin of the system.

For a celestial body such as the sun, it is more convenient to use the ECS. In this system the earth's equator is the reference plane, and the vernal equinox defines the reference direction. The coordinates of a body are given by the declination, which is the angle between the equatorial plane and the object, and the right ascension, which is the angle measured counter-clockwise relative to celestial north in the equatorial plane from the vernal equinox to the current projection to the object. The declination is positive if the object is north of the equator and negative if it is south. In this system, the coordinates of a distant celestial body are relatively fixed and independent of the earth's diurnal rotation. The system has an origin at earth's center.

The position of the sun at a given time is often determined by the right ascension and the declination in the ECS, as derived in the next section. For a distant object such as the sun, the transformation from the ECS to the HCS is the solution to the astronomical triangle [8], which is formed by arcs of great circles passing through the North Pole, the zenith, and the celestial body (Figure 1). The zenith is the point directly above the observer. The observer's meridian is the great circle through the North Pole and the zenith. The observer's latitude is measured between the observer's zenith and the equatorial plane on the observer's meridian. The latitude is negative if the observer's zenith is to the south of the equatorial plane; otherwise it is

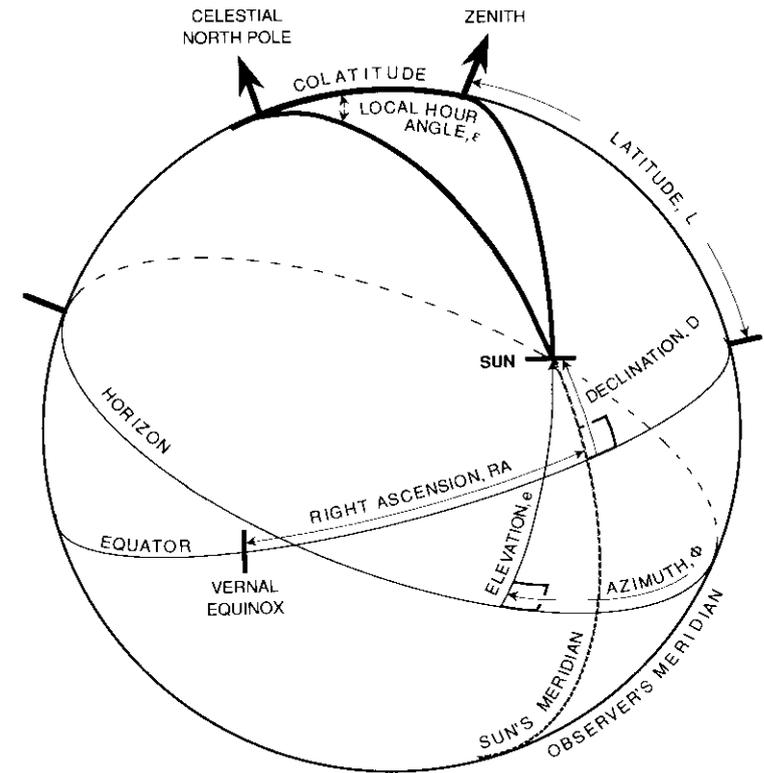


Figure 1. *The Astronomical Triangle*

positive. Colatitude is the arc segment complementary to the latitude. The horizontal plane is defined by the great circle perpendicular to the observer's zenith. Here, the parallax correction resulting from the different origins has been ignored.

The sun's meridian is the great circle through the pole and the sun. The local hour angle, which represents the hour difference from local noon, is the arc measured on the equatorial plane in the clockwise sense from the observer's meridian to the sun's meridian, expressed in hours or degrees. Before noon, the hour angle is greater than 12 hours, and after noon it is less than 12 hours. The local hour angle is also equal to the angular difference from the observer's longitude to the sun's longitude (which usually is not directly available).

From the law of cosines for sides in spherical trigonometry, the following basic formula is obtained for a spherical triangle with sides a , b , c and angle α opposite side a :

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha \quad (1)$$

where each side is specified in terms of the corresponding angle subtended at the center of the sphere.

The astronomical triangle may be related to this spherical triangle by letting α be the local hour angle (ϵ), a be the angle complementary to the elevation (e), b be the colatitude of the observer (latitude, L), and c be the angle complementary to the sun's declination (D). Substitution into equation (1) yields

$$\cos(90^\circ - e) = \cos(90^\circ - L) \cos(90^\circ - D) + \sin(90^\circ - L) \sin(90^\circ - D) \cos \epsilon \quad (2)$$

which reduces to

$$\sin e = \sin L \sin D + \cos L \cos D \cos \epsilon \quad (3)$$

and then to

$$e = \sin^{-1}(\sin L \sin D + \cos L \cos D \cos \epsilon) \quad (4)$$

Equation (4) expresses the elevation angle of the sun (referenced to the earth's center), as seen by an observer, in terms of the observer's latitude, local hour angle, and declination of the sun.

Now the spherical triangle may be rearranged by letting $\alpha = 360^\circ - \Phi$ (azimuth); $a = 90^\circ - D$; $b = 90^\circ - L$; and $c = 90^\circ - e$. Substituting into equation (1) gives

$$\cos(90^\circ - D) = \cos(90^\circ - L) \cos(90^\circ - e) + \sin(90^\circ - L) \sin(90^\circ - e) \cos(360^\circ - \Phi) \quad (5)$$

which reduces to

$$\cos \Phi = \sin D / (\cos L \cos e) - \tan L \tan e \quad (6)$$

and then to

$$\Phi = \cos^{-1}[\sin D / (\cos L \cos e) - \tan L \tan e] \quad (7)$$

Equation (7) expresses the azimuth of the sun, as seen by the observer, in terms of the observer's latitude, elevation, and declination of the sun.

Equations (4) and (7), which neglect parallax correction, are used to obtain the coordinates of the sun in the HCS based on those in the ECS.

Calculating time and the coordinates of the sun

The right ascension and the declination of the sun, as well as the earth's rotational orientation relative to vernal equinox, may be related to the universal time (UT) [6]. Time is measured according to the period of the earth's rotation. Usually, the unit of time is the solar day, which is the period between successive sun passages through the observer's meridian. However, since the earth also circles the sun with a period of 1 year, the solar day is not the true period of the earth's rotation. The true period is called a sidereal day and is the period between passages of the vernal equinox through the observer's meridian.

There are also variations (nutations and precessions) in the earth's rotation, and thus in the length of a sidereal day. The mean sidereal day is determined by the mean earth rotation. The difference between mean sidereal time and apparent sidereal time is called the equation of equinox. The apparent sidereal time also depends on the longitude of the observer. At the observer's location it is called the local apparent sidereal time (LAST), and at Greenwich it is called the Greenwich apparent sidereal time (GAST). Therefore, the true angle from the vernal equinox is implied in the apparent sidereal time, but universal time is referenced to the mean solar time at Greenwich. By neglecting the correction for the equation of equinox, the LAST and GAST can be obtained from the universal time [9] by

$$\text{LAST} = \text{GAST} \pm l_o \begin{cases} + & \text{for east} \\ - & \text{for west} \end{cases} \quad (8a)$$

$$\begin{aligned} \text{GAST (at } t^h \text{ UT)} &= 24,110.54841^s + 8,640,184.812866^s \text{ JC} \\ &+ 0.093104^s \text{ JC}^2 - 6.2 \times 10^{-6s} \text{ JC}^3 \\ &+ 3,600^s t \end{aligned} \quad (8b)$$

where l_o is the longitude of the observer; JC = (JD - 2,451,545.0)/36,525 (the interval of time measured in Julian centuries of 36,525 days of UT, elapsed since the epoch 2000 January 1, 12^h UT); and JD is the Julian date, which is the time interval measured in mean solar days elapsed since January 1, 4713 BC, 12^h UT. (For January 1, 2000, 12^h UT, JD = 2,451,545.0, i.e., the epoch J2000.0.) Multiples of 24^h must be added or subtracted as necessary to put LAST and GAST in the range of 0^h to 24^h.

By neglecting the ecliptic latitude of the sun, the right ascension and the declination of sun can be derived from the obliquity of the ecliptic (Ω) and the ecliptic longitude (τ) shown in Figure 2, as

$$\tau = l_s + 1.915^\circ \sin g + 0.020^\circ \sin 2g \quad (9)$$

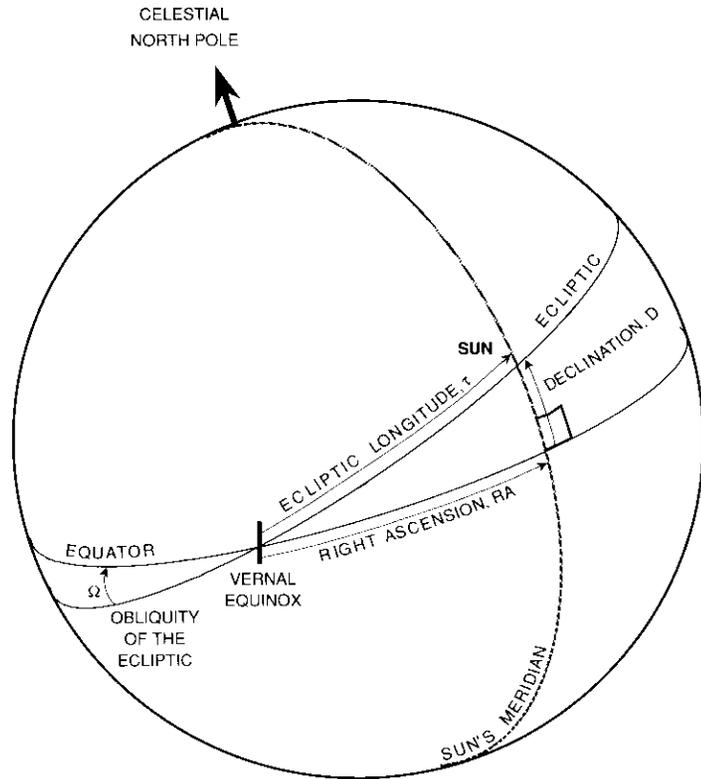


Figure 2. The Ecliptic Great Circle

where $l_s = 280.460^\circ + 0.9856474^\circ n$ (mean longitude of the sun); $g = 357.528^\circ + 0.9856003^\circ n$ (mean anomaly); $n = \text{JD} - 2,451,545.0$ (number of solar days to the epoch J2000.0), and

$$\Omega = 23.439^\circ - 0.0000004^\circ n \quad (10)$$

From the law of sines in spherical trigonometry we obtain

$$\sin \tau / \sin 90^\circ = \sin D / \sin \Omega \quad (11)$$

which reduces to

$$D = \sin^{-1}(\sin \Omega \sin \tau) \quad (12)$$

Now let $a = \tau$; $b = \text{RA}$ (right ascension of sun); $c = D$; $\alpha = 90^\circ$; and substitution into equation (1) yields

$$\cos \tau = \cos \text{RA} \cos D \quad (13)$$

Then let $a = D$; $b = \tau$; $c = \text{RA}$; $\alpha = \Omega$; and substitution into equation (1) gives

$$\cos D = \cos \tau \cos \text{RA} + \sin \tau \sin \text{RA} \cos \Omega \quad (14)$$

Combining equations (13) and (14), and eliminating $\cos D$, we obtain

$$\cos \tau \sin \text{RA} = \sin \tau \cos \text{RA} \cos \Omega \quad (15)$$

which further reduces to

$$\text{RA} = \tan^{-1}(\cos \Omega \tan \tau) \quad (16)$$

where the angle RA of equation (16) is selected to be within the same quadrant as τ . The local hour angle is the time (or angular) difference between the LAST and the right ascension of the sun, as

$$\epsilon = \text{LAST} - \text{RA} \quad (17)$$

Equations (12) and (16) can be used to obtain the coordinates of the sun in the ECS to a precision of 0.01° for any given time between the years 1950 and 2050.

Calculating the look angles of a satellite as seen from an earth station

In practice, a geostationary satellite is not completely stationary as viewed from the earth, but appears to drift with a diurnal pattern. Since the drift is usually very small, the diurnal motion of the satellite is ignored here. The satellite position is given by the geographic longitude, l_{gsa} , on the geostationary orbit, which is approximately 35,900 km (22,300 mi) above the earth's surface [10]. Let point A in Figure 3 be the subsatellite point at the earth's surface.

Let L and l_o be the latitude and longitude, respectively, of the earth station at point B in Figure 3. Also, let R_o be the radius of the earth and R be the distance from the satellite to the earth center, O . By assuming that the earth is a sphere and applying spherical trigonometry to Figure 3, we obtain

$$\cos \delta = \cos L \cos (l_o - l_{gsa}) \quad (18)$$

$$\cos (l_o - l_{gsa}) = \cos L \cos \delta + \sin |L| \sin \delta \cos \beta \quad (19)$$

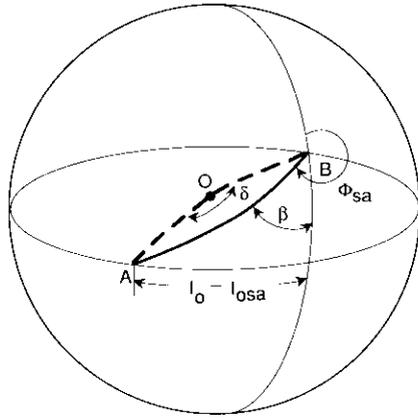


Figure 3. Geometry of the Earth Station on the Earth

where δ is the arc distance between points A and B on the sphere; β is the angle as shown in Figure 3; and l_{osa} is the longitude of the satellite. Equation (19) further reduces to

$$\cos \beta = \tan |L| \operatorname{ctn} \delta \quad (20)$$

Then the azimuth of the satellite as seen from a station located in the Northern Hemisphere is

$$\Phi_{sa} = \begin{cases} \pi + \beta & \text{satellite west of the station} \\ \pi - \beta & \text{satellite east of the station} \end{cases} \quad (21)$$

For a station in the Southern Hemisphere, the azimuth is

$$\Phi_{sa} = \begin{cases} 2\pi - \beta & \text{satellite west of the station} \\ \beta & \text{satellite east of the station} \end{cases} \quad (22)$$

Applying the law of cosines to the triangle formed by the satellite, the earth station, and the earth center, O (Figure 4), we obtain

$$d^2 = R_0^2 + R^2 - 2R_0R \cos \delta \quad (23)$$

where d is the distance from the earth station to the satellite. The figure shows a right triangle, SBC, and its relationship to the local horizontal plane at the earth station. The elevation angle is the angle between the horizontal plane and the direction to the satellite. From Figure 4, we obtain

$$R \sin \delta = d \sin (90^\circ - e_{sa}) \quad (24)$$

Solving for the elevation angle, e_{sa} , gives us

$$e_{sa} = \cos^{-1} (R \sin \delta / d) \quad (25)$$

Equations (20), (21), (22), and (25) are used to obtain the look angles of a satellite in the HCS.

Determining dates and time of day of sun transit

It is assumed that the direction of antenna pointing is known, and that the time is sought when the sun passes nearest to this particular direction. It is difficult to deduce the time from equations (4) and (7) by requiring that $e = e_{sa}$ and $\Phi = \Phi_{sa}$. Instead, a simple formula that approximates the declination of the sun, D' , in terms of days of the year is used:

$$D' = 23.5^\circ \sin [2\pi (p - x)/365] \quad (26)$$

where 23.5° is the approximate obliquity of the ecliptic, p is the day of the year, and the value of x is 80 for days near the vernal equinox and 83.5 for days near the autumnal equinox. This formula is similar to that used in Reference 2. Two different values are used for x because the number of days

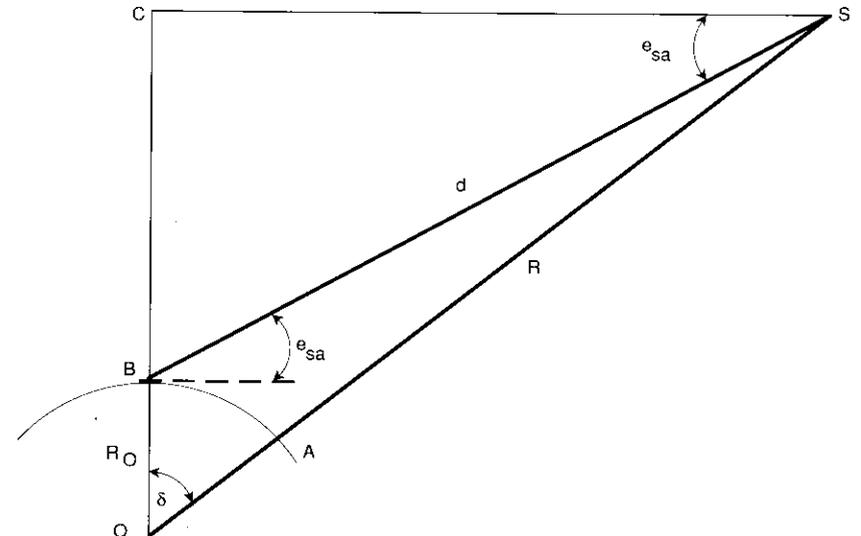


Figure 4. Planar View of the Earth Station-to-Satellite System

from vernal equinox to autumnal equinox is not equal to the number of days from autumnal equinox to vernal equinox.

At sun transit, the sun's declination, D , to the equatorial plane is very close to the equivalent declination, D_{sa} , of the satellite look angles in the ECS, which is calculated by replacing e with e_{sa} and Φ with Φ_{sa} in equation (5) to obtain

$$D_{sa} = \sin^{-1}(\sin L \sin e_{sa} + \cos L \cos e_{sa} \cos \Phi_{sa}) \quad (27)$$

The approximate date of peak sun transit can be solved by equating equations (26) and (27) to obtain the nearest integer of p , day P_0 .

The number of affected days can be approximated by [5]

$$\text{Affected days} = \frac{\text{Half-power beamwidth} + 0.48^\circ}{0.4^\circ/\text{day}} \quad (28)$$

where 0.48° is the approximate optical diameter of the sun. The half-power beamwidth (in degrees) may be estimated as

$$\text{Half-power beamwidth} = 70 \frac{\lambda}{d_{ant}} \quad (29)$$

where d_{ant} is the antenna diameter and λ is the operational wavelength, in the same units.

The approximate time of the transit peak on day P_0 can be derived from the following relationship:

$$\sin e_{sa} = \sin L \sin D_{sa} + \cos L \cos D_{sa} \cos \epsilon_{sa} \quad (30)$$

which reduces to

$$\epsilon_{sa} = \cos^{-1}(\sin e_{sa}/\cos L \cos D_{sa} - \tan L \tan D_{sa}) \quad (31)$$

where ϵ_{sa} is the equivalent local hour angle of the satellite look angles in the ECS. Equation (30) is obtained by replacing e with e_{sa} , D with D_{sa} , and ϵ with ϵ_{sa} in equation (3). A condition for the sun transit time is that both the sun and the satellite have the same local hour angle. As noted above, the local hour angle is the time difference from local noon. The local time, LT_0 , is

$$LT_0 = \begin{cases} 12 - \epsilon_{sa} & \text{for antenna pointing east} \\ 12 + \epsilon_{sa} & \text{for antenna pointing west} \end{cases} \quad (32)$$

Then the corresponding universal time UT_0 is

$$UT_0 = LT_0 \pm l_o \begin{cases} + & \text{for west} \\ - & \text{for east} \end{cases} \quad (33)$$

Substituting day P_0 and time UT_0 into equations (8), (9), (10), (12), and (16) gives the more accurate declination (D_0) and right ascension (RA_0) of the sun, and the LAST for that time, $LAST_0$. Since UT_0 and P_0 are derived from an approximate formula, D_0 and RA_0 will not be the declination and right ascension at peak sun interference, but will be very close to them. A more accurate date of peak sun interference is derived by substituting day $P = P_0 \pm I$ and time UT_0 into equations (9) and (12) to find a new day P with $|D - D_{sa}| \leq 0.2^\circ$, where I is an integer.

The equivalent declination, D_{sa} , of the satellite look angles does not vary with time, but the equivalent right ascension, RA_{sa} , is

$$RA_{sa} = LAST - \epsilon_{sa} \quad (34)$$

which varies as time progresses (at a rate of about 2.5° every 10 minutes). For a short period of time, the sun's declination and right ascension will not vary significantly (less than 0.003° and 0.007° , respectively, every 10 minutes).

Another criterion for the sun transit time is the coincidence of the sun's and the satellite's right ascensions (RA , RA_{sa}). Therefore, for a time (UT_0) close to the sun transit, the condition is

$$RA_{sa} = RA_0 \quad (35)$$

From equations (34) and (35), a new LAST is obtained, as

$$LAST_1 = RA_0 + \epsilon_{sa} \quad (36)$$

The corresponding universal time, UT_1 , at peak sun interference is derived as

$$UT_1 = UT_0 + (LAST_1 - LAST_0) \times 0.99726957 \quad (37)$$

For days immediately before or after day P , UT_1 for day P may be used as the input UT_0 to obtain the time of peak sun interference (UT_1) on those days, with the above algorithm.

The number of affected minutes for the peak day can be approximated by [5]

$$\text{Affected minutes} = \frac{\text{Half-power beamwidth} + 0.48^\circ}{0.25^\circ/\text{min}} \quad (38)$$

Determining offset angles

Once the periods of sun transit have been determined, the sun's coordinates (e and Φ) can be calculated from equations (4) and (7). The offset angle, θ_o , between the sun and the satellite can be computed in the HCS from

$$\theta_o = \cos^{-1} [\cos e \cos e_{sa} \cos (\Phi - \Phi_{sa}) + \sin e \sin e_{sa}] \quad (39)$$

Noise temperature of the sun

Assuming that the sun's temperature is constant over the visible disk, the solar flux density is equal to that of a blackbody radiator at 6,000 K for frequencies greater than 30 GHz, but exceeds this flux value at lower frequencies [7]. Furthermore, for the lower frequencies, the noise temperature also depends on the level of solar activity. For frequencies, f , between 1 and 20 GHz, the following relation may be used to estimate the noise temperature of the quiet sun [3]:

$$T_s = 120,000 f^{-0.75} \text{ (K)} \quad (40)$$

The radiated noise power is attenuated as it passes through the atmosphere because of gaseous absorption, and this attenuation increases with decreasing elevation angle. For the standard atmosphere, the one-way zenith path attenuation is about 0.036 dB at C-band and about 0.072 dB at Ku-band. For elevation angles greater than 5°, the slant path attenuation is given approximately by [11]

$$\text{atten} = \begin{cases} 0.036/\sin e & \text{dB at C-band} \\ 0.072/\sin e & \text{dB at Ku-band} \end{cases} \quad (41)$$

The solar noise temperature observed at an earth station is then approximated by

$$T'_s = T_s 10^{-\text{atten}/10} \quad (42)$$

extended over the sun's visible disk.

Calculating noise temperature and figure-of-merit degradation due to sun interference

The effective rise in noise temperature due to solar transit is calculated by averaging of the radiation pattern of the antenna over the sun's disk, weighted by the sun's temperature [3], as

$$\Delta T_a = \frac{1}{4\pi} \iint_{\text{sun's disk}} T'_s G(\theta, \phi) \sin\theta \, d\theta d\phi \quad (43)$$

The polarization factor is included implicitly [7], [12]. Here, it is assumed that the sun is a disk of optical size 0.48° in diameter and of constant temperature, T'_s . The radiation pattern, $G(\theta, \phi)$, of an antenna is derived from a simple form of the aperture illumination function [13] for circular symmetry

$$\left[1 - \left(\frac{2r}{d_{ant}} \right)^2 \right]^P \quad (44)$$

where r is the distance from the center of the antenna on the aperture. The pattern used is

$$G(\theta, \phi) = \begin{cases} 2P + 1 [2^P \Gamma(P + 1)] \left(\frac{2\pi d_{ant}}{\lambda} \right)^{-2} \left[\frac{J_{P-1} \left(\frac{k d_{ant}}{2} \sin\theta \right)}{\left(\frac{k d_{ant}}{2} \sin\theta \right)^{P-1}} \right]^2 \cos\theta & ; \text{for } 0 \leq \theta \leq \pi/2 \\ 0 & ; \text{otherwise} \end{cases} \quad (45)$$

where P is 1, 1.5, and 2 for aperture efficiencies of 75.0, 64.0, and 55.55 percent, respectively [14].

Equation (43) is integrated numerically in the θ direction. Since the integrand does not change in the ϕ direction, the result of the ϕ integration is simply the extent of the sun's disk, $2\phi_i$, in the ϕ direction (Figure 5), as

$$\Delta T_a = \frac{1}{4\pi} T'_s \int_{\substack{\theta = \theta_o, \alpha_s, \text{ or } 0 \\ (\text{whichever is larger})}}^{\theta_o + \alpha_s} G(\theta, \phi) \sin\theta (2\phi_i) \, d\theta \quad (46)$$

where

$$\phi_i = \begin{cases} \pi & ; \text{for } \theta_o - \alpha_s \leq 0 \text{ and } \theta \leq \alpha_s - \theta_o \\ \frac{\cos^{-1}(\cos\alpha_s - \cos\theta \cos\theta_o)}{\sin\theta \sin\theta_o} & ; \text{otherwise} \end{cases}$$

In the program, the Bessel functions, $J_{P+1}(\cdot)$, are computed using polynomial approximations [14].

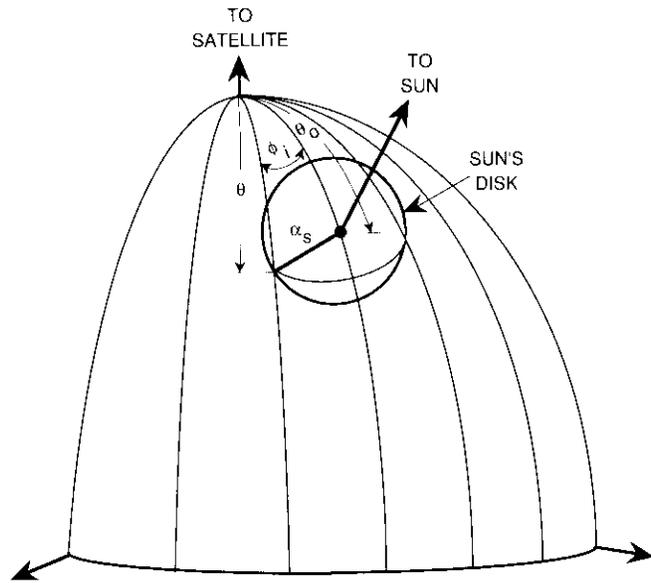


Figure 5. Offset Angle θ_0 Between the Sun and the Satellite

The receiving system itself also generates noise due to feed loss (L_f), the ambient temperature of the antenna (T_a), antenna noise from the sky (T_a), and receiver noise (T_r). The total noise temperature, T , during sun transit is

$$T = T_a + \Delta T_a + T_a(L_f - 1) + T_r L_f \quad (47)$$

The degraded G/T can then be expressed as

$$G/T = 10 \log_{10} [G(\theta = 0, \phi)/T] \quad (\text{dB/K}) \quad (48)$$

Verifying results

The predicted locations of the sun in terms of elevation and azimuth angles as functions of time have also been compared with results computed using another program [8]. The error in both elevation and azimuth is less than 0.1° . Since no measured data are available, no quantitative verification of the predicted antenna noise temperature increases and degradation in G/T can be made. The estimated accuracy in the date of peak interference is about 2 days, and in the time of peak interference is about 5 minutes.

Practical example

The program input data include the following:

- Year considered
- Satellite longitude
- Down-link carrier frequency (GHz)
- Earth station longitude and latitude
- Antenna diameter (m)
- Antenna noise temperature (K)
- Receiver noise temperature (K)
- Ambient temperature (K)
- Feed loss (dB)
- Efficiency of the antenna.

As an example, the program was executed for an earth station located at Taipei, Taiwan, pointing to a satellite at 180°E , with an antenna diameter of 1.8 m, antenna efficiency of 64.0 percent, antenna noise temperature of 58 K, receiver noise temperature of 230 K, feed loss of 0.25 dB, and receive frequency of 11.15 GHz. The program predicted that this station would experience peak sun interference at 23 hr, 41 min, 25 s on March 11, 1989, and also at 23 hr, 20 min, 34 s on October 3, 1989. Again, the predicted time of peak interference is expected to be accurate to within about ± 5 minutes, due to the simplified assumptions. For both events, a system degradation of about 9 dB in G/T at peak interference was estimated.

A sample of the output of the program for this case is shown in Figure 6. Graphical output (Figure 7) is available only from the IBM 3083 version, although the PC version could be extended to include this capability. The increase in antenna noise temperature, total system noise temperature, and G/T 's of the receive earth station during the sun transit periods are tabulated. The time increment in the computation is 1 minute for each affected day, covering 10 minutes before and after the peak interference for that day. The peak interference time of each affected day gradually changes during the sun transit periods; however, these small differences are ignored in the graphical outputs.

Conclusions

A method for estimating sun interference effects at an earth station has been developed and programmed. Noise variations due to solar disturbances, antenna pointing errors, small perturbations of the earth's rotation, and station

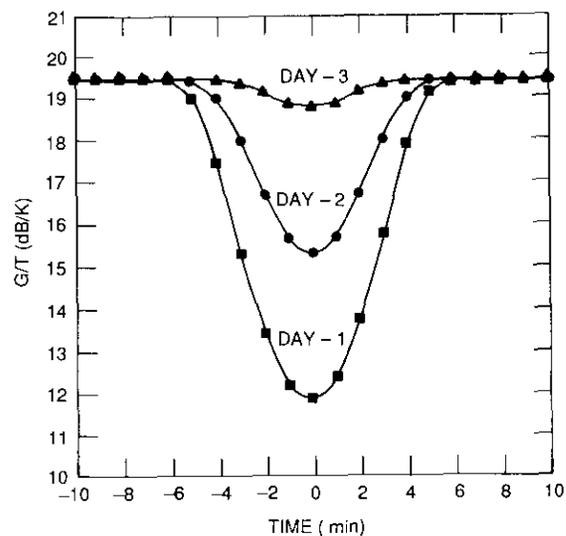
ANTENNA DIAMETER = 1.80 METER
 ANTENNA ESTIMATE GAIN = 44.51 DB
 ANTENNA EFFICIENCY = 64.00 %
 RECEIVED FREQUENCY = 11.15 GHZ
 ANTENNA NOISE TEMPERATURE : 58.0 DEGREE K
 FEED LOSS : 0.25 DB
 NOISE TEMPERATURE OF LOW NOISE AMPLIFIER : 230.0 DEGREE K
 AMBIENT TEMPERATURE : 290.0 DEGREE K
 SUN TEMPERATURE AT FREQUENCY 11.15 GHZ IS 19666.4 DEGREE K
 WAVELENGTH OF RECEIVE CARRIER : 2.69 CM
 THE LATITUDE OF THE EARTH STATION :
 N/S (+/-) (DD, MM, SS) = 25 3 0.00 = 25.05 DEGREE
 THE LONGITUDE OF THE EARTH STATION :
 E/W (+/-) (DD, MM, SS) = 121 18 0.00 = 121.30 DEGREE
 THE LONGITUDE OF THE SATELLITE :
 E/W (+/-) (DD, MM, SS) = 180 0 0.00 = 180.00 DEGREE
 THE EARTH STATION LOOK SATELLITE ANGLE :
 AZ = 104.44 DEGREE
 EL = 19.91 DEGREE
 VERNAL PEAK WILL OCCUR AT : MAR 11, 1989
 AUTUMNAL PEAK WILL OCCUR AT ; OCT 3, 1989
 PEAK DURATION IN MINUTES = +/- 4 MINUTES
 PEAK AFFECTED DAYS = +/- 2 DAYS
 ANTENNA BEAMWIDTH IS 1.05 DEGREE

Figure 6. Output File Produced by the SUNOUT Program

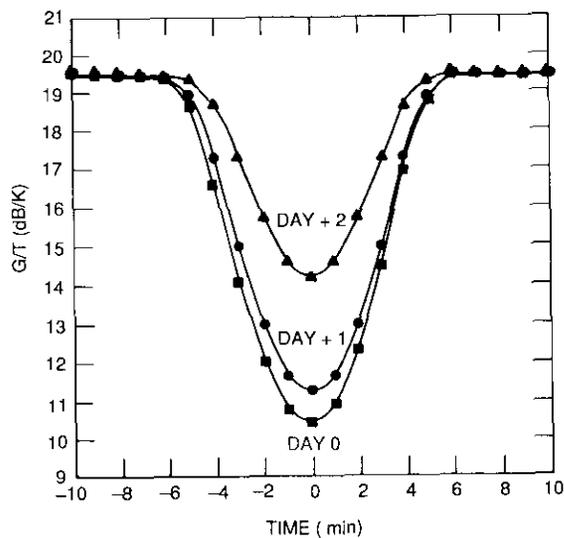
```

***** VERNAL PEAK *****
PEAK DAY + 0
MAR 11, 1989
PEAK SUN INTERFERENCE AT GMT (H:M:S) = 23:41:25
    GMT (MIN:SEC) = 31:25 32:25 33:25 34:25 35:25 36:25 37:25 38:25 39:25 40:25 41:25
    SUN TEMP (DEGREE K) = 0.1 0.9 2.9 2.7 4.3 52.8 249.5 689.7 1332.6 1948.6 2220.4
    RX SYS. TEMP (DEGREE K) = 319.0 319.7 321.7 321.5 323.1 371.6 568.3 1008.5 1651.4 2267.4 2539.2
    G/T (DB/DEGREE K) = 19.5 19.5 19.4 19.4 19.4 18.8 17.0 14.5 12.3 11.0 10.5
    GMT (MIN:SEC) = 41:25 42:25 43:25 44:25 45:25 46:25 47:25 48:25 49:25 50:25 51:25
    SUN TEMP (DEGREE K) = 2220.4 1997.1 1405.7 749.0 281.8 63.5 5.6 2.5 3.1 1.0 0.1
    RX SYS. TEMP (DEGREE K) = 2539.2 2315.9 1724.5 1067.8 600.6 382.4 324.4 321.3 321.9 319.8 319.0
    G/T (DB/DEGREE K) = 10.5 10.9 12.1 14.2 16.7 18.7 19.4 19.4 19.4 19.5 19.5
PEAK DAY + 1
MAR 12, 1989
PEAK SUN INTERFERENCE AT GMT (H:M:S) = 23:41:18
    GMT (MIN:SEC) = 31:18 32:18 33:18 34:18 35:18 36:18 37:18 38:18 39:18 40:18 41:18
    SUN TEMP (DEGREE K) = 0.2 0.6 2.6 3.1 2.3 27.0 144.2 419.7 828.5 1215.3 1372.7
    RX SYS. TEMP (DEGREE K) = 319.0 319.5 321.4 321.9 321.1 345.8 463.0 738.5 1147.3 1534.1 1691.5
    G/T (DB/DEGREE K) = 19.5 19.5 19.4 19.4 19.4 19.1 17.9 15.8 13.9 12.7 12.2
    GMT (MIN:SEC) = 41:18 42:18 43:18 44:18 45:18 46:18 47:18 48:18 49:18 50:18 51:18
    SUN TEMP (DEGREE K) = 1372.7 1208.5 819.4 411.2 139.2 25.3 2.2 3.2 2.5 0.6 0.2
    RX SYS. TEMP (DEGREE K) = 1691.5 1527.3 1138.2 730.0 458.0 344.1 321.0 322.0 321.3 319.4 319.0
    G/T (DB/DEGREE K) = 12.2 12.7 14.0 15.9 17.9 19.1 19.4 19.4 19.4 19.5 19.5
    (etc.)
    
```

Figure 6. Output File Produced by the SUNOUT Program (Cont'd)



(a) Before



(b) During and After

Figure 7. *G/T* Decrease Before, During, and After the Autumnal Peak Sun Transit Day

altitude are not included in the program. A spherical earth model and a mathematically workable antenna pattern are assumed. The satellite is also assumed to be perfectly stationary on the geosynchronous orbit. Calculations showing the quantitative effects of sun transit on the earth station *G/T* were presented. The program is designed mainly as an analysis tool for system impairments.

Acknowledgments

The authors wish to express their appreciation to W. Sandrin for his support, to D. Rogers for his encouragement, to D. Weinreich for his suggestions, and to V. Bagrodia, who documented the program and ported it to the IBM PC/AT.

References

- [1] F. I. Shimabukuro and J. M. Tracey, "Brightness Temperature of the Quiet Sun at Centimeter and Millimeter Wavelengths," *The Astrophysical Journal*, Vol. 152, 1968, pp. 777-782.
- [2] *CCIR Handbook on Satellite Communications (Fixed-Satellite Service)*, "Remark on Solar Interference," 1985, pp. 211-212.
- [3] X. T. Vuong and R. J. Forsey, "Prediction of Sun Transit Outages in an Operational Communication Satellite System," *IEEE Transactions on Broadcasting*, Vol. BC-29, No. 4, 1983, pp. 121-126.
- [4] E. J. Durwen, "Determination of Sun Interference Periods for Geostationary Satellite Communication Links," *Space Communication and Broadcasting*, Vol. 5, 1987, pp. 183-195.
- [5] F. Mohamadi and D. L. Lyon, "Effects of Solar Transit in Ku-Band VSAT Systems," *IEEE Transactions on Communications*, Vol. 36, No. 7, July 1988, pp. 892-894.
- [6] *The Astronomical Almanac for the Year 1984*, Nautical Almanac Office, U.S. Naval Observatory, Washington, D.C., p. C24.
- [7] J. D. Kraus, *Radio Astronomy*, New York: McGraw-Hill, 1966, pp. 34-42 and 97-102.
- [8] I. I. Meuller, *Spherical and Practical Astronomy as Applied to Geodesy*, New York: Frederick Ungar, 1969, pp. 37-51.
- [9] *The Astronomical Almanac for the Year 1984*, Nautical Almanac Office, U.S. Naval Observatory, Washington, D.C., p. B6.
- [10] W. L. Morgan and G. D. Gordon, *Communications Satellite Handbook*, New York: John Wiley, 1989, Ch. 6, pp. 774-870.
- [11] K. G. Johannsen and L. Titus, "Determination of Ku-Band Earth Station Antenna *G/T* Using the Moon as RF Source," *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-35, No. 3, 1986, p. 344-348.

- [12] J. H. Rainwater, "Radiometers: Electronic Eyes that 'See' Noise," *Microwaves*, Vol. 17, No. 9, September 1978, pp. 58-62.
- [13] A. F. Sciambi, "The Effect of the Aperture Illumination on the Circular Aperture Antenna Pattern Characteristics," *Microwave Journal*, Vol. 8, No. 8, August 1965, pp. 79-84.
- [14] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphics, and Mathematical Tables*, New York: Dover, 1972, pp. 369-370.



Kuan-Ting Lin received a Ph.D. in electrical engineering from the Ohio State University in 1986. He joined COMSAT Laboratories in 1986 as a Member of the Technical Staff in the Microwave Technology Division, and is currently in the Satellite Technologies Division. His interests include low-cost radiometer design, low-cost beacon receiver design, Ku-band up-link power control techniques, and a clear-air fade study for low elevation angle links. Recently he was involved in the design of the NASA ACTS beacon receivers for adaptive control. Dr. Lin is a member of IEEE.

Joseph Li-Jen Yang received a B.S. in computer science from Tamkang University, Taiwan, R.O.C. During the period of October 1987 through October 1988, he was assigned by the Directorate General of Telecommunications (DGT) of Taiwan to work in the Communications Technology Division at COMSAT Laboratories. Prior to this, he was an engineer at the Taipei Earth Station, Taiwan, for 12 years. Following his work at COMSAT, he joined the DOMSAT project of the Taiwan DGT.



Translations of Abstracts

Simulation de systèmes à étalement de spectre

L. C. PALMER ET P. Y. CHANG

Sommaire

Les systèmes à étalement de spectre agrandissent la largeur de bande des signaux de télécommunication de façon apparemment aléatoire, d'où protection du caractère confidentiel de la transmission et/ou résistance au brouillage. La simulation de ces systèmes exige la modélisation des produits bande-temps ($W.T_b$) élevés, caractéristiques de ces signaux. Un programme informatique, qui simule les systèmes à pseudo-bruit à séquence directe et les systèmes à sauts de fréquence est décrit dans l'exposé. Les résultats des essais sont présentés. Les simulations utilisent notamment le principe d'échantillonnage d'importance, ce qui permet de réduire les temps de simulation des systèmes à sauts de fréquence.

Modems programmables MDP-2, MDP-4, MDP-4 à décalage faisant appel à des processeurs de signaux numériques

L.-N. LEE, A. SIENOV ET M. K. ENG

Sommaire

La conception d'algorithmes d'une classe de modems programmables basés sur des processeurs de signaux numériques est présentée dans cette étude. Ces modems de coût modique peuvent travailler à des débits binaires de quelques centaines à plusieurs centaines de milliers de bits par seconde, dans une gamme étendue d'applications allant des petites stations terriennes à faible densité de trafic aux terminaux à porteuse monovoie, terrestres, maritimes ou aéronautiques. Les formats de modulation MDP-2, MDP-4 et MDP-4 à décalage, dans les modes de fonctionnement continu et par paquet ont été utilisés dans les modems prototypes. Pour pouvoir utiliser la vaste gamme de débits binaires, une architecture de multiprocesseur pouvant être adaptée à diverses configurations d'applications particulières a été conçue. La réalisation concrète de cette architecture est examinée pour certaines applications potentielles et les algorithmes pour l'acquisition dans le mode par paquets et le mode continu sont décrits. Les résultats des essais de performance typique, en terme de taux d'erreur binaire dans une voie en présence d'un bruit blanc gaussien, sont donnés pour chacune de ces applications. Les résultats corroborent les prédictions théoriques.