COMSAT
Technical Review

Volume 19 Number 1, Spring 1989

COMSAT TECHNICAL REVIEW is published twice a year by 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

COMSAT TECHNICAL REVIEW

Volume 19 Number 1, Spring 1989

1 N-MAC–A MULTIPLEXED ANALOG COMPONENT VIDEO TRANSMISSION FORMAT RASTER COMPATIBLE WITH NTSC **L-N. Lee** AND **M. D. Redman**

27 A MOTION-ADAPTIVE THREE-DIMENSIONAL COMB FILTER FOR THE TRANSMISSION OF NTSC-ENCODED VIDEO **D. W. Power** AND **L-N. Lee**

45 STATISTICAL DETERMINATION OF TEST CARRIER LEVELS FOR PASSIVE INTERMODULATION TESTING **M. H. Abdelaziz** AND **A. E. Atia**

63 A MODEL FOR PLANNING SATELLITE COMMUNICATIONS SYSTEMS **J. L. Wernimont, W. L. Cook** AND **G. J. H. Brown**

99 RANDOM ACCESS WITH NOTIFICATION—A NEW MULTIPLE-ACCESS SCHEME FOR VSAT NETWORKS **D. M. Chitre, A. C. Briançon** AND **R. Kohli**

123 DEVICE AND CIRCUIT MODELING FOR A 3.5- TO 6.5-GHZ GAAS MONOLITHIC DUAL-GATE FET SWITCH MODULE **M. C. Fu** AND **R. K. Gupta**

145 TRANSLATIONS OF ABSTRACTS
FRENCH 145 SPANISH 149

153 AUTHOR INDEX, CTR 1988

155 INDEX OF 1988 PUBLICATIONS BY COMSAT AUTHORS

Index: communication satellites, direct broadcast satellites, television, transmission

N-MAC—A multiplexed analog component video transmission format raster compatible with NTSC

L-N. LEE AND M. D. REDMAN

(Manuscript received February 9, 1989)

Abstract

A multiplexed analog component television transmission format raster (N-MAC) that is compatible with NTSC was developed for the Satellite Television Corporation in the early 1980s. This paper describes the detailed format of N-MAC and the implementation and performance of an engineering model. Subjective test results are also included.

Introduction

The National Television Standards Committee (NTSC) color television signal format design is based on two criteria: compatibility with monochrome receivers and conservation of transmission bandwidth. The three primary color signals (red, green, and blue) are linearly combined into a luminance signal of full bandwidth and two color difference signals of less bandwidth. If transmitted faithfully, the primary color signals can be recovered by linear combination of the luminance and color difference signals in the color receivers, whereas the luminance signal can be used directly by monochrome receivers.

To conserve bandwidth, the two color difference signals modulate the in-phase and quadrature components of a color subcarrier that occupies the higher portion of the video bandwidth. The frequency chosen for the color

subcarrier is an odd multiple of one-half of the line scanning frequency at the upper portion of the luminance bandwidth. This choice of frequency allows interleaving of luminance and chrominance signals, with the spectral energy of the chrominance signal being concentrated at odd multiples of the half-line frequency, as illustrated in Figure 1. Thus, luminance and chrominance can share the same frequency band. Since the color subcarrier must reverse its phase in every frame (as well as every line) for the same color, chrominance-to-luminance crosstalk is removed to some extent through spatial and time integration. Similarly, noncoherent luminance-to-chrominance crosstalk is also removed to some extent by coherent demodulation of the subcarrier phase.

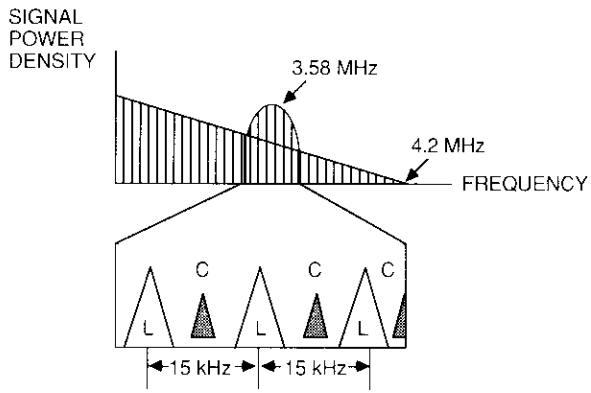


Figure 1. *Typical NTSC Television Spectrum*

As with any composite format, the frequency-multiplexed nature of NTSC has limitations. The spectrally interleaved luminance and chrominance signals are difficult to separate using inexpensive receivers, particularly for scenes that have vertical and horizontal details. The interleaved signals are especially difficult to separate after transmission through a nonlinear channel. It is also difficult to improve the resolution of the NTSC signal because of the presence of the color subcarrier and the incomplete separation of chrominance from the luminance signal.

The multiplexed analog component (MAC) format was first suggested by the Independent Broadcast Authority (IBA) in the United Kingdom as an alternative to the color subcarrier approach [1]. Instead of using a chrominance

subcarrier, the MAC format compresses the active portions of the luminance and chrominance in time and transmits them sequentially in a time-multiplexed format. A MAC receiver separates the components from the time-multiplexed format and decompresses them to their original form. For television sets with component inputs, the MAC format provides a means for direct component reception. For regular sets with composite input, the time-multiplexed transmission reduces the luminance and chrominance crosstalk caused by channel nonlinearities. However, the original IBA design is intended for a 625-line raster, and most of the design considerations assume a phase-alternate-line (PAL)-based video source, which makes it difficult to adopt directly in NTSC countries.

For direct broadcast satellite (DBS) application, the compression ratios of the original IBA proposal produce very wide baseband bandwidth, which in turn requires considerably greater RF bandwidth and satellite e.i.r.p. than NTSC. Potential DBS application in the U.S. or other NTSC countries requires a redesign of the original MAC format. For example, Goldberg *et al.* [2] have proposed a time-multiplexed component (TMC) format for transmitting the NTSC-compatible portion of a high-definition television signal.

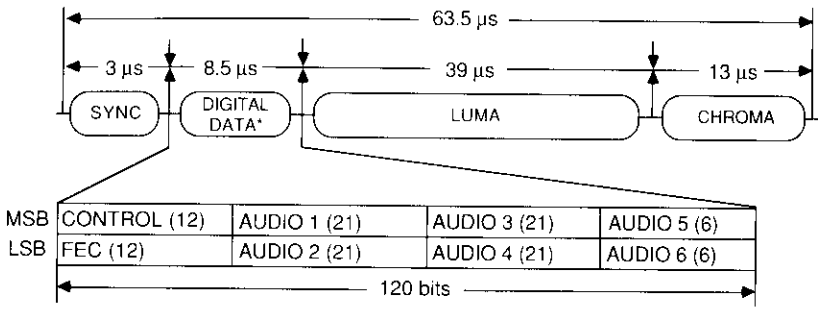
During 1982–1983, COMSAT Laboratories independently developed a MAC format (N MAC) that is compatible with the NTSC raster. An engineering model was developed, and objective and subjective tests were conducted over an IF link. The digital channel of this design was incorporated by Scientific Atlanta Digital Video Systems into their B-MAC product [3]. This paper describes the N-MAC format, including its design considerations, implementations, and performance.

Signal format

The N MAC signal format assumes a raster with 525 lines/frame, 30 frames/s, 2:1 interlacing, and 4:3 aspect ratio. The active video begins on line 22 in each field. The line time is approximately 63.5 μs. Each active line can be divided into periods for synchronization, digital data, luminance, and chrominance, as illustrated in Figure 2. The digital data portion includes control, audio, and parity checks for a forward error correction code. Any of the digital audio channels can be used to send data instead, if desired.

The luminance and chrominance signals are obtained by first linearly combining the red, green, and blue (RGB) signal according to the following equations:

$$Y = 0.30R + 0.59G + 0.11B$$
$$U = -0.30R - 0.59G + 0.89B$$
$$V = 0.70R - 0.59G - 0.11B$$



* Data modulated in 2 bits per ASK symbol: an MSB and an LSB.

Figure 2. *Time-Multiplexing Format for the N-MAC Signal Within a Scan Line*

The active portion of the Y, U, and V signals is assumed to be 52 μs. The bandwidth of the Y signal is limited by a high-order elliptical filter with an equiripple bandwidth of 4.2 MHz. The bandwidth of the U and V signals is limited by a transitional filter with 3-dB bandwidth of 1.25 MHz. The luminance signal is then time-compressed by a ratio of 4 to 3. The chrominance signal is first filtered vertically and then decimated by a factor of 2 to 1 (*i.e.*, the signal in every other line is dropped). The resultant signal is time-compressed by a ratio of 4 to 1, and the U and V signals are time-multiplexed in alternating lines. As a result of time compression, the luminance signal occupies 39 μs, one of the chrominance signals occupies 13 μs, and the total active video signal occupies 52 μs in each active scan line. Also due to time compression, the video bandwidth is expanded to 5.6 MHz for the luminance and to approximately 5 MHz for the chrominance.

The remaining 11.5 μs in each active scan line is divided into synchronization, digital data and audio, and guard time. The guard time is necessary to separate the signals in time. The synchronization signal establishes horizontal synchronization, DC level, and the receiver master clock frequency. The synchronization waveform is completely digital and is transmitted by the data modulator as part of the digital data, with a pulse rate of 7.16 MHz. The waveform consists of eight alternating pulses corresponding to a reference burst frequency of 3.58 MHz, and a unique-word pattern.

The remaining time is sufficient to transmit 60 data pulses. Using four-level amplitude-shift keying (ASK), two data bits can be transmitted per pulse, so that total of 120 bits of data per scan line can be supported. The outer levels of the four-level ASK signal are the same amplitude as the two-level

pulses used for synchronization. In the example shown in Figure 2, the 120 bits are divided into 24 bits of control data and 96 bits of audio (or data). To provide additional error protection, the control data are coded using a rate 1/2 convolutional code. Consequently the control channel has an information rate of 12 bits per scan line.

The 96 bits of uncoded data per line provide a data channel capacity of 1.5 Mbit/s and are capable of supporting up to four high-quality program audio channels and two supplemental audio channels, or any other combination of audio and data channels desired. To simplify the receiver hardware design, 10 μs in each of the unused lines will also be used for synchronization, data, and audio transmission.

The four levels of the digital signal correspond to four frequency tones after frequency modulation. The separation of the levels corresponds to the separation of the tones in the RF spectrum. For a given RF bandwidth and carrier-to-noise ratio (C/N), the bit error rate (BER) performance is a function of the separation, which can be optimized experimentally. For good BER performance, the peak-to-peak value of the digital signal normally is less than that used for the active video, and its mean is at the midscale value of the video amplitude range.

In one of the unused lines in each field, a unique word occupying the entire duration of active video (52 μs) is employed to provide field synchronization. This long unique word also provides a means for checking the horizontal synchronization. The video portions of all other unused lines are available for vertical interval test signal (VITS), teletex, and other services.

The video signal can be scrambled by interchanging the relative positions of luminance, chrominance, and audio within the line, as well as the sequence of video scanning. A total of up to 24 combinations can be obtained to achieve effective video scrambling. Eight of the 24 possible combinations are depicted in Figure 3, with the audio position fixed. As shown in the figure, the synchronization and control signals remain at a fixed position within the scan line in all cases. This ensures proper operation of the descrambler.

Design considerations

The most important considerations for a MAC format design are time-compression ratios for the video signals, means of multiplexing and transmitting the video signal, and video signal processing for bandwidth reduction. These considerations for N-MAC are discussed in this section.

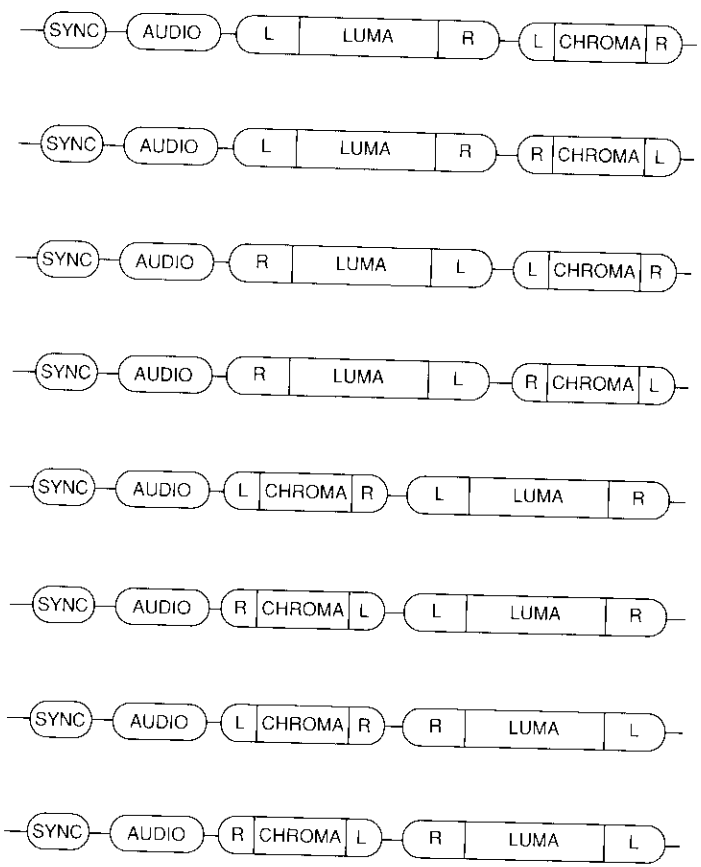


Figure 3. *Eight of the 24 Possible Signal Scrambling Sequences Within a Scan Line (R and L indicate the right and left sides of the original video signal)*

Compression ratio

The overriding rule for the choice of compression ratio is that the horizontal synchronization, the luminance, and the chrominance signal for each horizontal scan must be transmitted in one scan line. For the NTSC raster, each scan line has a duration of approximately 63.5 μs, of which the active video occupies approximately 52 μs.

The objective is to compress the active portion of the luminance and chrominance signals of each scan line so that the total time required to

transmit both of these signals and the synchronization signal sequentially is less than or equal to 63.5 μs. However, time compression increases the video baseband bandwidth of the signal. This in turn reduces the received baseband signal-to-noise ratio (S/N) for a given transmission link with a fixed carrier-to-one-sided-noise spectral density ratio, C/N_o, and a fixed RF bandwidth, B. This is easily seen as the received baseband S/N, given by

$$\frac{S}{N} = \frac{3}{2}\frac{C}{N_o}\left(\frac{\Delta F}{f_m}\right)^2\frac{1}{f_m}$$

where f_m is the video baseband bandwidth and ΔF is the frequency deviation, which is related to the baseband and RF bandwidth by Carson's rule:

$$B = 2(f_m + \Delta F) \quad .$$

This S/N reduction is worse than the cubic relationship shown explicitly in the formula because ΔF must also be reduced as f_m increases to maintain a constant channel bandwidth. Therefore, it is important that the compression ratios not be so excessive as to degrade performance unnecessarily.

Other considerations driving the choice of compression ratio include the following:

a. The luminance and chrominance S/N must be properly balanced.

b. The compression ratios selected must be realizable using practical implementations.

c. If possible, the active portion of the signal should be maintained within the 52-μs duration allocated for active video in NTSC so that it can easily be replaced by standard composite video if desired.

Without reliable subjective data to prove otherwise, it can be assumed that the human eye is equally sensitive to low-frequency noise in both the luminance and chrominance signals. Consideration (a) can then be interpreted as a desire to balance the S/N on the luminance and chrominance channels. A simple means of achieving this is to choose the compression ratio such that the S/N of the luminance and chrominance is about the same. Consideration (b) requires that the compression ratios be rational numbers of small integers, such as 3:2, 4:3, or 5:4. Consideration (c) may require that the sum of inverses of luminance compression ratio and chrominance compression ratio be unity if they are expressed in rational numbers less than unity. Based on these considerations, compression ratios of 4 to 3 for the luminance and 4 to 1 for the chrominance were selected for N-MAC. This choice clearly meets the requirements of considerations (b) and (c). It will be demonstrated that this choice also satisfies the requirements for consideration (a) in an FM link.

Program audio

Since the MAC format was developed to offer quality and versatility, high-fidelity stereo audio is a minimum requirement. The capability for an optional second language is also planned. Providing full stereo for both languages will require four audio channels. Instead of using multiple audio subcarriers, the program audio can be transmitted most efficiently by encoding it in digital form.

A high-quality adaptive delta modulation algorithm was selected based on two considerations. First, the delta modulation decoder is very easy to implement. The requirement for component precision is at least one order of magnitude less than for a pulse-code modulation (PCM)-based system achieving similar dynamic range. Also, the higher sampling rate associated with delta modulation requires no anti-aliasing filters. Second, delta modulation is inherently robust in noisy channels. For binary symmetric channels, no distortion can be detected by average listeners for a BER equal to or less than 10^{-3}. At a BER of 10^{-3}, the distortion is noticeable, but is not annoying until the BER is well above 5×10^{-3}.

To preserve the fidelity of the signal, the program audio must be encoded at a bit rate of at least 250 kbit/s. To simplify the receiver design, the bit rate chosen is 21 times the horizontal rate, or 330 kbit/s.

Data modulation

Digital audio and other digital data must be modulated and transmitted with the video. The following three options, corresponding to three versions of the IBA design, are available:

- *A-MAC*. A form of phase-shift keying (PSK) at a suitable subcarrier frequency.
- *C-MAC*. A form of PSK directly to an IF frequency transmitted in bursts in horizontal and/or vertical blanking intervals.
- *B-MAC*. ASK of baseband signal levels transmitted in bursts during horizontal and/or vertical blanking intervals.

The A-MAC option is not desirable from either the performance or cost standpoints. A greater baseband bandwidth is required, which in turn reduces the amount of frequency deviation allowed for a given FM link. This is particularly undesirable because the MAC signal already has a baseband bandwidth greater than normal video due to time compression. The presence of a digital subcarrier at the high end of the baseband also restricts the amount of overdeviation allowed for FM transmission. The performance of the digital subcarrier is adversely degraded by truncation noise, which is particularly severe near and below the FM threshold, and by the crosstalk between the video and the digital subcarrier. The cost of an additional PSK demodulator is also undesirable.

C-MAC is very desirable from a performance standpoint because a coherent demodulation performance close to that of theoretical additive white Gaussian noise (AWGN) may be achieved without crosstalk. However, the cost of a burst-mode PSK demodulator is higher than that of the continuous PSK demodulator required for A-MAC.

ASK/FM, which is used in B-MAC and is also known as pulse amplitude modulation (PAM)/FM, is a form of partially coherent demodulation. The signal format is equivalent to frequency-shift keying (FSK) with continuous phase. The performance is slightly better than with noncoherent FSK demodulation, but about 2 dB worse than with coherent demodulation. An ASK detector, which consists of only comparators and registers, is very inexpensive to implement. Since the video signal itself must be transmitted with reasonable S/N, the ASK data signal will be able to tolerate reasonable loss of power efficiency compared with the most efficient C-MAC case. Therefore, ASK/FM was selected for the N-MAC system.

Given that an ASK/FM type of modulation is selected, the number of ASK levels must be determined. The likely candidates are two, four, and eight levels. Lucky, Saltz, and Weldon [4] have shown that the RF spectra for two, four, and eight levels are quite similar for a given data throughput when optimized for the best BER performance. However, the selection of ASK pulse levels has a direct bearing on the shape of baseband spectral density. Therefore, the choice largely depends on the data rate to be supported and the maximum baseband bandwidth allowed.

The transmission data rate must be a multiple or submultiple of the video line rate. For convenience in hardware implementation, the sampling rate should be related to the data rate. The design choice for sampling rate is determined by Nyquist theory. For a 4:3 compression with 4.2 MHz of uncompressed video bandwidth, the compressed video has a bandwidth of 5.6 MHz. A sampling rate of approximately 14 MHz is required for a reasonably simple filter design. Thus, a data rate of approximately 14 Mbit/s is desirable. If the video is to be transmitted by satellite with a usable transponder bandwidth of 24 MHz for DBS application, the 14-Mbit/s data rate corresponds to a bandwidth-time product (BT) of 1.6, which is reasonably efficient for a suboptimal modulation such as ASK/FM.

With two-level ASK, the baseband bandwidth is very wide compared to the video bandwidth. This is undesirable, particularly if transmission of the N-MAC signal over the existing cable system is considered. Four-level ASK is better suited for this application because most of the signal energy is

concentrated within the video bandwidth. Although the baseband bandwidth of the eight-level ASK signal is even less, there appears to be no advantage in choosing eight-level signaling over four-level signaling because the baseband bandwidth of the four-level signal is already less than that required for video transmission. Furthermore, eight-level ASK detection requires greater implementation complexity. Thus, four-level ASK/FM appears to be the best choice for data modulation.

Video signal processing

The choice of compression ratio implicitly requires that the two color difference signals be transmitted in alternate lines. As mentioned previously, the baseband bandwidth is proportional to the time compression. If both color difference signals had to be transmitted in each line, greater time compression, and thus greater baseband bandwidth, would result. Observing that the horizontal resolution of the two color difference signals is only one-fourth that of the luminance signal in NTSC, it is reasonable to conclude that no adverse subjective effect will be observed by dropping the vertical resolution of the color difference signals to one-half that of the luminance signal. Thus, it was decided to transmit each of the color difference signals in alternate lines. Since the chrominance signals at the source have full vertical resolution, vertical prefiltering of the color difference signal is required before decimation, to prevent vertical aliasing. Proper interpolation is also required at the receive end to restore the decimated signal.

Base on computer simulation, it was determined that a three-tap finite impulse response (FIR) filter with coefficients of 1/4, 1/2, and 1/4 is most suitable for the transmit prefilter, whereas a two-tap FIR filter (*i.e.*, simple averaging) is sufficient for interpolation at the receive side.

Implementation

To demonstrate and test the performance of the N-MAC transmission format, an engineering model of the transmit and receive baseband processors was developed and used in an IF/RF testbed. At the outset, it was necessary to decide whether to use digital technology and random-access memory (RAM)-based memories, or analog technology with perhaps charge-coupled-device-based memories. The digital approach requires conversion of the analog video signal sources using analog-to-digital (A/D) and digital-to-analog (D/A) converters. The technology associated with the fabrication of these parts has been advancing rapidly, and consequently their cost has been decreasing. In comparison, the analog approach will always have an intrinsic testing and reliability cost associated with its inherent noise and temperature sensitivity.

Therefore, it was decided that the use of digital technology and RAM-based memory in the baseband processor would ultimately yield a more reliable engineering model and lower production costs. This approach also offers a shorter development cycle for the model.

Figure 4 is a functional block diagram of the transmit baseband processor unit (BPU), which was designed to accept video in an RGB-plus-synchronization format. The external composite synchronization signal is separated into horizontal and vertical synchronization signals, which are used by the transmit timing/control hardware to establish all input and output timing. The RGB signals are first clamped to a fixed DC level and passed through scaling/matrix electronics to generate YUV video components. The Y (luminance) signal is then passed through a 4.2-MHz elliptic low-pass filter (LPF) to prevent aliasing when sampled, and digitized at 10.7 MHz (or three times the subcarrier frequency, $3f_{sc}$) by the following A/D converter. The chrominance components (U and V) are filtered at 1.25 MHz before being sampled at the chrominance subcarrier rate of 3.58 MHz, f_{sc}.

After digitization, all video processing is performed in the digital domain. This consists of writing the Y data signal into static RAM, which serves as both a rate compression buffer and a delay compensation buffer. That is, in addition to rate compression, the delay through this transmit luminance buffer is chosen so that the total luminance delay from the transmit input to the receive output matches the total U/V (chrominance) delay, including that due to chrominance vertical filtering.

A second consideration was to minimize the receive memory requirement at the expense of the transmit memory, since the targeted environment was a point-to-multipoint application and the receive-side equipment cost had to be minimized. The U/V data streams are written into buffer memories at their 3.58-MHz sample rate. Due to the chrominance vertical decimation process, the U and V data streams are vertically filtered on alternate lines so that a single vertical filter is time-shared between the two components on an alternate-line basis. Thus, the N-MAC engineering model is actually based on a line-pair multiplex format; that is, every line-pair time interval of the composite N-MAC waveform consists of two chrominance bursts, two luminance bursts, and a data burst.

Figure 5 illustrates the digital video processing and memory architecture of the transmit processor in greater detail. The Y buffer memory consists of a simple two-line ping-pong memory architecture in which one bank is being written into at a 10.7-MHz sample rate, while the other is being burst out at a 14.3-MHz burst rate of $4f_{sc}$. The figure also depicts the inherent output time-division multiplexing, as all memories have tri-state outputs. One Read Address bus is used to sequentially access all video components/memories.

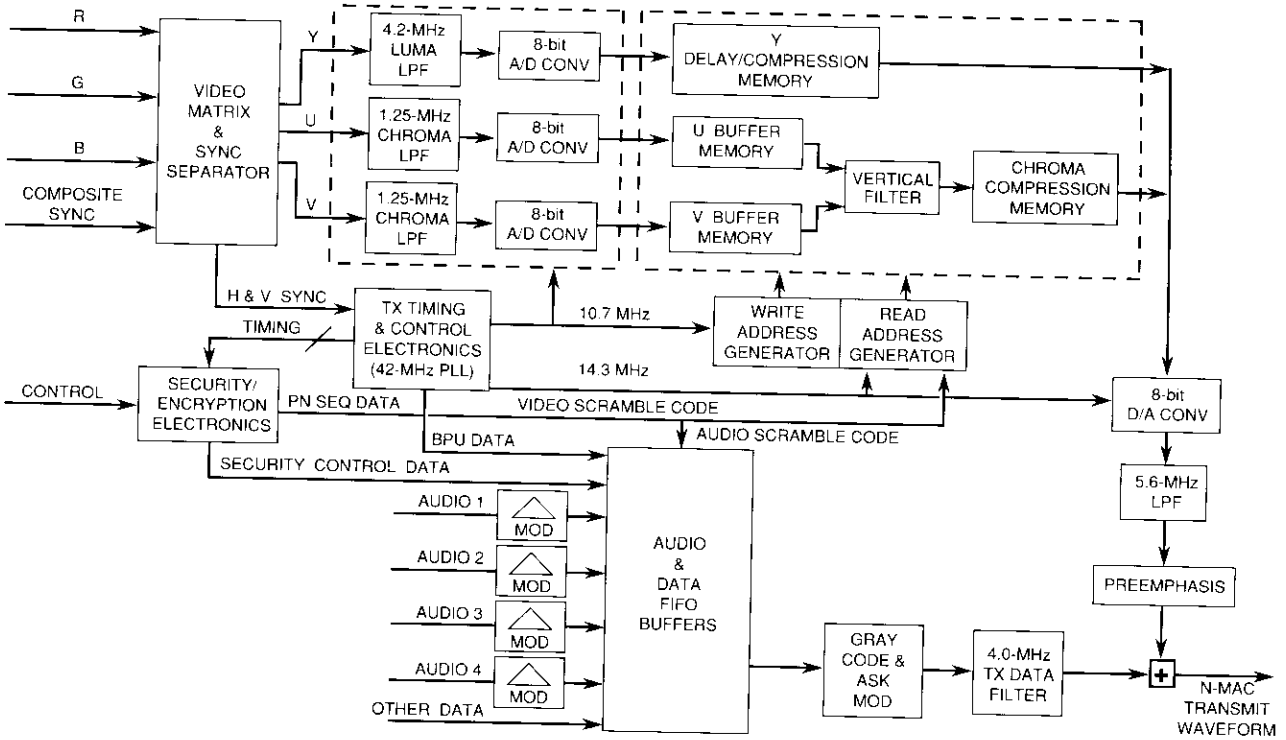


Figure 4. *Functional Block Diagram of the Transmit Baseband Processor*

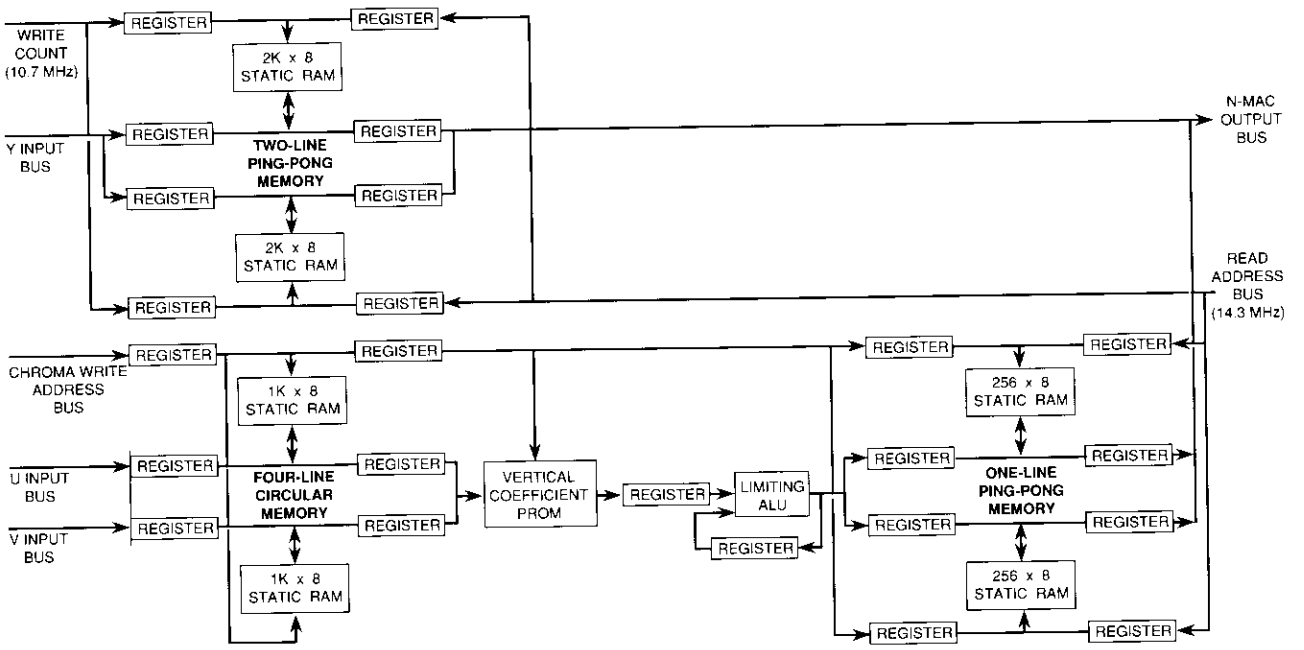


Figure 5. *N-MAC Transmit BPU Memory Architecture*

This control bus also provides for addressing the RAM in any of the possible scrambling modes under the control of a pseudorandom number (PN) sequence from a security subsystem. In the engineering model, the PN sequence is used to select/address a programmable read-only memory (PROM) which generates the selected address sequence.

The lower portion of Figure 5 illustrates the architecture of the chrominance processing hardware. Basically, both chrominance components are loaded into a four-line circular buffer memory at an input rate of f_{sc}. However, the input memory, together with the following vertical filter components, are clocked at a $3f_{sc}$ rate to permit the three-tap vertical filter to function at the f_{sc} rate. The input buffer and the following vertical filter could be considered as a simple state-machine which takes every input sample and the corresponding samples from the two previous video lines, passes them through a PROM-based multiplier, and accumulates the result in the arithmetic logic unit (ALU). This accumulation comprises the filter output, which is then loaded into the chrominance compression buffer. Although the final coefficients selected will not cause saturation, it should be noted that the ALU/accumulator has hard-limiting overflow protection to prevent the severe distortion caused by arithmetic rollover.

As shown in Figure 4, the video samples of the N-MAC waveform are converted again to an analog form by a D/A converter, aperture-corrected, and low-pass filtered by a 5.6-MHz elliptic filter. The signal from the filter output is then passed through a preemphasis network in preparation for transmission over an FM modem link. The four audio channels are first encoded using off-the-shelf delta-mod modules. They are then modulo-2 added bit by bit to a PN sequence or "scrambling code," and finally buffered into a first-in first-out (FIFO) memory for a whole line of 21 bits. The FIFO buffer output is read out at 14.3 MHz and grouped into 2-bit symbols, Gray-encoded, and converted into a four-level ASK signal. Additional security and control bits are handled in the same way.

The four-level ASK signal is filtered by an 8-MHz, four-pole Butterworth filter and time-multiplexed in the analog form with the preemphasized video. Note that this multiplexing function consists of simply adding the two waveforms, as long as the idle channel is forced to a midscale value during its idle period. Also, the digital data waveform is multiplexed with the preemphasized video without emphasis, because COMSAT's experiments have indicated that preemphasis of the digital data waveforms leads to poor BER performance.

The composite baseband waveform is normally sent to an FM modulator for transmission over a satellite link. The same portion of the waveform that is used for DC restoration at the receiver was used at the FM modulator to

achieve a DC-coupled effect. This is required in order to prevent the FM carrier from drifting during the data portion of the waveform due to the video content of the waveform. Omission of this DC coupling would result in temporary overdeviation of the FM carrier during scene changes and a "wandering" of the optimum data decision threshold at the receiver.

The transmit processor is partitioned into the following circuit boards:

- Video interface and matrix
- Anti-aliasing filters and A/D converters
- Digital signal processor
- D/A output encoder, filter, preemphasis, and analog multiplexer
- Audio encoder
- Four-level ASK modulator and data filter
- Transmit phase-locked loop.

Figure 6 is a functional block diagram of the receive BPU. The received N-MAC waveform first passes through the front-end electronics, which clamps the waveform to a specified DC level and adjusts the peak-to-peak signal to a specified level. The waveform is then fed into two separate paths: the video path and the data path. The front-end electronics are necessary to ensure that the four-level ASK data are optimally detected and that the received video waveform makes maximum use (*i.e.*, achieves maximum dynamic range and *S/N*) of the A/D converter which subsequently digitizes it.

The data path is first filtered by a 4-MHz, four-pole Butterworth LPF for noise reduction, and then demodulated by the four-level ASK demodulator. The demodulator uses the synchronization waveform to recover the clock. The receiver time base is based on a 14.3-MHz clock, which is phase-locked to the 3.58-MHz burst in the synchronization waveform. The demodulator also detects a horizontal unique word in the synchronization segment of the digital transmission in order to derive a horizontal timing reference for the timing and control electronics. During initial acquisition, horizontal synchronization is established prior to vertical synchronization. Vertical synchronization can then be more rapidly acquired by recognizing a vertical unique word in the first line of the vertical blanking interval.

The digital and audio data are demodulated by sampling the baseband data waveform, comparing the samples against three reference levels, and Gray-decoding the results. These demodulated data are then demultiplexed into separate FIFO expansion buffers for rate conversion. The typical receiver will have two audio stereo channels, even though four high-quality channels are available. Only selected channels will be buffered and delta-mod decoded. A separate low-speed data stream that contains security and control information

This control bus also provides for addressing the RAM in any of the possible scrambling modes under the control of a pseudorandom number (PN) sequence from a security subsystem. In the engineering model, the PN sequence is used to select/address a programmable read-only memory (PROM) which generates the selected address sequence.

The lower portion of Figure 5 illustrates the architecture of the chrominance processing hardware. Basically, both chrominance components are loaded into a four-line circular buffer memory at an input rate of f_{sc}. However, the input memory, together with the following vertical filter components, are clocked at a $3f_{sc}$ rate to permit the three-tap vertical filter to function at the f_{sc} rate. The input buffer and the following vertical filter could be considered as a simple state-machine which takes every input sample and the corresponding samples from the two previous video lines, passes them through a PROM-based multiplier, and accumulates the result in the arithmetic logic unit (ALU). This accumulation comprises the filter output, which is then loaded into the chrominance compression buffer. Although the final coefficients selected will not cause saturation, it should be noted that the ALU/accumulator has hard-limiting overflow protection to prevent the severe distortion caused by arithmetic rollover.

As shown in Figure 4, the video samples of the N-MAC waveform are converted again to an analog form by a D/A converter, aperture-corrected, and low-pass filtered by a 5.6-MHz elliptic filter. The signal from the filter output is then passed through a preemphasis network in preparation for transmission over an FM modem link. The four audio channels are first encoded using off-the-shelf delta-mod modules. They are then modulo-2 added bit by bit to a PN sequence or "scrambling code," and finally buffered into a first-in first-out (FIFO) memory for a whole line of 21 bits. The FIFO buffer output is read out at 14.3 MHz and grouped into 2-bit symbols, Gray-encoded, and converted into a four-level ASK signal. Additional security and control bits are handled in the same way.

The four-level ASK signal is filtered by an 8-MHz, four-pole Butterworth filter and time-multiplexed in the analog form with the preemphasized video. Note that this multiplexing function consists of simply adding the two waveforms, as long as the idle channel is forced to a midscale value during its idle period. Also, the digital data waveform is multiplexed with the preemphasized video without emphasis, because COMSAT's experiments have indicated that preemphasis of the digital data waveforms leads to poor BER performance.

The composite baseband waveform is normally sent to an FM modulator for transmission over a satellite link. The same portion of the waveform that is used for DC restoration at the receiver was used at the FM modulator to achieve a DC-coupled effect. This is required in order to prevent the FM carrier from drifting during the data portion of the waveform due to the video content of the waveform. Omission of this DC coupling would result in temporary overdeviation of the FM carrier during scene changes and a "wandering" of the optimum data decision threshold at the receiver.

The transmit processor is partitioned into the following circuit boards:

- Video interface and matrix
- Anti-aliasing filters and A/D converters
- Digital signal processor
- D/A output encoder, filter, preemphasis, and analog multiplexer
- Audio encoder
- Four-level ASK modulator and data filter
- Transmit phase-locked loop.

Figure 6 is a functional block diagram of the receive BPU. The received N-MAC waveform first passes through the front-end electronics, which clamps the waveform to a specified DC level and adjusts the peak-to-peak signal to a specified level. The waveform is then fed into two separate paths: the video path and the data path. The front-end electronics are necessary to ensure that the four-level ASK data are optimally detected and that the received video waveform makes maximum use (*i.e.*, achieves maximum dynamic range and *S/N*) of the A/D converter which subsequently digitizes it.

The data path is first filtered by a 4-MHz, four-pole Butterworth LPF for noise reduction, and then demodulated by the four-level ASK demodulator. The demodulator uses the synchronization waveform to recover the clock. The receiver time base is based on a 14.3-MHz clock, which is phase-locked to the 3.58-MHz burst in the synchronization waveform. The demodulator also detects a horizontal unique word in the synchronization segment of the digital transmission in order to derive a horizontal timing reference for the timing and control electronics. During initial acquisition, horizontal synchronization is established prior to vertical synchronization. Vertical synchronization can then be more rapidly acquired by recognizing a vertical unique word in the first line of the vertical blanking interval.

The digital and audio data are demodulated by sampling the baseband data waveform, comparing the samples against three reference levels, and Gray-decoding the results. These demodulated data are then demultiplexed into separate FIFO expansion buffers for rate conversion. The typical receiver will have two audio stereo channels, even though four high-quality channels are available. Only selected channels will be buffered and delta-mod decoded. A separate low-speed data stream that contains security and control information

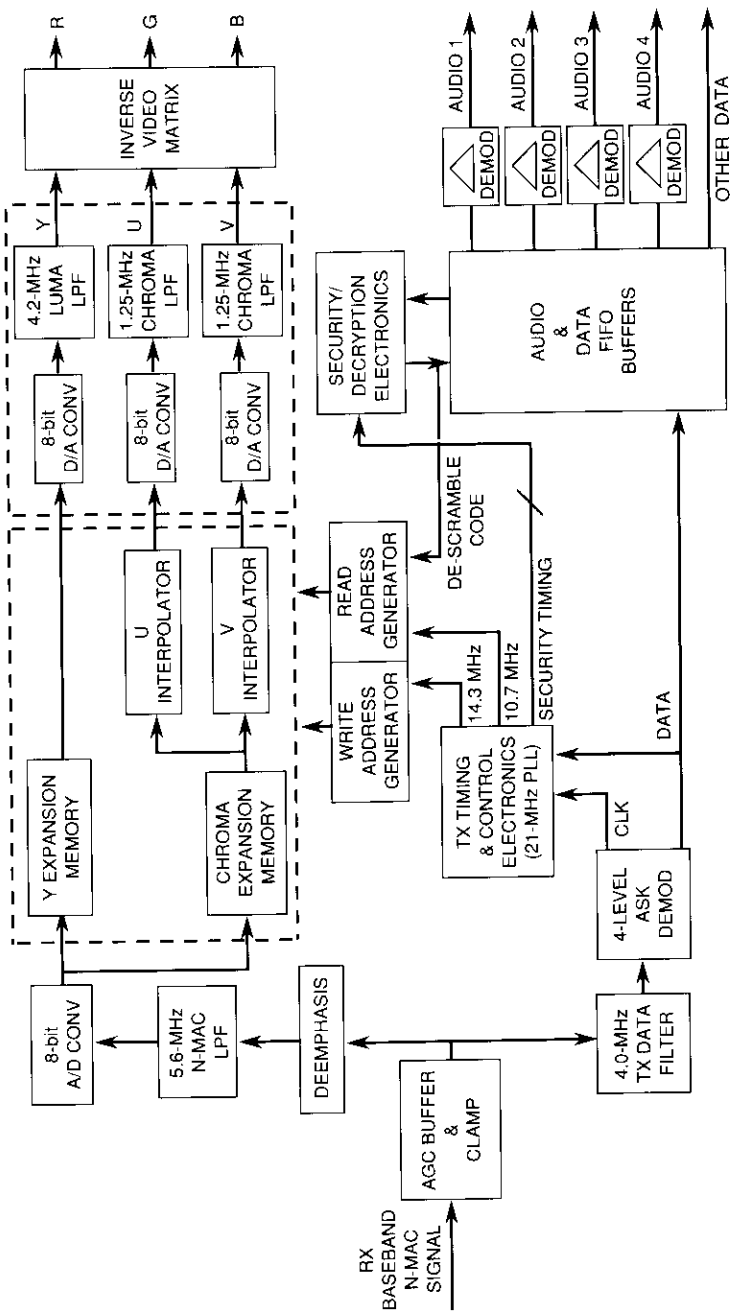


Figure 6. *Functional Block Diagram of the N-MAC Receive BPU*

is also demultiplexed. The receive security/decryption electronics uses this control information to regenerate the original PN sequence data needed to descramble the audio and video.

The video path is first deemphasized and low-pass filtered to 5.6 MHz to maximize the received *S/N*. The waveform is then sampled and quantized to 8 bits at 14.3 MHz. With proper synchronization, the Y and U/V portions of the MAC waveform are written into two separate time-expansion buffer memories. The Y expansion buffer is then read at the 10.7-MHz rate, D/A-converted to analog form, aperture-corrected, and low-pass filtered to 4.2 MHz. The U/V buffer is read at 3.58 MHz into a D/A converter, aperture-equalized, and filtered at 1.25 MHz. The missing chrominance component is generated by the interpolator electronics and transmitted along a path identical to that of the received component. All three video components are passed through an inverse matrix to regenerate an approximation of the original RGB video components.

Figure 7 illustrates the digital signal processing of the video component data streams in greater detail. First, the component is written to, and read from, a simple two-line ping-pong expansion memory with no processing other than the rate change. The single N-MAC input bus and the single Write Address bus illustrate a simple demultiplexing requirement in TDM systems. The Write Address bus must be sequenced in the same manner in which the Read Address was sequenced at the transmitter.

For U/V processing, the engineering model uses a simple one-line ping-pong expansion buffer and a two-line interpolation buffer. The same design used at the transmitter is used here; that is, all signals after the expansion buffer are read at the 10.7-MHz rate to yield a simple state-machine implementation. In every $3f_{sc}$ clock periods, one of the last-received chrominance components is sent to the output, and the other missing component is interpolated. Note that a single interpolator is used in the implementation, and simple registers are used as switches to time-share the interpolator between the two channels.

The receive BPU of the engineering model is partitioned into the following circuit boards:

- N-MAC input interface
- Input filter and A/D converter
- Digital video signal processor
- Three D/A converters with reconstruction filters
- YUV/RGB output matrix
- Data demodulator with filters and clock recovery
- Delta-mod decoders.

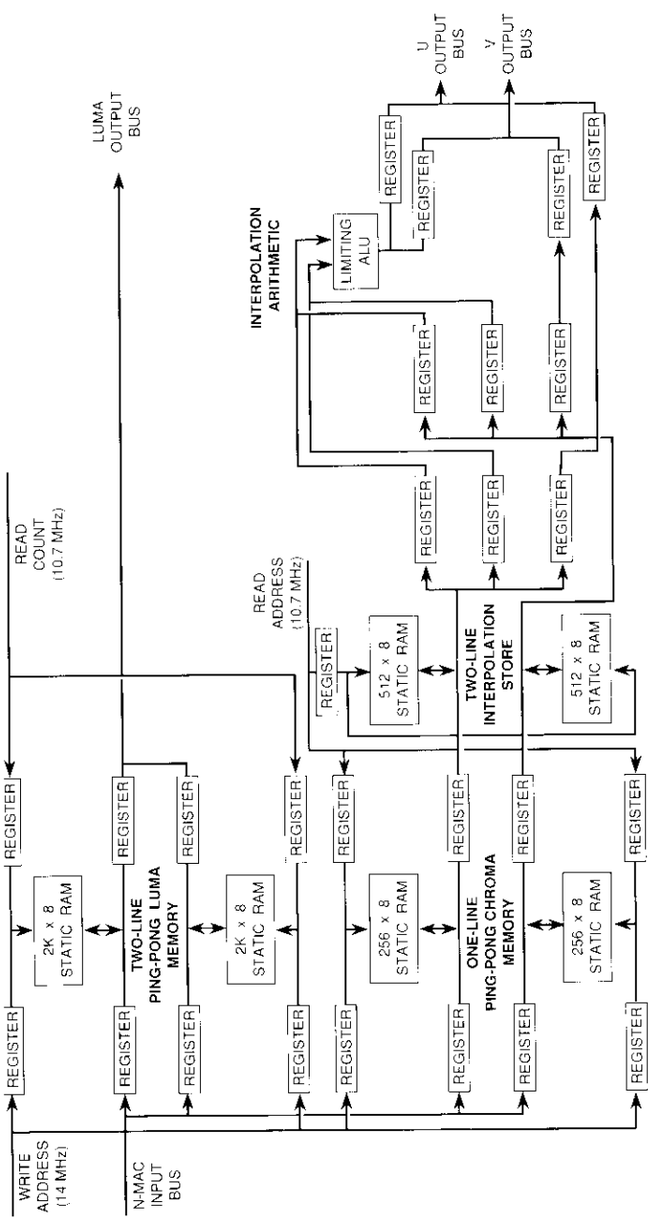


Figure 7. *N-MAC Receive BPU Memory Architecture*

Filtering delays within the system are compensated for digitally to within one 14.32-MHz sample. The longest delay within the system occurs during chrominance vertical filtering and interpolation. To minimize receiver complexity, this delay is preequalized at the transmit side.

The design of anti-aliasing filters for the chrominance channel also required special care. With a sampling frequency of 3.58 MHz and a chrominance bandwidth of 1.25 MHz, the filter must have reasonably sharp transition response. Elliptic filters, which have good passband ripple and sharp cutoff characteristics, produce excessive ringing that results in subjectively objectionable pictures. Gaussian filters, which have good impulse response, do not have the desired rolloff characteristics. A transitional filter that combines a Gaussian response in the passband and a Chebychev response in the stopband was selected and used effectively. The amplitude and group delay characteristics of this filter are shown in Figure 8.

The combination of an 8-MHz Butterworth filter at the transmit end and a 4-MHz Butterworth filter at the receive end was found to give the best BER performance for the ASK/FM signal among the class of Butterworth filters tested. To avoid the complexity of the extra filter, an experiment following completion of the engineering model demonstrated the feasibility of bypassing the 4-MHz data filter and relying completely on the 5.6-MHz elliptic filter for noise rejection, at the expense of slightly increased BER.

Performance measurements

Objective and subjective measurements were performed to evaluate the performance of N-MAC and the accuracy of the analytical model. Objective measurements included measurements of video *S/N*, frequency and delay response, linearity, component gain inequality, and delay inequality. Of these, only the video *S/N* is reported and discussed here.

The subjective measurements covered included quality rating and spectral truncation visibility tests, which entailed variation of *C/N* and deviation. The N-MAC transmit end was connected to the receive end through a 70-MHz IF link. The baseband signal was FM-modulated and combined with AWGN before being passed through a receive filter with noise bandwidth of 24 MHz. The filtered signal was then demodulated by a laboratory-quality limiter-discriminator before entering the N-MAC receive end. Figure 9 shows the test setup. It is assumed that the receive IF filter is the most severely band-limiting filter in the system. The effects of satellite nonlinearity, other transmission impairments, and the cascading of filters are not considered here.

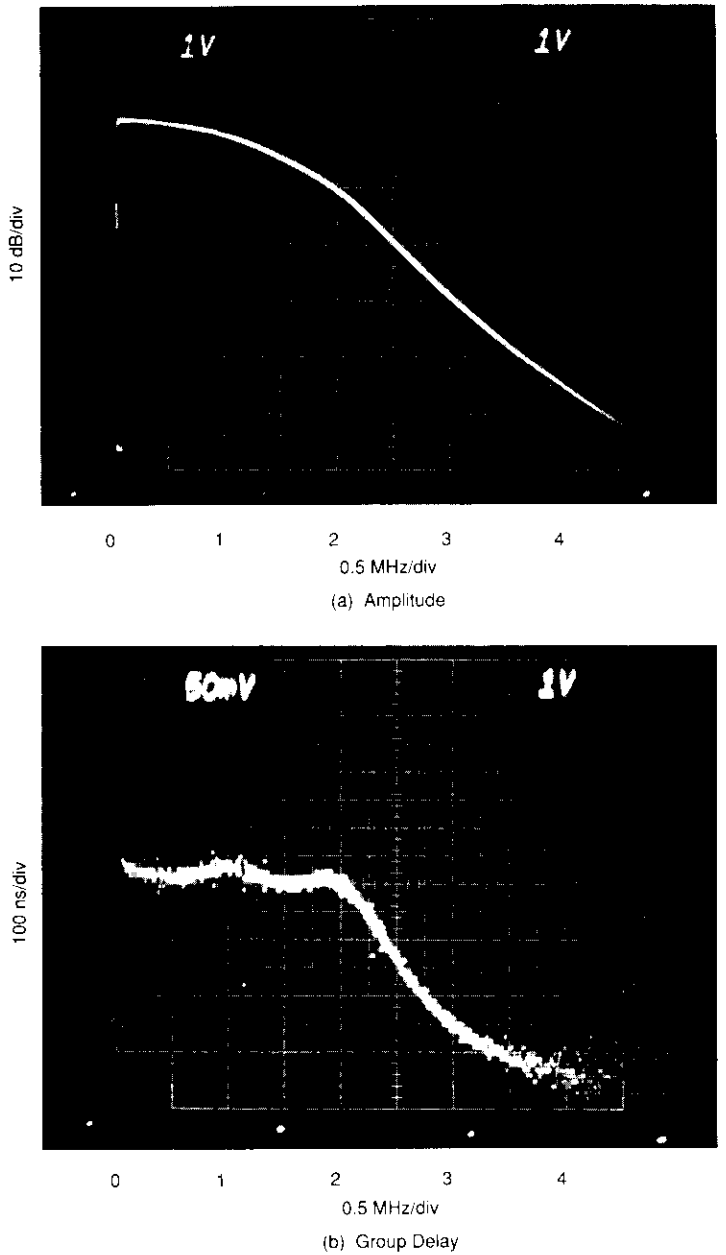


Figure 8. *Amplitude and Group Delay Response of the Transitional Filter Used in the Chrominance Channels*

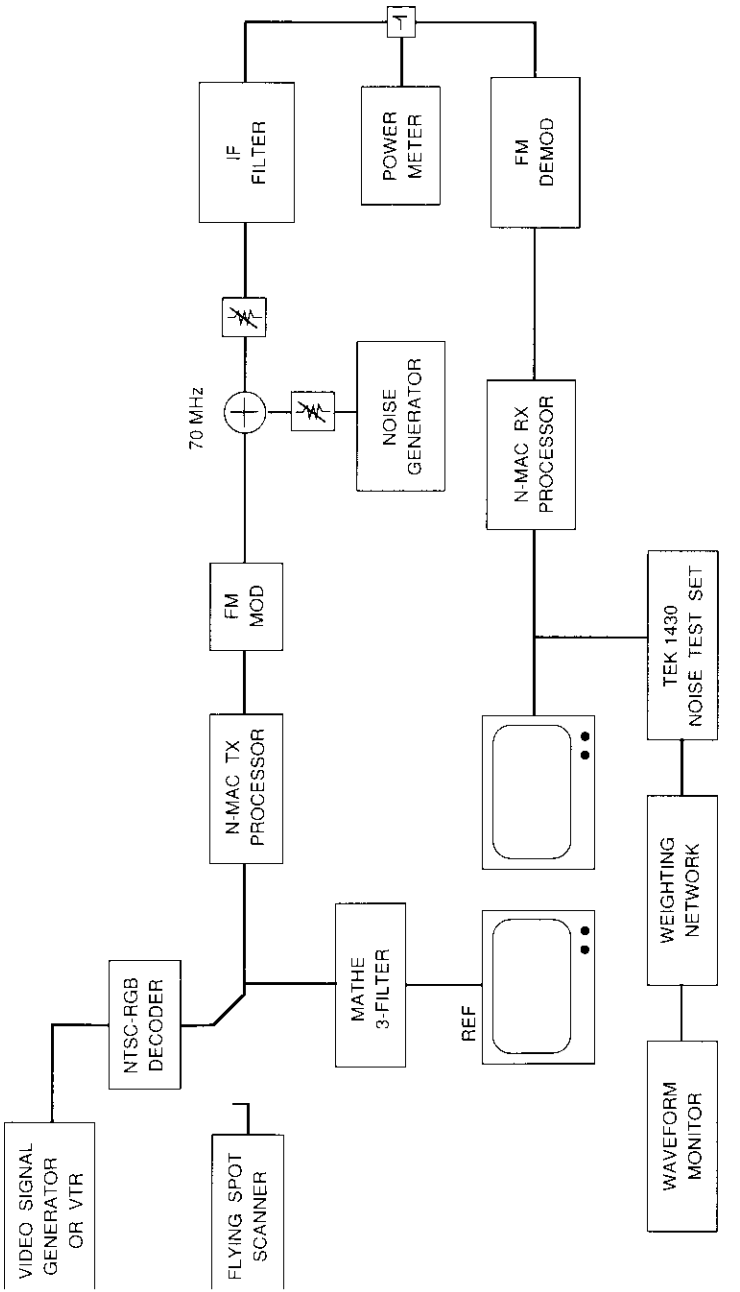


Figure 9. *N-MAC Test Setup*

Table 1 presents the measured weighted *S/N* as a function of *C/N*. The *S/N* is limited to 59 dB by the 8-bit quantization, as predicted. It is interesting to note that the chrominance signals experience vertical filtering and interpolation. Consequently, the interpolated lines of U and V generally exhibit 3-dB better *S/N* than the directly received lines. Thus, two values are given for each U and V measurement. If the averaging effect of human vision is considered, the weighted *S/N* of the luminance and chrominance channels differs by about 0.25 dB.

TABLE 1. WEIGHTED *S/N* MEASURED FOR THE LUMINANCE AND CHROMINANCE SIGNAL WITH 25-MHz IF BANDWIDTH AND 4-dB OVERDEVIATION

C/N	LUMINANCE *S/N* (dB)	CHROMINANCE *S/N* (dB)	
		ODD LINE	EVEN LINE
4	37	35	38
6	39	37	40
8	41	39	42
10	43	41	44
12	45	43	46
14	47	45	48
16	49	47	50
18	51	49	52
20	53	51	54
∞	59	57	60

Subjective tests were also conducted using the same test setup. The following five-point rating scale from CCIR Recommendation 500 was used:

Grade 5. Imperceptible
Grade 4. Barely perceptible
Grade 3. Perceptible but not annoying
Grade 2. Annoying
Grade 1. Very annoying

The reference picture was a component television signal from a high-quality film scanner, with the luminance and chrominance signals bandwidth-limited to 4.2 and 2.1 MHz, respectively.

Tests designed to rate quality as a function of *C/N* and truncation noise visibility are performed at a viewing distance of six times the picture height.

TABLE 2. TEST SLIDES AND NUMBER OF SESSIONS USED IN THE SUBJECTIVE TESTS

TEST SLIDES	SOURCE	QUALITY RATING (12 SESSIONS)	TRUNCATION VISIBILITY (8 SESSIONS)
Girl in Green Dress	SMPTE 14	x	x
Beach Scene	SMPTE 11	x	x
Toys and Blackboard	Phillips	x	x
Basket of Fruit	Phillips	x	x
Clown in Mirror	EBU6		x
Boats	EBU2		x

Table 2 summarizes the test slides used and the number of sessions conducted for each test. Each session consisted of sixty 10-second steps, with a 10-second separation between steps. Each session was observed by up to five viewers.

In the quality rating tests (Figure 10), the *C/N* ranged from 4 to 16 dB in 3-dB steps, and the FM deviation was 3, 5, and 7 dB over the Carson's rule deviation of 6.4 MHz in a Carson's rule bandwidth of 25 MHz with maximum baseband frequency of 5.6 MHz. The truncation noise visibility tests (Figure 11) were conducted with *C/N* set at either 9 or 14 dB. The overdeviation was 2 to 10 dB in 2-dB steps.

Based on the test results it can be concluded that, under the link parameters selected, the subjective quality at this viewing distance is basically dominated by the amount of impulse noise, rather than thermal noise. Therefore, a grade of 3.5 (50 percent of the observers considered the distortion barely perceptible, and the other 50 percent considered it perceptible but not annoying) is achieved at a *C/N* of 9 dB, which occurs at the onset of impulse noise for the particular FM demodulator used in the test. It was also observed that a large amount of overdeviation (up to 6 dB) is acceptable for typical picture material such as that used in the tests. This is because the baseband spectrum of the N-MAC signal typically contains very little high-frequency information.

Caution must be exercised in applying these results, particularly with regard to overdeviation. When critical test signals such as multiburst were used, considerable quality degradation was discerned by trained observers at 3- to 4-dB overdeviation. Large amounts of overdeviation may also cause multipath impairment due to signal leak through adjacent transponders and back [5], which may prove to be the limiting factor in practical operations.

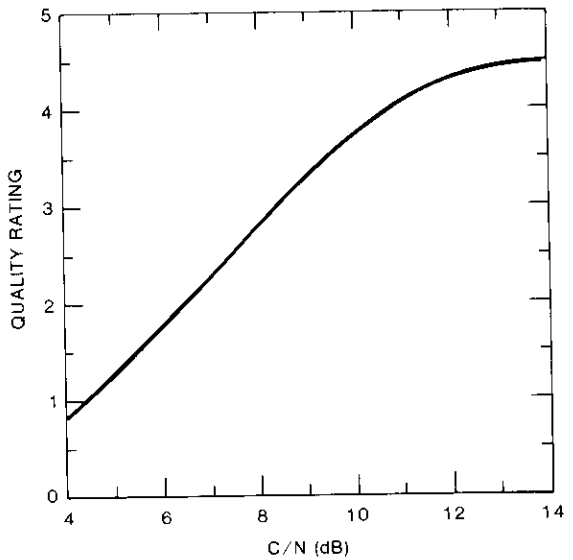


Figure 10. *Results of Subjective Quality Tests*

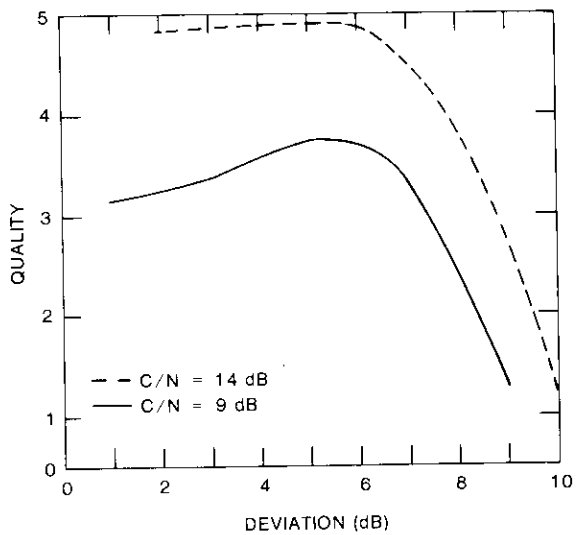


Figure 11. *Results of Truncation Visibility Tests*

Conclusions

The suitability and performance of the N-MAC transmission format have been successfully demonstrated using an engineering model. The 4:3 and 4:1 compression ratios for the luminance and chrominance signals, respectively, provided balanced *S/N* in the FM channel. Tests with the engineering model showed that the MAC format can tolerate greater overdeviation than NTSC, thus compensating for some of the *S/N* loss due to time compression. The four-level ASK/FM technique provided four high-quality program audio channels and other audio or data channels during the horizontal blanking interval. The digital synchronization is capable of maintaining synchronization at a *C/N* of about 0 dB. Because of the relatively shallow preemphasis and deemphasis used, the subjective performance of the video and audio is acceptable for a wide range of *C/N* below the FM threshold.

Acknowledgments

The development of N-MAC was a team effort. The authors would like to express their gratitude to many of their colleagues at COMSAT Laboratories and in other elements of COMSAT, particularly to R. Garlow for the selection of compression ratios, Y. Kao for the design of vertical filters, N. Becker for the design and implementation of the synchronization scheme, A. Shenoy for the design of the transitional filter, A. Gatfield for the objective and subjective tests, and P. Chang for the digital channel performance simulation.

References

[1] M. D. Windram, G. Tonge, and R. Morcom, "MAC-A Television System for High-Quality Satellite Broadcasting," IBA Experimental & Development Report 118/82, Independent Broadcasting Authority, Crawley Court, Winchester, Hants, U.K., August 1982.

[2] A. Goldberg, R. McMann, and J. Rossi, "A Two-Channel Compatible HDTV Broadcast System," CBS Technology Center Report 8/83-A, July 1983, CBS Technology Center, Stanford, Connecticut.

[3] J. D. Lowry, "B-MAC, an Optimal Format for Satellite Television Transmission," *SMPTE Journal*, Vol. 93, No. 11, November 1984, pp. 1034–1043.

[4] R. Lucky, J. Saltz, and E. Weldon, *Principle of Data Communication*, New York: McGraw-Hill, 1968.

[5] M. Wachs, "Analysis of Adjacent Channel Interference in a Multicarrier FM Communications System," *COMSAT Technical Review*, Vol. 1, No. 1, Fall 1971, pp. 139–170.

Lin-Nan Lee received a B.S.E.E. degree 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 processing (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.

Mark D. Redman received his B.Eng. (Electrical) and M.Eng. (Electrical) degrees from McGill University in 1973 and 1976, respectively. Previously employed as a Member of the Scientific Staff at Bell Northern Research in the Fiber Optic Systems Department, he joined COMSAT Laboratories as a Member of the Technical Staff in the Image Processing Department. His initial responsibilities primarily involved the research, design, test, and development of electronic telecommunications equipment in the areas of signal coding, multiplexing, and transmultiplexing. Subsequent work as a Staff Scientist included CAE systems, ASIC design, optical communications, and thin-route TDMA system design. He is currently Associate Manager of the Image Processing Department of the Communications Technology Division, where he is responsible for the development of TDMA video multiplex equipment utilizing high-speed digital signal processing techniques and electronics.

Index: television systems, image processing, signal processing, digital simulation, adaptive filters

A motion-adaptive three-dimensional comb filter for the transmission of NTSC-encoded video

D. W. POWER AND L-N. LEE

(Manuscript received February 1, 1989)

Abstract

An algorithm employing a motion-adaptive, three-dimensional comb filter for the suppression of luminance-chrominance crosstalk in NTSC-encoded video signals is presented. The filter, which consists of a motion-adaptive temporal comb for the static portions of the frame and a two-dimensional spatial comb for the time-varying portions of the frame, results in increased luminance bandwidth and greater luminance-chrominance separation than can be achieved with conventional comb filter designs. The proposed algorithm may be used to improve standard NTSC television transmission, or as part of a scheme by which high-definition television (HDTV) can be transmitted via an NTSC-compatible channel. The design principles were confirmed by computer simulation.

Introduction

The current U.S. broadcast standard was adopted by the National Television System Committee (NTSC) in 1953 to provide for the transmission of color television pictures while maintaining compatibility with the existing monochrome standard. The major problem facing video engineers at that time was how to transmit the approximately 2 MHz of color information necessary, along with 4 MHz of luminance information, and still meet the design criteria of compatibility and a 4-MHz total channel allocation. The result was a

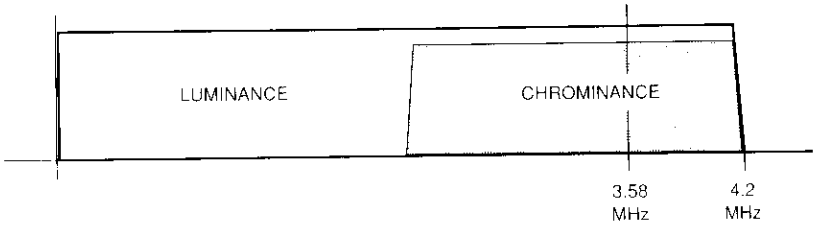


Figure 1. *NTSC Band-Shared Transmission Spectrum*

system in which chrominance and luminance information share the portion of the frequency spectrum from approximately 2 MHz to the band edge at 4 MHz, as illustrated in Figure 1.

This standard calls for the color difference signals to be quadrature amplitude modulated (QAM) about a suppressed subcarrier placed high in the luminance (or brightness) spectrum [1]. The use of color difference signals (which describe the difference between a color image and a monochrome image of the same brightness) with the suppressed subcarrier causes the subcarrier to exhibit the very desirable trait of disappearing in areas of pastel or neutral gray colors. The subcarrier chosen (3.579545 MHz) is an odd (455th) harmonic of one-half the line frequency. This causes the subcarrier phase to invert from one line to the next. As a result, the subcarrier appears as a barely visible checkerboard pattern on black-and-white monitors.

In theory, the spectral distribution of the luminance signal is much greater near harmonics of the line frequency [2]. Therefore, by placing the chrominance information in the region between these cusps (as shown in Figure 2), less of the luminance information is lost. NTSC encoding relies heavily on this principle, and simply modulates the chrominance information about the chrominance subcarrier and adds it to the luminance information, regardless of the spectral distribution of the luminance signal.

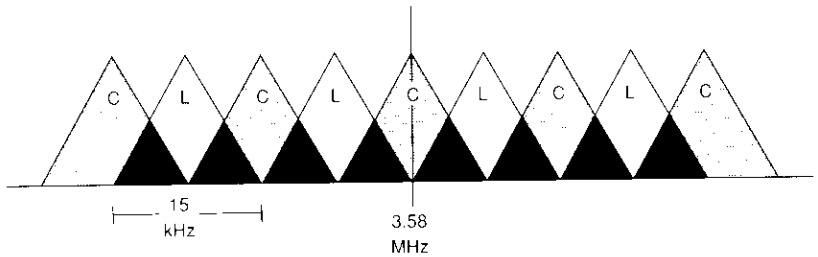


Figure 2. *Frequency Interleaving Near the Chrominance Subcarrier (based on Figure 2 of Reference 3)*

Figure 3 is a two-dimensional representation of the NTSC-encoded spectrum equivalent to the one-dimensional representation in Figure 1. Note that the chrominance subcarrier is at one-half the line frequency vertically, since it alternates its phase in each line. On the receive side, the chrominance information is removed by using either a simple notch filter or a two-dimensional filter commonly called a comb filter. The comb filter attempts to differentiate between the chrominance information placed about the odd multiples of one-half the line frequency, and the luminance centered about the even multiples.

Unfortunately, this band-sharing scheme leads to artifacts in the decoded color television signal. The presence of luminance in the areas specified for chrominance results in luminance information being incorrectly decoded as chrominance. This produces an artifact known as cross-chrominance, an effect often seen as a rainbow pattern in areas of high spatial frequency. Furthermore, the separation of the luminance at the receive side is not perfect because of hardware constraints placed on the filters. Typical NTSC sets use a simple notch filter which causes severe loss of luminance information near the color subcarrier. Use of a two-dimensional comb filter alleviates this problem somewhat, but loss of luminance resolution is unavoidable due to limitations on implementation complexity. As a result of this incomplete separation, portions of the chrominance signal can spill over into the decoded luminance, causing an artifact known as cross-luminance or dot-crawl.

A proposed improvement to this system involves "precombing," or removing those areas of the luminance that could be decoded as chrominance from the signal prior to adding the modulated chrominance [4]. This effectively

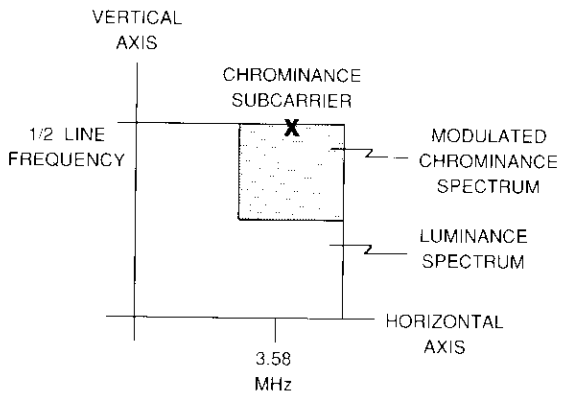


Figure 3. *NTSC Two-Dimensional Spectrum*

reduces the cross-luminance and cross-chrominance artifacts; however, the luminance bandwidth is necessarily limited in areas of high spatial frequencies.

The algorithm described here proposes to use the third, or temporal, dimension to eliminate cross-luminance and cross-chrominance artifacts without sacrificing the important elements of the high spatial frequencies.

System concept

Any video picture can be thought of as having resolution in three dimensions: horizontal, vertical, and temporal. The basic system concept presented here uses the temporal dimension to separate the luminance and chrominance during the still portions of the picture (areas of low temporal frequency), and by doing so allows the entire spatial frequency domain to be used for luminance information.

The NTSC chrominance subcarrier is characteristically high in frequency in the horizontal dimension (188 cycles per picture width), high in the vertical dimension (the subcarrier phase inverts on each line—this is in fact the Nyquist frequency in a sampled data system), and high in the temporal dimension (again, the subcarrier phase inverts from frame to frame). During still pictures, the modulated chrominance information can be thought of as occupying a single corner of the picture's three-dimensional frequency space, as indicated in Figure 4. In areas of low temporal frequency, complete luminance-chrominance separation can be achieved without loss of spatial resolution simply by applying a temporal low-pass filter to the received signal [5]. In fact, a two-tap temporal low-pass finite impulse response (FIR) filter can be implemented by computing the mean of the current frame and its predecessor.

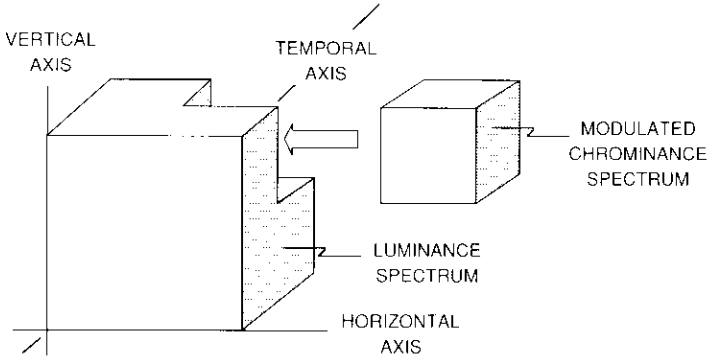


Figure 4. *Three-Dimensional Spectrum*

This temporal filtering approach yields optimum decoding in still pictures; however, in pictures containing motion (high temporal frequencies) the frames being averaged are not correlated, which results in serious errors in the decoded signals. Therefore, a motion-adaptive scheme which uses a temporal filter during still portions of the picture, and a more conventional two-dimensional spatial comb filter and transmit-side precomb during moving portions of the picture, must be implemented.

Transmit processing

The transmit processor can be thought of as two separate comb filters: one for still portions of the frame and one for moving portions. These two types of filters are multiplexed on a pixel-by-pixel basis, dependent upon a motion detector capable of predicting the temporal spectral content of the local video scene. This concept is illustrated in Figure 5.

The purpose of the transmit-side motion comb filter is to attenuate the luminance signal in areas of high diagonal resolution and thus avoid crosstalk with the modulated chrominance. This process is implemented as a seven-tap horizontal high-pass FIR filter in series with a five-tap vertical high-pass FIR filter. The result is then subtracted from the original signal. This process is known as precombing and results in lower luminance resolution at high spatial frequencies.

Precombing is necessary only during moving portions of the picture, because temporal processing completely removes any crosstalk in the still portions. In either case, the chrominance is band-limited in both the horizontal and vertical dimensions and modulated about the chrominance subcarrier frequency.

Since two types of processing are employed, an intelligent means of determining which to use at any particular time must be devised. A motion detector, which consists of a temporal high-pass filter and a digital comparator, will control the two multiplexed transmit preprocessors. A two-tap temporal high-pass filter can be realized by simply subtracting the current frame from its predecessor on a pixel-by-pixel basis. The absolute value of this difference is then compared with an empirically determined threshold. Differences that exceed this threshold are deemed *moving*; the remainder are deemed *still*. If a moving pixel is said to be still, multiple edges of an image may result. If a still pixel is said to be moving, some reduction of diagonal resolution may result, causing loss of picture details that contain the high frequency in both the horizontal and vertical directions. Since the artifacts caused by the former case are more severe, the threshold is skewed toward motion. Even so, experiments have shown that a typical frame contains three times as many still pixels as moving ones [6].

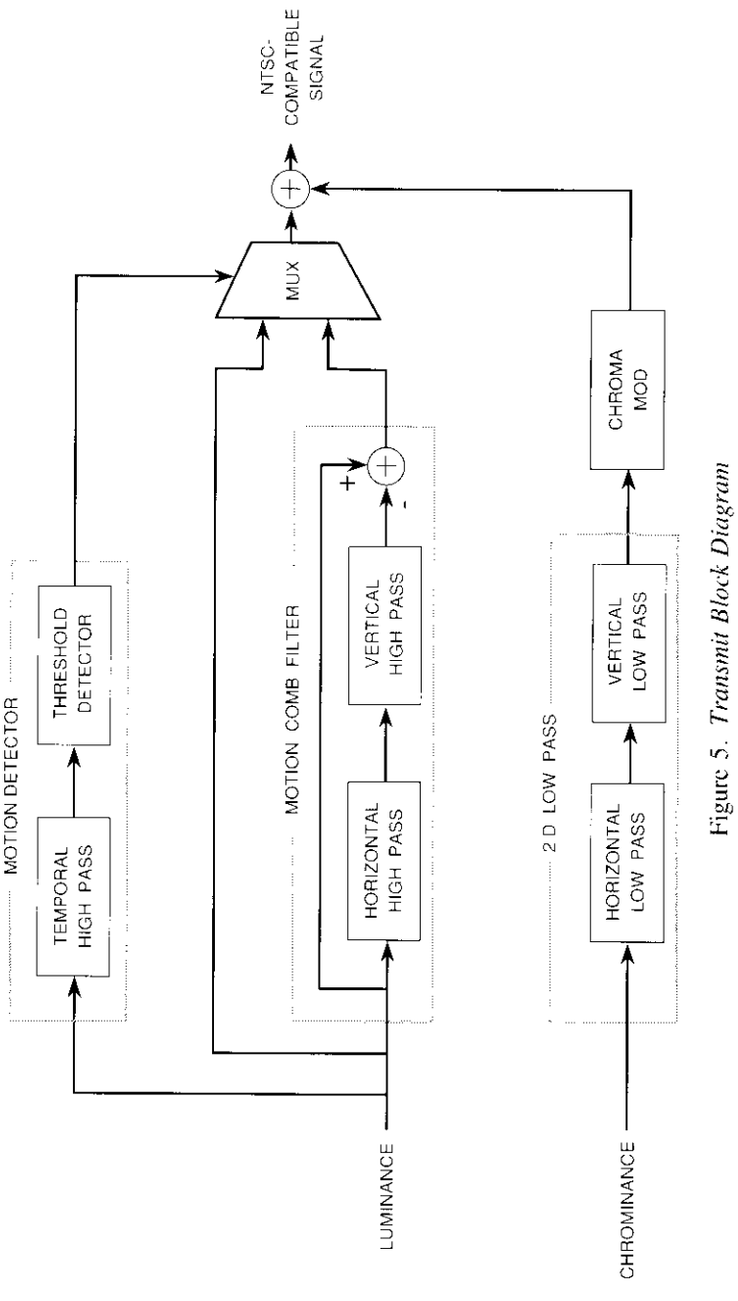


Figure 5. *Transmit Block Diagram*

The spectrum of the multiplexed luminance, as shown in Figure 6, now contains full spatial resolution in areas of low temporal frequency and is band-limited only in those areas where temporal resolution must be maintained. Naturally there is some loss of diagonal resolution in these areas; however, the human visual system is less sensitive to this loss during moving scenes. The chrominance, owing to its already narrow bandwidth, is simply band-limited without regard to its motion content, modulated about the chrominance subcarrier, and added to the precombed luminance. This produces a completely compatible NTSC signal.

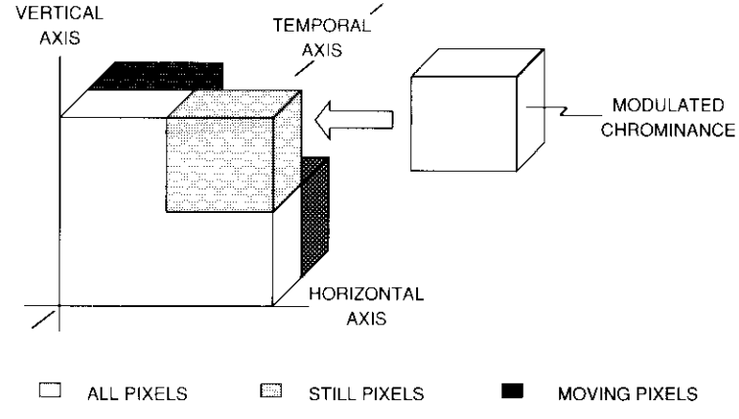


Figure 6. *Three-Dimensional Transmission Spectrum*

Receive processing

The receive processor architecture is similar to that of the transmit processor. The concept is the same: two comb filters are used—one for the still portions of the frame and a second for the moving portion. A motion detector controls the switching on a pixel-by-pixel basis. Figure 7 is a block diagram of the receive processor.

The received NTSC signal is presented to both comb filters. The temporal comb filter for "stills" is simply a two-tap temporal low-pass filter. The current frame is averaged with its predecessor, and the output is the still luminance. This output is subtracted from the received signal to obtain the still chrominance.

The comb filter for moving samples is a two-dimensional, high-frequency, band-stop filter. As before, this filter is implemented as two high-pass FIR

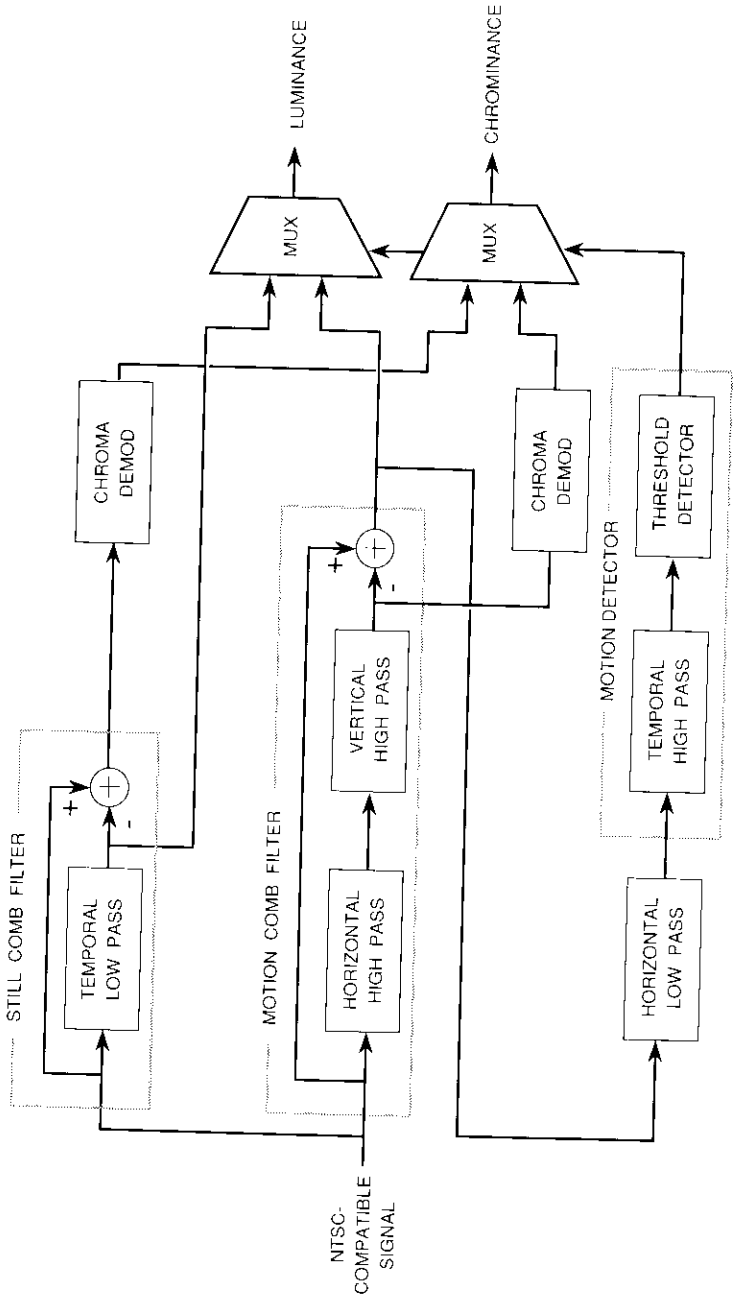


Figure 7. *Receive Block Diagram*

filters in series whose output is the moving chrominance. Subtracting this output from the received NTSC signal yields the moving luminance.

To avoid the need to transmit overhead information to control the switching of the receive comb filter multiplexer, the motion detector must predict the spectral content of the signal based on the received signal only. To accomplish this, the decoded moving luminance is also directed to the temporal high-pass filter and digital comparator of the motion detector. Because the decoded moving luminance contains the undesirable artifacts resulting from traditional NTSC comb filter decoding (*i.e.*, cross-luminance), a horizontal low-pass filter must be implemented prior to the temporal high-pass filter. Again, the motion detector is skewed toward motion to eliminate serious temporal artifacts.

Computer simulations indicate that two-tap temporal filters are sufficient in all cases discussed here. Therefore, only two frame memories are necessary for each transmitter and receiver. Given the current state of memory technology, this is not an unreasonable requirement. Similarly, the horizontal and vertical filters used to implement the two-dimensional spatial filters are of seven taps or fewer. Again, given the current state of technology, these filters can be realized in hardware with very few large-scale integrated (LSI) circuits.

System simulation

The simulation was performed in COMSAT Laboratories' Programmable Image Processing Facility. This facility is a minicomputer-controlled mass storage device capable of synchronizing to and storing any standard video input. A complete description of the capabilities of the facility is presented in the Appendix.

For this experiment, a two-channel configuration was used. The luminance signal (Y) was sampled at 14.31818 MHz, exactly four times the NTSC subcarrier frequency. The chrominance occupied the second channel. Two color difference signals (R-Y and B-Y) were each sampled at twice the subcarrier frequency and interleaved to form a single 14-MHz channel. These sampling frequencies were chosen because they satisfy the Nyquist criterion and enable NTSC chrominance modulation to be reduced to simply negating every other color difference sample.

Input video signals were generated from standard test slides using a wideband red/green/blue (RGB) flying spot scanner and a 525-line RGB camera. The inputs were converted into luminance and chrominance components and band-limited. The luminance signal was limited to 4.2 MHz, and each of the chrominance signals to 1.1 MHz. The video sequences were then digitized and stored for processing and evaluation.

A wide variety of sequences representing all types of video content were processed and evaluated to test the efficacy of each system component separately and as a unit. Because it is impossible to graphically display these video sequences on the two-dimensional printed page, a standard color bar is used to display some of the benefits of this processing technique. Figure 8a shows a standard color bar signal as it would be displayed using a monitor capable of accepting component (Y, R-Y, B-Y) inputs, and Figure 8b depicts a standard color bar display using conventional NTSC encoding and comb filter decoding techniques. Note the blurring on the edges of the bars, (particularly the green/magenta transition). This is caused by the incomplete luminance-chrominance separation inherent in this type of implementation. Figure 8c represents a standard color bar as it would appear when encoded and decoded using the adaptive three-dimensional comb filter decoder described in this paper.

Conclusions

An algorithm has been described which yields greater luminance-chrominance separation than conventional video processing techniques. The result is greater spatial resolution and the absence of the artifacts previously common to NTSC video transmission.

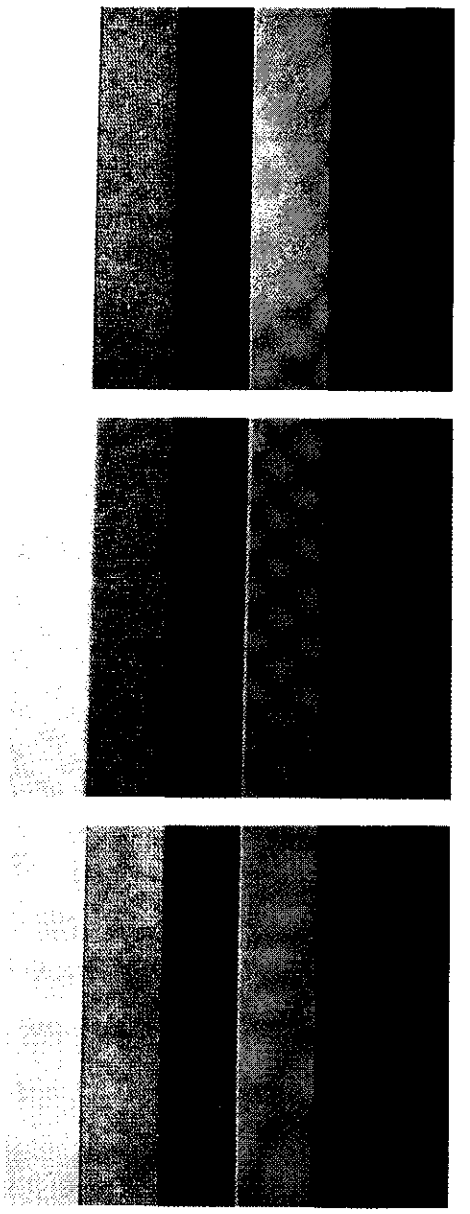
This technique may be applied simply to improve conventional NTSC transmission; however, its greater potential lies as part of an evolutionary technique to bring high-definition television (HDTV) into the home. One proposed application calls for the decoded video signal to be converted to a 525-line, 59.94-Hz sequentially scanned signal that would provide enhanced-definition television (EDTV), with no increase in transmission bandwidth. A second proposal calls for the transmission of a second augmentation channel to afford full HDTV fidelity by providing a 1,125-line, 60-Hz, 2:1 interlace, 16:9 aspect ratio signal at the output of the receiver.

Acknowledgments

The authors gratefully acknowledge the efforts of T. Lin, whose work on the COMSAT Programmable Image Processing Facility made this study possible.

References

[1] B. D. Loughlin, "Monochrome and Color Visual Information Transmission," *Television Engineering Handbook*, K. B. Benson, ed., New York: McGraw-Hill, 1986, pp. 4.1–4.88.



(a) Component (Y, R-Y, B-Y) Color Bar Source

(b) Conventional NTSC-Encoded/Decoded Color Bar

(c) Adaptive Three-Dimensional Comb Filter Encoded/Decoded Color Bar

Figure 8. *Component Color Bar*

[2] E. L. Hall, *Computer Image Processing and Recognition*, New York: Academic Press, 1979, pp. 305–360.

[3] Y. C. Faroudja, "NTSC and Beyond," *IEEE Transactions on Consumer Electronics*, Vol. 34, No. 1, February 1988, pp. 166–178.

[4] D. W. Power, "Computer Simulation of a Two-Dimensional Pre-Comb Filter Technique for Composite Transmissions," COMSAT Laboratories Technical Note CTD-87-221, January 27, 1987. COMSAT Data Catalog No. 87DC049.

[5] D. W. Power, "A Motion-Adaptive 3-Dimensional Comb Filter for Transmission of HDTV With a Compatible NTSC Channel," COMSAT Laboratories Technical Note CTD-87-126, August 5, 1987. COMSAT Data Catalog No. 87DC085.

[6] D. W. Power, "Computer Simulation of a Motion Adaptive Conditional Replenishment Algorithm for Transmission of Component Video," COMSAT Laboratories Technical Note CTD-87-132, August 25, 1987. COMSAT Data Catalog No. 87DC114.

Appendix. COMSAT Laboratories Programmable Image Processing Facility

A programmable image processing facility has been developed at COMSAT Laboratories to support research and development on HDTV, EDTV, and broadcast-quality television. This facility can substantially reduce the time and resources required to evaluate the effectiveness of various video processing techniques and algorithms and to optimize design parameters.

The Programmable Image Processing Facility can operate in three basic modes. First, it can capture up to 100 seconds of real-time component or composite video from video sources such as television cameras, videotape recorders, and film scanners. The video sequence is digitized and stored in digital form. Second, it can address each individual sample stored in the disks and process it according to the algorithm to be evaluated. The processed sample is again stored in a separate sequence. Third, upon completion of the processing, the new sequence can be displayed at real-time speed for subjective or objective evaluation.

Architecture

Figure A-1 is a functional block diagram of the Programmable Image Processing Facility. The facility consists of a Digital Equipment Corporation (DEC) VAX 11/750 computer and three IBIS 1400 disks, with in-house-developed hardware for control and interfacing. The selection of the DEC VAX 11/750 was based on the following considerations:

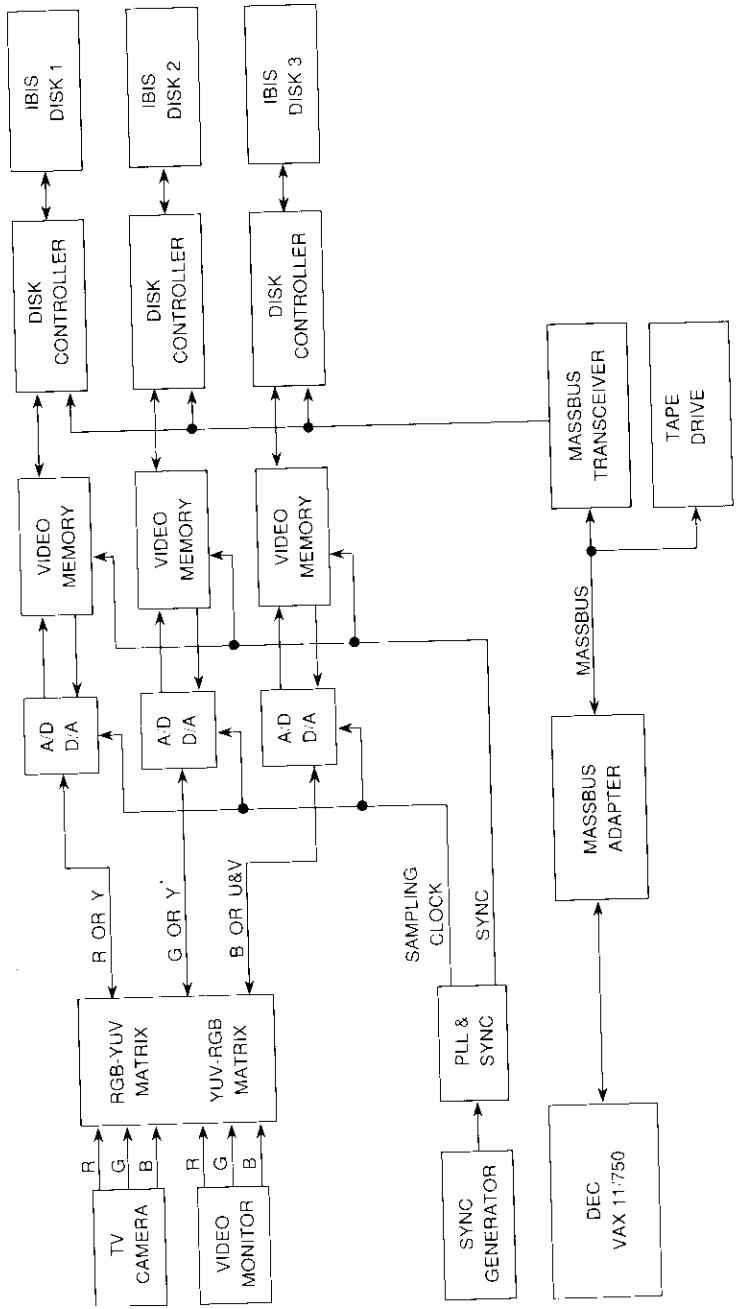


Figure A-1. *Functional Block Diagram of the Programmable Image Processing Facility*

a. Software compatibility with previous simulation software developed in the COMSAT Image Processing Laboratory on a DEC PDP 11/44 computer.
b. Capability of handling a high-speed tape drive backup for mass storage.
c. Availability of a large pool of VAX-based system and applications software.

The IBIS 1400 disks were selected because of their high storage capacity, and even more importantly their high data transfer rate. Each disk has 1,400 Mbytes of unformatted capacity, or 1,219 Mbytes of formatted storage. The burst transfer rate is 20 Mbyte/s, and the continuous transfer rate is about 10 Mbyte/s. The random data error rate is kept under 10^{-10} by a built-in error correction codec.

Because the three IBIS disks are interfaced to the VAX 11/750 through the DEC Massbus, the computer considers the disks as part of its mass storage peripheral. The Massbus has a transfer rate of 2 Mbyte/s; therefore, it will not be a bottleneck during data processing or data transfer operations such as disk loading or backup.

The in-house-developed hardware includes video interfaces, video memories, and disk controllers. The equipment is housed in a small 19-in. rack and can generally be divided into three separate channels with a common video interface.

The facility generally accepts video in the red, green, blue, and sync (RGBS) component format. A crystal-based phase-locked loop (PLL) is used to synchronize the internal time base with the external synchronization. If it is desired to store the video in the luminance (Y) and color difference (R-Y and B-Y) format, an optional RGB-to-(Y, R-Y, B-Y) matrix and its inverse are available, unless the signal is already in the (Y, R-Y, B-Y) format.

Each channel consists of a disk controller, a video memory, an analog-to-digital (A/D) converter, and a digital-to-analog (D/A) converter. The disk controller converts the commands from the VAX to instructions that are acceptable by the IBIS disk and returns the status of the disk to the VAX operating system. It also provides temporary storage for data transfer operations between the VAX and the IBIS disk. During sequence capture, the video memory serves as temporary storage for the continuous video input before it is loaded to the disk at a fixed burst rate. Similarly, during sequence display, the video memory is used as a rate-conversion buffer to convert the fixed-burst-rate disk output to continuous video. The memory also compensates for any discrepancy in response time among the three IBIS disks. Thus, synchronization among the three video signals is ensured.

Capabilities and operating modes

The facility is capable of accepting broadcast-quality television, EDTV, or HDTV signals in either RGB or (Y, R-Y, B-Y) format. Composite NTSC format can also be accepted by using only one of the video channels. This flexibility is achieved by inherent programmability in the hardware.

Two built-in voltage-controlled crystal oscillators can be selected to generate the basic sampling frequency and internal time base. The first group generates frequencies at 3, 4, 6, 8, and 12 times the NTSC color subcarrier frequency. The second group generates frequencies at 1, 2, 3, and 4 times 13.5 MHz. The 13.5 MHz frequency is often considered as the standard sampling frequency for digital television. Additional sampling rates can be arranged by using a different crystal oscillator. By selecting the proper horizontal and vertical counts, rasters of all television formats of interest can be generated. The highest sampling frequency of the system is 54 MHz, which is limited by the speed of the A/D converter and control logic.

The average data transfer rate of the IBIS disks is half of the 20-Mbyte/s burst data rate. In reality, the continuous throughput of the disks is further limited by the average track-to-track access time of 2.5 ms, and a maximum of 9.2-Mbyte/s is achieved. This maximum speed is lower than the desirable sampling rate of the video signals of interest. To circumvent this limitation, only a portion of the video raster (as shown in Figure A-2) is stored in the disks. The position and size of the window can be defined by the computer prior to sequence capture. Four parameters are required. Two values define the window location: vertical window start and horizontal window start. The

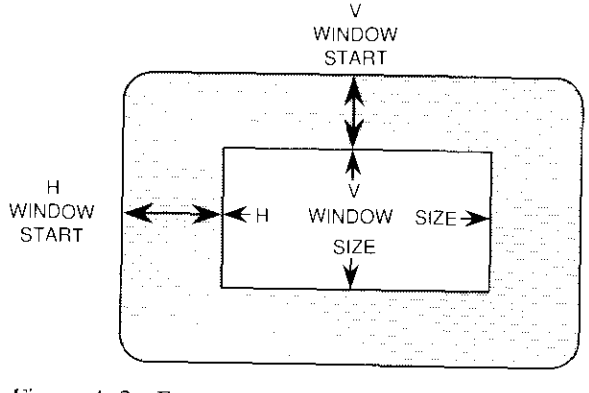


Figure A-2. *Four Parameters Defining the Size and Position of the Picture Area to be Stored*

other two values define the vertical window size and horizontal window size. Thus, any signal with any sampling rate acceptable to the video memory can be store in RGB form in the three disks. The differences lie in the size of the window.

For a normal broadcast signal with sampling frequencies of 13.5 or 14.3 MHz, the window is only slightly smaller than that of normal active video. However, for signals sampled at higher frequencies, the window can be considerably smaller than the normal screen. Therefore, a 2:1 multiplexer/demultiplexer is incorporated in the video channels to demultiplex the luminance signal into two channels, multiplex the two chrominance signals into one channel, and store them to the three IBIS disks. This approach effectively doubles the window area for high-resolution signals stored in (Y, R-Y, B-Y) format.

Each video memory is based on 128 256K dynamic random-access memory (DRAM) chips. These DRAMs are organized into eight-chip, 256-kbyte, single in-line packages (SIPs) to conserve board space. Sixteen SIPs are multiplexed to reduce the effective cycle time for read or write from 250 to 15.6 ns. The effective read or write operation can be executed at a maximum of 64 MHz. However, since the IBIS disk has a 20-MHz burst rate asynchronous to the memory cycle, the continuous throughput of the video memory during read and write operation is expected to be less than the maximum allowed for synchronous operation. The system has been successfully tested at 27 MHz for each channel. This is sufficient for HDTV operation at a 54-MHz sampling rate using the 2:1 multiplexing scheme described previously.

Each video memory has 4.19 Mbytes of capacity, which is more than sufficient to store 24 fields of video at 166 kbytes per field for each channel. Therefore, it is possible to store a short motion scene in the video memory for continuous display. This important feature quickly eliminates most of the problematic processing algorithms.

The video memory also supports full-screen display if desired. The number of fields stored in this mode will depend on the sampling frequency.

Image processing software development is a continuing part of COMSAT's research and development effort. The following is a partial list of the existing software library:

- Color bar generator
- Differential pulse-code modulation (DPCM)
- Vector DPCM and quantization
- Adaptive vector DPCM and quantization
- Two-dimensional discrete cosine transform

- Adaptive discrete cosine transform
- Multiplexed analog component processing
- Line and field comb filters
- Component-to-NTSC conversion
- Two-dimensional FIR filters
- Motion detection and analysis
- Time-multiplexed analog television

Six types of commands are available. Commands to support real-time storage to the IBIS disk, real-time display from the disk, and data transfer from disk to VAX and from VAX to disk are necessary for the basic operation of the facility. To support operations involving only the video memory, commands for real-time storage to the video memory and real-time display from the video memory are also implemented.

Donald W. Power is a Senior Technical Specialist in the Image Processing Department at COMSAT Laboratories. His work involves the design, development, and testing of advanced video systems. Prior to joining COMSAT in 1983, he specialized in the design of robotic and computer-aided manufacturing systems at Westinghouse Electric Corporation's Research and Development Center. He is currently studying for a B.S.E.E.

Lin-Nan Lee received a B.S.E.E. degree 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 processing (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.

Index: communication satellites; intermodulation, passive; models; stochastic processes

Statistical determination of test carrier levels for passive intermodulation testing

M. H. Abdelaziz and A. E. Atia

(Manuscript received November 24, 1987)

Abstract

When a single antenna is used to perform both the transmit and receive functions in multicarrier communications satellite systems, passive intermodulation products can be a serious problem if they fall within the receive band. Vigorous testing is necessary to determine if such a problem exists and to ensure the adequacy of the system. When the operational system has a large number of carriers, it is desirable to determine a testing methodology that uses fewer carriers (*e.g.*, two), and carrier levels which provide conditions equivalent to the worst-case operating conditions. Three such methodologies are described, including a new rigorous approach based on statistical analyses. Theoretical and numerical results are presented which demonstrate how the new method can be used to determine carrier test levels that more closely approximate realistic test conditions than do existing methods such as the total equivalent power method and the peak field (or voltage) method.

Introduction

Many microwave communications systems, especially those involving satellites, use a common antenna aperture for both the receive and transmit functions. For multicarrier operation, it is known that passive intermodulation products (PIMs) can be generated in a number of ways within components of the system that contain contacts of similar or dissimilar metals [1]–[4]. Frequency plans for communications systems are normally devised to prevent intermodulation products of the transmitted signals from falling within the receiving bands. However, this approach often wastes valuable frequency

spectrum. Thus, system designers often must compromise in frequency planning by allowing some higher-order intermodulation products of transmit-band carriers to fall within some of the receive-band channels. In such cases, it is necessary to ensure that the levels of any possible PIMs that could potentially fall within the receive bands are sufficiently low to avoid serious degradation of the communications system performance.

Vigorous system testing is usually necessary to determine whether the levels of undesirable intermodulation products falling within the receive frequency bands are acceptable. Numerous experimental studies have been performed to determine the nature and levels of PIMs generated by the nonlinear effects caused by contacting faces between similar and dissimilar metals under a variety of conditions [5],[6]. However, when the operational system has a large number of carriers, system testing under the same operating conditions is too complicated and expensive in terms of both the instrumentation and time required for tests, and the subsequent interpretation of results. In such cases, it is desirable to determine a testing methodology which uses fewer carriers (*e.g.*, two) than the operational system, and which yields results comparable to those for actual worst-case operating conditions.

This paper first describes the two most common methods for determining test carrier levels in order to measure PIMs: the total equivalent power method and the peak field (or voltage) method. The total equivalent power method uses a lower bound on the testing level to ensure that the test power does not exceed the total average power of the carriers. The peak field method, on the other hand, determines the test carrier power based on an upper bound of the maximum field strength that could exist in the system under operating conditions. These two methods yield testing levels that are either too optimistic (low, using the first method) or too pessimistic (high, using the second method).

A new methodology is then introduced which is based on the statistical properties of the signals encountered. The test carrier levels are determined using this approach, and the fact that the PIM is a threshold phenomenon. First, a threshold is determined according to a specified probability that the testing field levels exceed a certain value. The test carrier level is then determined based on that threshold. Consequently, the test levels obtained by this new approach lie between the levels obtained with the other two methods. As a special case, results obtained using this method reduce to the peak field method if the threshold value is chosen to be the maximum field.

The theoretical results presented show that the proposed new methodology could be used to provide more realistic and reliable testing. The scope of work under which this investigation was performed did not include experimental verification of the derived results.

Problem definition

Figure 1 is a simplified block diagram of a typical single-conversion channelized communications satellite transponder. The transponder consists of the combined transmit/receive antenna, the receiver, the input multiplexer, the high-power traveling wave tube amplifiers (TWTAs), the output multiplexer, and a diplexing network that separates the transmit and receive signals. The antenna has a common reflector and a feed system that consists of several feed elements connected to a beam-forming network (BFN), which provides the appropriate amplitude and phase weightings required at each feed to form the transmit and receive shaped beams. The diplexer allows the receive signal in the receive frequency band to be passed to the receiver from the output of the BFN, while simultaneously passing the transmit signals in the transmit frequency band from the output multiplexer's common port only to the antenna's BFN. The receive system amplifies the received signals in a low-noise amplifier (LNA), and then uses a mixer to down-convert (translate) the receive band to the transmit frequency band. The input multiplexer divides the receive frequency band into several narrowband channels, each of which is amplified by a separate power amplifier (usually a TWTA). In such a channelized system, the TWTAs may be driven close to saturation to achieve higher DC-to-RF efficiency, while avoiding the generation of any intermodulation products among the channelized bands due to the nonlinearities of the power amplifiers. The output multiplexer recombines the channels so that they can be fed to, and radiated from, a single antenna.

Although channelization is introduced to avoid the generation of intermodulation products among the channels due to the active devices in the transponder, the existence of high power in the circuits of the output multiplexer, the diplexer, and the BFN, and on the surface of the antenna reflector (all of which are passive devices) could produce PIMs due to the presence of small nonlinearities, such as the contacts between waveguide flanges or other contacts between similar or dissimilar metals. If some of these products have frequencies within the receive frequency band, the diplexer will permit their passage to the LNA. The levels of these products may well be above the levels of the desired received signals, and could even saturate the receiver.

To appreciate the problems involved in this type of transponder, consider the example given in Figure 2 which shows a typical frequency plan for a dual-polarized, multibeam, K_u-band communications satellite transponder. A simple calculation will show that many 3rd-, 5th-, 9th-, and 25th-order intermodulation products of several of the transmit channels could exist in some of the receive channels. For example, the third-order product of transmit

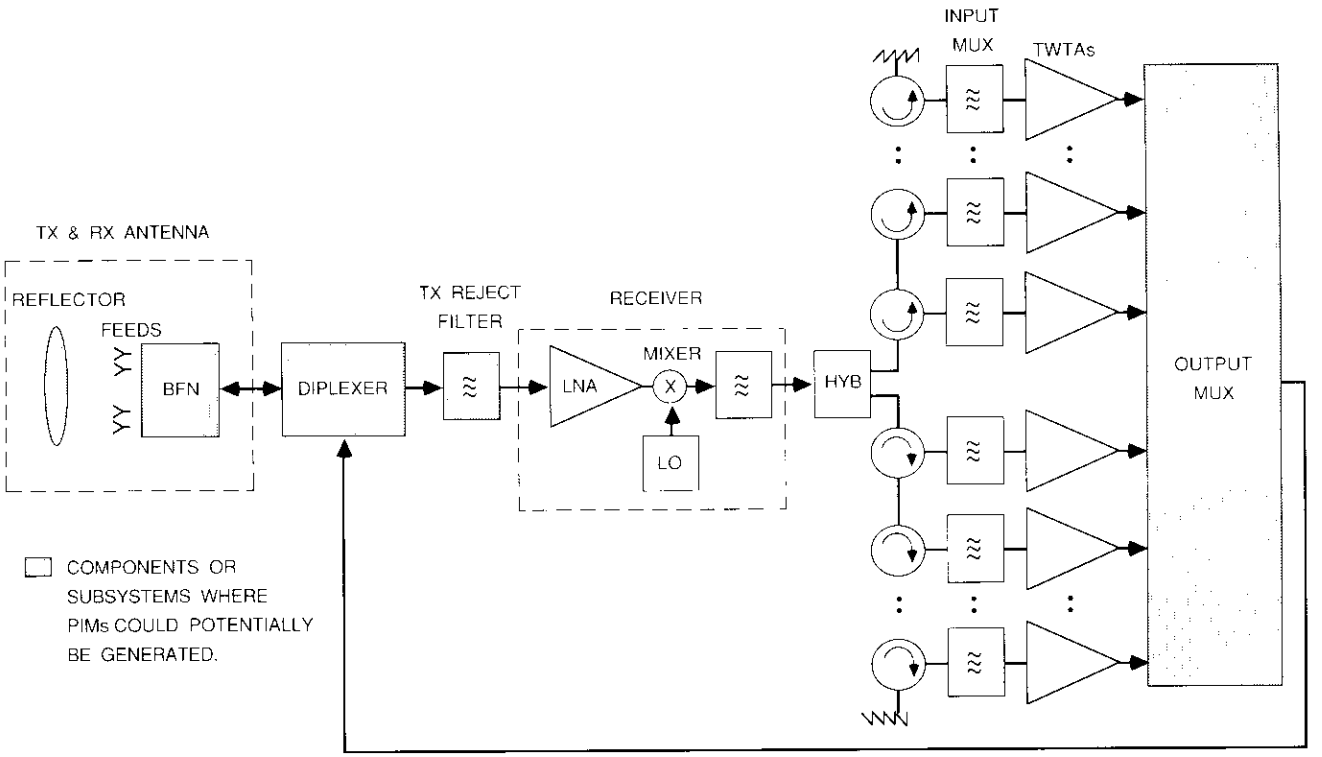


Figure 1. *Simplified Block Diagram of a Typical Communications Satellite Transponder*

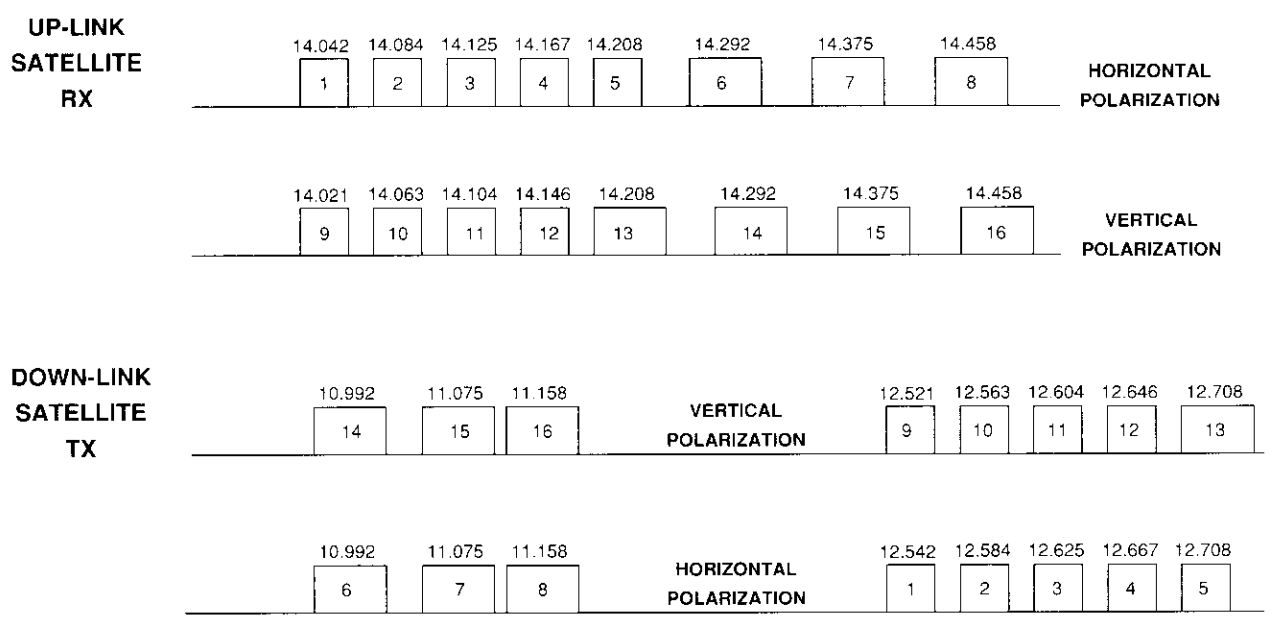


Figure 2. *Typical Frequency Plan of Dual-Polarized Multiple-Beam Satellite Transponders*

channels 10 and 14 ($2f_{10} - f_{14}$) would lie within receive channel 12; the third-order product of transmit channels 12 and 14 ($2f_{12} - f_{14}$) would lie within receive channel 14, and so forth. These intermodulation products would be generated in the passive portion of the transponder, since the carriers generating them are amplified in different amplifiers. The levels of these potentially harmful products can be relatively high, and could be as large or larger than the intended receive signals. This is because the power levels of the carriers near TWTA saturation can be on the order of tens of watts, while receiver sensitivity is in the picowatt range. Transponder gains of about 130 to 160 dB are not uncommon. Thus, PIM levels 160 to 180 dB below carrier levels can be quite damaging and can swamp the desired signals.

To ensure adequate PIM levels in satellite transponders under operating conditions, acceptable levels must first be defined. The levels are usually set at least several decibels below the level of thermal noise at the input to the receiver. Once all the transponder nominal parameters and link parameters are known, it is a simple matter to establish a reasonably safe allowable PIM level. A method must then be prescribed for testing the completed transponder to determine if the PIM levels are within the established limits.

Ideally, PIMs should be tested under worst-case operating conditions (for example, full saturated output power from all TWTAs that operate simultaneously). However, this process can be very complicated and costly in terms of the instrumentation required, the testing time consumed, and the ability to determine the worst operating conditions for PIM generation. Thus, it is highly desirable to develop a testing methodology that does not have the complications of the actual operating conditions, but yields high confidence that the measured PIMs are indeed representative of the worst operating conditions. This paper describes such a testing methodology and compares various commonly used methodologies.

Modeling the test

To analyze the problem mathematically, a model is defined that represents the physical phenomena leading to PIM generation. Consider a passive system, S, with small nonlinearities such as those encountered at the interfaces between switches, waveguides, and diplexers due to such factors as dissimilar materials or small unavoidable amounts of oxidation. This system has N input carriers, as shown in Figure 3a. Each input carrier is sinusoidal, with average power P_i. The instantaneous electric fields (or voltages) at any specific point within the system due to P_i can be represented by

$$E_i(t) = C_i\sqrt{P_i}\ \sin\ (\omega_i t + \Phi_i) \qquad (1)$$

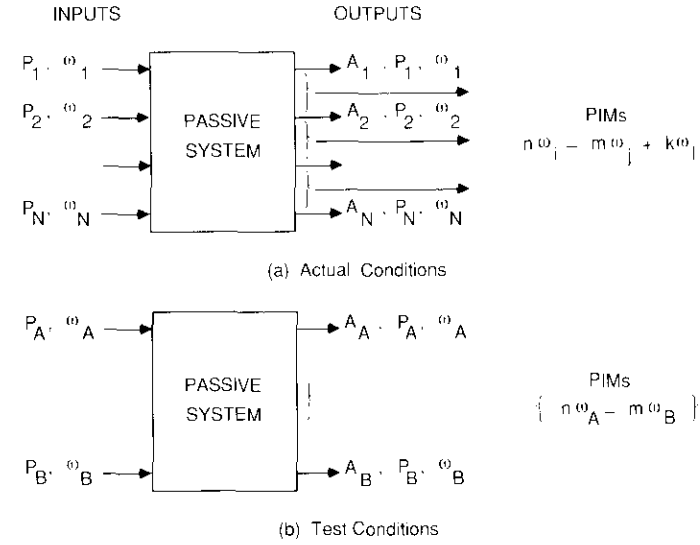


Figure 3. *Passive System With Small Nonlinearities*

where the C_i are constants, and ω_i and Φ_i are the angular frequency and phase, respectively, of the fields (or voltages).

The total field (or voltage) at this point produced by all carriers can be represented by

$$E(t) = \sum_{i=1}^{N} E_i(t) \quad . \qquad (2)$$

The output of the system will consist of the sum of the carriers attenuated and phase-shifted by certain amounts, plus a combination of intermodulation products (*i.e.*, signals with angular frequencies that are combinations of the integer sums and differences of the angular frequencies). These intermodulation products are generated due to small nonlinearities in the system.

As mentioned previously, testing the system to determine the PIM levels under actual operating conditions can be very complicated and costly. Figure 3b shows a less complicated test condition. In this case, the input to the system is chosen to be only two carriers of average powers P_A and P_B, and angular frequencies ω_A and ω_B. The test measures PIM levels ($n\omega_A - m\omega_B$). The values of P_A and P_B must be determined such that the measured intermodulation levels in the test case provide an upper bound on the level of the products under actual operating conditions, with a high degree of confidence.

Three statistical methods that can be used to define power levels P_A and P_B are discussed in the following sections. The first two methods—the total equivalent power method and the peak field (or voltage) method—are based on classical approaches that consider either the average carrier powers or the peak fields of the carriers to be independent random variables with different means and standard deviations. The third method is a novel approach that considers the total field, $E(t)$ (not the average power of the peak value) resulting from the superposition of individual fields $E_i(t)$ to be a stochastic process with a sample realization dependent on the phase of each individual carrier. This allows the selection of test carrier levels which are somewhere between the levels that result from the first two methods, thus yielding more realistic test conditions.

Total equivalent power method

In this method, the average powers of the carriers (P_i) are assumed to be independent random variables, normally distributed with mean $\overline{P}_i$ and standard deviation σ_i. Consequently, the average powers in the carriers (P_A and P_B) are taken to be equal, and each equals one-half of the total mean power of all the carriers; that is,

$$P_A = P_B = \frac{1}{2}\sum_{i}^{N} \overline{P}_i \quad . \tag{3}$$

The standard deviation, σ_t, of the total power would be given by

$$\sigma_t = \sqrt{\sum_{i=1}^{N} \sigma_i^2} \quad . \tag{4}$$

This choice corresponds to the most commonly used test conditions and assumes that the mechanism of PIM generation depends solely on the total average power in the system.

Peak field (or voltage) method

This method assumes that the mechanism producing the PIMs is principally related to the peak value of the fields in the system. The peak fields of the carriers ($C_i \cdot \sqrt{\overline{P}_i}$) are again assumed to be independent random variables normally distributed with mean values of ($C_i \cdot \sqrt{\overline{P}_i}$) and standard deviation $C_i\sqrt{\sigma_i}$. The test carrier peak fields, E_A and E_B, are taken to be equal, with each equaling one-half of the total average peak field of all carriers; that is,

$$E_A = E_B = \frac{1}{2}\sum_{i=1}^{N} C_i \sqrt{\overline{P}_i} \quad . \tag{5}$$

It is commonly assumed that all values of C_i are equal to a single constant, C. The average powers of each of the two carriers in this case would be given by

$$P_A = P_B = \frac{1}{C^2}E_A^2 = \frac{1}{C^2}E_B^2 = \frac{1}{4}\left(\sum_{i=1}^{N}\sqrt{\overline{P}_i}\right)^2 \quad . \tag{6}$$

Table 1 presents a comparison between the level of each carrier obtained using the total equivalent power method and those obtained with the peak field method. It is assumed that each of the N carriers has equal power (P_o), and the level of the carriers (P_A) is compared for various values of N. It can be seen that the two methods are equivalent for $N = 2$, as expected; however, for larger values of N, the peak field method provides higher testing levels, indicating its more stringent and conservative requirements. For example, for four carriers the peak field method requires testing levels that are 3 dB higher than with the total equivalent power method.

TABLE 1. COMPARISON OF CARRIER POWER TEST LEVELS

	Power Level of Each Test Carrier in a Two-Carrier Procedure for PIM	
Number of Carriers, N	Total Equivalent Power Method	Peak Field Method
2	P_o	P_o
3	1.5 P_o	2.25 P_o
4	2 P_o	4 P_o
5	2.5 P_o	6.25 P_o
6	3 P_o	9 P_o
7	3.5 P_o	12.25 P_o
8	4 P_o	16 P_o

Field-time threshold method

Since PIM generation often occurs when signals encounter two dissimilar materials in contact, it is treated as a threshold phenomenon [7]. The phenomenon happens whenever the instantaneous fields (or voltages) in the system exceed certain threshold values for a certain percentage of time.

Mathematically, the field-time threshold method treats the instantaneous fields of each carrier, $E_i(t)$, given by equation (1), as stochastic processes.

The parameters of each field (P_i and ω_i) are assumed to be deterministic constants, and the phases Φ_i are taken as independent random variables uniformly distributed over $(0, 2\pi)$. Thus, the total field, $E(t)$, in equation (2) is also, mathematically, a stochastic process. The statistics of this process must be determined and used to establish the level of testing by using two equal carriers.

Consider a sample function $E(t)$ from the sample space. This function is sketched *vs* time in Figure 4. Over the time interval T, $|E(t)|$ exceeds a threshold value E_o for the intervals $t_1, t_2, \ldots$, as shown. The threshold ratio $\rho(E_o)$ is defined as

$$\rho(E_o) = \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{N(T, E_o)} t_i \tag{7}$$

where t_i are the time intervals over which $|E(t)|$ exceeds E_o, and $N(T, E_o)$ is the number of these time intervals.

The equivalent testing field value, E_A, is chosen so that

$$\Pr\{\rho(E_A) \leq \rho_o\} = p_o \tag{8}$$

where ρ_o is a chosen ratio (less than unity) and p_o is a confidence probability (usually $0.99 \leq p_o < 1$).

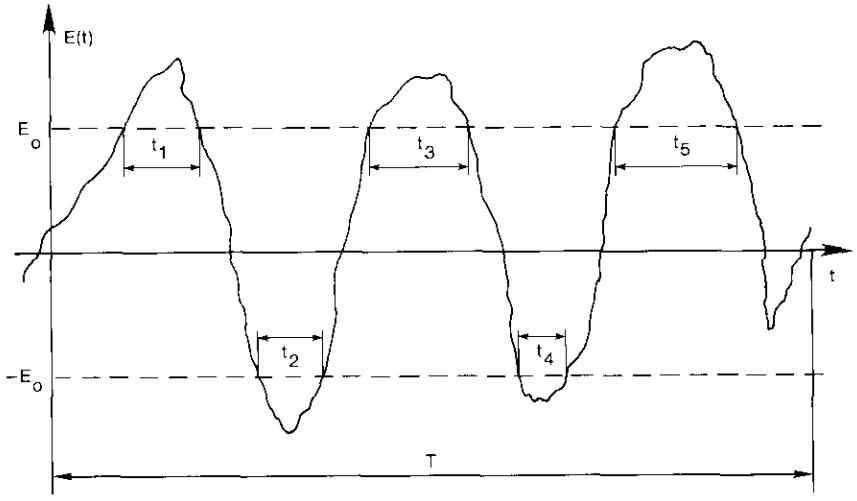


Figure 4. *Sample Function of E(t) Over a Time Interval, T*

Clearly, $\rho(E_o)$ is a random variable whose statistics are determined from the statistics of $E(t)$. Although in principle the derivations of these statistics are relatively straightforward, the numerical computations to evaluate the necessary quantities are quite involved. The following subsection outlines the analytical process involved in formulating the problem, and presents both formal and numerical results. A possible alternative approach to the numerical computation of the statistics is by pure simulation.

Determining the statistics of *E(t)*

The statistics of $E(t)$ are most easily analyzed in terms of the characteristic function of the sum of the random variables $E_i(t)$. The characteristic function is defined as

$$\Phi(\nu, t) = E\{e^{j\nu E(t)}\} \tag{9}$$

where $E\{\cdot\}$ is the expected value of the random variable in the parentheses. Using equation (1), and assuming that the phases Φ_i are independent random variables uniformly distributed over $(0, 2\pi)$, the characteristic function can be computed as follows:

$$\begin{aligned} \Phi(\nu) &= E\{e^{j\nu \sum_{i=1}^{N} A_i \sin(\omega_i t + \Phi_i)}\} \\ &= \prod_{i=1}^{N} E\{e^{j\nu A_i \sin(\omega_i t + \Phi_i)}\} \end{aligned} \tag{10}$$

where $A_i = C\sqrt{P_i}$. But

$$\begin{aligned} E\{e^{j\nu A_i \sin(\omega_i t + \Phi_i)}\} &= \frac{1}{2\pi} \int_o^{2\pi} e^{j\nu A_i \sin(\omega_i t + \alpha)}\, d\alpha \\ &= J_o(\nu A_i) \end{aligned} \tag{11}$$

where $J_o(\cdot)$ is the Bessel function of the first kind and zero order. Thus,

$$\Phi(\nu) = \prod_{i=1}^{N} J_o(\nu A_i) \quad . \tag{12}$$

The probability density function of the random variable $E(t)$ is then the inverse Fourier transform of equation (12); that is,

$$f_E(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\nu x} \Phi(\nu) d\nu \quad . \tag{13}$$

Determination of the testing threshold involves the evaluation of the probability that $|E(t)| \leq E_o$. This can be obtained by defining $P_N(\alpha)$ as

$$P_N(\alpha) \triangleq \Pr\{|E(t)| \leq \alpha\} = \int_{-\alpha}^{\alpha} f_E(y)\, dy \quad .$$

Thus,

$$\begin{aligned} \Pr\{|E(t)| \leq E_o\} &= P_N(E_o) = \int_{-E_o}^{E_o} f_E(y) \cdot dy \\ &= \frac{1}{\pi} \int_{\epsilon}^{x} \frac{\sin(\nu E_o)}{\nu} \prod_{i=1}^{N} J_o(\nu A_i)\, d\nu \qquad \text{for } E_o > 0 \\ &= 0 \qquad \text{for } E_o < 0 \quad . \end{aligned} \tag{14}$$

At this point, it can be observed that the process $E(t)$ is a stationary process, hence the characteristic function Φ, the probability density $f_E(\nu)$, and the threshold probability, given in equations (12), (13), and (14), respectively, are all independent of t. The only relevant parameters of the process are the number of different frequencies, N, and their amplitudes, A_i.

The threshold probability can be calculated numerically by evaluating the integral in equation (14). If all values of A_i are assumed to be equal to A_o, then the probabilities will be a function of N and A_o only. The value of this probability will be zero for $E_o < 0$, and 1 for $E_o > E_{\max}$, where

$$E_{\max} = C \sum_{i=1}^{N} \sqrt{P_i} \quad .$$

Numerical evaluation of the probability distribution function is given in the next section.

To determine the testing level, a threshold probability, P_o, must first be defined. The corresponding value of $(E_o/E_{\max})$ is then determined from the curve of $\Pr\{|E(t)| \leq E_o\}$. The testing power level is determined as (E_o^2/C^2). It is important to note that this level will generally depend on the choice of P_o. For practical purposes, this level will be less than the level obtained with the peak field method, but greater than that determined from the equivalent total power method.

Numerical results

To find the testing level for any number of carriers and with any required threshold probability, the probability function defined in equation (14) must

be numerically evaluated for different values of the number of carriers, N. In the following analysis, it is assumed that all individual carriers have the same levels; that is,

$$A_i = A_o \qquad i = 1, 2, \ldots, N \quad . \tag{15}$$

Equation (14) then becomes

$$P_N(x) = \frac{2}{\pi} \int_{o}^{x} \frac{\sin(\nu N x)}{\nu} |J_o(\nu)|^N\, d\nu \tag{16}$$

where

$$x = \frac{E_o}{NA_o} \quad . \tag{17}$$

The integral in equation (16) can be evaluated analytically for $N = 1$. The result is

$$\begin{aligned} P_1(x) &= 0 \qquad & x < 0 \\ &= \frac{2}{\pi} \sin^{-1} x \qquad & 0 \leq x \leq 1 \\ &= 1 \qquad & x > 1 \quad . \end{aligned} \tag{18}$$

For $N > 1$, the integral has been evaluated numerically, and the results are shown in Figure 5a for $N = 1$ to 20. The same results are shown in Figure 5b, but enlarged for the values of $P_N(x)$ which lie between 0.8 and 1.

To ensure accuracy, the numerical evaluation of the integrals in equation (16) is carried out as follows. The integration interval $o \leq \nu < \infty$ is divided into three sub-intervals: $o \leq \nu \leq \epsilon$; $\epsilon < \nu \leq Z$; and $Z < \nu < \infty$. Thus,

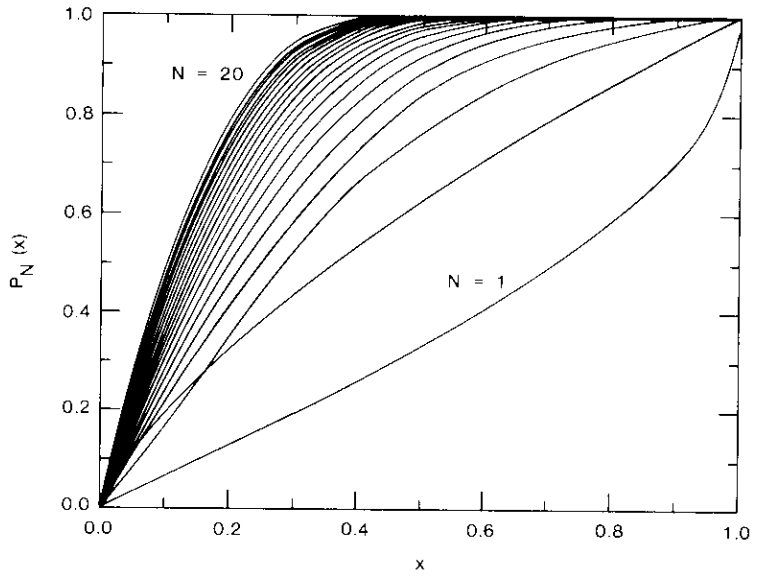
$$P_N(x) = \frac{2}{\pi} I_1 + \frac{2}{\pi} I_2 + \frac{2}{\pi} I_3 \tag{19}$$

where

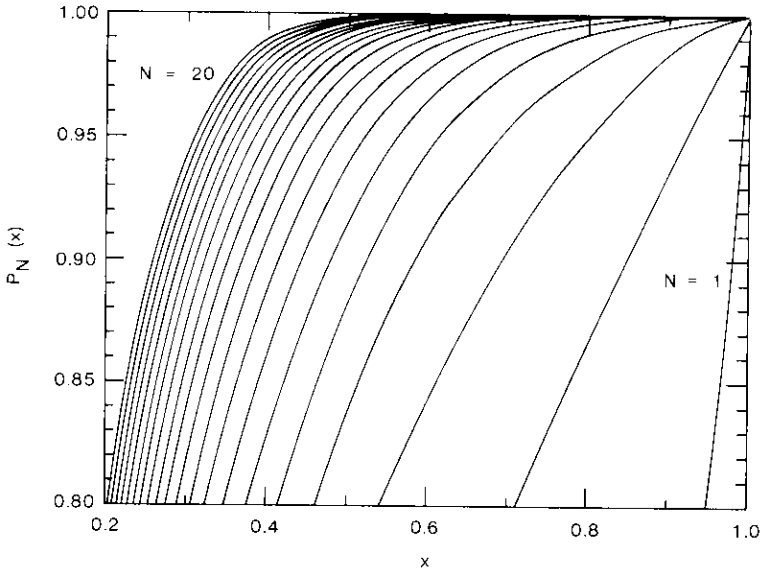
$$I_1 = \int_{o}^{\epsilon} \frac{\sin(\nu N x)}{\nu} [J_o(\nu)]^N\, d\nu \tag{20}$$

$$I_2 = \int_{\epsilon}^{Z} \frac{\sin(\nu N x)}{\nu} [J_o(\nu)]^N\, d\nu \tag{21}$$

$$I_3 = \int_{Z}^{\infty} \frac{\sin(\nu N x)}{\nu} [J_o(\nu)]^N\, d\nu \quad . \tag{22}$$



(a) For N = 1, 2, ..., 20



(b) For N = 1, 2, ..., 20 ; 0.8 ≤ $P_N(x)$ ≤ 1.0

Figure 5. *Evaluation of* $P_N(x)$

Integrals I_1 and I_3 are evaluated by expanding their integrals and finding bounds on theirs values as functions of ϵ and Z. Integral I_2 is evaluated using standard numerical integration subroutines [8].

For I_1, a Taylor series expansion of both the sine function and the Bessel function are used, as

$$I_1 = \int_o^{\epsilon} \frac{\sin(\nu Nx)}{\nu} [J_o(\nu)]^N \, d\nu$$

$$\simeq Nx \int_o^{\epsilon} \left(1 - N\frac{\nu^2}{4} + \ldots\right) d\nu \qquad (23)$$

$$= Nx\,\epsilon \left(1 - \frac{N\epsilon^2}{12} + \ldots\right) .$$

For the error in the evaluation to be less than δ, ϵ must be chosen such that $\epsilon < (\delta/Nx)$.

For I_3, a bound can be found by using the inequality $|\sin \nu Nx| \leq 1$ and replacing the Bessel function with its asymptotic expansion [9], as

$$|I_3| \leq \int_Z^{\infty} \left(\frac{2}{\pi}\right)^{N/2} \frac{d\nu}{\nu^{N/2+1}} = \left(\frac{2}{\pi}\right)^{N/2} \left(\frac{2}{N}\frac{1}{Z^{N/2}}\right) . \qquad (24)$$

For this remainder integral to be less than an allowable error Δ, Z must be chosen such that

$$Z > \frac{2}{\pi}\left(\frac{2}{N\Delta}\right)^{2/N} .$$

Conclusions

Three testing methodologies have been presented for predicting PIMs in a multicarrier system. These methodologies are the total equivalent power method, the peak field method, and the newly developed field-time threshold method. The three methods give equivalent power levels for two equal carriers to be used in testing, and yield PIMs that, in a sense, bound the operational case. For N equal carriers of power P_o each, the three methods yield two equal test carrier powers of

$$\left(\frac{N}{2}\right)P_o, \left(\frac{N}{2}\right)^2 P_o, \text{ and } \left(\frac{Nx}{2}\right)^2 P_o \quad .$$

In the field-time threshold method, x is determined from a given probability that the field of the operational carriers does not exceed the level of the test carrier fields. Graphs were presented that give the distribution function which can be used to determine x for any probability value and for any number of carriers, N, between 2 and 20. Compared to the overly optimistic test carrier levels of $\left(\frac{N}{2}\right)P_o$ and the overly pessimistic level of $\left(\frac{N}{2}\right)^2 P_o$, this new method yields a test level $\left(\frac{Nx}{2}\right)^2 P_o$ that is user-controlled. Depending on the level of confidence required in the test, a user can specify a certain probability, $P_N(x)$, from which x (and consequently the power level of the test carriers) can be determined. In the special case where $P_N(x)$ is chosen to be 1, the value of x is 1 and the test carrier powers reduce to $\left(\frac{N}{2}\right)^2 P_o$, the same as with the peak field method.

The statistical approach presented closely models the physical phenomena involved in PIM generation. Thus, it is expected that the testing levels predicted by this method will provide tight bounds for the required carrier test power.

Acknowledgments

The authors wish to acknowledge the helpful comments, suggestions, and discussions provided by G. Hyde of COMSAT Laboratories.

References

[1] W. Low, "An Investigation of the Production or Combination of Frequencies When Waveguide and Coaxial Components are Used With More Than One Transmission Frequency," *Frequenz*, 1963, Vol. 17, p. 94.

[2] J. A. Betts and D. R. Ebenezer, "Intermodulation Interference in Mobile Multiple-Transmission Communications Systems at High Frequencies," *Proc. IEE*, Vol. 120, No. 11, November 1973, pp. 1337–1344.

[3] E. Nuding, "Nonlinearities of Flange Connections in Transmission Lines Carrying High RF Power," 4th European Microwave Conference, Montreux, Switzerland, September 1974, *Proc.*, pp. 613–618.

[4] R. C. Chapman et al., "Hidden Threat: Multicarrier Passive Component IM Generation," 6th AIAA/CASI Communications Satellite Systems Conference, Montreal, April 1976. Published in *Satellite Communications: Advanced Technologies*, D. Jarett, ed., Vol. 55: Progress in Astronautics and Aeronautics, New York: AIAA, 1977, pp. 257–372.

[5] M. Bayrak and F. A. Benson, "Intermodulation Products From Nonlinearities in Transmission Lines and Connectors at Microwave Frequencies," *Proc. IEE*, Vol. 122, No. 4, April 1975, pp. 361–367.

[6] J. C. Lee "Intermodulation Measurement and Analysis of Some Conducting Materials Commonly Used in Aerospace," IEEE International Conference on Communications, Seattle, Washington, June 1980, *Conference Record*, pp. 25.6.1–25.6.6.

[7] C. Loo, "Calculation of the Suppression of Signals and Intermodulation Noise When Multiple Unequal Carriers are Amplified by a TWT," *Canadian Electrical Engineering Journal*, Vol. 2, No. 4, December 1977, pp. 29–32.

[8] A. S. Householder, *Principles of Numerical Analysis*, New York: McGraw Hill, 1969.

[9] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series No. 55, U.S. Government Printing Office, Washington, D.C. 20402, 1964.

Mohamed Abdelaziz received a B.S.E.E. and M.S.E.E. from Military Technical Academy, Cairo, Egypt, and a Ph.D. in electrical engineering from the George Washington University, Washington, D.C. He joined COMSAT in 1984, and has held the positions of Member of the Technical Staff and Senior Engineer in COMSAT Technical Services, where he was involved in the design and analysis of communications systems and networks. In particular, he was responsible for the design and analysis of the COMSAT-owned Satellite Data Network Services (SDNS), a software implementation of a design tool for future communications satellite systems, and traffic analysis for the second-generation ARABSAT system. He is currently Manager of the Network Engineering Department in COMSAT Systems Division, responsible for communications networks design, analysis, implementation, and testing under contracts with both the Government and commercial customers, as well as internal R&D projects.

Prior to joining COMSAT in 1984, Dr. Abdelaziz made original contributions to computer communications technology in the areas of protocol specifications, validation, testing, and performance evaluation. Most notably, he developed and evaluated the performance of a user-based priority protocol for use over CSMA/CD-based local area networks, which was implemented by E-Systems. Dr. Abdelaziz is a member of IEEE and Sigma Xi.

Ali E. Atia received a B.S.E.E. from Ain Shams University, Cairo, Egypt, and an M.S.E.E. and Ph.D. from the University of California at Berkeley. Prior to joining COMSAT in 1969, he held various research and teaching positions at both universities. As a Senior Scientist in the Microwave Laboratory at COMSAT Laboratories, Dr. Atia has made original contributions to satellite transponder and antenna technologies, most notably the development of dual-mode microwave filter technology, which is now the standard of the industry for satellite transponder channelization and many earth station applications. He also made significant contributions to several satellite programs, including INTELSAT IV-A, V, V-A, and VI, ARABSAT, and AUSSAT. He was responsible for the design, implementation, qualification, and testing of major subsystems in COMSAT's NASA ATS-F propagation experiment and the COMSTAR K_a-Band Beacon Experiment. He is currently Vice President of Communications Systems Engineering in the COMSAT Systems Division, responsible for all communications systems design, integration, implementation, and testing under contracts with various Government and commercial customers. Dr. Atia is a Fellow of IEEE, an Associate Fellow of AIAA, and a Member of Sigma Xi.

Index: communication satellites, models, software, system planning, economics

A model for planning satellite communications systems

J. L. Wernimont, W. L. Cook, and G. J. H. Brown

(Manuscript received December 23, 1988)

Abstract

The Communications System Planning Model (CSPM) is an interactive computer program that facilitates the planning of a satellite communications system. The need for CSPM stems from the large amount of data involved in the planning process, and from the desire to examine many system alternatives. The objective of this effort was to design and implement a package that a system planner would *want* to use.

The CSPM contains a number of analysis algorithms for evaluating the effective capacity and cost of satellite communications systems. In implementing the model, attention was focused on these algorithms, as well as on the user interface. This paper describes the software and its user interface, the data requirements and analysis algorithms of the model, and some of its specific applications.

Introduction

Extensive planning is necessary in order to establish the parameters of any new satellite communications system, as well as to effectively maintain and expand an existing system. Because of the long lead time associated with satellite procurement, system planners must look many years into the future. The planning process involves the iterative application of specific traffic requirements to alternative satellite system configurations and development scenarios. Planning also entails the evaluation of effective system capacity and cost. The Communications System Planning Model (CSPM) is a computer program that was developed to facilitate this planning process.

In using the model, the system planner defines the basic parameters of a satellite communications system, including earth stations (network nodes), ocean regions, satellite deployment data, transponder configurations, beam coverage regions, earth station equipment, and service categories. Traffic requirements, in the form of traffic matrices, represent the actual or forecast traffic for each communications service. By editing and manipulating the system and traffic data, the planner creates system configurations that are tested for feasibility or analyzed for performance (*e.g.*, effective capacity) and cost.

The software was implemented on an IBM mainframe using the Graphical Kernel System (GKS) and is accessed from a standard (unintelligent) graphics terminal. The CSPM interface incorporates several advanced features normally associated with software running on an intelligent workstation.

The most obvious contribution of the software is its systems planning capability, and another important aspect is its user interface. The user interface was a critical portion of CSPM because, regardless of the effectiveness of the underlying algorithms, they are worthless if they are not used. The interface is addressed first to give a sense of how the interactive system works, before moving on to a discussion of system requirements and the central concept of the planning software.

User interface

The user environment at COMSAT consisted of an IBM mainframe computer with standard graphics terminals. The use of a standard environment did not reduce the emphasis on the user interface, because it was recognized that the success or failure of a project relies heavily on the usability of the software.

The GKS was used as the basis for the graphical interface. This system provides graphic primitives and device input routines, but minimal support for operator interface development. With such a limited set of functions, a toolbox was needed that could provide additional features. Since no off-the-shelf software was available to provide these features for a standard terminal, the required toolbox was developed during the study. This reusable set of utilities is available for use in other software development work.

The Pascal language was chosen for implementing CSPM. The use of such standards as GKS, Pascal, and Fortran should reduce the effort required to transport CSPM to other environments in the future.

Interface design

The CSPM user interface design is based on an emerging "look and feel" in the software industry [1]. This interface style, first used by Xerox and eventually popularized by the Apple Macintosh computer, uses windows, dialogs, and scroll bars, as well as other direct manipulation features. Figure 1 is an example of a CSPM screen. New software based on Massachusetts Institute of Technology's X-Window follows a similar windowing interface style. These interfaces rely on hardware that supports bit-mapped graphics, and thus are most common on personal computers and workstations. But even with a standard terminal and a primitive graphics package, many of the advanced features can be implemented effectively if the software is properly designed.

The CSPM user interface incorporates features such as multiple screens, pop-up dialogs, scroll bars, drop-down menus, control buttons, and editable fields. The interface is primarily "modeless," with the exception of dialogs and alerts. A modeless interface allows the user to move freely around CSPM and select options from menus, controls on the screen, or fields to edit. This type of interface is important because it gives control to the user, and not the software. This freedom also presents a special problem in terms of managing the interface because the software must be ready to accept and handle any of these options at all times.

One problem which was encountered in using an output device that was not bit-mapped was the need to "refresh" the entire screen when removing

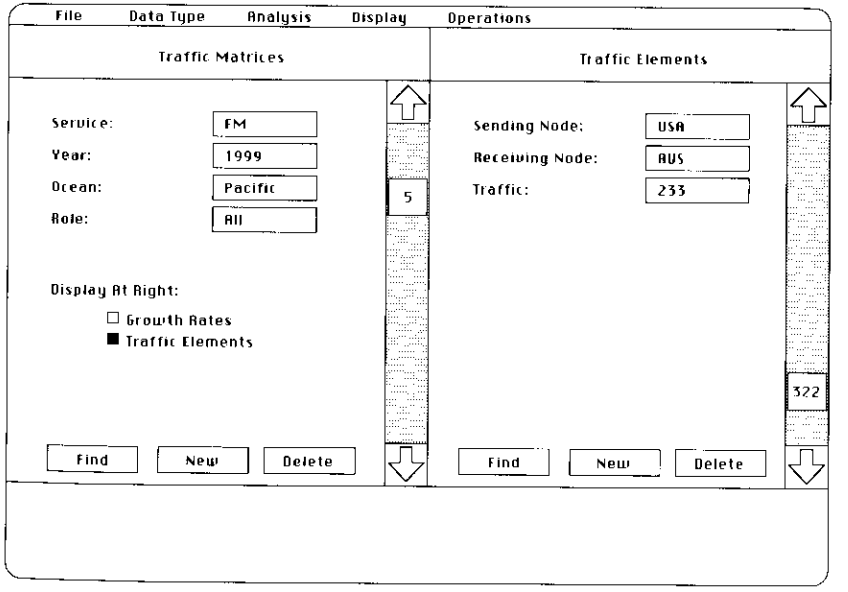


Figure 1. *Example of a CSPM Screen Layout*

displayed graphics (*e.g.*, messages), as opposed to just "turning off" pixels. Refreshing was sometimes unavoidable; however, the number of refreshes was reduced by displaying black rectangles over areas of the screen where graphics were to be erased. This method proved very effective in optimizing the performance of CSPM.

Managing the interface

Numerous options are available to the CSPM user. To manage this flexibility and still produce readable code, an event-driven design was selected. An event-driven interface requires a central point of control. With few exceptions, all user inputs flow through an event manager and are passed out to the "objects" that handle them. Conceptually, the objects are the screens, dialogs, and menus, which are represented internally by procedures that handle their related events. Events are passed between objects by means of event queues. Although an event manager is usually associated with object-oriented languages, CSPM demonstrates that it (and other object-oriented techniques) can be successfully implemented using a procedure-oriented language such as Pascal.

The most basic event handled is the pressing of the mouse button. The position of the cursor when the mouse button is pressed is passed through the hierarchy of objects until it reaches one that recognizes it. For example, a menu object might recognize the position as the selection of a command, and therefore the menu handler would produce a menu command event and post it in the event queue. The event would eventually be pulled from the event queue by the event manager and passed through the hiearchy until an object recognizes it. Figure 2 illustrates the flow of events through various objects in CSPM.

A screen or dialog recognizes a mouse position only if it falls within the boundaries defined by its internal representation (see below). By this method, a screen determines whether a user wants to edit a field, select an option, or invoke an associated function. When an object recognizes and handles an event, it does not pass it on to other objects; however, it may create new events which it places in the event queue for other objects to handle.

Defining screens

Each screen and dialog is described by an associated internal data structure which contains the locations of fields, buttons, controls, and scroll areas, as well as the text that is displayed. Utility routines are used to easily create and describe these features. After defining the structures, the programmer can use additional utilities to display, erase, and manage these screens and dialogs.

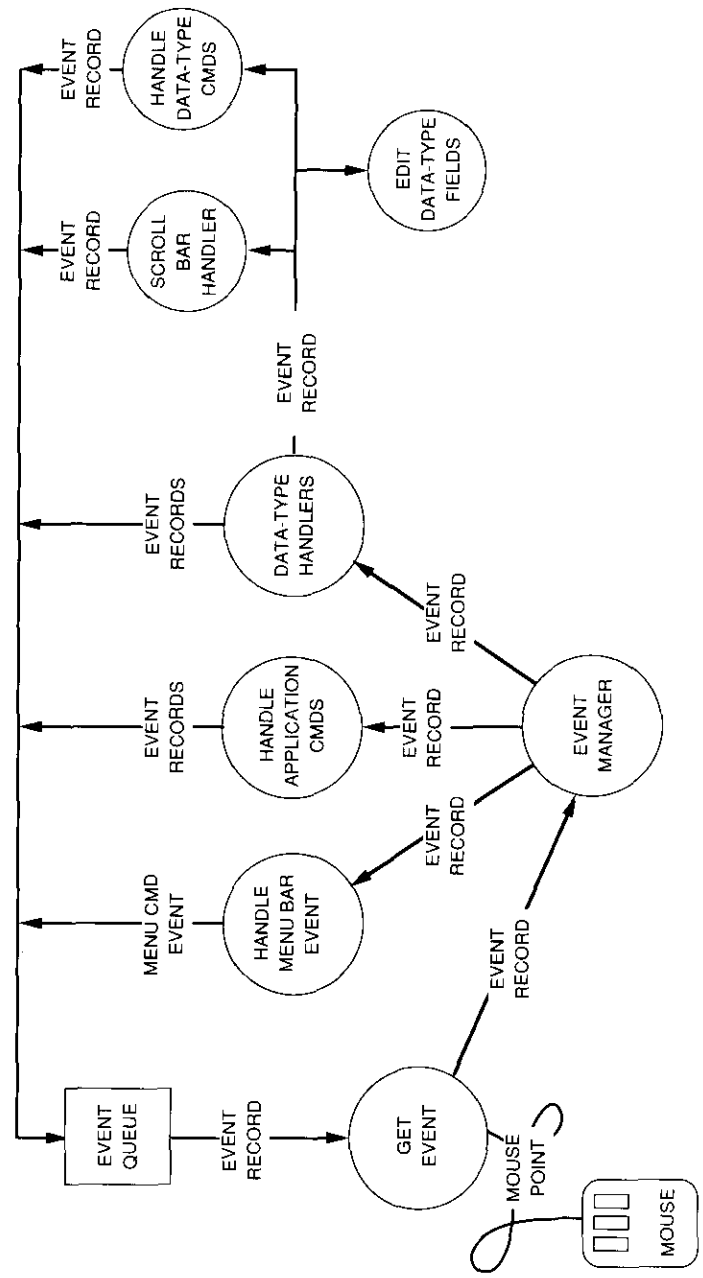


Figure 2. *Flow of Events Through the Event Manager and CSPM*

Expansion

Expansion of the model is facilitated by using this object-oriented approach. The addition of new commands, screens, fields, or dialogs does not affect the structure of the software. Rather, only the event manager need be made aware of new objects (*e.g.*, screens), and only an object need be made aware of the events that it must handle.

Without the event-controlled hierarchical design, each object would have to know about many other objects. The addition of new objects would affect all the objects, instead of only the controlling event manager.

Report output

CSPM generates a variety of output reports, including detailed satellite system data, equipment inventories, traffic, and analysis results. All output can optionally be directed to either a disk file or the printer. Examples of reports output by CSPM are included with the analysis algorithm descriptions.

System requirements

A system planner must gather and organize a large amount of data describing a satellite communications system, including traffic requirements data. The computer model must store, manipulate, and analyze these data in a manner that is easy to use and control. An interactive system is required in order to facilitate the iterative testing of system alternatives.

CSPM had to be flexible and expandable. Predefined parameters were avoided to ensure the most general representation possible. In addition, the model had to support multiple systems (each potentially containing numerous satellites) and multiple services. Current satellite communications systems handle numerous services within a single satellite, and often in a single transponder.

The model had to simultaneously address both system capacity and costs. Changes in the system configuration or traffic levels over time had to be represented, since satellites and equipment can be added or deleted on a yearly basis and the volume of traffic can grow, shrink, or be diverted to alternate paths or modes of transmission (services).

The first step in using CSPM is to define a satellite communications system and its traffic requirements. These system configuration data and traffic data are stored in separate databases and used as inputs to the analysis algorithms. The system configuration data describe a satellite communications system over a period of time, while traffic data define the number of communications channels between each pair of nodes for a specific service in a particular year. Multiple versions of the databases can be created to represent different system configurations, satellite systems, or traffic scenarios.

Ocean Regions

A global satellite communications network, such as the INTELSAT system, is composed of what can be considered to be independent satellite systems (*e.g.*, the Atlantic, Pacific, and Indian systems illustrated in Figure 3). For simplicity, CSPM considers one independent satellite system at a time when performing traffic analyses. Each system is called an *ocean region*. A valid network model might include only one ocean region representing an entire communications system. This configuration would be appropriate for a single satellite system, or for a system whose satellites are visible to all network nodes.

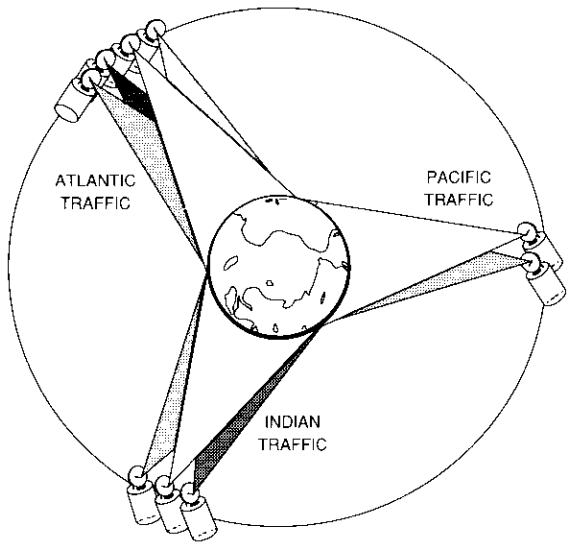


Figure 3. *Division of the Global System Into Three Independent Systems Represented by Ocean Regions*

Nodes

The sources and sinks of traffic, or *nodes* in CSPM, must be identified by name so that other satellite system data can reference them. The nodes can be earth stations, antennas, countries, or any other entity that represents a traffic source or sink. A latitude and longitude value is associated with each node.

Services

A *service* is a category of communications traffic such as frequency-division multiplex/FM/frequency-division multiple access (FDM/FM/FDMA) or time-division multiple access (TDMA). A service definition includes bandwidth utilization factors (in channels per unit bandwidth), which are used to determine effective satellite capacity when loading transponders. A system planner can use the concept of a service to define special classes or combinations of traffic, as well as to accommodate new methods of utilizing satellite bandwidth, such as companded FM (CFM) used in FDM/CFM/FDMA services, for example.

Satellite roles

Satellite *roles* in CSPM name a specific satellite function (or orbital location) within an ocean, such as "primary" or "major path." The role name permits a system designer to move or replace an existing satellite without changing the supporting system configuration data.

Satellite data

The space segment configuration is specified by using a combination of satellite types and specific satellite deployment actions for each type. A satellite is uniquely identified in CSPM by its *type* (*e.g.*, INTELSAT V) and its *flight number* (*e.g.*, F-3). A system planner defines each satellite type used in the communications system, optionally includes cost data (such as purchase, launch, and operating costs), and then specifies deployment actions for satellites of that type. Deployment actions define changes in the space segment configuration. Based on the deployment data, CSPM can generate an on-station plan that specifies the satellites occupying each role for each year in the study period.

Beam coverage regions

The area of the earth that a satellite antenna can transmit to or receive from is called a *beam coverage region*. Each transponder on a satellite is connected to both an up-link beam and a down-link beam. An earth station (*i.e.*, node) can transmit to or receive from a particular transponder only if the node is covered by the transponder's up-link or down-link beam, respectively. A system planner uses *regions* in CSPM to define beams by identifying the nodes that fall within the boundaries of a beam (Figure 4).

Satellite configurations

A *satellite configuration* defines the following for each transponder in a satellite: the service(s) carried, the available capacity, the beam coverage

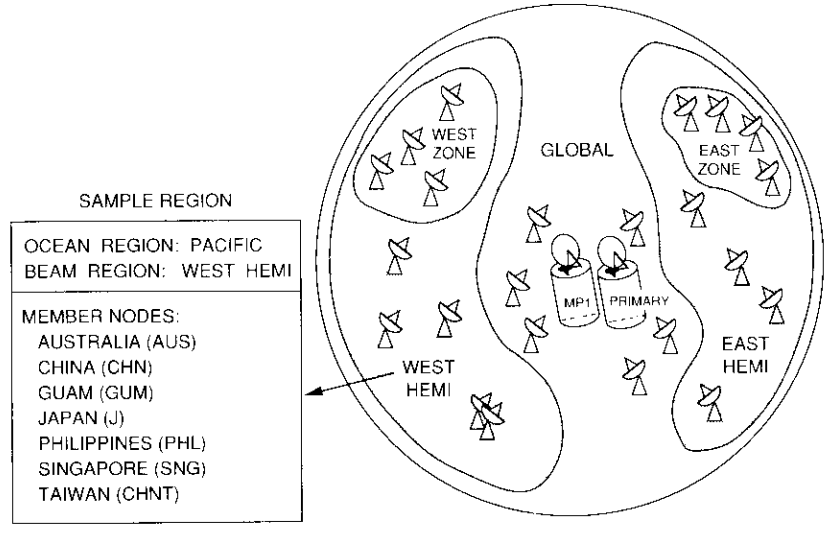


Figure 4. *Beam Coverage Regions for the Pacific Ocean Region*

regions connected, and the frequency bands used. Every satellite in a system has a satellite configuration associated with it for each year of its deployment.

Earth station equipment

The earth segment configuration is specified by using a combination of *equipment types* and *deployment actions*. The existence of equipment at a node in a given year provides that node with access to the satellite specified for the specific service and/or frequency band of the equipment type (Figure 5). Based on these data, CSPM generates the access information used in the traffic routing and assignment algorithms.

Service groups

Service groups in CSPM are used to specify multiple services per transponder. A service group definition includes a group name and a list of services belonging to it.

Traffic matrices

The traffic data define the actual or forecasted traffic requirements between network nodes. CSPM represents traffic data in the form of traffic matrices, each identified by a service, year, ocean region, and role. Each traffic matrix represents the traffic for a single defined service in a particular year. A matrix

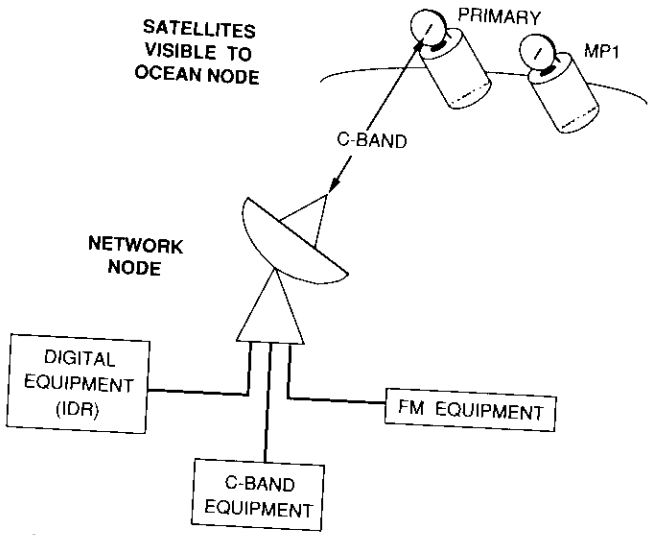


Figure 5. *Adding Equipment to a Network Node to Define Access to Satellites*

consists of many traffic elements, each defining the number of one-way communications channels from a sending node to a destination node. The number of channels can represent bearer channels, virtual channels, or any other user-defined unit, as long as the capacity values specified for services and transponders are consistent with the traffic. Associated with each traffic matrix is a set of forecasted growth rates for up to 20 years.

Although the term "traffic matrix" is used to describe the traffic data, it is not efficient to store the data as an $n \times n$ matrix. Given the potentially large number of nodes and the relative sparseness of traffic, a large amount of storage would be unused. Traffic data are stored internally using list structures that have no arbitrary size limitations and require minimal memory resources. Figure 6 illustrates this structure, with the A nodes representing the sending nodes and the B nodes representing all the receiving nodes for the particular sending node. "Nil" represents an end-of-list marker.

Because of the potentially large size of traffic matrices, a feature was implemented in the program that permits users to retrieve traffic data from a common database. The link to a central shared database also ensures consistency through the use of common data.

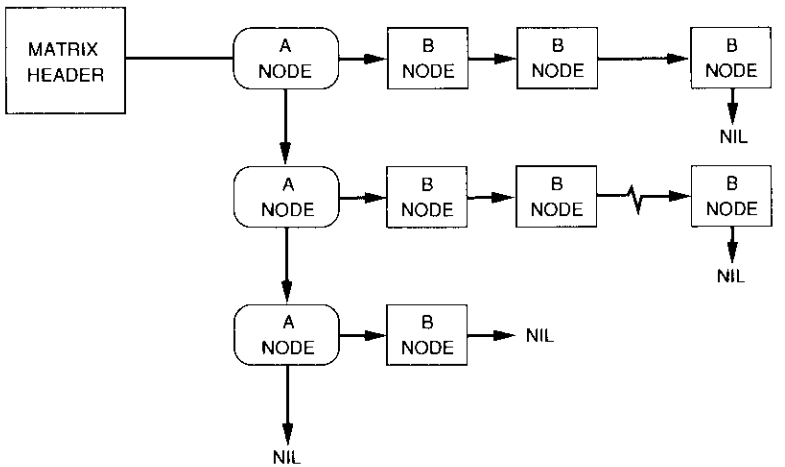


Figure 6. *Internal Structure of a Traffic Matrix*

Analysis algorithms

One of the first applications of CSPM was in evaluating the design alternatives for the INTELSAT VII generation of satellites in the Pacific ocean region. The analyses were performed to determine if the two planned satellites could handle the traffic forecasted for 1999 (the last year of their designed life), and also to determine the effect of employing an optional switchable east spot beam that could include Australia. The INTELSAT VII test case is used to demonstrate the major analysis algorithms in CSPM, including traffic manipulation, traffic routing, and traffic assignment. Note that bearer channels are used as the unit of capacity for the sample case.

Traffic manipulation

One of the most important features of CSPM is its ability to manipulate traffic matrices. The planner can use the following manipulation functions to derive the traffic matrices for a specific year, ocean region, or combination of services:

- Adding two matrices
- Subtracting two matrices
- Partitioning a matrix
- Scaling a matrix.

Partitioning and scaling are described in the following subsections.

PARTITIONING

The partitioning algorithm generates a traffic matrix that is a subset of an existing traffic matrix, based on user-specified parameters. The partitioning algorithm accepts as parameters either a satellite longitude and a minimum elevation angle, or a region of nodes.

For each element in the existing traffic matrix, both the sending and receiving nodes are tested against the parameters to ensure that they are visible, thus qualifying the traffic element to be part of the resulting matrix (Figure 7). The qualifying test takes one of two forms. If a region is specified, the algorithm determines if both nodes are members of that region. Alternately, if a satellite longitude and minimum elevation angle are specified, a test is made to determine if a satellite at that location is visible from both nodes at an elevation angle greater than or equal to that specified. If the applicable test succeeds, the traffic element is entered in the new matrix; otherwise, it is rejected and added to a residual matrix.

SCALING

In the traffic matrix scaling algorithm, each element in an existing traffic matrix is multiplied by a single scaling factor to produce a new matrix. This

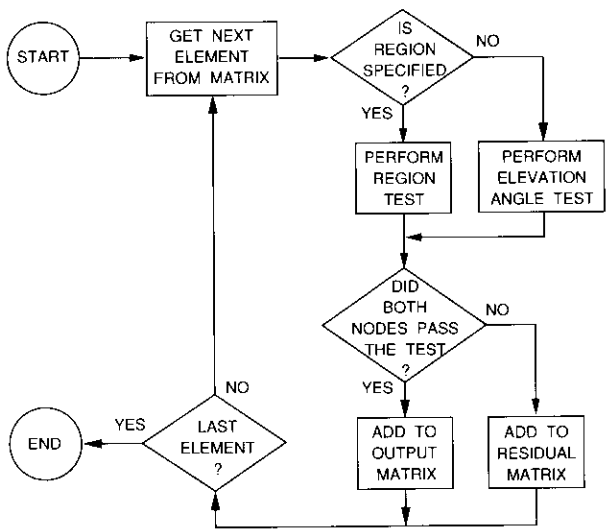


Figure 7. *Flowchart for the Traffic Partitioning Algorithm*

scaling factor is obtained in one of three ways. In the first method, the user specifies the scale factor directly, and no further computation is needed. In the second method, the user must specify the target traffic size. The computed scaling factor, SF, is the ratio of the target traffic to the current traffic in the matrix, represented as

$$SF = \frac{\text{Target traffic}}{\text{Current traffic}} \quad . \tag{1}$$

In the third method the user specifies the target year to which the matrix should be scaled (or more specifically, *grown*). In this case, CSPM calculates SF with the following equation:

$$SF = \prod_{i}^{TY-1} (1 + GR_i) \tag{2}$$

where MY = year of the matrix being grown
TY = target year
GR = array of yearly growth rates associated with the matrix.

In the INTELSAT VII test case, traffic matrices for the year 1993 were partitioned, added, and grown to obtain the required matrices for the analysis. Figure 8 shows the steps taken using the traffic manipulation functions.

Traffic routing

Traffic routing is the process of distributing the node-to-node traffic within an ocean region among the satellites serving that region, while satisfying specified traffic diversity requirements and ordering rules that govern satellite access. The process can be applied successively to several traffic matrices representing different services. For each input matrix, the process results in a set of new traffic matrices, one for each satellite in the ocean region. The traffic routing process is independent of the capabilities of the satellites involved.

SATELLITE ORDERING

The first step in the traffic routing process is to establish the order in which the satellites should be considered. CSPM first computes the number of nodes that can access each satellite, based on the equipment deployment data provided by the user. A node can access a satellite in a specific year if, prior to that year, the node has been assigned equipment that references the satellite. The satellites are then sorted in ascending order by the number of accessing nodes. CSPM eliminates those satellites that have no accessing nodes.

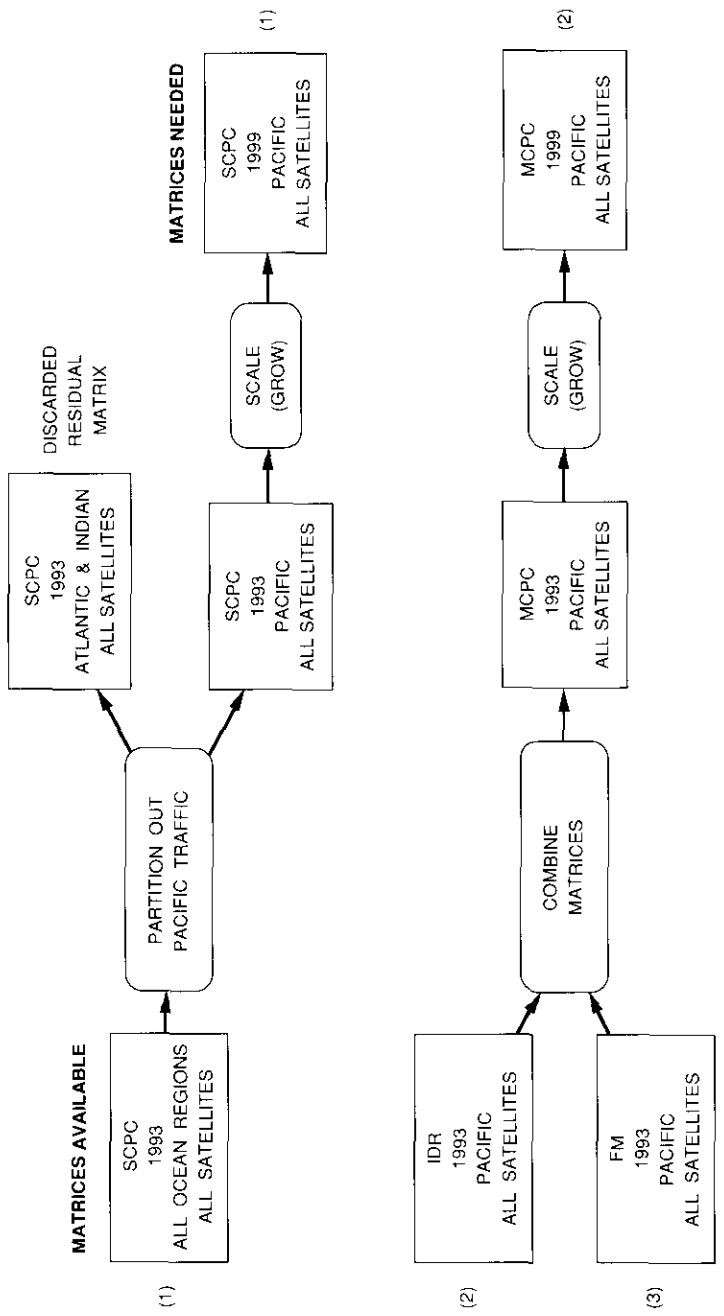


Figure 8. *Traffic Manipulation Steps Used in the Sample Case*

SATELLITE SELECTION

After satellite ordering, the sending and receiving nodes for each traffic element are tested against each satellite in turn to determine the satellites that can provide a routing path for the traffic. The nature of the test depends on whether frequency cross-strapping is assumed to be in effect. Frequency cross-strapping allows the frequency bands of the up- and down-link beams used by a particular traffic element to be different. If frequency cross-strapping is assumed, a test is made to determine whether each node of a traffic element can access the satellite in any frequency band. Otherwise, the test is more stringent in that each node must be able to access the satellite in a frequency band common to both nodes.

The subset of satellites for which this test succeeds comprise the set, A, of satellites available for routing the traffic between this pair of nodes. The traffic is then subjected to the diversity requirements described below.

APPLYING TRAFFIC DIVERSITY

It is sometimes desirable to split a traffic element between two or more satellites to prevent excessive dependence on a single satellite or pair of earth station antennas. The planner controls this process by specifying diversity thresholds and diversity rules.

Diversity Thresholds. Diversity thresholds are specified as a set of monotonically increasing traffic values, T_i, $(i = 1, \ldots, n)$, where n is the number of satellites in the ocean region and T_i is the number of channels in which i routing paths are desired. T_1 is set to zero, and T_i $(i = 2, \ldots, n)$ is set to ∞ if it is not specified by the user.

For each traffic element, CSPM searches for the largest diversity threshold T_i for which the traffic level is greater than T_i. T_i must meet the following requirements:

$$C \geq T_i$$

and

$$\nexists k, (i < k \leq n) : C \geq T_k$$

where $\nexists$ means "there does not exist," and C is the amount of traffic in channels for the current node pair. The number of satellites, s, among which the traffic is to be allocated is represented by

$$s = \min(i, j) \tag{3}$$

where j is the number of satellites available to both nodes in the traffic element (*i.e.*, the number of members in set A). The traffic is then allocated among s satellites according to the diversity rules.

Diversity Rules. Each diversity rule consists of a set of proportions, P_r ($r = 1, \ldots, n$), where $P_r \geq 0$. Then, $(P_r / \sum_{i=1}^{n} P_i)$ is the fraction of the traffic that should be routed to satellite role r if $\sum_{i=1}^{n} P_i$ is the sum of all proportions in the rule.

In applying diversity rules, CSPM looks for a rule which has exactly s nonzero P_i values, where i is a member of the set of available satellites, A, for the traffic element. If such a rule exists, the traffic (C_r) that should be routed to each satellite role (r) can be found with the equation:

$$C_r = \frac{C \cdot P_r}{\sum_{i=1}^{n} P_i} \tag{4}$$

where C is the total node-to-node traffic in the channels, and P_i is the proportion of traffic to assign to satellite role r. If a suitable rule does not exist, the traffic is allocated equally among the first s available satellites.

RESULTS

For each C_r that is greater than zero, a traffic element is created with the same send-receive node pair as the original traffic element, and with C_r channels of traffic. This traffic element is added to the traffic matrix created for satellite role r having the same service as the original traffic element.

The result of traffic routing is a set of traffic matrices, one for each service, routed for each satellite. Figure 9 is an example of the report generated by the routing algorithm.

SAMPLE CASE

The traffic matrices for the multiple channels per carrier (MCPC) and single channel per carrier (SCPC) services in the year 1999 were routed between the two INTELSAT VII satellites planned for the Pacific ocean region in that year (Figure 10). Iterative analyses of the sample case were performed using different diversity thresholds, diversity rules, and node accesses to obtain a balanced mix of traffic between the two satellites. The matrices generated by the routing process can also be manually edited, allowing the system planner to adjust the amount of traffic placed on each satellite in the system.

Traffic assignment

Traffic loading on satellite transponders is performed by the traffic assignment algorithm. The planner indicates the satellite to be used and the services to be assigned. The algorithm then determines the satellite configuration and traffic matrices to be used based on these user-entered paramaters.


```
                                                                    Date: 06/14/89
                                                                    Time: 09:53:12
                                                                    Page: 1
COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
                  Intelsat VII -config 4
                  TRAFFIC ROUTING SUMMARY

System File : CTR
Traffic File: OHDEMO

Traffic Routing Parameters:

Year     : 2000
Ocean    : Pacific
Services : BEARER, SCPC

Used access table for accesses

Diversity Thresholds:
 >= 0 channels, split between 2 roles

Satellites used for traffic routing:
pacific PRIMARY  using INTELVII Flight # 1  @ 174.0 Deg E
pacific MP1      using INTELVII Flight # 2  @ 177.0 Deg E

Satellite                Nodes       Links      Bearer
Role     Service       Accessing      Routed    Circuits
-------  -------       ---------   ---------   --------
PRIMARY  BEARER               28         236      16440
         SCPC                 27         122       1290
   Total PRIMARY              55         358      17730

MP1      BEARER               12         110       9300
         SCPC                  5           8         18
   Total MP1                  17         118       9318

Total For All Roles           72         476      27048

***** End Of Report *****
```

Figure 9. *Example of a Traffic Routing Summary Report*

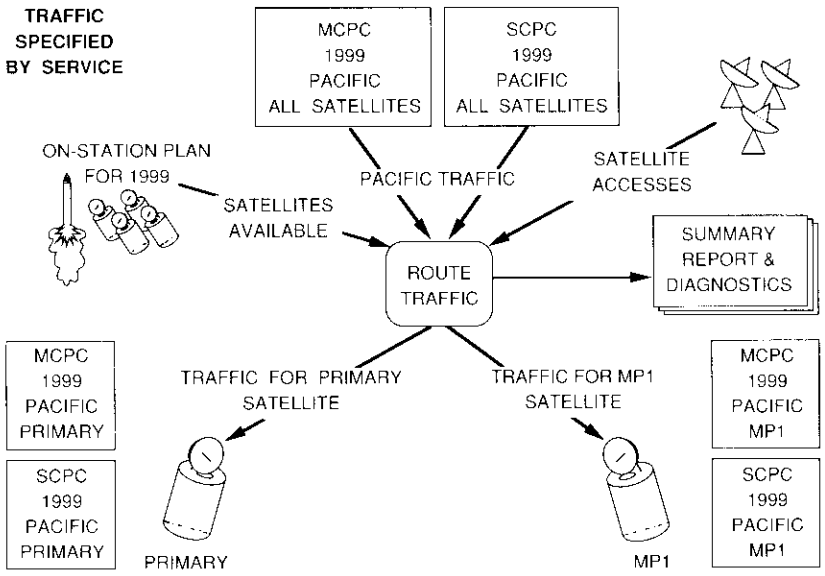


Figure 10. *Traffic Routing Process in the INTELSAT VII Test Case*

The main objective of the traffic assignment algorithm is to place as much traffic as possible onto transponders, leaving little or no traffic that cannot be assigned within the specified parameters. This objective is accomplished by filling the least-accessed transponders first, with traffic from nodes that access the fewest transponders (see below). The algorithm also allows the system planner to assign services in a specific order, thus providing the capability of assigning services with large capacity requirements (*e.g.*, TV and TDMA) first, and preventing the CSPM from splitting them.

In the INTELSAT VII study, the assignment process was performed twice, once for each of the two Pacific satellites.

ASSIGNMENT SETUP

Determining Satellite Configuration and Traffic. The satellite configuration to be used during the assignment process describes the satellite transponders. The configuration is determined based on the satellite on-station plan derived from satellite deployment data (or optionally specified by the system planner). The algorithm then determines the traffic matrices to be assigned. CSPM searches the traffic database to identify a traffic matrix for each specified service with the correct year, ocean region, and role (*i.e.*, satellite).

Determining Node Ordering. Each node that occurs in any selected traffic matrix is tested against all the transponders in the satellite configuration. A count is kept of the transponders that each node can access. A node is able to access a transponder if either of the following access rules applies:

a. the node is equipped to transmit in the up-link frequency band of the transponder and lies in the up-link beam coverage region for the transponder, or

b. the node is equipped to receive in the down-link frequency band of the transponder and lies in the down-link beam coverage region for the transponder.

The list of nodes is then sorted, with the nodes having the least accesses sorted to the front. The traffic elements are assigned using this node-ordering approach to minimize the amount of traffic that is left unassigned. In other words, those node pairs with the fewest opportunities to find an acceptable transponder are considered early in the process, when the number of unfilled transponders is greatest. CSPM does not permit the user to manually specify the order in which the nodes are considered.

The selected traffic matrices are then assigned one at a time, using the algorithm described below, in the order that the services are specified by the user.

TRANSMISSION GROUPS

During the traffic assignment process, the traffic in each transponder is divided into a number of transmission groups. Each group contains some or all of the traffic originating at a single sending node for a single service. For each service, the user must specify the maximum number of channels in a transmission group, and whether traffic for more than one destination (receiving node) can be assigned to the same group.

TRANSPONDER CAPACITY

The maximum amount of traffic that can be assigned to a transponder is governed by one of two factors: the effective transponder bandwidth, BW_T (in MHz); or, optionally, the transponder capacity in channels, C_{max}. If the transponder bandwidth is specified, it is used along with the bandwidth utilization efficiency to calculate the transponder capacity. For simplicity, factors such as the power level and noise density of the transponder are not considered when determining capacity.

Bandwidth Utilization Efficiency. For each defined service, s, there is a user-specified function, $U_s(n)$, referred to as the bandwidth utilization efficiency of that service. This function gives the number of channels of service s per megahertz of bandwidth as a function of the number, n, of transmission groups (for all services) assigned to the transponder. When n exceeds the largest number specified (n_{max}), the function is assumed to remain constant at the value $U_s(n_{max})$. The bandwidth utilization function is employed when the capacity of a transponder is calculated with the transponder bandwidth.

Determining Capacity. During the assignment process, the available capacity, C_A, of a transponder (in channels of service s) changes and can be determined by

$$C_A = C_{max} - C_U \quad ; \text{ if } C_{max} > 0, \tag{5a}$$

$$C_A = U_s(n) \cdot BW_A \quad ; \text{ otherwise,} \tag{5b}$$

where C_U is the number of channels already assigned to the transponder, n is the current number of transmission groups, and BW_A is the unused (available) bandwidth in the transponder.

BW_A is the difference between the total bandwidth of the transponder and the bandwidth used by all transmission groups, i, for all services, s, currently assigned to the transponder, as follows:

$$BW_A = BW_T - BW_U \tag{6}$$

where BW_U is computed from

$$BW_U = \sum_s \sum_i \frac{T_{s,i}}{U_s(n)} \quad .$$

$T_{s,i}$ is the amount of traffic in the ith transmission group of service s, and $U_s(n)$ is the bandwidth utilization efficiency for service s when n groups have been assigned to the transponder.

If the transponder accepts more than one service, the number of groups, n, includes the groups for all services, so that the function $U_s(n)$ more accurately reflects the bandwidth needed for the traffic. This method of determining bandwidth utilization does not consider any special relationship between the services assigned that could alter the effective capacity used.

ASSIGNING TRAFFIC

Sorting the Transponders. The transponders are first ordered according to a frequency priority associated with the service being assigned. Service frequency priorities are ordered pairs of up- and down-link frequencies that provide a method for the system planner to load transponders in a frequency-specific order. The transponders are then sorted by the number of nodes that can access them, with the least-accessed becoming the first transponder in the list of transponders having the same frequency pair. This ordering loads the least-accessed transponders first, thereby minimizing the amount of traffic that is left unassigned.

After the transponders are sorted, the traffic elements are assigned in turn, in the order specified by the sorted node list, onto transponders tested in the order specified by the sorted transponder list.

Determining a Suitable Transponder. CSPM attempts to put the remaining traffic, C_R, from the current traffic element on a suitable transponder. For the first iteration of this process with the current traffic element, C_R is the total traffic from the sending node to the receiving node of the element. A transponder is deemed suitable if it exhibits the following characteristics:

a. The sending node is equipped to transmit in the up-link frequency band of the transponder and lies in the up-link beam coverage region for the transponder.

b. The receiving node is equipped to receive in the down-link frequency band of the transponder and lies in the down-link beam coverage region for the transponder.

c. The transponder supports the service of the traffic matrix being assigned.

d. The transponder has available capacity ($C_A > 0$, as calculated above).

The last requirement can be waived by the user if desired. If no suitable transponder is found, a diagnostic message is logged and the process continues with the next traffic element.

Assigning the Traffic to a Transmission Group. If a suitable transponder is found, the traffic C_R must be placed in one or more transmission groups. The amount of the traffic, C_S, that can be assigned to any one group in the selected transponder is the smallest of the following three values:

- the capacity available on the transponder (C_A),
- the size of the traffic element to be assigned (C_R), and
- the maximum transmission group size allowed for this service ($T_{s,max}$)

so that

$$C_S = \text{minimum } (C_A, C_R, T_{s,max}) \quad . \tag{7}$$

A search is made for a transmission group already assigned to the selected transponder. A group i is considered suitable if the following conditions apply:

- The sending node is the same.
- The service is the same.
- The service allows multiple destinations for one transmission group.
- The availability capacity, T_A, in the transmission group is greater than or equal to the amount of traffic being assigned, C_S.

The available capacity, T_A, can be calculated using the expression

$$T_A = T_{s,\max} - T_{s,i} \tag{8}$$

where $T_{s,i}$ is the amount of traffic already assigned to group i.

If a suitable transmission group is found on the transponder, the traffic is assigned to it. Otherwise, a test is made to see if the transponder can accept a new transmission group. If a new transmission group must be created for the sending node, the transponder capacity is recalculated to account for its presence, as follows:

$$C_A = U_S(n + 1) \cdot BW_A \tag{9}$$

where

$$BW_A = BW_T - \sum_s \sum_i \frac{T_{s,i}}{U_s(n + 1)} \quad .$$

If C_A is zero, the traffic is not assigned to this transponder and the process continues by searching for a new transponder, as described above. Otherwise, the amount of traffic to assign is modified, and is the least of C_S (as calculated above) and the newly calculated C_A, thus

$$C_S = \text{minimum } (C_A, C_S) \quad . \tag{10}$$

A new transmission group is created, and the traffic, C_S, is placed in it. This group is then added to the current transponder.

Recalculating Values for Unassigned Traffic. When an amount of traffic, C_S, has been assigned to a transponder, new values for the total number of channels assigned to the transponder (C'_U), the used bandwidth (BW'_U) corresponding to C'_U, and the amount of the current traffic element remaining to be assigned (C'_R) are recalculated as follows:

$$C'_U = C_U + C_S \tag{11}$$

$$BW'_U = BW_U + C_S/U_S(n) \tag{12}$$

$$C'_R = C_R - C_S \quad . \tag{13}$$

If C'_R is zero, then all the traffic from the traffic element has been assigned and the process continues with the next traffic element, starting with the first transponder. Otherwise, the remaining traffic must be assigned and the process continues with the current transponder.

RESULTS

The output of the traffic assignment process consists of a transponder loading plan and a summary report on capacity utilization for the satellite (Figure 11). In addition, diagnostics are produced which identify the traffic elements that could not be assigned due to capacity or access limitations. The system planner can modify a transponder loading plan in order to fine-tune the results to meet system requirements.

SAMPLE CASE

In the INTELSAT VII case, the desirability of an optional switchable beam was investigated by adding Australia to the East Zone region and repeating the traffic assignment process. The results showed that some traffic from the heavily loaded West-Hemi-to-East-Hemi transponders was off-loaded to the transponder having the switchable beam. Figure 12 illustrates the traffic assignment process.

Economic analysis

The economic algorithms developed for CSPM originated from work performed at American Telephone and Telegraph Co. (AT&T) [2]. Economic evaluations of specific system configurations are based on the satellites and equipment deployed during the study period, as well as on cost data for each satellite and equipment type. These cost data include economic lifetime, purchase costs, installation costs, operation and maintenance (O&M) costs, and satellite launch costs. The analyses provide for an inflation rate, discount rate, and required rate of return on investment, which are considered constant through the period of the analysis. The following are available to all of the analyses:

$IRATE$ = annual inflation rate
$IYEAR$ = reference year for inflation
$DRATE$ = discount rate
$RATE$ = anticipated rate of return.

Additional information required in the economic model is the period (or time frame) over which the economic evaluations are to be performed. The specific algorithms discussed are present-worth analysis, system replacement costs, and revenue requirements. These algorithms are currently implemented

COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
Intelsat VII -config 4
TRAFFIC ASSIGNMENT SUMMARY

Date: 06/14/89
Time: 09:55:21
Page: 1

System File : CTR
Traffic File: OHDEMO

Transponder Loading Plan Identifiers:

Year : 2000
Ocean : Pacific
Role : PRIMARY
Config : VII-174

** NOTES: (1) Capacity values used during assignment
(2) Used access table for accesses
(3) Bandwidths in ()'s are overridden by Channels Available

Transponder	Up Region	Dn Region	Service	Txmission Elements	Traffic (Channels)	Bandwdth Used	Bandwidth Available	Channels Available	Fill Factor
11	VII-EH	VII-WH	BEARER	6	1100	———	———	———	1.00
			Total 11	6	1100	(72.00)	(72.00)	1100	1.00
12	VII-WH	VII-WH	BEARER	10	1098	72.00	72.00	———	1.00
			Total 12	10	1098	72.00	72.00	———	1.00
14	VII-WZ	VII-WH	BEARER	7	1098	72.00	72.00	———	1.00
			Total 14	7	1098	72.00	72.00	———	1.00
15	VII-EH	VII-WH	BEARER	2	648	36.00	36.00	———	1.00
			Total 15	2	648	36.00	36.00	———	1.00
21	VII-WH	VII-EH	BEARER	11	1200	———	———	———	1.00
			Total 21	11	1200	(72.00)	(72.00)	1200	1.00

Figure 11. *Example of a Traffic Assignment Summary Report*

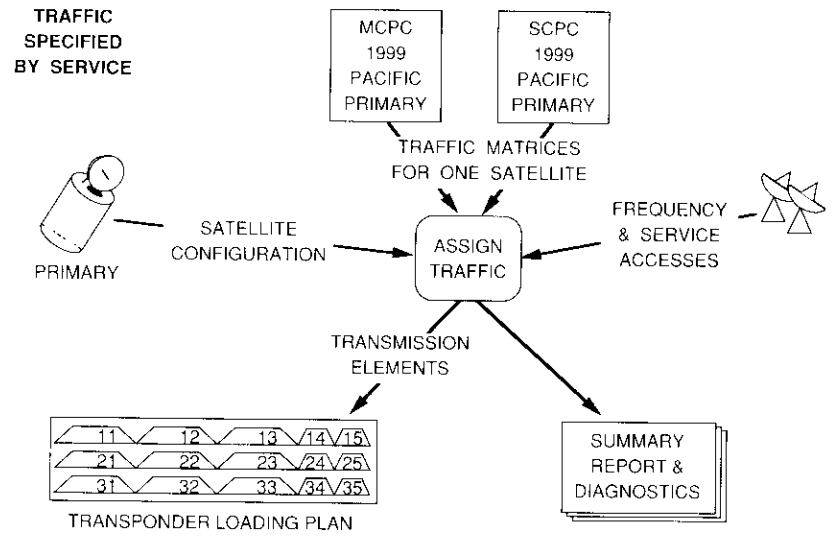


Figure 12. *The Traffic Assignment Process*

so that results are computed on a yearly basis. Figures 13 through 16 are examples of reports that are generated by the economic analysis.

SATELLITE AND EQUIPMENT COSTS

Cost information for each type of earth segment and space segment equipment is required in order to determine the economic consequences of planning decisions, and for comparing the costs of alternative systems. Each type of earth station equipment may have associated with it the following cost data:

$LIFE_i$ = economic lifetime in years
C_i = base unit cost
IC_i = installation cost
OMC_i = operation and maintenace costs.

Cost data for each satellite type i are as follows:

$LIFE_i$ = economic lifetime in years
C_i = base unit cost
LC_i = launch vehicle costs.

All costs are assumed to be incurred in the year in which the equipment is introduced into the system (or launched, in the case of satellites).

COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
Demonstration System
ANNUAL EXPENDITURES TABLE

Date: 06/30/88
Time: 08:43:42
Page: 2

(all $ amounts in thousands)

Item	Component	1980	1981	1982	1983	1984	1985	1986
Equipment:								
	EQUIP1	$ 350,900	$ 25,047	$ 0	$ 0	$ 5,049	$ 0	$ 0
	EQUIP2	0	0	0	1,164	0	0	0
	FMKU	612	0	0	0	0	0	0
	IBSEQ	0	0	0	0	0	0	0
	IDR	0	0	0	0	0	16,723	0
Equipment Investment		$ 351,512	$ 25,047	$ 0	$ 1,164	$ 5,049	$ 16,723	$ 0
Links:								
	LINK1	$ 590	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
Link Investment		$ 590	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
Satellites:								
	INTELVII	$ 100,000	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
	INTELV	120,000	77,625	0	83,154	172,128	0	0
Satellite Investment		$ 220,000	$ 77,625	$ 0	$ 83,154	$ 172,128	$ 0	$ 0
Total Investments		$ 572,102	$ 102,672	$ 0	$ 84,318	$ 177,177	$ 16,723	$ 0
Present Worth Equiv		$ 572,102	$ 96,860	$ 0	$ 70,795	$ 140,341	$ 12,496	$ 0
Annual O&M Costs		$ 40,600	$ 56,915	$ 58,907	$ 74,138	$ 104,345	$ 116,963	$ 121,056
Present Worth Equiv		$ 39,434	$ 52,152	$ 50,922	$ 60,460	$ 80,278	$ 84,892	$ 82,890
End Year Net Bk Val		$ 556,106	$ 632,771	$ 606,765	$ 655,499	$ 777,191	$ 737,738	$ 681,562
Present Worth Equiv		$ 524,628	$ 563,165	$ 509,452	$ 519,218	$ 580,764	$ 520,078	$ 453,279
Ann. Decline in Assets		$ 47,474	$ 58,324	$ 53,712	$ 61,029	$ 78,795	$ 73,182	$ 66,798
Cumulative A.D.A.		$ 47,474	$ 105,798	$ 159,510	$ 220,539	$ 299,334	$ 372,517	$ 439,315

Figure 13. *Example of an Annual Expenditures Report*

COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
Demonstration System
NET PRESENT VALUE TABLE

Date: 06/30/88
Time: 08:43:43
Page: 3

(all $ amounts in thousands)

Item	Component	Without Inflation	With Inflation	Residual Value	Op. & Maint. Costs	Amount Depreciated	Years of Depreciation	Depreciation Per Year
Equipment Costs:								
	EQUIP1	$ 379,500	$ 380,996	$ 227,190	$ 311,990	$ 119,169	2,377	$ 50
	EQUIP2	1,050	1,164	813	386	296	4	74
	FMKU	612	612	337	4,129	157	63	3
	IBSEQ	0	0	0	0	0	318	0
	IDR	14,080	16,723	13,820	18,245	1,382	256	5
All Equipment Types		$ 395,242	$ 399,495	$ 242,161	$ 334,751	$ 121,004	3,018	$ 40
Link Costs:								
	LINK1	$ 590	$ 590	$ 358	$ 18,997	$ 167	35	$ 5
All Link Types		$ 590	$ 590	$ 358	$ 18,997	$ 167	35	$ 5
Satellite Costs:								
	INTELVII	$ 100,000	$ 100,000	$ 100,000	$ 0	$ 0	0	$ 0
	INTELV	420,000	452,907	116,068	219,175	150,257	16	9,391
All Satellite Types		$ 520,000	$ 552,907	$ 216,068	$ 219,175	$ 150,257	16	$ 9,391
Total for Study Period		$ 915,832	$ 952,991	$ 458,587	$ 572,923	$ 271,428	3,069	$ 88
Present Worth Equiv.		$ 862,361	$ 892,594	$ 304,987	$ 451,026	$ 219,081	3,069	$ 71

System Net Present Value as of January 1, 1980: $ 1,038,633

Figure 14. *Example of a Net Present Value Report*

COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
Demonstration System
SYSTEM REPLACEMENT COSTS

Date: 06/30/88
Time: 08:43:43
Page: 4

(all $ amounts in thousands)

Item	Component	1980	1981	1982	1983	1984	1985	1986
Equipment Costs:								
	EQUIP1	$ 350,900	$ 388,221	$ 401,813	$ 415,876	$ 435,481	$ 450,722	$ 466,495
	Equip2	0	0	0	1,164	1,205	1,247	1,291
	FMKu	612	633	656	679	702	727	752
	IBSEQ	0	0	0	0	0	0	0
	IDR	0	0	0	0	0	16,723	17,308
All Equipment Types		$ 351,512	$ 388,854	$ 402,468	$ 417,719	$ 437,388	$ 469,419	$ 485,846
Link Costs:								
	LINK1	$ 590	$ 611	$ 632	$ 654	$ 677	$ 701	$ 725
All Link Types		$ 590	$ 611	$ 632	$ 654	$ 677	$ 701	$ 725
Satellite Costs:								
	INTELVII	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
	INTELV	0	139,725	144,615	299,353	619,661	641,348	663,795
All Satellite Types		$ 0	$ 139,725	$ 144,615	$ 299,353	$ 619,661	$ 641,348	$ 663,795
Totals For Period		$ 352,102	$ 529,190	$ 547,715	$ 717,726	$ 1,057,726	$ 1,111,467	$ 1,150,366

Figure 15. *Example of a System Replacement Costs Report*

COMMUNICATIONS SYSTEM PLANNING MODEL (VERSION 1)
Demonstration System
SYSTEM REVENUE REQUIREMENTS

Date: 06/30/88
Time: 08:43:43
Page: 5

(all $ amounts in thousands)

Item	1980	1981	1982	1983	1984	1985	1986
Beg Year Net Bk Value	$ 5,000	$ 561,016	$ 634,458	$ 608,700	$ 650,354	$ 753,015	$ 714,775
Investment During Year	572,102	99,200	0	76,050	154,400	14,080	0
Depreciation In Year	16,096	25,768	25,768	34,406	51,749	52,331	52,331
Incr In Work. Capital	10	10	10	10	10	10	10
End Year Net Bk Value	561,016	634,458	608,700	650,354	753,015	714,775	662,454
Average Book Value	283,008	597,736	621,578	629,527	701,684	733,895	688,614
Return Requirement	$ 31,131	$ 65,751	$ 68,374	$ 69,248	$ 77,185	$ 80,728	$ 75,747
Operating Expenses	$ 41,272	$ 55,912	$ 55,912	$ 67,996	$ 92,476	$ 100,156	$ 100,156
Depreciation In Year	16,096	25,768	25,768	34,406	51,749	52,331	52,331
Return Requirement	31,131	65,751	68,374	69,248	77,185	80,728	75,747
Total Revenue Required	$ 88,499	$ 147,431	$ 150,053	$ 171,650	$ 221,410	$ 233,215	$ 228,234
Additional Revenue	$ 150	$ 150	$ 150	$ 150	$ 160	$ 175	$ 180
Net Revenue Required	$ 88,349	$ 147,281	$ 149,903	$ 171,500	$ 221,250	$ 233,040	$ 228,054

Figure 16. *Example of a Revenue Requirements Report*

NET PRESENT WORTH

The function of the net present-worth analysis is to generate tables that summarize the yearly investment costs for each satellite or equipment type, the total annual expenditures, annual O&M costs, and end-of-year book value, as well as the present-worth equivalent of each. The annual decline in assets and cumulative decline in assets on a yearly basis are also determined. Additional summary information pertaining to each satellite or equipment type over the entire period is computed, including total investment cost (with and without inflation), residual value, total O&M costs, total depreciation, years of depreciation, and the average depreciation per year. Earth segment and space segment totals are accumulated, including the total net present worth of the system.

Present-Worth Factors. The first step in determining net present worth is to generate a table of present-worth factors for each year of the study period. This factor, PWF_j, for year j of the study period is

$$PWF_j = (1 + DRATE)^{k-j} \tag{14}$$

where $DRATE$ is the discount rate and k is the year in which the net present value is to be determined.

Annual Investments and Expenditures. The analyses for earth segment and space segment costs are generally identical. In either case, the number of items ($n_{i,j}$) of equipment or satellite type i introduced in year j is first determined. The annual investment, $INV_{i,j}$, in each type is then

$$INV_{i,j} = C_i n_{i,j} \tag{15}$$

where C_i is the cost associated with satellite or equipment type i. The total annual investment, $INV_{\text{tot},j}$, in year j is obtained from

$$INV_{\text{tot},j} = \sum_i INV_{i,j} \quad . \tag{16}$$

The inflated cost, $INV'_{i,j}$, of investments in item i in year j is

$$INV'_{i,j} = INV_{i,j} \cdot (1 + IRATE)^{(j-IYEAR)} \tag{17}$$

where $IRATE$ is the inflation rate and $IYEAR$ is the reference year for inflation. Then, the total annual inflated investment and its present-worth equivalent are

$$INV'_{\text{tot},j} = \left(\sum_i INV'_{i,j}\right) \cdot PWF_j \quad . \tag{18}$$

If O&M costs apply to item i, then the total annual O&M cost is

$$OMC_{\text{tot},j} = \sum_i OMC_{i,j} \quad . \tag{19}$$

The yearly depreciation of each satellite or equipment type is calculated from

$$DEPR_{i,j} = \frac{INV'_{i,k}}{LIFE_i} \qquad , \; k \le j \le (k + LIFE_i) \tag{20}$$

where $LIFE_i$ is the economic lifetime of equipment type i, and k is the year of introduction. Total yearly depreciation is then

$$DEPR_{\text{tot},j} = \sum_i DEPR_{i,j} \quad . \tag{21}$$

The net book value at the end of each year is found by subtracting the total depreciation in the year from the sum of the previous year's net book value and the total inflated investment for the year, that is,

$$NBV_j = NBV_{j-1} + INV'_{\text{tot},j} - DEPR_{\text{tot},j} \quad . \tag{22}$$

The residual value of each item of equipment is determined by

$$RES_i = INV'_{i,k} \cdot \frac{(LIFE_i - n + k - 1)}{LIFE_i} \tag{23}$$

where k is the year of introduction and n is the end of the year of the study period. The total residual value is then

$$RES_{\text{tot}} = \sum_i RES_i \quad . \tag{24}$$

The annual decline in assets is the decline in net book value during the year, as

$$ADA_j = NBV_{j-1} - NBV_j \tag{25}$$

and the cumulative annual decline in assets is

$$CADA_j = CADA_{j-1} + ADA_j \quad . \tag{26}$$

Present-Worth Equivalent. The present-worth equivalent of a value in year j can be found by multiplying the value by the present-worth factor, PWF_j, for that year. The present-worth equivalents for investments, O&M costs, and net book values for each year j in the study period are

$$\underline{INV}_{\text{tot},j} = INV_{\text{tot},j} \cdot PWF_j \quad (27)$$

$$\underline{OMC}_{\text{tot},j} = OMC_{\text{tot},j} \cdot PWF_j \quad (28)$$

$$\underline{NBV}_j = NBV_j \cdot PWF_{j+1} \quad . \quad (29)$$

The present-worth equivalent for the residual value, which is the value at the end of the study period n, is

$$\underline{RES}_{\text{tot}} = RES_{\text{tot}} \cdot PWF_n \quad . \quad (30)$$

The total net present worth of the system is then found by subtracting the discounted residual value of the system from the net present value of all investments and O&M costs during the study. Hence,

$$NPW_{\text{tot}} = \sum_j \underline{INV}_{\text{tot},j} + \sum_j \underline{OMC}_{\text{tot},j} - \underline{RES}_{\text{tot}} \quad . \quad (31)$$

REPLACEMENT COST

The replacement cost of the system in any year is the amount required in order to replace all satellites and earth segment equipment in place in that year with identical equipment. Earth segment replacement costs for earth station i in year j are given by

$$REPL_{i,j} = \sum_k C_k n_{k,i,j} \cdot (1 + IRATE)^{(j-IYEAR)} \quad (32)$$

where C_k is the base unit cost of equipment type k; $n_{k,i,j}$ is the number of items of equipment type k at earth station i in year j; and $IRATE$ and $IYEAR$ are the annual inflation rate and reference year for inflation, respectively. The space segment replacement cost is simply the sum of the costs of all satellites in the service region. Hence,

$$SREPL_j = \sum_k C_k n_{k,j} \cdot (1 + IRATE)^{(j-IYEAR)} \quad (33)$$

where C_k is the cost of satellite type k, and $n_{k,j}$ is the number of satellites of type k in the service region in year j.

REVENUE REQUIREMENTS

The revenue requirements algorithm generates a table that summarizes space segment and earth segment revenue requirements for each year of the study period. Certain additional inputs are required in order to account for the value of the system at the beginning of the study period, for the depreciation of the equipment, and for related operating expenses. Specifically, these inputs are as follows:

$RATE$ = required rate of return on investment
NBV_0 = net book value of the space segment at the beginning of the study period
$DEPR_{0,j}$ = depreciation in each year j of satellites in place at the beginning of the study period
IWC_j = increase in working capital for each year j
$OPER_j$ = system operating costs for each year j
$AREV_j$ = additional revenues required in each year j.

Space segment revenue requirements are determined for the entire system, including all service regions, since satellites may be shifted from one service region to another during their lifetime. Earth segment revenue requirements are calculated for each earth station in each service region. The net book value, NBV_0, for earth station equipment at the beginning of the study period is estimated for each earth station i by assuming that, on the average, all items of equipment in place at that time have depreciated to half of their original value. Hence,

$$NBV_{0,i} = \frac{\sum_k C_k n_{k,i}}{2} \quad (34)$$

where C_k is the cost of equipment type k, and $n_{k,i}$ is the number of items of equipment type k in place at earth station i at the beginning of the study period. Depreciation of this equipment in each year j is then found by summing the depreciation per year for all items of each equipment type throughout their lifetimes, as

$$DEPR_{0,i,j} = \frac{\sum_k C_k n_{k,i}}{LIFE_k} \quad , \; 1 \le j \le \frac{LIFE_k}{2} \quad . \quad (35)$$

Operating costs, $OPER_j$, are obtained from the O&M factor in the same manner as for net present-worth evaluation (see above). The increase in working capital and the additional revenues are neglected in determining earth segment revenue requirements.

Revenue requirements are calculated as follows. First, the net book value, NBV_j, at the end of each year j is calculated based on equipment additions occurring within the economic period (see above). The total net book value is then obtained by adding the net book value of equipment in place and the yearly increase in working capital, as

$$NBV_{\text{tot},j} = NBV_j + NBV_0 + IWC_j \quad . \quad (36)$$

The average net book value during each year is then

$$ANBV_{\text{tot},j} = \frac{NBV_{\text{tot},j} + NBV_{\text{tot},j-1}}{2} \quad . \tag{37}$$

The return requirement is equal to the rate of return times the average net book value during the year. Hence,

$$RET_j = RATE \cdot ANBV_{\text{tot},j} \quad . \tag{38}$$

Revenue requirements for each year are then simply the sum of the return requirement, operating expenses, and depreciation. The total depreciation in each year j is the sum of the depreciation due to equipment in place at the beginning of the study period and that added during the period. Hence,

$$DEPR_{\text{tot},j} = DEPR_{0,j} + DEPR_j \tag{39}$$

and the revenue requirements in year j are

$$REV_j = RET_j + OPER_j + DEPR_{\text{tot},j} \quad . \tag{40}$$

Conclusions

An interactive model was developed using a standard graphics terminal, and advanced interface features were successfully implemented. The resulting software provides system planners with numerous features for storing, manipulating, and analyzing the large amount of data required. The model has already been successfully applied to an actual system planning task and is currently being used in other applications.

The successful application of computer models will be invaluable for evaluating alternatives for the satellite communications systems of the future and for meeting the demands of changing communications requirements. It is expected that CSPM will evolve and expand to meet changing requirements for satellite system planning.

Acknowledgments

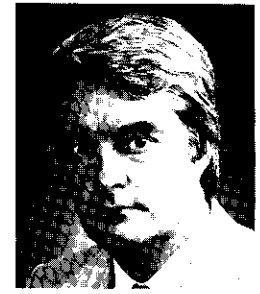
This paper is based on work performed at COMSAT Laboratories for the World Systems Division of the Communications Satellite Corporation. Contributions during the development of CSPM by W. Schnicke, E. Faine, C. King, D. Opiekun, S. Gambhir, and M. Stennett are especially appreciated.

References

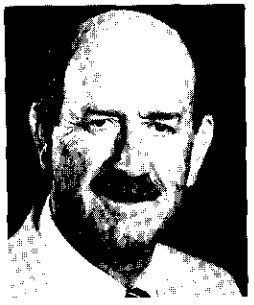
[1] B. A. Myers, "User-Interface Tools: Introduction and Survey," *IEEE Software*, Vol. 6, No. 1, January 1989, pp. 15–23.

[2] American Telephone & Telegraph Co., *Engineering Economy*, 2nd ed., New York: American Telephone & Telegraph, 1967.

Jerome L. Wernimont received a B.S. in computer science with a minor in mathematics from Virginia Polytechnic Institute in 1985. He is currently a Member of the Technical Staff in the System Development Division at COMSAT Laboratories, where his current work includes the research and development of a user interface management system, as well as the continued development of CSPM. He has also been responsible for the design and development of several interactive analysis programs and the user interface to a network monitoring and control system. Prior to joining COMSAT in 1986, Mr. Wernimont was a Programmer/Analyst for Satellite Business Systems.



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, and architectural design and software engineering for complex computer-based distributed systems. While at COMSAT, Dr. Cook has been responsible for the development of 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, he was a structural engineer at NASA's Goddard Space Flight Center and was a member of the group responsible for developing the NASA Structural Analysis Program (NASTRAN).



Geoffrey J. H. Brown received a B.A. degree in mathematical statistics from the George Washington University in 1968, and has completed post-graduate courses in computer science at the University of Maryland and the George Washington University. He joined COMSAT Laboratories in 1981 as a Member of the Technical Staff in the System Development Division, where he held the position of Project Manager at the time of his retirement in 1988. He was involved in the design, implementation, testing, and installation of the INTELSAT Operations Control TDMA Facility (IOCTF), and in developing several software products for COMSAT Maritime Services and INTELSAT System Services. He also developed a comprehensive set of test procedures for use with the INTELSAT TDMA (Earth Station) Terminal Simulator. Mr. Brown is currently retained as a consultant by the COMSAT World Systems Division. He is a Member of IEEE and the IEEE Computer Society.

Index: computer communications, earth terminals, multiple access, VSATs, protocols

Random access with notification—A new multiple-access scheme for VSAT networks

D. M. CHITRE, A. C. BRIANÇON, AND R. KOHLI

(Manuscript received June 21, 1988)

Abstract

A second-generation, multiple-access method for very small aperture terminal (VSAT) satellite networks is described and analyzed. The method, called random access with notification (RAN), has two main features that distinguish it from the slotted ALOHA multiple-access protocol commonly used for VSAT networks. First, RAN allows the hub station to identify collided packets and remove them from the random-access (contention) data slots to reserved slots dynamically allocated for their retransmission, thereby eliminating the major source of congestion and instability experienced in the slotted ALOHA scheme. Second, RAN can monitor traffic activity and adapt system parameters to traffic changes, including transitioning to a full-reservation multiple-access scheme under certain traffic conditions. A Markov model is developed for the RAN protocol and solved exactly to derive the throughput and delay expressions. Numerical results are presented showing the performance advantages of the RAN protocol compared to the slotted ALOHA and reservation protocols.

Introduction

Current very small aperture terminal (VSAT) networks are designed primarily for certain kinds of data communications applications. Typically, user networks consist of a number of distributed terminals (*e.g.*, IBM 3274 cluster controllers) accessing a central host computer or a front-end processor (*e.g.*,

an IBM 3705). The communication paths form a star connectivity, with no requirement for direct communications between the individual terminals. This connectivity requirement is exploited by designing an inexpensive VSAT terminal that communicates with a larger hub terminal. In a typical VSAT network, a single outbound carrier is used in a broadcast mode to transfer data from the hub to the VSATs, and one or more lower bit rate carriers operating in one of the first-generation multiple-access methods described below are used for traffic from the VSATs to the hub.

The performance and cost of VSAT networks are very sensitive to the multiple-access scheme selected. For a given grade-of-service delay requirement and a given traffic profile, the characteristics of the multiple-access protocol determine the space segment requirement, which in turn has an impact on both the space segment cost of the satellite channels and the cost incurred at the hub for managing and maintaining these channels.

In a large number of data applications, the duty cycle of the traffic generated by the data terminal is very low, often less than 5 to 10 percent. The methods applicable in VSAT systems for sharing satellite channels among a number of users with bursty packet arrival characteristics fall into four broad categories:

- Code-division multiple access (CDMA)
- Packet reservation time-division multiple access (TDMA)
- Fixed-allocation TDMA
- Slotted ALOHA

In the CDMA scheme, VSATs share the same inbound carrier by using a unique pseudorandom noise (PN) sequence of symbols to represent each bit of data to be transmitted. In many implementations, this method results in lower transponder utilization [1]; however, it is well suited for lower data rate applications and for applications that require spectrum spreading for frequency coordination and to meet interference restrictions (*e.g.*, C-band VSAT networks).

In reservation TDMA systems, the VSAT sends a request to the central hub for a data slot, and the data packet is transmitted only after a data slot assignment is made by the hub and received by the VSAT. Packet reservation TDMA has an inherent delay of three satellite hops for a data packet to reach the central hub from the VSAT. This access protocol also requires space segment capacity for the inbound reservation traffic.

The fixed-allocation TDMA method is quite inefficient for bursty data traffic with low (5 to 10 percent) duty cycle.

In the slotted ALOHA protocol, satellite channels are accessed by a VSAT in randomly selected time slots. A packet is retransmitted after a time-out

period for the packet (indicating no reception of an acknowledgment for that packet from the hub) and by randomizing the retransmission time. The slotted ALOHA scheme has low response delay characteristics, as long as shared random-access channels are operating at a throughput of less than 25 to 30 percent. However, this scheme has a well-known instability region where the throughput (successfully carried packet traffic) decreases with an increase in the offered traffic.

The main cause of throughput and delay performance degradation in the slotted ALOHA protocol is that collided and errored packets are retransmitted in data slots that are also used for the transmission of new packets. Numerous variations on the basic slotted random-access protocol have been presented in the literature [2],[3]; however, most of these schemes either do not address the stability problem or are not directly applicable to the VSAT networks because they typically require that each VSAT listen to its transmitting channel [4],[5].

The problem of instability is resolved in a new multiple-access scheme called random access with notification (RAN). In RAN, data channels have two types of slots: random-access (RA) and reservation (RES). Collided and errored packets are prevented from accessing the RA (contention) data slots in their retransmission attempts. New packets use RA slots for their initial transmissions. Over a separate signaling channel, every VSAT notifies the hub of the number of packets transmitted by the VSAT in a particular frame. If a packet experiences a collision, the hub recognizes that the number of packets successfully received is less than the number that were reported as sent. The hub then allots the required number of slots to the VSAT in some future frame for contention-free retransmission. The hub also monitors the total traffic rate generated by the VSAT terminals. For a given response delay requirement, the hub determines the key system parameters of the RAN protocol, namely the mix of RA and RES slots. When the traffic level exceeds a certain intensity (packets per second), the hub issues a command to all VSATs to switch over from the RAN protocol to full-reservation protocol. The hub orders a switch back to RAN after the traffic load has fallen to acceptable levels.

The scheduling of the collided and errored packets and the adaptation of the system parameters and protocols controlled by the hub for the entire network are the key features of the RAN protocol and are presented in detail in the next section. A Markov model is then developed for the RAN protocol and solved exactly to derive the throughput and delay expressions for an arbitrary distribution of packet arrivals. The effect of bit error rate (BER) on the operation of the protocol is also considered. Finally, numerical results are presented for typical frame and packet parameters.

Random access with notification

In this section, the basic assumptions and channel structure are presented, followed by a description and detailed analysis of the protocol.

Three basic assumptions are made about the network. First, the network is assumed to consist of a number of VSATs communicating with a hub. Second, a VSAT is unable to directly receive the transmissions of other VSATs; it can only receive transmissions from the hub. Finally, a VSAT is unable to listen to its own transmissions, and thus cannot determine whether a packet that it is transmitting is colliding with another packet.

For communications between the VSATs and the hub, the following features are required by protocol:

a. Communication is accomplished using a number of channels. Each channel consists of frames, and each frame consists of slots.

b. Communication from the hub to the VSAT is on a contention-free channel.

c. Communications from the VSATs to the hub are carried on two types of channels, which can be on the same frequency or on different frequencies:

- *Signaling channels* carry control packets. Every VSAT must have one contention-free signaling slot allotted to it in every frame period.
- *Data channels* carry user traffic and have two types of slots: RA and RES. Data packets first attempt transmission using RA slots. If there is a collision, the packet is retransmitted via RES slots. As an example, Figure 1 shows these channels on two different frequencies. The procedure is explained in detail below.

Protocol

The following is a detailed description of the RAN protocol and its operation.

VSAT TRANSMISSIONS

All initial packets are transmitted using RA slots, in the following manner. After a packet is generated by a VSAT, if the next data slot is an RA slot and does not overlap in time with the signaling slot for the VSAT, then the packet is scheduled in that RA slot. (This restriction is required if a VSAT cannot transmit on two frequencies simultaneously.) When there is an overlap, the signaling message transmission is given higher priority and the data packet is transmitted in the next RA slot. If the next data slot is not an RA slot, the data packet is randomly scheduled (see Figure 2) among the next cluster of

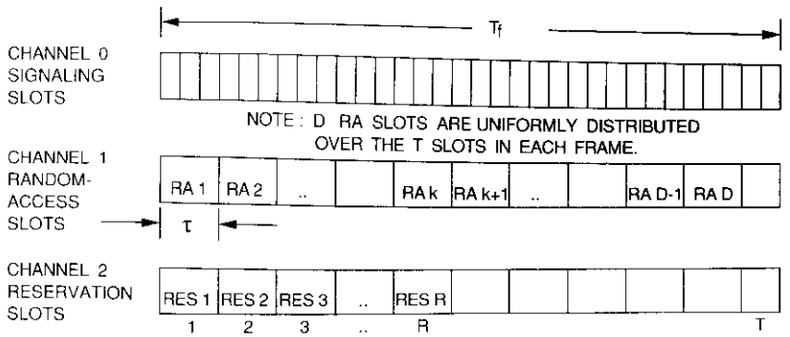


Figure 1. *Channels in RAN*

RA slots, excluding those slots that have already been assigned a packet by the same VSAT.

The VSAT keeps a copy of the data packet, along with a record of the frame number and time slot number used for transmission. This information allows the VSAT to resolve cases in which it sent several packets within a frame and some of those packets collided, or were lost because of bit errors. Based on information from the hub regarding successfully received packets, their frame numbers, and time slot numbers, the VSAT can determine which packets collided or were lost.

Data packet transmission in the RA slots is accompanied by a notification in a VSAT's signaling slot in the next frame. More precisely, the signaling message sent by a VSAT in one frame notifies the hub of the number of packets transmitted in the RA slots in the immediately preceding frame.

Once a data packet is successfully received at the hub, it is removed from the retransmission buffer. If it is not successfully received, it must be retransmitted in a RES slot. The manner in which successful reception is determined is described below.

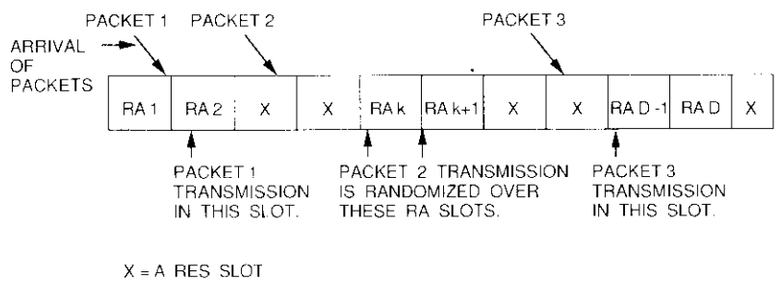


Figure 2. *Transmission Procedures for Data Packets*

Hub Processing of Data and Signaling Channels

The hub processes the information received on the signaling channel and the data channels. The hub is informed, via the signaling channel, of the number of transmissions made by each VSAT in the previous frame, and it determines the number of packets received from VSATs by monitoring the data channel. The hub then correlates these two sets of information to compute the number of packets lost (due to collision, bit errors, *etc.*) for each VSAT and schedules RES slots for their retransmission. The RES slots are randomly ordered and scheduled among the different VSATs that have lost packets which require retransmission, taking into account any processing and propagation delays that may be required before the VSATs can use the information to retransmit.

The hub transmits the following information on the return channel from the hub to the VSATs, which then determine the packets appropriate for retransmission:

- Acknowledgments of successful packets, indicating the slot numbers and frame numbers over which they were received.
- Scheduling information (the location of the RES slots) for the lost packets.

These two pieces of information are sufficient for the VSAT to determine which transmitted packets were lost and which RES slots should be used to retransmit them. The RES slots are located in the frame such that the scheduling messages from the hub are received at the VSAT before the start of the RES slots. A time margin is required in order to allow the VSAT to decode and process the information.

System Reservation Mode

In the RAN mode, the hub computes the number of packets, H, that were lost in each frame when transmitted in RA slots. Whenever that number exceeds the number of RES slots per frame (R) for a certain number of consecutive frames, the hub broadcasts a message to the VSATs (repeated a certain number of times) announcing a switchover from RAN to reservation TDMA, starting with a specific frame number. In reservation TDMA, all data slots (RA slots and RES slots) are accessed only as scheduled by the hub.

When the system is operating in the reservation TDMA mode, the hub computes the number of packet reservation requests per frame. Whenever, for a certain number of consecutive frames, the number of packet reservation requests falls below a prespecified percentage of the total number of data slots, T ($T = D + R$, where D is the number of RA slots), the hub first

determines the total network packet traffic per second from the VSATs to the hub, based on the average over the last few frames. Next, the hub computes the combination of D and R for the given number of data slots (T) and the above-determined VSAT packet traffic to give minimum average delay. The hub also decides at which frame number the switchover from reservation TDMA to RAN will take place and broadcasts that information and the location of RA slots and RES slots to all VSATs. The broadcast is repeated several times to ensure that all VSATs receive the information correctly. The exact number of repeats can also be determined by examining the probability of a VSAT not receiving the broadcast message *vs* the number of repeats.

Throughput and delay analysis

A detailed analysis of the RAN protocol can be carried out for any fixed combination of RA and RES slots per frame. The packet arrival statistics (*i.e.*, the total offered load) are assumed to be distributed uniformly over all the RA slots within a frame. The following variables define some of the system parameters:

D = number of RA slots per frame
R = number of RES slots per frame
G = offered load (packets) per RA slot
T_p = propagation delay (seconds)
T_f = frame period (seconds)
τ = data packet transmission time (seconds)

Markovian Model for Hub Scheduling

A Markov model was developed for the RAN protocol and solved exactly to derive the throughput and delay expressions for an arbitrary distribution of packet arrivals. Numerical results are presented in the next section for the special case of a Poisson distribution of packet arrivals within a frame period (to be distributed uniformly among the RA packet slots).

As before, let D denote the number of data slots per frame which are to be accessed by VSATs in a random-access mode for the transmission of new packets (RA slots). Let R denote the maximum number of data slots per frame which are used to schedule the retransmission of collided packets (RES slots). Consider the scheduling state at the hub following processing of the notifications of the $(k + 1)$th frame and correlation of that information with the data packets received in the RA slots of the kth frame. The hub schedules m collided packets ($m = 0, 1, 2, \ldots, R$) in some appropriate data frame I, where $I > k + 1$. Frame I does not directly follow frame $k + 1$ because

of the propagation delay between the hub and the VSAT, and because of the delay needed to perform the numerous computations required at the VSAT. Moreover, if the number of packets scheduled in frame I equals R, scheduling must be continued for the possible remaining collided packets, i. These packets are scheduled in the data frames subsequent to frame I. In other words, m is the number of packets scheduled for transmission in the R RES slots in data frame I ($0 \leq m \leq R$), and i is the number of packets still queued and awaiting scheduling in a subsequent frame for retransmission after m packets have been scheduled in frame I ($i = 0$ for $m < R$).

The Markov state is defined as the ordered pair (m, i) with the associated probability of $\Pi(m, i)$ where $m = 0, 1, 2, \ldots, R$; $i = 0, 1, 2, \ldots$; and $m = R$ for $i \geq 1$. For the steady state, the Markov equations are

$$\Pi(0, 0) = \Pi(0, 0)\, p(0) + \sum_{m=1}^{R} \Pi(m, 0)\, p(0) \tag{1}$$

$$\Pi(m, 0) = \sum_{n=0}^{R} \Pi(n, 0)\, p(m) + \sum_{i=1}^{m} \Pi(R, i)\, p(m - i) \qquad m = 1, 2, \ldots, R \tag{2}$$

$$\Pi(R, i) = \sum_{n=0}^{R} \Pi(n, 0)\, p(R + i) + \sum_{j=1}^{i+R} \Pi(R, j)\, p(i + R - j) \qquad i \geq 1 \tag{3}$$

where $p(i)$ denotes the probability that the hub received notifications for i lost packets transmitted in a single frame. The moment generating function $F(z)$ can be defined as

$$F(z) = \sum_{i=0}^{\infty} \Pi(R, i) z^i \quad .$$

Multiplying the right- and left-hand sides of equation (3) by z^i, summing over all i, and using equations (1) and (2) provides the generating function $F(z)$ for $\Pi(m, i)$,

$$F(z) = \frac{g(z) \sum_{m=0}^{R-1} \Pi(m, 0) - \sum_{i=0}^{R-1} \Pi(i, 0)\, z^i}{z^R - g(z)} \tag{4}$$

where the function $g(z)$ is defined by the series

$$g(z) = \sum_{i=0}^{\infty} p(i)\, z^i \quad . \tag{5}$$

Evaluating equation (4) at $z = 1$ yields the identity

$$\sum_{m=0}^{R-1} (R - m)\, \Pi(m, 0) = R - g'(1) \tag{6}$$

where the prime denotes the derivative with respect to z. Since $F(z)$ is an analytic function inside the unit disk ($|z| < 1$), the roots of $z^R - g(z) = 0$ must be the zeros of the numerator of equation (4). This fact yields a set of $R - 1$ equations for $\Pi(m, 0)$, $m = 0, \ldots, R - 1$. These equations, coupled with equation (6), can be solved to obtain $\Pi(m, 0)$. The probabilities $\Pi(R, i)$ can then be obtained from equation (4) by appropriately differentiating $F(z)$ and evaluating it at $z = 0$. The following recursion relationships can then be obtained:

$$\Pi(R, 0) = \frac{\Pi(0, 0) - g(0) \sum_{i=0}^{R-1} \Pi(m, 0)}{g(0)} \tag{7a}$$

$$\Pi(R, n) = \frac{\Pi(0, 0) - g^{(n)}(0) \sum_{i=0}^{R-1} \Pi(m, 0)}{n!} - \frac{\sum_{i=1}^{n-1} g^{(n-i)}(0) \Pi(R, i)}{(n - i)!} \qquad \text{for } i \leq n \leq R \tag{7b}$$

$$\Pi(R, n) = \frac{\Pi(R, n - R) - g^{(n)}(0) \sum_{i=0}^{R-1} \Pi(m, 0)}{n!} - \frac{\sum_{i=1}^{n-1} g^{(n-i)}(0) \Pi(R, i)}{(n - i)!} \qquad \text{for } n > R \tag{7c}$$

where $g^{(n)}(0)$ is the nth derivative of $g(z)$ evaluated at $z = 0$. Note that the above model is independent of the input traffic model or the mode of notification access. The probabilities $p(i)$, and hence $g(z)$, are computed explicitly for a specific input traffic model and notification access scheme.

The roots of the denominator of equation (4), which will determine the first R equations for $\Pi(m, 0)$, are computed using an iterative Fletcher-Powell-Davidon search in the complex plane to obtain a rough estimate of

the roots' locations, followed by a more traditional Newton-Raphson search to improve these estimates [6]. The values for $\Pi(m, 0)$ are then found by solving a set of R linear equations.

Average throughput efficiency and delay can now be computed from these generating functions.

Throughput Analysis

The steady-state distribution allows the throughput of the RAN protocol to be computed as a function of the traffic load, G. The average number, N_R, of packets transmitted in the R RES slots per frame is given by

$$N_R = \sum_{m=0}^{R-1} m\Pi(m, 0) + R\sum_{i=0}^{\infty}\Pi(R, i) = g'(1) \quad \text{[using equation (6)]}. \tag{8}$$

Assuming that input traffic follows a Poisson distribution with an average of G packets for each of the D RA slots, the average throughput efficiency, η, of the $D + R$ slots is given by

$$\eta = \frac{DGe^{-G} + N_R}{D + R} = \frac{DGe^{-G} + g'(1)}{D + R} \quad . \tag{9}$$

The first term of the numerator is the number of packets that do not experience collisions when submitted to the channel, and N_R is given by equation (8). Furthermore, if the notifications are assumed to be error-free, then

$$g(z) = [e^{-G}e^{Gz} + Ge^{-G}(1 - z)]^D \quad . \tag{10}$$

This result is found by first computing the generating function for the number of collided packets in each data slot. The probability distribution function for collisions involving k packets is given by

$$\begin{aligned} p\,(0\,\text{collision}) &= e^{-G} + Ge^{-G} \\ p\,(1\,\text{collision}) &= 0 \\ p\,(k\,\text{collision}) &= \frac{e^{-G}G^k}{k!} \quad . \end{aligned} \tag{11}$$

Substituting the value of $g'(1)$ into equation (8) yields the expected result for the throughput efficiency:

$$\eta = \frac{DG}{D + R} \quad . \tag{12}$$

Delay Analysis

To obtain an expression for the probability density function (pdf) of the total delay that a packet experiences, the pdf of delay for both collided and noncollided packets must be computed and then combined.

When a packet does not collide during the VSAT-to-hub transmission, the delay is simply given by

$$T_{nc} = T_p + \tau \quad . \tag{13}$$

Note that this delay is a constant, independent of the traffic load (G). The probability of no collision, P_s, is given by $P_s = e^{-G}$.

When a packet does collide, its delay is influenced by a number of factors. The following procedure must be carried out before a collided packet is retransmitted. Every packet sends a notification of transmission in its assigned notification slot. This is done on the notification channel in the frame that is the immediate successor to the frame in which the data packet was transmitted. A frame of notifications is received at the hub after T_p seconds. The hub processes the frame of notifications, determines which data packets collided, and schedules the retransmission of these packets randomly after scheduling the retransmission of queued packets that collided in previous frames.

Figure 3 is an example of the delay experienced by a collided packet in RAN. In this case, the RES slot is shown in the first frame after the VSAT is informed by the hub. While the system is designed to ensure that this is true in most situations, the hub can also assign slots for subsequent frames.

Figure 3 shows that the delay consists of a number of independent components. The first component depends on where in the frame the packet collided, while the second consists of the frame delay introduced by the fact that the hub must wait for all the notifications before doing any scheduling. The number of packets that have already been scheduled due to collisions in previous frames must also be addressed. This probability distribution is given by $\Pi(m, i)$, as defined above.

The number of packets in the frame under consideration must also be taken into account. It is necessary to know where the collided packet is likely to be scheduled among all the collided packets in the original frame. Finally, the other delay components indicated in Figure 3 must be added. The following subsections examine the different parameters and delays involved in this computation.

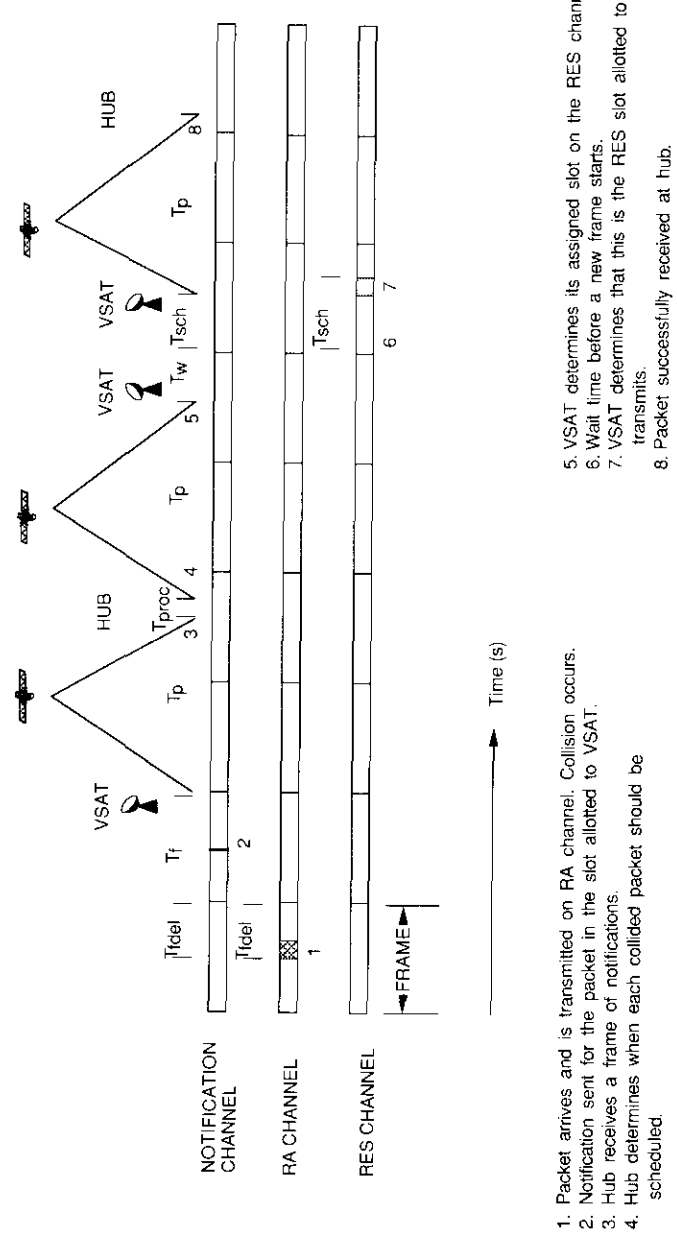


Figure 3. *Delay Experienced by a Collided Packet in RAN*

Position in Frame. Since Poisson arrivals are assumed, a packet's arrival time is uniformly distributed over the frame time. Because there are D RA slots, the probability that a packet collided in the $(D - k)$th slot in the frame, and the consequent delay, are given by

$$\begin{aligned} P_f(k) &= 1/D \\ T_{fdel} &= (k+1)\tau \qquad k = 0, 1, 2 \ldots, D-1 \end{aligned} \tag{14}$$

Scheduling Delay. The scheduling delay, T_{sch}, is determined by two factors. The first is the number of packets that collided prior to this frame (*i.e.*, the frame in which the collision of interest occurred) and are queued to be scheduled for retransmission. The second is the scheduling location of the packet of interest among the packets that collided in the current frame. The sum of the prior frame scheduling and the current scheduling is the overall scheduling delay.

Prior Scheduling. The probability of the number of packets remaining from previous collisions which are queued for retransmission, $P_{ps}(k)$, is obtained, by definition, from the second index of the state probabilities computed above, as follows:

$$P_{ps}(0) = \sum_{i=0}^{R} \Pi(i, 0) \tag{15}$$

$$P_{ps}(k) = \Pi(R, k) \qquad k = 1, 2, 3, \ldots$$

Current Frame Scheduling. To determine when a packet that collided will be scheduled relative to others that collided in the same frame, it is necessary to first determine how many packets collided in that frame. The approach taken is to examine the performance of an artificially introduced packet (test packet). This packet would collide only if at least one other packet was transmitted in the same slot. Since arrivals are Poisson-distributed, the probability of k collisions in the slot of the test packet is given by

$$P_{cs}(k) = \frac{G^{k-1} e^{-G}}{(k-1)!} \qquad k = 2, 3, 4, \ldots \tag{16}$$

Note that k collisions occur if there are only $k-1$ arrivals, because it is assumed that the test packet has already arrived. In the $k = 1$ case, the test packet is the only packet in the slot, and hence corresponds to the collision-free case mentioned above.

The hub must now resolve the collisions in the slot of the collided packet, as well as those that occur in the other $D-1$ slots. It can be shown that the probability of n collisions in $D-1$ slots, $P_{D-1}(n)$, is given by

$$P_{D-1}(n) = e^{G(D-1)} \sum_{k=0}^{D-1} \binom{D-1}{k} G^k \sum_{i=0}^{\min(n,k)} i \binom{n}{k} \times |G(D-1-k)|^{n-i} (-1)^i i! \binom{k}{i} \quad (17)$$

where $\binom{k}{i}$ equals $\dfrac{k!}{(k-i)!i!}$.

The total number of collisions that the hub must resolve when the test packet has collided is the sum of the collisions in the slot of the collided packet and in the other $D-1$ slots. Since these are two independent processes, the probability distribution for the total number of collisions in the frame, $P_{cf}(k)$, is obtained by convolving the two distributions. Hence,

$$P_{cf}(k) = P_{D-1} * P_{cs} \quad (18)$$

where $*$ represents convolution.

If k packets had collided, the probability that the test packet is scheduled as the jth packet among the k packets is simply $1/k$. This follows from the fact that collided packets are randomly scheduled. To determine the probability that the collided test packet will be the jth scheduled packet, the conditioning on the number of collisions must be removed. Therefore, the probability, $P_{sch}(j)$, that the test packet is the jth scheduled packet among the collided packets is given by

$$P_{sch}(j) = \sum_{k=j}^{\infty} \frac{P_{cf}(k)}{k} \quad . \quad (19)$$

The summation begins from j, since a packet can only be scheduled as the jth packet among the collided packets if there are at least j packets to begin with.

Overall Scheduling. The above calculations have determined the probability distributions, $P_{ps}(k)$, of the packets already scheduled due to collisions in previous frames, and $P_{sch}(k)$ of the relative position of the test packet among the packets that collided in its own frame. A collided packet will be retransmitted after all of the previously scheduled packets have been transmitted in RES slots on the reservation channel. The total number of previously scheduled packets is the sum of the prior scheduled packets and the current frame scheduled packets. Hence, the overall probability, $P_{tot}(k)$, can be obtained as follows:

$$P_{tot}(k) = P_{ps} * P_{sch} \quad . \quad (20)$$

Scheduling Delay in Real Time. The probability $P_{tot}(k)$ gives the probability distribution of having k packets scheduled in front of the test packet; however, this number must still be translated into seconds. The hub will inform the VSAT that it must transmit after $k-1$ RES slots (*i.e.*, in the kth RES slot). The frame size is assumed such that the next RES slot appears almost immediately after the hub has informed each VSAT of when it must transmit. Alternatively, the RES slots can be placed so that they begin as soon as the information arrives from the hub. Clearly, if $k \leq R$, then all the scheduled packets can be accommodated within the next frame. However, if $k > R$, then the packet must wait for one whole frame before it can transmit. Similarly, if $k > 2R$, the packet must wait two frames, and so forth.

Mathematically, the overall scheduling delay, T_{sch}, can be written as

$$T_{sch} = \mathrm{Int}\left(\frac{k}{R}\right) \cdot T_f + \left\{\left[k - R \cdot \mathrm{Int}\left(\frac{k}{R}\right)\right]\tau\right\} \quad . \quad (21)$$

The first term of the right-hand side represents the number of frames that a collided packet must wait, while the second term indicates its position in the frame in which it transmits.

Total Delay for a Collided Packet. The delay for the collided packets can now be computed based on the various delays introduced above. From Figure 3, the total delay for a collided packet, T_{ctot}, can be written as

$$\begin{aligned} T_{ctot} &= T_{fdel} + T_f + T_p + T_{proc} + T_p + T_w + T_{sch} + T_p \quad (22) \\ &= T_{fdel} + T_{sch} + T_c \end{aligned}$$

where $T_c = 3T_p + T_f + T_{proc} + T_w$ is a constant. Because T_{fdel} and T_{sch} are independent random variables whose probability distribution has already been obtained, the pdf for T_{ctot} can again be found by convolution of the two densities. The variables T_{ctot} and T_{nc} can now be substituted into the expression for the overall delay, T_{del}.

Overall Total Delay. The probability distribution for the overall total delay (T_{del}) that a packet experiences, whether or not it collides, is obtained by

combining two known probabilities: the probability for delay when a packet goes through without a collision (P_s); and the probability, after it collides, that it is scheduled kth [$P_{sch}(k)$; $k = 0, 1, 2, \ldots$].

ERRORED PACKETS

Thus far, the analysis has focused on the protocol itself. The effect of BERs is introduced here for completeness, although the impact of typical BERs on overall throughput is marginal.

If a packet collides, the protocol successfully retransmits using the notification procedure. However, successful retransmission may be prevented by any of the following events:

- The notification from the VSAT to the hub is lost due to bit errors or other factors.
- The notification is not lost, but the packet containing the hub's allocation of the RES slots is lost.
- Neither the notification nor the allocation packets are lost, but the data packet picks an error while being retransmitted in the RES slot.

If any of these situations occurs, the data packet will be lost and a higher layer of protocol will have to recover. Let

B_1 = BER on data channels (RA and RES slots)
B_2 = BER on the signaling channel from VSAT to hub
B_3 = BER on the signaling channel from hub to VSAT
L_1 = length of data packet
L_2 = length of notification packet from VSAT to hub
L_3 = length of allocation packet from hub to VSAT

The following probabilities can then be written:

P_{NED} = P(No error in transmission of data packet) = $(1 - B_1)^{L_1}$
P_{ED} = P(Error in transmission of data packet) = $1 - (1 - B_1)^{L_1}$
P_{NEN} = P(No error in notification) = $(1 - B_2)^{L_2}$
P_{EN} = P(Error in notification) = $1 - (1 - B_2)^{L_2}$
P_{NEA} = P(No error in allocation) = $(1 - B_3)^{L_3}$
P_{EA} = P(Error in allocation) = $1 - (1 - B_3)^{L_3}$

Let the probability that a packet is unsuccessful in its first attempt on an RA slot be P_u. Without considering the error probabilities, $P_u = 1 - e^{-G}$. However, taking errors into account, $P_u = (1 - e^{-G}) + P_{ED} * e^{-G}$.

A packet will be lost if it is unsuccessful in its first attempt and one of the three events indicated above occurs. Given the independence of the events, the probability that a packet will be lost (P_{lost}) is given by

$$P_{lost} = P_u * (P_{EN} + P_{NEN} * P_{EA} + P_{NEN} * P_{NEA} * P_{ED})$$

Typically, the notification packet is an order of magnitude smaller than either the data packet or the allocation packet. Therefore, it is assumed that $P_{NEN} \approx 1$ and $P_{EN} \approx 0$. Further, assuming that $G = 0.6$, $B_1 = 10^{-7}$, $B_3 = 10^{-7}$, $L_1 = 128$ bytes, and $L_3 = 128$ bytes, then $P_{lost} = 9.5 \times 10^{-5}$. Note that the effect of BER is negligible compared with the efficiency of the protocol over error-free channels. Furthermore, even if the notification or the allocation packet, or both, are lost, the protocol is stable and will not be affected beyond the loss of the packet(s) concerned.

Results

Three representative graphs are presented. Figure 4 shows the variation in mean delay with the number of RA slots at a fixed throughput. The throughput is fixed by considering a fixed packet arrival rate and a fixed number of total data slots ($T = D + R$). In the specific example considered here, the packet arrival rate is taken to be 30 pkts/s and the total number of data slots is 34. As the number of RA slots (D) is increased, the delay begins to decrease. This is because more slots are available for random-access transmission for the same number of packets, and consequently fewer collisions occur. At a certain combination of D and R (in this figure, at $D = 25$, $R = 34 - 25 = 9$) the delay reaches its minimum and then begins to increase once again. This is because the number of RES slots is now inadequate to handle even the limited number of collisions that occur in the larger number of contention slots.

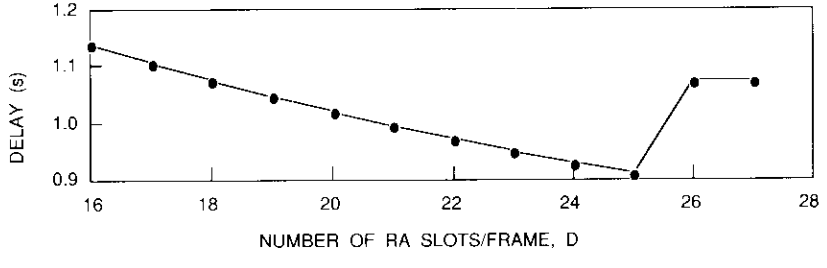


Figure 4. *RAN Mean Delay vs Number of RA Slots (Propagation delay = 270 ms, frame size = 512 ms, packet arrival rate = 30 pkts/s, packet length = 128 bytes, channel rate = 56 kbit/s, total slots = 34, throughput = 0.45)*

Figure 5 shows the delay throughput tradeoff for system design. Every point on this graph was obtained by first plotting graphs similar to the one in Figure 4 at a fixed throughput. As shown in Figure 4, for a throughput of 45 percent ($D + R = 34$), the combination that gives the minimum mean delay is $D = 25$, $R = 9$. This is optimum at a mean packet arrival rate of 30 pkts/s, as reflected in Figure 5. The procedure was repeated for different throughputs, which were modeled by keeping the packet arrival rate fixed at 30 pkts/s and varying the total number of data slots allocated for the traffic. The resulting graph is shown in Figure 5. As the total number of allocated data slots decreases, the throughput increases, but so does the mean delay. At a throughput of around 55 percent, the mean delay increases to a value greater than that of the reservation protocol, indicating that there is no benefit in designing a system that uses the RAN protocol at those throughput levels. However, for throughput values less than 55 percent, RAN provides a mean delay which is less than that available from the reservation method.

Assuming that it is decided to operate at a throughput of 45 percent for a packet arrival rate of 30 pkts/s, the designer would also like to see the projected variations in performance for different loads (traffic arrival rates) when the system is operational. This is reflected in Figure 6. Note that as the load drops, the number of collisions decreases, and therefore the mean delay decreases. Conversely, as the load increases above 30 pkts/s, the delay begins to increase. At a value of around 35 or 36 pkts/s, the RAN protocol will switch over to the reservation mode. Therefore, RAN can continue to provide mean delay performance better than the reservation protocol, even when the load offered exceeds the designed load by 15 or 20 percent.

In summary, Figure 4 indicates the best combination of D and R to provide acceptable delay at a given throughput for a given packet arrival rate.

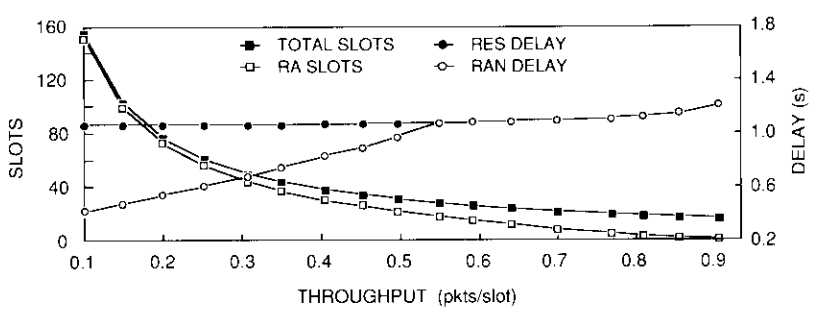


Figure 5. *Average Delay vs Throughput (Propagation delay = 270 ms, frame size = 512 ms, packet arrival rate = 30 pkts/s, packet length = 128 bytes, channel rate = 56 kbit/s)*

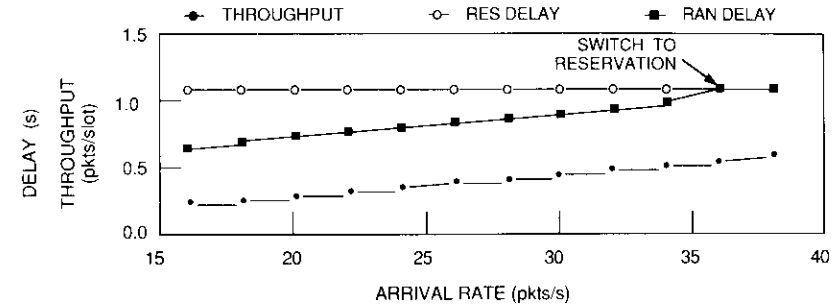


Figure 6. *Average Delay and Arrival Rate Variation Around Operating Point (Propagation delay = 270 ms, frame size = 512 ms, packet arrival rate = 30 pkts/s, packet length = 128 bytes, channel rate = 56 kbit/s, total slots = 34, RA slots = 25)*

Figure 5 shows the delay throughput characteristic for a given packet arrival rate. It provides parameters (D and R), and consequently different systems for different performance requirements. While Figure 5 is not an operating curve, it may be viewed as an encapsulation of different curves such as those in Figure 4.

Figure 6 is an operating curve which shows the variations in mean delay with packet arrival rate when the system is designed for a given packet arrival rate. It also shows the point at which a system in operation should switch over to the reservation mode.

Comparison with existing schemes

Comparison of the RAN and slotted ALOHA protocols is meaningful only when the signaling overhead of the RAN scheme is factored into the throughput. Because the RAN and reservation protocols require the same signaling overhead, the two schemes can be compared directly.

Reservation

The RAN and reservation schemes require the same overhead. In RAN, every time a VSAT transmits a certain number of packets in a frame, it must notify the hub of the number of transmissions by using the subsequent frame. In the reservation scheme, every time a VSAT needs to transmit a certain number of packets, it requests that the hub allocate the number of slots required. The same information is transmitted, but the difference is twofold:

a. In RAN, the signaling information is transmitted after the data packets have been transmitted, while in the reservation scheme the information is transmitted before the data packets.

b. In RAN, the signaling information serves to notify the hub of the number of packets sent, whereas in the reservation protocols the signaling channel requests data slots for packets to be transmitted in some future frame.

In terms of bandwidth required, both RAN and reservation overhead is the same. The difference is in the delay computation. In RAN, when the throughput is low, the delay characteristics far outperform those of the reservation scheme because in RAN the packets do not have to go through the negotiation required by the reservation protocol. The packet will simply be put on the channel and, since the throughput is low, it is likely that the packet will arrive at the hub without colliding. When the throughput increases, the delay begins to approach that of the reservation scheme. If the RAN protocol did not switch over to reservation, all the RA slots would be wasted in collisions. However, the hub monitors the traffic and the system switches over at the appropriate traffic level. Therefore, even at high throughput, the delay in RAN is at most equal to that in reservation.

Slotted ALOHA

Unlike RAN, pure slotted ALOHA does not have any signaling traffic from the VSAT to the hub. However, for status and maintenance, even slotted ALOHA may have some signaling overhead. To make a meaningful comparison between slotted ALOHA and RAN, an estimate of the RAN signaling traffic is first necessary. If there are N VSATs and the length of a signaling slot is L_s bits, then the amount of signaling bandwidth per frame is $N * L_s$. Expressed in bits per second, this is $N * L_s/T_f$. As far as RAN is concerned, L_s need be only 1 or 2 bytes. It can be assumed that the protocol is operating at 45-percent throughput if the signaling protocol is not included. If the signaling overhead is 20 percent of the data capacity, the 45-percent throughput is effectively reduced to 45/(100 + 20) or 37.5 percent. This is still better than the maximum throughput available from slotted ALOHA, and at this throughput the expected delays of the RAN and slotted ALOHA protocols are approximately equal.

Extensions to RAN

Under certain conditions, the delay characteristics of the basic RAN protocol can be improved further by making several changes in the protocol. In this "extended" RAN, the notification for the data packet transmitted in one of the RA slots is sent as soon as possible, instead of waiting for the notification slot in the following frame, as is done in basic RAN. The notifications are processed by the hub as they are received, and the information in each

notification is correlated with the information on the RA slots during T_f seconds (the frame period) prior to the reception of that notification. This correlation is then used to determine how many packets from that VSAT need to be retransmitted and which ones have been received successfully, in what specific data slots and frame numbers. Next, the hub schedules the retransmission of the lost packets, one after another, in the first available RES slot. (It is assumed that the processing power at the hub allows for the processing of the VSATs signaling packets as they reach the hub.)

Note that the major difference between basic RAN and extended RAN is in correlating the pieces of information received by the hub over the signaling channel and the data channel. This requires constant updating of the contents of the VSAT's signaling packet.

The advantage of extended RAN over basic RAN is that the scheduling delay for a lost packet decreases by a frame period, on the average. Hence, the extended RAN protocol is particularly appropriate, from a delay viewpoint, when the frame duration is large.

A second extension would allow VSATs to permanently request a RES slot. This RES slot would then be used as a circuit until the VSAT releases it back to the hub.

Conclusions

The RAN protocol has been described and analyzed by developing a Markov model. This protocol offers better performance where the criteria are the mean delay-throughput characteristics. It does not experience the instability associated with slotted ALOHA, nor does it have the high delay characteristic of the reservation scheme. RAN is particularly appropriate for VSAT networks that have signaling channel requirements for status and maintenance. In such a situation, the additional signaling overhead needed to implement the RAN protocol is relatively small. In cases where a signaling channel is required only for RAN, the protocol still provides better performance than slotted ALOHA for networks with a moderate number of VSATs. If the number of VSATs becomes extremely large, the signaling overhead may become prohibitive.

References

[1] A. J. Viterbi, "When Not to Spread Spectrum—A Sequel," *IEEE Communications Magazine*, Vol. 23, No. 4, April 1985, pp. 12–17.

[2] F. A. Tobagi, "Multiaccess Protocols in Packet Communication Systems," *IEEE Transactions on Communications*, Vol. COM-28, No. 4, April 1980, pp. 468–488.

[3] S. Tasaka and K. Ishida, "The SRUC Protocol for Satellite Packet Communication—A Performance Analysis," *IEEE Transactions on Communications*, Vol. COM-34, No. 9, September 1986, pp. 937–945.

[4] L. Kleinrock and S. S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *IEEE Transactions on Communications*, Vol. COM-23, No. 4, April 1975, pp. 410–423.

[5] D. Raychaudhuri, "Announced Retransmission Random Access Protocols," *IEEE Transactions on Communications*, Vol. COM-33, No. 11, November 1985, pp. 1183–1190.

[6] J. M. Ortega, and W. C. Rheibold, *Interative Solution of Nonlinear Variables*, New York: Academic Press, 1970.

Dattakumar M. Chitre received his B.Sc. from the University of Bombay, India; an M.A. in mathematics from the University of Cambridge, U.K.; and a Ph.D. in physics from the University of Maryland. He is currently an Associate Executive Director of the Network Technology Division at COMSAT Laboratories. He has been involved in research and development activities in ISDN, VSAT networks, data communications, network systems and architectures, multiple-access techniques, and network control and management. He also directs and participates in standards activities for ISDN and data communications. Prior to his current position, Dr. Chitre was a Principal Scientist in the Network Technology Division at COMSAT Laboratories.

Alain C. Briançon received the Diplôme d'Ingénieur from the Ecole Supérieure d'Electricité, Gif-sur-Yvette, France, in 1982; and an M.S. in electrical engineering and computer science and Ph.D. in electrical engineering from the Massachusetts Institute of Technology in 1983 and 1986, respectively. He joined COMSAT Laboratories in 1986 as a Member of the Technical Staff in the Network Systems Department of the Network Technology Division, where he was involved in the design and analysis of congestion, flow control, and routing mechanisms for communications networks; the use of distributed artificial intelligence for network control; and ISDN protocols. He represented COMSAT at ANSI T1S1 and CCITT Study Group XI meetings. Dr. Briançon is currently a Staff Engineer with GTE Spacenet Systems Development Department, McLean, Virginia, where he is involved in the design and development of VSAT and USAT networks.

Ranjit Kohli received his Bachelor of Technology degree in electrical engineering from the Indian Institute of Technology, Dehli, in 1984, and an M.S. from Rensselaer Polytechnic Institute, Troy, New York, in 1986. He joined COMSAT Laboratories in 1986 and is currently 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.

Index: integrated circuits, microwave switching, MMICs, models, solid-state devices

Device and circuit modeling for a 3.5- to 6.5-GHz GaAs monolithic dual-gate FET switch module

M. C. Fu and R. K. Gupta

(Manuscript received June 7, 1989)

Abstract

The design approach and measured performance of a broadband (3.5- to 6.5-GHz), dual-gate field-effect transistor (DGFET) switch module with an on-chip, TTL-compatible switch control circuit are presented. This MMIC switch module is a key element in the development of a highly reliable and lightweight microwave switch matrix. A computer program was developed for modeling the equivalent circuit element values for the DGFET from physical device parameters. The program is based on a cascode analysis of two single-gate FETs and is used to predict the small-signal, three-port, S-parameters for DGFET devices. These modeled parameters are then used to design a self-biased monolithic DGFET switch module with chip dimensions of 1.5 × 2.5 mm (60 × 100 mil). Measurements show that the switch module has an ON-state gain of approximately 11 dB over the 3.5- to 6.5-GHz range, with better than 50-dB ON-to-OFF isolation. The measured gain in the ON state is within 1 dB of the modeled performance over the 3.5- to 6.5-GHz band.

Introduction

Gallium arsenide (GaAs) metal-semiconductor field-effect transistors (MESFETs) are used extensively in the design and development of monolithic microwave integrated circuits (MMICs). MMICs are especially suited for microwave subsystems and communications satellite applications because of their small size and weight, improved reliability and reproducibility, and potentially lower cost [1].

The single-gate GaAs MESFET is the most common device used in the design of MMICs. However, the dual-gate GaAs FET (DGFET) has demonstrated even higher gain and greater reverse isolation than its single-gate counterpart [2],[3]. In addition, the DGFET possesses inherently increased functional possibilities due to the presence of two independent gates. As a result, the DGFET has found widespread application in the design of circuits such as mixers, variable-phase shifters, switch circuits, and variable-gain amplifiers [4]–[10]. The DGFET is almost universally operated in the common-source mode, with the second gate terminated in an appropriate impedance. However, the presence of the additional gate has complicated device modeling to the point of hampering DGFET circuit design and analysis [11].

This paper presents a device modeling effort for GaAs DGFETs which is based on physical device parameters. A software package has been developed in which single-gate device physics models are extended to DGFET analysis by assuming a cascode configuration of two single-gate FETs with identical channel currents [12]–[14]. Given a doping profile and other device design characteristics, the program generates the extrinsic and intrinsic lumped-element equivalent circuit models and calculates DC current-voltage relationships. The program model was tested against a commercial DGFET and then used in the design of an MMIC switch module with gain in the ON state. For a given set of device physics parameters, the DGFET model was derived and small-signal S-parameters were calculated. Second-gate termination was investigated to analyze its effect on the circuit design. Lumped-element components and transmission lines were used for the matching networks. The digital control and RF functions were integrated by incorporating a transistor-transistor logic (TTL)-compatible control circuit on the chip. The overall dimensions of the self-biased MMIC switch module, with integrated control, are 1.5 × 2.5 mm (60 × 100 mil).

Figure 1 shows a comparison of the MMIC DGFET switch module with the hybrid (MIC) PIN diode switch used in the design of a 3.5- to 6.5-GHz, 8 × 8 microwave switch matrix (MSM) for satellite-switched, time-division multiple-access (SS-TDMA) communications satellites [15]. In addition to a substantial size reduction, the MMIC switch offers the advantages of enhanced reliability and reduced assembly time because several substrates and manual bondwire operations are eliminated.

DGFET Modeling

The GaAs dual-gate MESFET, first introduced by Turner [2] in 1971, is simply a single-gate FET with another channel control (a second gate) added. The dual-gate structure has found widespread application since its introduction

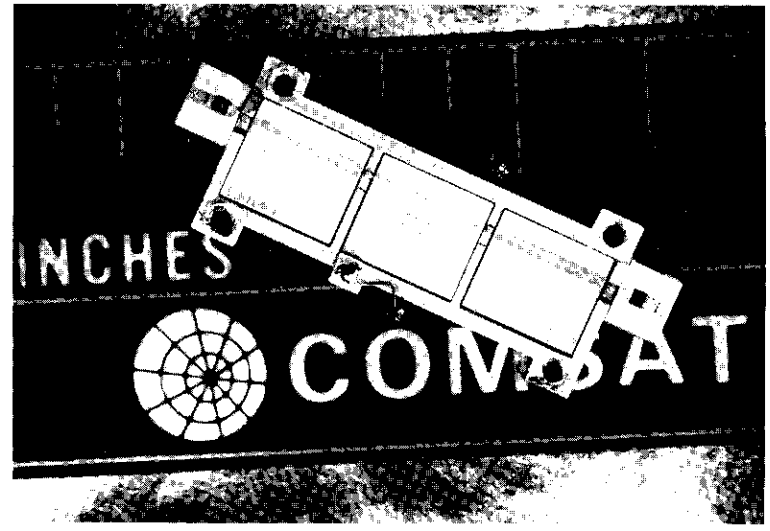


Figure 1. *Comparison of a 3.5- to 6.5-GHz PIN-Diode Switch With the MMIC DGFET Switch Module*

due to the number of advantages it holds over the conventional single-gate structure. The second gate provides additional functional possibilities in terms of signal control. Furthermore, the addition of the second gate between the first gate and the drain reduces the feedback capacitance between the two, leading to greater reverse isolation, improved stability, and higher gain [3],[4].

A qualitative picture of the DGFET, with applied bias voltages, reveals two distinct depletion regions under each of the two gates. By choosing an arbitrary point between the two gates, the dual-gate device can be thought of as two single-gate devices, with the arbitrary point serving as the drain of the first device and the source of the second device (Figure 2). The signal properties of the composite device can be analyzed by adopting the cascode configuration and determining operating point voltages/currents for each of the separate FETs. With the implicit assumption that the gate currents are negligible, the DGFET structure imposes the constraint that the current in the channels of the two individual single-gate FETs must be equal. By using this constraint, the floating source-2/drain-1 potential can be calculated. This potential is necessary to complete the calculation of the gate-source and drain-source potentials for each of the individual FETs. The following set of equations is obtained:

$$I_{d1} = I_{d2} \tag{1}$$

$$V_d = V_{d1} + V_{d2} \tag{2}$$

$$V_2 = V_{g2} - V_{d1} \tag{3}$$

$$V_1 = V_g \tag{4}$$

$$I_{d1} = f_1(V_{d1}, V_1) \tag{5}$$

$$I_{d2} = f_2(V_{d2}, V_2) \tag{6}$$

where I_{d1}, I_{d2}, V_1, V_2, V_{d1}, and V_{d2} are unknowns that are related by equations (1) through (6) and are determined by the external bias voltages for the DGFET (V_d, V_{g1}, and V_{g2}). The variables f_1 and f_2 represent the functional dependence of the drain current of a single-gate FET on its respective drain and gate potentials. These equations can be solved by using an iterative numerical analysis routine.

Once the operating points for each of the individual FETs are found, the small-signal response of the dual-gate MESFET can be determined using the cascode model. The composite lumped-element model simply combines the two single-gate FET models by cascading the circuits and including the parasitic elements. However, the small-signal parameters thus derived are functions of the four defined voltages (V_1, V_2, V_{d1}, and V_{d2}). These voltages, except

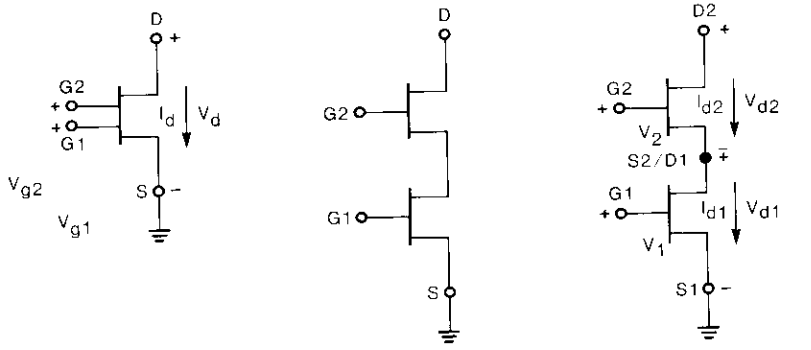


Figure 2. *Circuit Schematic for DGFET Analysis*

for V_1, are not directly measurable quantities of the dual-gate device, but were used for analytical convenience. The key dual-gate device parameters (g_m and r_o) can be estimated from DC measurements of I-V characteristics of the device, and are related to the equivalent single-gate device parameters as follows [16]:

$$(g_m)_{G1} = \frac{\partial i_d}{\partial v_{g1}} = \frac{g_{m1} r_{o1}(1 + g_{m2} r_{o2})}{R_D + r_{o2} + (1 + g_{m2} r_{o2})(R_{GG} + R_S) + r_{o1}(1 + g_{m1} R_S)(1 + g_{m2} r_{o2})} = \frac{g_{m1}}{1 + g_{m1} R_S + \dfrac{R_D + r_{o2}}{r_{o1}(1 + g_{m2} r_{o2})} + \dfrac{R_{GG} + R_S}{r_{o1}}} \tag{7}$$

$$(g_m)_{G2} = \frac{\partial i_d}{\partial v_{g2}} = \frac{g_{m2} r_{o2}}{R_D + r_{o2} + (1 + g_{m2} r_{o2})(R_{GG} + R_S) + r_{o1}(1 + g_{m1} R_S)(1 + g_{m2} r_{o2})} \tag{8}$$

$$r_o = \frac{\partial v_d}{\partial i_d} = R_D + r_{o2} + (1 + g_{m2} r_{o2})(R_{GG} + R_S) + r_{o1}(1 + g_{m1} R_S)(1 + g_{m2} r_{o2}) \tag{9}$$

where $(g_m)_{G1}$ and $(g_m)_{G2}$ are the transconductance parameters for the DGFET at gates 1 and 2, respectively, and r_o is the DGFET output resistance. Variables g_{m1}, r_{o1} and g_{m2}, r_{o2} are the corresponding parameters for the individual FET sections of the cascode model. R_{GG}, R_S, and R_D are the extrinsic gate, source, and drain resistances, respectively, of the FETs.

Program model description

A software package was developed so that the preceding analysis could be used for both discrete device analysis and MMIC design. Single-gate MESFET device models previously used at COMSAT were adopted, with several modifications and improvements. Some of the significant additional features were the ability to handle arbitrary channel doping profiles (to accommodate ion-implantation techniques), recessed gate structures, and multiple-fingered gate layouts. These enhancements allowed the program to be useful for a wider variety of device designs and fabrication processes now commonly found in practice.

For the single-gate FET components, the majority of the calculations in the program are based on equations derived by Pucel *et al.* [17]. The drain current is modified to account for additional voltage drops across the parasitic resistances. The procedure for implementing the dual-gate analysis is summarized by the flowchart shown in Figure 3. Iterative analysis with a bisection algorithm is used to force a match between the two single-gate FET currents. Once this is achieved, the program generates current-voltage relationships for the individual FETs, and hence for the composite dual-gate device, by using equations (1) through (6). Figure 4 summarizes the physical device geometries used in the analysis.

Once the drain and gate voltages and drain current for individual FET sections are determined, the program calculates the input resistance, gate-to-source and gate-to-drain capacitances, and transconductance and output resistance for individual FET sections, based on the Pucel equations. The output resistance is slightly modified to account for surface and substrate leakage due to a finite conductivity in the semi-insulating substrate [18]. A sidewall capacitance, which accounts for the extension of the depletion region toward the source, is added to the gate-to-source interelectrode capacitance in the FET model [1].

The source and drain parasitic resistances, which consist of a bulk semiconductor material resistive component in series with the ohmic metal contact resistance, are calculated from Fukui's empirical expressions [19]. Measured data for prototype single-gate FETs were used to modify the empirical constants in these equations. Interelectrode capacitances were then calculated using the geometries of two flat eletrodes on GaAs dielectric material [1],[19]. Six interelectrode capacitances were needed for the dual-gate, small-signal model—one for each pair of ohmic contacts. The output resistance and transconductance parameters at gates 1 and 2 of the DGFET are computed using equations (7) through (9).

Device measurements

To test the accuracy of the DGFET program model, two devices were measured: a commercially produced DGFET (Raytheon RDX-832) and a COMSAT-produced device. The active layer in the RDX-832 device was produced by epitaxial growth, resulting in theoretically uniform carrier concentration doping levels. The COMSAT device was processed using ion-implantation techniques. The ohmic contacts in both devices consisted of an N^+ layer, where N^+ indicates a peak density level one to two orders of magnitude higher than that implanted in the active region. This results in reduced drain and source parasitic resistances.

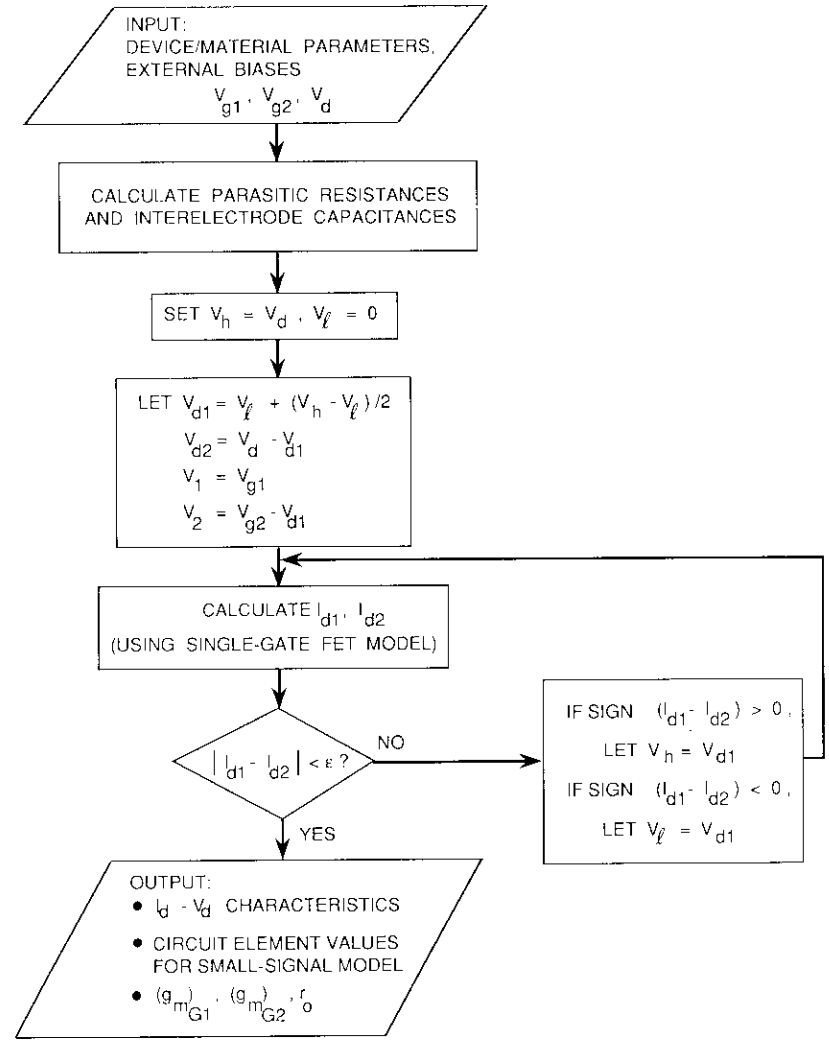


Figure 3. *Flowchart Summarizing the DGFET Analysis as Implemented by the Program*

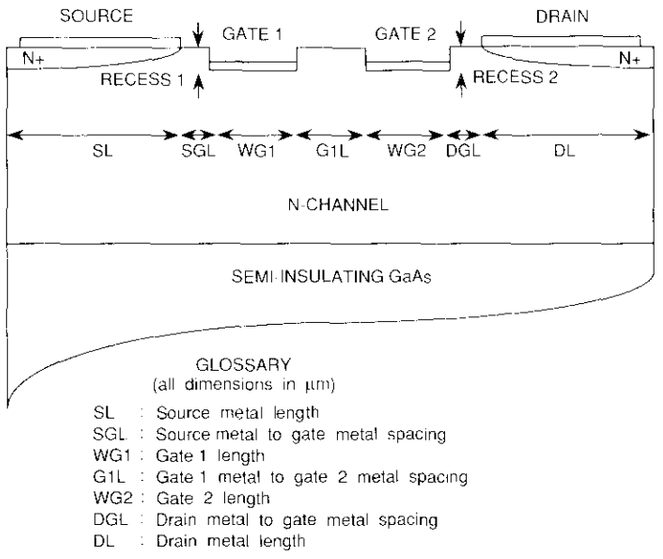


Figure 4. *Topology of a DGFET Showing Geometrical Parameters Used in the Program*

Table 1 summarizes the device geometries and doping characteristics for the RDX-832 device. Estimates of the range of doping density and of the active layer thickness and gate recesses were obtained from the manufacturer. For comparison of DC characteristics, I-V measurements were made on the DGFET using the HP4145A semiconductor parameter analyzer. In Figure 5, the drain current is plotted *vs* the gate 1 voltage with gate 2 grounded, and *vs* the gate 2 voltage with gate 1 grounded. Current-voltage relationships were generated by the program model. A number of calculated points are plotted against the measured curves. Considering that fluctuations in the drain current levels on the order of 20 percent are common even in devices from the same batch, the agreement was considered good. Transconductances corresponding to both gate 1 ($\partial I/\partial V_{g1}$) and gate 2 ($\partial I/\partial V_{g2}$) voltages are plotted in Figure 6. The uncertainty of the recess depth may have been a major contributor to the discrepency between the modeled and measured values. A similar comparison was obtained for the COMSAT device with an ion implantation profile.

S-parameter measurements were made for the DGFET devices at a number of bias points. A lumped-element equivalent circuit model (Figure 7) was fitted to the measured data using the SUPERCOMPACT program. Comparison

TABLE 1. RDX-832 DGFET PHYSICAL DIMENSIONS AND DOPING CHARACTERISTIC ESTIMATES

Active Gate Width	500 μm
Gate 1 Length	1.1 μm
Gate 2 Length	1.1 μm
Source Length	25 μm
Drain Length	25 μm
Source-to-Gate 1 Spacing	1.0 μm
Drain-to-Gate 2 Spacing	1.5 μm
Intergate Spacing	2.0 μm
Gate Recess Depth	0.1–0.2 μm
Gate Metal Thickness	0.5 μm
Active Layer Thickness	0.4–0.5 μm
Channel Carrier Concentration	10^{17}/cm^3
N^+ Doping Density	10^{18}/cm^3
Source Pad Area	0.036 mm^2
Drain Pad Area	0.018 mm^2
Gate 1 Pad Area	0.012 mm^2
Gate 2 Pad Area	0.029 mm^2

of the calculated and fitted equivalent circuit element values indicated much lower output resistance for the FET section. By reducing the output resistance by one-half to one-third of its model-generated DC value, much better agreement could be obtained for the S-parameters S_{21}, S_{22}, S_{23}, and S_{33}. Similarly, better fit could be obtained for the S_{11} parameter by reducing the input resistance. These modeling problems in the input and output resistance have been noted previously [20],[21], and further investigation is warranted. For example, the cascode model assumes two essentially decoupled single-gate FETs, with interaction only through an extrinsic parasitic capacitance. To improve the accuracy of the model, particularly at high frequencies, further investigation of possible interaction between the gates is necessary.

MMIC DGFET switch design

A GaAs MMIC dual-gate switch module is a key element in the development of highly reliable and lightweight microwave switch matrices (MSMs) [22]. For the switch application, the gate 1 and drain biases of the dual-gate MESFET are kept fixed, and the gain of the amplifier is controlled by varying the DC bias of the second gate. Forward bias voltages of ~0.5 V and a negative pinchoff voltage at gate 2 define the ON and OFF states, respectively, of the switch. A minimum 50-dB dynamic range was required for MSM application. A 3.5- to 6.5-GHz design frequency range was chosen to provide

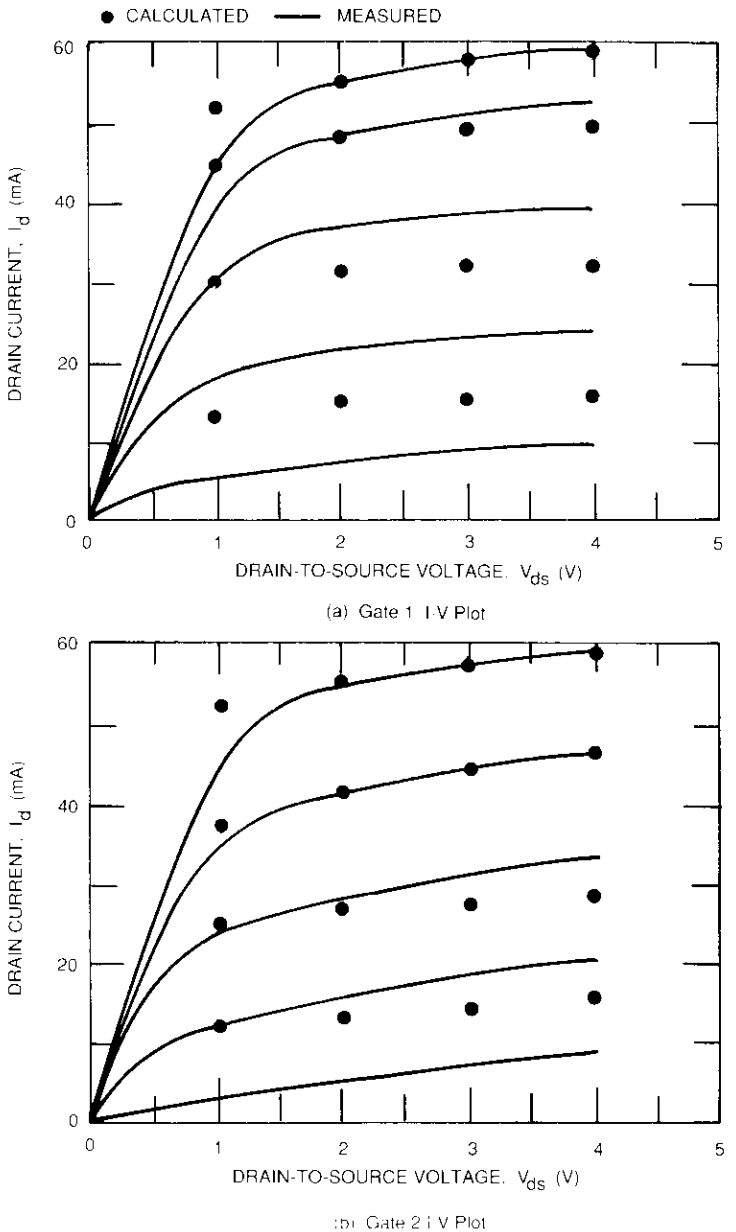


Figure 5. *Comparison of Measured and Computed Drain Current for the RDX-832 DGFET*

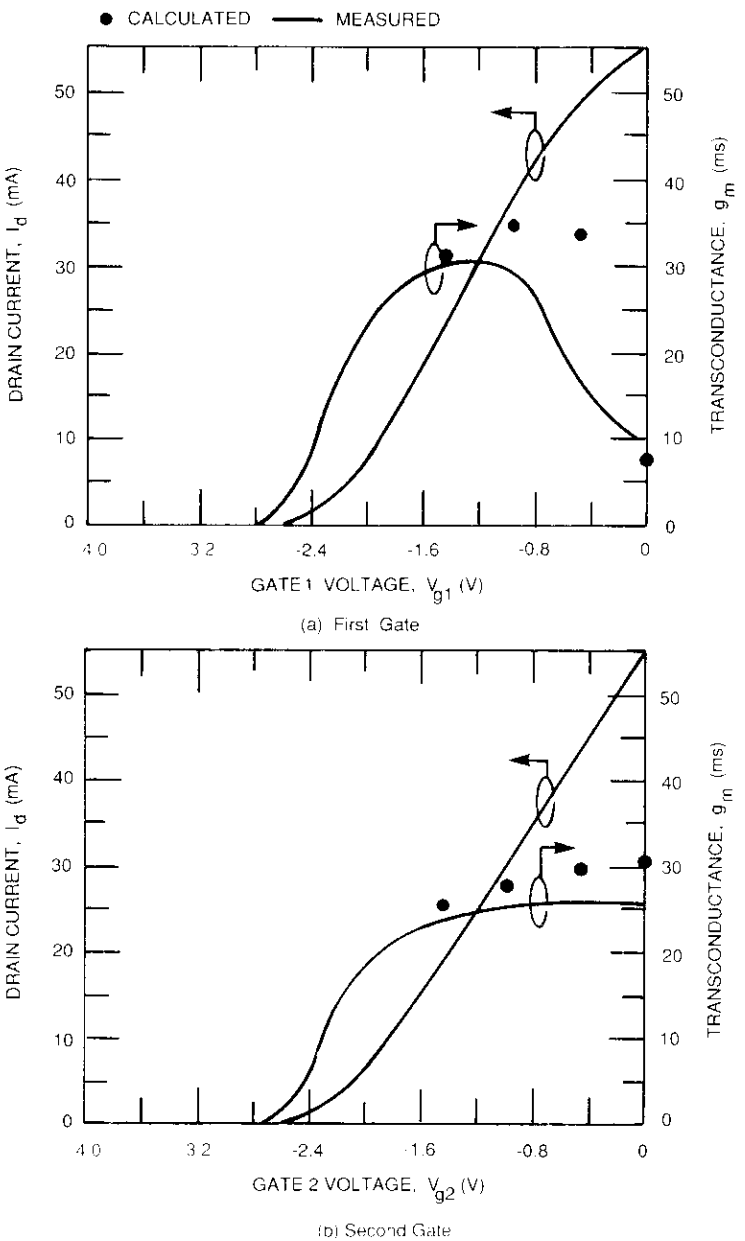


Figure 6. *Comparison of Measured and Computed Transconductance for the RDX-832 DGFET*

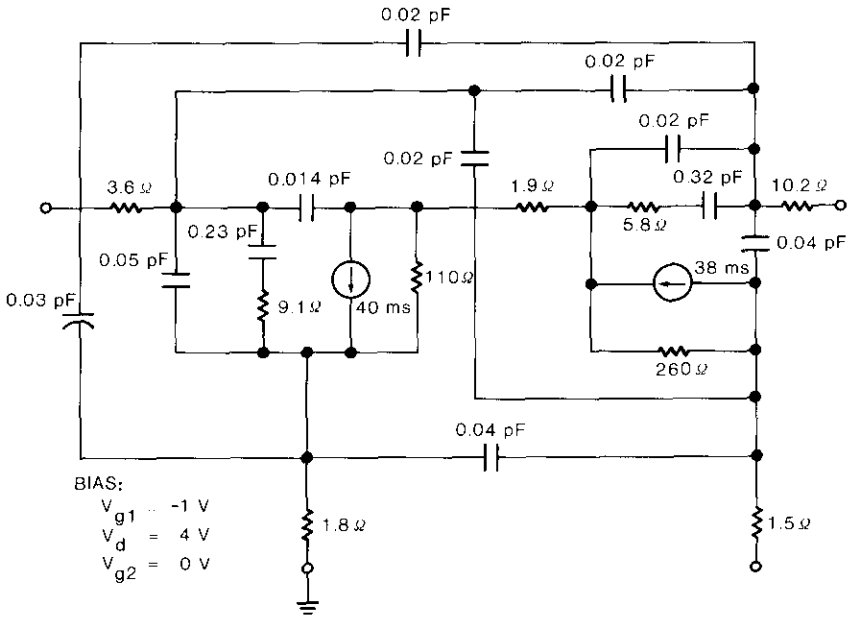


Figure 7. *DGFET Equivalent Circuit Model*

the flexibility of using a single design for all channels in the 6-GHz up-link and 4-GHz down-link satellite communications bands. Good input/output return loss in both the ON and OFF states was required in order to avoid the need for isolators. Finally, flat gain response in the ON state was required.

Design of a broadband DGFET switch module with ON-state gain requires devices with high transconductance (g_m). The g_m can be increased by increasing the device width; however, this also results in larger power dissipation in the ON state. For the MMIC switch module design, a dual-gate device with a nominal gate length of 0.5 μm and width of 300 μm was selected. The 0.5-μm gate length provided sufficiently high g_m, while the 300-μm device width provided more than 10 dBm of linear power with ON-state DC power dissipation of approximately 125 mW per DGFET. Other physical dimensions and profile parameters of the device are listed in Table 2.

The target ion-implantation profile [1] has a peak carrier density of approximately $n_p = 2.7 \times 10^{17}$ cm $^{-3}$. With a recess depth up to the peak of the profile, the program model predicted a pinchoff voltage of around -1.5 V. The program was then used to generate most of the intrinsic FET elements, including junction capacitance, input resistances, and transconductances. The output resistances and the parasitic capacitances and resistances were obtained from device geometries and measurements on prototype single-gate devices built with similar profile parameters. From the lumped-element equivalent circuit shown in Figure 7, the three-port S-parameters were calculated for the design of the DGFET switch circuit.

TABLE 2. PHYSICAL DIMENSIONS AND DEVICE PROFILE PARAMETERS OF THE DGFET

Active Gate Width	300 μm
Gate 1 Length	0.5 μm (nominal)
Gate 2 Length	0.5 μm (nominal)
Source-to-Gate 1 Spacing	1 μm
Drain-to-Gate 2 Spacing	1 μm
Intergate Spacing	1.5 μm
Gate Metal Thickness	0.5 μm
N^+ to Source (Drain) Ohmic Setback	1 μm
N^- Doping Density	1×10^{15} cm $^{-3}$
Peak Carrier Density	2.7×10^{17} cm $^{-3}$
Peak Location, dp	0.11 μm
Standard Deviation, σ	0.65 μm

A two-stage design was considered necessary in order to achieve the required 50-dB ON-to-OFF ratio. The use of resistive matching networks resulted in small circuit size, which made the circuit design relatively insensitive to device parameter variations and resulted in excellent stability. Furthermore, the circuit could be matched easily over a broad bandwidth. After extensive investigation of second-gate termination, a capacitive termination was selected because of its good dynamic range, good stability, small size, and convenience of implementation in MMIC form.

Figure 8 is a complete schematic of the switch circuit. The interstage network combined the output network of the first stage with the input network of the second stage to give a simpler realization, leading to a small circuit size. For operation with a single power supply, a parallel resistance/capacitance self-bias combination was added to the source terminals of each device. Circuit topologies were chosen to satisfy biasing requirements with a minimum number of elements. In the input circuit, the shunt resistance providing DC ground return for self-biasing, and the series capacitor providing input DC blocking, were also used in the circuit matching. Shunt inductances by which the drain bias voltages were fed to the FETs were incorporated into the interstage and output matching circuits. The output DC blocking capacitor was also part of the output matching network, and the interstage DC blocking capacitor was part of the interstage network.

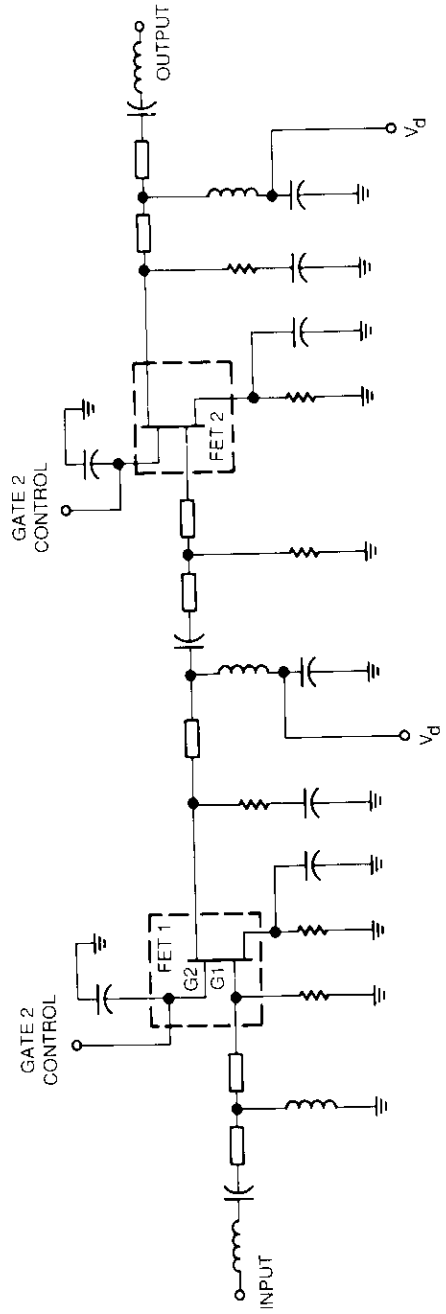


Figure 8. *Schematic of the DGFET Switch Circuit*

One of the unique features of the MMIC DGFET switch module design is the on-chip integration of a TTL-compatible logic/driver interface circuit. This requires that the pinchoff voltages of the devices in the control logic be the same as those of the DGFET devices (~ -1.5 V in this design). Furthermore, the control logic circuit design is restricted to using FETs and Schottky diodes that are compatible with the MMIC DGFET fabrication process. Figure 9 is a schematic of the TTL-to-GaAs interface circuit. The TTL control signal, V_{IN}, appears at the source of FET Q_1. The voltage at the gate of Q_1 is nominally 0 V. Therefore, this FET is turned on and off by V_{IN} voltages of 0 V and a positive voltage which is greater than the FET pinchoff. FETs Q_2 and Q_3 form the current sources and control the current through level-shifting Schottky diodes D_1 through D_4. Depending on the state of Q_1, the voltage at the output, V_{OUT}, is either slightly positive (0.5 V) or greater than -1.5 V. The width of Q_2 and Q_3 FETs in the interface circuit was designed to be small (10 to 20 μm) to minimize DC power consumption. Circuit simulation indicated power consumption in the control interface circuit of less than 10 mW in both the on and off states.

The control circuit performance was analyzed using the I-G SPICE program. A GaAs FET model was created to aid in the SPICE analysis of this interface circuit [22]. Because analysis indicated that a 10-pF capacitor to ground at the FET gate 2 (as assumed in the switch circuit design) resulted in large rise/fall times, a series 0.5-pF capacitor was included at the second gate. This

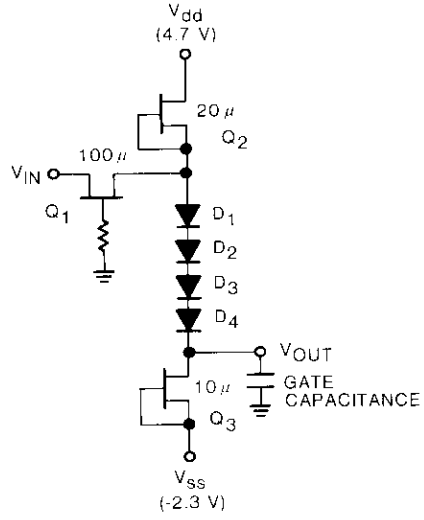


Figure 9. *TTL Control Interface Circuit Schematic*

resulted in predicted rise/fall times of less than 5 ns, at the cost of degrading the switch module gain in the ON state by approximately 2 dB. On the final chip, the 10- and 0.5-pF capacitors can be connected in parallel by bondwires for higher gain.

Circuit layout and fabrication

The resistors were fabricated using a resistive dielectric material with gold pads at the ends. Tantalum nitride was the resistive medium, with a sheet resistance of 50 Ω/□. The capacitors were realized by a thin-film overlay structure of metal/insulating dielectric/metal (MIM), using plasma-deposited silicon nitride ($\epsilon_r \sim 6.4$) as the dielectric. Two thicknesses (0.1 and 0.5 μm) of dielectric were employed: the thicker layer for realizing smaller capacitances (110 pF/mm²) and for passivation of the FET channel regions, and the thinner layer for the high-valued bypass capacitors (550 pF/mm²) which would otherwise occupy prohibitively large areas on the chip.

Higher-valued inductors in the circuit were realized as spirals to keep circuit size small, and were modeled using coupled transmission line models. Smaller inductor values were simply approximated by short lengths of high-impedance transmission lines. A 2-μm-thick layer of gold was used for plating the transmission lines, spiral inductors, top plates of capacitors, overlay metal of ohmic contacts, and other passive circuit metal areas. Air bridges were employed to interconnect the FET sources, to contact the top plate of capacitors, and to form crossovers for spiral inductors.

The entire circuit model was analyzed and optimized to the desired response by using SUPERCOMPACT. An extensive sensitivity analysis was performed to check the circuit performance variations with fabrication tolerances. For resistance and capacitance variations of ±20 and ±10 percent, respectively, the worst-case input/output return losses were predicted to better than 15 dB in both the ON and OFF states, with a gain variation within ±1 dB.

The circuits were fabricated using selective ion implantation into liquid-encapsulated Czochralski (LEC) material. The channel was formed by a 100-keV silicon implant, and the N^+ ohmic contact area was formed by a 200-keV silicon implant. Ohmic contacts were formed by vacuum deposition of an Au/Ge/Ni/Ag/Au film that was patterned by a dielectric-assisted liftoff process and then alloyed into the substrate. Direct-write, electron-beam lithography was used to fabricate the FET gate electrodes. The gate metal, a 0.5-μm-thick Ti/Pt/Au composite, also formed the bottom plates of the capacitors. The dual-gate MESFET had two parallel 150-μm-wide gate feeds, and the second gate was end-fed.

Figure 10 shows an enlargement of the realization of the entire module. The total chip size is 1.5 × 2.5 mm (60 × 100 mil) on a 0.3-mm (12.5-mil)-thick semi-insulating GaAs substrate, and includes two dual-gate MESFETs, three single-gate MESFETs, four Schottky diodes, seven resistors, nine capacitors, and four spiral inductors.

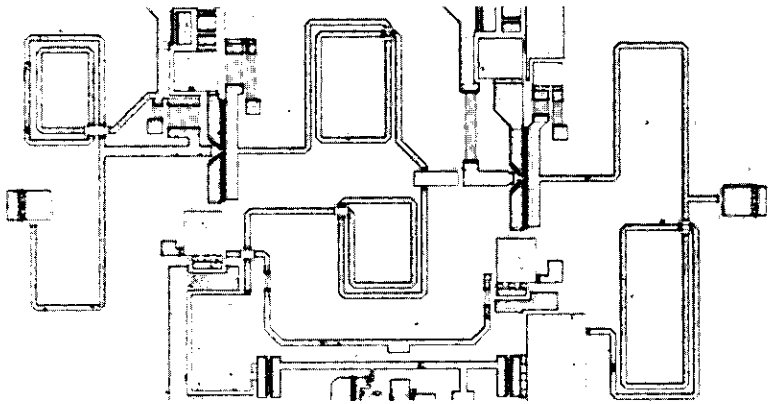
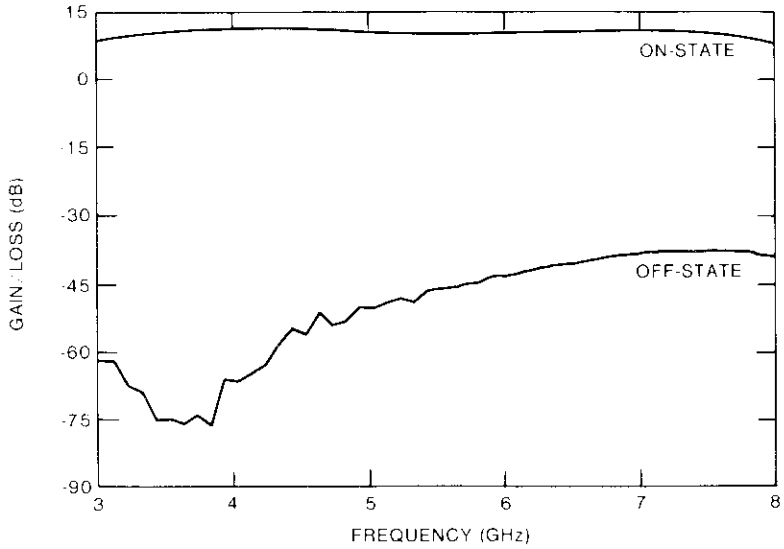


Figure 10. *Microphotograph of the MMIC DGFET Switch Module (Chip dimensions: 1.5 × 2.5 mm)*

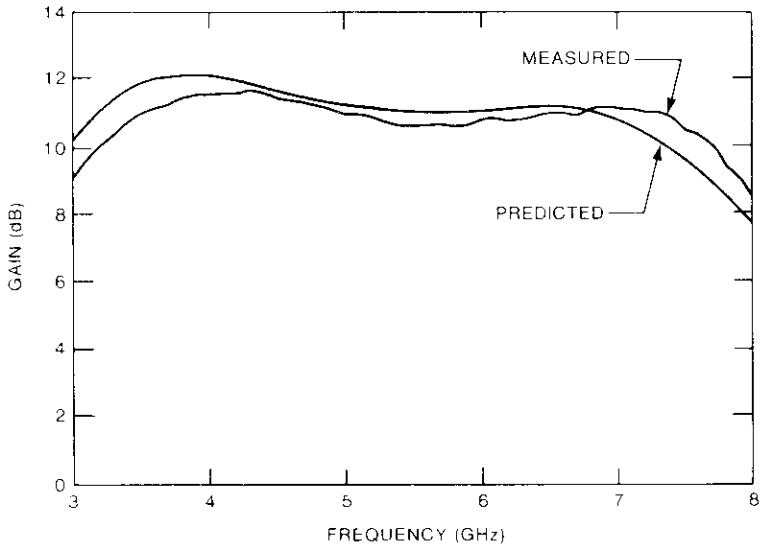
Test results

The transmission gain/loss and return losses were measured for both the ON and OFF states of the switch (Figure 11). For the first set of measurements, the driver interface was not included in the circuit. With the gate 2 bias voltage at 0.5 V, the switch module had an ON-state gain of 11.2 ± 0.4 dB from 3.5 to 6.5 GHz. The ON-to-OFF isolation was better than 50 dB up to 6.5 GHz. Predicted and measured gains agreed within 1 dB from 3.5 to 6.5 GHz. The worst-case input and output return loss values in the two states were approximately 12 dB over the band of interest.

The TTL control interface circuit was measured separately. The results (Figure 12) show that the control circuit provides approximately 0.5 V to gate 2 of the DGFETs for TTL '1' input (>2.4 V). The output voltage for TTL '0' input is below the pinchoff voltage to turn the DGFETs off. The gain of the switch element was measured to be about 2-dB lower with a 0.5-pF capacitor at the second gate.



(a) Measured ON-State Gain and ON-to-OFF Isolation



(b) Comparison Between Measured and Modeled ON-State Gain

Figure 11. *DGFET Switch Module Performance*

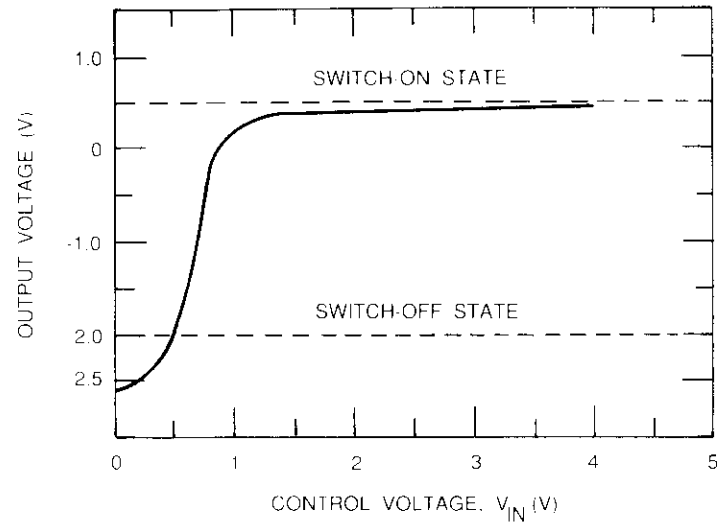


Figure 12. *TTL Interface Circuit Measured Performance*

Conclusions

A modeling approach for a dual-gate MESFET based on physical device parameters has been described. To demonstrate its usefulness, the model was used in the design of a broadband (3.5- to 6.5-GHz) DGFET switch module. Good agreement between the measured and modeled performance during the first fabrication iteration demonstrated the validity of both the model and the design approach. Integration of the RF and digital control functions on an MMIC chip was also demonstrated. This switch module can replace PIN diodes as the switching element in the MSM used in SS-TDMA communications satellites.

Acknowledgments

The authors wish to acknowledge the efforts of J. Allison, W. Baker, C. Cannon, G. Carlson, R. Edwards, K. Hogan, L. Holdeman, J. Kearney, R. Kroll, P. McNally, R. Porter, J. Reynolds, T. Sizemore, and J. Tyler during the design, CAD layout, fabrication, and test phases of this circuit development. They also wish to thank F. Assal and J. Potukuchi for many useful discussions. Support provided for this effort by C. Cotner of the ISS, World Systems Division is gratefully acknowledged.

References

[1] R. K. Gupta, J. H. Reynolds, M. C. Fu, and T. Heikkila, "Design and Modeling of a GaAs Monolithic 2- to 6-GHz Feedback Amplifier," *COMSAT Technical Review*, Vol. 17, No. 1, Spring 1987, pp. 1–22.

[2] J. A. Turner *et al.*, "Dual-Gate Gallium-Arsenide Microwave Field-Effect Transistor," *Electronics Letters*, Vol. 7, No. 22, November 1971, pp. 661–662.

[3] C. A. Liechti, "Characteristics of Dual-Gate GaAs MESFETs," 4th European Microwave Conference, Montreux, Switzerland, September 1974, *Proc.* pp. 87–91.

[4] S. Asai, F. Murai, and H. Kodera, "GaAs Dual-Gate Schottky-Barrier FETs for Microwave Frequencies," *IEEE Transactions on Electron Devices*, Vol. ED-22, No. 10, October 1975, pp. 897–904.

[5] C. A. Liechti, "Performance of Dual-Gate GaAs MESFETs as Gain-Controlled Low-Noise Amplifiers and High-Speed Modulators," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-23, No. 6, June 1975, pp. 461–469.

[6] M. Maeda and Y. Minai, "Microwave Variable-Gain Amplifier With Dual-Gate GaAs FET," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-22, No. 12, December 1974, pp. 1226–1230.

[7] C. Tsironis, R. Meierer, and R. Stahlmann, "Dual-Gate MESFET Mixers," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-32, No. 3, March 1984, pp. 248–255.

[8] J. L. Vorhaus, R. A. Pucel, and Y. Tajima, "Monolithic Dual-Gate GaAs FET Digital Phase Shifter," *IEEE Transactions on Electron Devices*, Vol. ED-29, No. 7, July 1982, pp. 1078–1088.

[9] W. C. Tsai, S. F. Paik, and B. S. Hewitt, "Switching and Frequency Conversion Using Dual-Gate FETs," 9th European Microwave Conference, Brighton, England, September 1979, *Proc.*, pp. 311–315.

[10] G. Ohm and J. Czech, "12-GHz Variable-Gain Amplifier With Dual-Gate GaAs FET for Satellite Use," 9th European Microwave Conference, Brighton, England, September 1979, *Proc.*, pp. 298–302.

[11] B. Kim, H. Q. Tserng, and P. Saunier, "GaAs Dual-Gate FET for Operation up to K-Band," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-32, No. 3, March 1984, pp. 256–261.

[12] J. R. Scott and R. A. Minasian, "A Simplified Microwave Model of the GaAs Dual-Gate MESFET," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-32, No. 3, March 1984, pp. 243–248.

[13] T. Furutsuka, M. Ogawa, and N. Kawamura, "GaAs Dual-Gate MESFETs," *IEEE Transactions on Electron Devices*, Vol. ED-25, No. 6, June 1978, pp. 580–586.

[14] S. Asai, F. Murai, and H. Kodera, "The GaAs Dual-Gate FET With Low Noise and Wide Dynamic Range," IEEE International Electron Devices Meeting, Washington, D.C., December 1973, *Digest*, pp. 64–67.

[15] F. Assal, R. Gupta, J. Apple, and A. Lopatin, "Satellite Switching Center for SS-TDMA Communications," *COMSAT Technical Review*, Vol. 12, No. 1, Spring 1982, pp. 29–68.

[16] M. C. Fu, "Dual-Gate GaAs MESFET Modeling and Designs for MMICs," S.M. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1985.

[17] R. Pucel *et al.*, "Signal and Noise Properties of Gallium Arsenide Microwave Field-Effect Transistors," *Advances in Electronics and Electron Physics*, Vol. 38, New York: Academic Press, 1975, pp. 193–265.

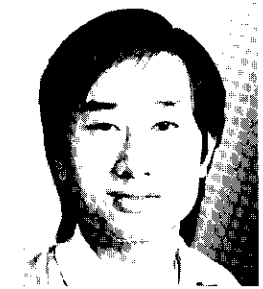
[18] T. Smith, "A Modeling System for Simulation of GaAs FET Performance," *COMSAT Technical Review*, Vol. 15, Number 2A, Fall 1985, pp. 237–258.

[19] H. Fukui, "Determination of the Basic Device Parameters of a GaAs MESFET," *Bell System Technical Journal*, Vol. 58, No. 3, March 1979, pp. 771–797.

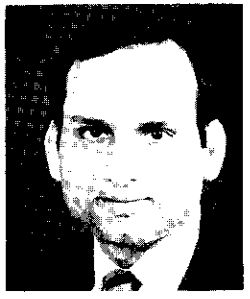
[20] R. Gupta, T. Smith, and J. Reynolds, "GaAs MMICs: CAD, Modeling, and Process Monitoring," *Microwave Journal*, Vol. 30, No. 8, August 1987, pp. 95–114.

[21] T. Smith, "Beyond-Punchthrough Current in GaAs MESFET's," *IEEE Electron Device Letters*, Vol. EDL-7, No. 3, March 1986, pp. 188–189.

[22] R. Gupta *et al.*, "A Broadband MMIC Dual-Gate FET Switch Module With On-Chip TTL Control Interface," 17th European Microwave Conference, Rome, Italy, September 1987, *Proc.*, pp. 243–248.



Michael C. Fu received S.B. and S.M. degrees in electrical engineering and an S.B. in mathematics from the Massachusetts Institute of Technology (MIT) in 1985, and a Ph.D. in applied mathematics from Harvard University in 1989. From 1982 to 1985, he worked in the Microwave Technology Division at COMSAT Laboratories under the MIT VI-A cooperative education program, where his research involved the modeling of microwave FET devices and the design of GaAs MMICs. From 1985 to 1989, he received Graduate Fellowships from the National Science Foundation (NSF) and the Center for Intelligent Control. He is currently Assistant Professor of Business and Management at the University of Maryland, College Park, where his research interests include perturbation analysis and stochastic optimization of discrete-event dynamic systems, with applications to manufacturing systems and communications networks. He is a member of Tau Beta Pi and Eta Kappa Nu.



Ramesh K. Gupta received a B.Sc. (Honors) in electronics and communications engineering from Punjab University, Chandigarh, India, in 1974, and an M.Sc. and Ph.D. in electrical engineering from the University of Alberta, Edmonton, Canada, in 1976 and 1980, respectively. He also received an M.B.A. from the Wharton School of Business, University of Pennsylvania, in 1989. In 1976, he was awarded a 3-year Alberta Government Telephones Centennial Fellowship for graduate research in telecommunications. He joined COMSAT Laboratories in 1980, where his responsibilities have included development of wideband 8 × 8 and 4 × 4 microwave switch matrix (MSM) arrays using hybrid MIC and GaAs MMIC technologies, respectively. He has also contributed to the spacecraft and ground network study and support effort for the INTELSAT VI satellite program. More recently, he has been involved with the development of active elements for K_u-band phased-array antennas and the design of miniaturized components for a 120-Mbit/s CQPSK on-board modem. Dr. Gupta is a Senior Member of IEEE and has served as Vice-Chairman (1987–88) and Chairman (1988–89) of the Washington, D.C./Northern Virginia Chapter of the IEEE MTT Society.

Translations of Abstracts

N-MAC—Canevas de format de transmission vidéo à éléments analogiques, compatible avec la norme NTSC

L-N. Lee et M. D. Redman

Sommaire

Un canevas de format de transmission vidéo à éléments analogiques multiplexés (N-MAC) qui est compatible avec la norme NTSC a été mis au point pour la Satellite Television Corporation au début des années 80. Le présent exposé décrit en détail le format de N-MAC ainsi que la mise en oeuvre et la performance d'une maquette d'identification. Les résultats d'essais subjectifs y figurent également.

Filtre-peigne tridimensionnel à compensation automatique du mouvement pour la transmission de la vidéo avec codage NTSC

D. W. Power et L-N. Lee

Sommaire

On présente un algorithme utilisant un filtre-peigne tridimensionnel à compensation automatique de mouvement pour la suppression de la diaphotie luminance/chrominance dans les signaux vidéo avec codage NTSC. Ce filtre, qui consiste en un peigne "temporel" à compensation automatique de mouvement pour les parties statiques de l'image et en un peigne "spatial" bidimensionnel pour les parties qui varient en fonction du temps, produit une largeur de bande de luminance accrue et une plus grande séparation entre la luminance et la chrominance que ne le permettent les conceptions de filtre-peignes classiques. L'algorithme proposé peut être utilisé pour améliorer la transmission de télévision NTSC standard ou dans le cadre d'un système grâce auquel la télévision haute définition (HDTV) peut être transmise via un canal compatible avec la norme NTSC. Les principes de la conception ont été confirmés par simulation sur ordinateur.

Author Index, CTR 1988

The following is a cross-referenced index of articles that appeared in the *COMSAT Technical Review* in 1988. The code number at the end of each entry is to be used for ordering reprints from Corporate Records, COMSAT Corporation, 22300 Comsat Drive, Clarksburg, MD 20871-9475.

ATIA, A. E., see Lee, Y. S. [CTR88/334].

CHANG, P. Y., see Palmer, L. C. [CTR88/331].

CREGGER, B. B., see McNally, P. J. [CTR88/330].

HAMPSCH, T., "Digital Controller for High-Speed Multibeam Antennas," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 255–268 [CTR88/337].

HEMMATI, F., "Reduced-Complexity Decoding of Convolutional Codes," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 239–253 [CTR88/336].

LEE, Y. S., Atia, A. E., and Ponchak, D. S.,* "Intersatellite Link Application to Commercial Communications Satellites," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 147–189 [CTR88/334].

MARSHALEK, R. G., "Comparison of Optical Technologies for Intersatellite Link Payloads. Part I: Mass, Prime Power, and Volume Estimates," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 191–214 [CTR88/335].

MARSHALEK, R. G., and Paul, D. K., "Annotated Bibliography of Optical Intersatellite Link Technology 1970–1988," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 283–348 [CTR88/339].

McNALLY, P. J., and Cregger, B. B., "Ion-Implanted Hyperabrupt Varactor Diodes for GaAs MMICs," *COMSAT Technical Review*, Vol. 18, No. 1, Spring 1988, pp. 1–20 [CTR88/330].

MEULENBERG, A., "Radiation-Hardened Deposited Oxides," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 269–282 [CTR88/338].

PALMER, L. C., and Chang, P. Y., "Simulation of a Random-Access-With-Notification Protocol for VSAT Applications," *COMSAT Technical Review*, Vol. 18, No. 1, Spring 1988, pp. 21–53 [CTR88/331].

PAUL, D. K., "Comparison of Optical Technologies for Intersatellite Link Payloads. Part II: Impact of Reliability Considerations on Technology Selection," *COMSAT Technical Review*, Vol. 18, No. 2, Fall 1988, pp. 215–238 [CTR88/335].

PAUL, D. K., see Marshalek, R. G. [CTR88/339].

SCHMITT, C. H.,* "Geostationary Satellite Log for Year End 1987," *COMSAT Technical Review*, Vol. 18, No. 1, Spring 1988, pp. 65–134 [CTR88/333].

SUYDERHOUD, H. G., "Dependence of Mean Opinion Scores on Differences in Lingual Interpretation," *COMSAT Technical Review*, Vol. 18, No. 1, Spring 1988, pp. 55–64 [CTR88/332].

*Non-COMSAT author.

Index of 1988 Publications by COMSAT Authors

The following is a cross-referenced index of publications by COMSAT authors for 1988, not including papers published in the *COMSAT Technical Review*. Copies of these publications may be obtained by contacting the COMSAT authors at COMSAT Corporation, 22300 Comsat Drive, Clarksburg, MD 20871–9475.

ALLISON, J., see Hegazi, G.

ANDERSON, W. T.,* Meulenberg, A., Beall, J. M.,* Kazi, A. H.,* Harrison, R. C.,* Gerdes, J.,* and Mittleman, S. D.,* "Radiation Induced Displacement Damage in GaAs Devices," *Gallium Arsenide and Related Compounds 1987*, A. Christou and H. S. Rupprecht, eds., Bristol, England: IOP Publishing, 1988, pp. 471–474.

ASSAL, F. T., and Mahle, C. E., "Hardware Development for Future Commercial Communications Satellites," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 332–349. AIAA Paper No.88–0815.

ASSAL, F. T., Zaghloul, A. I., and Sorbello, R. M., "Multiple Spot Beam Systems for Satellite Communications," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers, pp.* 322–331. AIAA Paper No. 88–0814.

ASSAL, F. T., Zaghloul, A. I., and Sorbello, R. M., "Active Phased Arrays for Multibeam Communications Applications," IEEE Military Communications Conference (MILCOM), San Diego, CA, October 1988, *Conference Record*, Classified Session No. 29.

ASSAL, F. T., see Gupta, R.

ASSAL, F. T., see Potukuchi, J. R.

ASSAL, F. T., see Sayegh, S. I.

ASSAL, F. T., see Zaghloul, A. I.

ATIA, A. E., see Zaki, K. A. (two papers)

BARGELLINI, P. L., and Hyde, G., "Electrical Communications Systems Approaching the Year 2000," *Satellite Integrated Communications Networks*, E. Del Re, P. Bartholomé, and P. P. Nuspl, eds., Amsterdam: Elsevier Science, 1988, pp. 11–21.

BASS, J. F., see Hung, H–L. A.

BASS, J. F., see Lee, T. T.

BONETTI, R. R., and Williams, A. E., "A TE Triple-Mode Filter," IEEE MTT-S International Microwave Symposium, New York, NY, May 1988, Digest, pp. 511–514.

* Non-COMSAT author.

BONETTI, R. R., see Williams, A. E.

CAMPANELLA, S. J., "Satellite Switches," *Satellite Integrated Communications Networks*, E. Del Re, P. Bartholomé, and P. P. Nuspl, eds., Amsterdam: Elsevier Science, 1988, pp. 57–64.

CAMPANELLA, S. J., and Pontano, B. A., "Economic Comparison of Cables and Satellites," *Satellite Integrated Communications Networks*, E. Del Re, P. Bartholomé, and P. P. Nuspl, eds., Amsterdam: Elsevier Science, 1988, pp. 35–46.

CAMPANELLA, S. J., and Sayegh, S. I., "A Flexible On-Board Demultiplexer/Demodulator," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 299–303. AIAA Paper No. 88–0811.

CAMPANELLA, S. J., and Dimolitsas, S., "Digital Voice Transmission Techniques for Aeronautical Communications," Radio Technical Commission for Aeronautics (RTCA) Annual Assembly Meeting and Technical Symposium, Washington, DC, November 1988, *Proc.*, pp. 51–73. RTCA Paper No. 405–88/AS-349.

CAMPANELLA, S. J., see Naderi, F. M.

CHEN, C., see Zaki, K. A. (two papers)

CHITRE, D. M., and McCoskey, J. S., "VSAT Networks: Architectures, Protocols and Management," *IEEE Communications Magazine*, Vol. 26, No. 7, July 1988, pp. 28–38.

CHITRE, D. M., "ISDN Challenges to Satellite Communications" (Abstract) IEEE 21st Electronics and Aerospace Conference (EASCON), Arlington, VA, November 1988, *Proc.*, p. 107.

CHITRE, D. M., see Coney, T. A.

CONEY, T. A.,* Dobyns, T. R., Chitre, D. M., and Lindstrom, R., "Service Offerings and Interfaces for the ACTS Network of Earth Stations," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 247–258. AIAA Paper No. 88–0800.

CORNFELD, A., see Hegazi, G.

CORNFELD, A., see Lee, T. T.

DiFONZO, D. F., see Sorbello, R. M.

DIMOLITSAS, S., and Gunn, J. E.,* "A Modular, Off-Line, Full Duplex Telephone Channel Simulator for High Speed Data Transciever Evaluation," *Proc. IEE, Part F: Communications, Radar and Signal Processing*, Vol. 135, No. 2, April 1988, pp. 155–160.

DIMOLITSAS, S., see Campanella, S. J.

DIMOLITSAS, S., see Suyderhoud, H. G.

DOBYNS, T. R., see Coney, T. A.

EARL, M. W., see Vaidyanathan, H.

EFFLAND, J. E., see Sorbello, R. M.

* Non-COMSAT author.

FANG, R. J. F., and Budinger, J.,* "Bandwidth-Efficient High-Speed Coded Trellis Modulation," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 313–321. AIAA Paper No. 88–0813.

GELLER, B. D., and Goettle, P., "Quasi-Monolithic 4-GHz Power Amplifiers With 65-percent Power-Added Efficiency," IEEE MTT-S International Microwave Symposium, New York, NY, May 1988, *Digest*, pp. 835–838.

GELLER, B. D., see Mahle, C. E.

GERSON, H., see Gupta, R.

GOETTLE, P., see Geller, B.

GOURLEY, S. E., see Mott, R. C.

GUPTA, R. K., Reynolds, J., and McNally, P., "Modeling and CAD of an Ultra-Broadband Monolithic 5-bit Digital Attenuator," 18th European Microwave Conference, Stockholm, Sweden, September 1988, *Proc.*, pp. 151–155.

GUPTA, R. K., Gerson, H., Ross, P., and Assal, F., "Design and Packaging Approach for MMIC Insertion in a Broadband 4 x 4 Microwave Switch Matrix," IEEE GaAs IC Symposium, Nashville, TN, November 1988, *Digest*, pp. 261–264.

GUPTA, R. K., see Mott, R. C.

GUPTA, R. K., see Potukuchi, J. R.

HEGAZI, G., Hung, H-L. A., Phelleps, F., Holdeman, L., Cornfeld, A., Smith, T., Allison, J., and Huang, H., "V-Band Monolithic Power MESFET Amplifiers," IEEE MTT-S International Microwave Symposium, New York, NY, May 1988, *Digest*, pp. 409–412.

HEGAZI, G. M., see Hung, H-L. A. (three papers)

HEMMATI, F., "Closest Coset Decoding," IEEE International Conference on Communications (ICC-88), Philadelphia, PA, June 1988, *Proc.*, pp. 21.2.1–21.2.5.

HOLDEMAN, L. B., see Hegazi, G.

HOLDEMAN, L. B., see Potukuchi, J. R.

HUANG, H. C., see Hegazi, G.

HUANG, H. C., see Hung, H-L. A. (four papers)

HUANG, H. C., see Lee, T. T.

HUANG, H. C., see Polak-Dingels, P.

HULLEY, M. A., "Ion-Noise Sidebands as a Possible Quality Control Tool in TWTs," Advisory Group on Electronic Devices (AGED) and IEEE Microwave Power Tube Conference, Monterey, CA, May 1988, *Proc.*, p. 3D.2.

HUNG, H-L. A., Lee, T. T., Phelleps, F. R., Singer, J. F., Bass, J. F., Noble, T. F., and Huang, H. C., "60-GHz GaAs MMIC Low-Noise Amplifiers," IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium, New York, NY, May 1988, *Digest*, pp. 87–90.

* Non-COMSAT author.

HUNG, H-L. A., Hegazi, G. M., Peterson, K. E., and Huang, H. C., "Design and Performance of MESFET Power Amplifiers at K- and K_a-Band," *Microwave Journal*, Vol. 31, No. 6, June 1988, pp. 177–190.

HUNG, H-L. A., Hegazi, G., Peterson, K. E., Phelleps, F. R., and Huang, H. C., "Monolithic Dual-Gate MESFET Power Amplifiers," IEEE GaAs IC Symposium, Nashville, TN, November 1988, *Digest*, pp. 41–44.

HUNG, H-L. A., Hegazi, G., Lee, T., Phelleps, F., Singer, J., and Huang, H., "V-Band MMIC Low-Noise and Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-36, No. 12, December 1988, pp. 1966–1975.

HUNG, H-L. A., see Hegazi, G.

HUNG, H-L. A., see Lee, T. T.

HUNG, H-L. A., see Meulenberg, A. (two papers)

HUNG, H-L. A., see Polak-Dingels, P. (two papers)

HYDE, G., see Bargellini, P. L.

HYDE, G., see Mahle, C. E.

INUKAI, T., Jupin, D., Lindstrom, R., and Meadows, D., "ACTS TDMA Network Control Architecture," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 225–239. AIAA Paper No. 88–0798.

INUKAI, T., see Sayegh, S. I.

JUPIN, D., see Inukai, T.

KELLY, W. H., Reisenweber, J. H., Robinson, J. A., and Abercrombie, D. B.,* "Program Overview and In-Orbit Performance Evaluation of the INTELSAT V Spacecraft Thermal Control Subsystem," AIAA Thermophysics, Plasmadynamics and Lasers Conference, June 1988, San Antonio, TX. AIAA Paper No. 88–2690.

KELLY, W. H., see Wise, P. C.

LEE, B. S., Zaghloul, A. I., and Sorbello, R. M., "Performance Comparison for Scanning and Hopping Beam Satellite Antennas" (Abstract), URSI Radio Science Meeting, Syracuse, NY, June 1988, *Program and Abstracts*., p. 435.

LEE, T. T., Hung, H-L. A., Cornfeld, A., Phelleps, F. R., Singer, J. L., Bass, J. F., and Huang, H. C., "GaAs MBE Monolithic Amplifiers for Millimeter-Wave Applications," Government Microcircuits Applications Conference (GOMAC), Las Vegas, NV, November 1988, *Digest*, pp. 369–372.

LEE, T. T., see Hung, H-L. A. (two papers)

LEE, Y. S., "Cost-Effective Intersatellite Link Applications to the Fixed Satellite Services," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 167–173. AIAA Paper No. 88–0770.

LINDSTROM, R., see Coney, T. A.

LINDSTROM, R., see Inukai, T.

* Non-COMSAT author.

MAHLE, C. E., Geller, B. D., Potukuchi, J. R., and Hyde, G., "Advanced Communications Satellite Technology," IEEE 21st Electronics and Aerospace Conference (EASCON), Arlington, VA, November 1988, *Proc.*, pp. 207–211.

MAHLE, C. E., see Assal, F. T.

McCOSKEY, J. S., see Chitre, D. M.

McNALLY, P., see Gupta, R.

MEADOWS, D., see Inukai, T.

MEULENBERG, A., Hung, H-L. A., and Tough, G., "Microwave Characterization of Bulk and Powdered High-T_c Superconductors," SPIE 13th International Conference on Infrared and Millimeter Waves, Honolulu, HI, December 1988, *Digest*, pp. 259–260.

MEULENBERG, A., Hung, H-L. A., Peterson, K. E., and Anderson, W. T.,* "Total Dose and Transient Radiation Effects on GaAs MMICs," *IEEE Transactions on Electron Devices*, Special Issue on Reliability, Vol. ED-35, No. 12, December 1988, pp. 2125–2132.

MEULENBERG, A., see Anderson, W. T.

MOTT, R. C., Potukuchi, J. R., Gupta, R. K., Zaghloul, A. I., Siddiqi, S., and Gourley, S. E., "Monolithic Transmit Modules for a Multibeam K_a-Band Phased Array Antenna," 18th European Microwave Conference, Stockholm, Sweden, September 1988, *Proc.*, pp. 759–763.

MOTT, R. C., see Potukuchi, J. R.

NADERI, F. M.,* and Campanella, S. J., "NASA's Advanced Communications Technology Satellite (ACTS): An Overview of the Satellite, the Network, and the Underlying Technologies," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 204–224. AIAA Paper No. 88–0797.

NOBLE, T. F., see Hung, H-L. A.

ONUFRY, M., "The New Challenges of DCME," *ITU Telecommunication Journal*, Special Issue on Speech Processing, Vol. 55, No. 12, December 1988, pp. 831–837.

PALMER, L., and White, L., "Demand Assignment in the ACTS LBR System," IEEE International Conference on Communications, Philadelphia, PA, June 1988, *Proc.*, pp. 16.3.1–16.3.6.

PAUL, D. K., "Status and Trends in Fiber Optic Communications," Guest Editorial in *Laser Focus Fall Product Bulletin*, August 1988, p. 20.

PAUL, D. K., "Reliable Fiber Optics" (Summary of the panel session), Fiber Optics Reliability: Benign and Adverse Environments II, Boston, MA, September 1988, *Proc. SPIE*, Vol. 992, D. K. Paul, R. A. Greenwell, and S. Wadekar, eds., pp. 223–229.

PAUL, D. K., "Broadband Analog Capability Opens Up New Fiber Optic Applications," Guest Editorial in *Laser Focus Magazine*, Annual Fiber Optics Review Edition, Vol. 24, No. 10, October 1988, p. 136.

* Non-COMSAT author.

PETERSON, K. E., see Hung, H-L. A. (two papers)
PETERSON, K. E., see Meulenberg, A.
PHELLEPS, F. R., see Hegazi, G.
PHELLEPS, F. R., see Hung, H-L. A. (three papers)
PHELLEPS, F. R., see Lee, T. T.
POLAK-DINGELS, P.,* Hung, H-L. A., Smith, T., Huang, H. C., Webb, K. J.,* and Lee, C. H.,* "On-Wafer Characterization of Monolithic Millimeter-Wave Integrated Circuits by a Picosecond Optical Electronic Technique," IEEE MTT-S International Microwave Symposium, New York, NY, May 1988, *Digest*, pp. 237–240.
POLAK-DINGELS, P.,* Hung, H-L. A., Webb, K. J.,* Lee, T. T., Smith, T., and Lee, C. H.,* "An Optoelectronic Technique for S-Parameter Measurements of GaAs Monolithic Integrated Circuits," SPIE 13th International Conference on Infrared and Millimeter Waves, Honolulu, HI, December 1988, *Digest*, pp. 69–70.
PONTANO, B. A., see Campanella, S. J.
POTUKUCHI, J. R., Gupta, R. K., Assal, F. T., Holdeman, L. B., Mott, R. C., and Zaghloul, A. I., "MMIC Modules for Active Phased-Array Applications in Communications Satellites," *Microwave Systems News*, Vol. 18, No. 11, November 1988, pp. 20–27.
POTUKUCHI, J. R., see Mahle, C. E.
POTUKUCHI, J. R., see Mott, R. C.
REISENWEBER, J. H., see Kelly, W. H.
REYNOLDS, J., see Gupta, R.
ROBINSON, J. A., see Kelly, W. H.
ROGERS, D. V., "Radio-Wave Propagation," Chapter 1, *Electronic Communications Handbook*, A. F. Inglis, ed., New York: McGraw-Hill, 1988.
ROSS, P., see Gupta, R.
SAYEGH, S. I., Assal, F. T., and Inukai, T., "On-Board Processing Architectures and Technology," IEEE 21st Electronics and Aerospace Conference (EASCON), Arlington, VA, November 1988, *Proc.*, pp. 217–223.
SAYEGH, S. I., see Campanella, S. J.
SCHMIDT, W. D., "The ACTS LBR System: A Technology Testbed for Future VSAT/TDMA Networking Applications," AIAA 12th International Communication Satellite Systems Conference, Arlington, VA, March 1988, *A Collection of Technical Papers*, pp. 259–263. AIAA Paper No. 88–0801.
SHARMA, S. P., see Wise, P. C.
SIDDIQI, S., see Mott, R. C.
SINGER, J. L., see Hung, H-L. A. (two papers)
SINGER, J. L., see Lee, T. T.
SMITH, T., see Hegazi, G.
SMITH, T., see Polak-Dingels, P. (two papers)

* Non-COMSAT author.

SORBELLO, R. M., Zaghloul, A. I., Effland, J. E., and DiFonzo, D. F., "An Efficient Flat Plate Antenna for Direct Broadcast Applications," 18th European Microwave Conference, Stockholm, Sweden, September 1988, *Proc.*, pp. 295–299.
SORBELLO, R. M., see Assal, F. T. (two papers)
SORBELLO, R. M., see Lee, B. S.
SORBELLO, R. M., see Zaghloul, A. I.
SUYDERHOUD, H. G., and Dimolitsas, S., "Impact of Noise and Encoder Decoder Mistracking in ADPCM System Performance," IEEE International Conference on Acoustics, Speech and Signal Processing, New York, NY, April 1988, *Proc.*, pp. 16.S6.4.1–16.S6.4.4.
TOUGH, G., see Meulenberg, A.
TZENG, F. F., "Multipulse Excitation Codebook Design and Fast Search Methods for CELP Speech Coding," IEEE Global Telecommunications Conference, Hollywood, FL, November 1988, *Conference Record*, Vol. 1, pp. 18.6.1–18.6.5.
VAIDYANATHAN, H., "Long-Term Storage of Nickel-Hydrogen Cells," *Journal of Power Sources*, Vol. 22, 1988, pp. 221–228.
VAIDYANATHAN, H., and Earl, M. W., "Capacity and Pressure Variation of INTELSAT VI Nickel-Hydrogen Cells With Storage and Cycling," 23rd Intersociety Energy Conversion Engineering Conference, Denver, CO, July–August 1988, *Proc.*, D. Groswami, ed., The American Society of Mechanical Engineers, New York, NY, pp. 471–475.
WHITE, L., see Palmer, L.
WILLIAMS, A. E., and Bonetti, R. R., "A Mixed Dual-Quadruple Mode 10-Pole Filter," 18th European Microwave Conference, Stockholm, Sweden, September 1988, *Proc.*, pp. 966–968.
WILLIAMS, A. E., see Bonetti, R. R.
WISE, P. C.,* Kelly, W. H., and Sharma, S. P., "Critique of the Thermal Design Verification Program for a High-Power Communications Spacecraft," AIAA Thermophysics, Plasmadynamics and Lasers Conference, June 1988, San Antonio, TX. AIAA Paper No. 88–2648.
ZAGHLOUL, A. I., Sorbello, R. M., and Assal, F. T., "Development of Active Phased Arrays for Reconfigurable Satellite Antennas," European Space Agency COST 213/KUL Phased Array Workshop, Leuven, Belgium, October 1988, *Proc.*, pp. 135–150.
ZAGHLOUL, A. I., see Assal, F. T. (two papers)
ZAGHLOUL, A. I., see Lee, B. S.
ZAGHLOUL, A. I., see Mott, R. C.
ZAGHLOUL, A. I., see Potukuchi, J. R.
ZAGHLOUL, A. I., see Sorbello, R. M.

* Non-COMSAT author.

ZAKI, K. A.,* Chen, C., and Atia, A. E., "Modeling of Coupling by Probes in Dual Mode Cavities," IEEE MTT-S International Microwave Symposium, New York, NY, May 1988, *Digest*, pp. 515–518.

ZAKI, K. A.,* Chen, C., and Atia, A. E., "A Circuit Model of Probes in Dual Mode Cavities," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MT-36, No.12, December 1988, pp. 1740–1746.

* Non-COMSAT author.