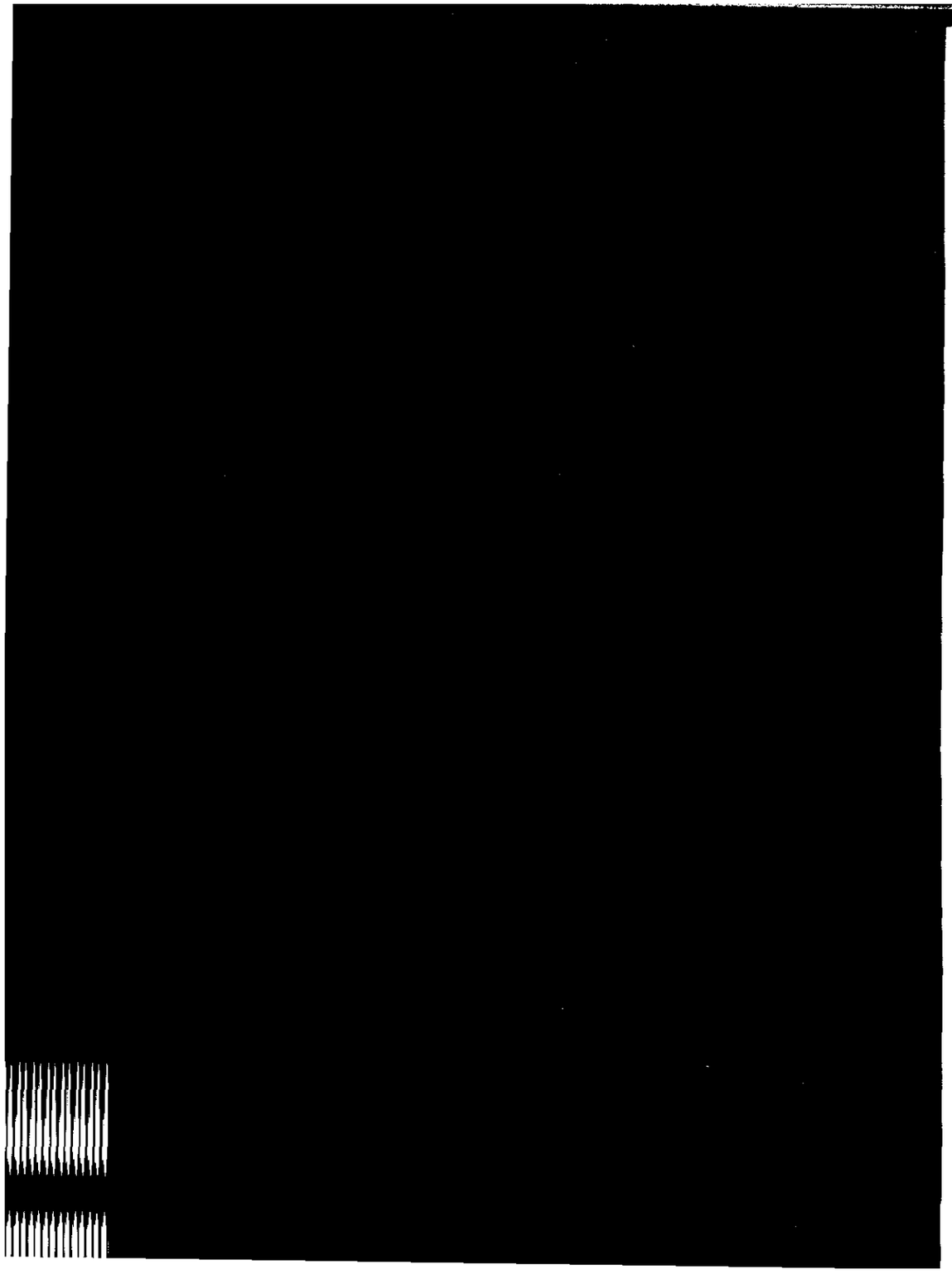


COMSAT Technical Review

Volume 9 Number 1, Spring 1979

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, $7 U.S.; two years, $12; three years, $15; single copies, $5; article reprints, $1. Make checks payable to COMSAT and address to Treasurer's Office, Communications Satellite Corporation, 950 L'Enfant Plaza, S.W., Washington, D.C. 20024, U.S.A.

COMSAT TECHNICAL REVIEW
Volume 9 Number 1, Spring 1979

Index: authentication, cryptosystem, one-time cipher, antispoofing, feedback communications

Authentication over low-data-rate channels with feedback

R. J. F. Fang and S. C. Lu

(Manuscript received July 17, 1978)

Abstract

A practical and unique authentication scheme is presented for use over low-data-rate channels with feedback. It employs one-time cipher to provide the highest possible level of security. The potential cryptosync problem existing in many one-time cipher cryptosystems is eliminated by using burst mode transmissions with preamble for burst synchronization. The effects of channel errors can be minimized in both the forward and the return links by using forward error correction (FEC) coding to detect and correct these errors. Such a scheme can be applied to access control of important data files as well as remote control of expensive or hazardous devices, apparatus, equipment, machinery, or processing plants.

Introduction

The information content of important and sensitive data files must be protected, and the authority of the accessing users must be checked. The data encryption standard (DES) [1] of the National Bureau of Standards (NBS) may be used for protecting the information content (or privacy) of these data files. The block-chaining technique of H. Feistal et al. [2] may be employed with DES to check the authenticity of the accessing users.

However, for some applications in which the total number of user accesses is limited or the total number of bits for transmission is relatively small, block-chaining of DES may not be desirable. Other simpler schemes for user authentication may be required.

In certain chemical, mechanical, electrical, or electrothermal plants, some operations and processes are very critical and/or irreversible. To avoid serious mistakes or to prevent the remote chance of sabotage, the control signal sent by the genuine controller must be authenticated before the operation or process is executed. Often the need for or the frequency of such an operation is small. Therefore, some simple encryption technique may be applied to authenticate the controller and its control signal.

There are also many expensive or hazardous devices, apparatus, equipment, machinery, or processing plants which must be controlled remotely. Some operations may significantly deteriorate the equipment or reduce the value or the life of the plants. Therefore, the control signals and their senders must be checked thoroughly before these operations are performed. Encryption techniques may be employed to authenticate the control signal and its sender and hence avoid possible errors and even third-party spoofings or sabotages.

Usually, only simple verification procedures are employed. For example, in an elementary computer security system, user authentication for data file access control may be performed as follows: Authorized users are given individual passwords. When a user wants to access the data file, the message word for accessing is prefixed by the user's own password and sent to the computer. The computer rejects the request if the password is not valid. Otherwise, the message word will be accepted and data file access will be permitted. Although channel errors can be avoided by using a cyclic-redundancy-check (CRC) code and automatic-repeat-request (ARQ) or simple FEC codes, this approach is not secure. For instance, if a transmitted signal is recorded and played back to the channel, the computer will consider this a legitimate new request and hence will grant access to the file.

Another example involves the remote control system of a hazardous or expensive apparatus. The controller is often given a secret but fixed password. When it is necessary to transmit a specific control word to the apparatus, the controller prefixes it with the password and sends it to the receiving apparatus. If the password corresponds to the word stored at the apparatus, the specific control word will be loaded into a control register and relayed back to the controller via the return link for verification. Otherwise, a rejection of the control word will be returned to the controller.

After the control word is verified, the controller sends the password followed by an execution word to execute the previously stored word at the apparatus. However, this remote control scheme, which has other possible weaknesses, is also not secure against playback attack.

This paper presents a simple authentication scheme that is suitable for low-data-rate channels with feedback. It employs the theoretically unbreakable one-time cipher [3] to achieve a high level of security and the feedback channel to return the received message to the sender for verification. With this scheme, the potential cryptosync problem existing in many one-time cipher cryptosystems can be avoided. It is useful for applications in which the date rate is low and the total amount of key bits to be used in enciphering and deciphering messages over a long period of time is reasonably small.

One-time cipher

One-time cipher [3] is an encryption scheme which performs bit-by-bit modulo 2 addition of the message bits with completely random, nonrepeating key bits. That is, when $\{m_i\}$ is a sequence of message bits and $\{k_i\}$ is a sequence of key bits, the cryptogram sequence $\{x_i\}$ is obtained as

$$x_i = m_i \oplus k_i$$

for all i. Deciphering the cryptogram requires knowledge of the key stream used for enciphering. The deciphering operation is as follows:

$$m_i = x_i \oplus k_i \quad , \qquad i = 1, 2, 3, \ldots .$$

One-time cipher is unbreakable both in theory and in practice [3]. Since the total number of key bits must equal the total number of message bits to be sent, one-time cipher is practical only if the size of the key or equivalently the total number of message bits is small. With modern large-scale integration (LSI) technology, a significant number of key bits may be stored in a few programmable read only memory (PROM) chips or other storage devices. Thus, the one-time cipher (or, as commonly known, the one-time pad) may be applicable to many secure communications systems. However, the key streams at the sender and the receiver must be completely synchronized, which may not be easily maintained. Therefore, cryptosync can be the most important problem in many one-time cipher cryptosystems.

Data formats

The data format for sending access request or remote control signals in the forward link from sender A to recipient B is depicted in Figure 1.

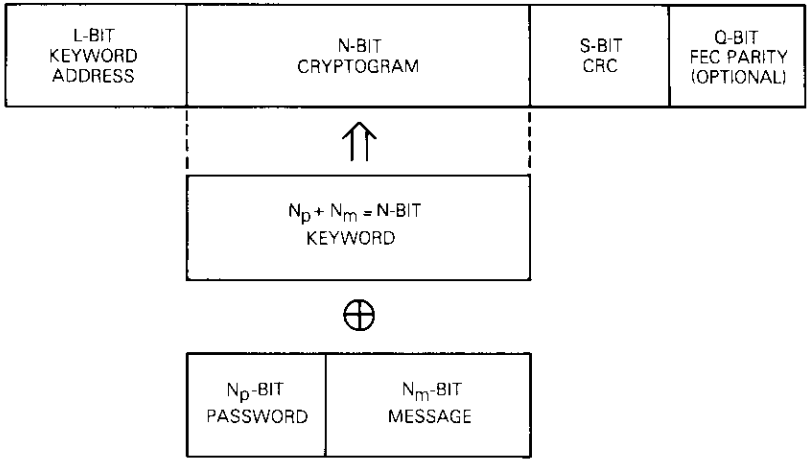


Figure 1. *Data Format in Forward Link*

The access request or control signal is represented by the N_m-bit message word. Prefixed to the message word is a password of N_p bits. These $N_p + N_m = N$ bits are then encrypted by an N-bit keyword from a one-time pad to form an N-bit cryptogram. This N-bit keyword is taken from the address represented by an L-bit address word. The L-bit keyword address is placed in front of the cryptogram. With the L-bit keyword address, sender A indicates to recipient B which of the keywords in the one-time pad has been selected in the generation of the cryptogram. Note that revealing the keyword address does not disclose the keyword itself. The size of the total number of keyword addresses 2^L must cover the total number of messages to be transmitted during the lifetime of the system.

To ensure a low probability of error in receiving the keyword address and the cryptogram, a CRC of S bits, which will yield an undetected error rate not exceeding 2^{-S}, is employed. For $S = 30$ bits, the CRC will ensure an undetected error rate no greater than 10^{-10}. The number of ARQs can be minimized by using FEC codes as an option. Thus, Q-bit parity checks for a rate $(L + N + S)/(L + N + S + Q)$ FEC code may be appended at the end of the block.

The return link may be shared with other transmissions. Figure 2 shows a data format in the return link excluding the other transmissions. The

feedback word may be encrypted by using only a portion of the same N-bit keyword used in the forward link.*

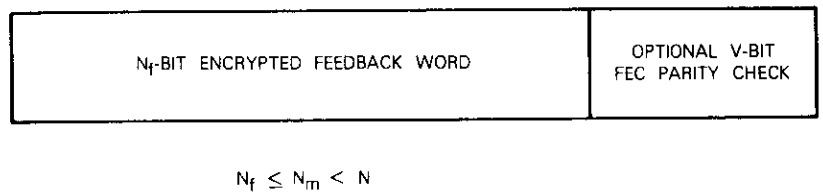


Figure 2. *Data Format in Return Link*

System description

Figure 3 is a simplified system diagram for authentication of messages over channels with feedback. The input message may be a computer data file access request or a remote control signal. Each message must be transmitted in at least three steps. The first involves sending the message to and storing the message at the receiving end. In the second step, the sender verifies the previously sent message as obtained by the recipient via the feedback channel. In the third step, the sender transmits the execution or confirmation word to the recipient.

The remote control problem can be considered as an example. The message of the desired control function and its associated parameters are first sent to the receiving end as shown in Figure 1. Namely, the N_m-bit message representing the desired control function and its associated parameters is first prefixed by an N_p-bit password to form a word of $N = N_p + N_m$ bits. From the location denoted by an L-bit keyword address, the sender obtains the N-bit keyword from a one-time pad to encrypt the N-bit message word bit-by-bit into an N-bit cryptogram. After prefixing the cryptogram with the L-bit keyword address, the sender uses an S-bit CRC code to encode the $L + N$ bits into an $(L + N + S)$-bit block, which is then transmitted to the recipient with or without FEC depending upon the application.

At the receiving end, the $L + N + S$ bits of input data are obtained either directly from the receiver if no FEC is used or from the FEC decoder if FEC is employed. These $L + N + S$ bits of input data are then processed

* In many applications, spoofing in the feedback link can be made impossible or infeasible. For simplicity, the reader may assume that the feedback link is secured from spoofing.

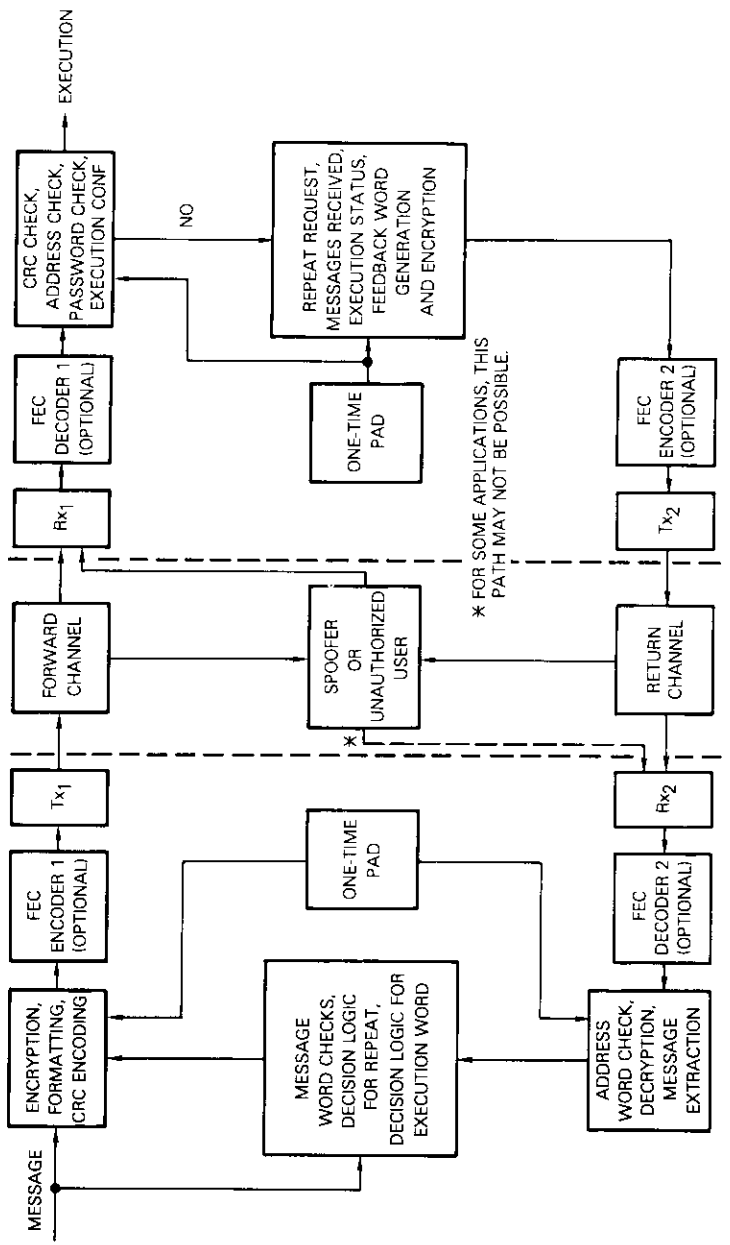


Figure 3. *System Block Diagram for Authentication Over Channels With Feedback*

by the authentication subsystem at the receiving end by using the algorithm shown in Figure 4 or a variation. At the receiving end, an identical one-time pad must be made available for processing the input data. If the CRC decoder does not detect an error, the received keyword address is permissible, and the received password is valid; then the present received keyword address will be used to obtain the deciphering key from the identical one-time pad. Otherwise, words representing the detection of errors, the reception of nonpermissible keywords, or the rejection of the received passwords will be loaded into the F-register for transmission to the sender via the feedback link. A portion of the deciphered N_m bits of message, representing the control function and its associated parameters, will also be loaded into the F-register, if the flag which indicates whether the recipient is executing the control has not been set. The content of the F-register can be relayed to the sender via the feedback link by encrypting these words with the key bits corresponding to the previous keyword address of the sender's identical one-time pad.

Upon receiving the control function and its associated parameters, the sender can verify whether they were received correctly. If not, the message will be retransmitted. Otherwise, a new message representing the execution or confirmation of the previous message will be transmitted, and execution will be performed at the receiving end. Upon receipt of a negative acknowledgment of the execution word, the sender will repeat the execution word. After each execution, the flag must be reset so that a new control function and parameters can be received and stored for further execution.

For various reasons, it may be necessary to abort an execution before its normal ending; a priority message word can be reserved for this purpose. (It is assumed that the execution can be terminated upon command.) Whenever this word is received, the ongoing operation will be stopped immediately.

The information stored in the F-register is relayed back to the sender via the feedback link by employing the key or a portion of the key contained in the location of the previous keyword address to encipher bit-by-bit the word contained in the F-register to form an encrypted feedback word. This word is then sent over the feedback channel with or without FEC using the format illustrated in Figure 2. Since the sender has an identical one-time pad and since the previous keyword address is also known, the contents of the F-register can be recovered from the encrypted feedback word. It should be noted that the feedback link can be quite different from the forward link and in general requires lower data rates because not everything transmitted in the forward link needs to be fed back.

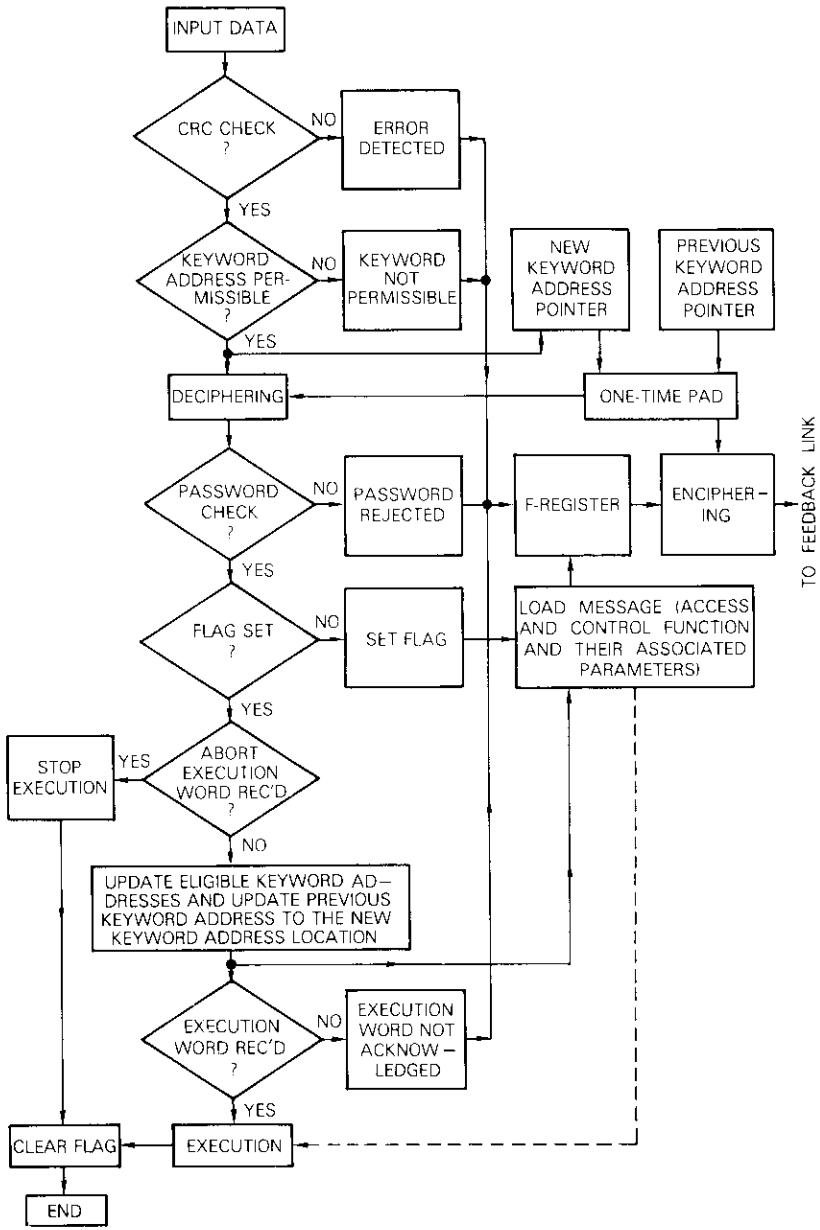


Figure 4. *Flow Chart for the Recipient's Authentication Subsystem*

Evaluation of the proposed authentication scheme

Since one-time cipher is unbreakable for one-way transmission both in theory and in practice, the amount of secrecy provided by this scheme is extremely high. Of course, the contents of the one-time pad must be randomly generated and should not be used more than once.

For the proposed authentication scheme, a playback attack by a spoofer in the forward link can be prevented, since each address of the one-time pad is used only once. The feedback link does not use a keyword address or password; therefore, only a portion of the pad is needed to encrypt certain status information as well as part of the message in the forward link. Thus, a playback in the feedback link can be prevented. A trial-and-error or exhaustive search attack in the forward link can be obstructed by the following methods:

a. Employing a long password to minimize the probability of a successful trial.

b. Informing the sender that use of a permissible keyword address has been attempted too frequently and should be temporarily blocked; therefore, this keyword should be placed at the bottom of the keyword list for future use and a new key word address should be selected.

c. Lengthening the transmission block by adding redundant bits so that the actual time available for the spoofer to spoof is minimized.

In the return link, trial-and-error or exhaustive search attack can be prevented because a spoofer can only perform such attacks when a control function and its associated parameters have already been loaded into the F-register at the receiving end. However, the time available for these attacks is simply too short compared to that required for the sender to execute the control or to confirm the previous message.

Since the total amount of required key bits in this authentication scheme is the same as that of message bits, the scheme is suitable only for low data rates and low-traffic applications. Otherwise, the one-time pad would be too large to be practical. For many applications the one-time pad can be implemented relatively simply with a few PROMS, and the authentication algorithm, such as the one in Figure 4, can be easily realized on a microprocessor. Cryptosystems employing one-time cipher usually encounter the problem of cryptosync. However, since burst mode transmission and the L-bit keyword address will automatically provide cryptosync whenever burst sync is achieved, there is no cryptosync problem in this scheme.

In the applications considered for the proposed scheme, transmission

efficiency should not be of any concern. Suppose that the execution word has N_e bits. Then the message expansion factor in the forward link is

$$\frac{(L + N + S) + (L + N_p + N_e + S)}{N_m + N_e} = \frac{2(L + N_p + S) + N_m + N_e}{N_m + N_e}$$

if FEC is not used. The message expansion factor becomes $[2(L + N_p + S + Q) + N_m + N_e]/(N_m + N_e)$, if FEC of Q parity check bits is employed.

When channel errors occur due to channel noise or external interference, these errors may or may not propagate depending upon whether or not FEC is used and the location of these errors. If FEC is used and there are more channel errors than can be corrected by the FEC codec, error propagation may occur; however, this is a phenomenon of FEC codes rather than the cryptosystem. If FEC is not employed and there are undetected errors in the keyword address, then error propagation may result because a different keyword will be used to decipher the cryptogram. Error propagation is limited within the message word and can be easily detected by the sender via the feedback information. If an undetected error occurs in the message portion, error propagation will not occur, since the deciphering process is bit-by-bit. In general, with adequate FEC and CRC, the probability of error propagation due to an undetected error is negligible. If it does occur, the sender can simply repeat the transmission.

Other variations of the baseline scheme

The proposed authentication scheme over channels with feedback may be modified to suit specific purposes depending upon the application and hardware implementation considerations. For example, in some anti-spoofing applications, message privacy is not important; thus, only the password requires protection. Because of implementation complexity as well as one-time pad size, it may be desirable to encipher only the password. Hence, a one-time pad of passwords can replace a fixed password being encrypted by a one-time pad. The password must be sufficiently long to minimize the probability of a successful random spoofing attack, 2^{-N_p+1}, to an acceptable level.

In another variation, the recipient (rather than the sender) selects the keyword according to a certain algorithm, *e.g.*, starting from the first keyword address and increasing automatically in ascending order. However, once a keyword is used, the recipient's keyword selector will never

return to this keyword address. The recipient informs the sender of the keyword to be used for enciphering the "next" message from the sender. This next keyword address is encrypted by the recipient with the last L bits of the present keyword, stored in a register, and then transmitted to the sender via the feedback link. The sender employs the last L bits of its present keyword to decipher the encrypted next keyword address. The L-bit keyword address can then be eliminated.

In feedback link transmission FEC may be employed. The sender can detect any errors in the keyword address by comparing the expected next keyword address according to the algorithm used by the recipient with the actual received next keyword address. The sender can use its expected next keyword to repeat the same message to the recipient. For example, assume that the simple ascending-order algorithm is used for selecting keywords. If the newly received next keyword address is advanced by one, errors have occurred in the previous keyword address. Otherwise, the pointer of the keyword address at the recipient's one-time pad has "jumped." The sender should then use the actual received next keyword address from the feedback link as its next keyword address.

In this variation, playback attack can be prevented, and cryptosync problems can also be avoided. Even though the keyword address pointer at the recipient's one-time pad could jump, this should not be of great concern because the probability of such a jump is extremely small and the new pointer location can be tracked relatively easily by searching in the vicinity of the expected next keyword address. In the worst case, it will require no more than 2^L trials to search out the correct new pointer location. Since the opponent does not have the password and the pad and since the pointer never revisits the same keyword address, the opponent has very little chance of moving the pointer.

Mode-switching considerations

Although the proposed scheme can be made reliable by incorporating sufficient redundancy, some applications may require clear-mode authentication as a backup. Therefore, automatic switching between clear and secure modes is necessary because the transmission of a control signal is impossible when a component failure occurs at the recipient's secure authentication subsystem. Mode switching can be implemented by using a highly reliable internal test program to periodically check the input/output relationship of the secure authentication subsystem. If an error is

detected and confirmed, the secure authentication subsystem is abandoned, and the clear-mode authentication subsystem will be used.

In some applications, "updating" messages can be periodically transmitted to maintain the authentication system in the secure mode. If an updating message is not received within a predetermined interval, the recipient's secure authentication system will automatically be switched to the redundant system or the clear mode. Upon command, the mode can be switched back to secure mode by the sender, if desired. This approach of sending updating messages seems simpler and more desirable. However, since a much larger one-time pad is required, it may not always be applicable.

Conclusion

A simple and unique authentication scheme has been proposed which uses one-time cipher for applications in low-data-rate channels with feedback. Since one-time cipher is unbreakable both in practice and in theory, this scheme appears very secure. Implementation complexity is simple and can be realized with a few ROMS (or PROMS) and microprocessors for many applications. The one-time pad should be properly prepared so that it closely represents a completely random sequence and should remain secret. The proposed authentication scheme and its variations can be directly applied to many access control or remote control systems without any cryptosync problems.

Acknowledgment

The authors wish to thank Dr. Lin-Nan Lee for some useful discussions and suggestions.

References

[1] "Data Encryption Standard, Federal Information Processing Standard (FIPS)," Publication 46, National Bureau of Standards, U.S. Department of Commerce, January 1977.

[2] H. Feistal, W. A. Notz, and J. L. Smith, "Some Cryptographic Techniques for Machine-to-Machine Data Communications," *Proc. IEEE*, Vol. 63, No. 11, November 1975, pp. 1545–1554.

[3] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, Vol. 28, October 1969, pp. 656–715.

Russell J. F. Fang is Manager of the Communications Systems Analysis Department of the Transmission Systems Laboratory at COMSAT *Laboratories. He is responsible for research and studies of advanced communications systems concepts and techniques for international and domestic satellite communications. Born in Chungking, China, he received a B.S. in electrical engineering from National Taiwan University in 1962, and an M.S. and a Ph.D. in electrical engineering from Stanford University in 1964 and 1968. Before joining* COMSAT *in 1968, he was employed by Stanford Electronics Laboratories (1965–1968), Stanford Research Institute (1964–1965), and the Chinese Air Force Electronics and Ordnance Division, Taiwan (1962–1963).*

Shyue-Ching Lu was born in Taiwan, April 1, 1949. He received the B.S. in engineering science from Cheng Kung University, Taiwan, in 1970, and the Ph.D. in electrical engineering from the University of Hawaii, in 1976. He was a member of the technical staff of the Telecommunication Laboratories, Ministry of Communications, Republic of China. During the 1976–1977 academic year, he was also an adjunct associate professor of the Institute of Electronics, National Chiao Tung University, Taiwan. From August 1977 to July 1978, he was on leave with the Communications Systems Analysis Department, COMSAT *Laboratories. He is now a project manager, responsible for the development of a digital transmission system in the Telecommunication Laboratories. He is a member of IEEE.*

Index: Chinese remainder theorem, cryptology, algorithm, cryptosystem

A simple and effective public-key cryptosystem

S. C. Lu and L. N. Lee

(Manuscript received August 12, 1978)

Abstract

A simple and effective method for realizing a public-key cryptosystem is presented. This scheme is based upon a variation of the Chinese Remainder Theorem and the fact that no efficient algorithm is known for factoring a large composite number. The encryption algorithm is simple; it consists of two multiplications and one addition modulo a large number. The decryption algorithm is also simple, consisting of two modulo operations and solving two linear simultaneous equations with two unknowns. The task of cryptanalysis appears to be difficult since it requires factoring a large composite integer. The proposed cryptosystem appears to provide a high level of security and can be used in high-data-rate communications systems.

Introduction

The concept of public-key cryptosystems was introduced by Diffie and Hellman [1] in 1976. In contrast to a conventional cryptosystem in which encryption and decryption keys are identical, a public-key cryptosystem contains distinct encryption and decryption keys. These keys must be designed so that it is computationally infeasible to obtain the decryption key from the encryption key. Since the encryption key can be publicly disclosed without compromising the decryption key, each network user can

place this encryption key in a public file and can send encrypted messages to any other user by utilizing the intended receiver's encryption key. This can be accomplished without the *a priori* exchange of a cryptographic key. The intended receiver recovers messages with a secret decryption key, to the exclusion of other users.

Three schemes have been proposed to construct such cryptosystems. Rivest, Shamir, and Adleman [2], by using the fact that no efficient algorithm for factoring a large composite number is yet known, proposed a method to implement a public-key cryptosystem. Although this scheme appears mathematically attractive, the encryption and decryption operations (exponentiation modulo a large number) are relatively complex; its computing time is about $T(r) \log_2 (r)$, where r is the large composite number used in the scheme and $T(r)$ is the time required to multiply two numbers, modulo r.

Merkle and Hellman [3], observing that knapsack problems are generally difficult to solve, proposed a scheme to implement the public-key cryptosystem. Although both the encryption and the decryption operations are simple in the Merkle-Hellman scheme, a message (or bandwidth) expansion factor of at least 2 is required when the suggested parameters are chosen. McEliece [4], recognizing that no efficient method for decoding a general linear code is yet known, also proposed a scheme to implement the public-key cryptosystem. However, the implementation of its encryption and decryption processes is extremely complex using current technology.

This paper presents a simple and effective method for realizing a public-key cryptosystem, based upon a variation of the Chinese Remainder Theorem and the fact that no efficient algorithm is yet known for factoring a large composite number. The encryption and decryption algorithms of the proposed scheme are simple to implement and can thus be used in high-data-rate communications systems while a very high level of security is provided.

Description of the proposed method

This section explains the encryption and decryption algorithms. All parameters and variables are integers, unless otherwise specified.

Encryption algorithm

It is assumed that m_1 and m_2 represent a message, where $0 < m_1 < M_1$ and $0 < m_2 < M_2$. Then, to encrypt the message with the proposed method using a public encryption key with parameters (c_1, c_2, r), the following

computation is performed to produce the cryptogram x:

$$x = (c_1 m_1 + c_2 m_2) \bmod r \quad . \tag{1}$$

Decryption algorithm

When the cryptogram x is decrypted using the secret decryption key with parameters $(a_{11}, a_{12}, a_{21}, a_{22}, p_1, p_2)$, the following computations are necessary to recover the message represented by m_1 and m_2. First,

$$x_1 = x \bmod p_1 \quad , \qquad x_2 = x \bmod p_2 \tag{2}$$

must be computed, and then it can be determined that

$$m_1 = \frac{x_1 a_{22} - x_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}} \tag{3}$$

$$m_2 = \frac{x_2 a_{11} - x_1 a_{21}}{a_{11} a_{22} - a_{12} a_{21}} \quad . \tag{4}$$

Choice of parameters

The constituents of the public encryption key (c_1, c_2, r) are chosen so that c_1 and r, as well as c_2 and r, are relatively prime and that r is the product of two large prime numbers, p_1 and p_2, *i.e.*,

$$r = p_1 p_2 \quad . \tag{5}$$

To ensure that every message is scrambled by the encryption operation, it is suggested that $c_1 + c_2$ be greater than r.

The secret decryption key consists of six parameters; two, p_1 and p_2, are prime factors of r, and the other four are derived from c_1, c_2 and p_1, p_2 as follows:

$$a_{11} = c_1 \bmod p_1 \quad , \qquad a_{12} = c_2 \bmod p_1 \tag{6}$$

$$a_{21} = c_1 \bmod p_2 \quad , \qquad a_{22} = c_2 \bmod p_2 \quad . \tag{7}$$

These four integers must satisfy the conditions that $a_{11} a_{22} - a_{12} a_{21} \neq 0$.

Upper limits M_1 and M_2 on integers representing a message must satisfy the following conditions:

$$M_1 \leq \left[\frac{1}{2} \min\left\{\frac{q}{a_{11}}, \frac{q}{a_{21}}\right\}\right] \tag{8}$$

$$M_2 \leq \left[\frac{1}{2} \min\left\{\frac{q}{a_{12}}, \frac{q}{a_{22}}\right\}\right] \tag{9}$$

where $q = \min\{p_1, p_2\}$ and $[y]$ denotes the integer part of a real number y.

The following equations reveal that the decryption algorithm correctly decrypts the cryptogram produced by the encryption algorithm. Taking modulo p_1 on both sides of equation (1) and noting that p_1 is a factor of r yields

$$x_1 = x \bmod p_1 = (c_1 m_1 + c_2 m_2) \bmod p_1 \tag{10a}$$

or

$$x_1 = x \bmod p_1 = (a_{11} m_1 + a_{12} m_2) \bmod p_1 \tag{10b}$$

and

$$x_1 = x \bmod p_1 = a_{11} m_1 + a_{12} m_2 \tag{11}$$

where the equality in equation (10) is obtained by using a property of modulo operation and equation (6), and the equality in equation (11) follows from equation (8). Similarly,

$$x_2 = x \bmod p_2 = a_{21} m_1 + a_{22} m_2 \quad . \tag{12}$$

Equations (11) and (12) are a pair of linear simultaneous equations with two unknowns, m_1 and m_2. Since $a_{11}a_{22} - a_{12}a_{21} \neq 0$, as required, the two unknowns can be uniquely solved, producing the results given in equations (3) and (6). Integer solution is guaranteed because such a solution exists.

Both the encryption and the decryption algorithms can be easily implemented. The encryption algorithm requires only two multiplications, one addition, and one modulo operation. The modulo r operation may be incorporated with each multiplication to simplify implementation. Hence, the computing time of the encryption operation is about $2T(r)$, where $T(r)$ is the time required to multiply two numbers (modulo r). The decryption algorithm requires two modulo operations, four multiplications, two subtractions, and two divisions. The denominators of equations (3) and (4) can be precomputed.

Thus, the computing time of the decryption operation is approximately the same as that required to perform eight multiplications of two integers, one with size p_1 and the other with size a_{11}. It should be noted that the computing time of the encryption algorithm dominates when r is large. Consequently, the proposed scheme can be operated at a data rate of about $(1/2) \log_2 (r)$ times that which can be achieved by the Rivest-Shamir-Adleman scheme. When r is a 100-decimal-digit number, the data rate improvement factor is about 150.

Selection of the encryption and decryption keys for each user is also easy. Large random prime numbers for p_1 and p_2 can be generated by using the efficient probabilistic algorithm developed by Solovay and Strassen [5]. When p_1 and p_2 are specified, c_1 and c_2 can be computed by first selecting a_{11}, a_{12}, a_{21}, and a_{22} such that $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Then two additional numbers, b_1 and b_2, should be computed using a variation of Euclid's algorithm such that

$$b_1 p_1 + b_2 p_2 = 1 \tag{13}$$

where b_1 and b_2 exist since p_1 and p_2 are relatively prime. Then c_1 and c_2 can be computed according to the following formulas:

$$c_1 = [(a_{21} - a_{11}) b_1 p_1 + a_{11}] \bmod r \tag{14a}$$

or

$$c_1 = [(a_{11} - a_{21}) b_2 p_2 + a_{21}] \bmod r \tag{14b}$$

and

$$c_2 = [(a_{22} - a_{12}) b_1 p_1 + a_{12}] \bmod r \tag{15a}$$

or

$$c_2 = [(a_{12} - a_{22}) b_2 p_2 + a_{22}] \bmod r \quad . \tag{15b}$$

Equations (14a) and (14b) are derived from equation (13) by multiplying both sides of the equation by $(a_{21} - a_{11})$ and then rearranging terms. Similarly, equations (15a) and (15b) are derived by multiplying both sides by $(a_{22} - a_{12})$ and then rearranging terms.

Discussion

The underlying mathematics in the encryption and decryption algorithms presented in the preceding section is a variation of the Chinese Remainder Theorem [6]. The following basic principles are used in the design of the proposed scheme:

a. Encryption uses a transformation [see equation (1)] in which a public encryption key (c_1, c_2, r) satisfies the conditions that r is the product of two relatively prime numbers, and that c_1 and r, as well as c_2 and r, are relatively prime numbers.

b. Decryption is equivalent to solving linear simultaneous equations derived from the encryption algorithm.

Several alternative designs of similar public-key cryptosystems based on these principles have also been successful. However, it is believed that the proposed scheme can be easily implemented and will provide a high level of security. Although the encryption key (c_1, c_2, r) is placed in the public file, no effective algorithm is known to find the secret decryption key (a_{11}, a_{12}, a_{21}, a_{22}, p_1, p_2) without first determining p_1 and p_2. In addition, no effective means is known to find p_1 and p_2 without factoring r. Factorization of a large composite integer is difficult; the most efficient algorithm currently known [2] requires about $\exp\{\sqrt{ln\,(r)\,ln\,[ln\,(r)]}\}$ steps to factor r. A computer with a speed of 1 μs per operation when r is a 100-decimal-digit number would require 74 years to factor r; when r is 200 decimal digits, 3.8×10^9 years would be necessary.

Cryptanalysis of the proposed scheme using a known plaintext attack does not appear to be rewarding, and a chosen plaintext attack does not seem to present a threat. Cryptanalysis by exhaustive search of a_{11}, a_{12}, a_{21}, and a_{22} can be easily denied when each of these parameters is at least 16 bits long.

There is a small factor of message (or bandwidth) expansion from messages to cryptograms due to the restrictions on integers representing a message. A cryptogram will require a log (r) bit representation and a message will require a log (M_1M_2) bit representation; therefore, the expansion factor is equal to (log r)/[log (M_1M_2)]. Since M_1 and M_2 are of the same order of magnitude, each about $q/(2a)$, and since q is in the order of $\sqrt{r}$, the denominator is about [log (r)] − [2 log (2a)]. The expansion factor is approximately equal to $1 + [2(1 + \log a)]/[\log r]$, where a is the largest of the a_{ij}'s, and the base of the logarithm is 2. When r is 320 bits and a is 16 bits, the expansion factor is approximately 1.1.

An example

A simple example can be used to illustrate the operation of the enciphering and deciphering procedures and to enhance the previous discussion. If two prime numbers, $p_1 = 97$ and $p_2 = 103$, and four arbitrary integers, $a_{11} = 3$, $a_{12} = 2$, $a_{21} = 5$, and $a_{22} = 4$, are chosen as the deciphering keys, it is necessary to verify whether $a_{11}a_{22} - a_{12}a_{21} = 0$. If this occurs, another set of four integers must be used. Since $a_{11}a_{22} - a_{12}a_{21} = 2$, the set of deciphering keys can be used to derive the enciphering keys. First, $r = p_1p_2 = 9991$, and

$$1 = 17(97) - 16(103)$$

can be derived by Euclid's algorithm. According to equation (13), this is equivalent to $b_1 = 17$ and $b_2 = -16$. When these values are substituted into equation (14), the enciphering keys

$$c_1 = 2(17)\,(97) + 3 = 3301$$

and

$$c_2 = 2(17)\,(97) + 2 = 3300$$

can be obtained. The upper limits on integers representing a message which can be obtained with equations (8) and (9) are also derived, *i.e.*,

$$M_1 \leq \left[\frac{1}{2}\min\left(\frac{97}{3}, \frac{97}{5}\right)\right] = 9$$

and

$$M_2 \leq \left[\frac{1}{2}\min\left(\frac{97}{2}, \frac{97}{4}\right)\right] = 12 \quad .$$

At this point, a basic public-key cryptosystem can be constructed. For example, if c_1 and c_2, as well as M_1 and M_2, are announced, a user desiring to send two blocks of messages, $m_1 = 7$ and $m_2 = 5$ (or $m_1 = 0111$ and $m_2 = 0101$ in binary form assuming a 4-bit block), can encipher these messages to obtain the cryptogram

$$x = [7(3301) + 5(3300)] \bmod 9991 = 9634$$

or $x = 10010110100010$ in binary form. With the deciphering key kept secret, only the legitimate receiver can determine the original message if the parameters are properly chosen. The legitimate receiver with the deciphering key can first calculate

$$x_1 = 9634 \bmod 97 = 31$$

and

$$x_2 = 9634 \bmod 103 = 55$$

and then solve the linear equations:

$$3m_1 + 2m_2 = 31$$

and

$$5m_1 + 4m_2 = 55$$

to obtain the original message by equations (3) and (4).

Several modifications may be made. For example, since it is desirable to block the messages in equally sized blocks, and it is necessary to exclude zero as a legitimate message for security reasons, the messages can be segmented into 3-bit blocks and incremented by one before enciphering. As long as the preprocessing scheme is announced publicly, there should be no difficulty in enciphering and deciphering. Furthermore, since $c_1 + c_2 = 6601$, $c_1 + 2c_2 = 9902$, and $2c_1 + c_2 = 9901$ are all less than $r = 9991$, the cryptogram appears obvious when $m_1 = 1$, $m_2 = 1$; or $m_1 = 1$, $m_2 = 2$; or $m_1 = 2$, $m_2 = 1$. One possibility is to find another set of enciphering keys such that $c_1 + c_2 > r$; the other alternative is to require that $m_2 \geq 3$. Since $M_2 = 12$, the second block may be incremented by three. Thus, every 3-bit message can be enciphered securely.

Since p_1 and p_2 are chosen to be small in this case, the upper limits M_1 and M_2 are even smaller, and the bandwidth expansion factor slightly exceeds two. It also can be argued that the system cannot be secure since the cryptanalyst can build a table of corresponding messages and cryptograms by precomputation and decipher the cryptograms by table look-up. In the actual situation when p_1 and p_2 are very large, the expansion factor is small and the total computation of the message-cryptogram pairs is simply infeasible because of the large size of the message space.

This example does reveal a potential problem if the keys are not chosen properly. Since c_1 and c_2 are of about the same magnitude, it is fairly easy to estimate the size of $m_1 + m_2$, primarily because $a_{21} - a_{11} = a_{22} - a_{12} = 2$, causing the dominant terms of equations (14) and (15) to be equal. Generally, when a_{ij} are reasonably large, the problem can easily be avoided.

References

[1] W. Diffie and M. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, IT-22, No. 6, November 1976, pp. 644–654.

[2] R. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, Vol. 21, No. 2, February 1978, pp. 120–126.

[3] R. Merkle and M. Hellman, "Hiding Information and Receipts in Trap Door Knapsacks," 1977 IEEE International Symposium on Information Theory, October 10–14, 1977, Cornell University, Ithaca, New York.

[4] R. McEliece, "A Public-Key System Based on Algebraic Coding Theory," JPL DSN Progress Report, 1978.

[5] R. Solovay and V. Strassen, "A Fast Monte-Carlo Test for Primality," *SIAM Journal on Computing*, March 1977, pp. 84–85.

[6] D. Knuth, *Seminumerical Algorithms: The Art of Computer Programming*, Reading, Massachusetts: Addison-Wesley Publishing Company, Vol. 2, 1969.

Shyue-Ching Lu was born in Taiwan, April 1, 1949. He received the B.S. in engineering science from Cheng Kung University, Taiwan, in 1970, and the Ph.D. in electrical engineering from the University of Hawaii, in 1976. He was a member of the technical staff of the Telecommunication Laboratories, Ministry of Communications, Republic of China. During the 1976–1977 academic year, he was also an adjunct associate professor of the Institute of Electronics, National Chiao Tung University, Taiwan. From August 1977 to July 1978, he was on leave with the Communications Systems Analysis Department, COMSAT Laboratories. He is now a project manager, responsible for the development of a digital transmission system in the Telecommunication Laboratories. He is a member of IEEE.

Lin-Nan Lee was born in Kaohsinng, Taiwan, on February 24, 1949. He received the B.S.E.E. degree (1970) from National Taiwan University, the M.S. (1973) and Ph.D. in electrical engineering (1976) from the University of Notre Dame. From 1970 to 1971 he was a communications officer in the Chinese Air Force. From 1971 to 1975, he was a research assistant at the University of Notre Dame. In 1975, he joined the Linkabit Corporation where he was involved in the design and development of packet communication protocols for the satellite networks. Subsequently, he joined COMSAT *in 1977 as a staff member of the Communications Systems Analysis Department. He is a member of IEEE and Sigma Xi.*

Index: Chinese remainder theorem, cryptology, algorithm, cryptosystem

A multiple-destination cryptosystem for broadcast networks

L. N. LEE AND S. C. LU

(Manuscript received March 1, 1978)

Abstract

An encryption and decryption scheme is presented which is particularly suitable for multiple-destination broadcast networks such as those of communications satellites. In a typical broadcast channel, information from different sources is multiplexed together at the transmit station before transmission, and subsequently demultiplexed at various receive stations. To ensure the privacy of communications between the transmit station and the receive stations, it is often desirable to encrypt each point-to-point link with a link-specific cryptographic key. However, the implementation of such a cryptosystem could be costly, or the privacy of individual links could be compromised.

This paper proposes an integrated enciphering and multiplexing scheme for the transmit station to perform a single enciphering and multiplexing operation on messages directed to different destinations. Without acknowledgment of the secret keys of other stations, a particular receive station cannot decipher messages for other earth stations. It is believed that this technique can provide a simple and cost-effective solution to the link encryption of multiple-destination broadcast communications networks.

Introduction

In a broadcast communications network such as that of a geostationary satellite, information from different sources and destinations is multi-

plexed together before transmission, and a complementary demultiplexing operation is accomplished at each receive location. To protect the privacy of communications between each pair of transmit and receive stations, it is often desirable to encrypt each point-to-point communications link with a link-specific cryptographic key. Figure 1 shows two commonly used approaches. One is to encipher messages before multiplexing and decipher them after demultiplexing; the other is to encipher after multiplexing and decipher before demultiplexing. (The locations of the enciphering and deciphering devices are shown by the circled numbers 1 and 2 in the figure.)

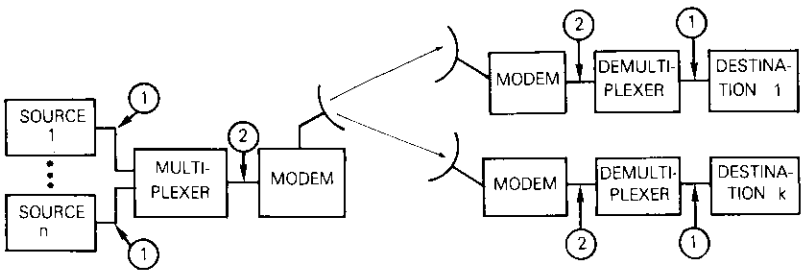


Figure 1. *A Typical Broadcast Network and Two Commonly Used Approaches to Provide Communications Security*

The disadvantage of the first approach is that it requires the same number of encipherers as input ports of the multiplexer, although the operating speed of the encipherers can be slower. In the first approach, since the messages from a given source in each session of conversation are usually intended for the same destination, the encryption keys can generally be set up on a session-by-session basis. In the second approach, however, it is essential that the encipherers are capable of recognizing the destination of each message and of changing the enciphering keys accordingly on an instantaneous basis. Although only one encipherer is required at each station, the capability of changing keys dynamically and of operating at a higher speed leads to costly implementation. The disadvantages of these approaches indicate the need for a new concept to implement a multiple-destination cryptosystem.

It is desirable that the cryptogram of a multiple-destination cryptosystem be a function of messages to all destinations, and that, with a particular deciphering key, each station can recover from the cryptogram only the messages destined to that station. The transmit station, with the knowledge of the key parameters of each destination, should be able to encipher all of the messages in a simple way such that the resultant cryptogram appears as a single entity and no correspondence between each message and individual segments can be observed. The concept of multiple-destination cryptosystems is quite similar to that of code-division multiplex (CDM) in which information streams are multiplied by a selected set of quasi-orthogonal codes and added together. The major difference is that the new concept requires only a small amount of bandwidth expansion and provides better privacy.

A multiple-destination cryptosystem

This section describes a multiple-destination cryptosystem that is based on the well-known theorem of Sun Tsu, the so-called "Chinese Remainder Theorem" [1], [2]. It is assumed that the message (a block of L symbols) directed to the ith user is represented by a positive integer, m_i. Also n users and n distinct prime numbers, p_i, are chosen so that $0 \leq m_i < p_i$ for $i = 1, 2, \ldots, n$. Let

$$P \equiv \prod_{i=1}^{n} p_i \quad ; \qquad d_i \equiv \prod_{\substack{j=1 \\ (j \neq i)}}^{n} p_j \tag{1}$$

and

$$c_i \equiv d_i(d_i^{-1} \bmod p_i) \qquad \text{for } 1 \leq i \leq n \tag{2}$$

where $d_i^{-1} \bmod p_i$ is the multiplicative inverse of d_i in the field of mod p_i such that $d_i(d_i^{-1} \bmod p_i) \bmod p_i = 1$. Then, the messages m_i are enciphered by the secret parameters c_i to generate the cryptogram

$$x = \sum_{i=1}^{n} m_i c_i \bmod P \quad . \tag{3}$$

The parameter p_i is the secret deciphering key at user i and must be sent to user i a priori via secure means.

At user i, the message m_i can be recovered with the secret deciphering key p_i by computing

$$m_i = x \bmod p_i \qquad \text{for } 1 \leq i \leq n \quad . \tag{4}$$

This process can be easily verified by noting that (y mod P) mod p_i

$= y \bmod p_i$ for any integer y, and that $[d_i(d_i^{-1} \bmod p_i)] \bmod p_j = 1$ if $i = j$, and 0 otherwise.

Therefore, the enciphering and deciphering processes are very simple. At the transmitting user, the messages m_i directed to user i are enciphered by first multiplying m_i by the secret enciphering parameters c_i in mod P and then summing the products in mod P to obtain the cryptogram x. The receiving user i divides the cryptogram x by its secret deciphering key p_i to obtain the remainder $x \bmod p_i$, which gives the original message m_i. However, for any other user, $j \neq i$, it is extremely difficult to determine m_i from x without the values of c_i and p_i.

The system configuration of such a multiple-destination cryptosystem is depicted in Figure 2, in which n users transmit different messages to each other. It is assumed that m_i^j denotes the message originating from user j and destined to user i, and that c_i^j represents the enciphering parameters of user j for messages destined to user i. Also, p_i^j are n distinct prime numbers for a given j, $P^j = \prod_i p_i^j$, $d_i^j = \prod_{k \neq i} p_k^j$, and $c_i^j = d_i^j[(d_i^j)^{-1} \bmod p_i^j]$.

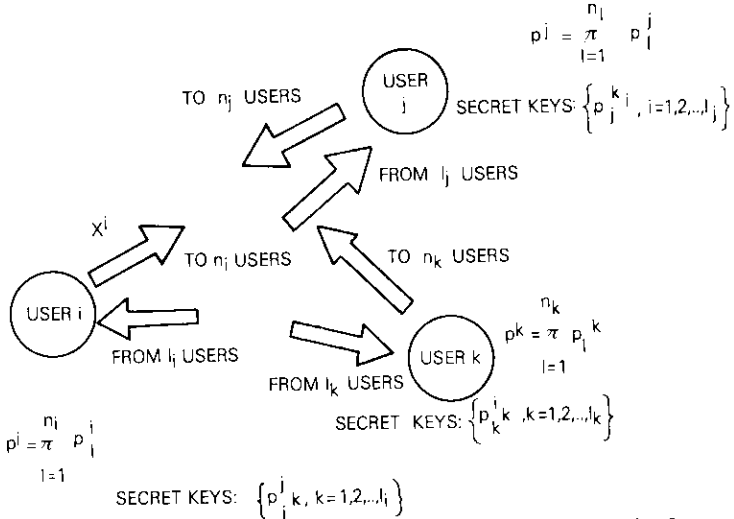


Figure 2. *System Configuration of the Proposed Multiple-Destination Cryptosystem*

Then, from user j, the messages m_i^j for all i, are enciphered by the secret parameters c_i^j to generate the cryptogram

$$x^j = \sum_{i=1}^{n_j} m_i^j c_i^j \bmod P^j \tag{5}$$

where n_j is the number of message destinations from user i. The parameters p_i^j for $j = 1, 2, \ldots, n$ are the secret deciphering keys which must be sent to user i from the other users via secure means. Since P^j are known to each user, the transmitting user j requires up to n_j multipliers and one adder in mod P^j.

The receiving user i may receive a total of l_i cryptograms from l_i transmitting users. The message m_i^j from each of these l_i transmitting users is recovered by using the secret deciphering key p_i^j, i.e., $m_i^j = x^j \bmod p_i^j$ for each j. Thus, at the receiving user i, up to l_i dividers divide the corresponding cryptograms from the l_i transmitting users by the deciphering keys p_i^j to obtain the remainders $m_i^j = x^j \bmod p_i^j$.

For FDMA applications, a total of l_i dividers at user i will be required; for TDMA applications a single divider may suffice if the divisor can be changed dynamically. The cryptosystem has the advantages of the two approaches (Figure 1) but not their disadvantages. In particular, the function of a multiplexer is replaced by a simple adder, and the functions of a demultiplexer and decipherers are replaced by dividers. Furthermore, if the bandwidth between the sources (or destinations) and the concentrator is sufficient, the multiplexers (or dividers) may possibly be placed at the sources to provide some degree of communications security between the sources and the concentrator.

Rivest et al. [3] observed that no efficient algorithm is known which factors a large composite integer. Consequently, they invented a powerful public-key cryptosystem based on the fact that a successful cryptanalytic attack of the cryptosystem is equivalent to factoring a large composite number. The multiple-destination cryptosystem proposed in this paper contends that successful cryptanalysis requires the knowledge of p_i, which must be obtained by factoring P, if neither c_i nor p_i are available. Therefore, it is usually unnecessary to conceal the product $P = \prod_i p_i$, although its concealment may be desirable for additional security.

If $P = p_1 p_2$, the product of two deciphering keys, is known, the strength of the cryptosystem is weakened, since two users may be able to derive the deciphering keys of each other. This difficulty can be avoided with a slightly modified implementation. For example, each of the two users may be given two pairs of enciphering and deciphering keys, and each can be treated as two separate users who are sending messages alternatively. That is, if $P = p_1 p_2 p_3 p_4$ is known, each user can obtain the product of the deciphering keys (e.g., $p_1 p_2$) of the other user, but not the individual keys, because the factorization of large composite numbers would be necessary.

Another potential weakness of the cryptosystem occurs when user j

broadcasts an identical message to all users or to a subgroup of destinations in the same cryptogram x^i, since the destinations (user i) may then be able to determine the enciphering keys c_i^j of each other. This difficulty can be avoided by broadcasting the same message to different users in different cryptograms. Because the majority of the traffic in a broadcast network is point-to-point, the additional bandwidth required for the broadcast messages is almost negligible.

It is believed that the cryptosystem is secure against ciphertext attack unless the cryptanalyst has more than one plaintext and cryptogram pair. For example, if x_i and x_i' are the corresponding cryptograms for messages m_i and m_i' from user i, that is,

$$m_i = x_i \bmod p_i \tag{6a}$$

and

$$m_i' = x_i' \bmod p_i \tag{6b}$$

then both $x_i - m_i$ and $x_i' - m_i'$ are multiples of p_i. The greatest common divisor of $x_i - m_i$ and $x_i' - m_i'$, which is usually a small multiple of p_i, can easily be calculated by using the Euclid algorithm. This difficulty can be circumvented by slightly modifying the enciphering keys, as described in the following section.

A multiple-destination cryptosystem against known plaintext attack

With the previously defined values of m_i, p_i, and d_i, the same basic enciphering procedure, and b_i as an arbitrarily chosen secret constant relatively prime to P, a multiple-destination cryptosystem can be constructed by choosing the enciphering keys

$$c_i = b_i d_i \bmod P \quad . \tag{7}$$

Since

$$x \bmod p_i = (m_i c_i) \bmod p_i \tag{8}$$

the intended decipherer, with knowledge of the secret parameters, c_i and p_i, and thus $c_i^{-1} \bmod p_i$, can easily compute $m_i = (x \bmod p_i) \cdot (c_i^{-1} \bmod p_i) \bmod p_i$. The decipherer implementation is similar to that described previously, except that an additional multiplier in mod p_i is required at the output of the divider. Since both c_i and p_i are secret, knowledge of the corresponding pairs of x and m_i does not facilitate the derivation of either

c_i or p_i, or $(c_i^{-1} \bmod p_i)$. Furthermore, to ensure that the cryptogram

$$x = \sum_i m_i c_i \bmod P \tag{9}$$

appears sufficiently random, it is desirable that

$$x \neq \sum_i m_i c_i \tag{10}$$

or equivalently,

$$\sum_i m_i c_i > P \tag{11}$$

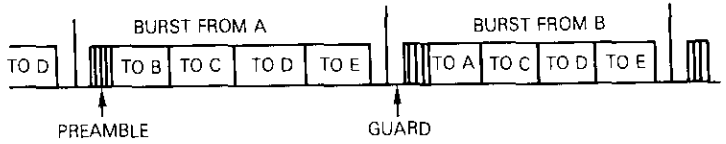
for all possible combinations of m_i. This can be achieved by excluding zero as a possible value for m_i and by requiring that the c_i are chosen such that $\sum_i c_i$ is at least greater than P.

This appears to be the first cryptosystem specifically designed for multiple-destination broadcast networks, for which messages to different destinations are blended together to increase the strength of the cryptosystem without interfering with each other. Although this cryptosystem is similar to a CDM, its bandwidth utilization is orders of magnitude more efficient, and its ability to resist known plaintext attack provides greater security than most CDMs. It is not a public-key cryptosystem [3], [4]; however, the deciphering parameters, p_i and $c_i^{-1} \bmod p_i$, can be derived from the enciphering key, c_i, although the appearance of the enciphering and deciphering keys is different.

Application

Because of its potential for multiple-destination applications, the cryptosystem is very attractive for the broadcast satellite networks. In TDMA applications, for example, information is transmitted in bursts. Conventionally, each TDMA burst contains a number of sub-bursts which are destined for different stations. For a cryptanalyst interested in messages to a particular destination, only the sub-bursts to the particular station must be recorded. The presence or absence of certain sub-bursts also provides information for traffic analysis. In the new system, in which the cryptogram x is transmitted by the entire burst as a single entity, the entire

burst must be recorded for cryptanalysis. Therefore, it is much more difficult for the opponent to perform cryptanalysis and traffic analysis. Figure 3 compares typical TDMA formats using the conventional approach and the new approach.



a. With conventional approaches, individual sub-bursts of each burst are clearly visible even if all traffic is encrypted

b. With multiple-destination cryptosystems, the entire burst appears as a single entity

Figure 3. *Typical TDMA Formats*

The Chinese Remainder Theorem, in the form most frequently presented, assumes that the numbers p_i^j are prime; however, it is also true if p_i^j are not prime provided that all p_i^j are relatively prime to each other. This leads to the application of the cryptosystem to large-scale digital communications networks. As shown in Figure 4, a typical digital communications network may include several levels of hierarchy in which small volumes of user traffic are multiplexed together by concentrators before entering larger trunk lines. To provide link encryption between concentrators, it is possible that all multiplexers, or concentrators may be built with the architecture shown in Figure 5, independent of the hierarchical level.

If all users have different primes, p_i, as their deciphering keys, the concentrator at a higher level can use a deciphering key which is equal to the product of all lower level deciphering keys. For example, if destinations A, B, and C in Figure 4 choose p_1, p_2, and p_3 as their deciphering keys, concentrator α can use them to demultiplex the information before forwarding it to them individually. Similarly, concentrator β can use the product $P = p_1p_2p_3$ to remove information which is irrelevant to concentrator α to conserve the bandwidth between α and β. Similar choices can also be applied at the transmit side. Thus, all links, except the link between a user and the directly connected concentrator, can be protected

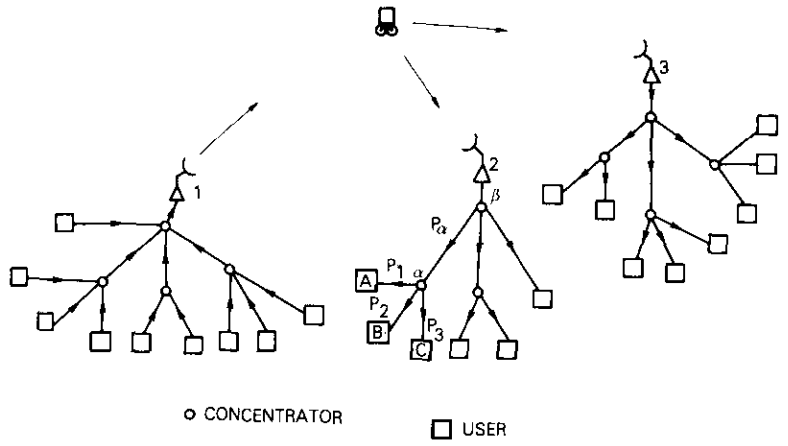


Figure 4. *A Typical Architecture of Digital Communications Networks with Levels of Hierarchy*

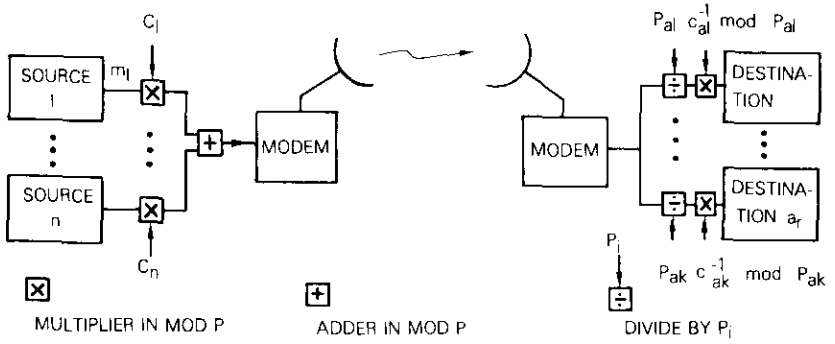


Figure 5. *The Implementation of the Proposed Multiple-Destination Cryptosystem for Broadcast Networks*

from intrusion without an additional bandwidth requirement. The required device is not much more complex than a regular concentrator.

It is not necessary that all the users are connected to the same level since it makes no difference whether the deciphering key belongs to a user or to a lower level concentrator. The deciphering keys can be sized according to relative traffic volume, provided that the deciphering key is large enough to ensure the required security. The advantage of this cryptosystem is that it performs the multiplexer/demultiplexer functions as well

as encipherer/decipherer functions. The integration of functions may permit substantial cost reduction in providing communications security to large-scale digital networks.

Acknowledgment

The authors would like to thank Dr. R. J. F. Fang, who suggested the possibility of constructing a multiple-destination cryptosystem and participated in many discussions during this research.

References

[1] A. Aho, J. Hopcroft, and J. Ullman, *The Design and Analysis of Computer Algorithms*, Reading, Massachusetts: Addison-Wesley Publishing Company, 1974, p. 294.

[2] D. Knuth, *Seminumerical Algorithms: The Art of Computer Programming*, Reading, Massachusetts: Addison-Wesley Publishing Company, Vol. 2, 1969.

[3] R. Rivest, A. Shamir, and L. Adleman, "A Method of Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, Vol. 21, No. 2, February 1978, pp. 120–126.

[4] W. Diffie and M. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, IT-22, No. 6, November 1976, pp. 644–654.

Lin-Nan Lee was born in Kaohsinng, Taiwan, on February 24, 1949. He received the B.S.E.E. degree (1970) from National Taiwan University, the M.S. (1973) and Ph.D. in electrical engineering (1976) from the University of Notre Dame. From 1970 to 1971 he was a communications officer in the Chinese Air Force. From 1971 to 1975, he was a research assistant at the University of Notre Dame. In 1975, he joined the Linkabit Corporation where he was involved in the design and development of packet communication protocols for the satellite networks. Subsequently, he joined COMSAT *in 1977 as a staff member of the Communications Systems Analysis Department. He is a member of IEEE and Sigma Xi.*

Shyue-Ching Lu was born in Taiwan, April 1, 1949. He received the B.S. in engineering science from Cheng Kung University, Taiwan, in 1970, and the Ph.D. in electrical engineering from the University of Hawaii, in 1976. He was a member of the technical staff of the Telecommunication Laboratories, Ministry of Communications, Republic of China. During the 1976–1977 academic year, he was also an adjunct associate professor of the Institute of Electronics, National Chiao Tung University, Taiwan. From August 1977 to July 1978, he was on leave with the Communications Systems Analysis Department, COMSAT *Laboratories. He is now a project manager, responsible for the development of a digital transmission system in the Telecommunication Laboratories. He is a member of IEEE.*

Index: secure communications, cryptosystem, redundancy, multiple substitution

Message redundancy reduction by multiple substitution: a preprocessing scheme for secure communications

S. C. Lu and L. N. Lee

(Manuscript received March 1, 1978)

Abstract

Shannon observed that the difficulty of cryptanalysis increases monotonically as the message redundancy decreases. As a result, a simple cryptosystem may be used to provide a high level of security against sophisticated cryptanalysis if message redundancy, as it appears to the cryptanalyst, is removed through certain kinds of preprocessing. Since redundancy can usually be removed by multiple substitution at the expense of bandwidth, a practical and optimal multiple substitution scheme is proposed to reduce the redundancy of messages with limited bandwidth expansion. It is believed that the security of cryptosystems can be significantly improved in messages preprocessed by such a scheme.

Introduction

The security of a cryptosystem can be improved substantially with certain kinds of message preprocessing just as the efficiency of a communications system can be improved by source encoding. Shannon [1] stated that the number of letters needed to successfully cryptanalyze a cryptogram (the unicity distance) is inversely proportional to the message re-

dundancy. Hence, the security of a cryptosystem can be strengthened with a good source coding scheme which reduces the message redundancy. Alternatively, if the only intent is to deceive the cryptanalysts, security can be strengthened by modifying the appearance of the messages.

Source coding schemes such as the Huffman codes and the run-length codes [2] have been studied extensively, resulting in a reduction in message redundancy as well as channel bandwidth requirements. If the channel bandwidth is not a concern, the "pseudoredundancy" of the message, as it appears to the cryptanalysts, can be reduced by simply expanding the alphabet size of the original messages. In such a scheme, the information remains unchanged; however, at least the same channel bandwidth may be required to transmit the expanded message. As a result, the actual redundancy of the message is generally increased. For clarity, the message redundancy, as it appears to the cryptanalysts, will be referred to as pseudoredundancy.

Figure 1 shows a cryptosystem using a preprocessing scheme. A message source emits a message stream $\underline{m}$, with an alphabet $A = (a_1, a_2, \ldots, a_n)$. A letter probability, p_i, is associated with each letter $a_i \epsilon A$. The set of letter probabilities $P = (p_1, p_2, \ldots, p_n)$ satisfies the usual constraints of a well-defined probability set, *i.e.*, $p_i > 0$ for $1 \leq i \leq n$, and $\sum_{i=i}^{n} p_1 = 1$.

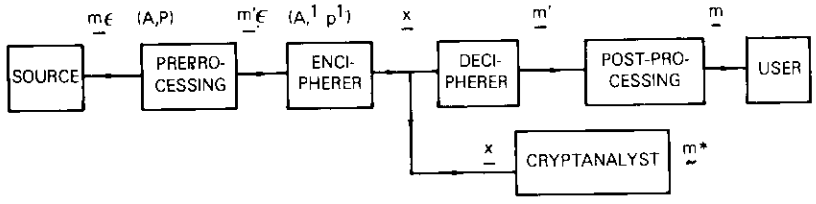


Figure 1. *A Cryptosystem Using A Message Preprocessing Scheme to Reduce Message Pseudoredundancy and to Improve Security*

The redundancy of the message stream $\underline{m}$, as defined by Shannon [1], is given by

$$r_{(A,P)} = \log n - h_{(A,P)} \tag{1}$$

where $h_{(A,P)}$ is the entropy of the source and is given by

$$h_{(A,P)} = -\sum_{i=1}^{n} p_i \log p_i \quad . \tag{2}$$

The message preprocessor accepts the message stream $\underline{m}$ and converts it to an expanded message stream $\underline{m}'$, with an alphabet $A' = (a_1', a_2', \ldots, a_t')$ and a set of letter probabilities $P' = (p_1', p_2', \ldots, p_t')$ such that the pseudoredundancy of the expanded message stream

$$r_{(A',P')} = \log t - h_{(A',P')} \tag{3}$$

is as small as possible.

For simple implementation at both the transmit and receive sides, a multiple substitution can be used in the preprocessing. This is achieved by partitioning the alphabet A' into n disjoint subsets, $A_i' \subset A'$, and by establishing a one-to-one correspondence between a letter in the alphabet A and a subset in the alphabet A' for all letters in the alphabet A. The preprocessor can then replace a letter in the message stream $\underline{m}$ by one of the letters in its corresponding subset in A' with a probability determined by P'. For example, if $A_i' = (a_j', a_{j+1}', \ldots, a_k')$, the letter a_i is replaced by the letter a_j with the probability $p_j'/(p_j' + p_{j+1}' + \ldots + p_k')$.

In general, the pseudoredundancy of a message stream can be reduced to zero if the letter probabilities of the message source are rational numbers. Since a common denominator d and integers e_i exist such that

$$p_i = \frac{e_i}{d} \quad , \qquad 1 \leq i \leq n$$

an alphabet A' of size d can be chosen and partitioned into n subsets with the ith subset containing e_i letters, and the letter a_i can be substituted for a letter in A_i' with probability $1/e_i$. As a result, all letters in A' have the same letter probability, $1/d$, and the pseudoredundancy is therefore minimized. This approach may be far from practical because considerable bandwidth expansion is usually involved; however, it is of academic interest because it is a feasible technique to achieve infinitely large unicity distance. In practice, it is necessary to choose an alphabet A' whose size t is only slightly larger than that of the original alphabet n, and, at the same time, to minimize the pseudoredundancy. In the next section, an algorithm will be described to find the optimal expansion for any size $t > n$ and for given source alphabet A and statistics P. It is believed that this approach has practical significance in applications in which adequate channel bandwidth is available, and a sophisticated source coding scheme is not warranted.

Optimal partial expansion

In the above described scheme, the pseudoredundancy of a message stream can be fully removed at the expense of very large alphabet size. In practical situations, such an alphabet expansion is probably intolerable. It is therefore important to choose an optimal expansion which minimizes the pseudoredundancy for a given size of the alphabet A'. Such an expansion will be designated optimal partial expansion to distinguish it from "full" expansion, which completely removes the pseudoredundancy.

Let $A = (a_1, a_2, \ldots, a_n)$ be the source alphabet, and $P = (p_1, p_2, \ldots, p_n)$ be the corresponding probabilities. If $n + 1$ is the desired size for the expanded alphabet A^1, a simple multiple substitution scheme is used to partition the alphabet into n disjoint subsets $A^1 = (A_1^1, A_2^1, \ldots, A_n^1)$ and to establish a one-to-one correspondence between a letter in A and a subset in A^1 for all letters in A. (The superscript "1" is used to indicate that the size of the new alphabet is one greater than the original source alphabet.) Naturally, among the n subsets in A^1, $n - 1$ subsets contain a single letter in A^1, and only one contains two letters. Without loss of generality, let $A_i^1 = (a_i^1)$ for $1 \leq i \leq n - 1$, and $A_n^1 = (a_n^1, a_{n+1}^1)$. Consequently,

$$p_i^1 = p_i \quad , \qquad 1 \leq i \leq n - 1 \tag{4a}$$

and

$$p_n^1 + p_{n+1}^1 = p_n \quad . \tag{4b}$$

To minimize the pseudoredundancy for A^1, the entropy

$$h_{(A^1,P^1)} = - \sum_{i=1}^{n+1} p_i^1 \log p_i^1 \tag{5}$$

must be maximized.

Substituting equation (4) into equation (5) and differentiating both sides with respect to p_n^1 demonstrates that the choice of

$$p_n^1 = p_{n+1}^1 = \frac{p_n}{2} \tag{6}$$

maximizes $h_{(A^1,P^1)}$ for a given p_n. The reduction of pseudoredundancy,

$$\Delta r = r_{(A^1,P^1)} - r_{(A,P)}$$

if the letters are not equally likely, is given by

$$\Delta r = \log \frac{n+1}{n} - p_n \quad . \tag{7}$$

From equation (7), it is obvious that the reduction is maximized if the most likely letter in A is replaced by two letters in A^1 with equal probability, as proven in the following lemma:

Lemma 1: If $A = (a_1, a_2, \ldots, a_n)$ is the source alphabet, $P = (p_1, p_2, \ldots, p_n)$ is the set of letter probabilities, and $p_n \geq p_i$ for $1 \leq i \leq n - 1$, the optimal partial expansion with an alphabet $A^1 = (a_1^1, a_2^1, \ldots, a_{n+1}^1)$ of size $n + 1$ is to replace a_i by a_i^1 for $1 \leq i \leq n - 1$, and a_n by a_n^1 and a_{n+1}^1 with equal probability.

With this technique, an alphabet of size n can be expanded with another alphabet of size $n + 1$, and maximum reduction of pseudoredundancy can be achieved. It seems possible to apply the technique recursively to expand the alphabet to an arbitrary size t. If the most likely letter a_i^1 in A^1 is the only letter in its subset A_i^1, the alphabet A^1 may be considered as another source alphabet, and the technique seems readily applicable. On the other hand, if the most likely letter a_i^1 in A is not the only letter in its subset A_i^1, all letters in the subset must be considered. In general, in an intermediate alphabet $A^k = (A_1^k, A_2^k, \ldots, A_n^k)$; any letter in subset A_i^k can be used to substitute the letter a_i in the alphabet A. (The superscript "k" indicates that the new alphabet is k greater than the original source alphabet.) If the partition of the disjoint subsets are fixed, and $P = (p_1^k, p_2^k, \ldots, p_{n+k}^k)$ are the letter probabilities,

$$\sum_{a_j^k \epsilon A_i^k} p_j^k = p_i \quad , \qquad 1 \leq i \leq n \quad . \tag{8}$$

With the Lagrange multipliers, it is easy to demonstrate the following lemma.

Lemma 2: For an expanded alphabet A^k of size $n + k$, $k > 0$, and a given partition $A^k = (A_1^k, A_2^k, \ldots, A_n^k)$, the pseudoredundancy is minimized if the letter probabilities of A^k satisfy

$$p_j^k = \frac{p_i}{l_i^k} \quad , \qquad \text{for all } a_j^k \epsilon A_i^k,\ 1 \leq i \leq n$$

where l_i^k is the number of letters in A_i^k.

To form a successful recursive algorithm, the optimal partition for the new alphabet must be determined whenever the size of the alphabet is incremented by one. As implied by *Lemma 1*, if the most likely letter, a_i^k in A^k, is the only letter in its subset, A_i^k, the optimal partition for A^{k+1} will contain two letters in the subset A_i^{k+1}, and the number of letters in any other subset $A_j^{k+1}, j \neq i, 1 \leq j \leq n$, is identical to that in A_j^k. That is, if a letter in A_i^k is the most likely letter in A^k, and $l_i^k = 1$, the optimal partition for A^{k+1}, is

$$l_i^{k+1} = 2$$

and

$$l_j^{k+1} = l_j^k \quad , \qquad 1 \leq j \leq n, j \neq i \quad .$$

In general, however, the subset which contains the most likely letter in A^k may have more than one letter. Intuitively, it seems possible that the same technique can be applied recursively. For example, if $A^k = (A_1^k, A_2^k, \ldots, A_n^k)$ is the partition for alphabet A^k, letters $a_j^k \epsilon A_i^k$ are the most likely letters in A^k, which may be expanded in the following manner:

a. The number of letters in subset A_i^{k+1} is one greater than that of A_j^k, i.e., $l_i^{k+1} = l_i^k + 1$.

b. The number of letters in other subsets remains unchanged, or $l_j^{k+1} = l_j^k, j \neq i$ and $1 \leq j \leq n$.

c. The letter probabilities for letters in each subset $A_j^{k+1}, i \leq j \leq n$, is given by p_j / l_j^{k+1}.

This procedure does not always lead to the optimal solution, as shown by the following counterexample.

Let A be a 3-letter alphabet, and let its letter probabilities be $P = (0.597, 0.300, 0.103)$, the optimal expansion for a 4-letter alphabet A^1 has three subsets, with the first subset containing two letters, and the others one letter. The letter probabilities for A^1 are given by $P^1 = (0.2985, 0.2985, 0.3000, 0.1030)$. According to this procedure, the expansion for a 5-letter alphabet A^2 has three subsets, with the first and second subsets containing two letters, and the other one letter. Its letter probabilities are $P^2 = (0.2985, 0.2985, 0.1500, 0.1500, 0.1030)$. The pseudoredundancy for the alphabet is 0.1218. However, the optimal expansion for the 5-letter alphabet should have a subset containing three letters, and two subsets containing one letter; the letter probabilities are (0.199, 0.199, 0.199, 0.300, 0.103), and the pseudoredundancy is 0.0726.

Lemma 2 shows that the letters in the same subset of an alphabet must be equally likely to achieve minimum redundancy. In a general case, the alphabet A^k is partitioned into n subsets $(A_1^k, A_2^k, \ldots, A_n^k)$ and the alphabet A^{k+1} expanded from A^k is also partitioned into n subsets. If the only difference between the two partitions is that subset A_i^{k+1} in the latter partition has one more letter than the corresponding subset A_i^k in the former case, the reduction in redundancy can be given as

$$\Delta r = \log \frac{n+1}{n} - p_i \log \frac{l_i^k + 1}{l_i^k} \tag{9}$$

if it is assumed that all letters in the same subset are equally likely for all subsets in both partitions. Hence, the following iterative optimization algorithm results:

Step 0: Set $k = 0$, $l_i^0 = 1$, $1 \leq i \leq n$, and $A_i^0 = (a_i)$.

Step 1: Calculate $p_i \log (l_i^k + 1)/l_i^k$ for the subsets A_i^k, $1 \leq i \leq n$. Find the subset A_j^k which maximizes $p_i \log (l_i^k + 1)/l_i^k$.

Step 2: Form the partition for A^{k+1} such that the number of letters in each subset is given by $l_j^{k+1} = l_j^k + 1$ and $l_i^{k+1} = l_i^k, i \neq j$, and $1 \leq i \leq n$. The letter probability for any letter $a_j^{k+1} \epsilon A_i^{k+1}$ is given by p_i / l_i^{k+1}, $1 \leq i \leq n$.

Step 3: Set $k = k + 1$ if $n + k$ is less than the maximum alphabet size desired, return to step 1; otherwise stop.

It should be noted that the pseudoredundancy for a given alphabet size is a function of entropy only, and the increase in entropy for each step is given by

$$p_i \log \frac{l_i^k + 1}{l_i^k} \quad .$$

This increase is a function of the particular subset to be expanded, and is independent of previous expansions in other subsets. In addition, it is a monotonic decreasing function of l_i^k as k increases, *i.e.*,

$$p_i \log \frac{l_i^{k+1} + 1}{l_i^{k+1}} < p_i \log \frac{l_i^k + 1}{l_i^k} \quad .$$

The independency among expansions in different disjoint subsets and the

monotonic decreasing property makes it impossible for any other expansion which has not been examined by the algorithm to contribute more entropy gain than the particular expansion selected by the algorithm at any given iteration. The optimality of the algorithm is therefore ensured.

The pseudoredundancy at each step is given by

$$r = \log (n + k) + \sum_{i=1}^{n} p_i \log \frac{p_i}{l_i^k}$$

or equivalently by

$$r = \log (n + k) - h_{(A,P)} - \sum_{i=1}^{n} p_i \log l_i^k \quad . \tag{10}$$

Taking the derivative of equation (10) with respect to l_j^k yields

$$\frac{\partial r}{\partial l_j^k} = \frac{\partial}{\partial l_j^k} [\log (n + k)] - \frac{p_i}{l_i^k} \quad . \tag{11}$$

Since the increase in $n + k$ is due to the expansion at the particular subset A_j^k,

$$\frac{\partial (n + k)}{\partial l_j^k} = 1 \quad . \tag{12}$$

Substituting equation (12) into equation (11) results in

$$\frac{\partial r}{\partial l_j^k} = \frac{1}{n + k} - \frac{p_i}{l_i^k} \quad . \tag{13}$$

This equation shows that the reduction of pseudoredundancy continues as long as the selected disjoint subset satisfies

$$\frac{p_i}{l_i^k} > \frac{1}{n + k} \tag{14}$$

or as long as all letters are not equally likely.

Example

The relative letter frequency in English [3] is shown in Table 1. The

TABLE 1. THE RELATIVE LETTER FREQUENCY IN ENGLISH

Letter	Frequency	Letter	Frequency	Letter	Frequency
e	0.13105	d	0.03788	w	0.01539
t	0.10468	l	0.03389	b	0.01440
a	0.08151	f	0.02924	v	0.00919
o	0.07995	c	0.02758	k	0.00420
n	0.07098	m	0.02536	x	0.00166
r	0.06832	u	0.02459	j	0.00132
i	0.06345	g	0.01994	q	0.00121
s	0.06101	y	0.01982	z	0.00077
h	0.05259	p	0.01982		

redundancy of written English is 0.56 bits per letter. According to the expansion procedure outlined in the previous section, the optimal expansion for an alphabet of 64 letters is to replace "*e*" by 7 letters in the new alphabet, "*t*" by 6 letters, "*a*" by 5 letters, "*o*" by 5 letters, "*n*" by 4 letters, "*r*" by 4 letters, "*i*" by 4 letters, "*s*" by 4 letters, "*h*" by 3 letters, "*d*" by 2 letters, "*l*" by 2 letters, "*f*" by 2 letters, "*c*" by 2 letters, and "*m*" by 2 letters, and to substitute any other letter in English by a single letter. The entropy for the new alphabet is 5.91 bits, and the pseudoredundancy is 0.09 bits per letter. Therefore, the pseudoredundancy of the new alphabet is only 1/6 of the redundancy of English. That is, communications security, measured in terms of unicity distance, is increased by a factor of 6 at the expense of about 20-percent bandwidth expansion. A similar expansion with an alphabet of 32 letters has a pseudoredundancy of 0.32 bits per letter, which is about 4/7 of the redundancy of English, with practically no bandwidth expansion.

Conclusions

After the partition of the new alphabet has been determined, the implementation for the preprocessor and the postprocessor is straightforward. The preprocessor may consist of a read-only-memory (ROM) table which

stores the corresponding subset of letters in the new alphabet for each letter in the original alphabet, and a uniform random number generator which selects a particular letter in the corresponding subset in the new alphabet. The postprocessor may consist of an ROM table which stores the corresponding letter in the original alphabet for each letter in the new alphabet. The preprocessing involves random number generation and a table look-up, and the postprocessing involves only a table look-up.

It appears that this scheme is particularly attractive for improving the security of teletype communications. For instance, computer algorithms using source statistics exist for successful cryptanalysis of any single substitution cipherer [4]. With slight bandwidth expansion and multiple substitution, a single substitution cipherer can provide very high security. Since the usual source coding schemes map the original letters into a block of digits of variable length in a smaller alphabet, letter synchronization must be maintained for successful operations. In addition, buffer storage must be provided at both the source encoder and source decoder. In situations such as teletype communications, a good source coding scheme, which reduces the message redundancy of English to 0.09 bits per letter, converts each letter to 4.23 bits, on the average, which is about a 15-percent bandwidth reduction. However, such a source coding scheme is considerably more complex than multiple substitution. If the difference between bandwidth requirements is not a concern, the additional complexity required for buffering and letter synchronization is not justified, and multiple substitution is certainly more attractive than source coding.

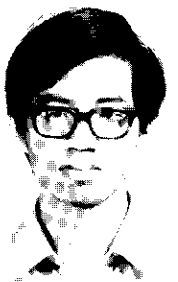
References

[1] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, Vol. 28, October 1949, pp. 656–715.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.

[3] F. Pratt, *Secret and Urgent*, New York: Blue Ribbon Books, 1942.

[4] L. Bahl, "An Algorithm for Solving Simple Substitution Cryptograms," IEEE Symposium on Information Theory, October 10–14, 1977, Ithaca, New York.



Shyue-Ching Lu was born in Taiwan, April 1, 1949. He received the B.S. in engineering science from Cheng Kung University, Taiwan, in 1970, and the Ph.D. in electrical engineering from the University of Hawaii, in 1976. He was a member of the technical staff of the Telecommunication Laboratories, Ministry of Communications, Republic of China. During the 1976–1977 academic year, he was also an adjunct associate professor of the Institute of Electronics, National Chiao Tung University, Taiwan. From August 1977 to July 1978, he was on leave with the Communications Systems Analysis Department, COMSAT Laboratories. He is now a project manager, responsible for the development of a digital transmission system in the Telecommunication Laboratories. He is a member of IEEE.

Lin-Nan Lee was born in Kaohsinng, Taiwan, on February 24, 1949. He received the B.S.E.E. degree (1970) from National Taiwan University, the M.S. (1973) and Ph.D. in electrical engineering (1976) from the University of Notre Dame. From 1970 to 1971 he was a communications officer in the Chinese Air Force. From 1971 to 1975, he was a research assistant at the University of Notre Dame. In 1975, he joined the Linkabit Corporation where he was involved in the design and development of packet communication protocols for the satellite networks. Subsequently, he joined COMSAT in 1977 as a staff member of the Communications Systems Analysis Department. He is a member of IEEE and Sigma Xi.

Index: cryptosystem, secure communications, noise, channel (data transmission), convolutional codes, sequential decoding

An integrated system for secure and reliable communications over noisy channels

S. C. Lu, L. N. Lee, and R. J. F. Fang

(Manuscript received January 11, 1979)

Abstract

A cryptosystem based on convolutional encoding and sequential decoding techniques is proposed for reliable and secure communications over additive noise channels. Before encoding the information with a convolutional code, the proposed system employs a linear expansion function known only to the sender and the recipient. The actual rate of information transmission is therefore lower when compared with that observed by the cryptanalysts who have no knowledge of the expansion function. The expansion ratio is selected such that the equivalent convolutionally encoded data are transmitted to the desired recipient at a rate slightly lower than the computational cutoff rate, R_{comp}, of the noisy channel and hence can be decoded reliably with a sequential decoding algorithm. However, the cryptanalysts cannot reliably decode these data since the transmission rate which they observe is above R_{comp}, and no feasible algorithm exists to decode a convolutional code with a sufficiently long constraint length at a rate above R_{comp}. The integration of encryption and channel coding might have some impact on the design of future communications systems and might reduce the overall systems complexity.

Introduction

Certain modern data or computer communications systems not only require that messages are correctly (or reliably) transmitted but also that their privacy (or security) is protected. Forward error correction (FEC) coding techniques have been employed, with or without automatic repeat request (ARQ), to ensure correct and reliable message transmission over noisy channels. Cryptographic techniques have been used to protect message privacy and security. Figure 1 is a typical system block diagram for reliable and secure communications. Usually, FEC channel codecs are employed to reduce the transmission errors caused by the channel noise such that minimal errors are present at the input of the decipherer. However, this separation of channel coding and data encryption may impose an unnecessary constraint on the design of a secure communications system for noisy channels, and consequently, may lead to a complex and costly system design.

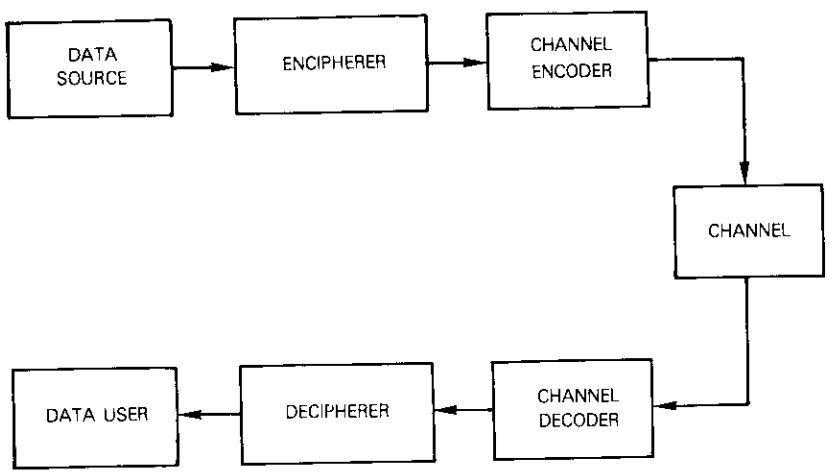


Figure 1. *Block Diagram of a Typical System For Reliable and Secure Communications*

Recently, a public-key cryptosystem [1] was proposed based upon the fact that no efficient algorithm is known to decode a general linear FEC code [2]. It employs a Goppa code, which can be decoded with an efficient algorithm, and transforms it into a general linear code with a secret, linear transformation. The encoding procedure for the general linear code is

announced for public use, and an artificially generated random noise is added to the encoded information for further protection. The intended receiver with the secret linear transformation can apply the inverse transformation to the received sequence such that it can be decoded with the decoding algorithm of Goppa codes. In a similar manner, a cryptosystem using the computational properties of sequential decoding may be constructed. In particular, a cryptosystem may be designed which will not only have some error-correction capability against channel errors, but will also ensure message privacy. Although such a system may not possess the properties required by a public-key cryptosystem, the integration of the cryptosystem with the channel coding system might reduce the overall system complexity compared to separate implementation of the cryptosystem and the channel coding system.

It is well known in coding theory that there exists a threshold rate known as the computational cutoff rate, R_{comp}, for sequential decoding [3], [4]. Information transmitted with a rate below this threshold can be decoded reliably by sequential decoding with reasonable complexity, whereas the required number of computations to decode the information grows as a Paretial function of r/R_{comp} if the rate r is above R_{comp}. In general, the computational cutoff rate is a function of the channel error rate. Convolutional encoding with sequential decoding is the most efficient method to achieve reliable communications over noisy channels. The computational cutoff rate is often considered the practical limit on the rate for reliable information transmission.

In a cryptosystem, the recovery of information from processed data must be relatively straightforward for the user with key information whereas extraction of information must be computationally impossible for a cryptanalyst without key information. This paper will demonstrate that the threshold effect of the computational cutoff rate can be used to achieve the seemingly conflicting requirements of a good cryptosystem. Furthermore, because the cryptosystem is based on an FEC coding system, it possesses the inherent error-correcting capability in addition to that of communications security.

Description of the cryptosystem

In the block diagram of the proposed cryptosystem (Figure 2), the encipherer consists of the message expansion function, F; the convolutional encoder, G, of code rate r; and an artificial noise generator which emits the random sequence, $\underline{n}$. The expansion function, F, is a linear func-

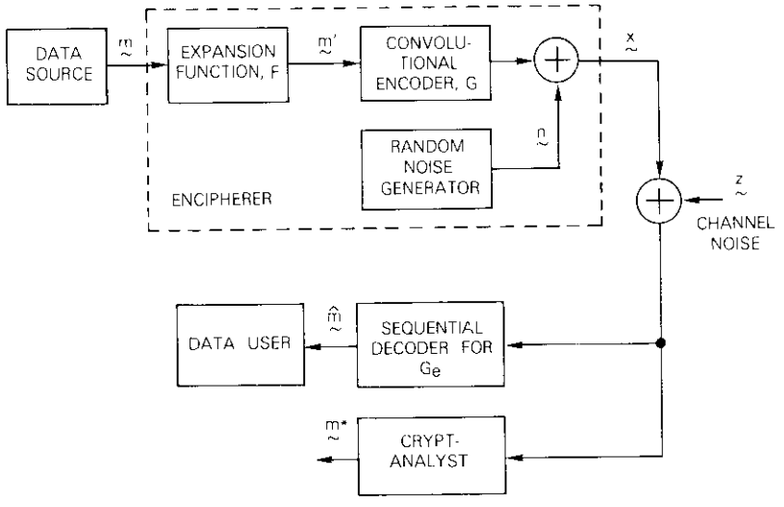


Figure 2. *Block Diagram of the Proposed Cryptosystem*

tion which must be kept secret from all except the sender and the recipient. In general, a linear expansion function transforms k bits of information linearly into n bits of data, $n > k$. These linear transformations may be dependent upon previous information. Thus, message sequence $\underset{\sim}{m}$ is first transformed into sequence $\underset{\sim}{m}'$ by the expansion function, F, and then encoded by the convolutional encoder, G. Cryptogram $\underset{\sim}{x}$ is the sum of the encoded message and the artificial noise, $\underset{\sim}{n}$, which can be either a pseudorandom noise or a true random noise. In terms of operator notations, the cryptogram to be transmitted can be expressed as the sequence

$$\underset{\sim}{x} = G[F(\underset{\sim}{m})] + \underset{\sim}{n} \quad . \tag{1}$$

After the cryptogram is transmitted over a noisy channel characterized by the additive noise sequence $\underset{\sim}{z}$, a corrupted version of the cryptogram is received, *i.e.*,

$$\underset{\sim}{y} = \underset{\sim}{x} + \underset{\sim}{z} \quad . \tag{2}$$

The substitution of equation (1) into equation (2) yields

$$\underset{\sim}{y} = G[F(\underset{\sim}{m})] + \underset{\sim}{e} \tag{3}$$

where $\underset{\sim}{e} = \underset{\sim}{n} + \underset{\sim}{z}$ is the sum of the artificial noise and the channel noise. A reinterpretation of equation (3) reveals that the expansion function and the convolutional encoder can be considered an equivalent channel encoder, $G_e = GF$; the sum of the artificial noise and the channel noise can be represented by the equivalent channel noise sequence $\underset{\sim}{e}$. Since the expansion function, F, is linear, the equivalent encoder G_e also constitutes a convolutional encoder, and the code rate r_e of the equivalent convolutional code is lower than r. When the equivalent convolutional code G_e has reasonable distance properties and is noncatastrophic [5] in the sense that a limited number of channel errors can cause only a limited number of decoding errors, the message sequence can be reliably recovered from the received sequence by sequential decoding provided that the cutoff rate of the equivalent channel is greater than the code rate. Thus, deciphering the cryptogram by sequential decoding is straightforward.

Strength of the cryptosystem

In the proposed cryptosystem, the parameters specifying the expansion function F must be kept secret and distributed from one end to the other. These parameters can therefore be considered as the key of the cryptosystem. In addition, if a pseudorandom number generator is used to produce the artificial noise, either the initial content (known as the "seed") or the algorithm must be kept secret. Thus, the pseudorandom number generator, or its seed, may also be part of the key. (The receiver, however, does not require the knowledge of the pseudorandom number generator or its seed to decipher the messages. Hence, only the parameters specifying F must be distributed in advance.)

Generally, the security strength of a cryptosystem is based on the size of the key space. Since ample expansion functions can be used by the proposed system, it is not difficult to design a particular configuration with sufficiently large key space. The computational cutoff rate of the equivalent channel can be controlled by the artificial noise generator, which is controlled by the message sender. Thus, it is impractical for the cryptanalyst to estimate the code rate, r_e, of the equivalent code, and thereby the rate of the expansion function F.

The fact that the equivalent convolutional code must possess reasonable distance properties may seem to impose some limitation on the key space.

However, this limitation should not cause a problem because the random coding theorem ensures that most of the FEC codes are good. The requirement that the equivalent code must be noncatastrophic can easily be satisfied by choosing a noncatastrophic convolutional code for G and an expansion function whose inverse can be implemented by a feed forward linear sequential circuit (FFLSC) [5]. An FFLSC is defined as a network with a finite number of inputs and outputs consisting of delays and adders without any feedback connections (see Figure 3). Since a convolutional

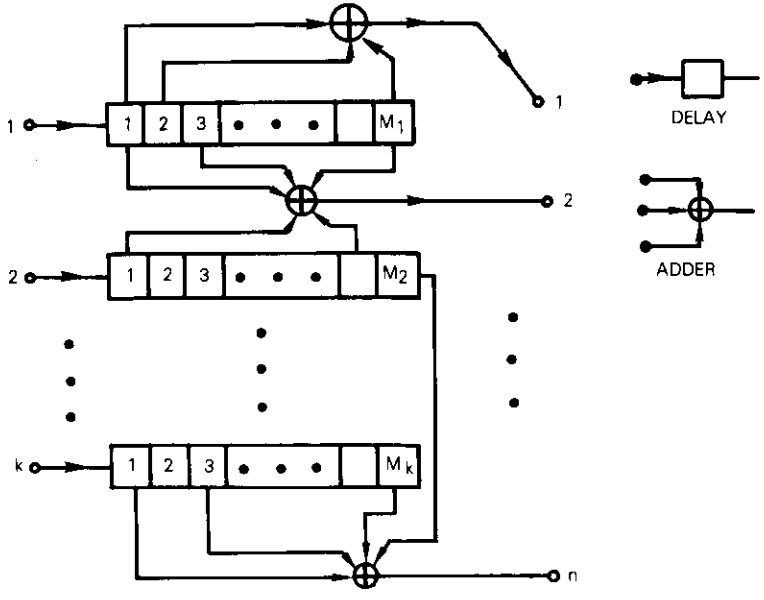


Figure 3. *k-Input, n-Output Feed Forward Linear Sequential Circuit (FFLSC)*

code has an FFLSC inverse if and only if it is noncatastrophic [5], the inverse of the equivalent code

$$G_e^{-1} = [G(F)]^{-1} = F^{-1}\, G^{-1}$$

can also be implemented by an FFLSC. It is therefore noncatastrophic. Also since there are numerous expansion functions with this property, the noncatastrophic requirement does not seem to limit the key space significantly.

From an operational point of view, it is also straightforward for the user to select an expansion function and test the code properties of the equivalent code. It can be assumed that the user has an encipherer and a decipherer connected in the loopback mode. Thus, he can adjust the artificial noise generator in the encipherer to include the effect of the actual channel noise, and select an expansion function with a rate below the R_{comp} of the equivalent channel. A random bit stream is first enciphered with the encipherer and then deciphered with the decipherer. By comparing the input and output sequences, the user can quickly determine whether or not the equivalent code is noncatastrophic and has reasonable distance properties. From the random coding theorem, it may be deduced that a satisfactory key can be selected with few trials.

The success of known plaintext attack by the cryptanalyst to directly determine the equivalent code seems unlikely because of the artificial noise $\underline{n}$ introduced at the encipherer. With knowledge of the encoder, G, a seemingly easier attack is to estimate the expanded sequence by decoding the convolutional code and then to determine the expansion function, F. However, such an attack can be voided by controlling the artificial noise generator such that the R_{comp} of the equivalent channel is less than the rate r of the convolutional code G, but is still greater than r_e of the equivalent code G_e, *i.e.*,

$$r_e < R_{\text{comp}} < r \quad .$$

With these selections of R_{comp} and r, it is infeasible to decode the long convolutional code G due to the lack of an algorithm to function at transmission rates higher than R_{comp}. For example, the number of computations of a sequential decoder grows as a Paretial function of r/R_{comp} when the code rate is above R_{comp} [3], [6]. A threshold decoder [7] may not experience a computational problem, but its estimates of the transmitted sequence are too erroneous to be useful at rates above R_{comp}. The maximum-likelihood Viterbi [8] decoder may provide a better estimate than the threshold decoders, and its computational complexity is independent of the channel quality. However, this complexity grows exponentially with the constraint length of the convolutional code; hence, it is practically infeasible to employ the Viterbi decoder to decode convolutional codes of sufficiently long constraint length.

The complexity of the Viterbi decoder may be reduced by using the "reduced states" technique [9] at the expense of performance. However, at rates above R_{comp}, the operation of the reduced states decoder is ex-

tremely difficult because the states in the Viterbi decoder must be reset frequently to prevent error propagation. The usefulness of the estimates obtained with this decoder is questionable for cryptanalysis because of excessive errors.

The proposed cryptosystem is not secure against chosen plaintext attack. The difference between two frames of cryptograms resulting from two frames of identical information is the difference between the noise sequences used in each frame. If the cryptanalyst inputs the same frame of information for a large number of frames, the noise effect can be eliminated by using majority decision on each bit location of the cryptogram. The expansion function F can then be determined.

This cryptosystem is best suited for constant traffic operations. Constant traffic not only contributes additional security to the messages but also increases the difficulty of off-line cryptanalysis. However, care must be taken in generating constant traffic while no real traffic is present; the only precaution is to avoid repetition of identical message frames.

Applications to satellite communications

The proposed cryptosystem can achieve secrecy only at the expense of bandwidth expansion, and consequently, throughput reduction, which must be sufficiently small for the system to be of any practical value. For simplicity, the following discussion considers only the hard-decision case in which the receiver makes a binary decision at each time instant. However, the same discussion can easily be extended to the soft-decision case in which the receiver outputs not only its decision but also reliability information about its estimate.

The equivalent channel can be regarded as a binary symmetric channel (BSC) with crossover probability ϵ, as shown in Figure 4a. It can also be considered as the concatenation of two BSCs (Figure 4b). One has a crossover probability p representing the artificial noise, $\underline{n}$, introduced by the encipherer, and the other has a crossover probability q representing the actual channel noise, $\underline{z}$. It can be shown that $\epsilon = p + q - 2pq$. The cutoff rate of the equivalent BSC, R_{comp}, is given by $1 - \log_2[1 + 2\sqrt{\epsilon(1-\epsilon)}]$[3]. Table 1 lists the values of the cutoff rates of the equivalent BSC, which correspond to crossover probabilities ϵ. It seems that a threefold-to-fourfold bandwidth expansion may be sufficient for moderate security, and a sixfold to tenfold expansion may achieve a high level of security. The cost increases rapidly if a higher crossover probability is desired.

If the integrated cryptosystem and communications system is designed

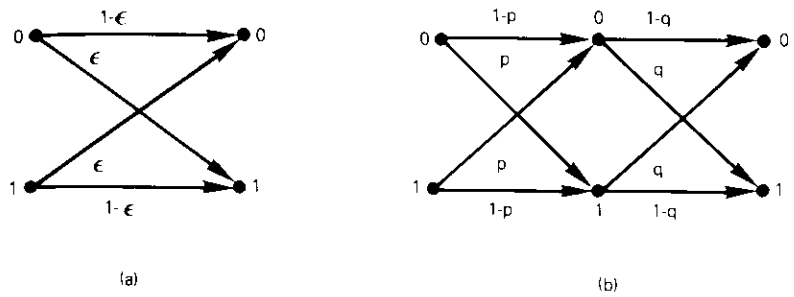


Figure 4(a) *Binary Symmetric Channel with Crossover Probability* ϵ (b) *Concatenation of Two Binary Symmetric Channels with Crossover Probabilities p and q, Respectively*

TABLE 1. BSC CUTOFF RATES CORRESPONDING TO SEVERAL VALUES OF CROSSOVER PROBABILITY ϵ

ϵ	R_{comp}
0.08	0.375
0.10	0.322
0.15	0.225
0.20	0.152
0.25	0.100

to operate at a channel crossover probability ϵ, it is not desirable to design a communications system that will yield a bit error rate much less than ϵ. Since $\epsilon = p + q - 2pq \approx p$ if $q \ll \epsilon$, reducing q by increasing the transmitted power of the communications will not improve the quality of services delivered to the user after decryption. Therefore, considerable cost savings may be realized by properly balancing p and q, without using excessive transmitted power to reduce q substantially, as long as p remains dominant to control R_{comp}. For example, if a 10-percent crossover probability ϵ is desired, the bit error rate of the channel q may be kept at approximately 3 percent provided that the encipherer noise is adjusted to yield a crossover probability p of approximately 7 percent (since $\epsilon \cong p + q$ for small p and q). Therefore, this cryptosystem appears to be particularly attractive for wideband, power-limited applications such as deep space and satellite channels.

Degradation due to precipitation may be offset by the dynamic adjustment of the encipherer's artificial noise generator, $\underline{n}$. However, for satellite

communications applications in which down-link noise dominates, care must be exercised to ensure that a safety margin is maintained. For instance, the cryptanalyst may be at a location with a clear-sky condition, or may have a receiver with improved performance; therefore, the rate r must be high enough such that the R_{comp} under the clear-sky condition is still lower than r. Otherwise, the artificial noise generator at the encipherer may no longer be adequate if the up-link noise level is not high enough to inhibit sequential decoding. Therefore, this proposed scheme may have potential applications in secure satellite communications involving up-link noise that is reasonably significant.

Generally, the expansion function, F, can be any linear function, time-varying or time-invariant, with or without memory. Also, it can be a block or convolutional encoder. The strength of the proposed cryptosystem is based on the size of the key space, and there are ample time-invariant functions. Thus, system security is not reduced even if the key space is limited to time-invariant functions, which are much simpler than time-varying functions in terms of key distribution and cryptosynchronization. Furthermore, sequential decoders for convolutional codes at data rates up to 1 Mbit/s are readily available, and it seems relatively easy and straightforward to incorporate the expansion function into the decoders. Thus, the cryptosystem may find applications in high-speed data channels.

This paper has emphasized that reliable transmission of information at a code rate above R_{comp} is not feasible. To provide additional margin, the system can be designed such that $r_e < R_{comp} < C$ and $r > C$, where C is the channel capacity. Shannon [10] has shown that reliable transmission of information at rates above channel capacity is impossible. Thus, the security of the cryptosystem can be further enhanced.

References

[1] W. Diffie and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, IT-22, No. 6, November 1976.

[2] R. J. McEliece, "A Public Key Cryptosystem Based on Algebraic Coding Theory," JPL DSN Progress Report, 1978.

[3] J. M. Wozencraft and I. M. Jacobs, *Principle of Communications Engineering*, New York: John Wiley and Sons, Inc., 1965.

[4] R. M. Fano, "A Heuristic Discussion of Probabilistic Decoding," *IEEE Transactions on Information Theory*, IT-9, No. 1, April 1963, pp. 64–74.

[5] J. L. Massey and M. K. Sain, "Inverses of Linear Sequential Machines," *IEEE Transactions on Computers*, C-17, April 1968, pp. 330–337.

[6] I. M. Jacobs and E. R. Berlekamp, "A Lower Bound to the Distribution of Computation for Sequential Decoding," *IEEE Transactions on Information Theory*, IT-13, No. 2, April 1967.

[7] J. L. Massey, *Threshold Decoding*, Cambridge, Mass.: M.I.T. Press, 1963.

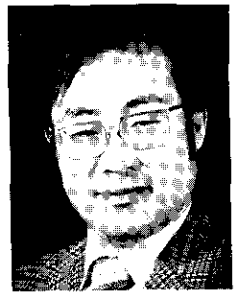
[8] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, IT-13, No. 2, April 1967, pp. 260–269.

[9] C. W. Zeoli, "Coupled Decoding of Block-Convolutional Concatenated Codes," *IEEE Transactions on Communications*, COM-21, March 1973, pp. 219–226.

[10] C. E. Shannon, "The Mathematical Theory of Communication," *Bell System Technical Journal*, Pt. 1, Vol. 27, July 1948, pp. 379–423; Pt. 2, Vol. 27, October 1948, pp. 623–656.

Shyue-Ching Lu was born in Taiwan, April 1, 1949. He received the B.S. in engineering science from Cheng Kung University, Taiwan, in 1970, and the Ph.D. in electrical engineering from the University of Hawaii, in 1976. He was a member of the technical staff of the Telecommunication Laboratories, Ministry of Communications, Republic of China. During the 1976–1977 academic year, he was also an adjunct associate professor of the Institute of Electronics, National Chiao Tung University, Taiwan. From August 1977 to July 1978, he was on leave with the Communications Systems Analysis Department, COMSAT Laboratories. He is now a project manager, responsible for the development of a digital transmission system in the Telecommunication Laboratories. He is a member of IEEE.

Lin-Nan Lee was born in Kaohsinng, Taiwan, on February 24, 1949. He received the B.S.E.E. degree (1970) from National Taiwan University, the M.S. (1973) and Ph.D. in electrical engineering (1976) from the University of Notre Dame. From 1970 to 1971 he was a communications officer in the Chinese Air Force. From 1971 to 1975, he was a research assistant at the University of Notre Dame. In 1975, he joined the Linkabit Corporation where he was involved in the design and development of packet communication protocols for the satellite networks. Subsequently, he joined COMSAT *in 1977 as a staff member of the Communications Systems Analysis Department. He is a member of IEEE and Sigma Xi.*



Russell J. F. Fang is Manager of the Communications Systems Analysis Department of the Transmission Systems Laboratory at COMSAT *Laboratories. He is responsible for research and studies of advanced communications systems concepts and techniques for international and domestic satellite communications. Born in Chungking, China, he received a B.S. in electrical engineering from National Taiwan University in 1962, and an M.S. and a Ph.D. in electrical engineering from Stanford University in 1964 and 1968. Before joining* COMSAT *in 1968, he was employed by Stanford Electronics Laboratories (1965–1968), Stanford Research Institute (1964–1965), and the Chinese Air Force Electronics and Ordnance Division, Taiwan (1962–1963).*

Index: INTELSAT V, transmission, standard C, earth terminal, construction, rain attenuation

*Transmission planning for the first U.S. Standard C (14/11-GHz) INTELSAT earth station**

L. F. GRAY AND M. P. BROWN, JR.

(Manuscript received May 26, 1978)

Abstract

The first INTELSAT V satellite will operate in both the 6/4- and 14/11-GHz frequency bands. Operation at 14/11 GHz will occur for the first time in the INTELSAT system and will require earth stations to function under degraded propagation conditions while maintaining C.C.I.R. quality service. This paper describes initial U.S. transmission planning for the construction of a pair of antennas operating in a space diversity mode which will meet the performance objectives given in BG-28-73, "Standard C Performance Characteristics of Earth Stations in the INTELSAT V System (14/11-GHz frequency bands)," August 18, 1972. Relevant factors affecting the design, such as transmission margins, channel performance, diversity improvement, and the use of propagation models to convert rain statistics to attenuation statistics, are discussed. Actual attenuation data for October 1977 to October 1978 are given elsewhere in this issue. The measurements show that the values selected for the model during the

* This is an updated version of a paper presented at the IEEE Electronics and Aerospace Systems Convention (EASCON '77), Arlington, Virginia, September 26, 1977. The information was derived from work sponsored by INTELSAT and the U.S. Earth Stations Owners Consortium (ESOC). The views expressed in this paper are not necessarily those of INTELSAT or ESOC.

original planning stage were optimistic compared to the measured values taken in 1977–78, which are substantially larger than predicted. This emphasizes the need for diversity to meet internationally agreed upon performance criteria established by INTELSAT and the C.C.I.R. COMSAT, the U.S. Signatory to INTELSAT, plans to locate the first Standard C facility in West Virginia. Authority to construct the facility described herein was pending before the FCC when this paper was submitted to the *COMSAT Technical Review*.

Introduction

Present operational plans require the first INTELSAT V to be deployed as the Primary Path satellite located at 335.5° east longitude in the Atlantic Ocean Region. Later, the Major Path 1 satellite will be located at 325.5° and Major Path 2 at 340.5°, followed by a spare located at 330.5°. The antenna elevation angles from the planned West Virginia locations could therefore vary from 14° to 25°.

The first five INTELSAT Standard C earth stations, to be located in France, Germany, Italy, the United Kingdom, and the U.S., will communicate among themselves in the 14- and 11-GHz bands and, through the same satellite, with Standard A stations (6/4-GHz) by "cross-strapping" transponders. For example, selected 6-GHz earth station transmitters can be connected through the satellite to 11-GHz down-path receivers, and certain 14-GHz transmitters can be connected to 4-GHz down-path receivers. The approximate size and location of the INTELSAT V 14/11- and 6/4-GHz beam coverages used to accomplish these connections are illustrated in Figure 1.

With cross-strapped and direct beams, the Primary Path satellite should achieve a saturation capacity of approximately 24,500 telephone channels as shown in Table 1 [1]. In addition to this capacity, there will be direct

TABLE 1. INTELSAT V CHANNEL CAPACITY ESTIMATES
(Primary Atlantic Satellite)

Frequency Band	Channels	Beams
Direct 6 → 4 GHz	11,000	Hemi, zone-to-hemi, or zone
Cross-strapped (6 → 11 and 14 → 4)	7,500	Hemi or zone-to-spot and vice versa
Direct 14 → 11 GHz	6,000	Spot-to-spot
Total	24,500*	

* Half-circuits (12,250 circuits).

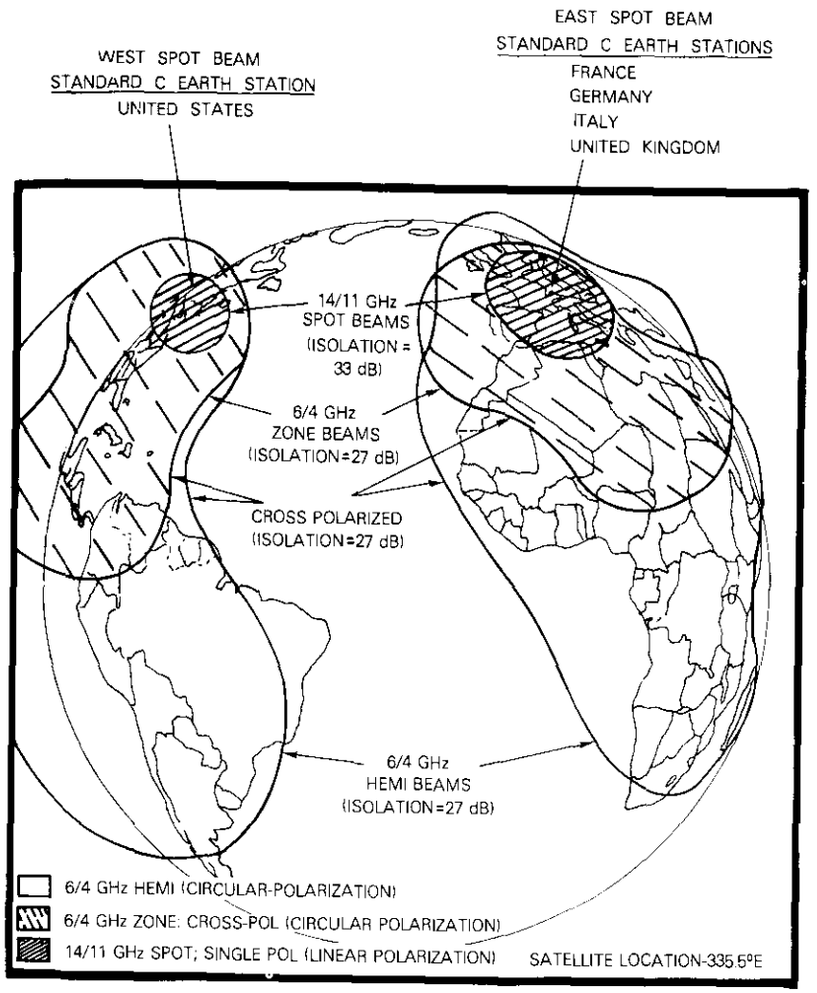


Figure 1. *INTELSAT V Atlantic Satellite Transmit Capabilities for Standard C Earth Stations, Primary Satellite (14/11-GHz spot beams are steerable and may be moved to meet traffic requirements)*

6/4-GHz global beams capable of handling television, and up to 800 SPADE channels. Some global capacity will also be used for FDM/FM carriers.

Standard C station requirements

INTELSAT has established the criterion that the value of the carrier-to-noise (C/N) ratio at demodulator inputs should exceed 10 dB for at least 99.98 percent of a year, except for equipment and scheduled sun interference outages. This percentage value includes propagation effects at both ends of a link. The antenna figure-of-merit that has been chosen for Standard C stations is 39.0 dB/K plus attenuation and noise temperature increase values that are expected for specified percentages of time. To account for this condition, INTELSAT has specified figures-of-merit according to two categories: those that must be exceeded for all but 10 percent of the time (long-term), and those that must be exceeded for all but 0.017 percent of the time (short-term).

For the west spot beam, the long-term requirement specified in BG-28-73 is "$G/T_1 - L_1 \geq 39.0 + \log_{10}(f/11.2)$ dB/K for all but 10 percent of the time for which statistics are available." The short-term requirement, representing degraded propagation conditions is "$G/T_2 - L_2 \geq 29.5 + \log_{10}(f/11.2)$ dB/K for all but 0.017 percent of the time for which statistics are available." In these requirements, G is the antenna gain; T_1 is the system noise temperature for all but 10 percent of the time and L_1 is the additional attenuation over a clear sky for all but 10 percent of the time; T_2 and L_2 are corresponding values for 0.017 percent of the time. These conditions represent performance in the direction of the satellite with the understanding that there is proper linear polarization alignment between the earth station and satellite antennas.

The earth station owner must determine the use of available rain statistics* to obtain the required figure-of-merit values. Then he must determine the optimum antenna size, receiver noise temperature, and transmitter power, based on factors such as cost, reliability, and equipment availability. He must also decide whether the outage criteria can be met with a single station. In general, single stations appear practical in Europe; however, a space diversity pair is indicated at U.S. locations where rainfall is quite heavy along the Eastern seaboard. When two stations are used, the noise contribution of the interconnecting link must be included in the overall performance analysis.

* In the case of the U.S., the statistical median of data obtained at the closest available West Virginia locations over a 12-year period was employed. The procedure used for applying these data to the actual locations and for obtaining values of attenuation is discussed later.

The above requirements apply to the down-link. INTELSAT has similarly specified the up-link power requirements in terms of the required values of e.i.r.p. for various carrier sizes for a lossless sky condition. To maintain the same power flux density at the spacecraft for all but 10 percent of the time, the earth station transmitter power must be increased by a factor corresponding to the 14-GHz propagation loss. For this purpose, INTELSAT has specified the short-term up-link performance in terms of permissible power flux density reductions at the spacecraft (from the lossless sky values) for specified percentages of time. Table 2 gives these margins for the west spot beam.

TABLE 2. INTELSAT V WEST SPOT BEAM UP-PATH MARGINS

Link (Frequency Band)	Margin (dB)	Percentage of Year
14 → 4 GHz (80 MHz)*	6.0	0.017 (89.4 min)
14 → 11 GHz (80 MHz)	6.5	0.010 (52.5 min)
14 → 11 GHz (240 MHz)	12.5	0.010 (52.5 min)

* Transponder bandwidth.

It is not permissible to simply provide sufficient excess power to satisfy short-term margins because the transponders can be shared by more than one station and transponder overloading must be avoided. Therefore, the power flux density at the spacecraft has been specified not to exceed the lossless sky value by more than 1 dB. Automatic power control must be used to meet this requirement if the expected attenuation increase exceeds the margins under the specified conditions. An alternate procedure is to switch to a diversity station to avoid excessive attenuation conditions.

Overview of U.S. Standard C station

Engineering studies have been conducted in planning for the procurement of an operational facility for the first INTELSAT V satellite [2]. The main Standard C antenna is to be collocated with the existing Standard A

station at Etam, West Virginia, while a diversity antenna is to be located about 35 km to the northeast at Lenox, West Virginia. A microwave heterodyne repeater located on Laurel Mountain (near Etam) is to provide a double hop diversity interconnection link (DIL) between the two stations, and traffic is to be routed to the station with favored propagation conditions. The interface with the terrestrial link will be located at Etam with the diversity switching performed on the COMSAT side. It is intended that the diversity station at Lenox will be an unmanned site fully controlled by a monitoring and control subsystem. The following design parameters have been used to formulate the configuration of the station.

Propagation statistics

To meet transmission margin and circuit availability requirements, the earth station owner must deduce attenuation statistics from measured data or rainfall information since long-term attenuation data at 11 and 14 GHz is unlikely to be available. Contrary to the conditions in the 6/4-GHz bands, where rain attenuation is small, climatic conditions will govern 14/11-GHz station design. Earth station G/T, transmitter power, and space diversity requirements are all determined by the propagation conditions of the geographic area selected. Table 3 reveals that weather

TABLE 3. COMPARISON OF U.S. AND EUROPEAN RAIN ATTENUATION AT 14/11 GHz[a] (0.01% time in a year)

Station (Country)	Elevation Angle (deg)	11.20-GHz Attenuation (dB)	14.25-GHz Attenuation (dB)	ITU Zone[b]
Barcenay-en-Othe (France)	23	4.1	6.8	4
Usingen (Germany)	19	6.2	10.2	3
Lario (Italy)	21	5.5	9.1	3–4
Madley (U.K.)	23	4.6	7.6	3
Etam (U.S.)	18	12.5	24.0	1

[a] Excess over clear sky attenuation toward a satellite at 335.5°E [3].

[b] ITU Radio Regulation, Appendix 28, Figure 17 shows rain-climatic zones of the world. Zones are categorized from 1 to 5, where 1 has the heaviest rainfall and 5 the lightest.

conditions in the eastern U.S. are worse than those in Europe in terms of rainfall attenuation. Rain-rate data over a 12-year period, extracted from

U.S. Weather Service records for two locations (Rowlesburg and Elkins, West Virginia), were analyzed to produce the results shown in Table 4 [3].

TABLE 4. 12-YEAR RAIN STATISTICS FOR ETAM AND LENOX AREA[a]

Rain Collection Location (West Virginia)	Median Yearly Rainfall, M (mm)[b]	Maximum Excessive Short-Duration Rainfall, Median Year (mm/hr)			Beta Factor, β (deduced)[b]
		5 min	10 min	15 min	
Elkins	916	104	79	60	0.06
Rowlesburg	1165	—	—	—	0.20[c]

[a] Values shown in this table represent the statistical median of all data examined over the 12-year period [3].

[b] Two values are required for the COMSAT Labs Propagation Model: "M" representing total annual rainfall and "β" the thunderstorm ratio. β is estimated by two techniques: (1) a U.S. map of thunderstorm ratios and (2) maximum excessive short-duration rainfall. Elkins' β was derived using (2) and Rowlesburg with (1) and Reference 3.

[c] See Reference 4.

Rowlesburg is approximately 10 km from Etam while Elkins is approximately 50 km. The Rowlesburg statistics were used as the most representative for the thunderstorm ratio of the area (*i.e.*, $\beta = 0.2$) [4] and the total average annual rainfall was taken as 1,165 mm (usually identified by the letter "M"). With these values of M and β, the deduced values of attenuation shown in Figures 2 and 3 were generated with propagation models prepared by COMSAT Laboratories based on the Rice-Holmberg method [5]. Observed values of attenuation, using radiometers at 11.6 GHz, during 1977–78 at Etam and Lenox, West Virginia, are reported elsewhere in this issue [6]. These data show that a larger value for the beta factor would have been appropriate and thus underscore the need for actual data collection.

The attenuation estimates given in Figures 2 and 3 were generated from rain-rate data and, therefore, are given with reference to clear sky. With the conditions set forth in the Standard C specification, there is a total of three cases of sky conditions needed to describe the Standard C station as defined below:

a. CLEAR SKY. Under this condition, the sun is not obscured by overcast, but average conditions of water vapor are present. Assumed average conditions for the earth's surface are a temperature of

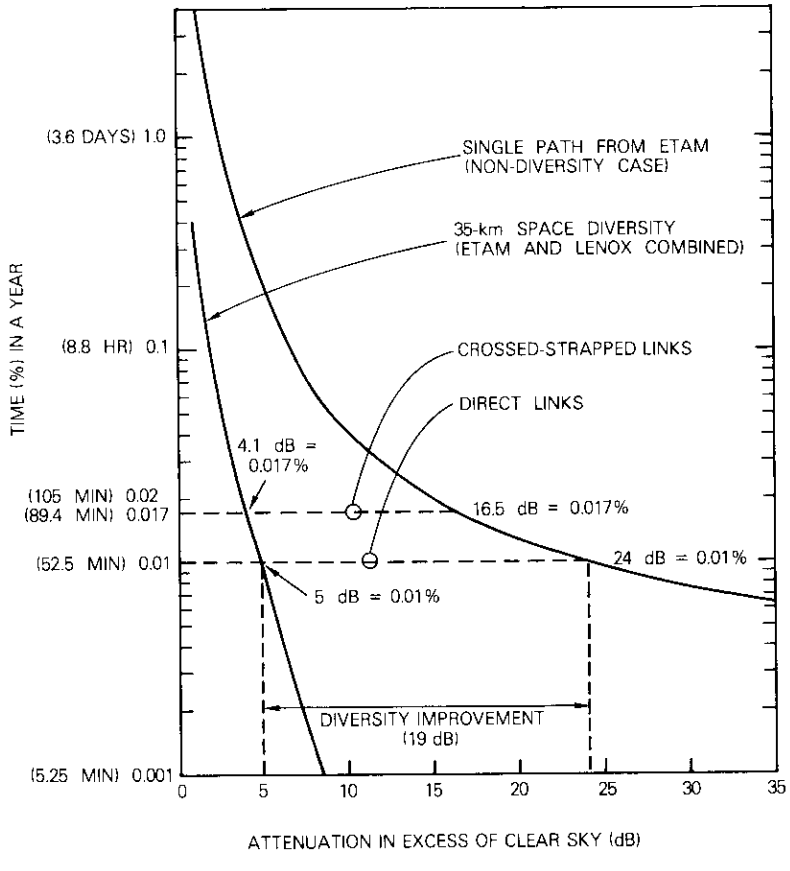


Figure 2. *Estimated Rain Attenuation at Etam-Lenox, W. Va., 14.25 GHz, 18° Elevation (Based on COMSAT Labs propagation models; attenuation represents levels derived from the median value of 12-year statistics)*

20°C, 760-mm mercury pressure, and 10 g/m³ of water vapor. For the U.S. station: 0.40 dB at 11.20 GHz and 0.65 dB at 14.25 GHz.

b. 10-PERCENT SKY (CLOUDY SKY). Clear sky plus attenuation from rain expected over the long-term (*i.e.*, 90 percent of the year). For

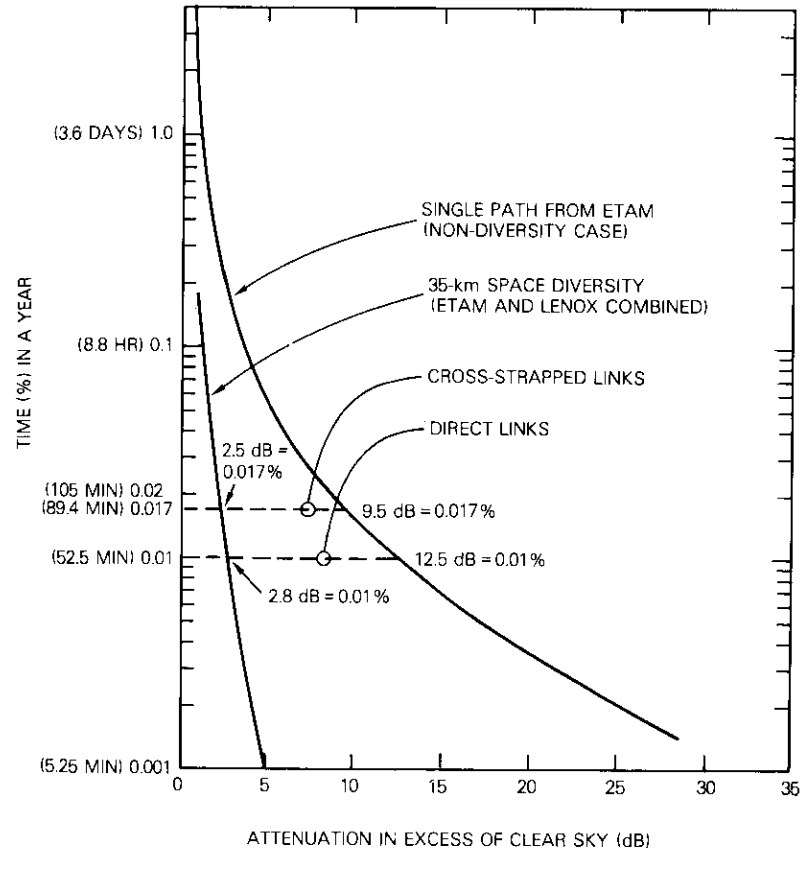


Figure 3. *Estimated Rain Attenuation at Etam-Lenox, W. Va., 11.2 GHz, 18° Elevation (Based on COMSAT Labs Propagation models; attenuation represents levels derived from the median value of 12-year statistics)*

the U.S. station: 0.40 + 0.55 = 0.95 dB at 11.2 GHz and 0.65 + 0.85 = 1.5 dB at 14.25 GHz.

c. Degraded sky (short-term). Clear sky plus attenuation for the

short-term of year. For the U.S. station: 0.40 + 9.5 (0.017%) = 9.9 dB at 11.20 GHz and 0.65 + 16.5 (0.017%) = 17.15 dB at 14.25 GHz.

The clear sky values include mist, fog, and high strato-cumulus clouds. These atmospheric phenomena are not observed by the rain-bucket techniques used to collect rain-rate data as required by the cloudy and degraded sky conditions. For clear sky, the design assumption for water vapor content was taken from the C.C.I.R. Draft Study Report 234-3 (Rev. 1976) in which the estimated value of sky noise temperature (at 11.2-GHz and 18° elevation angle) is 17 K. The attenuation values shown above for both clear and 10-percent sky were derived for the 14° to 18° elevation range expected at Etam, West Virginia.

Transmission margins

The Standard C earth station criteria established by INTELSAT were chosen to be adaptive to varying worldwide rain conditions and to provide a constant value for transmission planning in a similar manner to that in which the G/T value of 40.7 dB/K was used for Standard A stations. The objective is to provide sufficient carrier power at a station having a minimum G/T of 39.0 dB/K so that the basic telephone channel noise power is maintained at better than 8,000 pW0p for the space segment except for specific outages in a year. Based on this philosophy, the resulting INTELSAT V link budgets associated with the Standard C antenna in the cross-strapped modes are presented in Tables 5 to 7 as they would apply to a station located in the U.S.

Channel performance requirements

The performance criteria applicable to analog FM telephony for the Fixed Satellite Service, stated in C.C.I.R. Recommendation 353-2 are given in terms of percentages of any month [7]. One difficulty in applying these criteria to the 14/11-GHz band is relating yearly propagation statistics to monthly requirements. By correlating typical worst-month statistics with those expected for an entire year, INTELSAT generated short-term outage criteria for the 14/11-GHz band as summarized in Table 8 [8].

Criterion 1 was intended to realistically adapt the INTELSAT V transmission design and the C.C.I.R. criteria to the performance of a typical FM demodulator given in Table 9. The actual relationship between C/N and worst channel noise varies with channel capacity, test tone deviation, and the type of demodulator. For example, the value of C/N corresponding to 8,000 pW0p is 12.7 dB for 24 channels and 22.3 dB for 792 channels.

TABLE 5. LINK BUDGET DERIVATION
(Cross-strapped 14 → 4 and 6 → 11 GHz) (80-MHz Transponders)

Parameters	14 U.S. → 4 Hemi	6 Hemi → 11 U.S.
Saturation Flux Density (dBW/m²)	−75.3 [a]	−64.5 [b]
G/T of Satellite (dB/K)	+3.3	−11.6
C/N at Saturation (dB)	34.1	37.4
Input Backoff (dB)	−7.5	−10.0
Up-path Thermal C/N (dB)	26.6	27.4
Up-path Frequency Reuse C/I (dB)	27.7	21.8
Carrier/Intermod C/N (dB)	25.0	20.6
Saturated e.i.r.p. (dBW)	29.0	44.4
Earth Station Sensitivity (dB/K)	40.7	39.0
C/N at Saturation	23.5	28.1
Output Backoff	−4.5	−5.8
Down-path Thermal C/N (dB)	19.0	22.3
Down-path Frequency Reuse C/I (dB)	21.8	34.3
Total Space Segment Clear Sky C/N (dB), (7,500 pW0p)	15.8	16.3
Total Up-link e.i.r.p. (dBW)	80.6	89.1

[a] 5-dB gain step attenuation inserted.
[b] 7.5-dB gain step attenuation inserted.

TABLE 6. C/N SUMMARY
(Cross-strapped 14 → 4 and 6 → 11 GHz) (80-MHz Transponders)

Link Configuration	14 U.S. → 4 Hemi G/T = 40.7 dB/K	6 Hemi → 11 U.S. G/T = 39 dB/K
Up-path Thermal C/N (dB)	26.6	27.4
Up-path Frequency Reuse C/I	27.7	21.8
Carrier/Intermod C/N	25.0	20.6
Down-path Thermal C/N	19.0	22.3
Down-path Frequency Reuse C/I	21.8	34.3
Net Space Segment C/N (7,500 pW0p)	15.8	16.3
Internetwork C/I (1,000 pW0p)	25.4	25.2
Terrestrial C/I (1,000 pW0p)	25.4	25.2
Out-of-band Emission (500 pW0p)	28.4	28.2
Total Link C/N (10,000 pW0p)	14.6	15.0
Criterion 1 (C/N = 10 dB)	10.0	10.0

Similarly, the value of C/N corresponding to 50,000 pW0p varies from 5.6 dB for 24 channels using a threshold extension demodulator to 15 dB for 792 channels for both a conventional and a threshold extension

TABLE 7. MARGIN SUMMARY
(Cross-strapped 14 → 4 and 6 → 11 GHz) (80-MHz Transponders)

Link Configuration	14 U.S. → 6 Hemi	6 Hemi → 11 U.S.
Up-link Margin	5.4	16.0
Down-link Margin	7.3	9.5
System Margin	4.6	5.0

TABLE 8. INTELSAT V SHORT-TERM CHANNEL PERFORMANCE CRITERIA

Criterion 1	C/N at demodulator ≥10 dB, for all but 0.02% (105 min) of a year
Criterion 2	50,000 pW0p for ≤0.1% (8.8 hr) of a year
Criterion 3	1,000,000 pW0p for ≤0.01% (52.5 min) of a year

demodulator. The total link values of C/N shown in Table 6 are based upon a full transponder bandwidth, which will not be occupied by a single carrier, but rather by several carriers whose power and bandwidth have been adjusted to provide a C/N value in the range of 15 dB. In recognition of these variations, an average C/N value of 15 dB was selected as the basis for formulating INTELSAT V performance criteria.

TABLE 9. TYPICAL DEMODULATOR PERFORMANCE ASSUMED FOR INTELSAT V PERFORMANCE CRITERIA [a]

C/N at Input to Demodulator (dB)	Performance of Worst Channel (pW0p)
15	10,000
10	32,000 [b]
9	50,000
<6	560,000 [c]

[a] INTELSAT will provide adequate power to meet performance objectives for the few carrier sizes which have a required C/N below 15 dB.
[b] Assumed threshold of conventional demodulator.
[c] Considered equivalent to 1,000,000 pW0p.

INTELSAT V channel performance must be further described according to the specified transmission path since some links are cross-strapped

(6 → 11 and 14 → 4 GHz) and others are direct (14 → 11 and 6 → 4 GHz). The channel performance standards for these modes of operation are given in Table 10.

TABLE 10. INTELSAT V CHANNEL PERFORMANCE CRITERIA
(According to Path)

A. Long-Term (nominal conditions occurring for 90 percent of the year)

Noise power in any telephone channel will be 10,000 pW0p or less. It is assumed that 10,000 pW0p will be obtained for 80 percent of the time in the worst month of the year.

B. Short-Term (adverse weather conditions) *

Satellite Connection	Link and Transponder BW (MHz)	Percent/Minutes		
		Up-Path	Down-Path	Total
Cross-Strapped	14 → 4 (80)	0.017/89.4	0.003/15.8	0.02/105
	6 → 11 (80)	0.003/15.8	0.017/89.4	0.02/105
Direct	14 → 11 (80)	0.010/52.5	0.010/52.5	0.02/105
	14 → 11 (240)	0.010/52.5	0.010/52.5	0.02/105

* The percentages shown refer to the time in a typical year during which the C/N at the demodulator input is below 10 dB or exceeds 50,000 pW0p, whichever occurs first

Application of INTELSAT Standard C specification to first U.S. earth station

The propagation statistics presented for Etam and Lenox will be used to illustrate the transmission design procedure. As a starting point, Tables 11 and 12 summarize rain attenuation values for the 14- and 11-GHz links corresponding to the appropriate outage criteria.

Required antenna gain

Antenna gain is selected by examining the trade-off between the low-noise amplifier noise temperature and antenna size. The first step is to determine the values of T_1, T_2, L_1, and L_2 from the INTELSAT specification. As discussed above, the clear sky temperature is taken as 17 K. It was assumed that the contributions attributed to the antenna, such as spar blockage, spillover, and ground noise pickup due to sidelobes, were 10 K.

The earth station system temperature value also includes contributions resulting from losses between the antenna and the low-noise amplifier, which have been assumed as follows:

Transmit reject filter	0.15 dB
Terrestrial link reject filter	0.05 dB
Waveguide and coupler	0.10 dB
Switch	0.05 dB
Total	0.35 dB

TABLE 11. 14-GHz ATTENUATION VALUES APPLICABLE TO THE WEST SPOT BEAM (Etam/Lenox, West Virginia, U.S.)

Criteria	Link	% of Year	Up-Link (14.25 GHz)					
			Single			Diversity		
			A*	Clear	Total	A*	Clear	Total
Short-Term	14 U.S. → 4	0.017	16.5	+ 0.65 =	17.15	4.1	+ 0.65 =	4.75
	14 U.S. → 11	0.01	24.0	+ 0.65 =	24.65	5.0	+ 0.65 =	5.65
	6 → 11 U.S.	0.003	—	—	—	—	—	—
	14 → 11 U.S.	0.01	—	—	—	—	—	—
Long-Term	All	10	0.85	+ 0.65 =	1.50	0.85	+ 0.65 =	1.50

* A represents attenuation in excess of clear sky (obtained from rain-rate data).

TABLE 12. 11-GHz ATTENUATION VALUES APPLICABLE TO THE WEST SPOT BEAM (Etam/Lenox, West Virginia, U.S.)

Criteria	Link	% of Year	Down-Link (11.20 GHz)					
			Single			Diversity		
			A*	Clear	Total	A*	Clear	Total
Short-Term	14 U.S. → 4	0.003	—	—	—	—	—	—
	14 U.S. → 11	0.01	—	—	—	—	—	—
	6 → 11 U.S.	0.017	9.5	+ 0.40 =	9.90	2.50	+ 0.40 =	2.90
	14 → 11 U.S.	0.01	12.5	+ 0.40 =	12.90	2.80	+ 0.40 =	3.20
Long-Term	All	10	0.55	+ 0.40 =	0.95	0.55	+ 0.40 =	0.95

* A represents attenuation in excess of clear sky (obtained from rain-rate data).

Based on these values and the assumptions listed below, the 10-percent system noise temperature can be derived using equation (1):

$$T_{\underset{(10\%\ \text{sky})}{sys}} = T_1 = 0.7\frac{T_{sky}}{L_1 L_b} + \frac{T_V(L_1 - 1)}{L_1 L_b} + \frac{T_a}{L_b} + \frac{T_o(L_b - 1)}{L_b} + T_r + T_f \quad . \tag{1}$$

where
T_1 = system noise temperature for all but 10 percent of the year (10-percent sky) referred to LNA flange
T_{sky} = clear sky temperature (17 K)
T_a = antenna noise contribution (10 K)
T_V = water vapor temperature (285 K)
T_o = component temperature (290 K)
L_1 = attenuation excess over clear sky, for all but 10 percent of year (1.14 or 0.55 dB)
L_b = component losses (1.08 or 0.35 dB)
T_f = contribution of receiving stages following low-noise amplifier (3 K)
T_r = low-noise amplifier temperature (150 K).

If equation (1) is solved for different values of low-noise amplifier temperature and for two cases of rain attenuation, the G/T reduction due to rain can be plotted as a function of amplifier temperature as shown in Figure 4. Note that the G/T drops 1 to 1.5 dB more when the amplifier temperature is 50 K compared to when it is 150 K. In addition, the cost of a 150-K amplifier is substantially lower than that of a 50-K amplifier. Recent advances in field effect transistor technology indicate that a 150-K amplifier is feasible if Peltier-type cooling is used, and thus allows a simplification of LNA hardware with a corresponding improvement in reliability. For these reasons, a value of 150 K was chosen for T_r, which provides a 10-percent sky system noise temperature from equation (1):

$$T_{\underset{(10\%\ \text{sky})}{sys}} = T_1 = 225.8\ \text{K}(23.5\ \text{dB})$$

The same method can be used to determine T_2 (system noise temperature expected under degraded sky conditions) based on the following values:

$$L_2(0.017\%) = 9.5\ \text{dB, single station}$$

$$L_2(0.017\%) = 2.5\ \text{dB, diversity pair}$$

which yield

$$T_{sys}\ \text{(short-term)} = T_2 = 419.3\ \text{K(26.2 dB), single station}$$

$$T_2 = 307.2\ \text{K(24.9 dB), diversity pair}$$

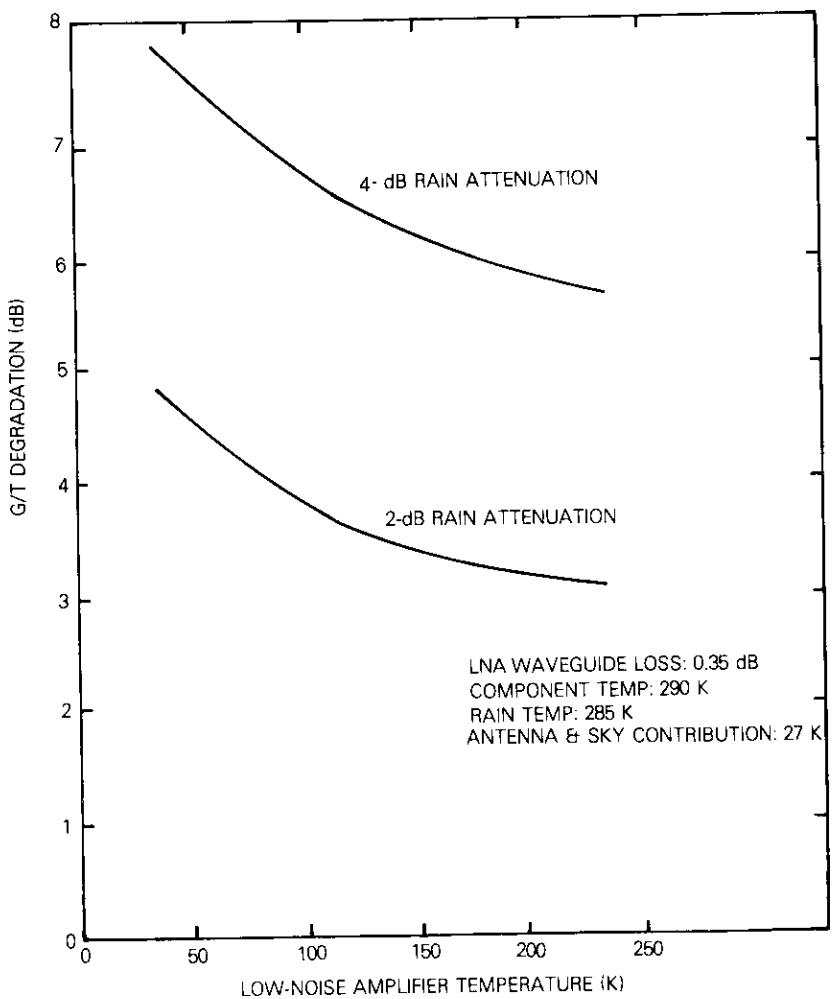


Figure 4. *G/T Degradation vs LNA Temperature with Rain Attenuation as a Parameter*

The required antenna gain can now be determined as follows:

$$G_a = G/T + T_i + L_i \qquad (2)$$

where G_a = antenna gain measured at the LNA flange
G/T = Standard C specification (G/T = 39.0 dB/K)
T_i = system noise temperature for the specified time
L_i = attenuation values for the specified time.

With equation (2), the following values of receive gain are obtained at 11.2 GHz.

a. Long-term: $G_a = 39.0 + 23.5 + 0.55 + 0.40 = 63.5 + 0.5 + 0.2^* = 64.2$ dBi;†

b. Short-term (single site): $G_a = 29.5 + 26.2 + 9.5 + 0.40 + 0.7 = 66.3$ dBi;

c. Short-term (diversity pair): $G_a = 29.5 + 24.9 + 2.5 + 0.40 + 0.7 = 58.0$ dBi.

It is apparent that the antenna size for the West Virginia site is governed by the short-term condition when diversity is not used and by the long-term condition when diversity is used. The antenna gain is about 2.0 dB larger without diversity and, as shown in the next section, a much higher transmitter power would also be required.

Based on the above values, the final specifications, to be measured under a clear sky (not 10 percent), for each antenna in the Etam-Lenox diversity pair are :

$$G/T = 41.3\ \text{dB/K}$$

$$G_a = 64.2\ \text{dBi (receive)}$$

$$T_e = 196\ \text{K(22.9dB)} \quad .$$

Transmitter power

The required transmitter power is determined by considering several

*An additional 0.5 dB has been added to the 10-percent attenuation value to allow for uncertainty in rain attenuation estimates. Another 0.2 dB has been used for tracking error (step-track system).

† G_a required to compensate for 10-percent cloudy sky.

factors, including losses between the transmitter and the antenna, allowed margins, and the estimated values of attenuation corresponding to the outage time limits.

One example of a fully loaded satellite in the West spot beam would be four 612-channel carriers in the cross-strapped links and another four 792-channel carriers in the direct links. However, since numerous changes are expected in the number of carriers (and their size*) over the INTELSAT V lifetime, planning for the U.S. Standard C has considered a transmitter design capable of handling up to eight carriers, each with a maximum bandwidth of 36 MHz and an e.i.r.p. of 84.0 dBW (a total required e.i.r.p. for eight carriers of 93 dBW).

To provide this capability, several assumptions were made regarding the transmit antenna gain:

$$
\begin{aligned}
G_{\text{(dry sky)}} &= 66.0 \text{ dBi at } 14.25 \text{ GHz } (64.2 \text{ dBi at } 11.2 \text{ GHz}) \\
&\quad -1.5 \text{ dB (10-percent cloudy sky attenuation)} \\
&\quad -0.2 \text{ dB (tracking error)} \\
G_{\text{(effective)}} &= 64.3 \text{ dBi (transmit)} \ .
\end{aligned}
$$

Therefore, the required transmitter power at the antenna flange, per carrier, would be 19.7 dBW (84—64.3) or 28.7 dBW (93—64.3) for eight carriers. A total of 740 W would be required for eight carriers with a diversity pair of antennas in operation. In comparison, without diversity, with the 2.0 dB additional antenna gain, and 24.0 + 0.65 = 24.65 dB (0.01 percent) short-term rain attenuation, the total power required would be 43.2 dBW (allowing for a 6.5-dB margin):

93.00 dBW	(required e.i.r.p.)
−68.10 dBi	(nondiversity Tx gain, 66.3 + 1.8 = 68.10)
+24.00 dB	(short-term attenuation over clear sky)
+ 0.65 dB	(clear sky attenuation)
− 6.50 dB	(link margin)
+ 0.20	(tracking error)
43.25 dBW	(nondiversity power for eight carriers) .

* FDM/FM carriers will be used ranging in size from 12 channels/1.25 MHz to 972 channels/36 MHz.

Obviously, the power requirements for a nondiversity site would be at an unacceptable level, approaching 22 kW at the antenna flange and substantially more at the transmitter output to compensate for waveguide losses.

With diversity, two methods are available to provide 740 W at the antenna flange. One method would use a single wideband (500-MHz) transmitter in the 5-kW class, allowing 2.0 dB for losses and an 8.5-dB output backoff. The other method, which has merits of economy and reliability, uses several low-power transmitters coupled to the antenna through a filter-combiner. If each transmitter were assigned only two carriers, then the filter-combiner could be utilized to avoid transmission of intermodulation products and allow the amplifiers to be operated close to saturation.

When the amplifiers operate near saturation, it is important to ensure that the selected high-power amplifier will meet relevant crosstalk specifications. A preliminary examination of klystrons excluded these devices from consideration during the U.S. Standard C design stage because they could not meet the crosstalk specification with high drive levels and large carrier sizes. Even with a traveling wave tube amplifier (TWTA), gain slope values must be minimized throughout the transmitting chain.

The actual computation of required transmitter power for Etam and Lenox is based on the assumption that only two carriers will be assigned per transmitter and that wideband TWTs will be procured. Transmitter size has been derived based on the assumptions given in Table 13. From

TABLE 13. ASSUMPTIONS RELATIVE TO TRANSMITTER SIZE (U.S. Diversity Station)

Power Required at Antenna Flange (93 W per carrier) (dBW)	19.7
TWT End-of-Life Degradation (dB)	+0.5
Waveguide Losses	+0.5
Spare Transfer Switch	+0.2
Test Signal Injection Coupler	+0.5
Two Carriers	+3.0
Intermodulation Loss Near Saturation	+1.3
Isolators	+0.35
Filters	+0.35
Filter Combiner	+0.80
Output Backoff	+0.50
Total TWT Saturated Power Required to Support Two Carriers (dBW)	27.7 (≃600 W)

these calculations, the final specified tube power for each of the four transmitters will range from 600 to 700 W, depending on actual losses. This planning is based on the premise that the west spot beam is not pointed directly at West Virginia to allow for the possibility that Canada may construct a Standard C antenna. If the beam were pointed directly at West Virginia, the required transmitter power would be 1 to 2 dB less. Although power control is not required until the margins in Table 7 are exceeded, the excess power in a beam center situation would be used to improve the overall circuit availability during periods of heavy rain.

Diversity Interconnection link (DIL)

When the diversity station is operative during adverse weather conditions, the transmitted and received carriers must be carried to and from

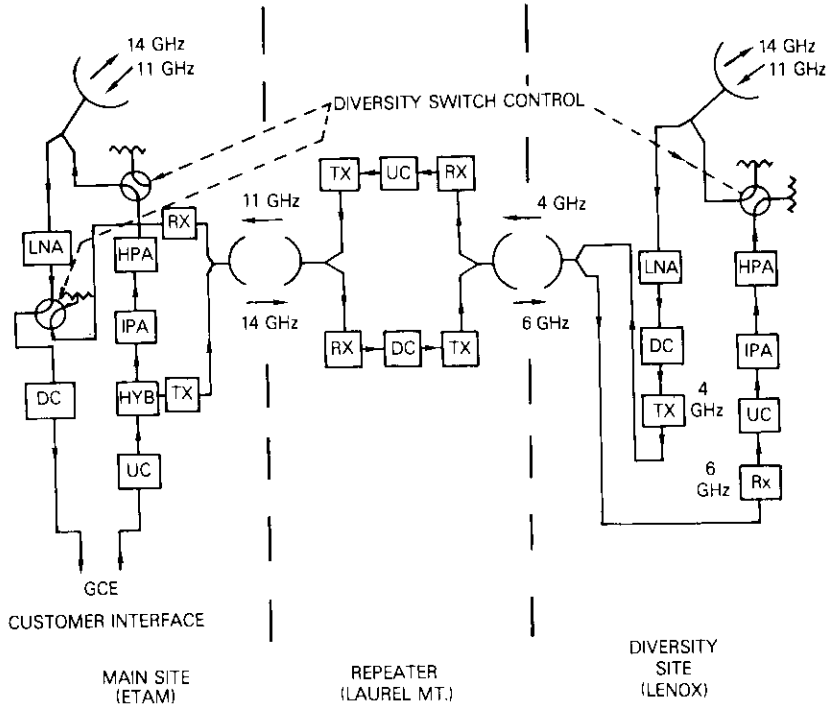


Figure 5. *U.S. Standard C Earth Station Block Diagram Showing Diversity Interconnection Link (Etam-Lenox, W. Va.)*

this station unless a second customer interface is provided. The most economical method of transmission is to avoid modulation and demodulation at the diversity station and to retransmit signals with the same bandwidths and deviations that are used on the satellite link.

Figure 5 is a simplified block diagram, representing current planning for the DIL. Pending FCC approval, the first hop will be at 14/11 GHz to avoid terrestrial interference with other nearby systems. The second hop to the Lenox station will be in the 6/4-GHz bands. Frequency plans and noise budgets for each link are shown in Figures 6 and 7, and Table 14 [9].

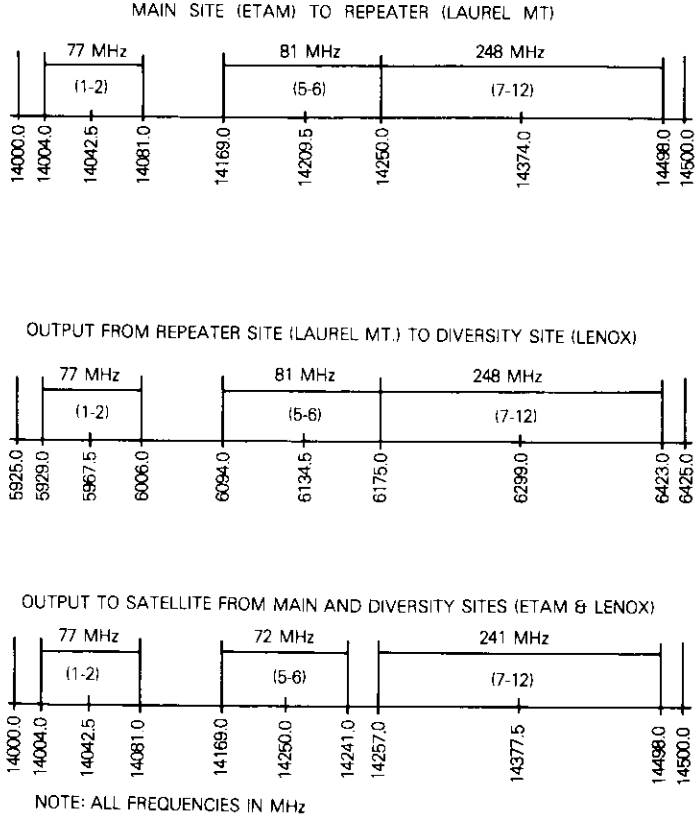


Figure 6. *Main Site to Diversity Site Frequency Plan (Etam to Laurel Mt. to Lenox)*

With reference to Table 14, INTELSAT and the C.C.I.R. include the DIL noise contribution in the overall earth station noise budget. For the U.S. DIL, 300 to 350 pW0p of noise is planned and allowance for this noise contribution will be taken from other earth station impairment budgets. For example, the normal allocation of 500 pW0p for out-of-band emissions from other earth stations will not be required since there will be a maximum of two stations occupying the west spot beam. In the 11- and 14-GHz portions of the DIL, planning must also account for heavy rainfall. For this purpose sufficient power margin has been incorporated into the 2-km link as a safeguard during these periods.

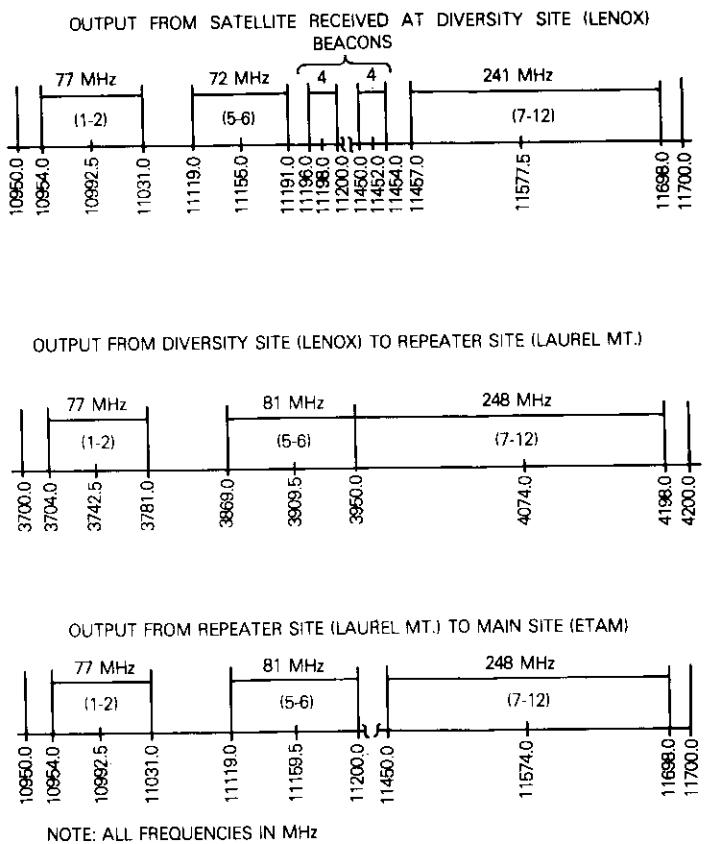


Figure 7. *Diversity Site to Main Site Frequency Plan (Lenox to Laurel Mt. to Etam)*

Overall long-term transmission performance

Figure 8 gives the estimated performance of the receiving system using a diversity pair. This considers only rain degradation in the U.S. and assumes that a possible heavy European rain event will not coincide with one in West Virginia. Figure 9 shows the reduction in power reaching the satellite from the west spot beam.

Conclusion

This paper has described the initial transmission planning employed by the U.S. to meet INTELSAT's Standard C (14/11-GHz) earth station specifications. Since this paper deals with the first U.S. Standard C earth station, which is required for a 1979 launch of INTELSAT V, the design procedure has been outlined so that it can be applied to other stations. The 14/11-GHz propagation parameters, as well as channel performance

TABLE 14. NOISE BUDGET FOR A SINGLE REPEATER WIDEBAND INTERCONNECT SYSTEM (pW0p)
(Etam/Lenox Diversity Interconnect Link)

Noise Source	Link 1		Link 2		Total
	Etam	Laurel Mountain	Laurel Mountain	Lenox	
Feeder Echo	1.3	1.3	1.3	1.1	5.0
Radio Frequency Thermal Noise Including 5-dB Fade					
Etam-Lenox	—	1.0	—	46.8	—[a]
Lenox-Etam	1.5		74.1		75.6
Equipment Intermodulation[b] (typical)	—	52.0	—	5.0	57.0
Multiplex Noise	—	—	—	—	—
Radio Frequency Inference					200.0
Total System Downlink Noise					338 pW0p[c]

[a] Lenox-Etam is larger.
[b] After any necessary group delay equalization.
[c] Not to be exceeded for more than 20 percent of any month.

requirements, have been discussed in detail and used to derive the transmitting and receiving system of a diversity pair of antennas to be located at Etam and Lenox, West Virginia. Traffic is to be routed to the antenna with favored propagation conditions through a DIL, which is to be controlled by an automated monitoring subsystem. Figure 10 provides an artist's conception of the facilities planned for Etam and Lenox while Table 15 lists the physical features expected for both antennas.

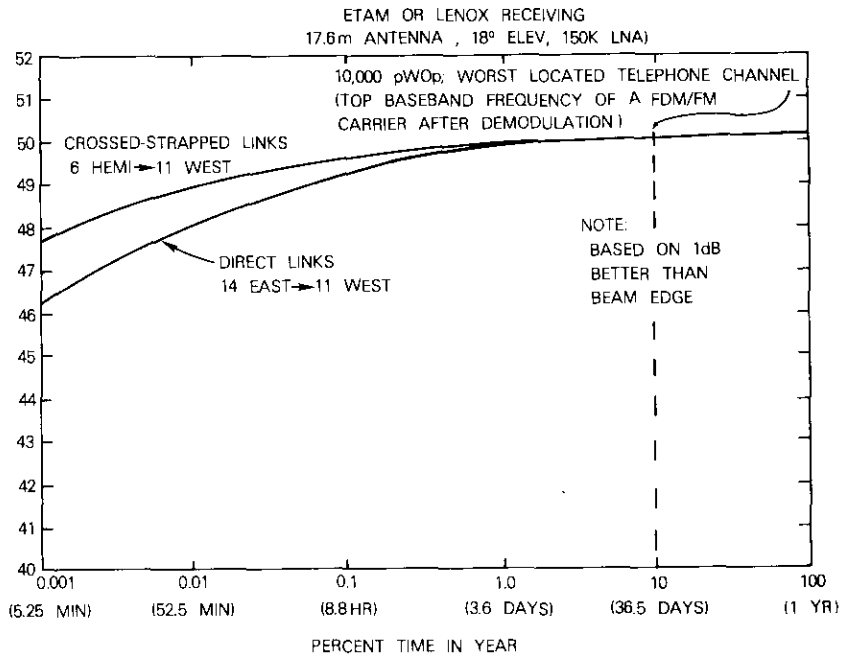


Figure 8. *Telephone Channel Signal-to-Noise Ratio vs Time*

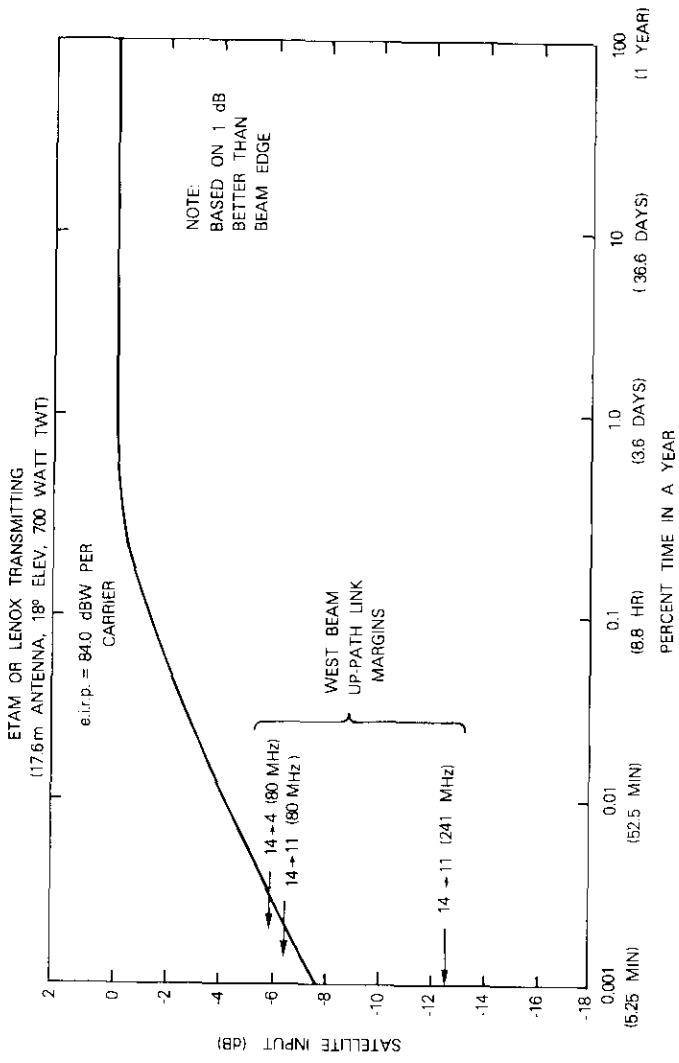


Figure 9. *Relative Power Received at Satellite vs Time (U.S. to Europe link, 14 → 4 and 14 → 11 GHz)*

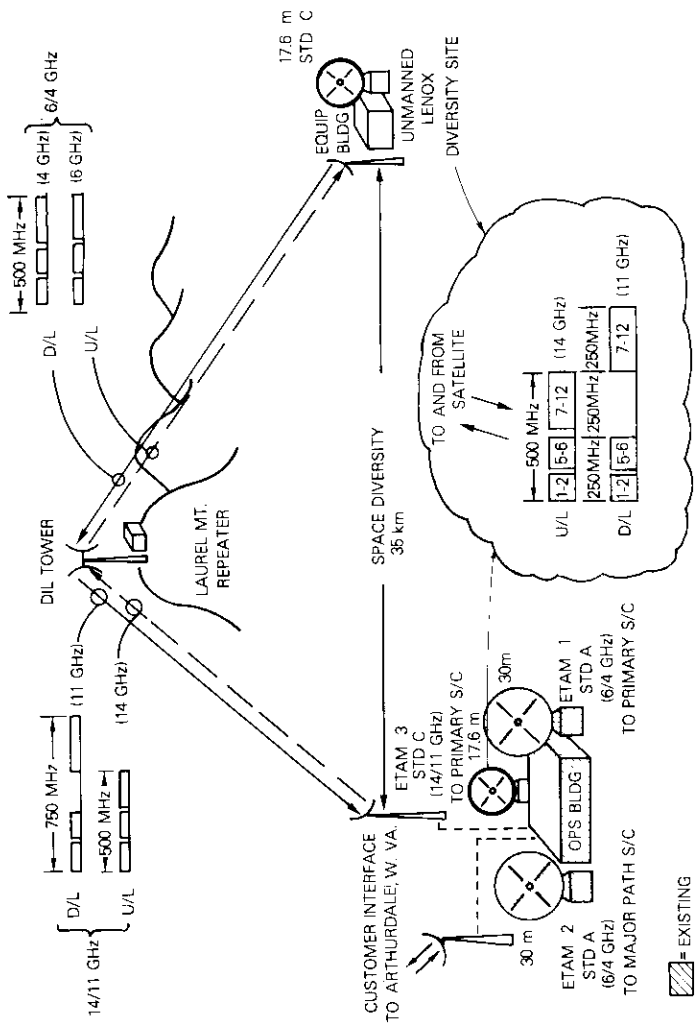


Figure 10. *U.S. Standard C Earth Station, Diversity Pair (Etam and Lenox, W. Va.)*

TABLE 15. FINAL CONFIGURATION OF U.S. STANDARD C EARTH STATION PLANT (Etam and Lenox, West Virginia)

Parameter	Main Station Etam	Diversity Site Lenox
Terrestrial Interface	Yes	No
Manned	Yes	No
Polarization	Linear	Same
Power Axial Ratio	>30 dB	Same
Transmit Frequency	14.0–14.5 GHz (500 MHz)	Same
Receive Frequency	10.95–11.20 GHz (250 MHz) 11.45–11.70 GHz (250 MHz)	Same Same
Transmit Gain (clear sky) (dBi)	66.0 (14.25 GHz)	Same
Receive Gain (clear sky) (dBi)	64.2 (11.20 GHz)	Same
System Noise Temperature (clear sky) (K)	196	Same
G/T (clear sky) (dB/K)	41.3 (11.20 GHz)	Same
Antenna Diameter [a]	58 ft (17.6 m)	Same
HPA Size (without power control) [b] (W)	600 to 700	Same
Carriers per HPA	2	Same
Number of HPAs (maximum), on Line	4	Same
Number of Carriers (maximum, up to 972 channel/36 MHz each)	8	Same

[a] Estimate based on a surface tolerance of 0.03 in. (0.08 cm) at 14.5 GHz. Actual construction size may differ depending upon contractor design to meet specification requirements for gain and G/T, and will probably be 60 ft (18.3 m).

[b] Assumes beam edge operation. With the west spot beam pointed directly at West Virginia, 1 to 2 dB of power control can provide circuit availability beyond that required by the Standard C specification.

Acknowledgments

The authors are indebted to numerous contributors for the material presented in this document. The performance criteria relative to rain attenuation were developed by H. Weiss. The transmission engineering and margin selections were established by J. Dicks and his co-workers. Rain attenuation models were prepared by G. Hyde, J. Harris, and J. Maas, of COMSAT *Laboratories. Criteria for equipment design including antennas, low-noise receivers, transmitters, and the diversity link were developed in a task force study headed by R. Benedict, with substantial contributions from* COMSAT *Laboratories personnel including F. Assal, C. Cotner, W. English, W. Sandrin, and W. Sones, under the guidance of L. Pollack. Engineering,* FCC *application, and specification preparation were handled by a group led by B. Williams, which included G. Caprio, E. Carpenter, A. Donahoe, E. Magnusson, R. Price, K. Rasmusson, and L. Smith.*

References

[1] J. Dicks, "Transmission Planning for INTELSAT V," *Proc. INTELSAT V Earth Station Technology Seminar*, June 1976, Munich, Federal Republic of Germany.

[2] R. N. Benedict, "Earth Station Engineering," *Proc. INTELSAT V Earth Station Technology Seminar*, June 1976, Munich, Federal Republic of Germany.

[3] Statistics prepared by the Propagation Studies Department of COMSAT Laboratories.

[4] G. Hyde and E. Dutton, Private Communication.

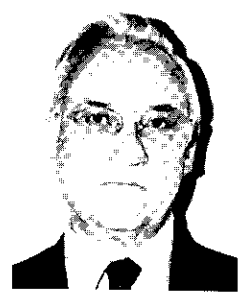
[5] P. L. Rice and N. R. Holmberg, "Cumulative Time Statistics of Surface Point-Rainfall Rates," *IEEE Transactions on Communications*, COM-21, No. 10, pp. 1131–1136.

[6] D. V. Rogers and G. Hyde, "Diversity Measurements of 11.6-GHz Rain Attenuation at Etam/Lenox, West Virginia," *COMSAT Technical Review*, Vol. 9, No. 1, Spring 1979.

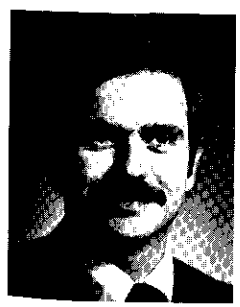
[7] C.C.I.R. Rec. 353-2, Volume IV, C.C.I.R. Green Book, Geneva, 1975. A third revision to this recommendation has been proposed in C.C.I.R. Doc. 4/215, "Conclusion of the Interim Meeting of Study Group 4," Geneva, June 1976.

[8] H. Weiss, *Proc. INTELSAT V Earth Station Technology Seminar*, June 1976, Munich, Federal Republic of Germany.

[9] Application of COMSAT, AT&T, ITT World Communications, RCA Global Communications, Inc., and Western Union International, Inc. to Federal Communications Commission, Filing No. 291-CSG-P-78, October 14, 1977.



L. F. Gray received the B.A.Sc. from the University of British Columbia and the M.S. in Telecommunications Operations from George Washington University. He was with Canadian Marconi from 1938 to 1943 and served in the Royal Canadian Navy until 1945. From 1945 until 1951, he worked on transmitter development at Federal Telephone and Radio and Air Associates. From 1951 to 1964, he was employed by Federal Telecommunication Laboratories where he was concerned with equipment development for television broadcasting, troposcatter, and space communications. He joined COMSAT *in 1964 and is presently manager of Systems Engineering in the Earth Station Engineering Division. He is a registered Professional Engineer in the District of Columbia.*



Martin P. Brown, Jr. received a B.E.E. from the Georgia Institute of Technology in 1966. Before entering military service, he was employed by IBM at Cape Kennedy where he was involved in the construction of the first Saturn V launch vehicle. From 1967 to 1971, he served as an engineering communications officer with the U.S. Air Force assigned to HQ USAFSS. He joined COMSAT *Laboratories in 1971, and later became part of the Engineering Division where he was primarily involved in transmission planning for INTELSAT IV, IV-A, and V. After leaving* COMSAT *in August 1978 as Manager of Transmission Systems, he joined* INTELSAT *to become Chief of Satellite Transmission Modeling.*

Mr. Brown is Chairman of the IEEE Communications and Broadcast Satellite Systems Committee, a member of AIAA, and a registered Professional Engineer.

Index: statistical analysis, antenna, sidelobe, radiation, interference

*Statistical properties of antenna sidelobes**

P. R. KARMEL

(Manuscript received November 30, 1978)

Abstract

This paper investigates interference effects due to antenna sidelobe radiation when a geostationary satellite is operating in the environment of multiple interfering radiators. A theoretically derived probability density function for sidelobe levels is obtained, and the power density statistics are found to be exponential for typical parameters. A general approach to calculating probability levels for interference-to-carrier ratios is given, and formulas for an exponential distribution and several examples are presented.

Introduction

Geostationary satellite systems, because of the limited orbital spacing available, operate in the presence of interference due to earth stations communicating with other geostationary satellites. As the number of earth stations and satellites increases, the problem of interference from antenna

*This paper is based upon work performed at COMSAT Laboratories under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Views expressed are not necessarily those of INTELSAT.

sidelobes becomes more serious. Current C.C.I.R. recommendations [1] require that for off-axis angles θ greater than 1° no more than 10 percent of the peak antenna sidelobe gain values can exceed $32 - 25 \log \theta$. This standard is not suitable for assigning a probability level to interference, particularly with multiple interfering sources.

The calculation of expected values of interference in the presence of multiple sources requires a description of the statistical properties of the antenna sidelobes. A C.C.I.R. Study Group Document [2] presented experimental data which suggested a Rayleigh distribution law for the sidelobe gain. This paper also perspicaciously noted a similarity in the mathematical relations between the far field antenna pattern and the aperture field distribution, and the mathematical relations between the amplitude envelope of narrowband noise plus sine wave signal and the filter input spectrum. The statistics of the latter have been studied by S. O. Rice [3] and shown to follow a Rayleigh distribution law (within certain limits of parameters). B. D. Steinberg [4] has similarly derived the probability density function of the sidelobe field amplitude for a random array of radiators. A paper by M. C. Jeruchim [5] develops some interference implications of a Rayleigh distribution for two interfering sources, utilizing the data from the C.C.I.R. Study Document.

This paper theoretically develops the sidelobe probability distribution and the mean values of the field and power in the sidelobes. For certain ranges of parameters, the sidelobe field statistics are Rayleigh, and the power density statistics are consequently exponential. For other parameter values, these distributions are closely related to the Rayleigh and exponential distributions. A method of calculating the interference-to-carrier ratio is also presented based on the statistics of the antenna's gain probability density function in the sidelobe regions. Results are given for various arrangements of interfering sources, each with an exponential probability density function.

Derivation of sidelobe statistical distribution

Field amplitude probability density function (PDF)

The PDF of the sidelobe field amplitude for a random array of radiators has been derived by B. D. Steinberg [4]. This derivation, which follows the work of S. O. Rice [3] for the statistics of the envelope of a noise plus sine wave signal, has been adapted to the present problem.

Attention can be focused on the normalized field pattern for a plane aperture [6] given by

$$g(\theta, \phi) = \iint_S F(r', \phi')\, e^{jkr' \sin\theta \cos(\phi - \phi')}\, dS' \quad . \tag{1}$$

The radiation intensity $K(\theta, \phi)$, which is the time average power per unit solid angle, is

$$K = K_1(\theta)\, |g(\theta, \phi)|^2 \quad , \qquad K_1 = \frac{\cos^4 (\theta/2)}{2\lambda^2 Z_0} \quad . \tag{2}$$

In equations (1) and (2), θ and ϕ are the spherical coordinates of a far field point in space, and r' and ϕ' are the cylindrical aperture coordinates. The aperture illumination can be written in terms of real amplitude and phase functions as

$$F(r', \phi') = [f_0(r', \phi') + f(r', \phi')] \exp\{j[\psi_0(r', \phi') + \psi(r', \phi')]\} \quad . \tag{3}$$

The amplitude and phase of the desired aperture illumination are represented by f_0 and ψ_0; f and ψ are the random amplitude and phase error over the aperture.

In the Appendix, it is demonstrated, with certain assumptions, that the PDF for the field amplitude, $|g|$, where g is given in equation (1), is

$$W_{|g|}(|g|) = \frac{2|g|}{\sigma_g^2} \exp\left[- \frac{|\bar{g}|^2 + |g|^2}{\sigma_g^2} \right] I_0 \left(\frac{2|\bar{g}|\, |g|}{\sigma_g^2} \right) \tag{4a}$$

In this equation

$$|\bar{g}| = |E\{g\}| = \sqrt{[E\{Re(g)\}]^2 + [E\{Im(g)\}]^2} \tag{4b}$$

is the magnitude of the mean value of g,

$$\sigma_g^2 = E\{|g|^2\} - |E\{g\}|^2 = E\{gg^*\} - |\bar{g}|^2 \tag{4c}$$

is the variance of g, and $I_0(z)$ is the modified Bessel function of the first kind of order zero. In general, $|\bar{g}|$ and σ_g are functions of angular position θ, ϕ.

For those values of θ, ϕ (generally, off the main beam), where $|\bar{g}| \ll \sigma_g$, equation (4) for the field amplitude PDF reduces to the Rayleigh distribution

$$W_{|g|}(|g|) = \frac{2|g|}{\sigma_g^2} \exp\left(-\frac{|g|^2}{\sigma_g^2}\right) \quad . \tag{5}$$

Radiation intensity PDF and gain

The PDF of K, equation (2), is determined from equation (4) or (5), respectively:

$$W_K(K) = \frac{1}{(K_1\sigma_g^2)} \exp\left[-\frac{K_1|\bar{g}|^2 + K}{K_1\sigma_g^2}\right] I_0\left(\frac{2|\bar{g}|\sqrt{K}}{\sqrt{K_1}\sigma_g^2}\right) \tag{6}$$

$$W_K(K) = \frac{1}{(K_1\sigma_g^2)} \exp\left(-\frac{K}{K_1\sigma_g^2}\right) \quad , \quad |g| \ll \sigma_g \quad . \tag{7}$$

Equation (7) represents an exponential distribution that has a single parameter $K_1\sigma_g^2$, which is its mean value, $\bar{K} = E\{K\} = K_1\sigma_g^2$.

Antenna gain, G, is related to K by $G = 10 \log_{10} K$. The probability that the gain G exceeds some given level G_1 is

$$Pr(G > G_1) = \int_{G_1}^{\infty} Ce^{CG} W_K(e^{CG})\, dG \tag{8}$$

where $C = 0.2303$ and either equation (6) or (7) may be used for W_K.

In terms of field amplitude, G is equal to $20 \log_{10} |g|$ and

$$Pr(G > G_1) = \int_{G_1}^{\infty} \frac{C}{2\sqrt{K_1}} e^{CG/2} W_{|g|}\left(\frac{1}{\sqrt{K_1}} e^{CG/2}\right) \mathrm{dG} \quad . \tag{9}$$

Normalized distribution function

Equations (4)-(7) may be normalized by changing variables

$$t = \frac{|g|^2}{\sigma_g^2} = \frac{K}{K_1\sigma_g^2} \tag{10a}$$

$$p = \frac{|\bar{g}|^2}{\sigma_g^2} \tag{10b}$$

to yield

$$W_t(t) = \exp[-(p + t)]\, I_0(2\sqrt{pt}) \tag{11}$$

$$W_t(t) = \exp(-t) \quad , \quad p = 0 \quad . \tag{12}$$

The parameter p is the ratio of average field amplitude to field amplitude variance. Equation (11) has only a single parameter p and is therefore concisely represented in graphical form. Figure 1 shows the probability that t is greater than some value t_1 with p as a parameter:

$$Pr(t > t_1; p) = \int_{t_1}^{\infty} \exp[-(p + t)]\, I_0(2\sqrt{pt})\, dt \quad . \tag{13}$$

Derivation of parameters

This section presents theoretically obtained expressions for $|\bar{g}|$ and σ_g^2 and attempts to relate these to measurable quantities. In summary, equations (4b) and (4c) give

$$\sqrt{K_1}\,|\bar{g}| = \sqrt{K_1}\,|E\{g\}| \tag{14a}$$

$$K_1\sigma_g^2 = K_1[E\{|g|^2\} - |E\{g\}|^2] \quad . \tag{14b}$$

The first quantity, $\sqrt{K_1}\,|\bar{g}|$, is the magnitude of the expected or mean value of the complex scalar electric far field phasor. This quantity is measured by determining the magnitude and phase of the field (relative to the field at beam peak) at a given angle for many statistically identical antennas, averaging the real and imaginary parts, and then finding the magnitude of these means. Even if it is assumed that sampling over a limited angular range for a single antenna will correctly estimate this statistic, field strength and phase are not normally measured; therefore, this statistic cannot be obtained from the usual recorded power pattern.

A theoretical expression for $|\bar{g}|$ can be derived by simplifying the model, equation (3), by neglecting random variations in the amplitude of the aperture field, f. Small amplitude variations will have little effect on the field, equation (1), in comparison with the effect of small phase variations. Substituting equation (3) with $f = 0$ into equation (1), reversing the order of spatial integration and averaging gives

$$E\{g(\theta, \phi)\} = \iint_S f_0(r', \phi')\, E\{e^{j\psi(r',\phi')}\}\, e^{j\psi_0(r',\phi')} e^{jkr'\sin\theta\cos(\phi-\phi')}\, dS' \quad .$$

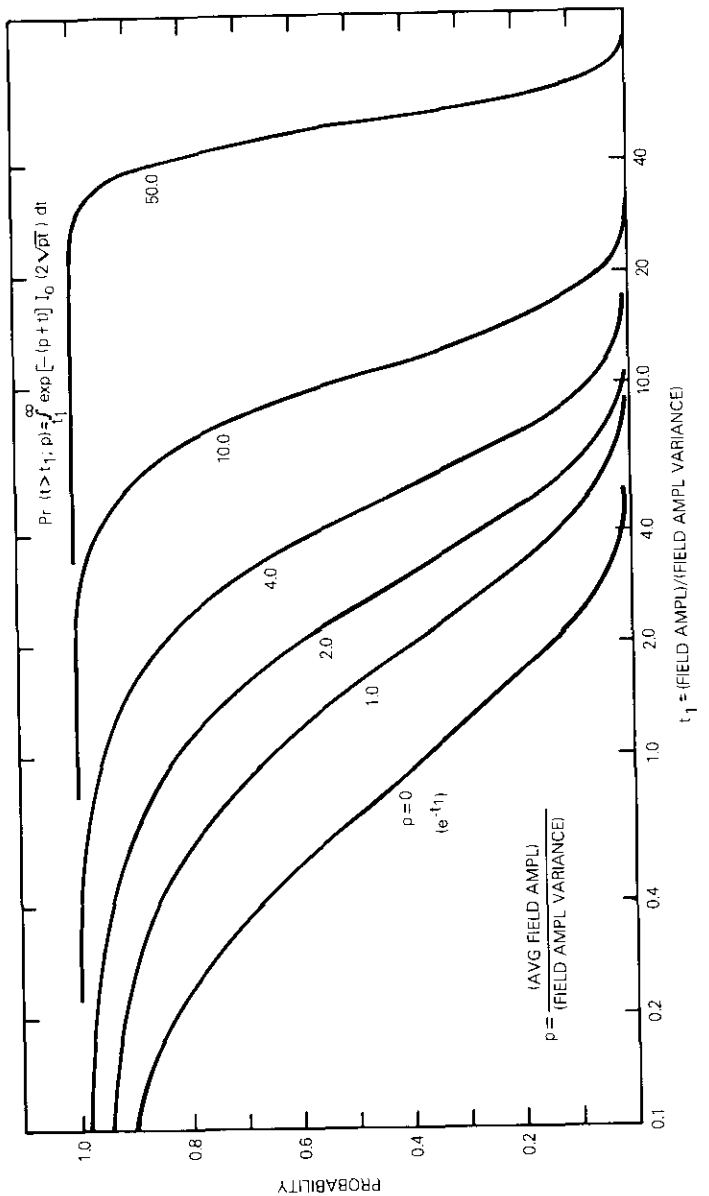


Figure 1. *Probability That t Exceeds t_1 with Parameter p*

If the random phase angle ψ is normal with zero mean and with variance σ^2, then $E\{e^{j\psi}\} = e^{-\sigma^2/2}$ and

$$\sqrt{K_1}\,|\bar{g}| = \sqrt{K_1}\,|g_0(\theta, \phi)|\, e^{-\sigma^2/2} \tag{15}$$

where

$$g_0 = \iint_S f_0(r', \phi)'\, e^{jkr'\sin\theta\cos(\phi - \phi')}\, dS' \tag{16}$$

is the desired field pattern, or the field pattern in the absence of phase error. Equation (15) shows that as the variance of the normal PDF of the phase increases (*i.e.*, probable phase angles occur over a broader range), the expected field decreases exponentially.

The variance term $K_1\sigma_g^2$ in equation (14b) is the difference between the mean of the squared field magnitude, which is the mean power density, and the squared mean field. The quantity $K_1E\{|g|^2\}$ should be estimated by averaging the measured power pattern at a given angle for many statistically identical antennas. If it is assumed that this quantity can be estimated from the mean value of the power pattern of a single antenna taken over a limited angular range, then it is directly obtainable from the usual recorded power pattern.

Ruze [7] has developed an approximate analytic expression for this quantity for a circular aperture of radius a assuming no random amplitude variation and a uniform desired phase function [$f = 0$ and $\psi_0 = 0$, equation (3), respectively]. Ruze assumes that $E\{(\psi_1 - \psi_2)^2\}$, where ψ_1 and ψ_2 are the random phases at two points on the aperture, is given by $2\sigma^2(1 - e^{-\tau^2/c^2})$, where c is a phase correlation distance and τ is the distance between the points. For a separation which is large compared to the correlation distance, $\tau/c \gg 1$, the two random phases are independent. Since each phase is normally distributed with zero mean and variance σ^2, their difference is also normal with zero mean and variance $2\sigma^2$. Conversely, for $\tau/c \ll 1$, ψ_1 and ψ_2 should be closely correlated so that their difference approaches zero. Ruze's function satisfies these criteria and is computationally convenient. The result is

$$E\{K\} = K_1\,|g_0(\theta, \phi)|^2\, e^{-\sigma^2} + K_1\pi c^2\phi_0 S\left(\frac{kc\,\sin\theta}{2}, \sigma^2\right) \tag{17a}$$

where

$$\phi_0 = \int_0^a \int_0^{2\pi} [f_0(r', \phi')]^2 \, r' \, dr' \, d\phi' \tag{17b}$$

and

$$S(x, \sigma^2) = e^{-\sigma^2} \sum_{n=1}^{\infty} \frac{(\sigma^2)^n}{n(n!)} e^{-x/n} \quad . \tag{17c}$$

The first (diffraction) term on the right side of equation (17a) is the desired power pattern reduced by the exponential factor. Equation (15) shows that this first term is $(\sqrt{K_1}\,|\bar{g}|)^2$. Equation (17b) then indicates that the second (scattering) term is $K_1\sigma_R^2$, which represents the noise level produced by the random phase error. As σ^2 increases, peak power is reduced [due to the $\exp(-\sigma^2)$ in the first term], and this power enters the scattered sidelobes. Figure 2 is a plot of the summation term in equation (17c). The argument is $x = [0.5(ka \sin \theta)\ (c/a)]$; hence, for a small ratio of correlation distance to radius (c/a), x will be small compared to ka. The low-slope region, $x < 1$, shown in Figure 2 will cause the scattering term to dominate the diffraction term as illustrated more clearly in the following example.

A numerical example: Uniform illumination

For illustration, a uniform aperture illumination of $f_0 = E_0$ is assumed. From equation (16) the well-known result is obtained:

$$|g_0| = 2\pi a^2 \left| \frac{J_1(u)}{u} \right| E_0, \; u = ka \sin \theta \quad . \tag{18}$$

From equation (15),

$$|\bar{g}| = 2\pi a^2 \left| \frac{J_1(u)}{u} \right| E_0 e^{-\sigma^2/2} \tag{19}$$

and, from equation (17),

$$E\{gg^*\} = \pi^2 a^4 \left[2 \frac{J_1(u)}{u} \right]^2 E_0^2 e^{-\sigma^2} + \pi^2 a^4 \left(\frac{c}{a} \right)^2 \left[\sum_{n=1}^{\infty} \frac{(\sigma^2)^n}{n(n!)} e^{-(cu/2a)/n} \right] E_0^2 e^{-\sigma^2} \quad . \tag{20}$$

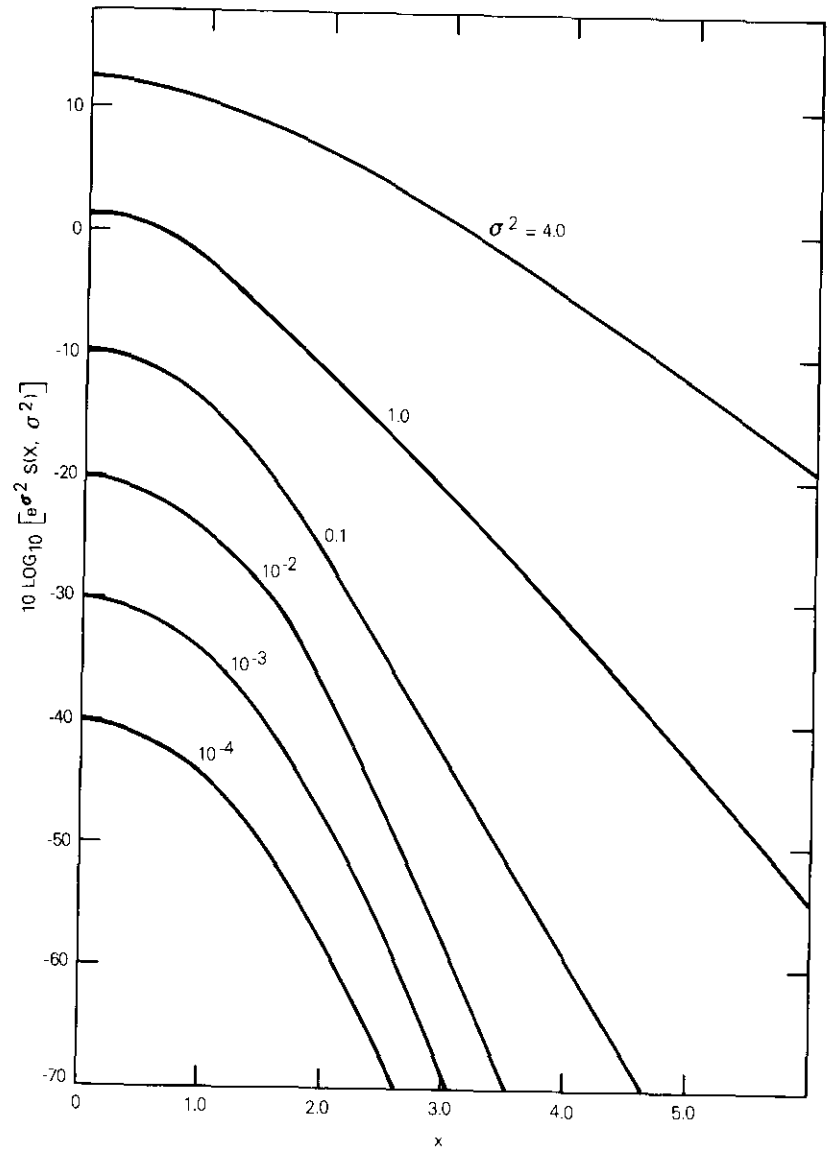
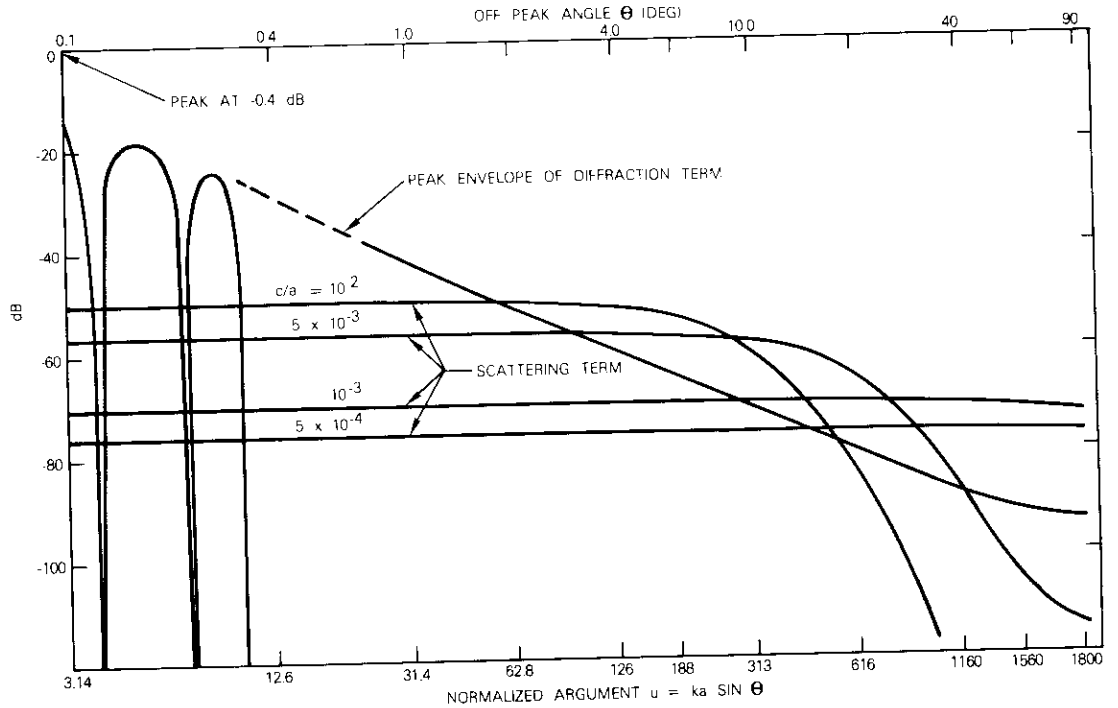
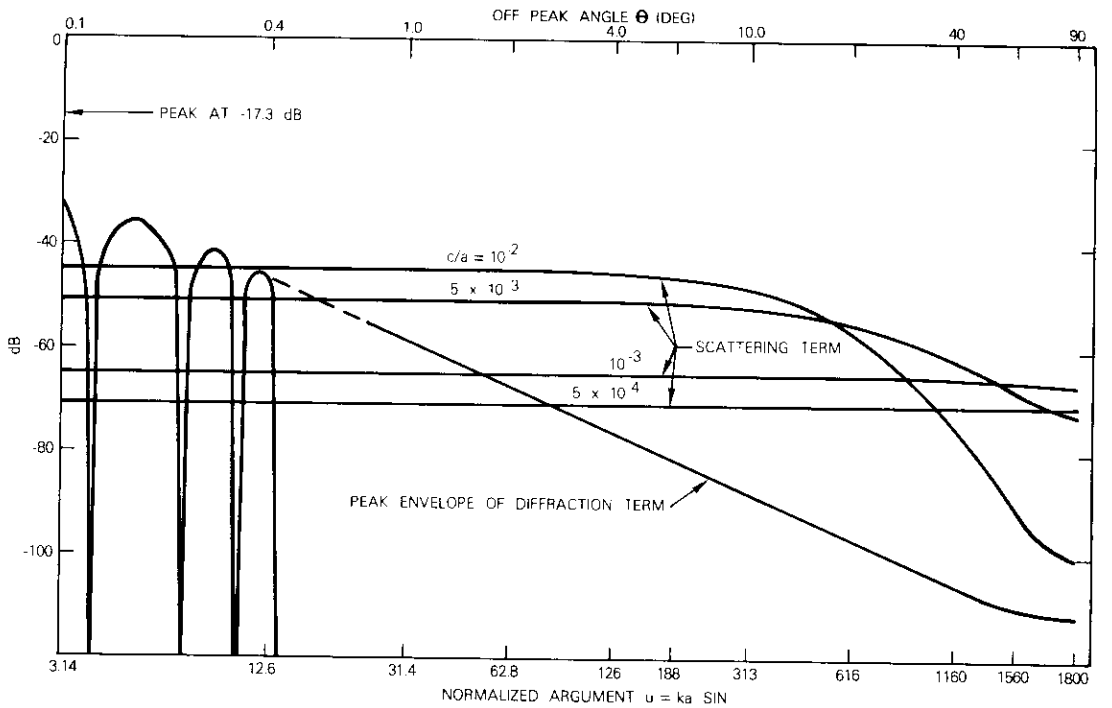


Figure 2. *Normalized Summation*

Figures 3a and 3b show the normalized $E\{gg^*\}/\pi^2a^4E_0^2$ plotted against u for two values of σ^2 (0.1 and 4.0, respectively) and several values of c/a.



a. $ka = 1800, \sigma^2 = 0.1$



b. $ka = 1800, \sigma^2 = 4.0$

Figure 3. *Diffraction and Scattering Terms of $E\{gg^*\}$ vs Angle*

The first three peaks of the diffraction term and the envelope of the remaining peaks are indicated. Main beam normalized gain without any phase error ($\sigma^2 = 0$) is at the 0-dB level. When the scattering term level is above that of the diffraction term, the scattering term predominates and determines the power pattern. In any case, the scattering term sets the level of the nulls. The upper scale in each figure exhibits an off-peak angle θ for $ka = 1800$, which represents a 30-m diameter aperture at 6 GHz. For a normal distribution of random phase, a value of $\sigma^2 = 0.1$ implies that 68 percent of the random phase angles are in the $\pm 18°$ range. As indicated in Figure 3a, peak gain is therefore reduced by 0.4 dB. The value $\sigma^2 = 4.0$ indicates that 68 percent are within $\pm 115°$ and a reduction of peak gain by 17.3 dB. The latter is a clearly unrealistic value.

These results may be stated in terms of the PDF of equation (11). Substituting equations (19) and (20) into equation (4c) shows that the variance of g and σ_g^2 is given by the scattering term of equation (20). The parameter p [equation (10b)] is the ratio of diffraction radiation intensity to scattered radiation intensity, while the normalized variable t [equation (10a)] is the ratio of radiation intensity to scattered radiation intensity. For a small off-axis angle, p is therefore large ($p \gg 1$), and the probability that t is close to the value of p is high. For large p, the PDF of t peaks sharply at $t = p$, and the total radiation intensity has a high probability of being close to the diffraction value. However, for large off-axis angles, p is much less than 1, and the probability that t lies between about 0.1 and 2.5 is high. The radiation intensity is likely to assume values between 0.1 to 2.5 times the diffraction value (10 dB less to 4 dB greater than the diffraction gain).

Interference from multiple sources

General statistical relations

A particular receiver may be subject to interference from N independent sources. Let G_i be the power of the ith source and let its PDF be $W_{G_i}(G_i)$. Then the total interfering power is

$$G = G_1 + G_2 + \ldots + G_N \quad . \tag{21}$$

The N interfering sources are statistically independent; hence, the joint PDF is the product of the N individual PDFs. The PDF of the total interfering power G, $W_G(G)$, is given by

$$W_G(G) = \int_0^G dy_2 W_{G_1}(G - y_2) \int_0^{y_2} dy_3 W_{G_2}(y_2 - y_3) \ldots \int_0^{y_{N-1}} dy_N \left[W_{G_{N-1}}(y_{N-1} - y_N)\, W_{G_N}(y_N) \right] \quad . \tag{22}$$

The probability that the total interfering power G from N sources will be greater than some level G_p is

$$Pr_N(G > G_p) = \int_{G_p}^{\infty} W_G(G)\, dG \quad . \tag{23}$$

Relations for an exponential PDF

Each independent source is assumed to have an exponential PDF as in equation (7). For the ith source,

$$W_{G_i}(G_i) = \frac{1}{\gamma_i} e^{-(G_i/\gamma_i)}, \qquad G_i > 0 \quad . \tag{24}$$

In equation (24), γ_i is the mean value of G_i, which is generally a function of off-axis angle θ. The median value of G_i is $\gamma_i \ln 2 \doteq 0.69\gamma_i$. Applying the general statistical relations, equations (22) and (23), to the exponential PDF gives

$$Pr_N(G > G_p) = \frac{\gamma_1^{(N-1)}\, e^{-G_p/\gamma_1}}{\prod_{i=2}^{N} (\gamma_1 - \gamma_i)} + \frac{\gamma_2^{(N-1)}\, e^{-G_p/\gamma_0}}{\prod_{\substack{i=1 \\ i \neq 2}}^{N} (\gamma_2 - \gamma_i)} + \ldots + \frac{\gamma_N^{(N-1)}\, e^{-G_p/\gamma_N}}{\prod_{i=1}^{N-1} (\gamma_N - \gamma_i)} \quad . \tag{25}$$

Equation (25) must be used with caution when $\gamma_i = \gamma_j$. In particular, if all the γ_i are equal, equation (25) then becomes

$$Pr_N(G > G_p) = \left\{ \frac{1}{(N-1)!} \left(\frac{G_p}{\gamma}\right)^{N-1} + \frac{1}{(N-2)!} \left(\frac{G_p}{\gamma}\right)^{N-2} + \ldots + 1 \right\} e^{-G_p/\gamma}, \quad \gamma_i = \gamma,\ i = 1, 2, \ldots N. \tag{26}$$

Figure 4 shows $Pr_N(G > G_p)$ versus (G_p/γ) for $N = 1, 2, \ldots, 6$.

For a single source $N = 1$,

$$Pr_1(G > G_p) = e^{-G_p/\gamma} \quad . \tag{27}$$

With two interfering sources ($N = 2$), let the ratio of the mean power levels be $\gamma_1/\gamma_2 = k$ and let $x = G_p/\gamma_1$. Then

$$Pr_2(G > G_p) = \begin{cases} \dfrac{ke^{-x} - e^{-kx}}{k - 1}, & k \neq 1 \\ (x + 1)\, e^{-x}, & k = 1 \end{cases} \tag{28a}$$

$$k = \frac{\gamma_1}{\gamma_2}, \quad x = \frac{G_p}{\gamma_1} \quad . \tag{28b}$$

Figure 5 shows $Pr_2(G > G_p)$ versus x for several values of k. The curve for $k = 1$ in Figure 5 and the curve for $N = 2$ in Figure 4 are the same. Note that as $k \to \infty$, Pr_2 converges to Pr_1, *i.e.*, a single source exponentially distributed.

Interference to carrier (I/C) ratio

Normalization of the distributions

For the calculation of interference power levels or I/C ratios, it is necessary to choose the parameter γ of the exponential distribution. An expression for γ has been presented previously as the scattering term in equation (17a); however, there are undetermined parameters, and the evaluation depends on the aperture illumination. In this subsection, the value of the distribution parameter is selected to yield a 10-percent probability that the gain of a single interfering source will exceed the level G_p when G_p is given by the C.C.I.R. standard envelope. Thus, it is assumed that

$$Pr_1(G > G_p) = 0.1 \tag{29}$$

when $10 \log G_p = 32 - 25 \log \theta$.

The 10 percent for the standard envelope refers to the peak value while equation (29) refers to all values of gain. Also, the C.C.I.R. Study Group Document [2] estimated that 10 percent of the peaks conformed to $31.8 - 23.5 \log \theta$ rather than to the standard envelope. The general nature of the results will not be affected by differences between these values.

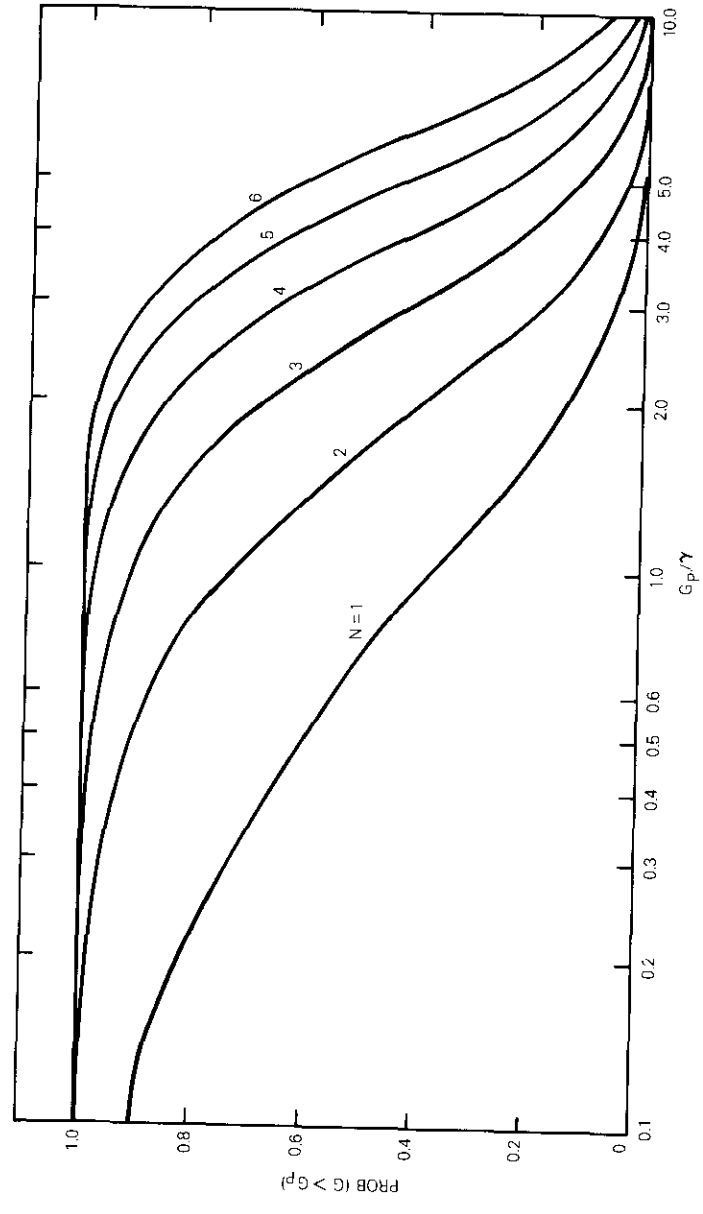


Figure 4. *Probability That $G > G_p$ vs G_p/γ for N Independent Sources, Each Exponentially Distributed with Mean γ*

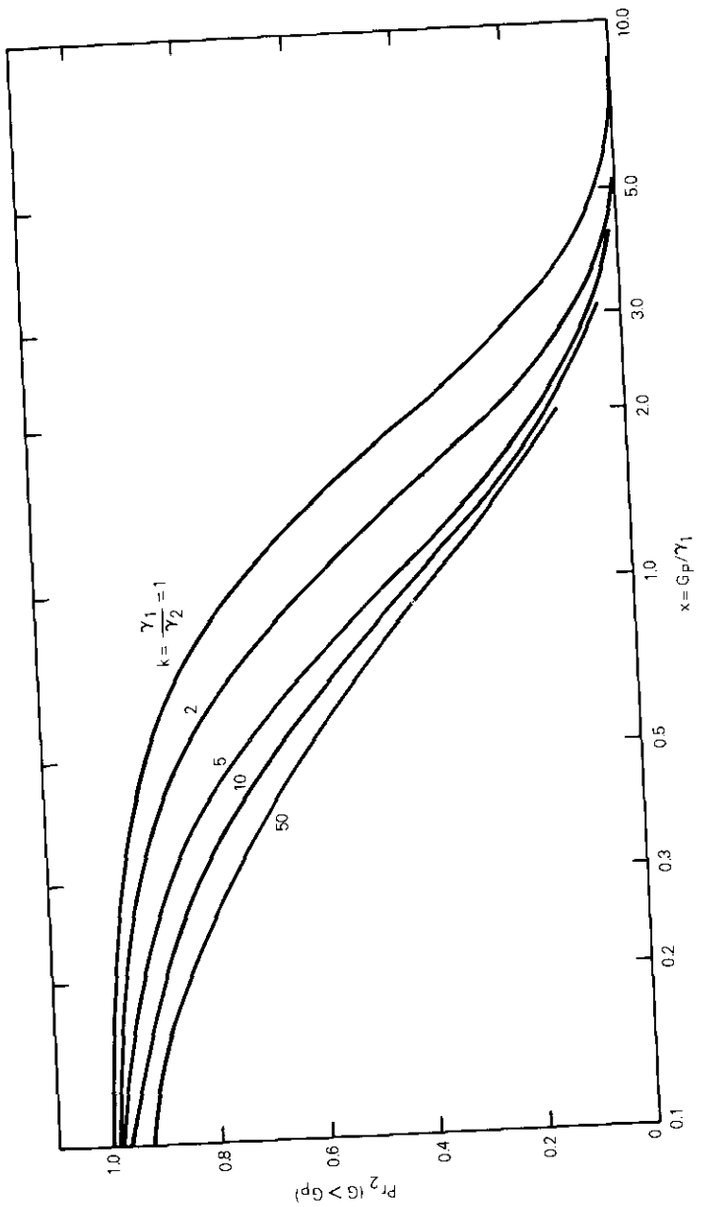


Figure 5. *Probability That $G > G_p$ for Two Exponentially Distributed Sources With Means γ_1 and γ_2*

From equation (27), $Pr_1 = 0.1$ when $G_p/\gamma = -\ln 0.1$. Hence, from equation (29)

$$\gamma = \frac{10^{3.2}}{-ln\ 0.1}\,\theta^{-2.5} \doteq 689\theta^{-2.5} \tag{30}$$

for the exponential PDF.

Interference examples

Two examples are presented to illustrate the calculation of the I/C ratio. These are simplified by assuming that most of the factors which comprise the link relationship are constant from one source to the other. It is thus assumed that the transmit powers from sources and path lengths are the same for the desired and all interfering sources, and that the satellite antennas are broad beam for identical gain toward all earth stations.

In the first example there is one desired signal and N interfering signals, each spaced at angle θ. Figure 6a illustrates the up- and down-links for two interfering sources. The I/C ratio at either satellite or earth station is then

$$\frac{I}{C} = \frac{G(\theta)}{G_0}$$

where G_0 is the peak antenna gain and $G(\theta)$ is the gain in the direction θ. The probability that I/C exceeds some particular level $(I/C)_p$ is

$$Pr[(I/C) > (I/C)_p] = Pr(G > G_p) \tag{31}$$

where

$$G_p = (I/C)_p\, G_0 \tag{32a}$$

or

$$G_p = (I/C)_p + G_0\ \text{dB} \quad . \tag{32b}$$

In the last expression, G_p and G_0 are expressed in dBi. The $Pr(G > G_p)$ is given by equation (26) with γ given by equation (30). The solid lines in Figure 7 plot $Pr[(I/C) > (I/C)_p]$ on a logarithmic scale versus $(I/C)_p G_0/\gamma$ in *dB* for $N = 1, 2, \ldots, 5$ interfering sources. For example, if there is one

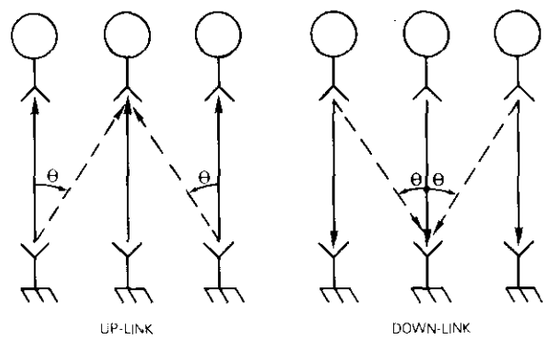


Figure 6a. *Two Interfering Sources at Equal Angles*

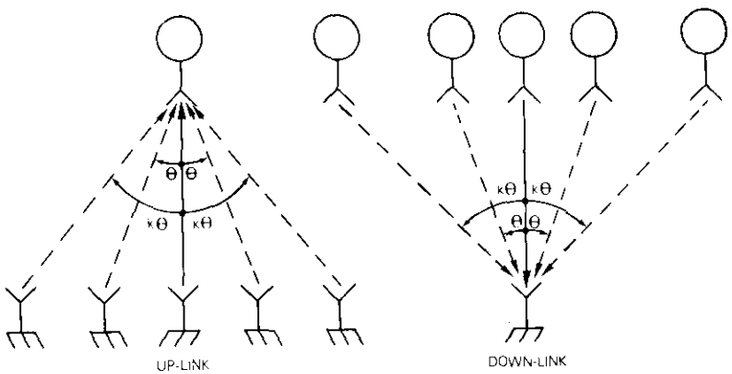


Figure 6b. *Two Interfering Sources at Angle θ and Two at kθ*

interfering source with antenna gain $G_0 = 65$ dBi and with $\theta = 6°$ spacing, then the I/C ratio exceeds -50 dB with probability 0.017 (point A, Figure 7). Two interfering sources with $G_0 = 65$ dBi at 6° spacing have an $I/C > -50$ dB with probability 0.083 (point B). By increasing the spacing to 7.1°, the Pr $(I/C > -50$ dB) for the two interfering sources can be reduced to that for one interfering source (point C).

A second example (Figure 6b) has one desired source, two interfering sources at angle θ, and two additional interference sources at angle $k\theta$. The $Pr(G > G_p)$ is given by equation (25) with $N = 4$, $\gamma_1 = \gamma_2 = 689\theta^{-2.5}$, and $\gamma_3 = \gamma_4 = 689\ (k\theta)^{-2.5}$; the evaluation must be carefully conducted because of the equal γ's. The results are plotted as the dashed lines in

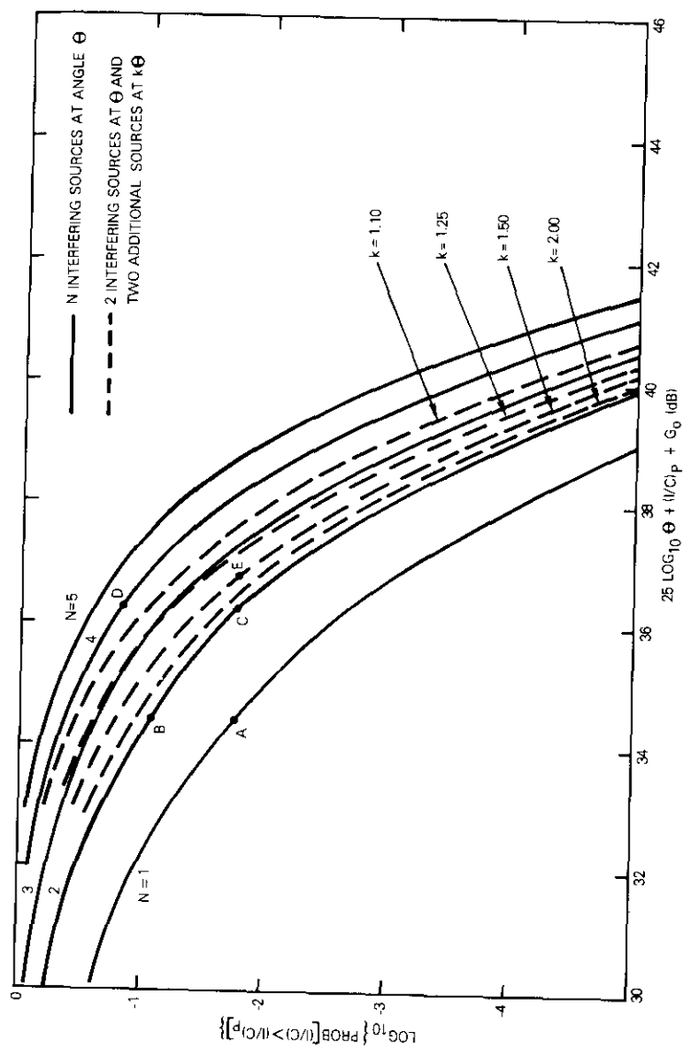


Figure 7. *Probability That the I/C Ratio is Greater Than $(I/C)_P$*

Figure 7. With $G_0 = 65$ dBi and $\theta = 7.1°$, the $Pr(I/C > -50\ dB)$ increases from 0.017 (point C) to 0.155 (point D) as k decreases from infinity to unity. The two interfering sources at $k\theta$ have only a small effect on the probable I/C ratio if k is greater than about 1.5. The $P_r(I/C > -50$ dB) is equal to 0.017 for four sources, two sources at 7.4° and two at 11.2° (point E).

General observations

Curves, such as those in Figure 7, can be constructed using the closed-form expression for the exponential PDF for any number of interfering sources at any angular arrangement. For other distributions, they can be drawn for any angle for two sources generally by using numerical integration; more sources would probably be cumbersome. The curves depend upon the parameters. For the exponential PDF, the mean and median are γ and 0.69γ, respectively. The distribution has only a single parameter and hence only one arbitrary level can be selected. However, if this distribution correctly represents the physical situation, then establishing one level should correctly set all others.

Data collection

Whether the power gain has an exponential PDF or some other PDF significantly affects the conclusions concerning probable I/C ratios. There is no *a priori* reason which indicates that the PDF should not be a function of angle. (This is the direct theoretical result.) The general statistical relationships and normalization techniques discussed previously are general, but the specific results depend on the actual distributions.

Therefore the use of antenna patterns, which are measured to accommodate the extraction of the data required and are sufficiently numerous for different classes of antennas, is suggested so that the results can be obtained with reasonable confidence levels. Antenna manufacturers and users routinely perform these measurements, but the patterns must be expanded in resolution to supply meaningful data. In the gain ordinate, this implies that the main beam will saturate the system or, equivalently, that the power level should be increased by a known 20 or 30 dB after the pattern is outside the main beam and the first one or two sidelobes. In the angular abscissa, the pattern must be expanded (increased chart speed or reduced skew rate) so that individual sidelobes are well resolved (perhaps of the order of 4 to 5 cm per sidelobe angular width). Antenna size and type, frequency, angle, and power levels must be clearly documented.

If an equivalence between frequency and angle expressed by $f \sin \theta$ is assumed, then data can be accumulated from different antennas at small regions about many different $f \sin \theta$ products. It is important to note that grouping together populations of data points from different $f \sin \theta$ ranges is not generally valid, even if the means or medians of the data within each range are adjusted to be the same because the PDF changes with angle. The PDF of the resultant overall population will be different from that of any of the individual populations. The statistical process is not stationary in angle: angle averages and ensemble averages are not equivalent.

Measurement Program

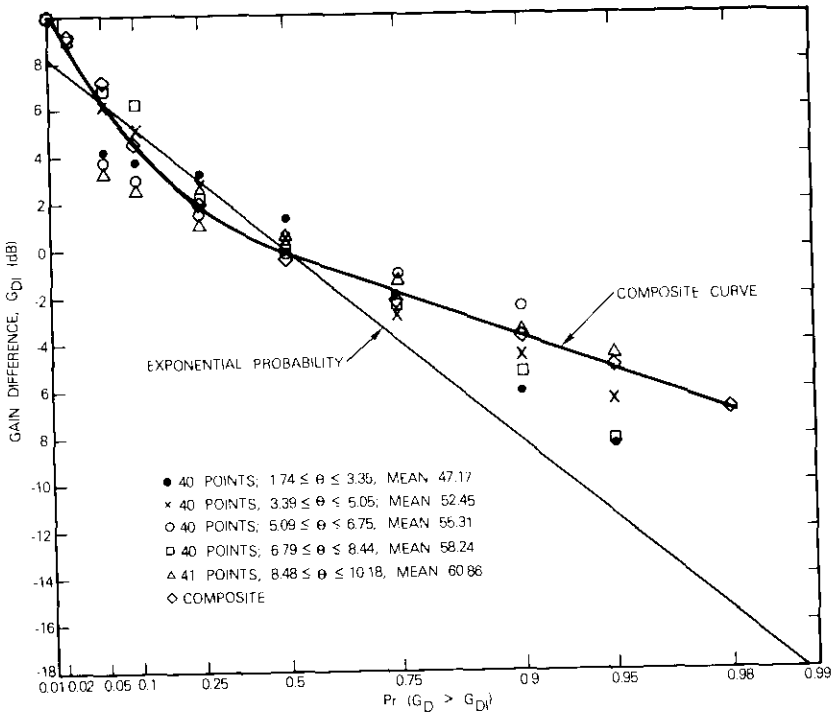
From April 10 to April 14, 1978, pattern measurements were performed at up- and down-link frequencies on the PAM-1 antenna located at Paumalu, Hawaii. Patterns were recorded with as large a dynamic range below beam peak as possible, over a wide angular range from the peak, and with an angular scale sufficiently large to resolve individual peaks and nulls. All patterns recorded involve rotation of the antenna in azimuth at an elevation of about 32°. Table 1 summarizes the recorded patterns. Each pattern was taken by slewing the PAM-1 antenna in azimuth across the INTELSAT IV-A F-6 satellite which functioned as a relay to the TT&C antenna at Paumalu.

Sidelobe statistics

For pattern 1, points were read every 0.05° from 253.00° to 263.00° azimuth (201 points) and transferred to a computer file. The angles were then corrected to true off-peak angles by

$$\theta_{\text{true}} = 2 \arcsin \left\{ \cos (\text{Elev}) \sin \left[\frac{az - \text{peak } az}{2} \right] \right\} \tag{33}$$

resulting in a range from 1.70° to 10.18° off peak. The points have been separated into various groups of consecutive points and ordered in decreasing gain about the mean value of each group. The probability distribution (PD), $Pr(G > G_p)$, for these groups is plotted in Figure 8a; the dots, crosses, circles, squares, and triangles correspond to the PD in angular ranges shown in the figure. Only the triangular points (for the angles farthest from the beam) can be described as forming a straight line. The composite curve is obtained by taking the five groups of 40 points with the mean value subtracted and combining them in a single set of data.



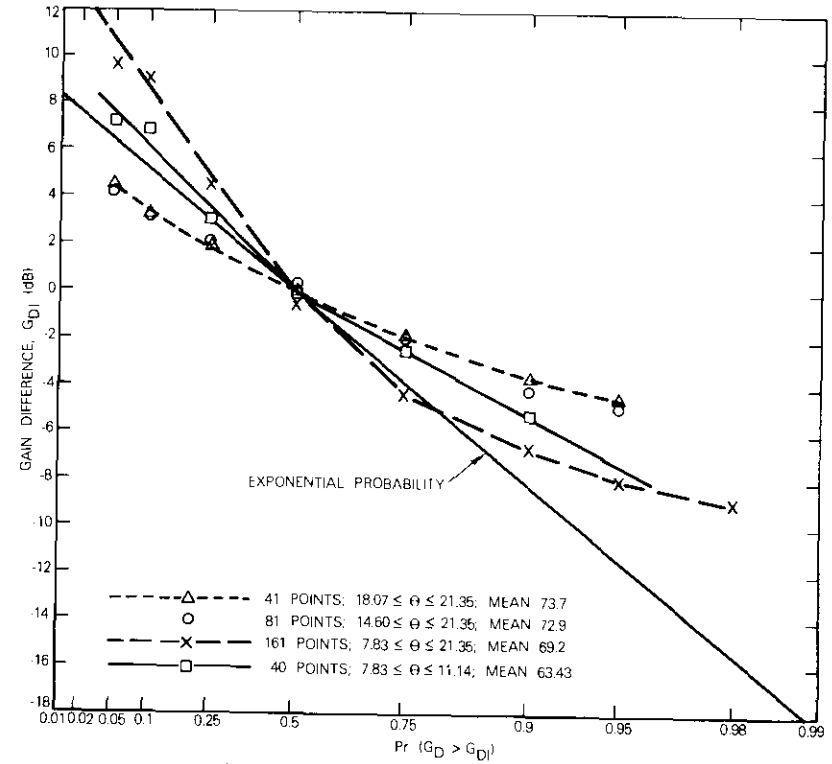
a. *Pattern No. 1, f = 5,950 MHz*
Figure 8. *Experimental PD*

All of the curves fall above the exponential curve for probabilities greater than 50 percent, indicating that the pattern is less likely to have low values of gain than would be predicted by the exponential PDF. This may be attributable to the low recorded signal level corresponding to noise level.

Figure 8b is similar for pattern number 9 at 6,205 MHz. These points were spaced 0.1° farther from the peak (7.83° to 21.35°). The crosses and triangles in Figure 6b represent the groups farthest from the peak and most closely approximate a straight line. Again, all the curves show fewer low values of gain than the exponential curve predicts.

Angle-frequency equivalence

The far field pattern of an antenna is expressed in terms of a variable which is proportional to frequency and to a trigonometric function of the spherical angular coordinates θ and ϕ of the far field point. For a circular aperture with illumination which is independent of angle about the center,



b. *Pattern No. 9, f = 6,205 MHz*
Figure 8. *Experimental PD*

the variable is $u = ka \sin\theta = (2\pi a/c)(f \sin\theta)$. Hence, f and $\sin\theta$ appear as a product to determine u: As f decreases, the pattern should broaden or spread away from the axis. A null or peak which occurs at θ_1 at f_1 should appear at θ_2 at f_2, where

$$f_1 \sin\theta_1 = f_2 \sin\theta_2 \quad . \tag{34}$$

To verify this relationship with the patterns recorded (Table 1), the following procedure was used. An examination of the patterns revealed that certain nulls had a recognizable or characteristic shape. Nine such nulls were chosen and labeled 1 to 9, with null 1 closest to the main beam (but not the first null) and null 9 farthest from the main beam (but many more than 9 nulls away). For each pattern except 3 and 15, these nulls were located and the azimuth angle recorded. Data from pattern 7 were

TABLE 1. SUMMARY OF TEST PATTERNS OF THE PAM 1 ANTENNA IN PAUMALU, HAWAII, FROM APRIL 10–14, 1978

Pattern No.	Freq (MHz)	Beam Az (deg)	Peak Elev (deg)	Final Az (deg)	Az Range (dB)	Dynamic Range (dB)	Approx Chart Speed (cm/deg)	
1	5950	251.00	31.92	278	27	70+	14	
2	5955	250.00	31.92	278	27	70+	14	
3	5960	250.88	31.92	259	8	70	14	FREQ LOCK LOST RESTART
4	5960	250.85	31.92	278	27	70	14	
5	5965	250.80	31.92	278	27	70	14	
6	5940	251	32.07	283	32	80	7	264° – 268.5°
7	6185	251.01	32.07	306	54	80	7	FREQ LOCK LOST
8	6195	250.97	32.01	288	38	80	7	
9	6205	250.76	32.01	307	57	80	7	
10	6215	250.73	31.99	297	47	80	7	
11	6345	250.68	31.98	288	38	80	7	
12	6355	250.62	31.96	287	37	80	7	
13	6365	250.65	31.94	282	32	80	7	
14	6375	250.67	31.88	287	37	80	5	
15	6375	250.67	31.88	370	120	80	0.4	FOR ILLUSTRATION PURPOSES
16	3710	250.61	31.79	301	51	70	4.7	
17	3740	250.60	31.78	290	40	70	4.7	
18	3960	250.65	31.78	280	30	70	4.9	
19	3990	250.64	31.77	280	30	70	4.9	
20	4120	250.66	31.77	274	24	70	4.9	
21	4150	250.68	31.77	270	20	70	4.9	

discarded because the recorded angles were not correct. The angles were corrected [using equation (37)], and the frequency sin θ product computed. These are displayed in Figure 9, where frequency times sin θ for each null is the ordinate and frequency (not to scale) is the abscissa. The range of values in each transponder is shown. Null 1 is at about 0.37°, and null 9 is at about 10° at 4,120 MHz. The curves are essentially flat over the entire band from 3,700 MHz to 6,400 MHz for the nulls confirming the $f \sin \theta$ product relationship.

Conclusions

A theoretical basis for assuming an exponential distribution for the PDF of the sidelobe power level has been established. This theory indicates that the power level at a given angle for many nominally identical antennas is exponentially distributed. The distribution of power levels has been compared for a single antenna over a range in angles. Although the results are not particularly supportive, they do not contradict the theory. Data on several (or many) generally similar antennas should be recorded to verify the theory or to empirically determine a different distribution as a better fit to reality.

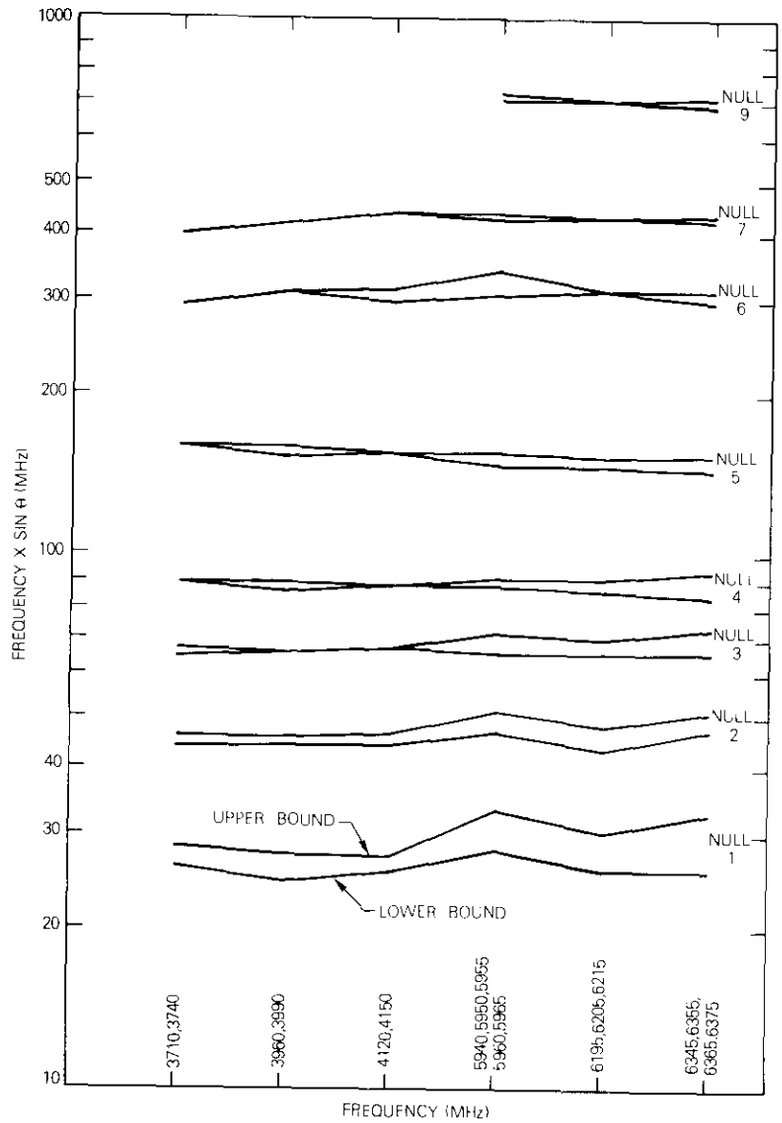


Figure 9. *Freq-Sin θ Product for Selected Nulls at Frequencies Across the Up- and Down-Bands*

Statistical relationships for finding the probability of exceeding an I/C level from multiple interfering sources have been developed and applied to the specific case of an exponential PDF. Interference-to-carrier ratios have been calculated for several examples. This theory should provide a tool for system planners to estimate the likelihood of multiple interfering sources causing error producing noise-to-carrier ratios.

Acknowledgments

The author wishes to thank D. F. DiFonzo and A. E. Atia for their helpful comments and discussions, as well as R. W. Gruner and K. K. Yamashita (COMSAT Earth Station, Paumalu, Hawaii) for their help in performing the measurements.

References

[1] C.C.I.R. XIIIth Plenary Assembly, Geneva, 1971, Recommendation 465-1.
[2] C.C.I.R. Study Group, Period 1974–1976, December 4/32-E, 19 January 1976.
[3] S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vol. 23 and 24; Nelson Wax, ed., *Noise and Stochastic Processes*, New York: Dover, 1954.
[4] B. D. Steinberg, *Principles of Aperture and Array System Design*, New York: Wiley, 1976, pp. 147–148.
[5] M. C. Jeruchim, "A Statistical Approach to Satellite Interference Levels," *ICC'78 Conference Record*, Vol. 3, 1978, pp. 35, 3.1–35, 3.4.
[6] S. Silver, *Microwave Antenna Theory and Design*, New York: McGraw Hill, 1949, p. 173.
[7] J. Ruze, "Antenna Tolerance Theory—A Review," *Proc. IEEE*, Vol. 54, No. 4, April 1966, pp. 633–640.

Appendix. Derivation of sidelobe statistical distribution

This appendix derives the PDF for the antenna field amplitude. The normalized field pattern for a circular aperture is given by equation (1), and the radiation intensity is given in equation (2). The normalized field pattern can be expressed as

$$g(\theta, \phi) = a(\theta, \phi) + jb(\theta, \phi) \tag{A-1}$$

where a and b are the quadrature real and imaginary components.

The illumination can be written as in equation (3), where each function on the right side of the equation is real. The real part a of equation (1) is

$$a = \iint_S [f_0 + f] \cos [kr' \sin \theta \cos (\phi - \phi') + \psi_0 + \psi] \, dS'$$

$$\cong \sum_n [f_0(r_n, \phi_n) + f(r_n, \phi_n)] \cos [kr_n \sin \theta \cos (\phi - \phi_n)$$

$$+ \psi_0(r_n, \phi_n) + \psi(r_n, \phi_n)] \, \Delta S_n \tag{A-2}$$

where the surface integral has been approximated by a summation over a large number of elementary areas ΔS_n. Regardless of the elementary nature of ΔS_n, each ΔS_n is assumed sufficiently large (compared to any correlation distance which may be present) to consider that each random variable $f(r_n, \phi_n)$ and $\psi(r_n, \phi_n)$ is independent. Each term is therefore an independent random variable. Furthermore, because each of the random amplitudes must be small and because the maximum magnitude of the cosine is unity, it is reasonable to assume that no single term or small group of terms can be so large as to dominate the sum. According to the central limit theorem [A-1], [A-2], the PDF of a is a normal (Gaussian) distribution. Thus, the PDF of a, W_a, is given by

$$W_a(a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp - \frac{(a - \mu_a)}{2\sigma_a^2} \tag{A-3}$$

where $\mu_a = \bar{a}$ is the mean of a and $\sigma_a^2 = \overline{a^2} - \bar{a}^2$ is the variance of a. An identical argument holds for the imaginary part b. Also, the parameters μ_b and σ_b of the normal PDF $W_b(b)$ are equal to μ_a and σ_a, respectively.

The mean of the normalized field is

$$\bar{g} = \bar{a} + j\bar{b} = \mu_a + j\mu_b \tag{A-4}$$

and the squared magnitude of the mean is

$$|\bar{g}|^2 = \mu_a^2 + \mu_b^2 \quad . \tag{A-5}$$

The value of this term is related to the illumination f_0 and the PDF of the random phase [see equations (15) and (16)] and amplitude error.

The variance of g is given by

$$\sigma_g^2 = \overline{gg^*} - \bar{g}\,\bar{g}^* = \overline{(a^2 - \bar{a}^2)} + \overline{(b^2 - \bar{b}^2)} = \sigma_a^2 + \sigma_b^2$$
$$\sigma_a^2 = \sigma_b^2 = \tfrac{1}{2}\sigma_g^2 \quad . \tag{A-6}$$

It may be shown that a and b, the quadrature components of f, are independent [A-3]. Hence, the joint PDF, $W_{a,b}$, of a and b is

$$W_{a,b}(a, b) = W_a(a)\, W_b(b) = \frac{1}{\pi\sigma_g^2} \exp\left[\frac{-(a - \mu_a)^2 + (b - \mu_b)^2}{\sigma_g^2}\right] \tag{A-7}$$

where equations (A-3) and (A-6) have been used. The quadrature components a and b are transformed to polar components $|g|$ and α, where

$$a = |g| \cos\alpha \qquad \text{and} \qquad b = |g| \sin\alpha \quad . \tag{A-8}$$

The joint PDF of b and α is determined by the transformation

$$W_{|g|,\alpha}(|g|, \alpha) = W_{a,b}(|g| \cos\alpha, |g| \sin\alpha)\, |\mathrm{J}|$$

where

$$\mathrm{J} = \begin{vmatrix} \dfrac{\partial a}{\partial |g|} & \dfrac{\partial a}{\partial \alpha} \\ \dfrac{\partial b}{\partial |g|} & \dfrac{\partial b}{\partial \alpha} \end{vmatrix} \quad .$$

Substituting equation (A-8) into $|\mathrm{J}|$ yields $|\mathrm{J}| = |g|$ so that the joint PDF is given by

$$W_{|g|,\alpha} = \frac{|g|}{\pi\sigma_g^2} \exp\left[-\frac{|g|^2 + |\bar{g}|^2}{\sigma_g^2}\right] \cdot \exp\left[\frac{2|g|}{\sigma_g^2}(\mu_a \cos\alpha + \mu_b \sin\alpha)\right] \quad . \tag{A-9}$$

The PDF $W_{|g|}(|g|)$ is found from the joint PDF by integrating α over its complete range

$$W_{|g|}(|g|) = \int_0^{2\pi} W_{|g|,\alpha}(|g|, \alpha)\, d\alpha \quad .$$

The argument of the right most exponential is rewritten

$$\frac{2|g|}{\sigma_g^2}(\mu_a \cos\alpha + \mu_b \sin\alpha) = 2\,\frac{|g|\sqrt{\mu_a^2 + \mu_b^2}}{\sigma_g^2} \sin(\alpha + \alpha_1)$$
$$= \frac{2|g|\,|\bar{g}|}{\sigma_g^2} \sin\alpha'$$

where $\alpha_1 = \arctan(\mu_a/\mu_b)$, $\alpha' = \alpha + \alpha_1$ and equation (A-5) has been substituted. Since [A-4]

$$\int_0^{2\pi} \exp(t \sin V)\, dV = 2\pi I_0(t)$$

where I_0 is the modified Bessel function of the first kind and zero order, then

$$W_{|g|}(|g|) = \frac{2|g|}{\sigma_g^2} \exp\left[-\frac{|g|^2}{\sigma_g^2}\right] \exp\left[-\frac{|\bar{g}|^2}{\sigma_g^2}\right] I_0\left(\frac{2|g|\,|\bar{g}|}{\sigma_g^2}\right) \tag{A-10}$$

which is the PDF of the normalized field pattern amplitude.

This can be simplified if

$$|\bar{g}| \ll \sigma_g \tag{A-11}$$

and hence $\exp(-|\bar{g}|^2/\sigma_g^2) \cong 1$ and $I_0 \cong 1$, at least for values of $|g|$ less than or similar to σ_g. If equation (A-11) is applicable but $|g| \gg \sigma_g$, then the asymptotic expression for I_0 increases as

$$\frac{\sigma_g}{2}\sqrt{\frac{1}{\pi|\bar{g}|\,|g|}} \exp\left[\frac{2|g|\,|g|}{\sigma_g^2}\right] \quad .$$

However, the multiplicative term $\exp[-g^2/\sigma_g^2]$ dominates and $W_g \to 0$. Thus, when equation (A-11) is satisfied, equation (A-10) reduces to the Rayleigh distribution

$$W_{|g|}(|g|) = \frac{2|g|}{\sigma_g^2} \exp\left(-\frac{|\bar{g}|^2}{\sigma_g^2}\right) \tag{A-12}$$

for the field amplitude.

The PDF equation (A-10) is called the "non-central Rayleigh" or the "Rician" density function and its complementary cumulative distribution

$$Q(|\bar{g}|, g_1) = \int_{g_1}^{\infty} W_{|g|}(|g|)d(|g|)$$

is known as the "Q-function" [A-5].

References

[A-1] G. J. Hahn and S.S. Shapiro, *Statistical Models in Engineering*, New York: Wiley, 1967, pp. 72–73.

[A-2] N. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3d ed., New York: Wiley, 1968, p. 254.

[A-3] D. Middleton, *An Introduction to Statistical Communications Theory*, New York: McGraw Hill, 1960, p. 399.

[A-4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, 1965, p. 376.

[A-5] C. W. Helstrom, *Statistical Theory of Signal Detection*, 2nd ed., New York: Pergamon Press, 1968, pp. 449–453.

Paul R. Karmel received a B.E. (E.E.) from Cornell University in 1956, an M.S. (E.E.) from the Massachusetts Institute of Technology in 1956, and a D. Engr. Sc. from Columbia University in 1964. Since 1964, he has been a professor in the Department of Electrical Engineering at The City College of the City University of New York, where he has taught undergraduate and graduate courses in electromagnetics and microwave engineering. He also served as Acting Dean of the School of Engineering for two years. In September 1977, he joined the Antenna Department of the Microwave Laboratory at COMSAT *Laboratories, while on leave from The City College.*

Index: microwave integrated circuit, converter, down-converter, field effect transistor

Design and performance of low-cost integrated MIC up- and down-converters for earth station applications

R. STEGENS

(Manuscript received November 14, 1978)

Abstract

This paper describes the design and performance of low-cost microwave integrated circuit (MIC) up- and down-converters developed for earth station applications. The down-converter is an all-MIC double conversion 11.7- to 12.2-GHz receiver realized in a 28-cm × 13-cm × 2.5-cm enclosure. Overall gain is 68 dB, with a noise figure of less than 4.0 dB. Any one of 10 channels, each of 43-MHz bandwidth, may be received by properly selecting the second local oscillator (LO) input frequency. All of the MIC components within the down-converter are designed for both low cost and high performance; these include the field effect transistor (FET) LO and RF amplifiers on fused silica, a 12/1-GHz image enhanced mixer, 1-GHz IF filters and amplifiers, and a 1-GHz/70-MHz IF processor.

The up-converter translates a 43-MHz baseband centered at 70 MHz to a 1-GHz IF, eliminates the image and LO signals, amplifies, up-converts to any one of 10 channels located in the 14.0- to 14.5-GHz band, and produces an output level (+5 dBm) sufficient to drive an earth station traveling wave tube amplifier (TWTA). At this output level, the linearity is excellent with a two-tone carrier-to-intermodulation (C/I) ratio exceeding 30 dB, while the phase shift is less than 1°.

The success of the converter designs represents the culmination of extensive effort directed toward the development of high-performance MIC circuits, including a new class of filters; microstrip interdigital filters (MIDFS) allow realization of compact multisection high-Q bandpass structures near 1 GHz in a low-cost printed form.

Introduction

Satellite communications systems operating in the 11.7- to 12.2-GHz (down-link) and 14.0- to 14.5-GHz (up-link) frequency bands will emphasize the use of low-cost earth terminal components which achieve high levels of performance and reliability. Two of the most challenging components to be developed for these future earth stations are the down- and up-converters. Their use in a typical earth station is illustrated in Figure 1.

Both up- and down-converters require the realization of complex filtering, frequency translation, and amplification functions at widely differing frequencies with low noise and exceptional frequency and amplitude stability. Production costs can be reduced by using a manufacturing technique that is reproducible with minimal adjustments and leads to unconditionally stable, reliable, and conservatively designed circuit configurations which can be cascaded. These requirements are fulfilled by MICS. The up- and down-converter developmental program has demonstrated that extensive computer modeling and optimization of each circuit could substantially reduce production costs.

Thirteen up-converters and five down-converters have been manufactured, and the ease of reproducibility of the assemblies surpassed expectations. Integration of the individual subassemblies proceeded without complication and later adjustments were not required.

Down-converter configuration

The earth station down-converter receives 12-GHz signals from a remote low-noise amplifier, selects one 50-MHz channel for reception, down-converts it to a center frequency of 70 MHz, and amplifies it to an

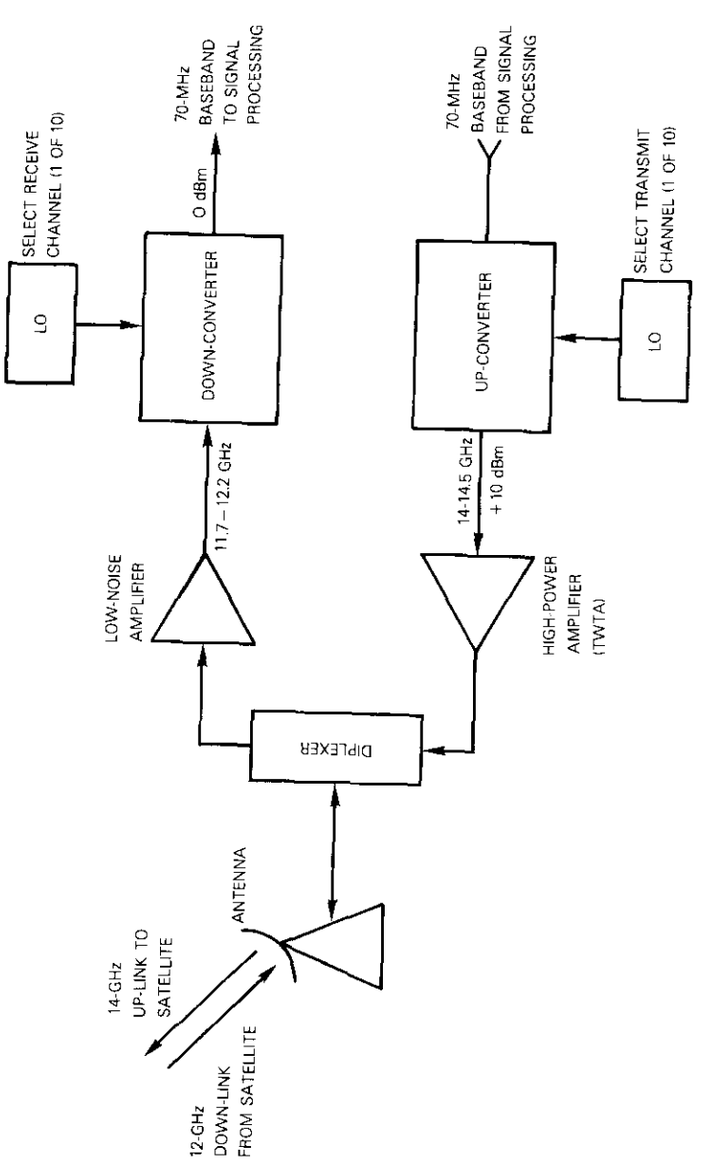
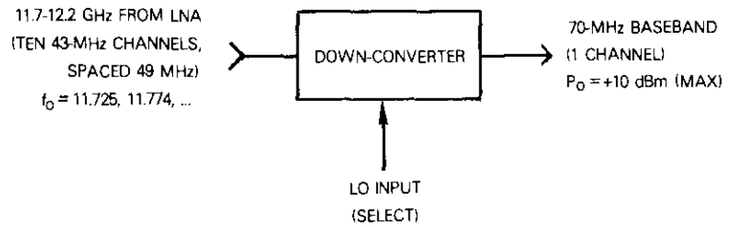


Figure 1. *A Typical 12/14-GHz Earth Station Configuration Illustrating the Use of Up- and Down-Converters*

output level suitable for demodulation at a remote site. Thus, the down-converter is an 11.7- to 12.2-GHz receiver, exclusive of the LO sources and demodulation equipment. Figure 2 shows the major characteristics of the down-converter and its performance requirements. An external low-noise parametric cooled FET amplifier, with 14-dB gain and 1.5-dB noise figure is assumed.



GAIN: 60 dB NOMINAL; ±10 dB REMOTELY ADJUSTABLE
NOISE FIGURE: <4 dB ACROSS 11.7-12.2 GHz
IMAGE REJECTION: >45 dB
GAIN VARIATION: ±1 dB ACROSS 11.7-12.2 GHz (LO CHANGED)
±0.25 dB ACROSS ANY 43-MHz CHANNEL (LO CHANGED)
GROUP DELAY: <3 ns PEAK-TO-PEAK ACROSS 43-MHz CHANNEL

Figure 2. *Down-Converter Requirements*

The most difficult requirements concern the rejection of image signals by 45 dB and the reception of any 43-MHz channel with a change in LO frequency only. The latter requirement dictates the use of a dual-conversion scheme so that filter modifications are not required when changing channels; the large image suppression limits the range of choices for the first IF. An IF which is too high results in a very difficult filtering problem because the fractional bandwidth is small. On the other hand, an IF that is too low results in larger filter structures which cannot be integrated easily and increases the difficulty of designing the LO, which must be operable at any one of 10 frequencies without retuning.

Figure 3 shows the configuration chosen for the RF simulator versions of the up- and down-converters after extensive analysis of the trade-offs resulting from various IF selections.

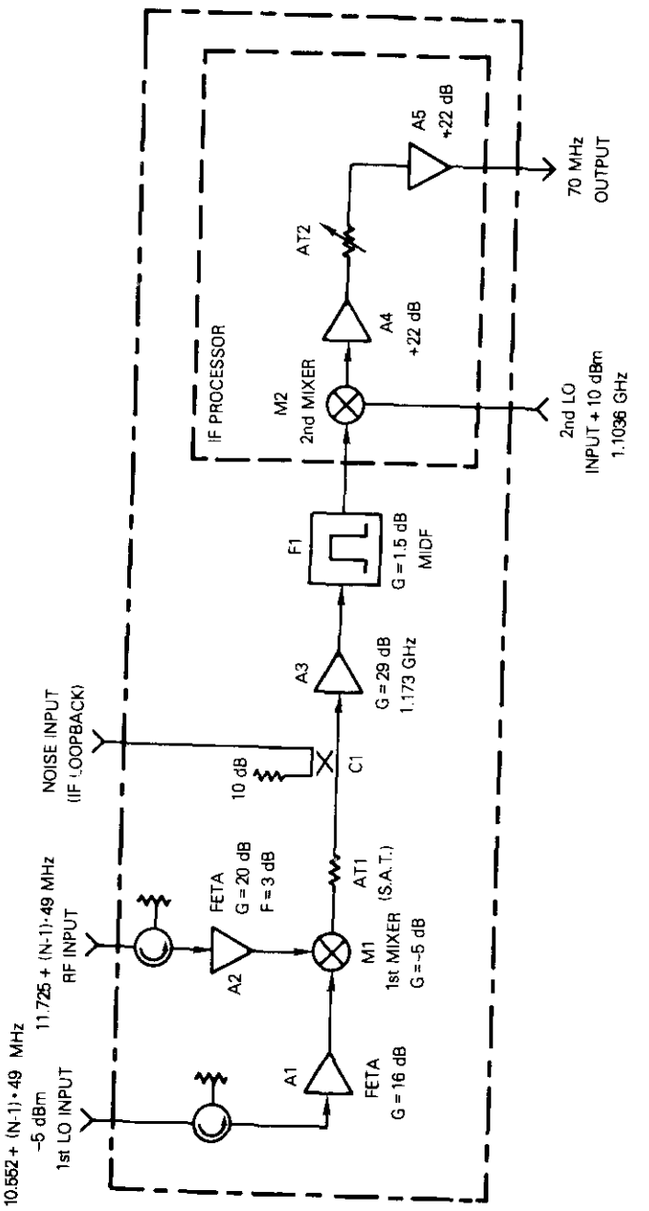


Figure 3. *Selected Configuration for COMSAT Labs Down-Converter (n, the channel number, is from 1 to 10.)*

Component requirements

Amplifier A1 is an MIC FET type which isolates the mixer LO port from the LO input connector, providing 50-dB isolation in the reverse direction. The forward small signal gain of 16 dB guarantees that an input level of −5 dBm will result in saturated operation so that LO drive variations will not affect mixer operation. An MIC isolator is also included at the LO input port to ensure good matching and to further isolate the mixer and its spurious products. The amplifier and circulator must have transmission properties constant to within ±0.5 dB over the entire LO band of 10.552 to 10.993 GHz.

The first mixer, M1, must produce a conversion loss of less than 8 dB across the entire 11.7- to 12.2-GHz band, while remaining insensitive to temperature-induced variations in the LO level of ±1 dB. More importantly, thorough control of all the mixing products is necessary so that the impedances presented to the three mixer ports by the RF, LO, or IF amplifiers do not affect performance. This often neglected feature of mixer design must be addressed if the mixers are to be integrated with other components.

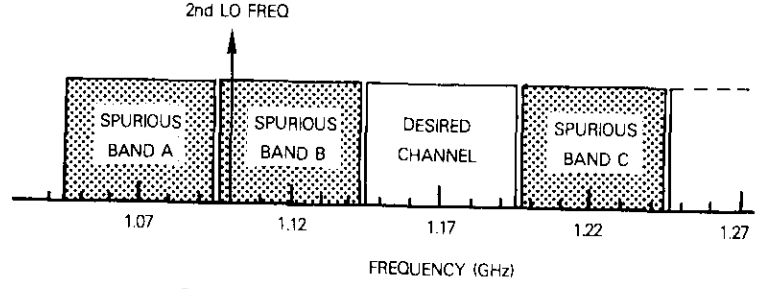
Since the 12-GHz RF input amplifier A2 establishes the noise figure of the down-converter, it must have a noise figure of 3 dB across the 11.7- to 12.2-GHz receive band. An MIC input isolator is included because of the large input VSWR associated with FET mismatching to achieve minimum noise figure. An output MIC isolator guarantees a broadband, low-VSWR impedance match between the mixer and A2.

Attenuator AT1 is a select-at-test MIC type which sets maximum down-converter gain and, in the RF simulator version, also allows the injection of noise into the IF at the same level as the signal. Coupler C1 provides a test port for measuring down-converter performance from the IF amplifier input to the 70-MHz output. In operational earth stations, it also allows for system testing by looping back the IF from the up-converter to the down-converter.

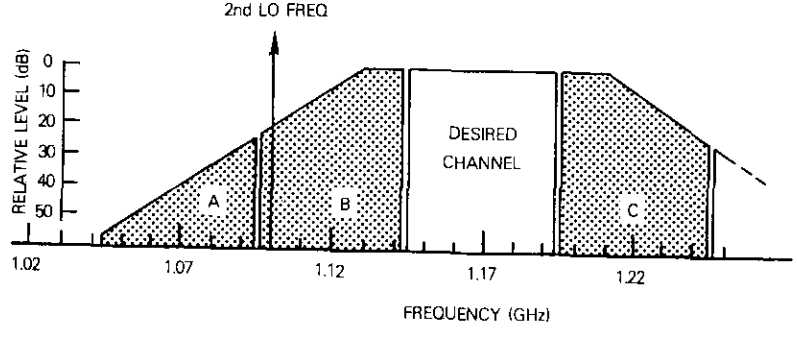
IF amplifier A3, which must provide at least 25-dB gain to establish the noise figure of the IF system, isolates the two mixers (M2 and M1) from one another and establishes a low-VSWR nominal 50-ohm impedance across the IF band of 1173 ±21.5 MHz. This is important in digital modulation applications because of the large equivalent electrical length of IF filter F1, which is approximately 6 m. Mismatches at the input of M2 and output of A3 would result in amplitude and group delay variations

across the channel bandwidth, causing distortion and increasing bit error rates.

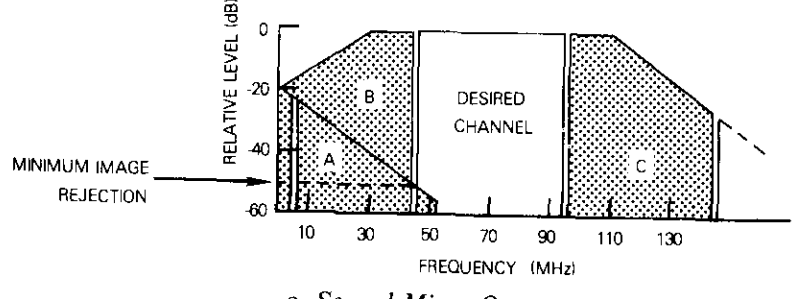
Filter F1 must deal with the image rejection problem shown in Figure 4 without producing excessive group delay variation across the 43-MHz



a. *Output Spectrum of First Mixer Module*



b. *Output of 1.173-GHz Filter*



c. *Second Mixer Output*

Figure 4. *Frequency Plan of Down-Converter*

passband. Since it must be small, inexpensive, and easily reproduced, a microstrip version of the well-known interdigital filter was developed. The final design is printed on a 2.5-cm × 5-cm × 1.27-mm alumina substrate.

Mixer M2, amplifiers A4 and A5, and an electronically variable attenuator AT2 comprise the 70-MHz IF processor and are mounted on a single 5-cm × 5-cm substrate with a low dielectric constant. Because the operating bandwidth is 61 percent, interactions between components must be minimized and the line lengths, which contribute substantial group delay variation and gain ripple across such a large fractional bandwidth, must be limited.

The design of the entire down-converter assembly addressed the following problems which have previously made large-scale integration of MIC subassemblies difficult.

a. Protecting all circuits from the damage which results from the differing thermal coefficients of expansion of each MIC substrate and other materials used in the assembly;

b. Ensuring low-VSWR interfaces between each MIC circuit regardless of the relative looseness of mechanical tolerances typical of production environments;

c. Preventing intercircuit coupling of energy at the many frequencies present within the assembly.

Figure 5 is a photograph of a completed down-converter. The design and performance of the individual components as well as the down-converter assembly will be described in the following sections.

FET amplifier descriptions

The RF and LO amplifier designs incorporate NEC 38800 FETs, self-biased with series source resistors; metal-insulator-silica (MIS) bypass capacitors; beam-lead blocking capacitors between stages; and MIC circuits constructed on 0.38-mm-thick fused silica. Computer-optimized, "tapped" resonator type bandpass matching structures are used to restrict out-of-band gain and ensure stability. The circuit is adjusted by varying the length of open-circuited shunt transmission lines, thus minimizing losses. These design techniques have been extensively developed and proven on previous FET amplifier programs [1], [2].

The thin fused-silica substrates limit radiation from the circuits to low levels while allowing low losses to be achieved. The surface of the fused silica is optically flat (less than 2.5×10^{-8} m irregularities) which minimizes losses, and the dielectric constant is within ±0.5 percent or less from lot to lot which ensures reproducibility of each circuit in production. Special techniques were developed to metallize and etch the substrates; these techniques produce excellent adhesion and etching accuracies of $\pm 2.5 \times 10^{-4}$ cm on the MICs.

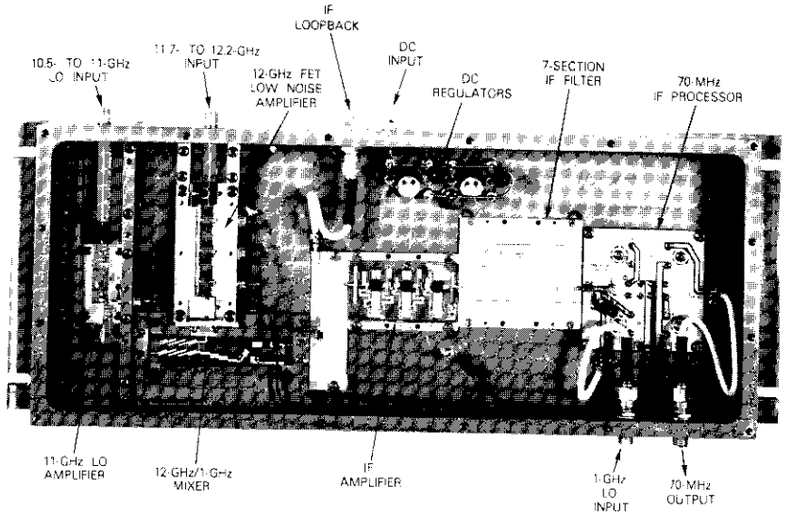


Figure 5. *Photograph of Down-Converter*

Each circuit is mounted on a gold-plated INVAR circuit carrier, designed to match the thermal coefficient of expansion of the fused silica. Contact ridges beneath each MIC input/output line guarantee low-VSWR interfaces between circuits and provide sufficient sliding motion to accommodate thermal expansion differences between the down-converter frame and the circuit carriers. Figure 6 shows the details of this solution to the ground-plane continuity problem.

The FET chips are mounted on metal ridges on the carriers which provide low inductance from source to ground. The 12-GHz low-noise RF amplifier consists of three stages, including three ridges and six substrates on three subcarriers. Therefore, each stage can be measured individually, and the amplifier with the lowest noise figure can be selected for the input stage. The LO amplifier uses two stages consisting of two ridges and three substrates on a single carrier. Reliability is maximized and costs are mini-

mized by utilizing inexpensive monolithic ceramic capacitors wherever their larger size and highly temperature-sensitive capacitance can be tolerated.

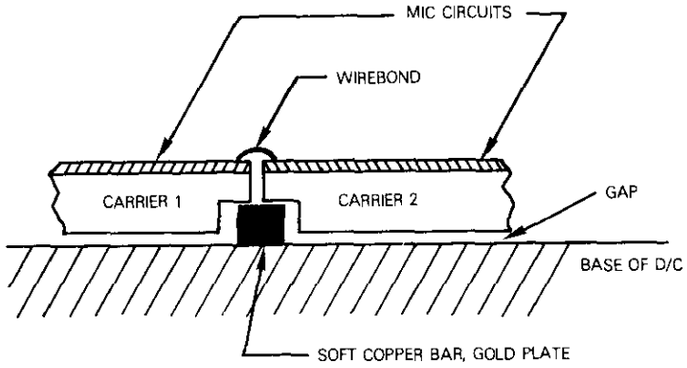


Figure 6. *Details of MIC Interconnection Technology*

Since exhaustive computer modeling of each FET amplifier design was completed before a mask was cut, only minimal circuit adjustment was necessary. After one mask iteration, tuning is performed only to accommodate the FET's varying S-parameters. Computer modeling of the RF and LO FET amplifiers included analysis of the circuit performance from 2 to 18 GHz followed by optimization for out-of-band gain, gain flatness, input and output VSWR, and stability. As for any multistage amplifier, stability was checked by examining the internal impedances of the amplifier computer model over very wide bandwidths and verifying that the real parts were never negative. Specialized programs, which have been verified against observed results over many years, were used in the computer-aided design process, including GCP15, HVDUTY, and OPTI13 [3], [4].

Figures 7 through 9 show typical measured performance of the LO amplifier. The gain is 15.7 ± 0.3 dB across the LO band, and at the nominal input level of −5 dBm the amplifier is 2.5 dB into saturation. The design provides a particularly well-matched impedance at the mixer (output) port. Figure 10 is a photograph of the low-noise RF amplifier, and Figure 11 shows its measured performance. The gain is 21.5 ± 0.2 dB across the 11.7- to 12.2-GHz band.

During fabrication of the four down-converters, it was observed that little or no adjustment of the LO amplifiers was necessary; however, the

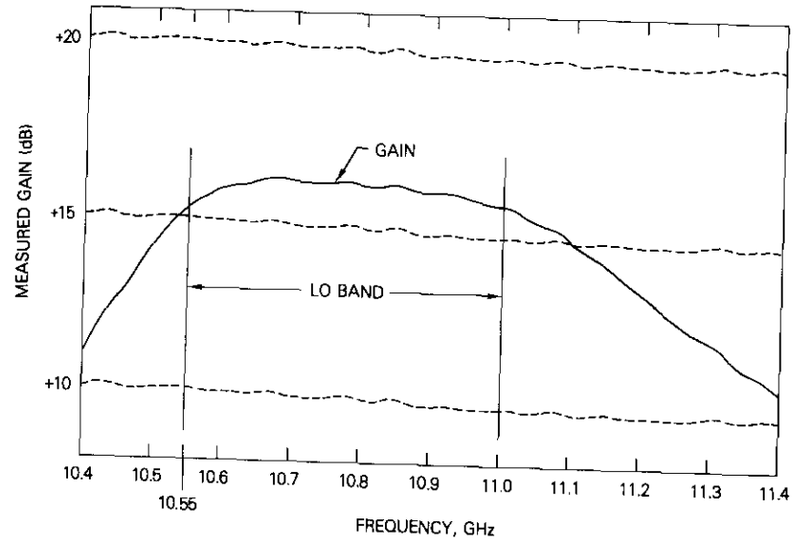


Figure 7. *Measured Gain of FET LO Amplifier*

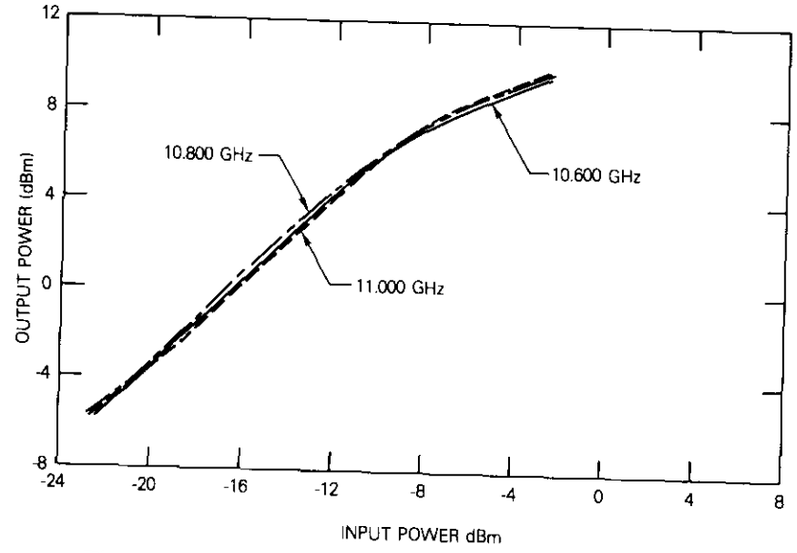


Figure 8. *Saturation Characteristics of FET LO Amplifier*

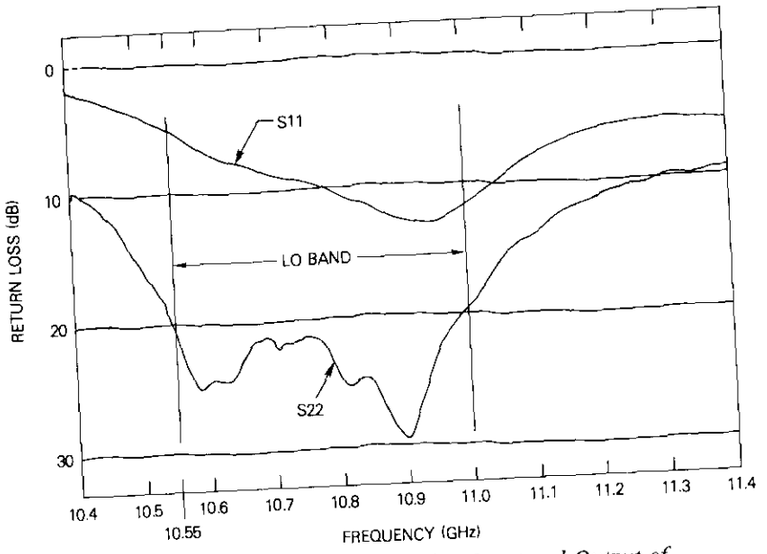


Figure 9. *Return Losses Measured at Input and Output of FET LO Amplifier*

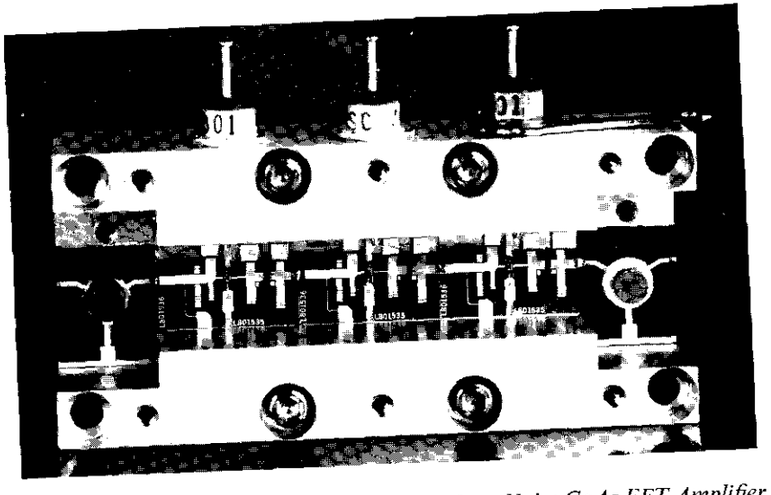


Figure 10. *Photograph of 11.7- to 12.2-GHz Low-Noise GaAs FET Amplifier*

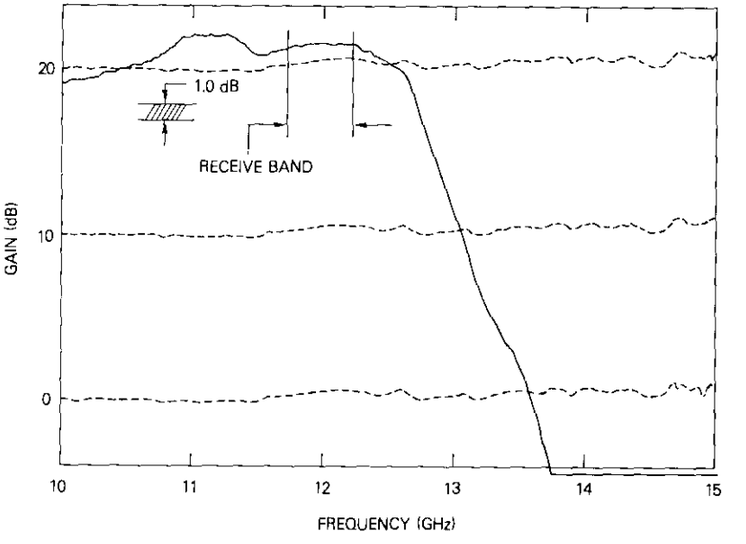


Figure 11. *Wideband Measured Response of a 3-Stage Low-Noise Amplifier*

low-noise stages for A2 generally required some optimization before integration. A noise figure of 3.6 dB maximum, including circulator loss, was achievable without using special selection techniques.

MIC mixer

The MIC mixer design uses a proven LO/RF diplexer arrangement and IF filter, as shown in Figure 12, to completely control all mixing products, thus making the mixer performance insensitive to loading effects at the three ports and preventing the propagation of unwanted frequencies. The 3-dB quadrature hybrids are interdigitated and have been designed for use on 0.381-mm fused silica [5]. A Schottky barrier beam lead diode is used.

Because of the extreme repeatability of the edge-coupled MIC filters on fused silica, the required tuning of the mixer circuit is negligible and confined entirely to the diode impedance matching structure. Special techniques have been developed to accurately design these MIC filters [6]. Figure 13 is a photograph of the MIC mixer, and Figure 14 shows typical performance. Since the circuit is image enhanced, conversion losses near 5 dB are typically observed across the 12-GHz receive band.

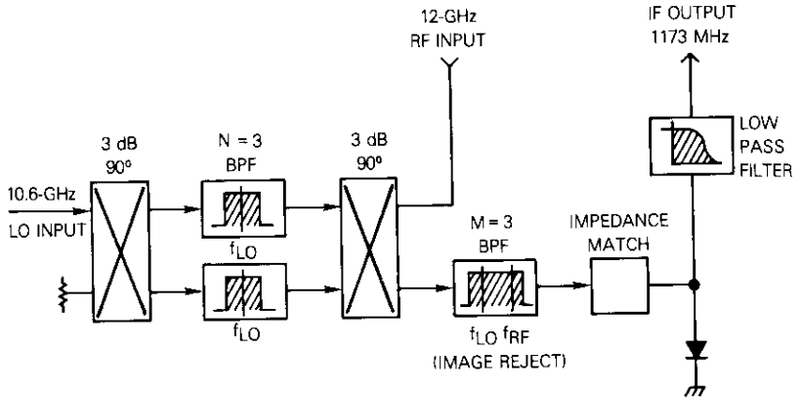


Figure 12. *MIC Mixer Block Diagram*

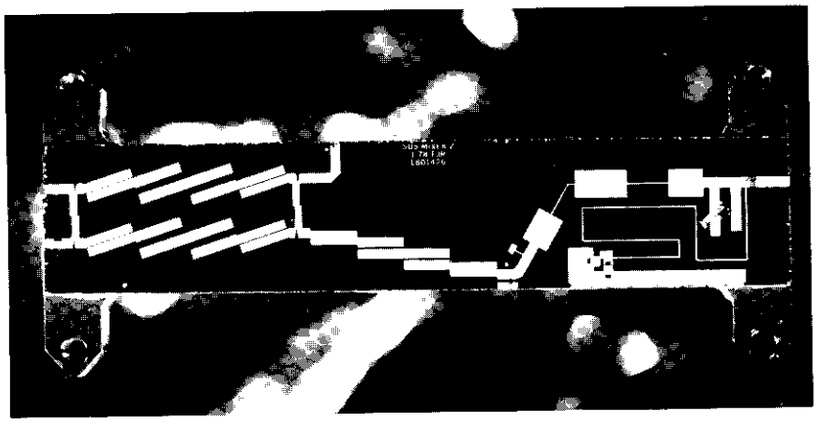


Figure 13. *Photograph of Down-Converter MIC Mixer*

This circuit is an excellent example of a successful trade-off between developmental effort and ease of reproduction. Although the design procedure is complex and requires the use of a sophisticated nonlinear analysis program, the result is a circuit which can be easily duplicated with minimal tuning. This contrasts with alternate mixer designs which, although simple, require extensive adjustment of each circuit to achieve proper performance and sufficient stability margins in a production run.

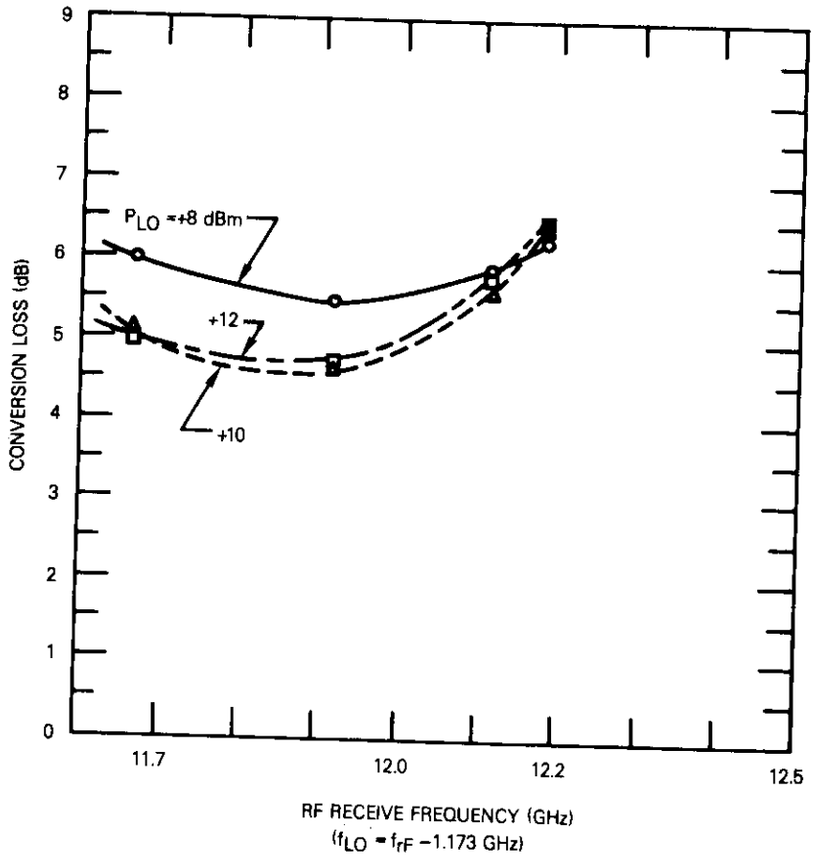


Figure 14. *Measured MIC Mixer Performance*

IF components (1173 MHz)

Image filter

Figure 15 shows the computed performance for a 7-section 0.1-dB-ripple Chebychev filter with a 120-MHz bandwidth and resonator Q's of 200, which is the expected value for microstrip filters operating at L-band. The computer model, which is based upon a transverse electromagnetic (TEM) filter, indicates that seven sections are required if the group delay

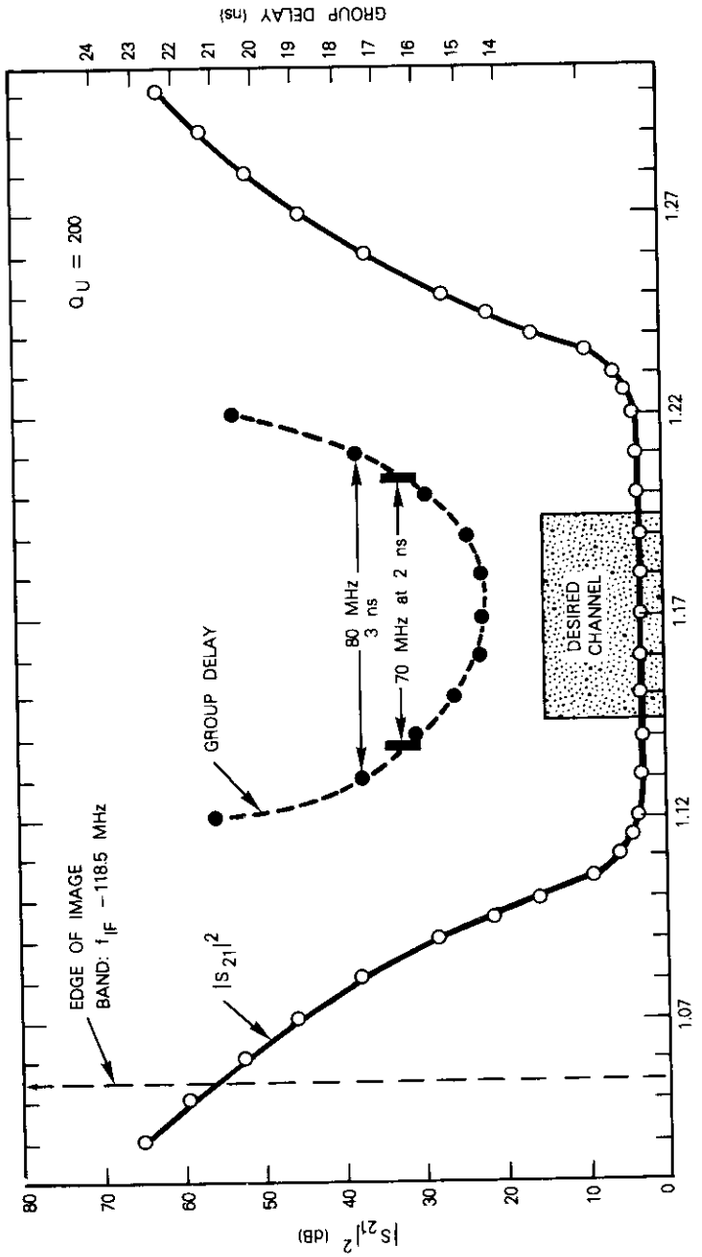


Figure 15. *Predicted Performance of a TEM Interdigital Chebychev Filter with N = 7, Δ = 0.1 dB, BW = 120 Hz, f_o = 1173 MHz, Q_u = 200*

variation across a 43-MHz band is not to exceed 2 ns (allowing ±1-percent center frequency error in production), while the image band is rejected by 50 dB or more. Filters with fewer sections and narrower bandwidths could be used, but they would require substantial tuning to meet the 2-ns requirement because of the smaller allowable error in center frequency. The 2-ns value is based upon two-thirds of the entire down-converter group delay specification.

Various filter configurations were investigated; however, none of the candidates were acceptable. An air dielectric interdigital filter was rejected on the basis of cost and size; an inexpensive microstrip edge-coupled 180° filter was too large on alumina ($\epsilon_r \cong 10$) and too lossy on barium tetratitanate ($\epsilon_r \cong 38$).

This problem was eventually solved by developing a microstrip interdigital filter* with parameters chosen by a design procedure based upon admittance synthesis rather than capacitance synthesis [7]. Figure 16 is a photograph of a completed 7-section microstrip interdigital filter, and Figure 17 shows its performance at the 1173-MHz IF. A top shield affected the skirt performance of the filter at frequencies below f_o and could, in fact, be made to produce a transmission zero at the edge of the image band. This results from the coupling between nonadjacent resonators provided by the shield; the response under these conditions resembles that of an elliptic function filter.

The measured insertion loss of 1.5 dB indicates that an unloaded resonator Q greater than 200 was achieved. The addition of the top shield also partially compressed the bandwidth; nevertheless, the measured group delay variation was within specification. The image rejection exceeded 60 dB because of the quasi-elliptic function shape achieved with the shield.

IF amplifier

The IF amplifier was designed to be inexpensive while realizing a low noise figure, high gain, large bandwidth, and a very low ouptut VSWR. The finished amplifier, as shown in Figure 4, consists of three bipolar transistor stages on an Epsilam-10 MIC board. Distributed elements and lumped resistors are used to match impedances and stabilize the amplifier at all frequencies. The amplifier contributes negligible gain slope and less than 0.3 ns of group delay variation to the overall down-converter performance. The intercept point for third-order intermodulation products is +24 dBm.

*Patent application has been filed.

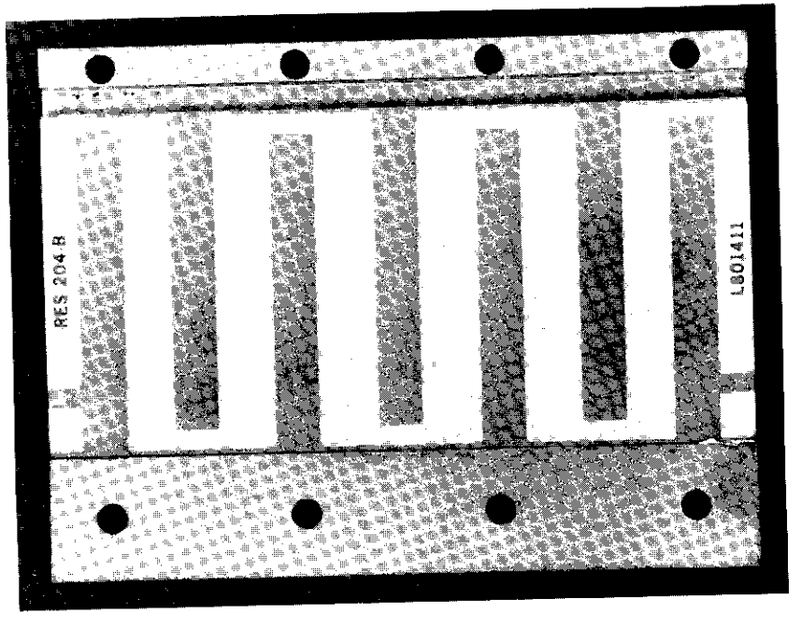


Figure 16. *Photograph of 7-Section Microstrip Interdigital Filter*

The 29-dB gain and 3-dB noise figure extend from 0.5 to 1.3 GHz.

The production cost of this amplifier is substantially less than that which could be achieved using a cascade of commercially available "unit" amplifiers. Figure 18 shows the excellent agreement between measured and computed gain of the IF amplifier. As a result of the large operating bandwidth, no tuning or adjustment has been necessary on most of the 40 amplifiers fabricated to date. The same amplifier is used in the up-converter and 1.103-GHz LO chain.

IF processor

Components of the miniaturized hybrid IC IF processor are incorporated on a 5-cm × 5-cm epoxy-glass printed circuit board. A double-balanced mixer, which down-converts to 70 MHz, is followed by two amplifiers in *TO*–8 cans and a variable attenuator consisting of a balanced mixer that is unbalanced intentionally by DC bias. The input to the mixer contains a variable capacitor, allowing optimization of the input return loss at

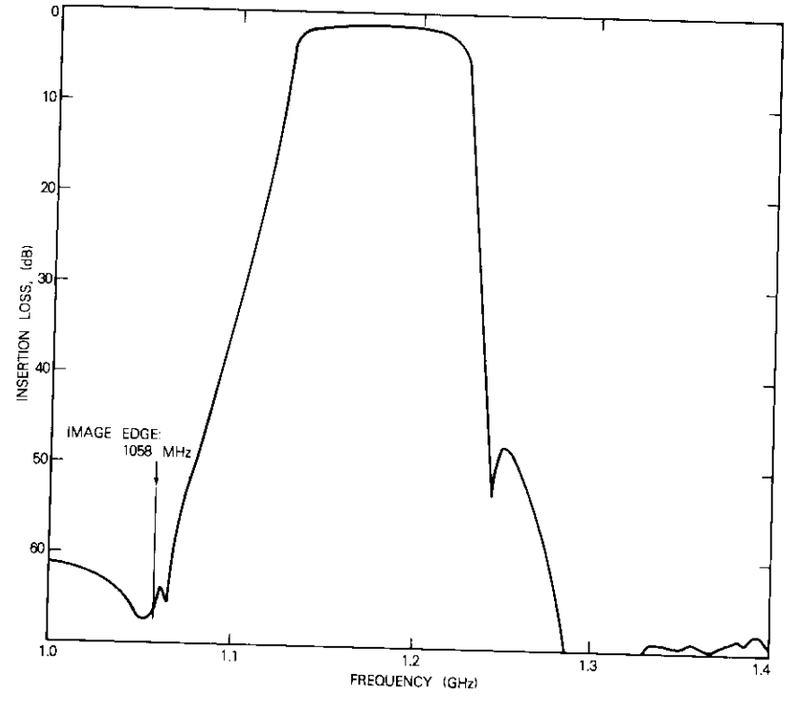


Figure 17. *Measured Insertion Loss of IF Filter with N = 7, δ = 0.1 dB, BW = 120 MHz, f_o = 1173 MHz*

1.173 GHz. This is necessary because of the potential for rapid gain ripple inherent in the very "long" IF filter when terminated by a poor impedance match.

Down-converter integration and results

During initial manufacture of the down-converter units, RF tests were performed on each FET amplifier, mixer M1, amplifier A3, and filter F1 before integration. By the conclusion of the production run, this pretesting had been limited to units A2 and F1. Adjustment of F1 was necessary because no preselection or screening procedure had been used to verify that the dielectric constant of the alumina substrate was within 2 percent from lot to lot.

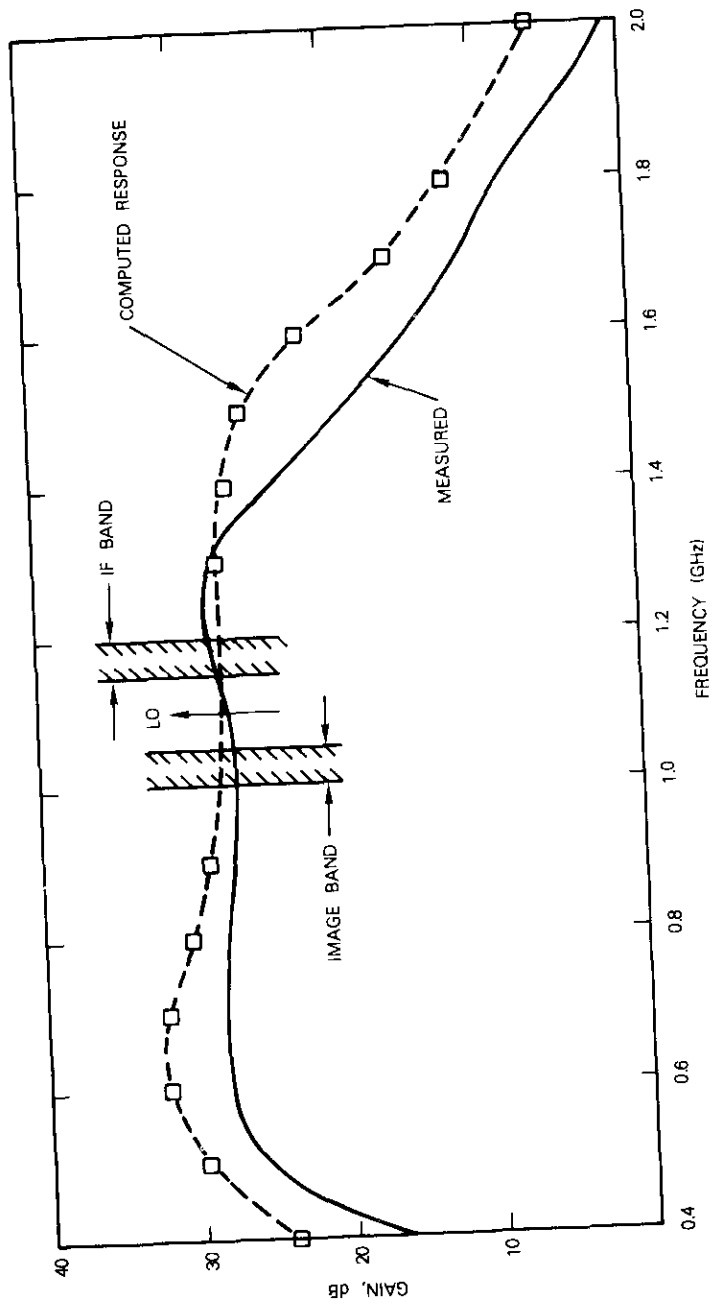


Figure 18. *Measured and Computed Response of a 3-Stage IF Amplifier*

After the components were integrated into the down-converter, virtually no additional adjustments were necessary. Tuning of IF filter F1 or the second mixer input was occasionally necessary to obtain the flattest possible passband. This ease of integration can be attributed to the effort which minimized the discontinuities between subassemblies and ensured that each subassembly was well matched at all frequencies of interest and unconditionally stable.

Covers were necessary for amplifiers A1 and A2 to prevent low-level LO spurious frequency signals, which propagate within the large down-converter frame, from gaining access to the down-converter's "front end" and appearing in the baseband. In production, such covers may be eliminated in favor of resistive loading within the frame, since this approach would reduce mechanical fabrication costs.

Figures 19, 20, and 21 show typical measured performance of the integrated MIC down-converter. In Figure 19, the frequency response is essentially that of the IF filter. The additional notch results from the low-frequency roll-off of amplifiers A4 and A5, which do not exhibit gain below 5 MHz. At the image band edge, an overall rejection of greater than 50 dB has been obtained. Figure 20 shows about ±0.2-dB typical gain ripple inband. Group delay (Figure 21) was, as expected, contributed entirely by the IF filter and IF processor and ranged from 1 to 3 ns for the four down-converters fabricated.

Temperature behavior

Each component was designed for minimal performance variation over a temperature range of −20°C to 50°C. Therefore, it was unnecessary to compensate for overall down-converter performance. Gain was stable to within ±1 dB over the above range, and the noise figure degraded to 4.25 dB maximum at 50°C.

Up-converter design and performance

The up-converter translates the 70-MHz baseband signal to any one of 10 channels in the 14.0- to 14.5-GHz SATCOM up-link band and delivers an output level (+5 dBm) sufficient to drive the earth station transmit TWTA chain. Spurious outputs, such as noise, images, and LO feedthrough, must be adequately suppressed. This is especially important for the first LO spurious output, which appears 70 MHz below the center of the output channel regardless of whether the up-converter is being driven. For a

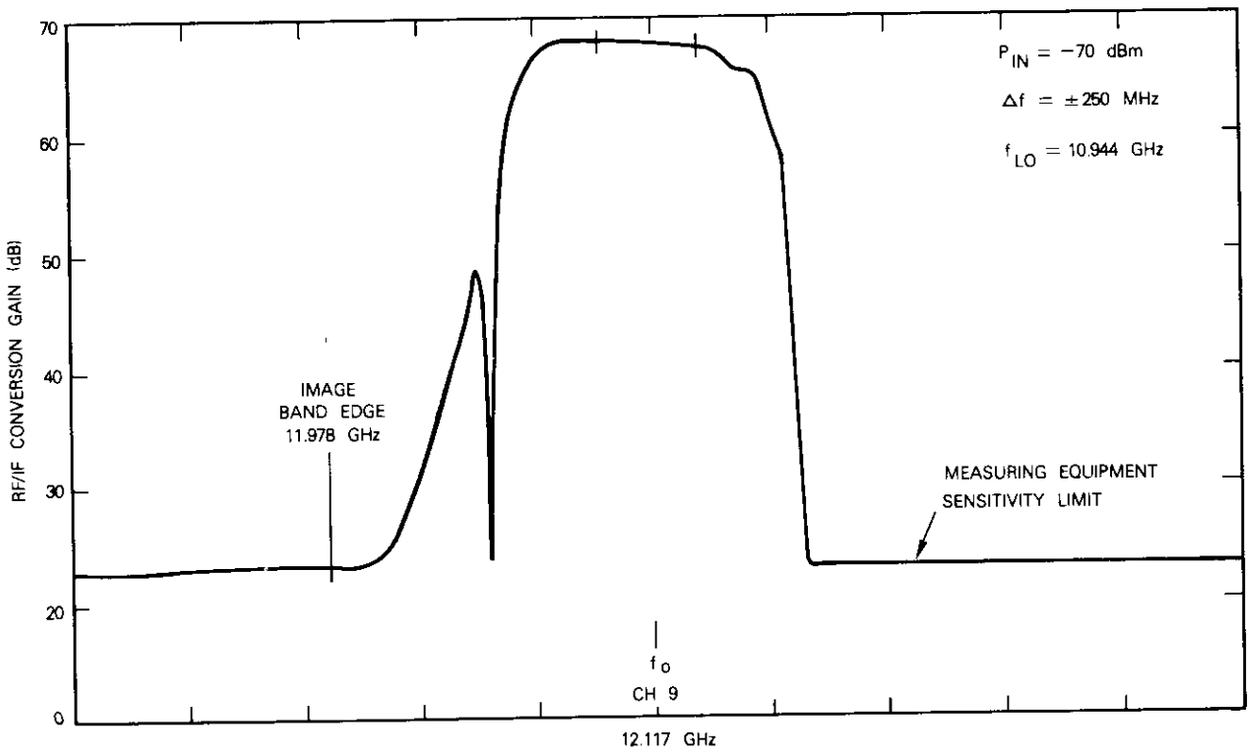


Figure 19. *Down-Converter Conversion Gain vs Input Frequency (LO Chosen for Channel 9)*

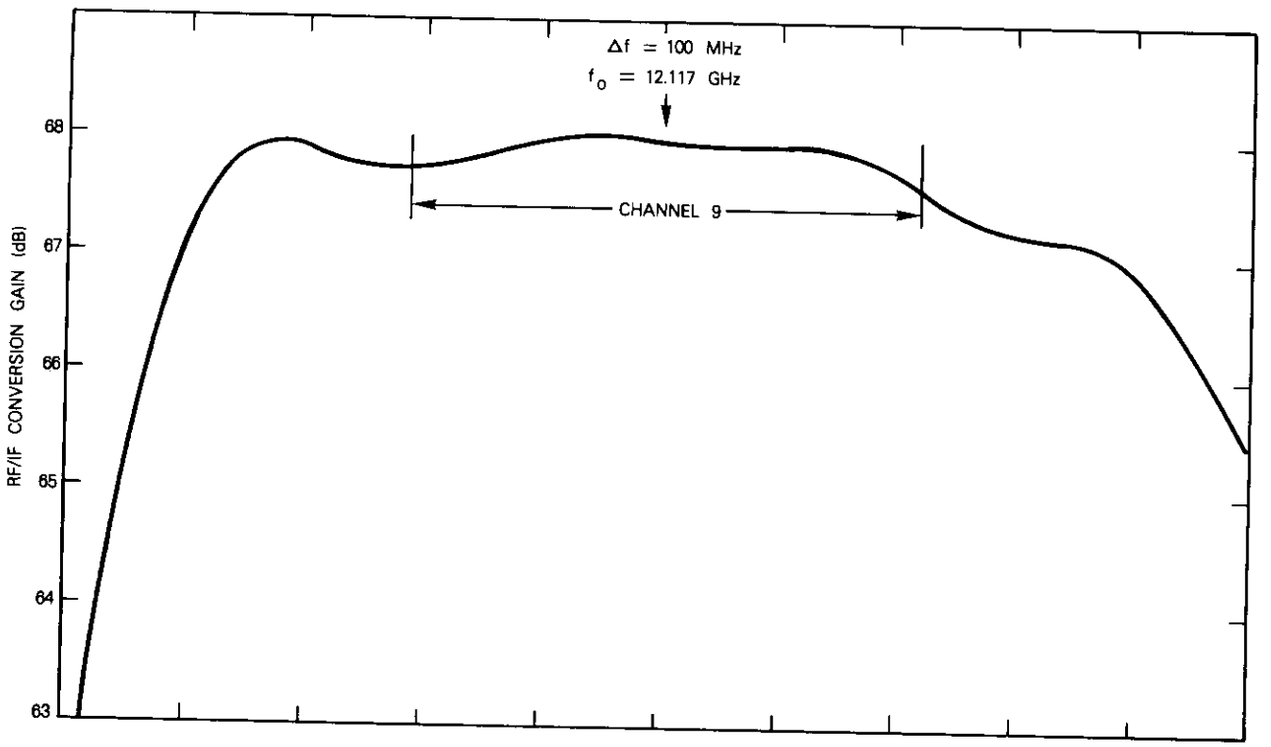


Figure 20. *Inband RF/IF Conversion Gain of Typical Integrated Down-Converter*

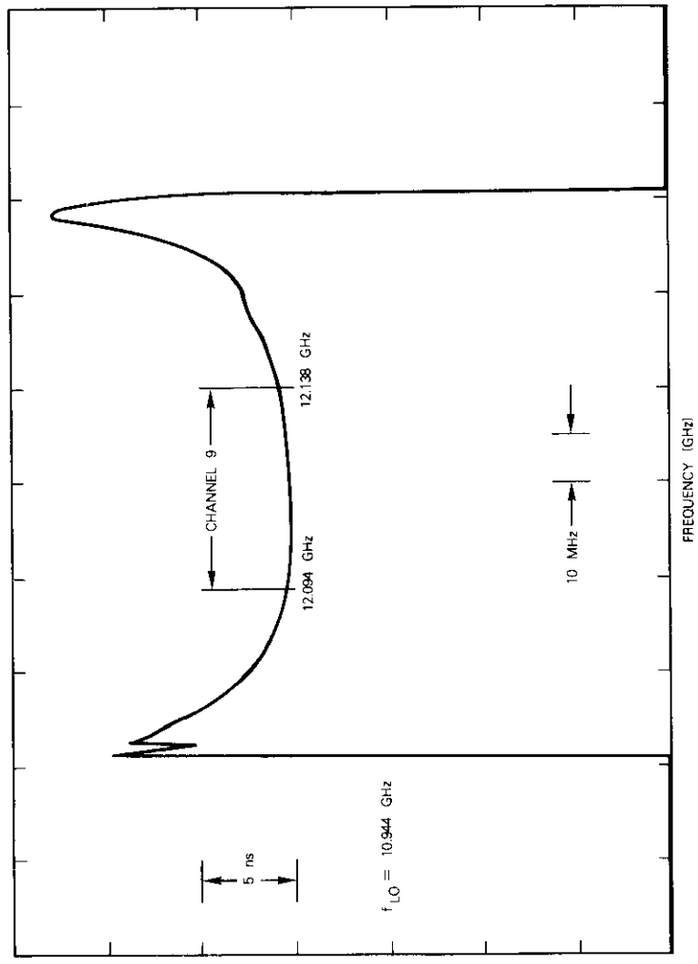


Figure 21. *Measured Group Delay of Integrated Down-Converter Across the Single Receive Channel (Specification is 3 ns.)*

system employing a significant number of earth stations, this interfering signal is cumulative and must consequently be suppressed to a level at least 50 dB below the nominal up-converter output. The effect is intensified by the increase in gain of the earth station high-power TWTA when no carrier is present. The spurious image signal (140 MHz from the output channel) must be suppressed to a level at least 45 dB below the desired output.

Noise generated within the up-converter and transmitted by the high-power amplifier is especially objectionable under no-signal conditions; therefore, the gain and noise figure product is important. However, the version described was not designed to minimize transmitted noise. Because of these considerations, the technical problems encountered in designing the up-converter are more formidable than those associated with the down-converter. Figure 22 is a block diagram of the MIC up-converter.

The LO at 1.1036 GHz is suppressed by 50 dB or more relative to the signal at the 14-GHz output by selecting a first mixer (M1) which exhibits a typical LO rejection of 50 dB and by incorporating a 3-section MIC LO bandstop (notch) filter (F1) in the *L*-band IF section. The IF bandpass filter (F2) contributes an additional 15 dB of rejection for a theoretical total LO suppression of 115 dB measured from the LO port of M1 to the input port of amplifier A1. In actual practice, the leakage from the LO input connector, cable, and mixer M1 is higher and establishes the LO power at the input port of A1 near −80 dBm, or 50 dB below the level of the up-converter baseband.

The bandstop filter (F1) must contribute negligible group delay at the edge of the IF band, which is 48.5 MHz (or 4.4 percent) above the LO frequency; 40-dB suppression of the LO signal must be maintained over the operating temperature range. The IF bandpass filter requirements are similar to those of the down-converter IF filter. The up-converted image centered 140 MHz below the signal must be suppressed by a minimum of 45 dB.

An IF amplifier raises the signal level, isolates the mixers, establishes the noise figure of the up-converter, and provides a broadband impedance match to the 1/14-GHz MIC mixer. FET Amplifier A3 produces an output level of +5 dBm at a 2-tone third-order C/I ratio of 30 dB. Bandpass filter F3 reduces the feedthrough of the second LO by 25 dB. Amplifier A2 elevates the second LO level to nominally +10 dBm, isolates the LO port of mixer M2 from the up-converter LO port, and operates in a saturated

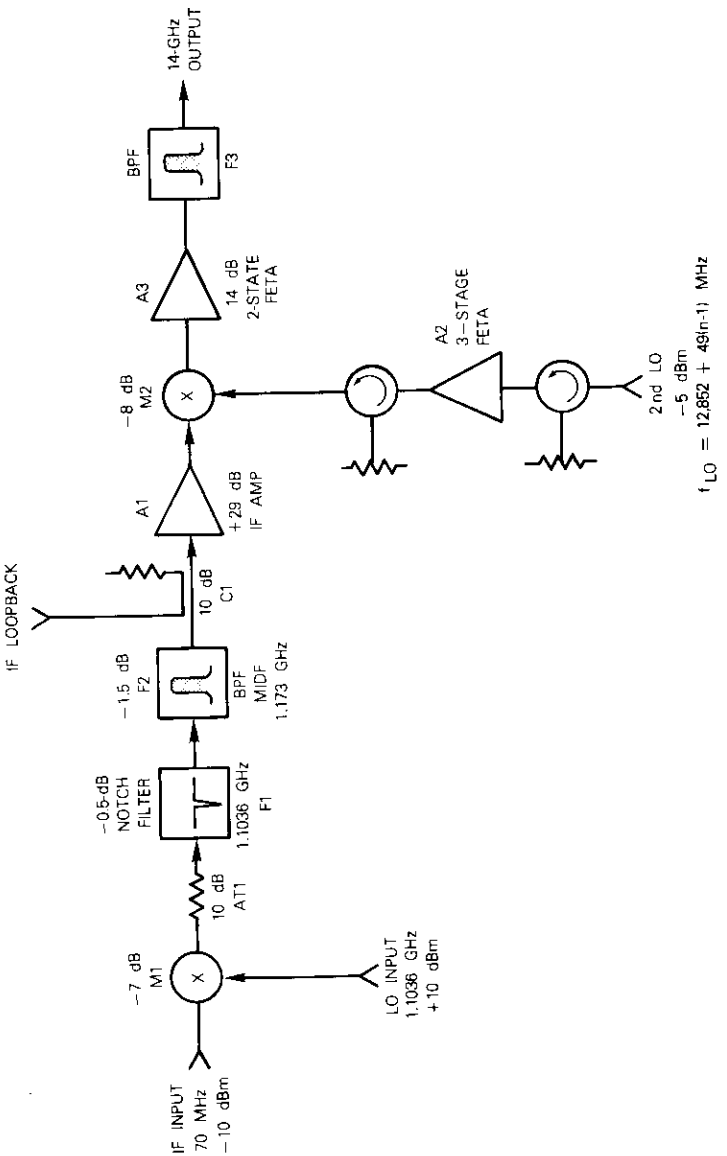


Figure 22. *MIC Up-Converter Block Diagram*

mode to minimize LO variations. Figure 23 is a photograph of a completed up-converter.

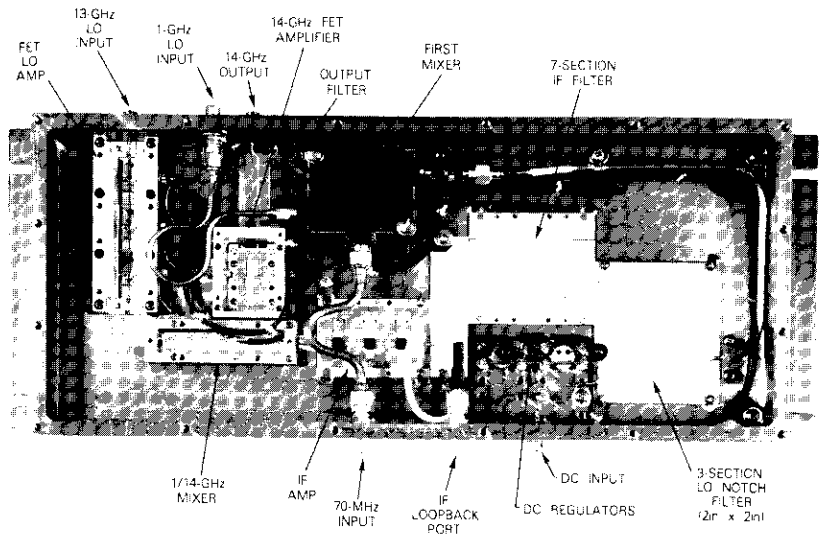


Figure 23. *Photograph of MIC Up-Converter*

Up-converter component description

Double-balanced diode mixer M1, which is realized in strip-line form, was chosen for its high LO rejection, varying from 50 to 65 dB. This mixer, which is the only subassembly in either the up- or down-converter that was not designed and fabricated at COMSAT Laboratories, is not easily integrated with the other MIC components in the up-converter because it utilizes SMA connectors. The 10-dB attenuator, AT1, is a coaxial unit employed to resistively terminate the image of mixer M1. If it is not used, complete reflection of the image by filter F2 results in substantial amplitude ripple across the 43-MHz signal band.

The band reject or "notch" filter is a 3-section MIC design on 1.27-mm alumina with 90° edge-coupled sections terminated by 90° open-circuited transmission lines. It is based upon a 0.1-dB Chebychev bandpass filter transformation with a 3-percent bandwidth. A computer model of the design included the effects of losses and the nonequal even- and odd-

mode phase velocities inherent in microstrip edge-coupled lines. Performance of the final design (Figure 24) was virtually identical to the predicted results; 40 dB of LO rejection was obtained over the operating temperature range. The very narrow bandwidth combined with the variation in dielectric constant of the alumina substrate necessitated individual adjustment of each notch filter before integration into the up-converter. This adjustment could be eliminated if the substrate dielectric constant was prescreened to be within ± 0.1 percent of the required value.

Bandpass filter F2 and amplifier A1 are identical to the microstrip interdigital filter and IF amplifier, respectively, used in the down-converter. Mixer M2 is very similar in design and construction to the down-converter MIC mixer; the same nonlinear analysis computer program was employed in the design. Performance data and a photograph of this mixer are presented in Figures 25 and 26, respectively. The fused-silica substrate allows the edge-coupled filters within the mixer to be reproduced consistently with center frequency accuracies of a fraction of one percent.

FET amplifiers

The FET amplifiers in the down-converter, as in the up-converter, utilize NEC's NE38800 devices in chip form, bonded to metal ridges which separate the fused-silica substrates. Computer modeling is used during the design phase to eliminate the possibility of out-of-band instabilities and optimize the passband shape. RF amplifier A3 shown in Figure 27 consists of four FETs configured as two balanced stages; fused-silica interdigital hybrids [5] developed at COMSAT Laboratories are used to sum the outputs. This amplifier has a measured third-order intercept point of +23 dBm, 14-dB gain across the 14.0- to 14.5-GHz band, and input and output return losses exceeding 20 dB.

Figure 28 is a photograph of LO amplifier A2. The three stages guarantee that an input level of −5 dBm will compress the output stage, thus stabilizing the LO input level to the mixer. MIC isolators are included in the up-converter assembly to ensure low-VSWR impedance matches at both ports.

Up-converter performance

The measured up-converter flatness and group delay typically achieved for any fixed second LO frequency, as shown in Figures 29 and 30, respectively, fulfill the system requirements with considerable margin. Rejection of the image exceeds 50 dB and leakage from the first LO is 40 dB below the signal when observed at the 14-GHz output. A 3-section MIC bandpass filter following the output FET amplifier was used to achieve a 35-dB total sup-

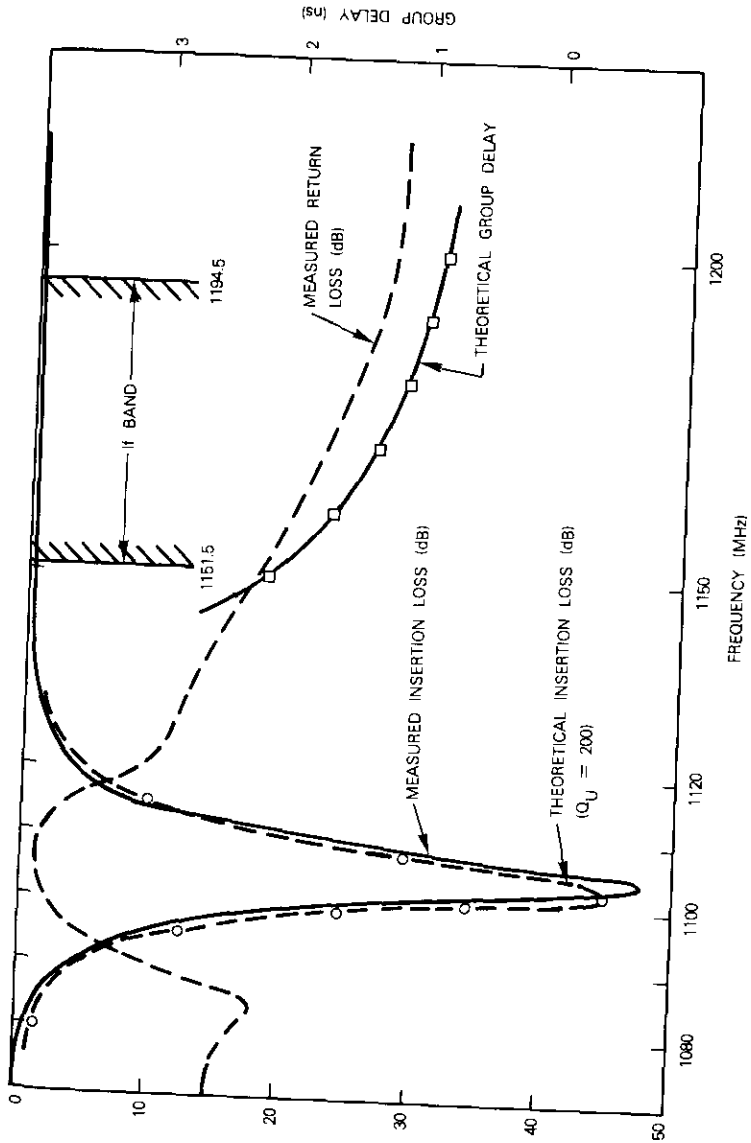


Figure 24. *Theoretical and Measured Performance of a 3-Section MIC Notch Filter*

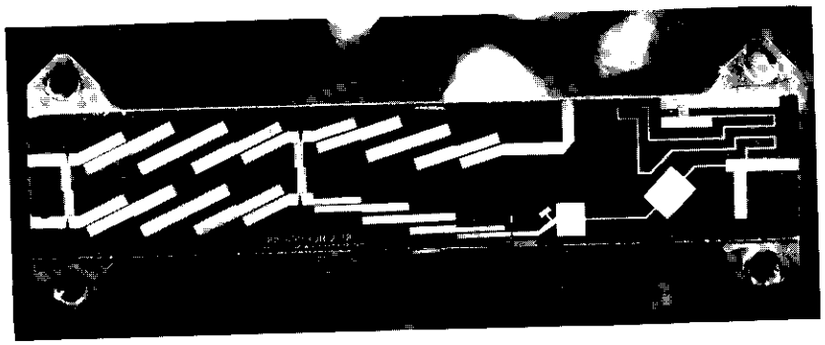


Figure 25. *Photograph of the Up-Converter Mixer*

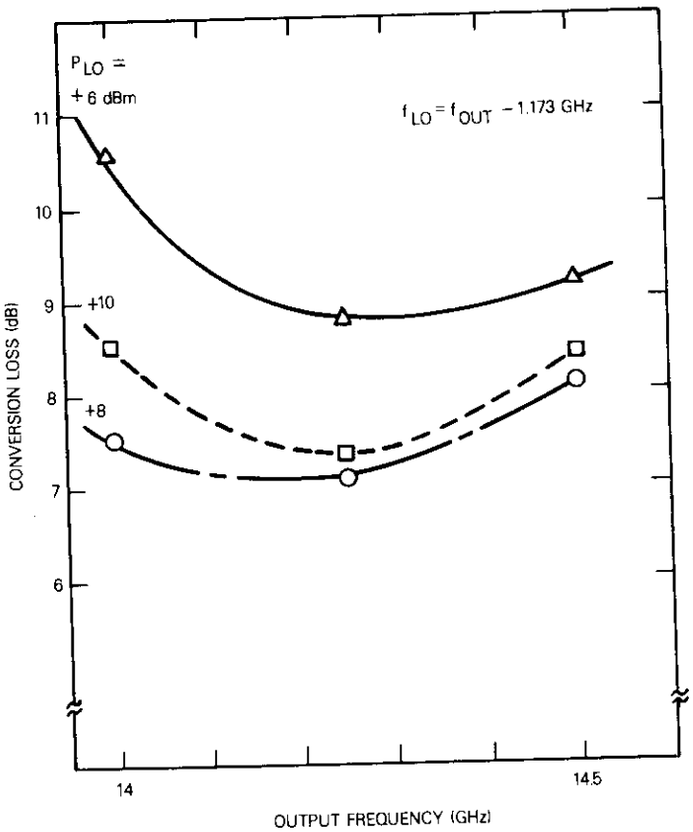


Figure 26. *Measured Performance of the MIC Up-Converting Mixer*

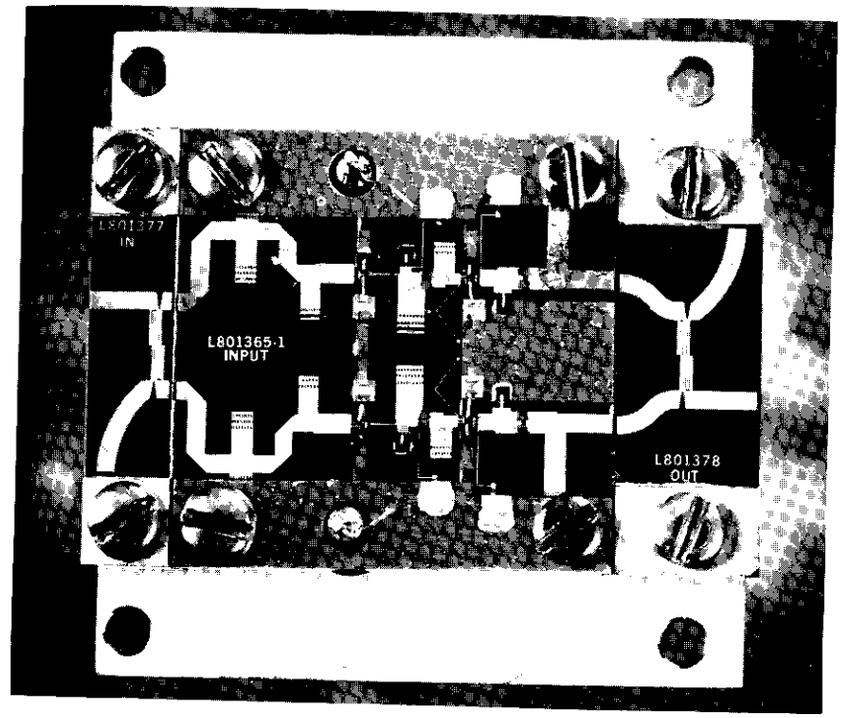


Figure 27. *Photograph of 14-GHz Balanced FET Amplifier*

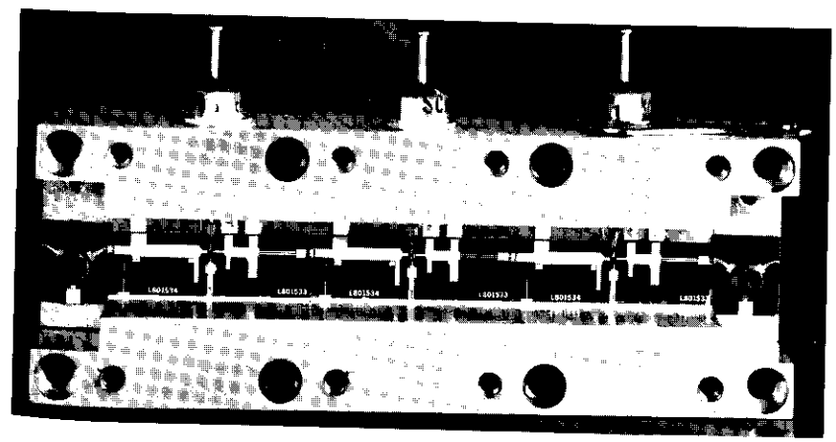


Figure 28. *Photograph of 13-GHz GaAs FET LO Amplifier*

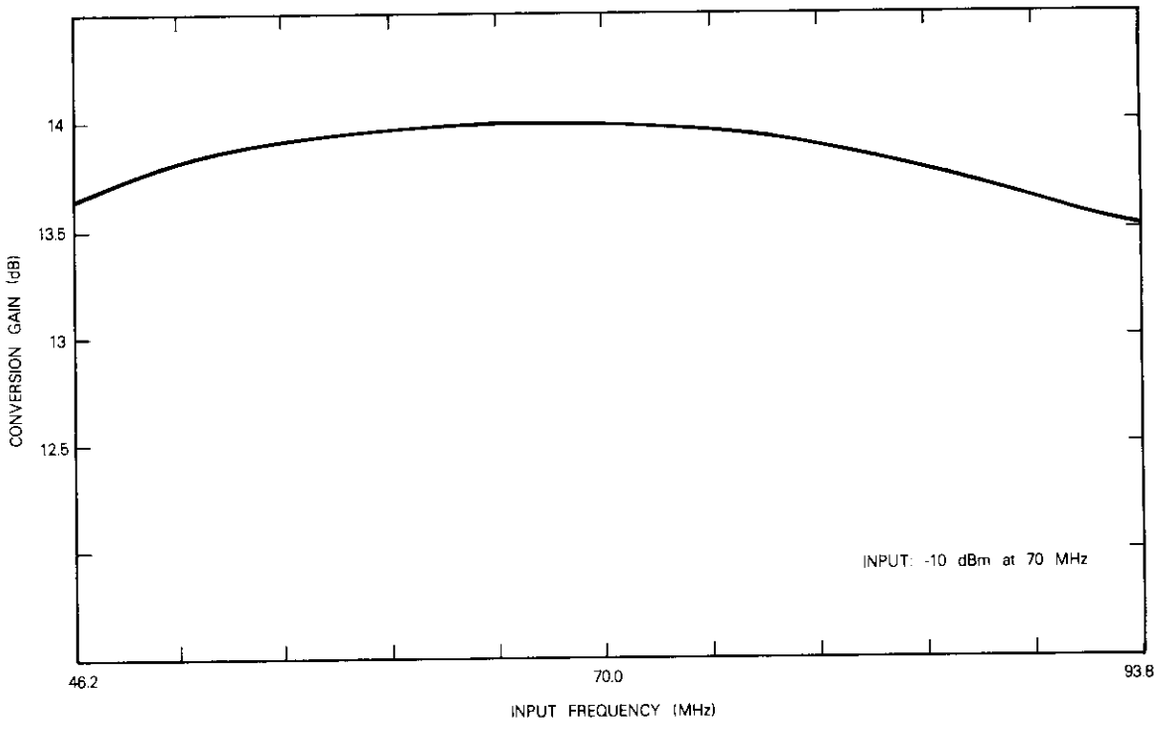


Figure 29. *Measured Up-Converter Conversion Gain vs Input Frequency with an Output Frequency of 14 GHz (Second LO is fixed.)*

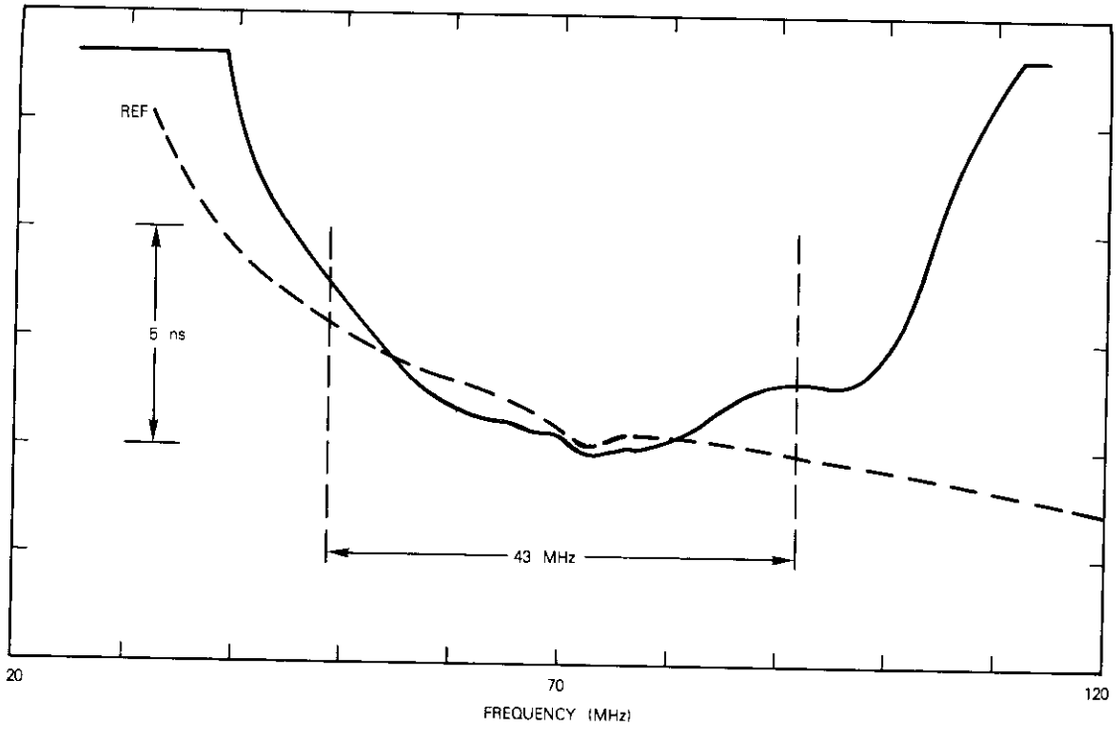


Figure 30. *Measured Group Delay*

pression of the second LO relative to the signal. As with the down-converter, no significant readjustment of any subassembly was required after up-converter integration.

Summary and conclusion

Up- and down-converters of the type needed for 14/12-GHz earth station applications and which exhibit state-of-the-art performance have been realized as integrated assemblies of low-cost MICs. Production costs have been minimized by applying computer-aided design techniques to each circuit and by carefully designing the integrated assembly to minimize discontinuities between subassemblies and coupling of spurious energy. In a production run of 18 complete up- and down-converters, minimal adjustment of the integrated assembly was necessary to meet the specifications.

In addition to the benefits of low cost, small size, and high reliability, performance is superior to that which could be achieved from the interconnection of separate assemblies. This is expected, since transmission lines connecting the subassemblies are reduced to very short lengths. Efforts are continuing to further reduce fabrication costs as well as to improve performance with respect to image and LO suppression.

Acknowledgments

Initial studies of the up-converter design were performed by W. Childs. The FET amplifiers were designed by H. Hung (13-GHz LO and 12-GHz low-noise amplifier) and M. Dressler (14-GHz up-converter output and 10-GHz LO amplifiers). Both MIC mixers were designed by F. Rieger and W. Childs. The down-converter IF processor subassembly was designed by J. Molz. Design of the up- and down-converter chassis was completed by W. Burch and C. Barrett.

The author also wishes to acknowledge the significant contributions of J. Molz and S. Taylor in the fabrication and testing phases. The assistance of C. Mahle is also gratefully acknowledged.

References

[1] H. L. Hung, R. E. Stegens, and M. Dressler, "Low-Temperature Performance of GaAs MESFET Amplifiers at 14.25 Hz," *Proceedings of the Cornell Electrical Engineering Conference: Active Microwave Semiconductor Devices and Circuits*, August 1977, pp. 331–337.

[2] P. Estabrook et al., "A Low Noise Single-Ended GaAs FET Amplifier for a 14 GHz Satellite Communication Application," *IEEE MTT-S 1978 International Microwave Symposium Digest*, pp. 129–131.

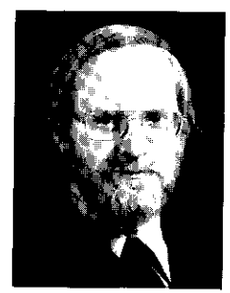
[3] A. Atia and R. E. Stegens, Private Communication.

[4] W. Childs, Private Communication.

[5] W. H. Childs and P. A. Carlton, "A 3-dB Interdigitated Coupler on Fused Silica," *IEEE MTT-S 1977 International Microwave Symposium Digest*, pp. 370–372.

[6] W. H. Childs, "Design Techniques for Bandpass Filters Using Edge-Coupled Microstrip Lines on Fused Silica," *IEEE MTT-S 1976 International Microwave Symposium Digest*, pp. 194–196.

[7] R. E. Stegens, "A Design Technique for Microstrip Interdigital Filters," to be published.



Ronald E. Stegens received a B.S.E.E. degree from the Case Institute of Technology in 1964 and an M.S.E.E. from the State University of New York (Buffalo) in 1969. He participated in the design of electronically steerable antenna arrays at Sylvania Electronic Systems, Buffalo, New York. After joining COMSAT *Laboratories in 1969, he developed microwave integrated circuit components for the ATS-6 experiment and the Beacon program. He is presently Assistant Manager of the Microwave Circuits Department of the Microwave Laboratory, responsible for development of FET amplifiers, oscillators, and mixers for use in future satellite communications systems.*

Index: attitude control, spaceborne microcomputer, reaction wheel, momentum control

Air bearing testing of a skewed reaction wheel system for attitude control*

A. RAMOS

(Manuscript received June 15, 1978)

Abstract

Tests of an attitude control system, using skewed reaction wheels and a spaceborne microcomputer, are described. Stability, accuracy, and speed of response were evaluated in the normal and backup modes, mode transitions, various submodes, and during wheel failure; tests were supported by analyses and computer simulations. For system operation with the minimum of three wheels, peak attitude errors are acceptable even for a maximum motor-torque failure. The importance of feeding the momentum of a failed wheel into the distribution matrix is proven. It is shown that momentum control, in contrast to traditional torque control, offers many advantages in the design and operation. A skewed reaction wheel system appears attractive for high accuracy and reliability in geostationary communications satellites.

*This paper is based upon work performed at COMSAT Laboratories under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Views expressed in this paper are not necessarily those of INTELSAT.

Introduction

An attitude control system which uses an array of skewed reaction wheels and a spaceborne microcomputer was tested with the air bearing facility at COMSAT Laboratories. The testing, supported and verified by analyses and digital and analog computer simulations, was part of an effort to investigate and develop the technology required to apply reaction wheels in a long-life, high-accuracy control system with maximum reliability and minimal mass and cost, for possible use on future spacecraft. Two primary objectives were given. The first was to evaluate system performance for all modes of satellite operation, including the normal mode, simulated stationkeeping, and mode transitions. Performance was measured in terms of stability, accuracy, and speed of response. The second objective was to study operational procedures associated with using skewed reaction wheels and a spaceborne microcomputer.

The tests demonstrated that a skewed reaction wheel system is capable of achieving high reliability and attitude accuracy for geostationary communications satellites. The performance is compatible with accuracies of 0.05° in pitch, roll, and, if necessary, yaw under all modes of operation. Even if high yaw accuracy is not required, this system provides fast yaw response, which facilitates the transition from stationkeeping to the normal mode. In contrast, two-sensor momentum bias systems perform poorly during the transition interval.

System techniques and procedures for resolving reaction wheel failure to utilize the skewed arrangement of wheels were tested under simulated failures. The test results confirm that the characteristics of the microcomputer-based control logic unit (CLU) are well matched to this application. Similarly, the quantization and scaling of input, internal and output variables, processor speed, and the instruction set are shown to be appropriate. Procedures for switching modes, updating parameters in the control program, and unloading momentum by ground command were developed and tested.

R&D effort requirements

Reaction wheels have been used on numerous spacecraft. Notable examples are the Broadcast Satellite Experiment (BSE), Orbiting Geophysical Laboratory (OGO), Nimbus, Orbiting Astronomical Observatory (OAO), Earth Resources Technology Satellite (ERTS), Molniya, and ATS-6 spacecraft. These satellites represent a wide range of applications in terms

of orientation (earth and astronomical), attitude accuracy (2° for the early OGOs to about 0.1 arc-second for the astronomically oriented OAOs), and orbit (low-altitude, highly elliptical, and synchronous).

A characteristic common to these satellites is that the wheels are arranged in an orthogonal triad, aligned with the attitude control (pitch, roll, and yaw) axes. Therefore, if this arrangement is followed for redundancy purposes, one-for-one redundancy is required. A potentially better method in terms of reliability, mass, and cost is to arrange the reaction wheels so that any three can supply the required control torques. This method is referred to as a skewed reaction wheel arrangement.

The ATS-6 and BSE are the only geostationary communications satellites among the examples given. Their basic attitude accuracies are better than 0.1° when thrusters are inhibited from firing. It was decided to study the application of the skewed reaction wheel concept to potential communications satellites requiring attitude accuracies of 0.05° under all operating conditions. This accuracy must be maintained continuously, even when thrusters are fired, *i.e.*, during stationkeeping and momentum unloading. This type of control system and the overall mission involve multiple operating modes, axes that are actively controlled, and torquers. Therefore, a spaceborne control processor with the structure of a programmable digital computer was considered a desirable concomitant development. Advantages of this technology over the traditional use of dedicated, hardwired circuitry are enhanced reliability, greater versatility, and mass and power reductions realized by time sharing hardware. To implement the control processor, it was necessary to employ a multiple-chip design approach, based on new large-scale integration (LSI) bit-slice microprocessor technology, in which each part is space qualified.

Typical flight application

A typical flight application was defined as a mission for the attitude control system. Table 1 briefly summarizes the detailed mission requirements given in References 1 and 2.

The modes of operation are attitude acquisition, the normal mode, and a backup mode. The normal mode uses reaction wheels for control torquing and consists of nadir pointing, offset pointing, momentum unloading, and stationkeeping. The backup mode uses thrusters for torquing. The specified accuracies are maximum or 3 σ levels, and must be met during all normal and backup modes of operation.

TABLE 1. TYPICAL FLIGHT REQUIREMENTS

Mission	Earth-oriented communications
Orbit	Geostationary
North-South Stationkeeping Limits	±0.1°
East-West Stationkeeping Limits	±0.05°
Moments of Inertia	Of the order of 1000 kg-m²
Low-Level Disturbance Torques	Of the order of 10^{-6} N·m
Disturbance Torques During Stationkeeping	Due to ±5% thrust imbalance and a shift of the center-of-mass equal to 3 cm
Pitch and Roll Accuracy	±0.05°
Yaw Accuracy	±0.1°

Operation

Attitude control laws can be selected so that an array of reaction wheels stores the system angular momentum accrued from external disturbance torque. Each reaction wheel is an accurate velocity (equivalently, momentum) servo, which uses tachometer feedback. These servos, driven by momentum commands that are proportional-plus-integral functions of the attitude errors, functionally differentiate the momentum commands, producing reaction torques between the spacecraft body and the wheel rotors. The control laws that are selected yield a stable system. Therefore, the control system exchanges momentum between the reaction wheels and the external world while keeping body rates at zero average value and attitude errors sufficiently small.

Since a part of the external disturbance torque is constant in inertial space, the stored momentum tends to increase. When the stored momentum approaches the capacity of the array of reaction wheels, it must be unloaded by exerting appropriate external torque on the system (by thrusters). During the unloading process, attitude errors are minimized by commanding the wheels to obtain a reaction torque opposing the thruster torque.

The thrusters that are used for unloading serve as a backup system to the reaction wheel system if more than one wheel fails; they control attitude during stationkeeping, function as a secondary system for large disturbances when enabled with coarser thresholds in the normal mode, and can be used for attitude acquisition. Figure 1 [3], a block diagram of the control system, shows the main signal flow in the operating modes.

The wheels in a skewed system are not aligned with the control axes. However, the momentum commands are generated on a control-axis

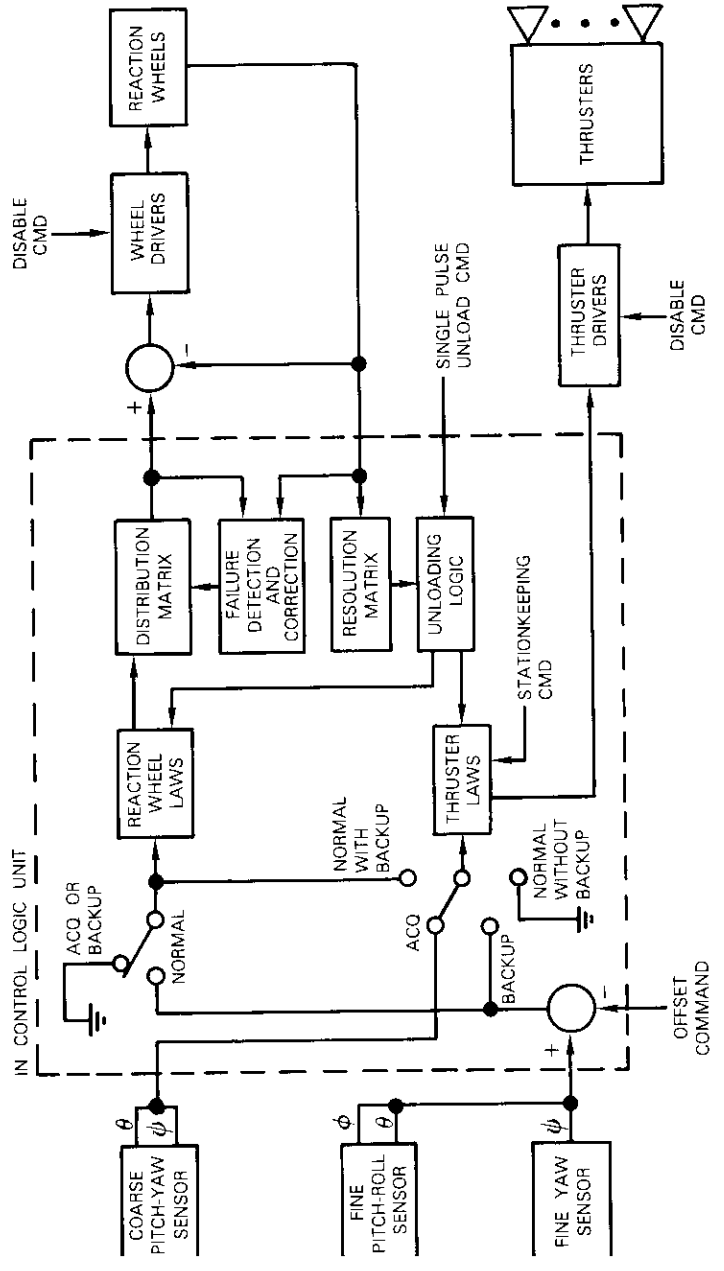


Figure 1. *Simplified Block Diagram of the Control System*

basis, rather than on a wheel basis, for the following reasons:

a. in the main, the control axes are dynamically decoupled (for this type of control system);

b. the attitude errors are sensed on a control-axis basis;

c. the multifunction thrusters are aligned on a control-axis basis.

Therefore, the momentum commands H_{cx}, H_{cy}, and H_{cz} must be combined to form the velocity commands to the individual wheels. This is the function of the distribution matrix, which depends on the wheel orientation and the particular wheels that are operating. Conversely, the processed tachometer signals are combined by a resolution matrix to calculate the momentum stored in the reaction wheels on a control-axis basis. These components of momentum are then used in the unloading logic, as well as other applications.

Wheel failure is detected for each wheel. The command to each wheel servo is also used to drive a simplified mathematical model of the wheel servo. Failure of a reaction wheel channel is detected by comparing actual wheel velocity with the velocity estimated in the model. If this difference exceeds a threshold value for longer than a certain grace period, then the corresponding wheel channel is deemed to have failed. The failure is corrected by turning off the faulty wheel and using the distribution matrix that is appropriate for the operating wheels. The momentum of the faulty wheel is subtracted from the momentum command to minimize the effect of its rundown friction.

Figure 2 [3],which is a more detailed block diagram of the normal mode control with wheels, explicitly shows only the momentum control loop (the outer loop) for the pitch axis and the velocity servo for wheel 0. The roll and yaw loops and the other wheel channels are similar. For concreteness, four wheel channels are shown, the same number used for the air bearing test.

When the angular momentum about a particular axis exceeds a given threshold, a sequence of thruster (unload) pulses of equal width is automatically fired until the momentum component decreases beyond a given smaller threshold. The pulse width and the thruster torque are selected so that the momentum is decreased by a given amount with each unload pulse.

The simplest scheme for unloading is to fire the thrusters and to null the resulting attitude error through the action of the outer loop, thereby achieving the desired change in momentum. A novel method (Reference 4, pp. 60–61) of reducing the peak attitude error is to accompany each unload pulse with a pre-emphasis pulse that commands the desired change in wheel

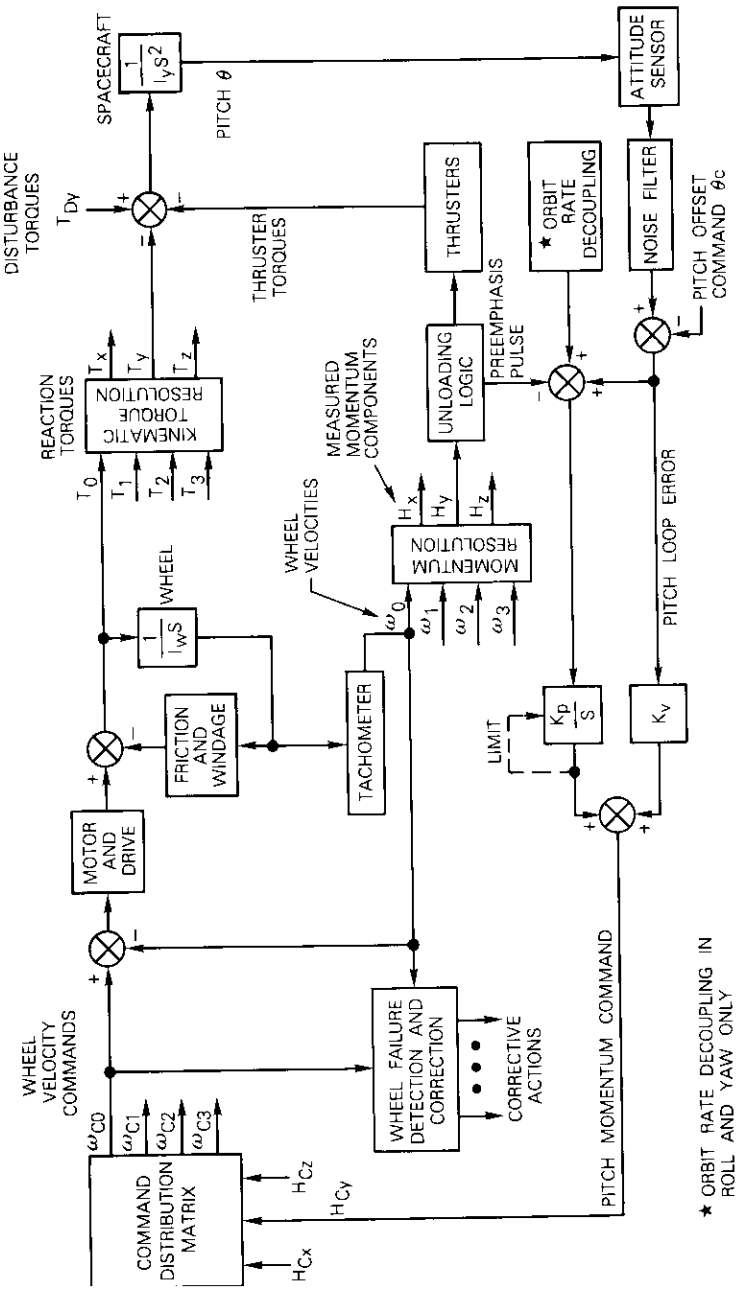


Figure 2. *Block Diagram for Normal Mode Control*

momentum, as shown schematically in Figure 2. This technique reduces the error because the pre-emphasis (reaction wheel) pulse and the unload (thruster) pulse yield torques on the spacecraft body of opposite sign. Ideally, the change in momentum is equal to the angular impulse (pulse width multiplied by torque) of the pre-emphasis pulse. Although many factors detract from this ideal condition, attitude errors can be minimized by careful design.

Precession torques ($\omega \times H$) originate from the orbit rate ω_o and the stored angular momentum. These torque terms occur because the angular momentum of a momentum-conservative system must remain fixed in inertial space and therefore oscillate in the roll-yaw plane of the spacecraft (Reference 4, pp. 58–59). If considered to be disturbances, these torque terms would result in large attitude errors in the air bearing test because of the fast orbit (one revolution per hour) and the considerable momentum that may be stored. The orbit rate is a known constant, and the momentum of the system can be estimated from the tachometer outputs. Therefore, the precession torques $\omega_o H_x$ and $\omega_o H_z$ can be considered known and opposed by so-called orbit rate decoupling (ORD) terms in the control functions, as shown in Figure 2. These terms command reaction wheel velocities such that the roll and yaw components of momentum are properly interchanged, without generating attitude errors.

The pitch momentum command H_{cy} is the sum of proportional and integral feedback terms. The former is obtained by multiplying the pitch loop error by the rate loop gain, K_v. The loop error, modified by the compensatory preemphasis pulse and the ORD term (for the roll and yaw loops only), is integrated and then multiplied by the position loop gain, K_p, to obtain the integral feedback term.

Description of design and of test facility

Figure 3 [3] shows the control system that was tested with the air bearing facility. The local vertical frame (x_o, y_o, and z_o) was used as a reference. The y_o-axis points down in the laboratory (and south in orbit), and the z_o-axis points toward the pitch-roll target (and toward the earth in orbit). The x_o-axis completes the right-handed set of orthogonal axes (and points along the orbit velocity for a circular orbit). Ideally, the body-fixed frame (x, y, and z) is aligned with the reference frame. The figure shows the pitch θ, roll ϕ, and yaw ψ attitude angles, and the directions of geostationary orbit quantities that were being simulated. Orbiting was simulated by mounting the pitch-roll target on a rotating arm, whose axis of rotation is

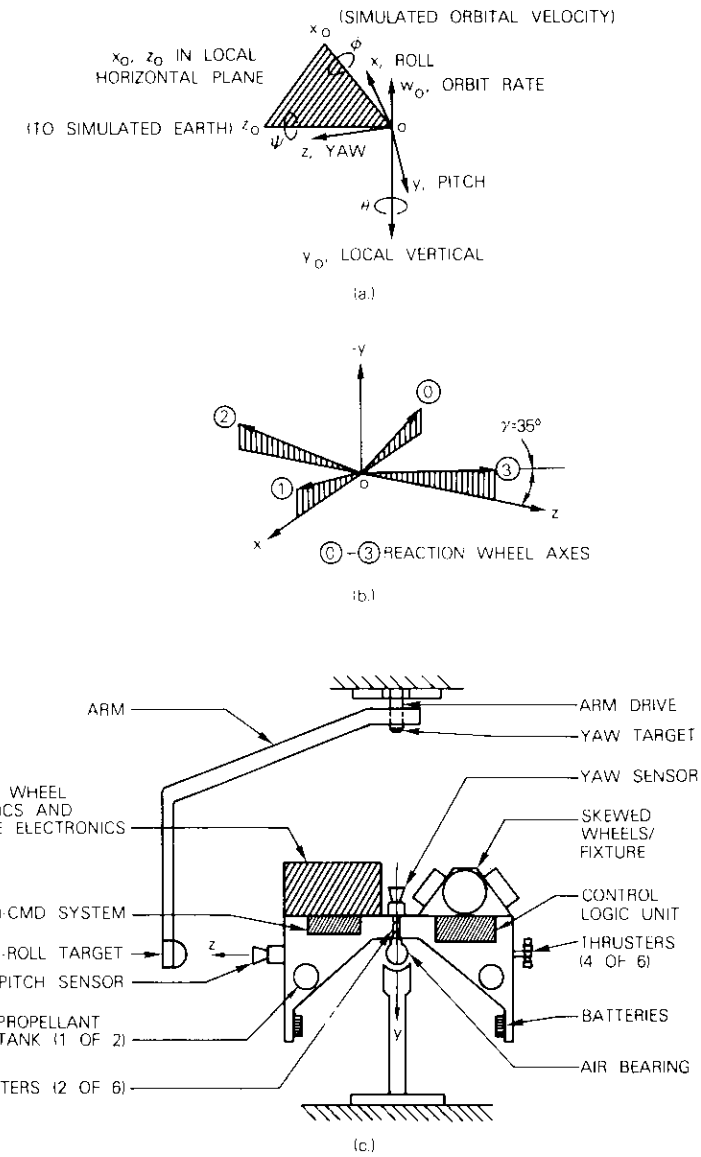


Figure 3. *Control System on Test Facility*

colinear with the y_o-axis. For simulation, its velocity was one revolution per hour, that is, 24 times greater than the synchronous orbit rate. The air bearing platform, which could be tilted by as much as 45° from the vertical, had unlimited rotation about the vertical.

Four reaction wheels of identical design are mounted on a pyramidal fixture. As shown in Figure 3b, their spin axes are either in the xy- or the yz-plane and are tilted by an angle $\gamma = 35°$ from the xz-plane. Their identification and directions of positive angular velocity are also shown. Each wheel has an angular momentum of 6.1 N·m·s at the maximum operating speed of 2,100 rpm. Motor torque is 0.1 N·m for speed less than 2,100 rpm. The tachometer is built with slots that are accurately cut into the steel rim of the rotor and sensed by two eddy-current sensors. The outputs of these sensors are electronically processed to produce a signal with 200 pulses every revolution, and a signal indicating the direction of rotation. These signals are further processed electronically to produce the tachometer output.

The CLU contains an 8-bit high-speed bipolar microprocessor, specially developed for attitude control. It features microprogramming, fixed-point fractional arithmetic, expandable memories, and nearly 100 instructions. One-third of the instructions are general purpose and the others are specially created for attitude control. Double precision arithmetic is necessary for control. Add time is about 7 μs and multiply time is about 9 μs. The attitude control and logic program, including default parameters, is stored in a 4,000-word LSI programmable read-only memory (PROM). The program cycles every 0.12 s. A 512-word LSI random-access memory (RAM) is used for scratch pad and working storage. The CLU also includes circuitry for digital and analog input, digital output, driving the telemetry modulator, and receiving various command signals.

Upon initial execution of the program, the default values of selected parameters are copied from PROM to RAM. Thereafter, the values in RAM are used. Since all locations in the RAM can be changed via the command system, parameters can be modified to conform to actual conditions that exist on-orbit. This feature can compensate for wrong polarity due to crossed wires, perhaps saving the mission.

The nucleus of the air bearing platform is a 25-cm ball of hardened aluminum. In addition to the CLU, reaction wheels, and drive electronics, the platform contains six nitrogen gas thrusters with drive electronics; two gas tanks connected through a common line to a pressure regulator; two nickel-cadmium batteries; a PCM/PSK/FM command receiver and decoder; an FM telemetry transmitter; balance wheels and micrometer

heads for balancing; and a complement of optical, analog sensors consisting of a pitch-roll sensor, a yaw sensor, and a wide-angle pitch-yaw sensor.

The ground support equipment consists of the controller for the air bearing, telemetry and command equipment, strip chart recorders, magnetic tape recorders, battery chargers, an external power supply, and a minicomputer. The minicomputer is used to select, scale, and convert telemetry data, to store telemetry on IBM-compatible tape, and to synthesize 32-bit words which are sent through the command system to the CLU as address/data.

Tests and descriptions of selected results

A series of tests was performed to evaluate system performance and to study operational procedures associated with using skewed reaction wheels and a spaceborne microcomputer. All of the different modes of operation of a communications satellite, including the normal mode, simulated stationkeeping, mode transitions, various submodes, and wheel failure modes, were tested. Specific tests were conducted for the following aspects of the application:

a. transient performance in normal mode;
b. steady-state and long-term performance in normal mode;
c. zero speed crossing of the wheels;
d. momentum unloading;
e. orbit rate decoupling;
f. stationkeeping, including transitions;
g. detection and correction of wheel failure;
h. transient performance of the thruster backup system;
i. steady-state performance of the thruster backup system;
j. attitude acquisition.

The test results for wheel failure and mode transition are described in this paper; however, a complete test report has been compiled [5].

Balancing the air bearing platform

The major cause of disturbance torque on the platform is imbalance. Aerodynamic drag and air bearing imperfection cause torques less than 5×10^{-5} N·m. The offset between the center of mass of the platform and the center of pressure of the air ball must be sufficiently small to keep the imbalance torques within acceptable limits. These varied according to the type of test, from 0.002 N·m (corresponding to a 0.5-μm offset) to 0.0005

N·m. The imbalance torques are considered to be the disturbance torques for the control system designed for the test.

The platform was first manually balanced to about 0.02 N·m. It was found that the balance systematically changes by 0.06 N·m about the roll axis and 0.02 N·m about the yaw axis after a 3-hr warm-up. Therefore, it was necessary to compensate for these biases.

The attitude control system was used for further balancing. In the steady state, the rate of change in the roll and yaw axis components of wheel momentum accurately measured horizontal imbalance. The pendulousness of the platform (the amount that its center of mass is below the center of the air ball) was similarly measured, except that a target aligned 16° off the negative y_o-axis in the $x_o y_o$-plane was used for the yaw sensor. Thus, pendulousness was found by measuring the rate of change in momentum along the z-axis. The accuracy of the balancing procedure described is 5×10^{-5} N·m.

Data presentation

The tests provide data in the form of various functions of time. Sensor outputs and significant signals in the reaction wheel electronics are fed to the CLU. These variables and others that are generated internally are collected in various groupings of 15 variables to form different telemetry tables in RAM. The desired telemetry table is selected by ground command. The CLU appends the synchronization word and sends a PCM stream of sixteen 16-bit words every 0.12 s over the telemetry link. This signal is simultaneously recorded on magnetic tape (for playback), on strip charts after digital-to-analog conversion by a minicomputer, and on digital tape by the same minicomputer. The data on the digital tape are later processed on the COMSAT system computer to produce the graphs in this paper. The variables on these graphs and their labels are defined in Table 2, which also shows the quantization of each variable, as telemetered.

Pseudo-inverse distribution matrix

THEORY

The dynamic characteristics of the 3-axis control system should be determined when the wheels are operating properly, as well as when there is an undetected failure. The latter case is extremely important because the attitude of the spacecraft must be well behaved, although a wheel

TABLE 2. DEFINITIONS OF PLOTTED VARIABLES

Label	Unit	Symbol	Definition	Quantization
ROLL	deg	ϕ	Roll sensor input to the CLU	0.0015
PITCH	deg	θ	Pitch sensor input to the CLU	0.0018
YAW	deg	ψ	Yaw sensor input to the CLU	0.0012
HX CMD		H_{cx}	Momentum commands resolved in	
HY CMD	N·m·s	H_{cy}	body frame	0.00043
HZ CMD		H_{cz}		
ROLL MOM		H_x	Actual momentum of wheel array re-	
PITCH MOM	N·m·s	H_y	solved in body frame	0.00043
YAW MOM		H_z		
WH0 CMD		ω_{c0}		
WH1 CMD	rpm	ω_{c1}	Wheel velocity commands	1.27
WH2 CMD		ω_{c2}		
WH3 CMD		ω_{c3}		
WH0 VEL		ω_0		
WH1 VEL	rpm	ω_1	Wheel velocities as measured by the	1.27
WH2 VEL		ω_2	tachometers	
WH3 VEL		ω_3		
X JET WIDTH		τ_x	Pulse width commands to thrusters	
Y JET WIDTH	s	τ_y	(positive for positive torque,	0.00092
Z JET WIDTH		τ_z	negative for negative torque)	

failure is not yet detected or corrected.

Four reaction wheels arranged as shown in Figure 3b are assumed. The system can be described by the following vector-matrix equations in Laplace transformation form (after wholesale simplification for clarity):

a. Spacecraft to be controlled:

$$\begin{bmatrix} I_x S^2 & 0 & 0 \\ 0 & I_y S^2 & 0 \\ 0 & 0 & I_z S^2 \end{bmatrix} \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + \begin{bmatrix} T_{dx} \\ T_{dy} \\ T_{dz} \end{bmatrix} . \qquad (1)$$

b. Momentum control laws:

$$\begin{bmatrix} H_{cx} \\ H_{cy} \\ H_{cz} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{K_{px}}{s} + K_{vx} & 0 & 0 \\ 0 & \frac{K_{py}}{s} + K_{vy} & 0 \\ 0 & 0 & \frac{K_{pz}}{s} + K_{vz} \end{bmatrix}}_{G(s)} \begin{bmatrix} -\phi \\ -\theta \\ -\psi \end{bmatrix} . \quad (2)$$

c. Distribution matrix:

$$\begin{bmatrix} H_{c0} \\ H_{c1} \\ H_{c2} \\ H_{c3} \end{bmatrix} = D \begin{bmatrix} H_{cx} \\ H_{cy} \\ H_{cz} \end{bmatrix} . \quad (3)$$

d. Wheel servos:

$$\begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \end{bmatrix} = \frac{as}{s+a} E_4 \begin{bmatrix} H_{c0} \\ H_{c1} \\ H_{c2} \\ H_{c3} \end{bmatrix} . \quad (4)^*$$

*E_4 is the 4 × 4 identity matrix.

e. Mounting matrix:

$$\begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = C \begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \end{bmatrix} \quad (5)$$

The distribution matrix D (a 4 × 3 matrix) relates the wheel momentum commands to the control-axis momentum commands. The mounting matrix C (3 × 4) relates the control-axis torques to the wheel torques, and depends solely on the orientation of the wheels. Again, with reference to Figure 3b,

$$C = \begin{bmatrix} -\cos\gamma & \cos\gamma & 0 & 0 \\ -\sin\gamma & -\sin\gamma & -\sin\gamma & -\sin\gamma \\ 0 & 0 & -\cos\gamma & \cos\gamma \end{bmatrix} \quad (6)$$

where $\gamma = 35°$. Figure 4 shows the block diagram of the simplified system described by equations (1) through (6), where

$$I^{-1} = \begin{bmatrix} \frac{1}{I_x} & 0 & 0 \\ 0 & \frac{1}{I_y} & 0 \\ 0 & 0 & \frac{1}{I_z} \end{bmatrix} . \quad (7)$$

The manner in which the momentum commands are distributed among the wheels is not unique when more than three wheels (the minimum) are operating; that is, the distribution matrix D is not unique when four or more wheels operate. The D matrix used in the design was the pseudo-inverse given by the following formula [6]:

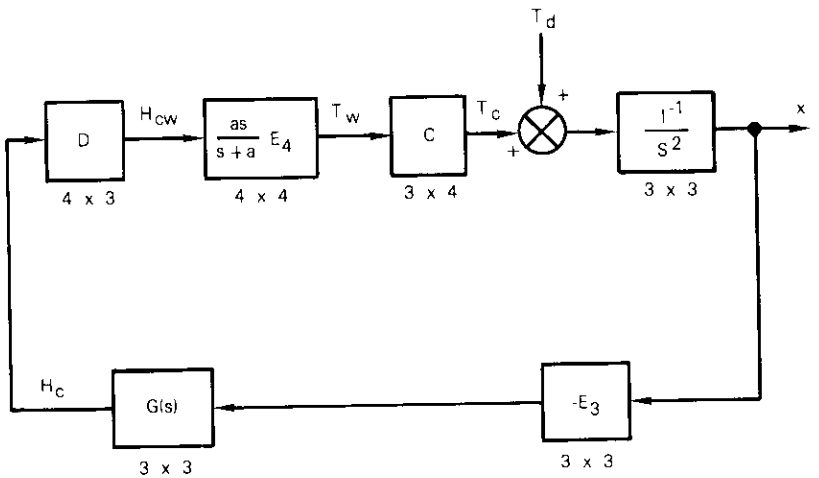


Figure 4. *Block Diagram of the Simplified Control System*

$$D = C^T(CC^T)^{-1} \tag{8}$$

if CC^T is nonsingular. This specific distribution matrix is called C^+. An advantage of the pseudo-inverse is analytic in that it minimizes the sum of the squares of the wheel momenta (Reference 2, p. 3-40). If one of the wheels is unpowered because it has failed, or is in unpowered standby, the corresponding column in C is nulled.

Examples of distribution matrices for the test are given below for two cases:

$$C^+_{\text{(4 wheels)}} = \begin{bmatrix} \frac{1}{2\cos\gamma} & -\frac{1}{4\sin\gamma} & 0 \\ \frac{1}{2\cos\gamma} & -\frac{1}{4\sin\gamma} & 0 \\ 0 & -\frac{1}{4\sin\gamma} & -\frac{1}{2\cos\gamma} \\ 0 & -\frac{1}{4\sin\gamma} & \frac{1}{2\cos\gamma} \end{bmatrix}$$

$$C^+_{\text{(wheel 0 off)}} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{\cos\gamma} & 0 & 0 \\ -\frac{1}{2\cos\gamma} & -\frac{1}{2\sin\gamma} & -\frac{1}{2\cos\gamma} \\ -\frac{1}{2\cos\gamma} & -\frac{1}{2\sin\gamma} & \frac{1}{2\cos\gamma} \end{bmatrix} .$$

The characteristic equation of the simplified system is readily shown to be

$$\left| s^2E_3 + \frac{as}{s+a} CDG(s)\, I^{-1} \right| = 0 \quad . \tag{9}$$

When D is chosen to be the pseudo-inverse matrix, equation (8) yields

$$CD = E_3 \tag{10}$$

which is the 3×3 identity matrix. Therefore, when the proper pseudo-inverse matrix is used according to the particular wheels that are operating, the control axes become decoupled. In addition, the linear response of the system does not depend on specific wheels or the number of wheels that are operating. The characteristic equation can be resolved into three equations:

$$s^2(s+a) + \frac{a}{I_x}(K_{px} + K_{vx}s) = 0 \tag{11}$$

$$s^2(s+a) + \frac{a}{I_y}(K_{py} + K_{vy}s) = 0 \tag{12}$$

$$s^2(s+a) + \frac{a}{I_z}(K_{pz} + K_{vz}s) = 0 \quad . \tag{13}$$

Each control loop is third order. They were designed to have identical characteristic roots s_1, s_2, and s_3 given by

$$s_1 = -\xi\omega_n - j\omega_n\sqrt{1-\xi^2} \tag{14}$$

$$s_2 = -\xi\omega_n + j\omega_n\sqrt{1-\xi^2} \tag{15}$$

$$s_3 = -\omega_r \quad . \tag{16}$$

Numerically,

$$\omega_n = 0.2 \text{ rad/s}, \quad \xi = 0.7, \quad \omega_r = 1.72 \text{ rad/s} \tag{17}$$

so that $s_1, s_2 = -0.14 \pm j0.14$ and $s_3 = -1.72$.

The response of each control loop is dominated by a critically damped second-order mode. The real pole equal to -1.72 rad/s can be attributed to the effect of the wheel servos (each with real pole $-a = -2$ rad/s) when they are coupled to the system. Therefore, in terms of attitude control, the wheel servos can be assumed to be perfect differentiators; *i.e.*, the reaction torque is simply the derivative of the momentum command, without any time delay.

To find the response of the system with an uncorrected wheel failure, the proper column of the mounting matrix C is nulled; however, the distribution matrix is not changed. For example, if wheel 0 fails, then

$$CC^+ = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4}\cot\gamma & 0 \\ -\frac{1}{2}\tan\gamma & \frac{3}{4} & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{18}$$

Wheel 0 is in the roll-pitch plane. Hence, as expected, its failure does not affect the response of the yaw axis. The on-axis gains are reduced, and roll and pitch become strongly coupled to each other. The dominant roots of the coupled roll and pitch axes have been calculated as

$$-0.13 \pm j0.13 \quad \text{and} \quad -0.03 \pm j0.09 \quad .$$

The most important result is that the system remains stable even with an uncorrected wheel failure. One complex pair of roots is essentially unchanged by the uncorrected failure. The associated mode is termed the fast mode. The other complex pair of roots corresponds to one-half the damping ratio and the natural frequency. The associated mode is referred to as the slow mode. For arbitrary initial conditions or disturbances, both modes appear in the response of each axis. However, for single-axis disturbances, the fast mode dominates the on-axis response, whereas the slow mode dominates the cross-axis response. Other wheel failures have effects similar to those described.

TRANSIENT RESPONSE WITHOUT UNCORRECTED FAILURES

Figure 5 shows the transient response of the attitude control system with four reaction wheels active. (The thrusters are inhibited.) This test proves the basic integrity of the control system, including the use of the pseudo-inverse law to calculate the distribution matrix. First, attitude offsets of 2° are commanded simultaneously in roll and pitch by ground command; then the offset commands are removed, again by ground command. The initial response is nonlinear because of wheel servo saturation. Initially, large step changes in momentum are commanded; however, the wheel servos can respond only at maximum motor torque. The limiting of roll at 3° is due to saturation of the roll channel of the analog-to-digital converter. The coupling into yaw is due to large products of inertia. The overshoot in roll and pitch is 60 percent. The attitude settles to within 5 percent of the desired value (2° or 0°) in 30 s.

TRANSIENT RESPONSE WITH UNCORRECTED FAILURE

Figure 6 shows the stability and transient response of the system with an uncorrected wheel failure. Power to wheel 0 was turned off while the system was configured for four reaction wheels. Thus, wheel 0 is decelerating in the graphs shown. Since failure correction was inhibited, the distribution matrix used was the erroneous all-up matrix C^+. These tests confirm the analysis that the system is stable even with a failure. The natural frequency of the linear response is 0.1 rad/s, which is near the value of 0.093 rad/s predicted by the analysis. This mode appears on both coupled axes (roll and pitch). The 5-percent settling time is 60 s. There is 35-percent overshoot in pitch and 60-percent overshoot in roll.

Stationkeeping including transitions

For stationkeeping, one or more thrusters are used to apply a net force on the spacecraft's center of mass. Also, imperfect thruster alignment, thruster imbalance, shift in the center of mass, or other factors cause these thrusters to exert nonzero torques on the spacecraft. If the disturbance torques are small enough, the reaction wheel system can be used to control attitude. However, since the torques are not known *a priori*, thrusters

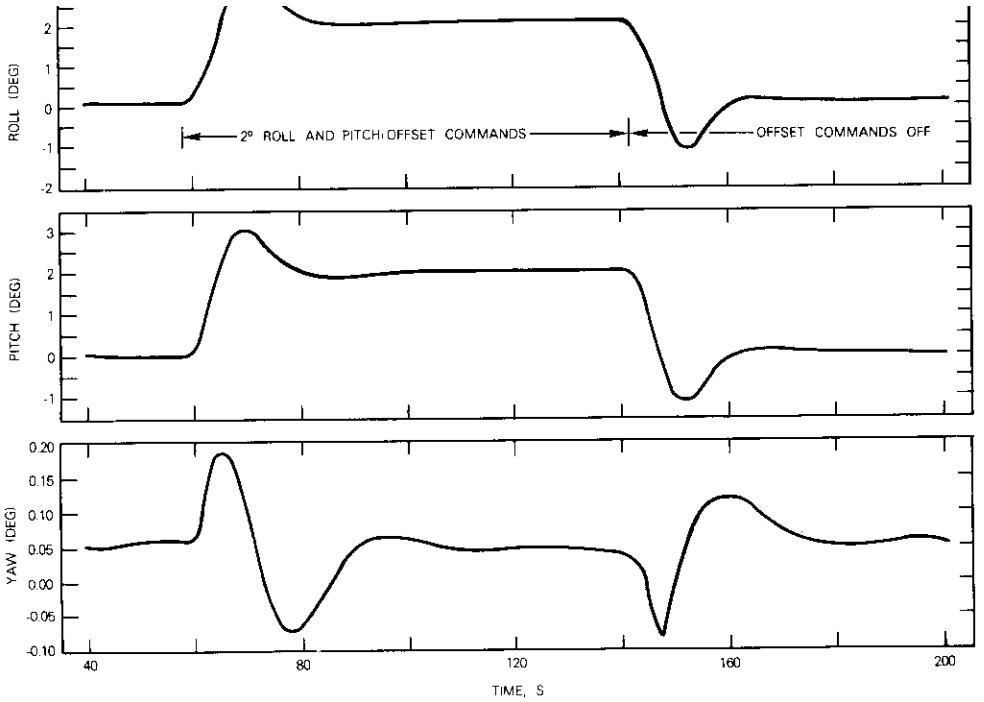


Figure 5. *Transient Response for Four Reaction Wheels*

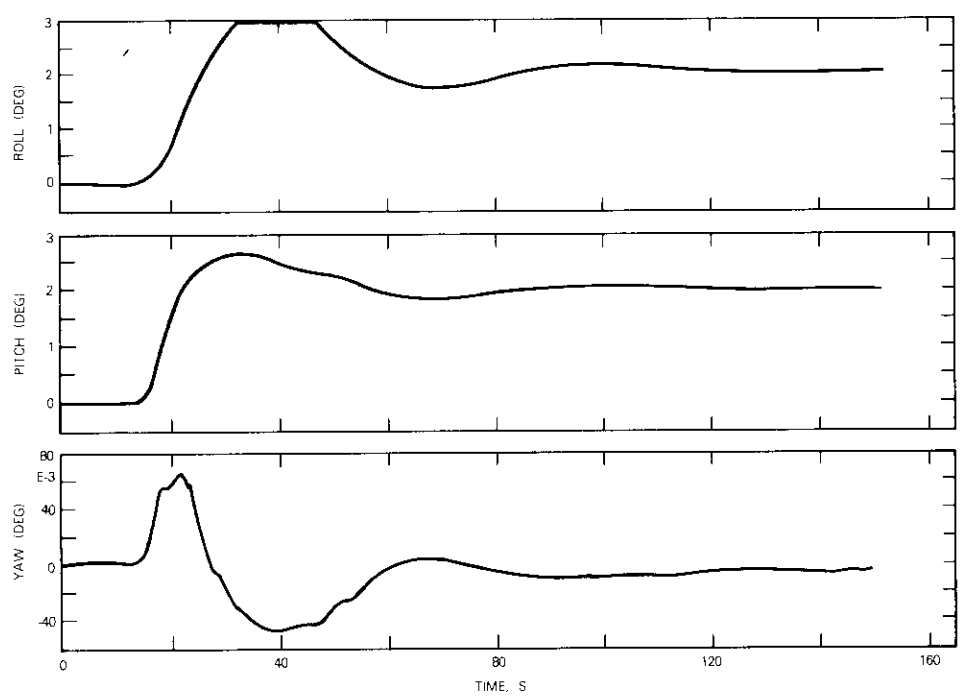


Figure 6. *Transient Response with Uncorrected Failure of Wheel 0 (offset commands)*

must be used for attitude control. Some thrusters can serve the dual function of stationkeeping and control. If a pair of north-pointing thrusters were used for north-south stationkeeping, they could be off-pulsed, *i.e.*, turned off in response to control commands, to simultaneously obtain x- or z-axis control torque and stationkeeping force. Control torques about the other axes would be obtained by on-pulsing, as usual.

When thruster channels are used for control during stationkeeping the reaction wheels are caged, *i.e.*, held at their velocities upon entry into the thruster mode. Stationkeeping was simulated during the test by firing opposed thrusters. North-south stationkeeping was simulated by firing the opposed plus or minus roll thrusters, thereby off-pulsing them for roll torque. Pitch and yaw torque was obtained by on-pulsing. Analogously, east-west stationkeeping was simulated by firing the opposed plus or minus pitch thrusters. Because of imperfect alignment and thrust imbalance, simulated stationkeeping generated disturbance torques on the platform.

In Figure 7, north-south stationkeeping is commanded (by a keydown ground command) from 61 to 86 s, and east-west stationkeeping is commanded from 116 to 126 and from 152 to 176 s. The disturbance torque can be estimated from the duty cycle of thrusting in the steady state. For example, the pulse-width command about the x-axis during north-south stationkeeping is 0.08 s. Since the period of the pulse-width modulators for the thrusters is 1.4 s, then the duty cycle of off-pulsing is 0.06. The duty cycle multiplied by the thruster torque of 0.087 N·m yields 0.005 N·m for the x-axis disturbance torque, which would lead to a roll offset of 0.082° (using the position loop gain for the roll axis, $K_{tpx} = 3.44$ N·m/rad). The actual offset is 0.09°. The disturbance torques about the y- and z-axes are not changed substantially by stationkeeping. In contrast, east-west stationkeeping in this case generates x-axis disturbance torque in addition to y-axis torque. For reference, the wheel momenta during this test are constants as follows: $H_x = 0.8$, $H_y = 0.3$, and $H_z = 0.6$ N·m-s.

It appears that the design, including the off-pulsing logic, is correct. However, to achieve better control of experimental conditions, it was necessary to hang a 5-g mass over the platform with a fine-gauge wire. The mass was then lowered onto the platform to apply x- or z-axis disturbance torque.

Figure 8 is an example of this test, which was performed according to the following steps:

a. The reaction wheels are in control.

b. At 34 s, thruster control is initiated by ground command. The momentum commands are frozen at the entry values of $H_x = 1.7$,

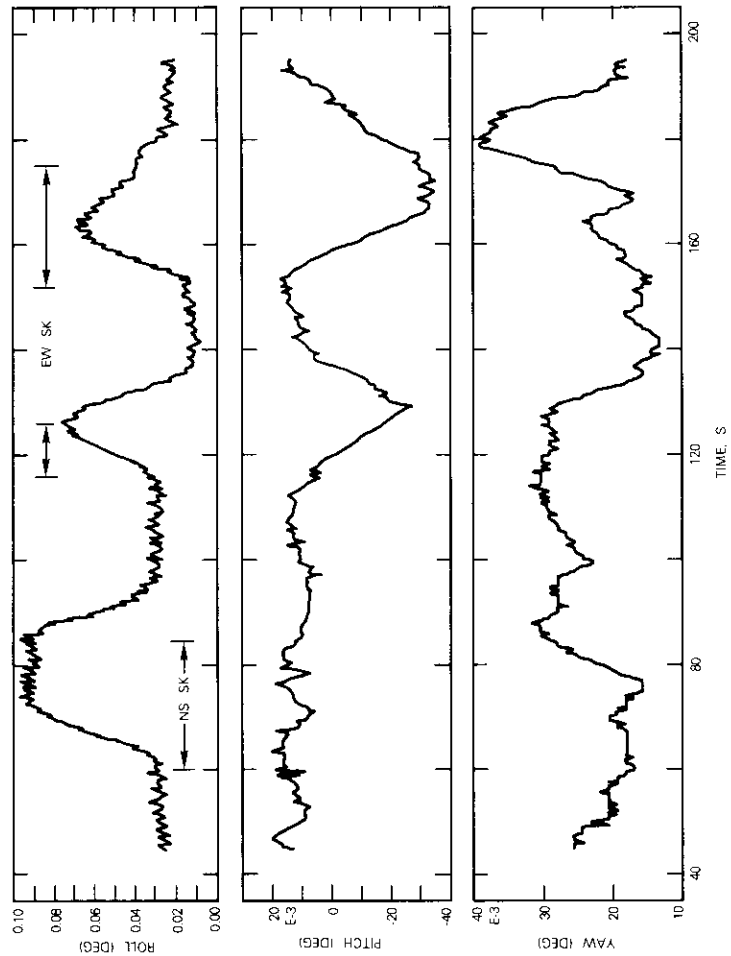


Figure 7. *Stationkeeping with Thrusters in Control (sheet 1 of 2)*

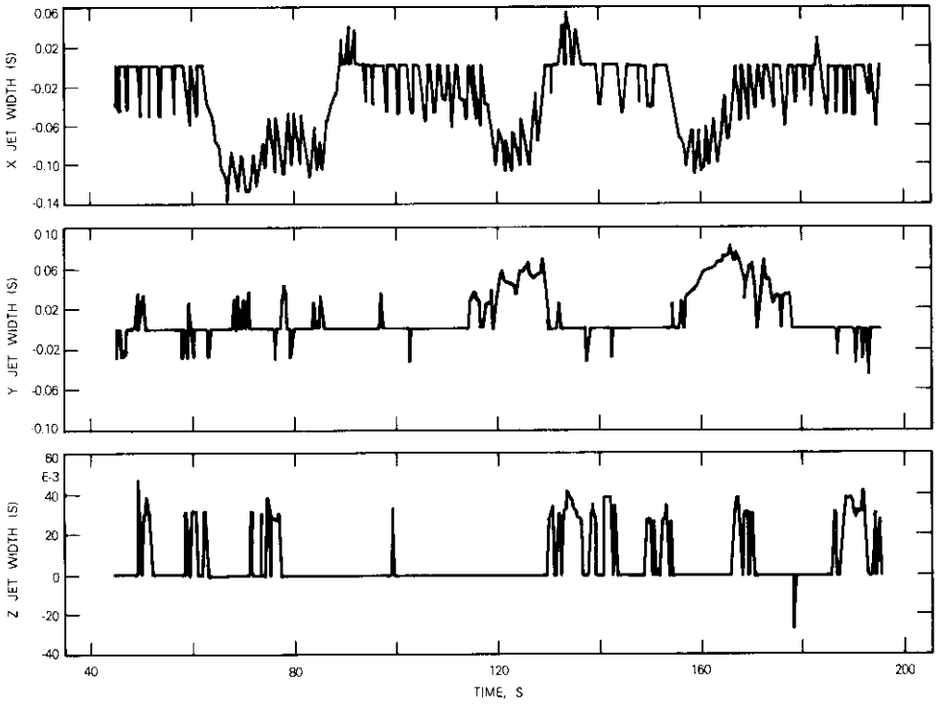


Figure 7. *Stationkeeping with Thrusters in Control (sheet 2 of 2)*

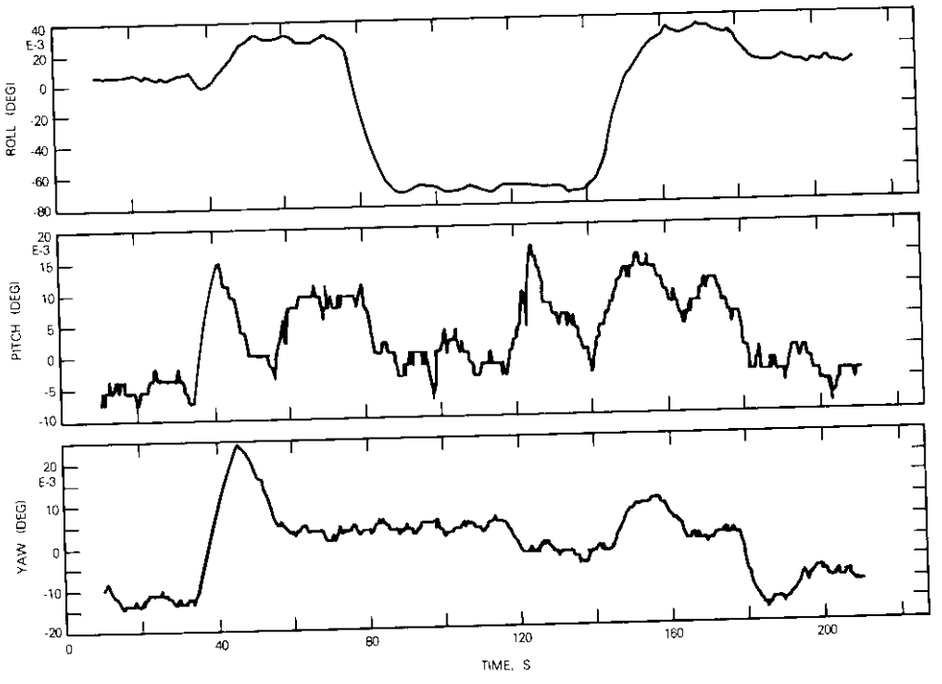


Figure 8. *Stationkeeping by Applying a Weight on the Platform (thruster control) (sheet 1 of 3)*

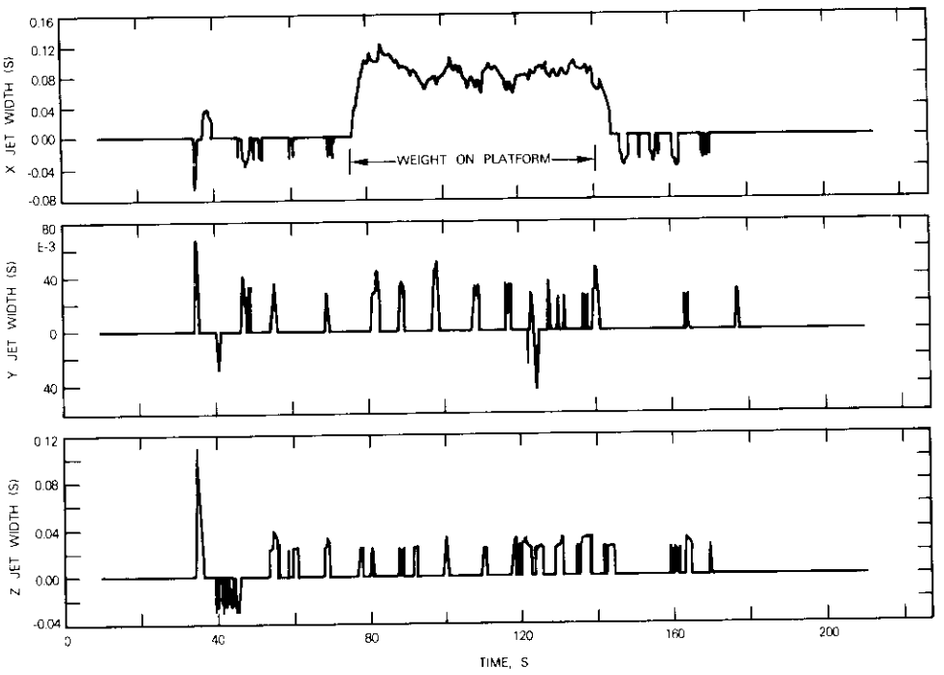


Figure 8. *Stationkeeping by Applying a Weight on the Platform (thruster control) (sheet 2 of 3)*

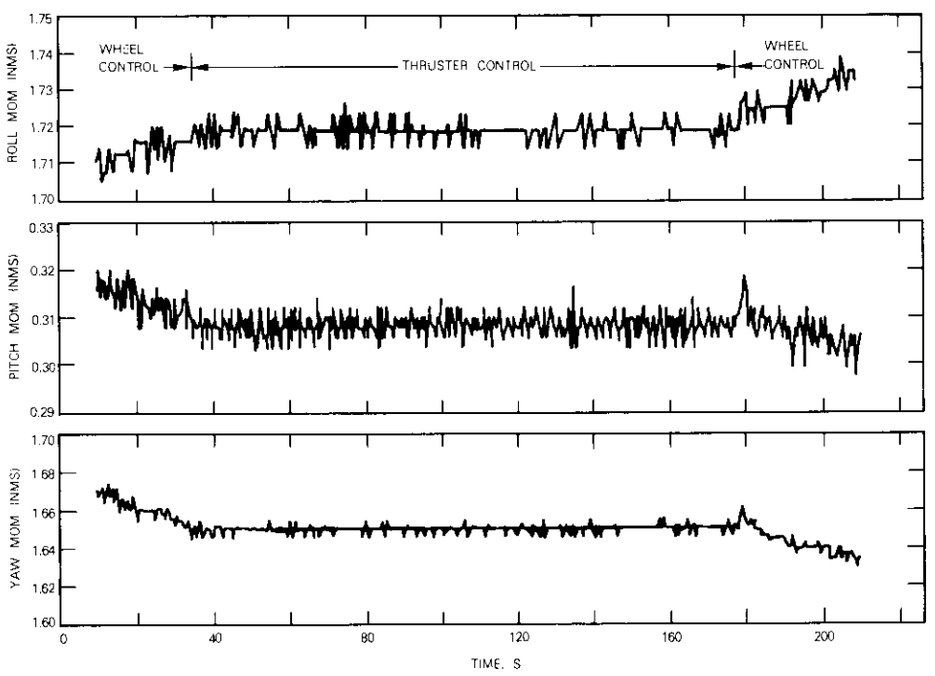


Figure 8. *Stationkeeping by Applying a Weight on the Platform (thruster control) (sheet 3 of 3)*

$H_y = 0.3$, and $H_z = 1.6$ N·m·s. The actual momenta, revealing the effect of wheel servo jitter, are shown in Figure 8, sheet 3.

c. The weight is lowered onto the platform at 75 s.

d. The weight is raised off the platform at 140 s.

e. A ground command is given to switch to reaction wheel control at 178 s.

In another test, the torque applied by the weight was measured to be -0.0071 N·m (along the x-axis). The imbalance torque along the same axis is obtained from the rate of change of H_x when the wheels are in control. This torque is 0.0003 N·m, which is verified by the value of the associated roll offset. Therefore, the net external torque during the simulated stationkeeping is $T_{dx} = -0.0068$ N·m. This would yield a roll offset of $-0.11°$ if the system were linear. The actual value is $-0.07°$. The average value of the pulse width of x-axis thrusting is given by

$$\text{Average } PW = \frac{T_{dx}\ x\ (\text{modulator period})}{(x\text{-axis thruster torque})} = 0.11 \text{ s} \quad .$$

The actual value is about 0.09 s.

It is important to note that the transients associated with the application or removal of stationkeeping torque and with mode switching are well behaved. This is due to the high bandwidth, well-damped design of the thruster and wheel control channels. Therefore, no transition mode is required in switching from stationkeeping to nadir or offset pointing.

Automatic detection and correction of wheel failure

Many types of wheel failure [2] can be categorized in two broad groups:

a. soft failures (*e.g.*, failure to respond to commands, causing the wheel to decelerate or to hold its speed, decreased motor torque, and increased friction);

b. hard failures (*e.g.*, bearing seizure and exertion of maximum motor torque).

In the system designed for test, rundown and constant speed failures were simulated. Some hard failures, such as bearing seizure, may require reacquisition of attitude from large errors and rates with the thruster control system. In any case, the objective of the design is for normal mode requirements to be met soon after the failure is corrected, even if the failed wheel is still decelerating, because the wheel can continue to decelerate for approximately one hour if it has been running near maximum operating speed.

The dynamic performance of the control system before a failure and after failure correction is the same whether it is operating with the minimum number of wheels (three) or with more than the minimum. However, the performance during the interval between a failure and its correction is slightly different for each case. This aspect of the problem will now be developed.

OPERATION WITH MINIMUM NUMBER OF WHEELS

If the normal mode of operation is with three wheels, a wheel failure results in the loss of controllability about a satellite axis between the time of failure and correction. The uncontrollable axis is defined by the normal to the plane determined by the two remaining wheels.

The design that has been tested in the laboratory can be considered as a specific example. Suppose that the system is operating with wheels 0, 1, and 2 when wheel 2 fails. Based on the way that the wheels are mounted, the failure means that yaw is uncontrolled; however, roll control is unaffected. System dynamics can be represented by the following vector-matrix equation in terms of the Laplace transformation variable s:

$$s^2 Ix + \frac{as}{s + a} CDG(s)\ x = T_d \quad . \tag{19}$$

For this case, $D = C_3^+$. Without any failure, the matrix CC_3^+ equals the identity matrix so that the control axes are decoupled (and controlled). Between the time of failure of wheel 2 and its correction, the distribution matrix C_3^+ remains in use. Then

$$CC_3^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\tan\gamma \\ 0 & 0 & 0 \end{bmatrix} \quad . \tag{20}$$

This proves that roll control is unchanged, whereas yaw is uncontrolled. It also shows that pitch control is unchanged, except that any yaw dynamics are coupled into pitch. The differential equations of motion for pitch and yaw are

$$I_y\ddot{\theta} + K_{vy}\dot{\theta} + K_{py}\theta = T_{dy} + \tan\gamma(K_{vz}\dot{\psi} + K_{pz}\psi) \tag{21}$$

$$I_z\ddot{\psi} = T_{dz} \quad . \tag{22}$$

For constant disturbance torque T_{dz}, yaw grows parabolically. Therefore, pitch also tends to grow parabolically. The failure correction interrupts these growing errors. A procedure for estimating the maximum error [Reference 2, equation (3 71)] is based on determining the time at which the speed threshold for failure is first violated, according to the following equation:

$$\epsilon_F = \frac{T_F t_T}{I_w}\left(1 + \frac{K_v t_T}{2I_v} + \frac{K_p t_T^2}{6I_v}\right) \tag{23}$$

where ϵ_F = speed threshold
T_F = net disturbance torque during the failure transient
t_T = time between failure and first violation of speed threshold
I_v = spacecraft moment of inertia about the axis of the failed wheel
I_w = spin moment of inertia of wheel rotor
K_v = velocity loop gain of the outer loop
K_p = position loop gain of the outer loop.

Figure 9 shows the test results of the example used in this section. At 71 s, power to wheel 2 is turned off by ground command, and the wheel begins to decelerate (Figure 9, sheet 3). The deceleration rate yields the friction torque, $T_f = 0.0025$ N·m, whose components are $-T_f \sin\gamma = -0.0014$ N·m along the y-axis and $-T_f \cos\gamma = -0.002$ N·m along the z-axis. The imbalance torque along the z-axis, derived from the yaw offset, $\psi_{os} = -0.035°$, is -0.002 N·m. Therefore, the net disturbance torques are $T_{dy} = -0.0014$ and $T_{dz} = -0.002 + (-0.002) = -0.004$ N·m. Between the failure and its correction, the change in yaw angle is given by

$$\psi(t) - \psi_{os} = \frac{T_{dz}}{2I_z}(t - 71)^2 \tag{24}$$

which is checked at two points with the following results:

Change in Yaw Angle, deg	$t = 81$	$t = 85$
Expected value	0.12	0.24
Actual value	0.1	0.19

At 85 s, the failure is corrected automatically with an algorithm in the control system program in the CLU by placing wheel 3 on line, by accounting for the momentum of the failed wheel, and by using the correct distribution matrix C_2^+. The last step automatically nulls the wheel 2 command

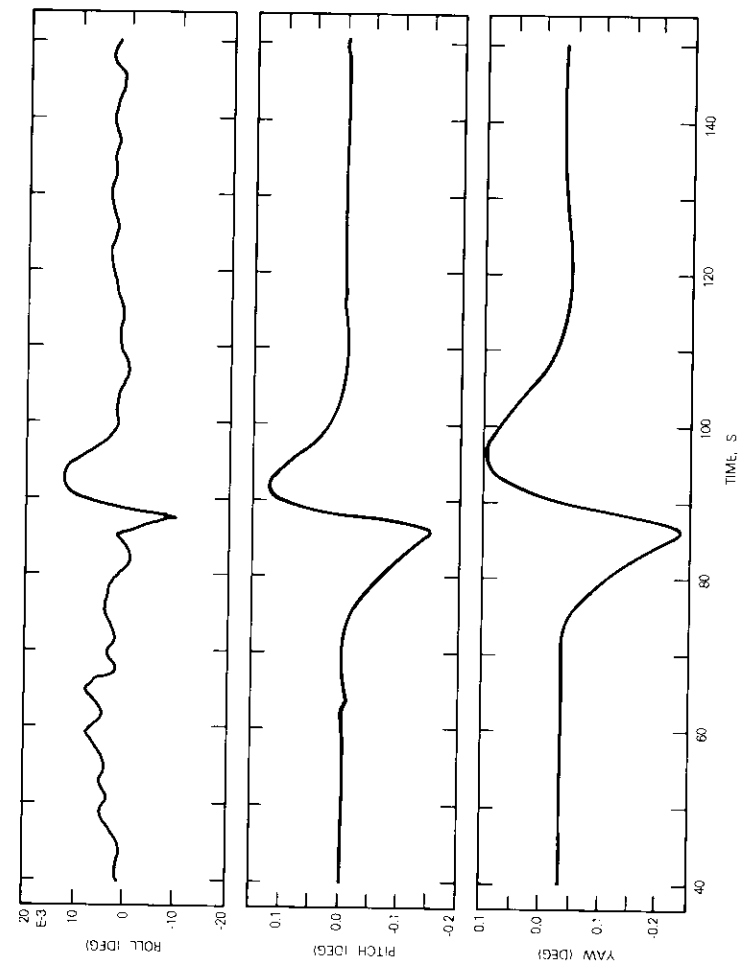


Figure 9. *Wheel Failure with Three Reaction Wheels (sheet 1 of 3)*

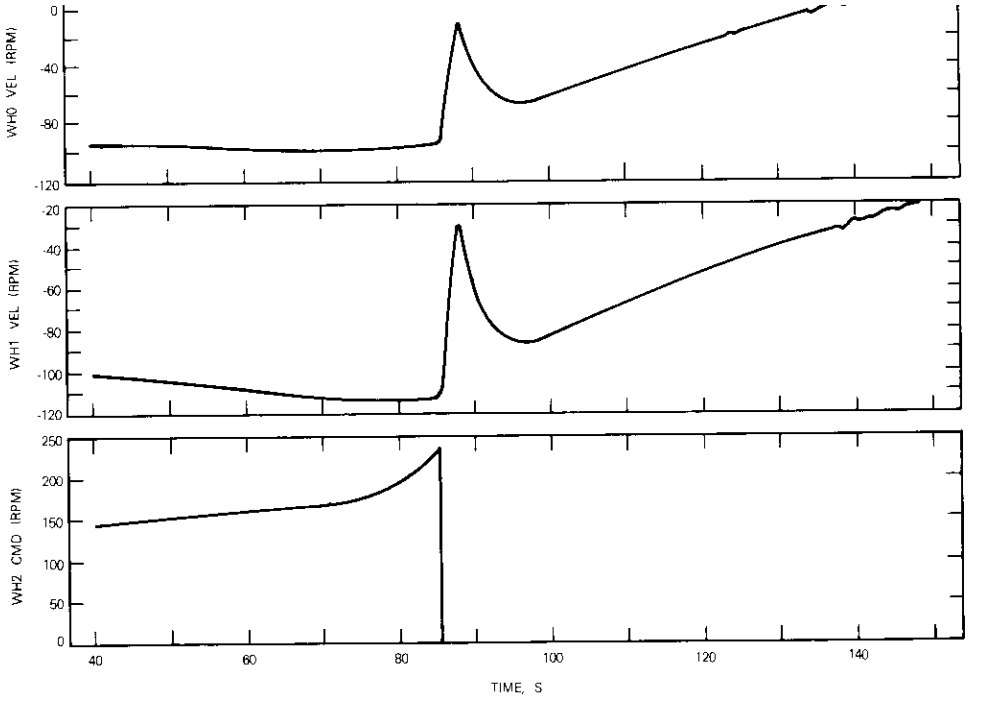


Figure 9. *Wheel Failure with Three Reaction Wheels (sheet 2 of 3)*

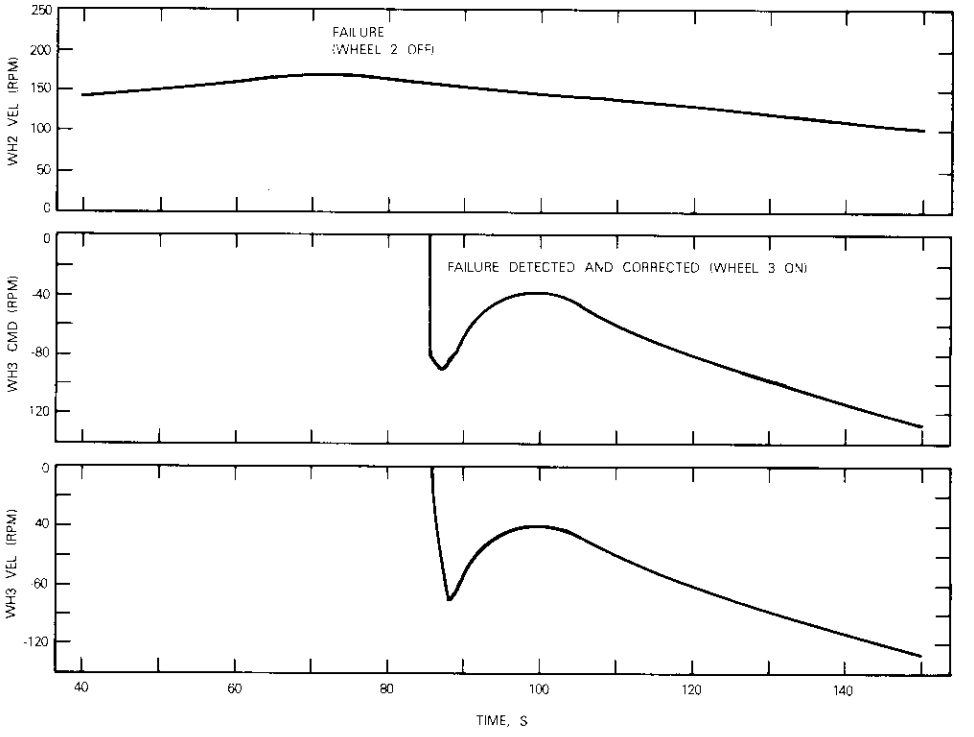


Figure 9. *Wheel Failure with Three Reaction Wheels (sheet 3 of 3)*

(Figure 9, sheet 2) and activates wheel 3 (Figure 9, sheet 3). The attitude errors are rapidly reduced as expected for a healthy reaction wheel system. The settling time is about 25 s. The effectiveness of accounting for failed wheel momentum is affirmed by the absence of pitch or yaw offsets due to friction torque. That is, performance after the correction transient is the same as before the failure.

The logic and computation for failure detection can be checked by backing up the grace period of 8 s from the correction instant to 77 s. Then the velocity of wheel 2 is 164 rpm, whereas its commanded velocity is 180 rpm. The difference is almost equal to the speed threshold of 15 rpm. Therefore, the first time that the threshold was violated was near 77 s. This would be 6 s after the failure. Solving equation (23) for t_T, using $T_F = T_{dz} = -0.004$ N·m, yields 5.7 s. Finally, the maximum yaw error is estimated by evaluating equation (24) at the correction instant. The expected value is 0.27°, compared to the actual value of 0.23°.

The differences between the expected and actual values in the preceding analyses are small. It can be concluded that this aspect of the design is well understood. Therefore, a prediction will be made which was not tested; that is, the maximum error expected for a failure wherein full motor torque of 0.1 N·m is commanded. The value $T_F = 0.1$ N·m in equation (23) can be used to determine that the time of the first threshold violation is 0.4 s after the failure. The maximum error, which occurs at correction after the grace period, would be about 2.3° *for the laboratory experiment.*

OPERATION WITH MORE THAN THREE WHEELS

When the reaction wheel system operates normally with more than three wheels, the system is stable even after an uncorrected failure, as proven earlier. Therefore, the attitude errors will tend toward their steady-state values, which may actually be attained if the disturbances are small enough.

The steady-state solution for pitch, roll, and yaw is given by the limiting form of equation (19) as the variable s approaches zero:

$$CD \begin{bmatrix} K_{px} & 0 & 0 \\ 0 & K_{py} & 0 \\ 0 & 0 & K_{pz} \end{bmatrix} \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} T_{dx} \\ T_{dy} \\ T_{dz} \end{bmatrix} . \tag{25}$$

In this equation, the appropriate column of C is nulled to account for a failed wheel, and $D = C^+$.

In Figure 10, the system is operating with all four wheels when wheel 2 is turned off by ground command at 37 s. The failure is corrected at 84 s. The new distribution matrix, C_2^+, is used instead of C^+. The command to wheel 2 is automatically clamped to zero, as shown in Figure 10, sheet 3. The response after the correction instant is typical.

The speed threshold is first violated at about 76 s, which is an unusually long time after the failure. At this time instant, the velocity of wheel 2 is −220 rpm, whereas its command is −236 rpm. The difference of 16 rpm is close to the threshold value of 15 rpm.

The attitude errors reached their steady-state values before the failure was corrected. While the failure remains uncorrected,

$$CD = CC^+ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{4} & -\frac{\tan\gamma}{2} \\ 0 & -\frac{\cot\gamma}{4} & \frac{1}{2} \end{bmatrix} . \tag{26}$$

Therefore, the solution of equation (25) is given by

$$\phi = \frac{T_{dx}}{K_{px}} \tag{27}$$

$$\theta = 2\frac{T_{dy} + T_{dz}\tan\gamma}{K_{py}} \tag{28}$$

$$\psi = \frac{3T_{dz} + T_{dy}\cot\gamma}{K_{pz}} . \tag{29}$$

The rate of change of the command to wheel 2 prior to failure is +0.0022 N·m. The rundown torque after failure is +0.0019 N·m. (Torque on wheel 2 is considered positive in the direction shown in Figure 3 b.) Therefore, during failure, there is a deficiency equal to 0.0003 N·m in torque on the body, which can be considered as the disturbance torque. Its components are

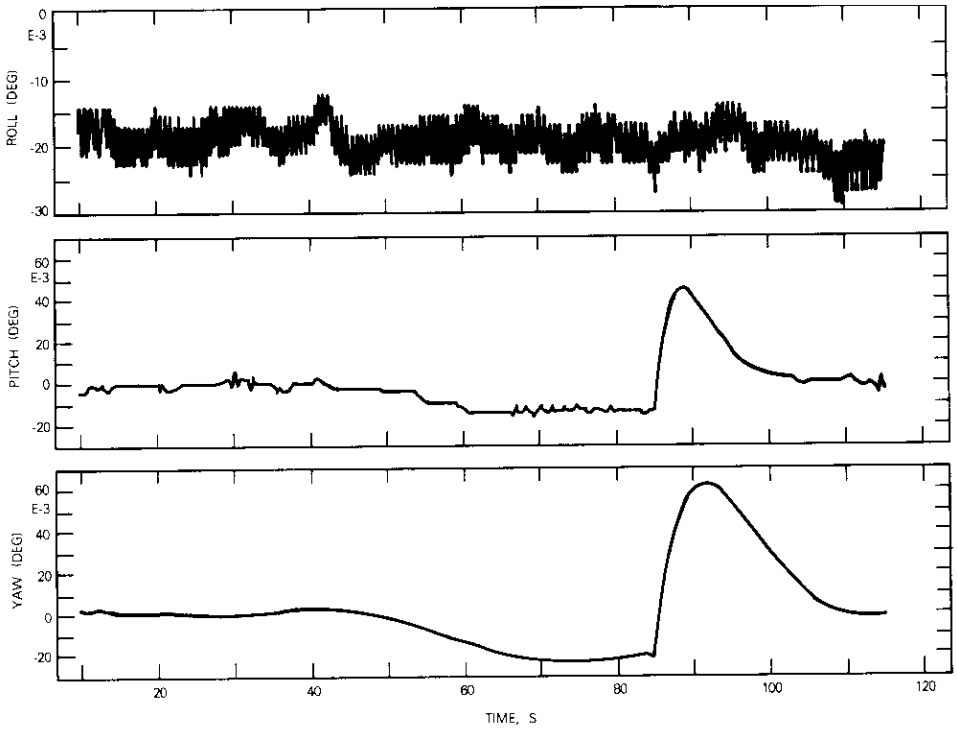


Figure 10. *Wheel Failure with Four Reaction Wheels (Example 1) (sheet 1 of 3)*

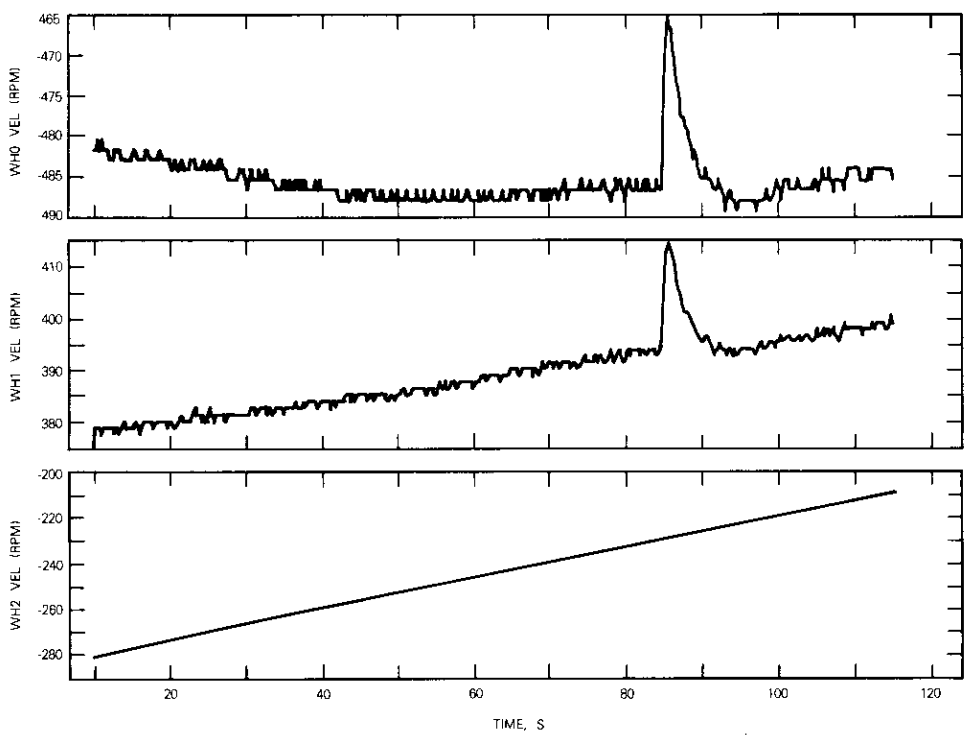


Figure 10. *Wheel Failure with Four Reaction Wheels (Example 1) (sheet 2 of 3)*

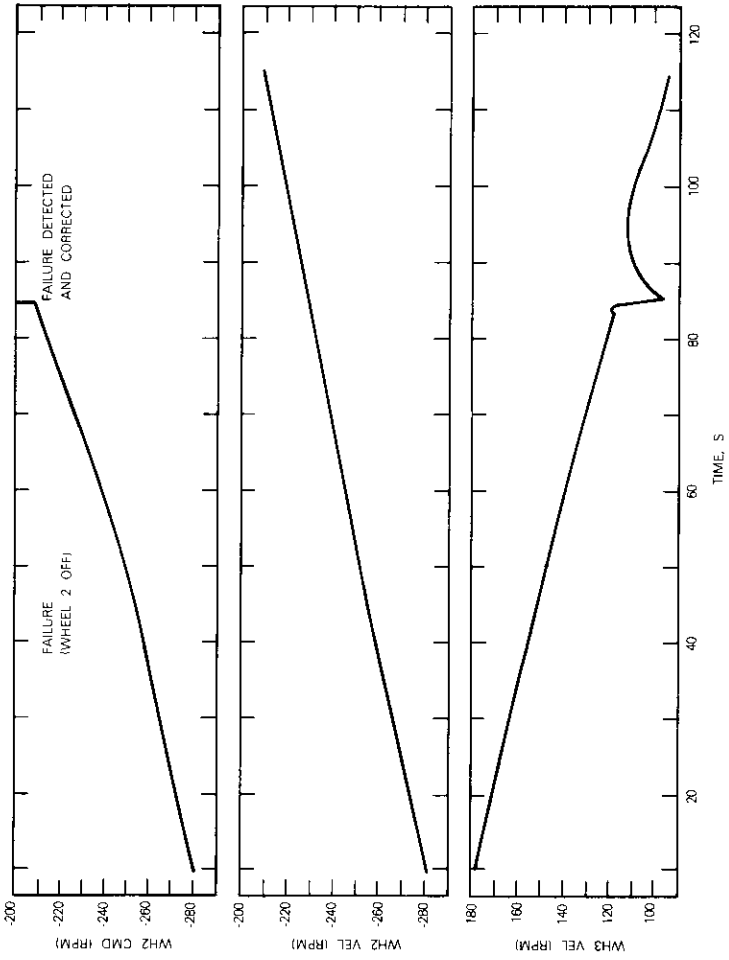


Figure 10. *Wheel Failure with Four Reaction Wheels (Example 1) (sheet 3 of 3)*

$$T_{dy} = -0.0003 \sin 35° = -0.00017 \text{ N·m}$$

$$T_{dz} = -0.0003 \cos 35° = -0.00024 \text{ N·m} \quad .$$

Using these values in equations (28) and (29) yields steady-state values of $-0.01°$ in pitch and $-0.017°$ in yaw. The actual values are $-0.013°$ and $-0.02°$, respectively. Therefore, the errors during the failure transient can be readily calculated; however, this case is special because it corresponds to insignificant net disturbance torque on the spacecraft body. The more general case of large disturbance torque is illustrated by the following example.

At the start of the run shown in Figure 11, the velocity of wheel 0 tracks its command accurately. Power to the wheel is turned off at 192 s. The velocity and the command to wheel 0 diverge rapidly as shown in Figure 11, sheet 2. In only 7 s after the failure, the speed threshold is violated. After an 8-s grace period, the failure is corrected at 207 s.

The peak errors during the failure transient are 0.17° in roll and 0.1° in pitch. The growth of the attitude errors (in roll and pitch only) before the correction can be characterized as ramps. This growth is contrasted with the parabolic growth of errors for the 3-wheel case. However, since the peak errors associated with the 4-wheel case are not sufficiently small, the 4-wheel case is no more advantageous than the 3-wheel case.

Summary of performance

The techniques developed in this project were applied to the spacecraft with characteristics defined for the typical flight application. The results of this projection to flight (Reference 2, Table 4-2) and of the air bearing testing are shown in Table 3.

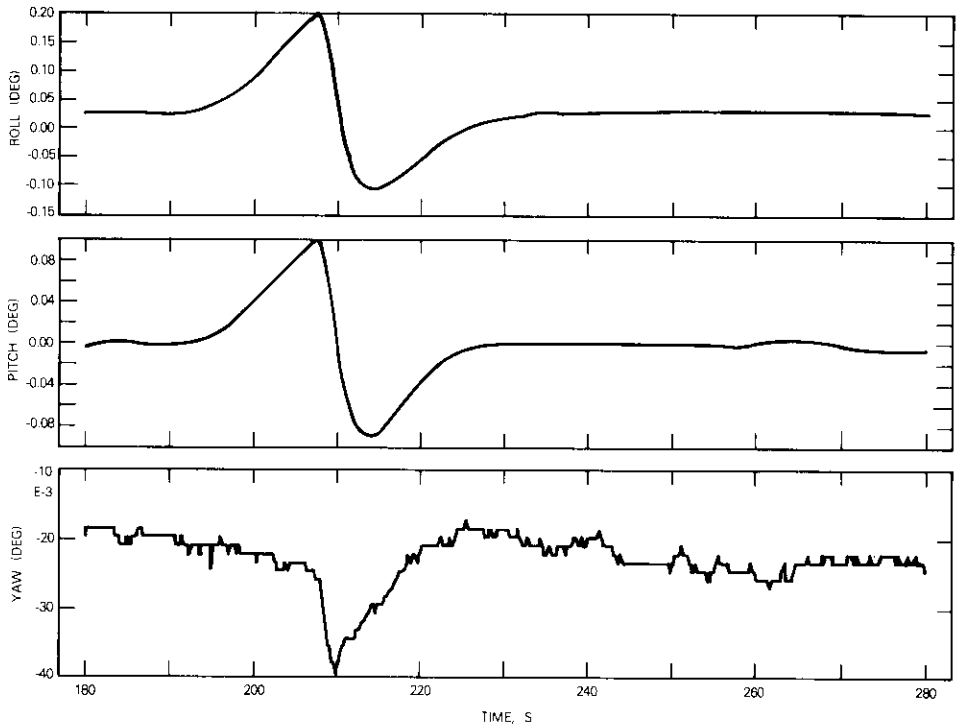


Figure 11. *Wheel Failure with Four Reaction Wheels (Example 2) (sheet 1 of 2)*

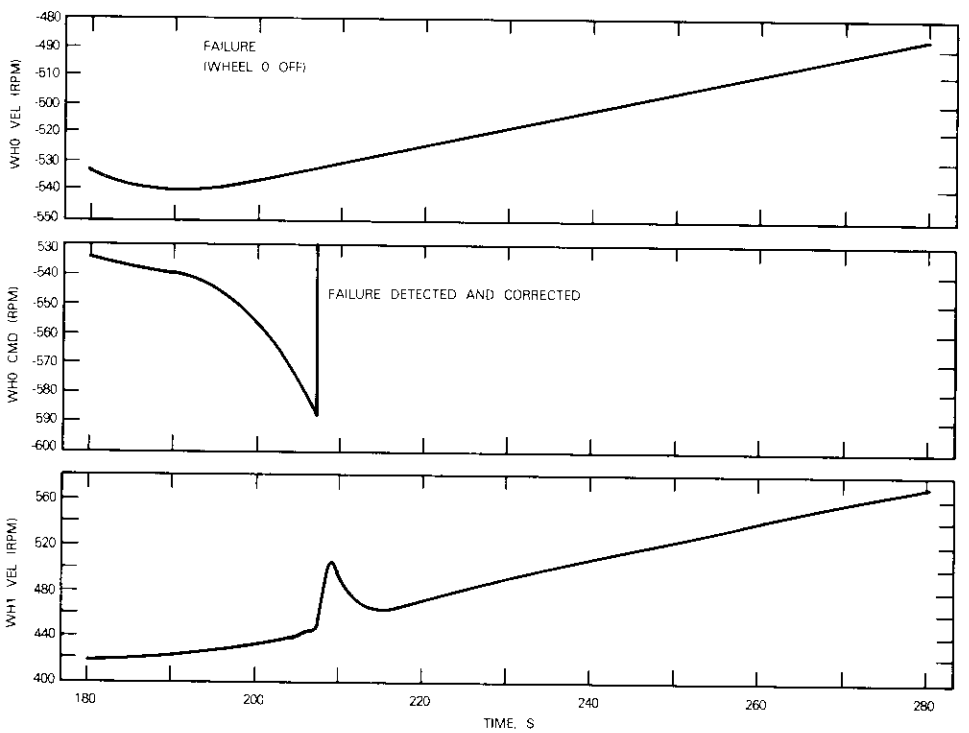


Figure 11. *Wheel Failure with Four Reaction Wheels (Example 2) (sheet 2 of 2)*

TABLE 3. SUMMARY OF PERFORMANCE IN NORMAL MODE

Parameter	Flight Application (3σ or peak, deg)		Laboratory Experiment (3σ or peak, deg)	
	Roll or Yaw	Pitch	Roll or Yaw	Pitch
Disturbance Torques	0.002	0.0002	0.04	~0
CLU Quantization	0.003	0.003	0.004	0.008
Tachometer Quantization	0.0008	0.0002	0.003	0.002
Zero Speed Crossing				
(0.018-N·m stiction)	0.002	0.001		
(0.002-N·m stiction)			0.003	0.002
Stationkeeping			0.04	0.04
(0.16-N thrusters)	0.015	0.001		
Momentum Unloading	0.012	0.012	0.022	0.02
Benign Failure (3 wheels)				
(0.018-N·m friction)	0.02	0.012		
(0.003-N·m friction)			0.24	0.15
Maximum Torque Failure (3 wheels)				
(0.12-N·m motor)	0.047	0.028		
(0.1-N·m motor)			2.3	1.3

Conclusions

The work on the skewed reaction wheel concept, ranging from analysis to the air bearing test, has proved that the concept offers an attractive means of achieving high attitude accuracy and reliability in future communications satellites. The correspondence between the results of the analyses, simulations, and tests has been very good; as usual, the process was an iterative one. Some adjustments to the design, computer simulations, and analyses were necessitated or illuminated by the tests.

The design is robust in that the transition from stationkeeping to nadir or offset pointing is simple and accurate, 3-axis response to unexpected upsets is rapid, the system is easy to operate, and the design is insensitive to parameter variations.

Momentum control, in contrast to traditional torque control, has many advantages in terms of design and operation (Reference 3, p. 5-2). For example:

a. it is not required to measure or estimate attitude rates;

b. the design is desensitized to nonlinearities and variations in wheel characteristics;

c. failure detection is rapid and correction is accurate.

Selection of the pseudo-inverse for the distribution matrix contributes to high performance and produces a design with high control integrity; as long as at least three wheels are operating, the system is stable even with uncorrected failures.

The use of wheel pre-emphasis is required for high attitude accuracy during momentum unloading. Compensation of precession torques due to the orbit rate (with ORD terms), which is required for the air bearing testing, may be necessary for flight, depending on the values of certain parameters. If it is required, the use of momentum control and an onboard microcomputer allows ORD terms to be easily implemented.

Initially it was thought necessary to operate with more than the minimum of three wheels to prevent a pointing outage. Analyses and tests proved that, with three wheels, peak attitude errors are acceptable (for flight) even for a failure as severe as a maximum-motor-torque failure. However, proper failure detection and correction are required. For a bearing seizure, even operation with more than three wheels cannot prevent an outage. Therefore, normal operation with three reaction wheels is recommended.

The importance of feeding the momentum of a failed wheel into the input of the distribution matrix was proven. This is also necessary when wheels that are on standby are temporarily activated for lubrication or monitoring. The validity of the design was demonstrated by the fact that the performance after failure correction was the same as before failure.

The CLU, *i.e.*, the microprocessor, memories, and the associated input/output circuits, is well matched to the skewed reaction wheel concept. The hardware features (*e.g.*, the quantization and the instruction set) and the design of the control system program (including sampling frequency and scaling) were properly chosen. The program is efficient in that it requires less than 4,000 words of memory and permits a realistic time margin. Such efficiency is important for high reliability and low programming cost.

The skewed reaction wheel concept for attitude control has several advantages. The most important of these are that yaw accuracy can be as good as the roll and pitch accuracies (perhaps for intersatellite links), overall mass and cost could be less than that for other types of system, certain types of wheel failure need not cause pointing outage, and opera-

tion during the mission life is straightforward (*e.g.*, transitions between stationkeeping and nadir or offset pointing are simple). This kind of system may be used for its fast yaw response even if high yaw accuracy is not required; in this case a medium-accuracy low-cost scheme for detecting yaw error would be attractive.

TRW is qualifying the programmable control processor for use on spacecraft. Also, the use of the momentum control approach will probably increase in the future.

Acknowledgment

Part of the hardware and software for the skewed reaction wheel system was designed, fabricated, tested, integrated into the air bearing facility, and acceptance tested by TRW under INTELSAT *R &D contract. Sperry Flight Systems served as a major subcontractor for the reaction wheels. The author would like to acknowledge the outstanding work of TRW and Sperry personnel in conjunction with the contractual work that preceded the tests. Special acknowledgment is given to Dr. A. W. Fleming, who served as project manager for TRW. The author wishes to give special thanks to J. Austin, who provided technical support during these tests and wrote the minicomputer program for command, telemetry, display, and digital tape recording; and R. Curtis, who provided technical support during the tests.*

List of symbols

a	Loop gain of wheel servo
C	Wheel mounting matrix
C^+	Pseudo-inverse distribution matrix for all wheels operating
C_i^+	Pseudo-inverse distribution matrix for ith wheel off or failed
D	Generic distribution matrix
E_n	$n \times n$ identity matrix
G	3×3 matrix of control laws
$H_c = [H_{cx}H_{cy}H_{cz}]^T$	Vector of momentum commands resolved in the body-fixed frame (T represents transposition)
$[H_xH_yH_z]^T$	Vector of actual momentum of wheel array resolved in the body-fixed frame
$H_{cw} = [H_{c0} \ldots H_{cm}]^T$	Vector of momentum commands along m wheel axes

I	Matrix for moments of inertia
I_x, I_y, I_z	Moments of inertia about body-fixed axes
I_w	Spin moment of inertia of wheel rotor
K_p	Generic position loop gain for momentum control (specifically K_{px}, K_{py}, K_{pz})
K_v	Generic rate loop gain for momentum control (specifically K_{vx}, K_{vy}, K_{vz})
K_{tp}	Generic position loop gain for thruster control
K_{tv}	Generic rate loop gain for thruster control
s	Laplace transformation variable
t	Problem time or run time
$T_c = [T_xT_yT_z]^T$	Vector of control torques (from wheel or thrusters) resolved in the body-fixed frame
$T_d = [T_{dx}T_{dy}T_{dz}]^T$	Vector of generic disturbance torque resolved in the body-fixed frame
$T_w = [T_0 \ldots T_m]^T$	Vector of wheel reaction torques along m wheel axes
γ	Mounting angle for reaction wheels
ϵ_F	Speed threshold for detecting wheel failure
θ	Pitch angle
θ_c	Pitch offset command
τ_x, τ_y, τ_z	Pulse width commands to thrusters (positive for positive torque, negative for negative torque)
ϕ	Roll angle
ψ	Yaw angle
$[\omega_{c0} \ldots \omega_{cm}]^T$	Vector of wheel velocity commands along m wheel axes
$[\omega_0 \ldots \omega_m]^T$	Vector of wheel velocities as measured by the tachometers
ω_o	Orbit rate

References

[1] Statement of Work (Exhibit A), INTELSAT Contract IS-759, "Skewed Reaction Wheel-Thruster Attitude Control and Stabilization," January 1976.

[2] A. Fleming, "Analysis and Conceptual Design for a Typical Flight Application and Laboratory Experiment," TRW Report for Contract IS-759, August 1977.

[3] A. Fleming, "A Skewed Reaction Wheel/Thruster Momentum Management Attitude Control System," TRW Final Report for INTELSAT Contract IS-759, October 1977.
[4] A. Ramos, "Air Bearing Platform Testing of a Double-Gimbaled Momentum Wheel Attitude Control System," *COMSAT Technical Review*, Vol. 5, No. 1, Spring 1975, pp. 53–72.
[5] A. Ramos, "Air Bearing Testing of a Skewed Reaction Wheel System for Attitude Control," COMSAT Laboratories Technical Report CL-TR-3-78, August 8, 1978.
[6] T. N. E. Greville, "Some Applications of the Pseudo-Inverse of a Matrix," *SIAM Review*, Vol. 2, No. 1, January 1960, pp. 15–22.

Alberto Ramos received a B.S. degree in electrical engineering from the University of the Philippines in 1958, and an M.S. degree in electrical engineering from Purdue University in 1961. In October 1967 he joined COMSAT, *where he is presently Assistant Manager of the Stabilization and Structures Department. He is a member of the IEEE and the IEEE Control Systems Society.*

Index: protocol, high-level data link control, time-division multiplexing, mathematical model

Performance of data link control protocols over synchronous time-division-multiplexed communications channels

A. K. KAUL

(Manuscript received January 1, 1978)

Abstract

Analytic models are presented for predicting the throughput efficiency obtainable with high-level data link control (HDLC) or similar protocols using a reject (go-back-N) error recovery mode in synchronous time-division-multiplexed (STDM) communications channels. These models, which are presented for systems transmitting a data frame in multiple bursts as well as those transmitting multiple data frames per burst, are applicable to both terrestrial time-division-multiplexing (TDM) channels and satellite time-division multiple-access (TDMA) channels. The computations needed to apply the models are summarized, and a simple satellite TDMA system is presented to illustrate the effect of various system parameters on throughput efficiency.

Introduction

The performance of high-level data link control (HDLC) protocols over point-to-point communications links has been examined, with particular emphasis on satellite communications links [1]. The models developed for

this purpose are applicable to similar protocols such as the advanced data communications control procedure (ADCCP) and level 2 of the X.25 protocol, and assume that a communications channel is available as a continuous server during the transmission period. For links providing discontinuous transmission capability at a transmitting station, these models are not valid, although in some cases the performance may be approximately the same.

Discontinuous transmission service can be attributed to several factors. In the most common case, a single communications channel is time-shared by several terminals so that the periods when the channel is not available to a terminal correspond to the time that the channel is servicing other terminals. This technique is generally classified as time-division multiplexing (TDM). When the transmission time and duration of bursts from individual terminals are predetermined, the multiplexing can be considered synchronous as in terrestrial TDM and satellite TDMA systems; otherwise the multiplexing may be considered asynchronous or demand assigned as in polling systems.

This paper develops models for the throughput performance of HDLC-type protocols for the asynchronous response mode (ARM) over synchronous time-division-multiplexed (STDM) channels. The application of these models is illustrated using a hypothetical satellite TDMA system with four stations.

Mathematical model

The development of the mathematical model is based upon several assumptions concerning the format and framing of the data transmitted. It is assumed that all data to be transmitted over a particular transmission channel are organized into superframes as shown in Figure 1. The data within a superframe are framed by HDLC-type "flag" sequences (01111110) at each end, with a 16-bit (extendable) address and control field, a 16-bit cyclical redundancy check (CRC) field, and a 16-bit overhead for additional controls for a total overhead of 64 bits. Zero bit stuffing is assumed after each occurrence of five consecutive 1's within a superframe to prevent inadvertent occurrences of flag sequences.

For communications channels with point-to-point capability the data carried by a superframe are intended for a single destination. However in point-to-multipoint channels, individual superframes may be addressed to a single destination, or the superframe itself can be multidestinational, with data to each destination organized in the form of basic logic units

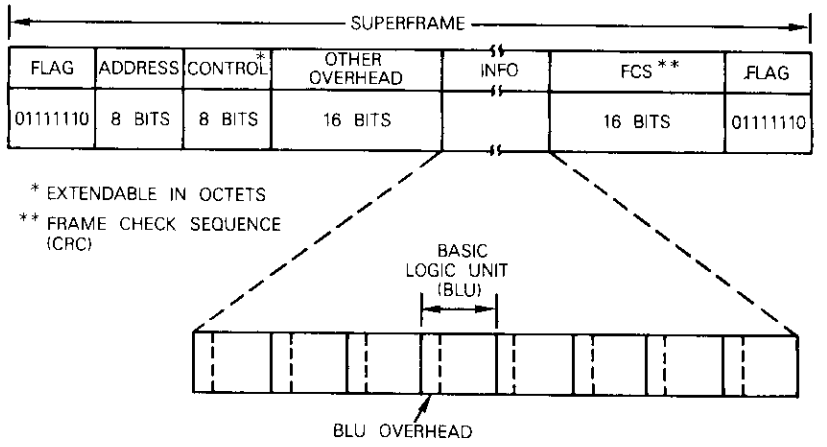


Figure 1. *Superframe Format and BLU Structure (Basic logic units may all be for the same destination or for different destinations.)*

(BLUs) within the superframe. The latter assumption is made in this paper, so that transmission throughput can be considered at the superframe level. For superframes with a multidestinational BLU structure, the performance at the superframe level is determined for each link (source-to-destination station pair) at the superframe level, and the effect of BLU structures can be considered independently.

A random frame error process is assumed for the channel, and only the HDLC reject mode [go-back-N automatic repeat request (ARQ)] is assumed for error recovery. The selective reject mode is not discussed. The "checkpoint" mechanism using time-outs is not included since, in a good system design, this mechanism should exist only for extra protection against loss of data and should not be invoked too often.

A link L, defined as a source-to-destination data path, can be assumed to have a complement L', which provides the reverse path for acknowledgments or for data flow in the opposite direction. Figure 2 shows a link and its complement in a TDM system, with transmission occurring in bursts having a definite periodicity (cycle time). A typical burst structure is also illustrated, showing the guard time and preamble for burst acquisition. Bursts for L and L' can be of unequal size.

Two separate cases are examined: first, the bursts are assumed to be short so that more than one burst is required to transmit an entire superframe; in the second case, it is assumed that the burst is long to enable the transmission of one or more complete superframes. Figure 3 illustrates both cases.

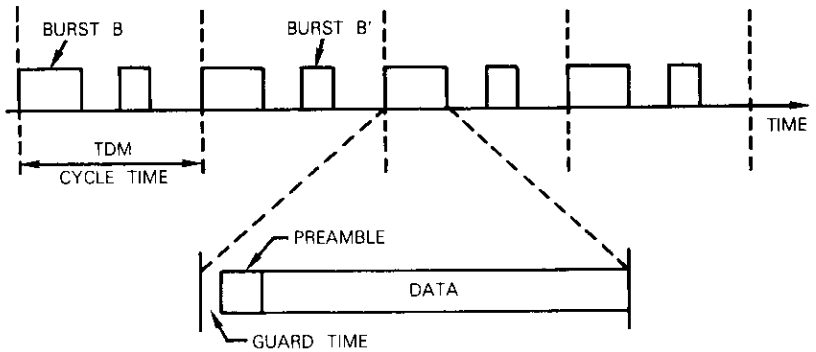


Figure 2. *Representation of a TDM System With Bursts B Containing Link L and B' Containing Complementary Link L'*

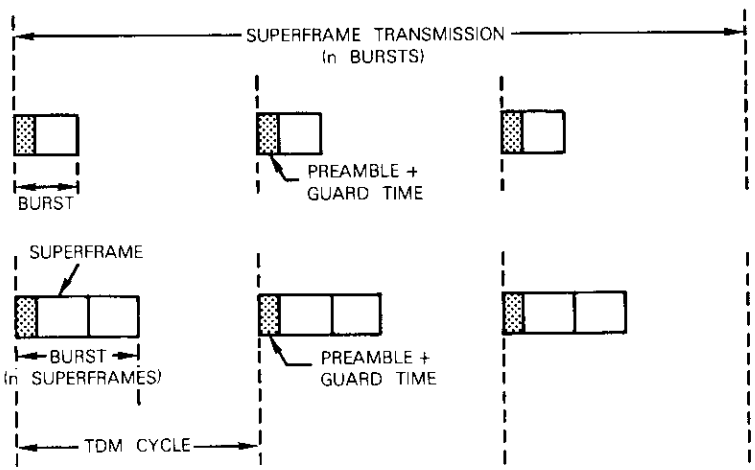


Figure 3. *Illustration of the Case of Multiple Bursts per Superframe and the Case of Multiple Superframes per Burst*

Multiple bursts per superframe

Two limiting cases are examined. In the first case (Figure 4), responses are "piggybacked" to superframes sent in the reverse direction. In the second case (Figure 5), a one-way data flow is assumed so that short response frames in the reverse direction can be accommodated in a single burst. In addition to the fact that the effective round-trip delay can be

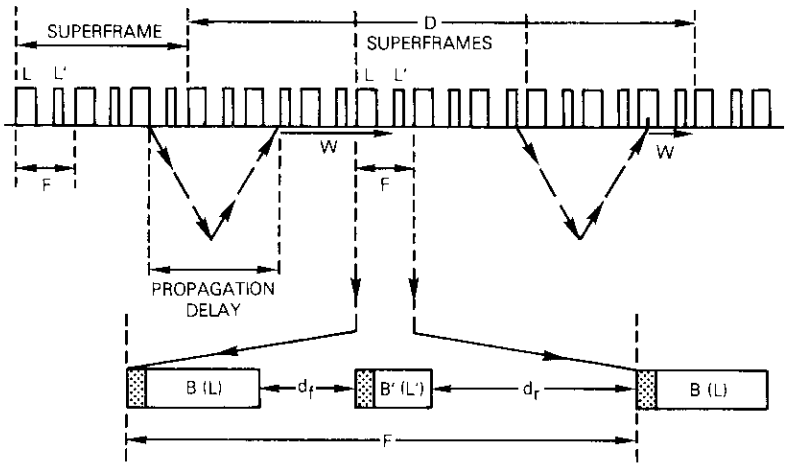


Figure 4. *Representation of TDM System Case of Multiple Bursts per Superframe Operation With Acknowledgments Piggybacked on Data*

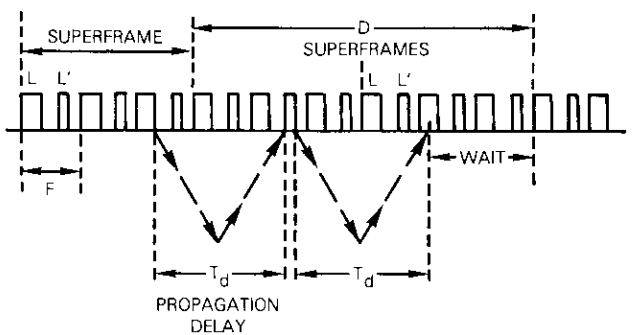


Figure 5. *Representation of TDM System Case of Multiple Bursts per Superframe Operation With Short Acknowledgments*

different in each case, the block transmission success probability is lower when responses are piggybacked on the longer data superframes over the reverse channel, resulting in a non-negligible error rate for responses for long superframe sizes. These consequences will be evident as the models are developed.

Figures 4 and 5 represent a TDM system with link L and its complement L', in which n bursts of L are required to transmit a single superframe.

The first step is to determine the effective round-trip delay, D, in superframes, *i.e.*, the total number of superframes transmittable from the end of transmission of the current superframe to the start of the superframe transmission immediately following receipt of an acknowledgment at the transmitting station. The first case implies that an acknowledgment is piggybacked to a superframe transmitted in the reverse direction. If it is assumed that n bursts per superframe are also required for the L' link and that superframe transmissions for both L and L' are initiated in the same cycle F, then D may be computed as follows:

$$N_f = \left[\frac{T_d - d_f - P - G}{F}\right]_+ \geq 0 \tag{1}$$

$$N_R = \left[\frac{T_d - d_R - P - G}{F}\right]_+ \geq 0 \tag{2}$$

$$n_f = \left[\frac{N_f + n - 1}{n}\right]_- \geq 0 \tag{3}$$

$$n_R = \left[\frac{N_R + n - 1}{n}\right]_- \geq 0 \tag{4}$$

and

$$D = n_f + n_R \tag{5}$$

where T_d = one-way propagation delay (270 ms for satellites)
P = preamble in seconds
G = interburst guard time in seconds
F = TDM cycle time in seconds
d_f = time interval between end of L burst and start of next L' burst (see Figure 4)
d_R = time interval between end of L' burst and start of next L burst (see Figure 4)
n = number of bursts per superframe
$[x]_+ = x$ if $x =$ integer or next higher integer if $x \neq$ integer
$[x]_- = x$ if $x =$ integer or truncated integer if $x \neq$ integer.

For the second case of one-way data flow with short response frames (Figure 5), equations (1) and (2) are still valid; however, equations (3)–(5) must be replaced by

$$D = \left[\frac{N_f + N_R}{n}\right]_+ \geq 0 \quad . \tag{6}$$

A close examination of Figures 4 and 5 at the superframe level reveals a similarity to a full-duplex asychronous response mode (FDX-ARM) continuous server link, with a constant round-trip delay of D [equation (5) or (6)] superframes. In the appendix, the throughput efficiency for this case is shown to be

$$\eta_o = \frac{S(1 - S^m)}{S(1 - S^m) + (1 + D - mS^m)(1 - S)} \tag{7}$$

where $m = \min(N, 1 + D)$; N represents the "maxout" count, or the size of the transmit buffer in superframes, and D is given by equation (5) or (6). For the large* buffer case ($N > 1 + D$), this equation reduces to

$$\eta_o = \frac{S}{S + m(1 - S)} \tag{8}$$

with $m = 1 + D$, which is identical to the expression for a continuous server link in an FDX-ARM operation. However, the expression for D is different in the two cases. For the continuous server,

$$D \simeq \left[\frac{T_R}{t}\right]_+ = \left[\frac{2T_d}{t}\right] \tag{9}$$

where t is the transmission time per superframe. For the TDM server, the evaluation of D also requires information concerning the parameters d_f, d_R, F, P, and G of the TDM system.

When $F \ll T_d$, $N_f \simeq N_R \simeq T_d/F$ are large numbers and hence $n_f \simeq n_R \simeq T_d/nF$, yielding $D \simeq 2T_d/nF \simeq 2T_d/t$ for both cases, which is the value for the continuous server case. *Therefore, for short-frame (i.e., $F \ll T_d$) TDM systems operating in the multiple burst per superframe mode, the performance is equivalent to that of a continuous server link in an FDX-ARM operation.*

In the other extreme case, if $F \gg T_d$ and $d_f > T_d$ and $d_R > T_d$ then,

$$N_f = N_R = n_f = n_R = D = 0 \tag{10a}$$

*The minimum buffer requirement is $(1 + D)$ superframes.

and

$$\eta = S \quad . \tag{10b}$$

Theoretically, this is the best throughput efficiency achievable for any system using ARQ, and would be achievable for links with no propagation delay. This suggests that it is advantageous to make TDM cycles long enough to increase throughput efficiency. However, in a practical sense, for a channel with a fixed data transmission rate and a fixed burst size, increasing the cycle time reduces the capacity available to the link. Hence, although throughput efficiency improves, the net throughput over the link may not. In addition, there may not be enough transmission bursts to fill a long TDM cycle so that the channel "idle" time increases, thereby reducing the overall channel utilization.

Multiple superframes per burst

Two limiting conditions are examined in the case of multiple superframes per burst. In the first case, responses are piggybacked on superframes sent in the reverse direction. In the second case, a one-way data flow is assumed and acknowledgments are sent in short (*e.g.*, 50- to 80-bit) supervisory frames in the reverse direction.

Figure 6 represents the TDM bursts corresponding to a link and its complement (reverse) in the multiple superframe per burst mode. For a given TDM system and superframe size, the effective round-trip delay in superframes will generally depend on the superframe used as a starting point and its relative location within a burst (as shown below) and also

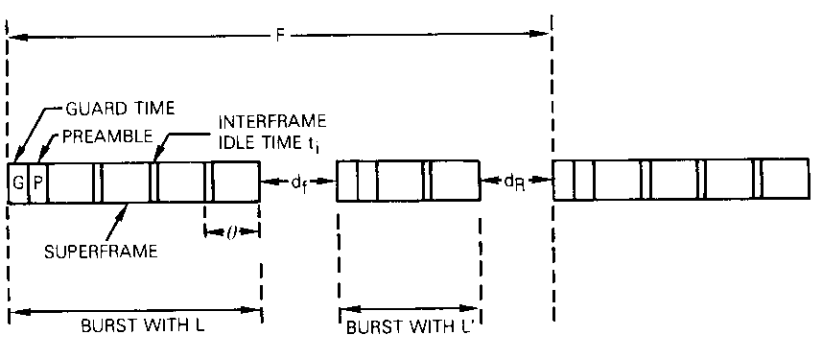


Figure 6. *Representation of TDM System Case of Multiple Superframes per Burst Operation and Burst Structure*

on the manner in which the acknowledgments are handled.

It is assumed that there are n superframes in the burst associated with the link and n' superframes in the burst associated with the complementary link. The following quantities are also defined:

θ = transmission time per superframe (including interframe idle time)
t_i = interframe idle time
P = burst preamble in seconds
G = guard time
B = burst time for burst containing link, including preamble and guard time
B' = burst time for burst containing complementary link
n = number of superframes in burst B
n' = number of superframes in burst B'.

If the effective round-trip delay $D(i)$ in superframes is to be associated with the ith superframe in the burst, then for the case of piggybacked responses (Figure 7), the receiving station sends an acknowledgment in the rth superframe of the complementary burst in the n_fth TDM cycle from the current burst. It can be shown that

$$n_f = \left\{ \frac{T_d - [(n - i - 1)\,\theta + B' + d_f]}{F} \right\}_+ \geq 0 \tag{11}$$

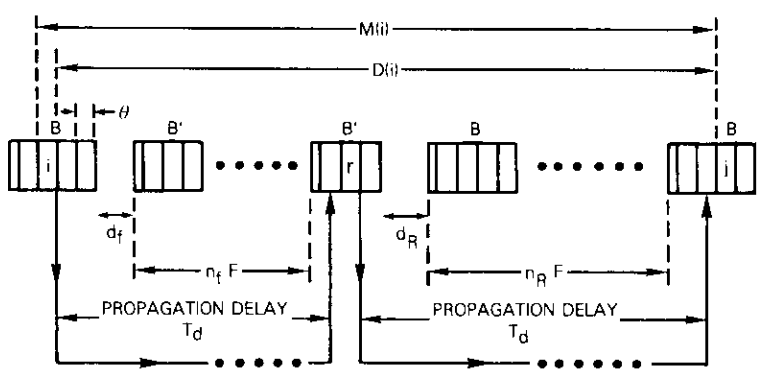


Figure 7. *Illustration of Effective Propagation Delay D(i) in a TDM System in the Case of Multiple Superframe per Burst Operation With Acknowledgments Attached to Superframes*

and

$$r = 1 + \left\{\frac{T_d - [d_f + n_f F + (P + G) + (n - i)\,\theta]}{\theta}\right\}_+ \geq 1 \quad (12)$$

where d_f and d_R are defined as in the case of multiple bursts per superframe.

In the case of one-way data flow (Figure 8), the supervisory frame of b_o bits containing the acknowledgments is transmitted in time t_o starting at time t from the start of the burst B' after the preamble has been transmitted. Idle portions of the burst are assumed to be filled by the transmission of the idle bit patterns (continuous flags). In this case,

$$n_f = \left\{\frac{T_d - [(n - i)\,\theta + B' + d_f + t_o]}{F}\right\}_+ \geq 0 \quad (13)$$

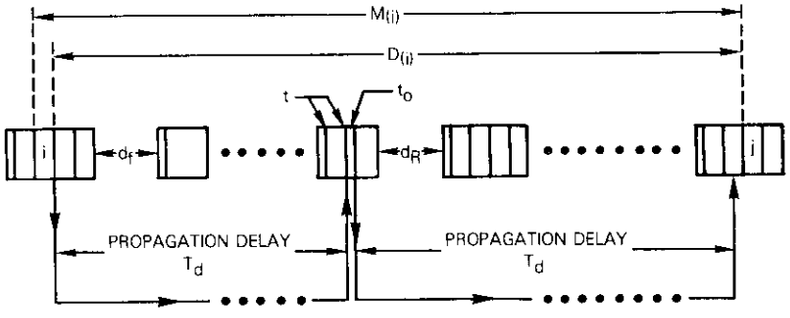


Figure 8. *Illustration of Effective Propagation Delay D(i) in a TDM System in the Case of Multiple Superframe per Burst Operation and One-Way Data Flow*

and

$$t = T_d - [(n - i)\,\theta + (P + G) + d_f + n_f F] \geq 0 \quad . \quad (14)$$

For the case of piggybacked responses, if a retransmission is required, it will occur in the jth superframe of the burst located in the n_Rth TDM cycle following the complementary burst which contains the acknowledgment in its rth superframe:

$$n_R = \left\{\frac{T_d - [(n' - r - 1)\,\theta + B + d_R]}{F}\right\}_+ \geq 0 \quad (15)$$

$$j = 1 + \left\{\frac{T_d - [(n' - r)\,\theta + d_R + n_R F + (P + G)]}{\theta}\right\}_+ \geq 1 \quad . (16)$$

For the case of one-way data flow,

$$n_R = \left\{\frac{T_d - [B' - (P + G + t + t_o) + B - \theta + d_R]}{F}\right\}_+ \geq 0 \quad (17)$$

and

$$j = 1 + \left\{\frac{T_d - [B' - (t + t_o) + n_R F + d_R]}{\theta}\right\}_+ \geq 1 \quad . \quad (18)$$

The total number of superframes transmitted in this round-trip period can be called the effective round-trip delay, $D(i)$, in superframes, and is given by

$$D(i) = n - i + n(n_f + n_R) + j - 1 \quad (19)$$

and $M(i) = 1 + D(i)$ is

$$M(i) = (1 + n_f + n_R)\,n + j - i \quad . \quad (20)$$

It can be observed that $D(i)$ depends on i both explicitly and implicitly through n_f, n_R, and j, which also depend on i. Because the effective delay is not a constant, the model derived in the appendix cannot be directly used; a modified approach is presented below.

The service cycle is defined as in the appendix. However, $M(I)$ depends on the superframe location (I) within the burst. It is necessary to consider each service cycle originating at the ith ($i = 1, 2, \ldots, n$) superframe in the burst containing the particular link. For such a service cycle originating at i, $M(i)$ is computed using equation (20). If $m(i)$ consecutive superframes are transmitted without error,

$$P_{0,i} = S^{m(i)} \quad (21a)$$

$$C_{0,i} = M(i) \quad (21b)$$

$$n_{0,i} = m(i) \quad (21c)$$

where

$$m(i) = \min[N, M(i)] \quad (22)$$

and P, C, and n are defined in the appendix.

For a first transmission error in location k [$k = 1, 2, \ldots, m(i)$]

$$P_{k,i} = S^{k-1}(1 - S) \tag{23a}$$

$$C_{k,i} = k + M(l) \tag{23b}$$

$$n_{k,i} = k - 1 \tag{23c}$$

where

$$l = k - (n - i + 1) - Kn \tag{24}$$

$$K = \left[\frac{k - (n - i + 1)}{n}\right]_{-} \geq 0 \quad . \tag{25}$$

The average number of successful frames, $\bar{n}_i$, and total average capacity, $\bar{C}_i$, per service cycle can then be computed as

$$\bar{n}_i = m(i)\, S^{m(i)} + \sum_{k=1}^{m(i)} (k - 1)\, S^{k-1}(1 - S) \tag{26}$$

or

$$\bar{n}_i = \frac{S[1 - S^{m(i)}]}{(1 - S)} \quad . \tag{27}$$

Similarly,

$$\bar{C}_i = M(i)\, S^{m(i)} + \sum_{k=1}^{m(i)} [k + M(l)]\, S^{k-1}(1 - S) \quad . \tag{28}$$

Unlike equation (27), equation (28) cannot be reduced to a closed form because of the $M(l)$ term, and hence must be computed numerically using equations (20), (24), and (25). For a service cycle starting at the ith superframe of a burst, the average efficiency is

$$\eta_i = \frac{\bar{n}_i}{\bar{C}_i} \quad . \tag{29}$$

The next step is to evaluate the steady-state probability p_i that a service cycle starts at the ith superframe of a burst. It can be assumed that A_{ij} is the transition probability (in the steady state) for a service cycle starting at i; the next service cycle can start at the jth superframe of some future

burst. For each i, the value of j may depend on the value of k of the first transmission error. Thus each A_{ij} is the sum of the probabilities in equations (21) and (23) for a value i, and each value of k, which, upon using equations (11)–(18), yields the specified value of j. These computations can be repeated for each value of i ($i = 1, \ldots, n$) and k [$k = 0, 1, \ldots, m(i)$], and the probabilities in equations (21) and (23) could be added to the corresponding A_{ij} to obtain the coefficient matrix (A_{ij}) of the equation

$$(A_{ij})[p_i] = [p_i] \quad . \tag{30}$$

The matrix equation (25) can then be solved numerically in conjunction with

$$\sum_{i=1}^{n} p_i = 1 \tag{31}$$

to yield the set of probabilities $[p_i]$, and the throughput efficiency η can then be evaluated as

$$\eta = \sum_{i=1}^{n} \eta_i p_i \tag{32}$$

where η_i have been computed for each value of i from equation (29).

The minimum buffer requirement for maximum throughput efficiency may be expressed as

$$B_u = \max[M(1), M(2), \ldots, M(n)] \tag{33}$$

where $M(i)$ is given by equation (20).

Effect of overhead and BLU structure

The throughput efficiencies, η_o, account for the average ratio of superframes successfully transmitted to the maximum number of superframes which are transmittable over the specific TDM system. The true throughput efficiency, η, can be defined as the average ratio of information bits successfully transmitted per second to the channel capacity in bits/second allocated to the link. This can be written as

$$\eta = \eta_o f_T f_S(f_B) \tag{34}$$

where f_T is an overhead factor due to the TDM system, f_S is the superframe

overhead, and f_B is an optional factor for BLU overhead when the BLU structure must be considered.

The TDM overhead is due to the preamble, P, and guard time, G; hence,

$$f_T = \frac{n\theta}{B} = \left(\frac{B - P - G}{B}\right) \quad . \tag{35}$$

In a real system, interframe idle times, t_i, may exist during a burst transmission, and in some systems the tail end of the burst of duration Δ may not accommodate a full superframe, and therefore, transmissions occurring during this period are lost. Consequently, the general expression for f_T can be written as

$$f_T = \frac{n(\theta - t_i)}{B} = \left\{\frac{B - [P + G + \Delta + (n-1)\, t_i]}{B}\right\} \quad . \tag{36}$$

The superframe overhead of A bits gives

$$f_S = \left(\frac{b - A}{b}\right) \tag{37}$$

where b is the total superframe size including overhead bits.

For performance at a superframe level, f_T and f_S given by equations (36) and (37) are multiplied by η_o to yield true throughput efficiency, η. The BLU overhead of A' bits can be determined using a similar multiplicative factor

$$f_B = \left(\frac{b' - A'}{b'}\right) \tag{38}$$

where b' is the BLU size including A' overhead bits, if it is assumed that a superframe is constructed from an integral number of BLUs.

Application of the models

The explicit computation of throughput efficiency for an individual link in a TDM system requires knowledge of the following parameters:

a. channel transmission rate;

b. superframe size and overhead;

c. preamble, guard times, and interframe idle times;

d. number of bursts per superframe or the number of superframes per burst, *i.e.*, n;

e. bit error rate (BER) or superframe success probability, S;

f. parameters d_f and d_R, and the TDM cycle time, F;

g. available transmission buffer size, N (superframes);

h. mode of transmitting acknowledgments (*i.e.*, piggybacked or not).

Based on these parameters, the computation can proceed as follows for the case of multiple bursts per superframe:

a. N_f and N_R are computed using equations (1) and (2).

b. D is computed using equation (5) or (6).

c. The throughput efficiency, η_o, is computed using equation (7) or (8).

d. The effect of the various kinds of overhead is computed using equations (34)–(38).

For the case of multiple superframes per burst, the following procedures are used:

a. For each value i ($1 \leq i \leq n$) n_f, r (or t), n_R, j, $D(i)$, and $M(i)$ are computed using equations (11)–(20), and $m(i)$ is determined from equation (22).

b. For each value of k [$0 \leq k \leq m(i)$] with probabilities defined in equations (21) and (23), the values of $M(l)$ are computed using equations (11)–(20). The initial i and the final j corresponding to each k are used to add the corresponding probability P_k [equations (21) and (23)] to a probability matrix A_{ij}.

c. The averaged $\overline{n}_i$ and $\overline{C}_i$ are computed using equations (27) and (28), from which η_i is derived using equation (29).

d. The above procedure is repeated for each i to derive the set of η_i.

e. The matrix equation (30) is solved using equation (31), and the final throughput efficiency, η_o, is determined with equation (32).

f. The effect of overhead is included by using equations (34)–(38).

The complex computational requirements can be simplified to a large extent when specific systems are considered which may correspond to the models operating at certain limits. However, the computational algorithm outlined above could be easily implemented as a computer program, with the required data and TDM system parameters as input. Such a program is applicable to a wide range of TDM systems. The example system used for illustrative purposes in the next section was evaluated using such a program.

Example system

To illustrate the effect of the various system parameters on the throughput efficiency and the prediction of TDM system performance characteristics using an HDLC type protocol, the case of a hypothetical 4-node satellite TDMA network is presented, based on the following assumptions:

a. The one-hop satellite propagation delay is 270 ms.

b. The satellite channel is a random error channel with the same BER, β, for all links. The superframe success probability, S, is given by

$$S = (1 - \beta)^b (1 - \beta)^{b_r}$$

where b is the total superframe size in bits and b_r is the size in bits for the frame containing the acknowledgments. For small b_r (one-way data flow cases) $S \simeq (1 - \beta)^b$; for the piggybacked acknowledgment cases $S \simeq (1 - \beta)^{2b}$.

c. A TDMA frame (cycle) consists of four bursts of equal size; one burst is transmitted by each node.

d. The superframe overhead is 64 bits.

e. A 1.544-Mbit/s channel transmission rate is assumed.

f. A large transmit buffer (larger than minimum requirements) is assumed at each node.

g. Preamble plus guard times of 50 μs and 5 ms have been used. Although 5 ms may be considered excessive for satellite TDMA systems by conventional standards, it may be typical of a less precisely synchronized TDMA system, for which reasonably high throughputs could still be achieved. Since performance for the 50-μs case is close to that expected for a dedicated link, smaller values than 50 μs will yield performances similar to the 50-μs case for the 1.554-Mbit/s transmission rate.

h. No interframe idle times are assumed.

i. All superframes are of the same size in all of the links.

Figure 9 represents the TDMA frame burst pattern for four nodes, A, B, C, and D. The total system contains 12 individual simplex links. Since the links are distinguishable only on the basis of the d_f and d_R parameters in terms of evaluating throughput efficiencies, the system consists of three link types labeled L_{0-2}, L_{1-1}, and L_{2-0} as shown in Table 1; each link type consists of four links in the system.

For the limiting case of one-way data flow, Figures 10–12 represent the net throughput efficiency (including TDMA and superframe overhead)

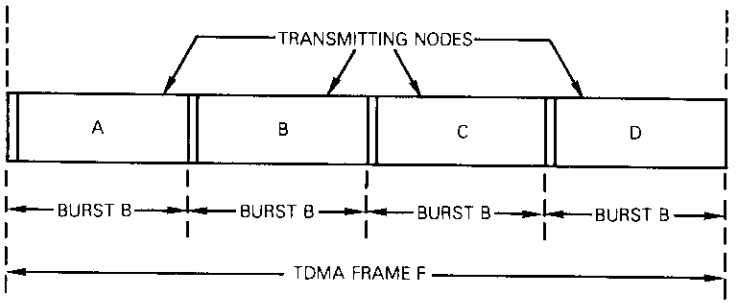


Figure 9. *Representation of Example Satellite TDMA System With Four Nodes*

TABLE 1. LINK TYPES AND PARAMETERS

Link Types	Parameters d_f	Parameters d_R	Links	Total Number of Links
L_{0-2}	0	2B	A → B, B → C, C → D, D → A	4
L_{1-1}	B	B	A → C, B → D, C → A, D → B	4
L_{2-0}	2B	0	A → D, B → A, C → B, D → C	4

for the three link types for BERs of 10^{-4}, 10^{-6}, and 10^{-8} and a preamble plus guard time of 50 μs. The dashed curves represent the case of multiple bursts per superframe while the solid curves represent the region where the case of multiple superframes per burst is applicable. Figures 13–15 show the performance for the same parameters with a preamble plus guard time of 5 ms. The family of curves in each figure represents superframe sizes of 1, 2, 4, 8, 16, and 32 K bytes. Figures 16–21 represent the performance of the same systems with all nodes transmitting data and assuming piggybacked acknowledgments.

The general behavior of the throughput efficiency is as follows. First, the piggybacked cases generally yield lower throughput, with the difference increasing with the BER and superframe size. This is primarily attributed to an increase in acknowledgment errors due to piggybacking on large data superframes. Further, in short TDMA frame lengths where bursts are small, the efficiency is low because of the overhead due to preamble and

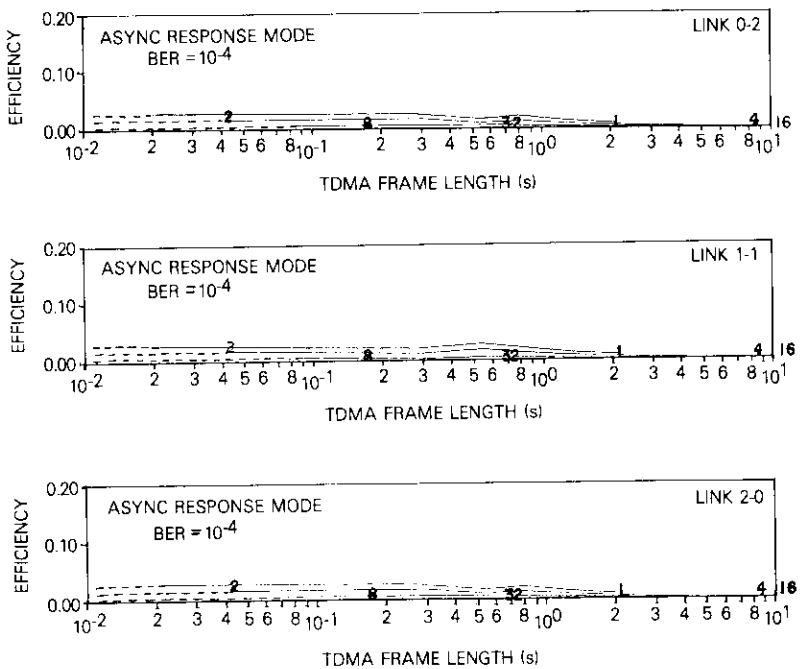


Figure 10. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 50 μs (BER = 10^{-4}; curves represent different superframe sizes in K bytes)*

guard time. As the burst length (and hence TDMA frame length) is increased, the overhead factor becomes insignificant and the throughput efficiency increases until the additional retransmissions at longer TDMA frame lengths cause it to decline again. Thus, all performance curves exhibit a sharp or broad peak corresponding to maximum achievable efficiency, the broader peaks occurring at either a low BER or a short preamble plus guard time. In Figures 10–12, the reduction in efficiency due to 50-μs preamble plus guard time overhead is at very short TDMA frames, which are outside the scale used. The reduction of the BER therefore causes the efficiency peaks to be higher and broader. In addition, the sensitivity to superframe size is reduced as the BER is reduced.

In general throughput efficiencies are higher for smaller superframe sizes, although the difference becomes negligible at low BERs. However, as superframe sizes are reduced below 1 K byte, the throughput efficiency

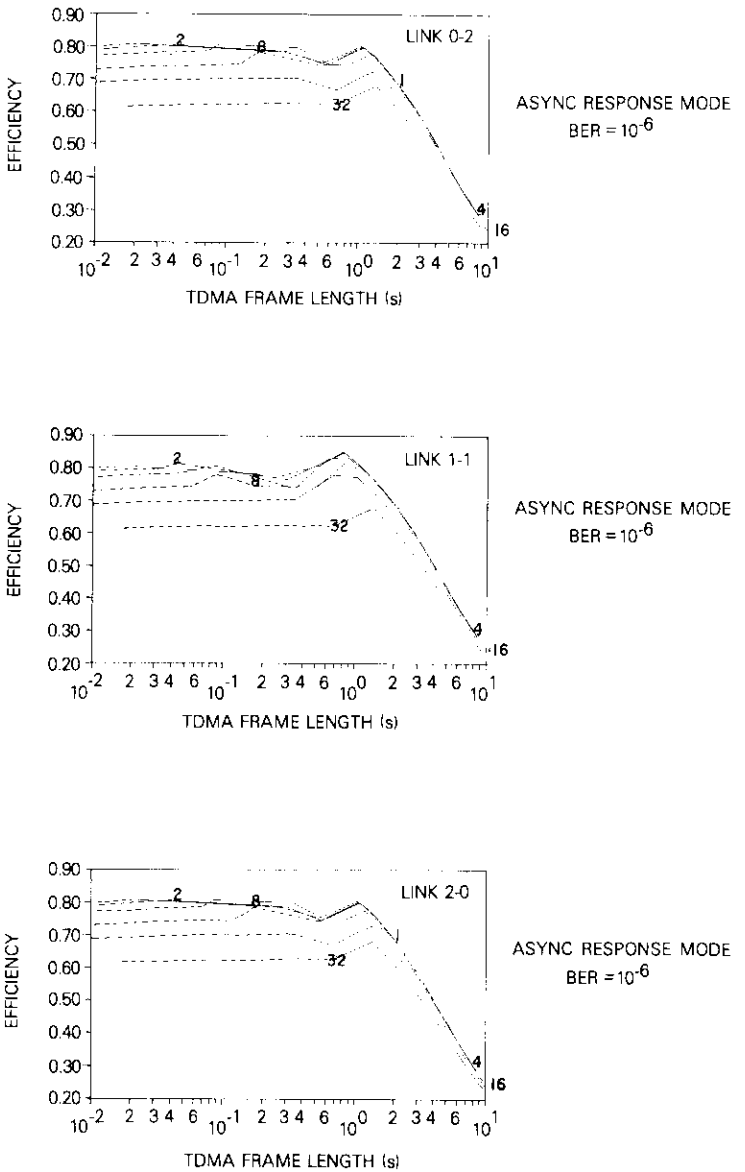


Figure 11. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 50 μs (BER = 10^{-6}; curves represent different superframe sizes in K bytes)*

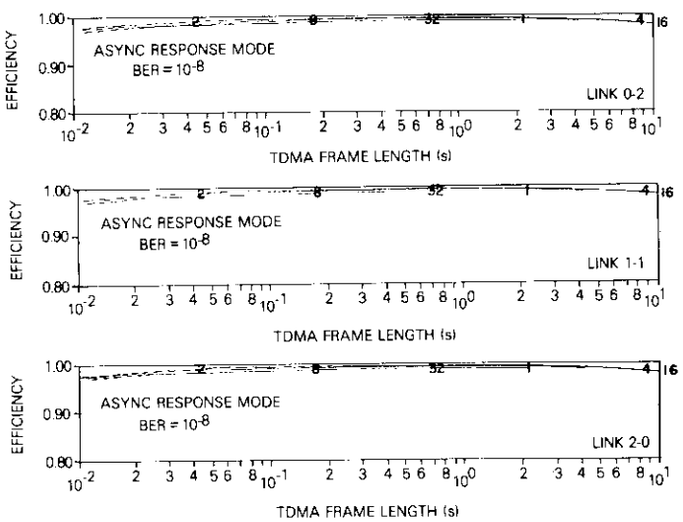


Figure 12. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 50 μs (BER = 10^{-8}; curves represent different superframe sizes in K bytes)*

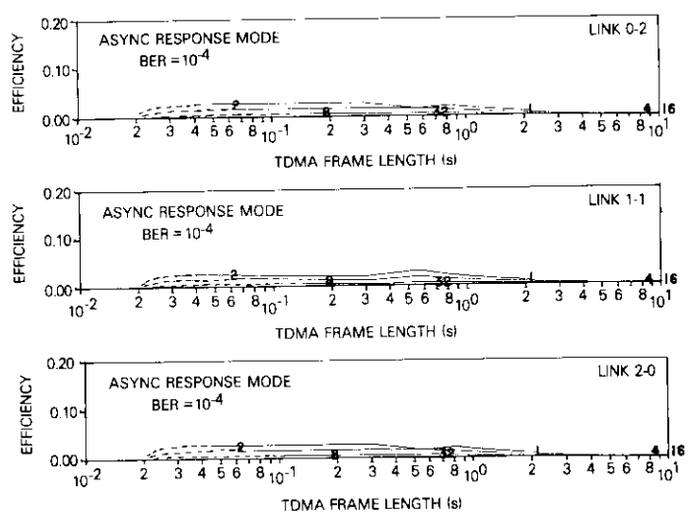


Figure 13. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 5 ms (BER = 10^{-4}; curves represent different superframe sizes in K bytes)*

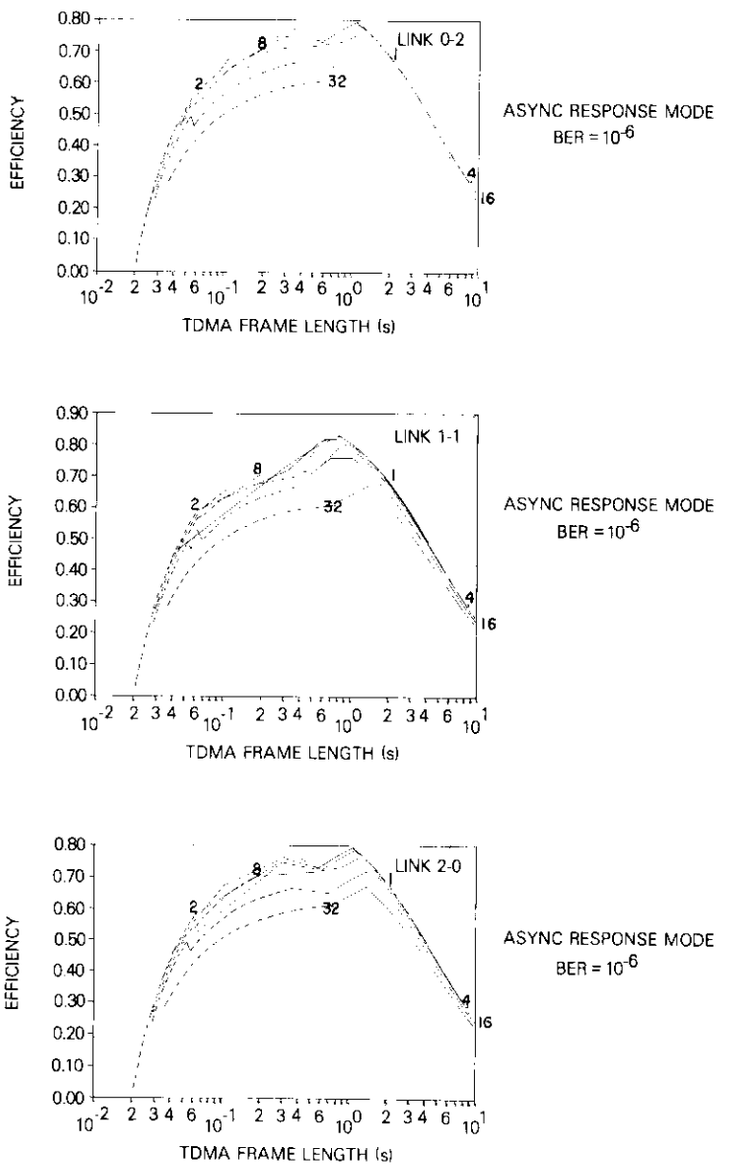


Figure 14. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 5 ms (BER = 10^{-6}; curves represent different superframe sizes in K bytes)*

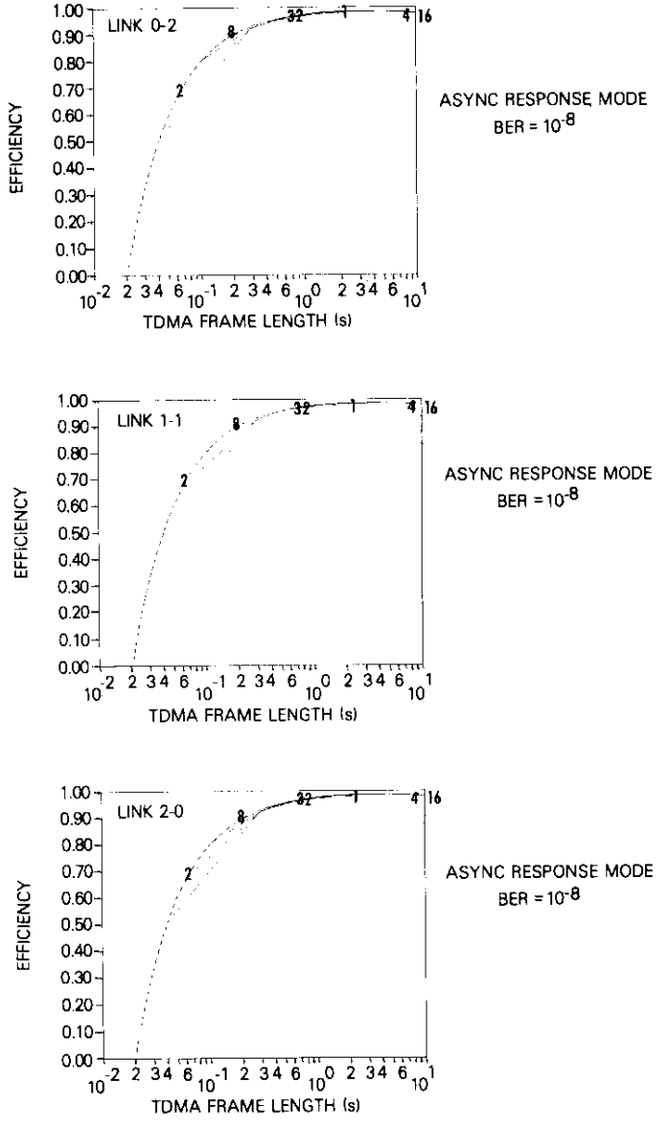


Figure 15. *Net Throughput Efficiency for One-Way Data Flow and a Preamble Plus Guard Time of 5 ms (BER = 10^{-8}; curves represent different superframe sizes in K bytes)*

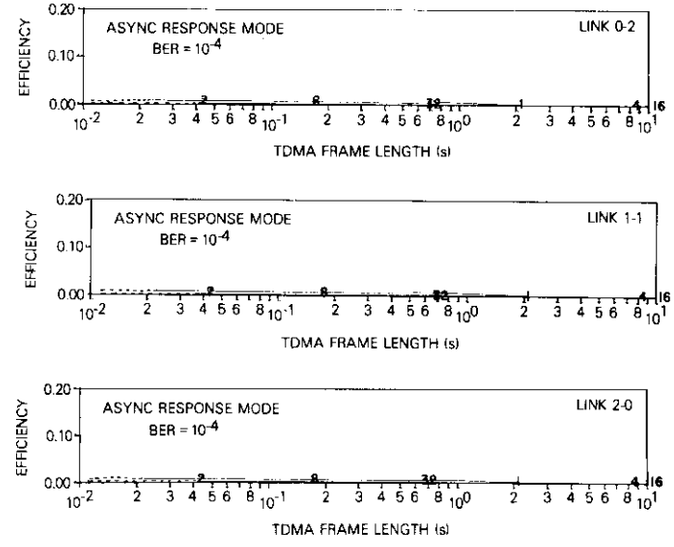


Figure 16. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 50 μs (BER = 10^{-4}; curves represent different superframe sizes in K bytes)*

will eventually start decreasing (not shown) as the superframe overhead of 64 bits will become significant. For this system, the performances for the three types of links are almost identical at low BERs. Differences in the performance become apparent as the BER is increased due to different effective round-trip delays; however, this becomes a significant factor only with a significant number of retransmissions and is negligible at a low BER.

Acknowledgments

The author is indebted to Dr. W. L. Cook and Mr. R. C. Davis for their invaluable encouragement and support.

Reference

[1] A. K. Kaul, "Performance of High-Level Data Link Control in Satellite Communications," *COMSAT Technical Review*, Vol. 8, No. 1, Spring 1978, pp. 41–87.

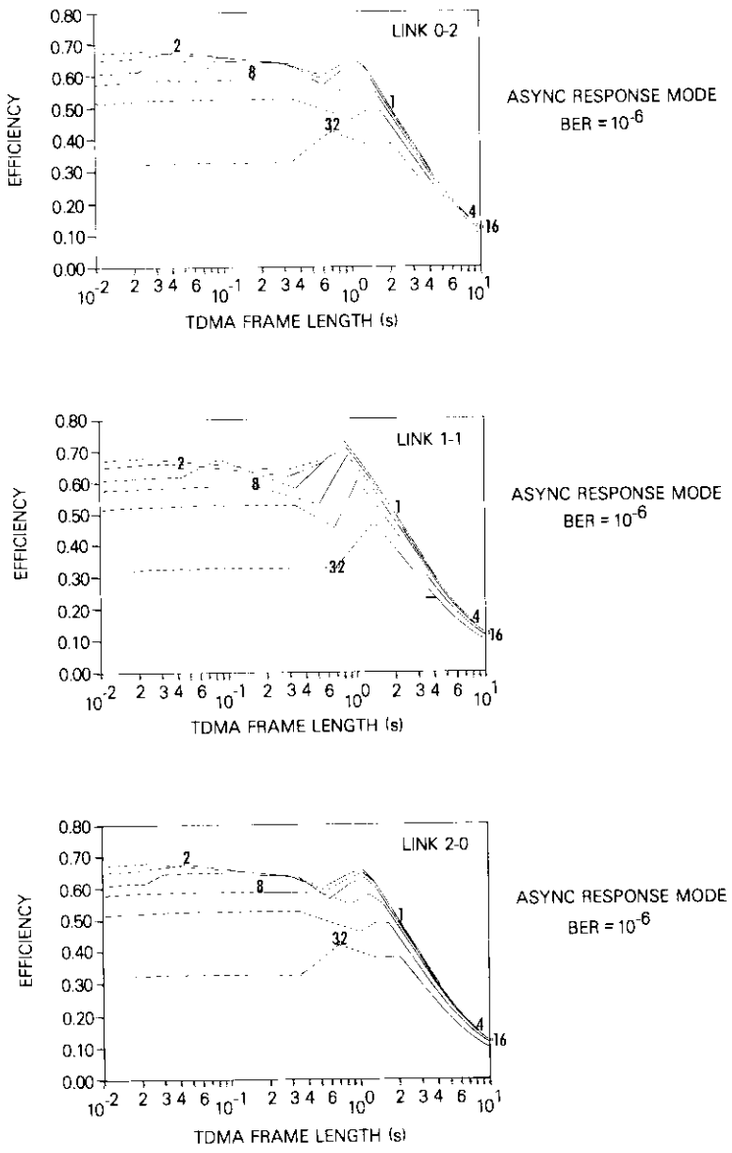


Figure 17. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 50 μs (BER = 10^{-6}; curves represent different superframe sizes in K bytes)*

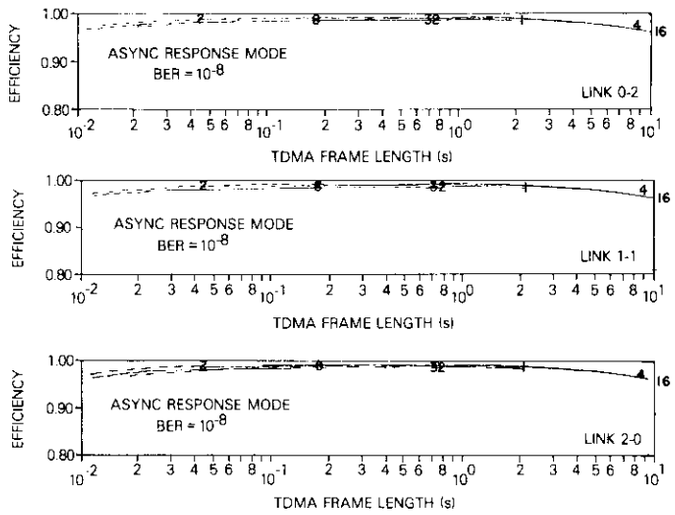


Figure 18. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 50 μs (BER = 10^{-8}; curves represent different superframe sizes in K bytes)*

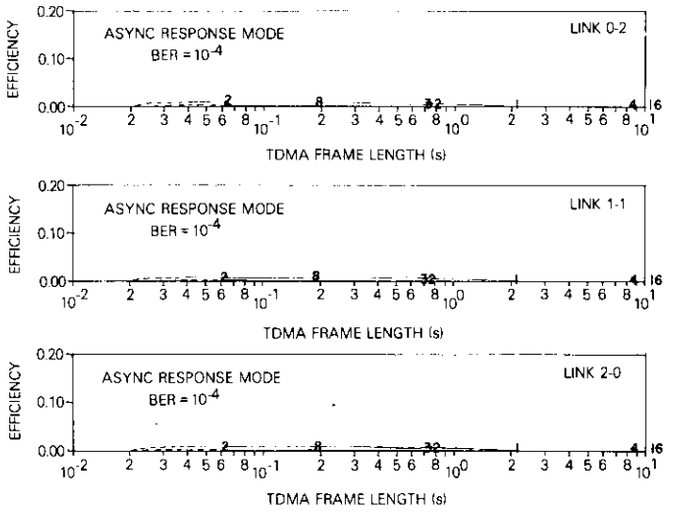


Figure 19. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 5 ms (BER = 10^{-4}; curves represent different superframe sizes in K bytes)*

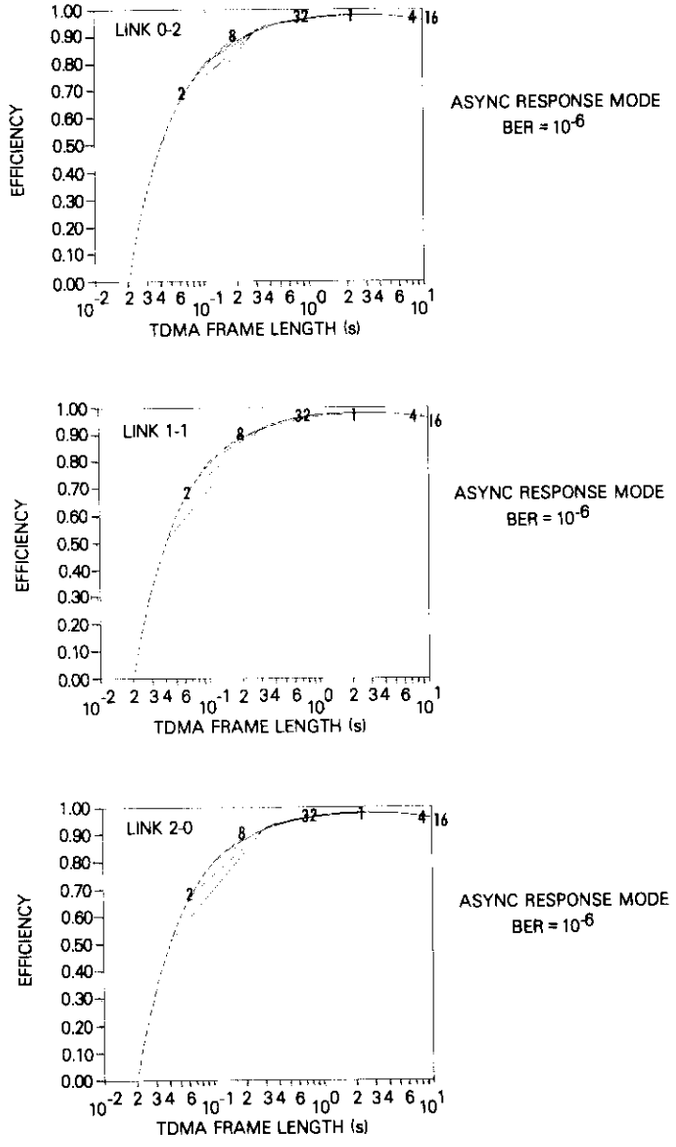


Figure 20. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 5 ms (BER = 10^{-6}; curves represent different superframe sizes in K bytes)*

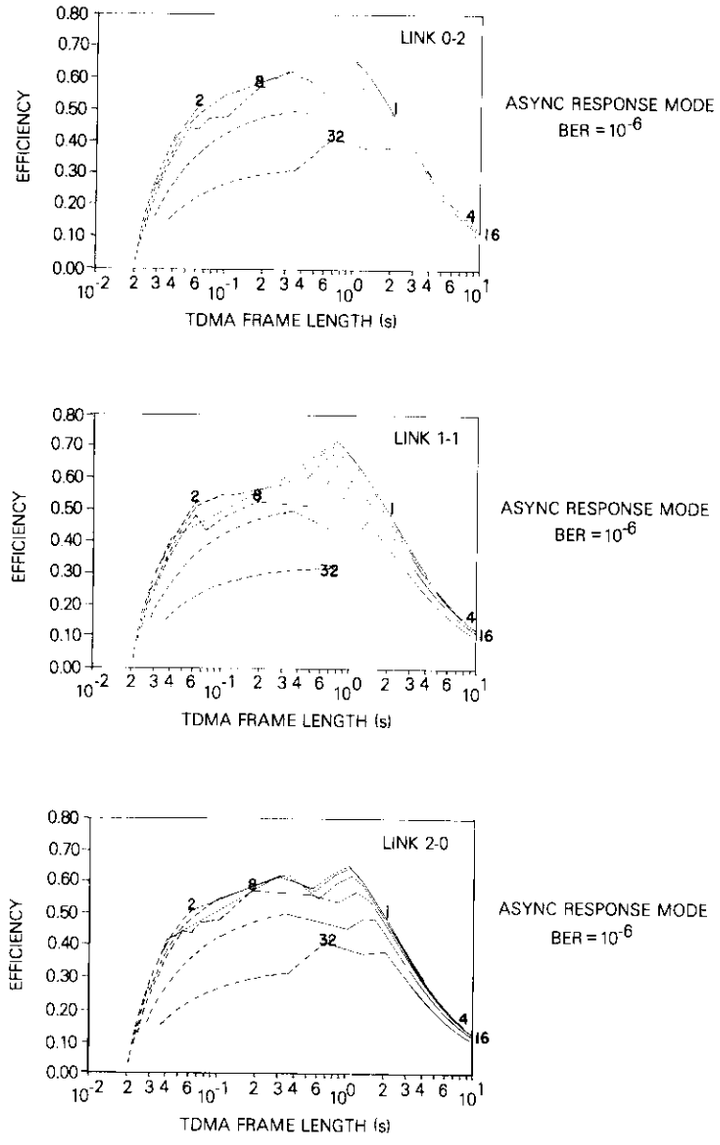


Figure 21. *Net Throughput Efficiency for Piggybacked Acknowledgments With All Nodes Transmitting and a Preamble Plus Guard Time of 5 ms (BER = 10^{-8}; curves represent different superframe sizes in K bytes)*

Appendix: Derivation of throughput efficiency for the case of multiple bursts per superframe

A service cycle can be defined as originating with the first superframe of a retransmission sequence or the first superframe transmitted after m consecutive superframe transmissions without error, where

$$m = \min[N, 1 + D] \quad . \tag{A-1}$$

Service cycles can be characterized by k, the position of the first superframe with error ($k = 0$ implies no errors); P_k, the probability of occurrence; C_k, the capacity in the service cycle, *i.e.*, maximum possible superframes transmittable in the service cycle; and n_k, the number of superframes successfully transmitted in the service cycle. For $k = 0$,

$$P_0 = S^m \tag{A-2a}$$

$$C_0 = 1 + D \tag{A-2b}$$

$$n_0 = m \tag{A-2c}$$

and for $k = 1, 2, \ldots, m$,

$$P_k = S^{k-1}(1 - S) \tag{A-3a}$$

$$C_k = k + D \tag{A-3b}$$

$$n_k = k - 1 \quad . \tag{A-3c}$$

These equations can be used to compute averages:

$$\bar{n} = \sum_{k=0}^{m} n_k P_k = mS^m + \sum_{k=1}^{m} (k - 1)\, S^{k-1}(1 - S) \tag{A-4a}$$

or

$$\bar{n} = \frac{S(1 - S^m)}{(1 - S)} \tag{A-4b}$$

and

$$\bar{C} = (1 + D)\, S^m + \sum_{k=1}^{m} (k + D)\, S^{k-1}(1 - S) \tag{A-5a}$$

that is,

$$\bar{C} = 1 + D - mS^m + \frac{S(1 - S^m)}{(1 - S)} \tag{A-5b}$$

or

$$\bar{C} = \bar{n} + 1 + D - mS^m \quad . \tag{A-5c}$$

Hence, throughput efficiency is

$$\eta = \frac{\bar{n}}{\bar{C}} = \frac{S(1 - S^m)}{S(1 - S^m) + (1 + D - mS^m)(1 - S)} \quad . \tag{A-6a}$$

For $N < 1 + D$, (i.e., $m = N$),

$$\eta = \frac{S(1 - S^N)}{S(1 - S^N) + (1 + D - NS^N)(1 - S)} \tag{A-6b}$$

and for $N \geq 1 + D$ (i.e., $m = 1 + D$),

$$\eta = \frac{S}{S + m(1 - S)} \quad . \tag{A-6c}$$

Ashok K. Kaul received B.Sc. and M.Sc. degrees in physics from the University of Jammu in Kashmir, India, and a Ph.D. in physics from the University of Maryland. In 1973, he joined COMSAT *Laboratories, where he is currently Manager of the Computer Communications Department of the Transmission Systems Laboratory, responsible for directing research and experiments in computer communications.*

CTR Notes

Integrated circuits in communications satellites

A. G. Revesz

(Manuscript received December 1, 1978)

Integrated circuits (ICs) have been used in many civilian and military space vehicles. The more recently developed small, medium, and large scale integrated (SSI, MSI, and LSI) circuits are also entering space electronics hardware [1]. An important consideration in this application is radiation sensitivity. Although ICs based on bipolar devices are less sensitive to radiation than those utilizing MOS devices, their power requirement is much larger. The very low quiescent power consumption, excellent noise immunity, and cost effectiveness of MOS-based MSI and LSI circuits also offer great advantages over bipolar ICs. Complementary (*i.e.*, both n- and p-channel) MOS (CMOS) ICs, which are the essential elements of current microprocessors and similar systems, are especially advantageous.

The use of CMOS circuits, which have been employed in several military space programs, has been suggested for the telemetry and command subsystems of communications satellites. The benefits realized from these circuits in space-borne data reduction applications are expected to be extended to onboard signal processing of future sophisticated communications satellites. Various functions such as adaptive collection of telemetry

---

A. G. Revesz is a Senior Staff Scientist in the Applied Sciences Laboratory of COMSAT Laboratories.

data, stationkeeping, and traffic routing could be performed by central microcomputers and distributed microprocessor systems.

The irradiation behavior of ICs is fundamentally different from that of the silicon solar cells used extensively in communications satellites. The degradation of solar cells due to irradiation is well understood and completely predictable; hence, it does not represent a reliability problem because end-of-life power output can be considered in the design of the satellite power system [2]. However, the irradiation degradation of MSI and LSI circuits is not well understood, and the technology of radiation-tolerant devices is still in an evolutionary stage. Consequently, ICs represent a potentially serious reliability problem in communications satellites.

This note evaluates the present (1978) radiation resistance of MSI and LSI circuits, as well as the need for a better understanding of the fundamental processes which determine radiation resistance and for improved control of the fabrication technology.

The cumulative radiation dose during seven years in geostationary orbit has been calculated for silicon shielded by aluminum or lead [3]. Because X-rays are generated by "Bremsstrahlung," the total radiation absorbed in silicon during seven years cannot be reduced below $\sim 3 \times 10^4$ rads or $\sim 4 \times 10^3$ rads by using lead or aluminum shields, respectively, even if the thickness of the aluminum shield is increased beyond $\sim$10 mm (corresponding to $\sim$2.7 g/cm^2). Commercially available MSI and LSI circuits, which have not been especially fabricated to be radiation hard, fail catastrophically even when exposed to radiation levels lower than these minimum values. For example, the currently very popular commercial NMOS microprocessors exhibit a 100-percent failure rate after exposure to 3×10^3 rads (Si) irradiation, although failures began at levels as low as 10^3 rads (Si) [4]. These failures are caused by an excessive shift in the threshold voltage of MOS FETS arising from irradiation-generated charges at the Si/SiO_2 interface.

The shift in threshold voltage is the primary effect of irradiation on MOS devices that is attributed to the positive charge accumulation at the Si/SiO_2 interface resulting from the trapping of holes generated by irradiation in the SiO_2 film. One reason for preferential hole trapping is that the electrons of the radiation-generated electron-hole pairs are much more mobile than holes in SiO_2. Also, the Si/SiO_2 interface contains hole traps, whose density depends on subtle details of oxide formation and post-oxidation heat treatments. In addition to the positive charge build-up, trapping of holes at the Si/SiO_2 interface may also change the density of electronically active interface states. Since the holes are less mobile than

electrons in the SiO_2 film and have a high probability of being trapped at the Si/SiO_2 interface, irradiation under positive bias is usually more detrimental than that under negative or zero bias. Therefore, n-channel MOS devices with a positive turn-on threshold voltage are, in principle, more sensitive to irradiation than p-channel MOS devices which operate under negative gate voltage. It was suggested in a recent review of the chemical and structural aspects of irradiation effects in Si/SiO_2 interface structures that SiH groups act as hole traps at the interface [5].

The irradiation behavior of the Si/SiO_2 interface structure is different from that of bulk semiconductors in two important aspects. Ionizing radiation is sufficient to change the properties of the Si/SiO_2 interface; this change is probably associated with a chemical reaction involving hydrogen. In contrast, ionizing radiation does not affect the defect structure of bulk semiconductors because defects are generated by more energetic radiation and hydrogen is not involved in this process.

The effect of irradiation on the characteristics of an n-channel MOS FET is illustrated in Figure 1 [6]. The threshold voltage, V_T, shifts with increasing dosage level. The logic operation of a CMOS gate would probably tolerate dosage level 1 but not level 2 since V_T crosses the zero axis, thus increasing the leakage current. At dosage level 3, speed and noise immunity may also be seriously affected. Finally, at dosage level 4, the $V_{in} - V_{out}$ characteristics of the device collapse completely. The onset of $V_T = 0$ is most commonly interpreted as the point of maximum acceptable degradation or failure, and the corresponding dose is the maximum acceptable dose, D_A (Max). In the preceding example, D_A (Max) is 10^3 rads (Si) [4]. When the hole trapping at the Si/SiO_2 interface is associated with an increase in the interface state density, the shape of the $I_d - V_g$ (drain current *vs* gate voltage) curve also changes. Although such a change may sometimes appear "advantageous" (Figure 1b), interface states eventually become potential sources of undesirable instability effects. A partial annealing ("recovery") of the charges present in the interface states may also occur (Figure 1c).

For long-term geostationary orbit applications, radiation tolerance control of MSI and LSI circuits is critical as evidenced in Figure 2. A device with a D_A (Max) value of 10^3 rads, even with a thick shield (3 kg), will survive for only two years. Elevating D_A (Max) to 5×10^3 rads or 10^4 rads allows a 7-year survival by applying only 1.5-kg or 0.5-kg shielding, respectively. This "hardening" may be achieved by a relatively simple preselection process. Additional hardening, resulting in increased life and/or decreased shield weight, can be achieved by using specially fabricated

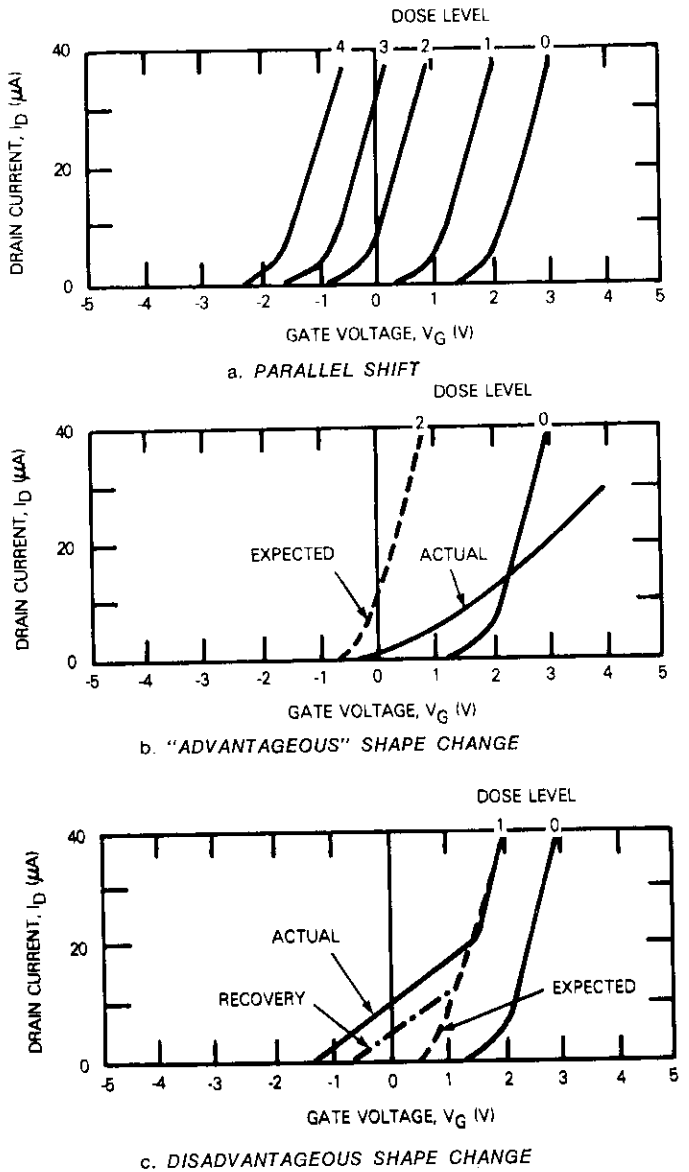


Figure 1. *Effect of Irradiation on the Characteristics of an n-Channel MOS FET. The numbers representing dosage levels are only indicators; the actual dose level depends on many parameters (from Reference 3).*

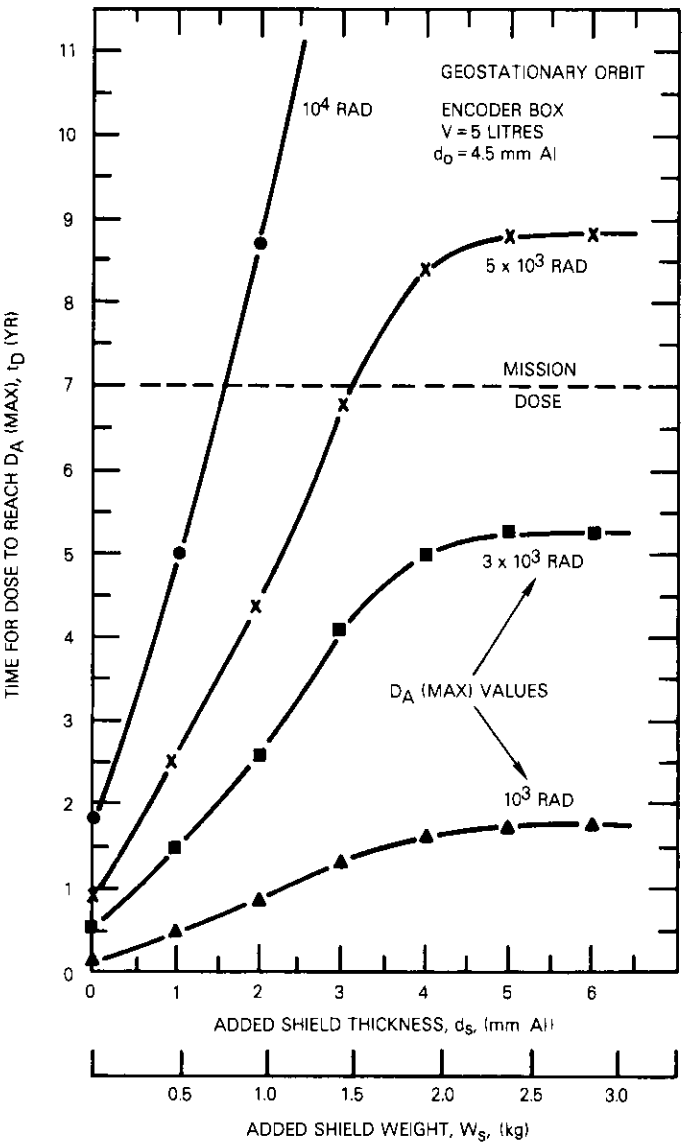


Figure 2. *Survival of CMOS Circuits in a Particular Location of a Communications Satellite (from Reference 3)*

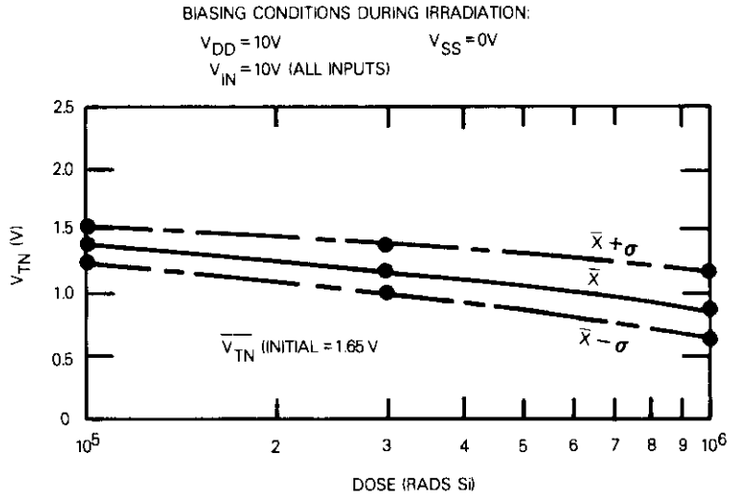
MSI and LSI circuits.

The general trend is toward CMOS circuits. It is inadvisable to mix the various circuits (*e.g.*, TTL, CMOS, and PMOS) because their compatibility may be degraded by radiation. Therefore, the following discussion of radiation-hard devices is confined to CMOS circuits. These circuits use either bulk silicon or Si films grown epitaxially on a sapphire (Al_2O_3) substrate; the latter is termed the SOS (silicon on sapphire) structure. The advantages of the SOS structure include additional speed and packing density, and reduced dynamic power consumption. The disadvantage is that the Si/Al_2O_3 interface also degrades during irradiation, resulting in increased leakage current. This increase must be considered in the circuit design and the available power.

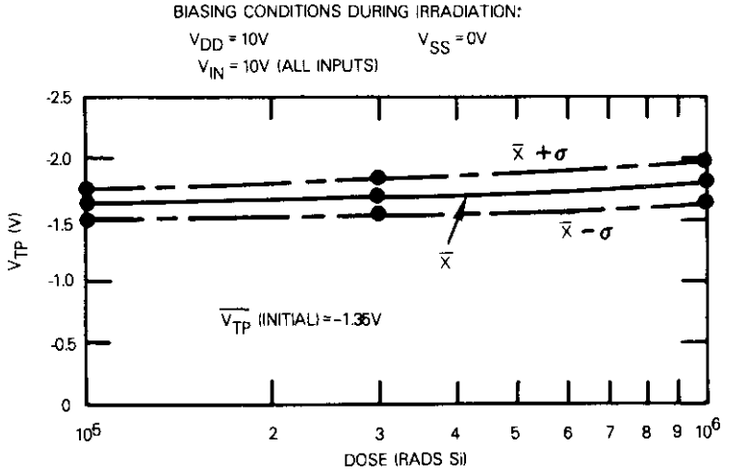
Replacement of the Al gate by the self-aligned Si gate (P- or B-doped) increases the speed at low power as well as the packing density. With respect to irradiation, B-doped Si-gate CMOS circuits are disadvantageous, since they may exhibit an instability even without irradiation [6] probably because of the exposure of the initial oxide film and Si/SiO_2 interface to several potentially harmful steps during processing. Therefore, their technology appears to be very critical. Apparently, the P-doped Si-gate devices do not exhibit this instability but, as yet, cannot be made radiation hard [7].

Presently, only radiation-hardened bulk silicon CMOS devices with Al gate are acceptable for space use without changing the circuit design (*e.g.*, minimizing positive voltages on n-channel MOS gates) [8]. The performance of such ICs, including SSI, MSI, and LSI levels of integration, which is shown in Figure 3, guarantees post-radiation circuit functionality even after 10^5 rads (Si) irradiation [9]. During a reliability test, a total of 346 SSI devices exhibited ~350,000 operating hours without a failure.

The radiation-hard Al-gate CMOS/SOS LSI circuits exhibit greater shifts in threshold voltage than the bulk Si CMOS circuits. The shift in the threshold voltage is reflected in the shift of the minimum operating voltage, V_{DD}, of a multiplexer control circuit as shown in Figure 4 [10]. This LSI circuit has been designed for use in telemetry and command subsystems in communications satellites. Changes in V_{DD} are an excellent indication of the composite degradation of all devices in the LSI circuit. Even after 10^7 rads (Si) irradiation, V_{DD} values are below the intended nominal operating voltage of 10 V. The range of the multiplexer control chip supply current increases from 62 72 μA (before irradiation) to 200–350 μA (after 10^6 rads). A significant portion of this current probably results from the charge at the Si/Al_2O_3 interface (back channel) of the n-channel transistors. Although these results are encouraging, extensive reliability

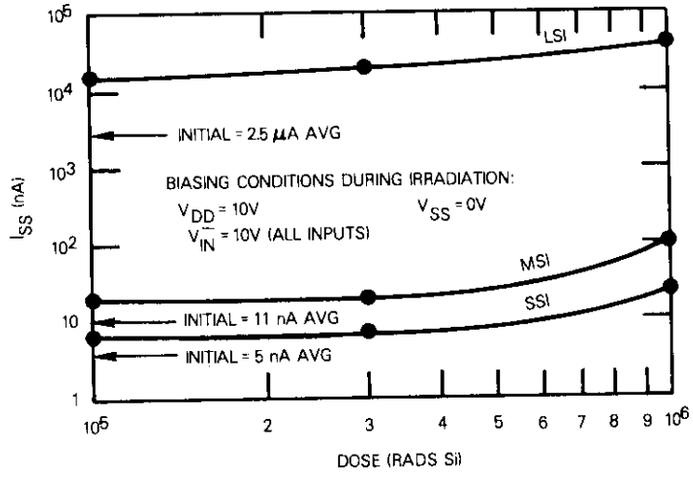


a. *THRESHOLD VOLTAGE OF N-CHANNEL DEVICES, V_{TN} (INCLUDES ALL THREE LEVELS OF INTEGRATION; SOLID LINE INDICATES MEAN VALUE AND DASHED LINES INDICATE ONE STANDARD DEVIATION)*

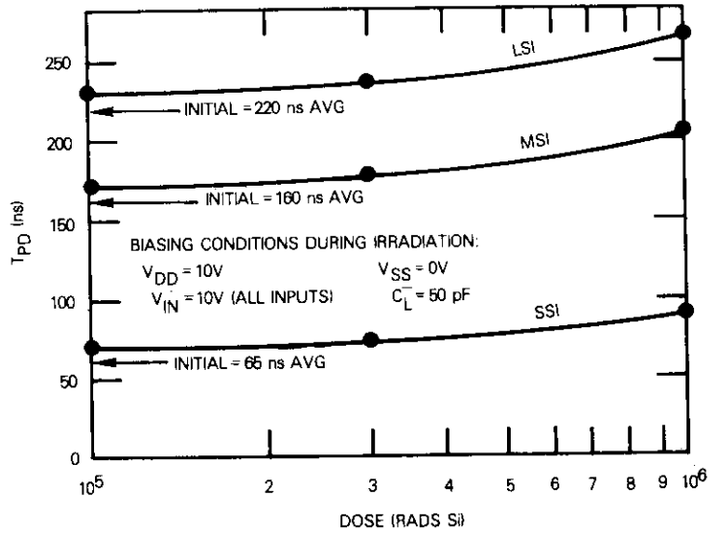


b. *THRESHOLD VOLTAGE OF P-CHANNEL DEVICES, V_{TP} (INCLUDES ALL THREE LEVELS OF INTEGRATION; SOLID LINE INDICATES MEAN VALUE AND DASHED LINES INDICATE ONE STANDARD DEVIATION)*

Figure 3. *Electrical Characteristics of Bulk Si CMOS (Al-Gate) ICs of the CS 4000 Series as a Function of Co 60 Irradiation Dose (from Reference 9)*



c. *QUIESCENT CURRENT, ISS, FOR THREE LEVELS OF INTEGRATION*



d. *PROPAGATION DELAY, T_{PD}, FOR THREE LEVELS OF INTEGRATION*

Figure 3 (continued). *Electrical Characteristics of Bulk Si CMOS (Al-Gate) ICs of the CS 4000 Series as a Function of Co 60 Irradiation Dose (from Reference 9)*

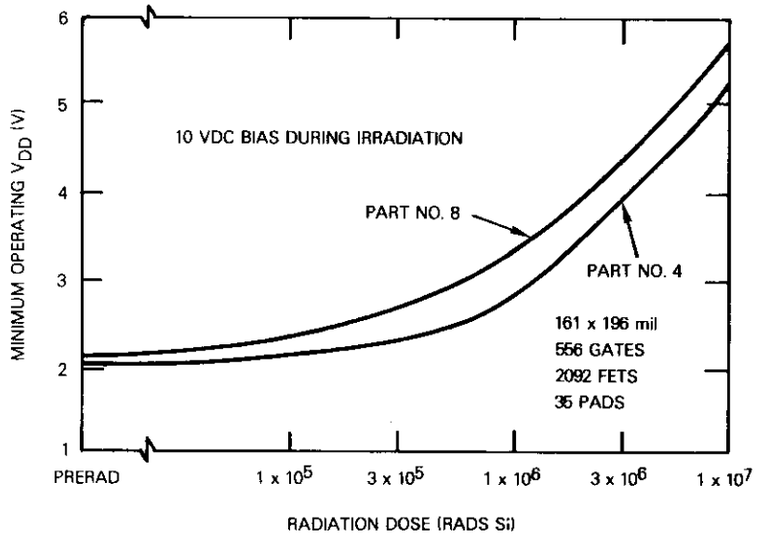


Figure 4. *Minimum Operating Voltage of a Multiplexer Control LSI Circuit (curves represent minimum and maximum shifts exhibited by six test chips) (from Reference 10)*

and reproducibility evaluation is needed before CMOS/SOS ICs can be used on communications satellites. This is particularly applicable to Si-gate CMOS circuits (both bulk Si and SOS), since these LSI circuits can exhibit a room temperature instability even before irradiation [6], [11].

The technology of radiation-hard ICs has been evolving rather empirically; the main features include the conditions of oxidation and post-oxidation heat treatments. In this respect, the technology employed by different manufacturers varies significantly as, for instance, in the water content of the oxidizing ambient. The unintentional presence of hydrogen during most of the oxidation processes and heat treatments is probably responsible for the wide variation in the processing steps and the reported results. Hydrogen is known to play a paramount, but very complicated, role in determining the oxidation kinetics of silicon and the defect structure of SiO_2 film including its interface with silicon [12]. Various instability effects occurring in Si/SiO_2 interface structures have also been attributed to the hydrogen in SiO_2 films[6], [11], [13]. Therefore, the irradiation behavior of MOS devices may also be determined by the presence of

hydrogen in various forms in thermally grown SiO_2 films. The understanding of the subtle role of hydrogen in SiO_2 films is meager, and the control of the various processing steps, with respect to hydrogen contamination, is relatively poor. However, these appear to be the most important factors for developing radiation-hard ICs which will have the high reliability required for application to communications satellites.

References

[1] P. R. Kurzhals, "New Directions in Space Electronics," *Journal of Astronautics and Aeronautics*, February 1977, pp. 32–41.

[2] H. Y. Tada and J. R. Carter, *Solar Cell Radiation Handbook*, NASA JPL Publication 77-56, 1977.

[3] A. Holmes-Siedle and R. F. A. Freeman, "Improving Radiation Tolerance in Space-Borne Electronics," *IEEE Transactions on Nuclear Science*, NS-24, No. 6, December 1977, pp. 2259–2265.

[4] D. K. Myers, "Ionizing Radiation Effects on Various Commercial NMOS Microprocessors," *IEEE Transactions on Nuclear Science*, NS-24, No. 6, December 1977, pp. 2169–2171.

[5] A. G. Revesz, "Chemical and Structural Aspects of the Irradiation Behavior of Sio_2 Films on Silicon," *IEEE Transactions on Nuclear Science*, NS-24, No. 6, December 1977, pp. 2102–2107.

[6] F. Faggin, D. D. Forsythe, and T. Klein, "Room Temperature Instabilities Observed on Silicon Gate Devices," Reliability Physics 8th Annual Symposium, Las Vegas, April 7–10, 1970, *IEEE Proceedings*, pp. 35–41.

[7] H. Hughes, Private Communication.

[8] A. Stanley, Private Communication.

[9] A. London, D. A. Matteucci, and R. C. Wang, "Establishment of a Radiation Hardened CMOS Manufacturing Process," *IEEE Transactions on Nuclear Science*, NS-24, No. 6, December 1977, pp. 2056–2059.

[10] D. P. Shumake, R. A. Kempke, and K. G. Aubuchon, "Hardened CMOS/SOS LSI Circuits for Satellite Applications," *IEEE Transactions on Nuclear Science*, NS-24, No. 6, December 1977, pp. 2177–2180.

[11] H. Nakayama, "Room Temperature Instabilities of p-channel Silicon Gate MOS Transistors," *Journal of the Electrochemical Society*, Vol. 125, No. 8, August 1978, pp. 1302–1306.

[12] A. G. Revesz, "The Role of Hydrogen in SiO_2 Films on Silicon," *Journal of the Electrochemical Society*, Vol. 126, No. 1, January 1979, pp. 122–130.

[13] K. P. Jeppson and C. M. Svensson, "Negative Bias Stress of MOS Devices at High Electric Fields and Degrading NMOS Devices," *Journal of Applied Physics*, Vol. 48, No. 5, May 1977, pp. 2004–2014.

Diversity measurements of 11.6-GHz rain attenuation at Etam and Lenox, West Virginia

D. V. ROGERS AND G. HYDE

(Manuscript received February 2, 1979)

Introduction

Sky noise temperature variations at 11.6 GHz are being measured simultaneously along slant paths from the Etam, West Virginia, earth station and a site near Lenox, West Virginia, about 35 km north-northeast of Etam, at an elevation angle of 18° and azimuth of 114° (clockwise from north). Rainfall and rain rate data are also being collected. The data collected from October 25, 1977, through October 24, 1978, were reduced and analyzed. The results show that, during the 1-year measurement period, excess attenuation was greater than 11 dB at Etam for about 0.033 percent of the time, and at Lenox for about 0.021 percent of the time. By extrapolation of the cumulative distributions of rain attenuation, it is estimated that for 0.017 percent of the year excess attenuation surpassed 14.5 dB at Etam and 12 dB at Lenox. Under a site diversity configuration, rain attenuation in excess of 2.5 dB was measured for 0.017 percent of the time. In addition to these cumulative distributions of attenuation fading statistics, diurnal distributions and fade duration histograms are presented.

The purpose of these measurements is to verify previous estimates of rain-induced attenuation at 11 and 14 GHz along paths to the proposed INTELSAT V satellite orbit stations, and to determine the efficacy of site diversity in overcoming rain impairments. The 11.6-GHz sky noise temperature variations are measured radiometrically along the slant path between these sites and the proposed INTELSAT V satellite orbit station at 24.5°W longitude. Such measurements and the conversion of sky noise temperature variation measurements into attenuation fade data have been previously described [1]. Figure 1 shows site and path geometry, and

D. V. Rogers is a member of the Propagation Studies Department at COMSAT Laboratories.

G. Hyde is Manager, Propagation Studies Department, COMSAT Laboratories.

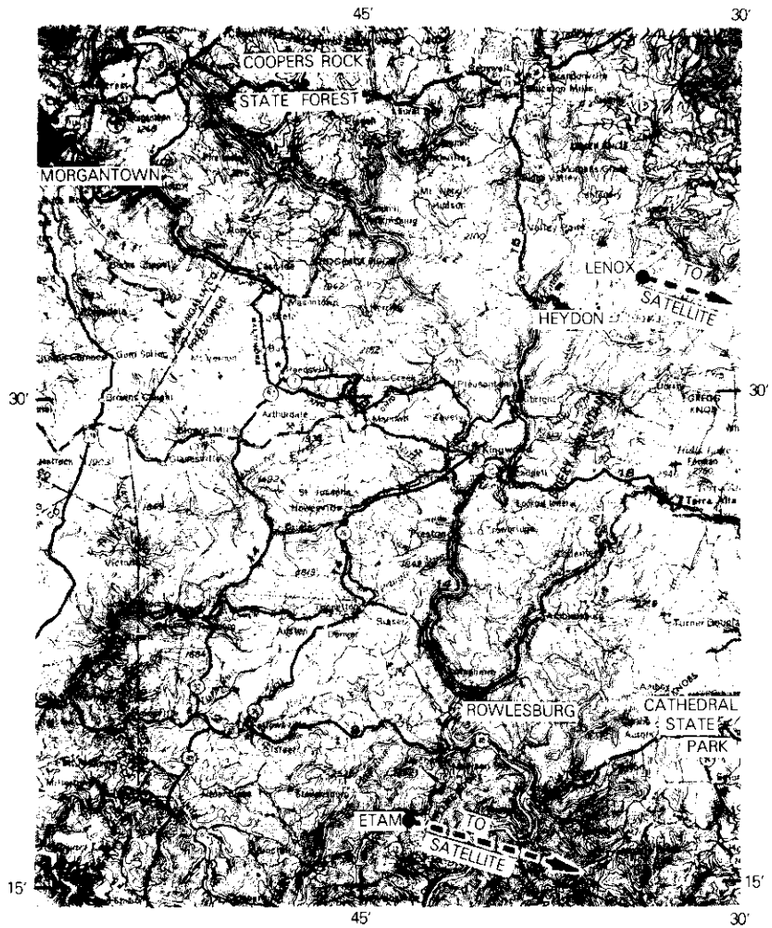


Figure 1. *Etam and Lenox Site Geometry*

Figure 2 is a block diagram of the equipment configuration. Principal system parameters are listed in Table 1. (Note that the IF radiometric receiver at Etam is an AIL Model 777 while that at Lenox is an AIL Model 2392C.) The rain data are collected with a tipping bucket rain gauge.

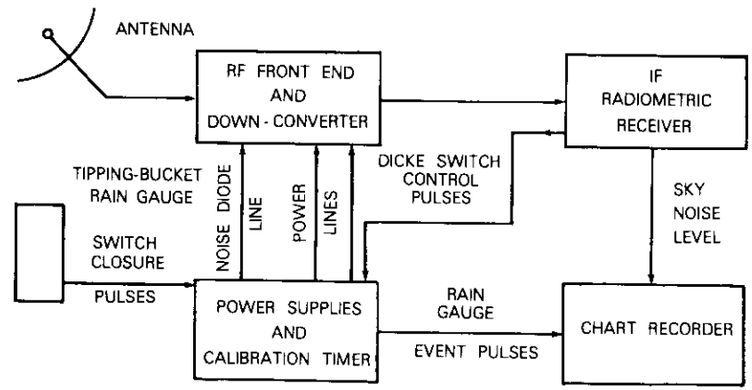


Figure 2. *Block Diagram of Radiometric Receive Terminal*

TABLE 1. PERFORMANCE PARAMETERS OF RADIOMETRIC RECEIVE TERMINAL

	AIL 2392C (Lenox)	AIL 777 (Etam)
IF Bandwidth	100 MHz (5–105 MHz)	100 MHz (5–105 MHz)
Noise Figure	7 dB	<5 dB
Input Impedance	50Ω	50Ω
Input VSWR	<2:1 over entire passband	<2:1 over entire passband
IF Gain	85 dB (maximum)	38 dB (input power level ≥ −60 dBm)
Dicke Switch Rate	5–500 Hz (400 Hz used)	25–500 Hz (400 Hz used)
Integration Time Constant	0.1, 1, 3, 5, 10 and 30 s (3 s used normally)	0.1, 0.5, 1, 3, 5, and 10 s (3 s used normally)
Antenna Gain	44 dB (at 11.6 GHz)	
Antenna Beamwidth	1° (half power beamwidth)	
Mixer Noise Figure	7 dB	

Measured data

The data are collected on 2-channel chart recorders at each site. Sky noise variations are recorded on one analog channel, and rain data are recorded on the other, using a summing circuit which counts the number of bucket tips in one minute and generates a "step" of corresponding height. Once every hour the radiometer cycles through a self-calibration

during which the receive system is first latched for about a minute to an oven-controlled reference load held at 343 K, which appears to the radiometer as a reference load at about 337 K after allowing for small losses. The receive system is then switched back to the antenna while a known increment of noise power from a noise diode is coupled for one minute into the RF front end and summed with the antenna and sky noise contributions. The radiometer is then restored to normal operation. In addition, about once a month the system is manually calibrated with a "cold" load (RF absorber) soaked in liquid nitrogen to provide a reference temperature of about 77 K and then with a "hot" load of RF absorber at the ambient temperature. At present, both radiometer and rain data are being obtained at Etam and Lenox. However, rain statistics for Lenox were unavailable for the measurement period analyzed.

Data reduction and analysis

The chart paper traces were reduced with a digitizer to record samples on computer cards of the amplitude of the events, *i.e.*, changes in the 11.6-GHz sky noise temperature or rain gauge steps. Calibration and event identification information was also encoded. The resultant card files were processed on the COMSAT IBM 360/65 to form standard files that can be processed by a collection of standard propagation measurement analysis subroutines to produce statistical analyses of the data base. Outputs include the cumulative distributions of excess attenuation, both for the individual slant paths from Etam and Lenox and for path diversity, *i.e.*, selecting the lesser excess attenuation from either data file for a given sample time. Also, the data were analyzed to provide the diurnal dependence of attenuation, which is presented as the number of minutes during the measurement year that attenuation fade events exceeded a given level of attenuation for each hour of the day, and the duration histogram of fade events, which is the number of events that exceeded a given level of attenuation as a function of event duration.

Results and conclusions

Cumulative excess attenuation statistics at 11.6 GHz from radiometric measurements along slant paths from Etam and Lenox and from path diversity analysis of the joint 11.6-GHz attenuation statistics are presented in Figure 3. These curves evidence that 11.6-GHz rain attenuation

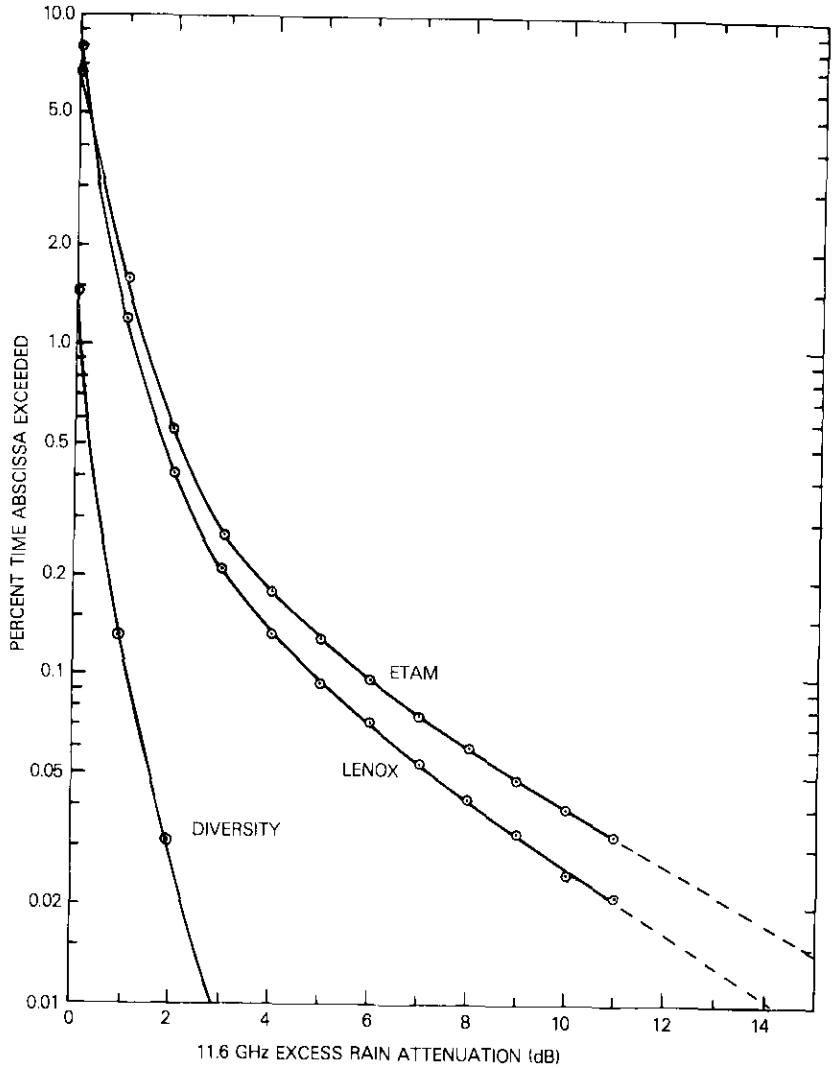


Figure 3. *Cumulative Distributions of 11.6-GHz Excess Rain Attenuation Observed at Etam and Lenox from October 25, 1977, to October 24, 1978, and the Distribution Observed for Path Diversity Between These Sites During the Same Period (Straight-line extrapolations for the single-site distributions for attenuations above 11 dB are designated by the dashed lines.)*

in excess of 11 dB occurred for about 0.033 percent of the year along the slant path from Etam and for about 0.021 percent of the year along the slant path from Lenox. If straight-line extrapolation of these curves to the 0.017-percent excess attenuation level is assumed to be valid, then, for this time percentage, attenuation levels of about 14.5 dB and 12 dB were exceeded at Etam and Lenox, respectively. (The extrapolation method probably yields somewhat optimistic results.) Rain-induced attenuation fading was generally more severe for Etam than for Lenox, as shown in Figure 3.

With diversity switching between the Etam and Lenox paths, 11.6-GHz rain attenuation exceeded about 2.5 dB for 0.017 percent of the year. For the Etam and Lenox single-site fading distributions, therefore, diversity gains of about 12 dB and 9.5 dB, respectively, were observed for 0.017 percent of the year. (The highest diversity attenuation level for which significant data were measured was about 4 dB; this level was exceeded for about 0.005 percent of the year.) Hence, path diversity would have provided a significant improvement over the single path performance achievable at Etam or Lenox.

These two sites were selected not only because they were about 35 km apart, but also because they were separated by well-defined geographic features, including the Cheat River Valley and Laurel and Briery Mountains. It was believed that these features would help to prevent the simultaneous interception of both propagation paths by intense rain. While the synoptic scale climate is the same for both sites and mesoscale systems usually affect both sites simultaneously, terrain features between the sites may diminish the deleterious effects of smaller structures such as intense rain cells. Since the diversity gains measured at Etam and Lenox were greater than the corresponding gains computed with Hodge's empirical expression [2] for diversity gain (exceeding the calculated diversity gain by approximately 1 dB for levels below 1 percent of the time), there may have been beneficial effects on path diversity due to the mountainous terrain in this region.

Figure 4 shows the cumulative distribution of rain rate measured at Etam. The total accumulation of precipitation measured by the heated rain gauge was 1,506 mm for the measurement year. It is estimated that about 1,250 mm was rain, and that the remainder was snow. An analysis of these data shows that a Rice and Holmberg [3] rain rate distribution with $M = 1{,}250$ mm and $\beta \cong 0.4$ (see Figure 4) fits acceptably down to approximately the 0.05-percent time level, but significantly underestimates the observed rain rate below this level.

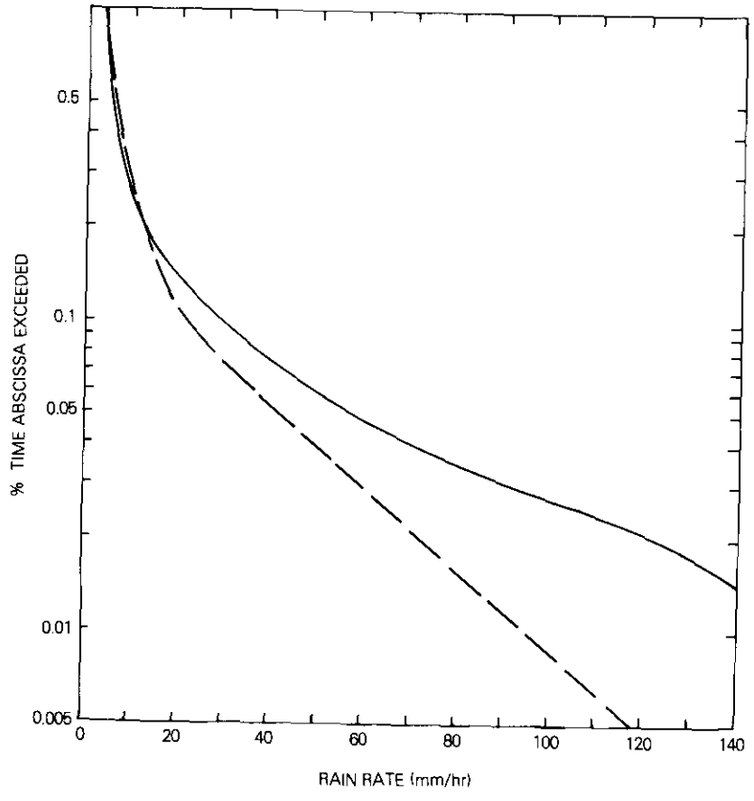


Figure 4. *Cumulative Distribution of Rain Rate Observed at Etam, West Virginia, from October 25, 1977 to October 24, 1978 (For comparison, the rain rate distribution predicted by the Rice-Holmberg model for a yearly rainfall M of 1,250 mm and "thunderstorm" component β of 0.4 is also shown as a dashed line.)*

The diurnal distributions of attenuation fading for the Etam and Lenox paths are given in Figures 5 and 6, respectively. Each vertical bar represents the total number of minutes of fading greater than the indicated level, as measured during a particular hour of the day. No consistent diurnal pattern appears in these distributions. Overall, the diurnal distributions of fades at the two sites show only that during the hours of 4–5,

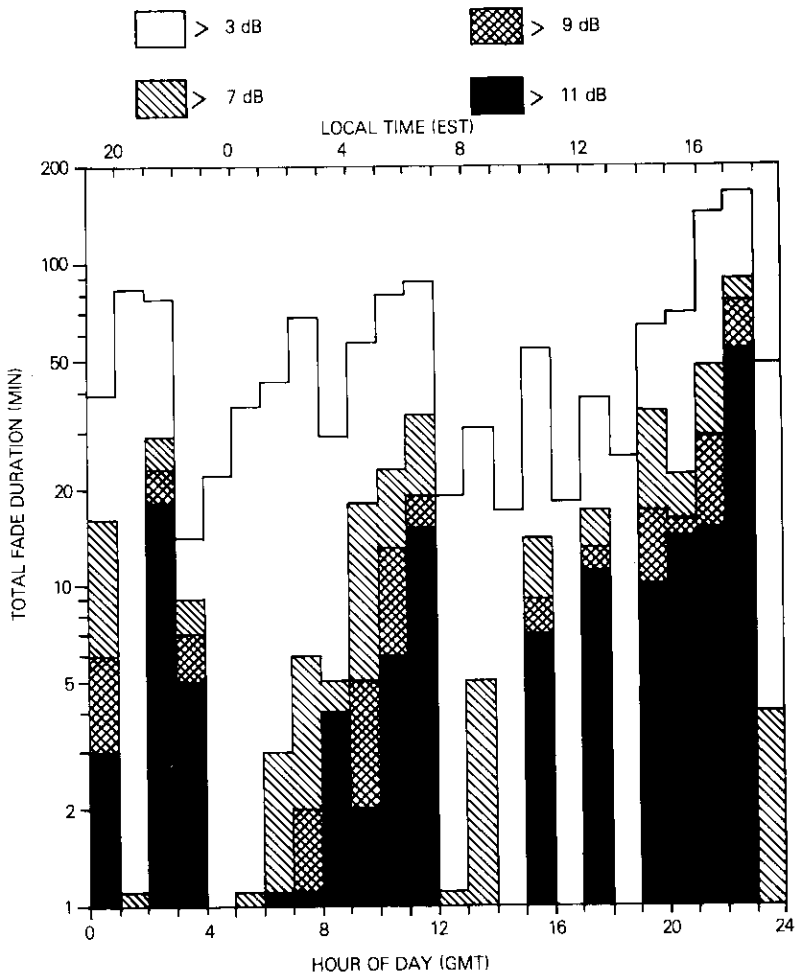


Figure 5. *Diurnal Distribution of Durations (min) that 11.6-GHz Excess Rain Attenuation was greater than the value on ordinate at Etam, West Virginia, from October 25, 1977, to October 24, 1978*

12–13, 14–15, 18–19, and 23–24 GMT almost no fading above about 7 dB occurred at either site. Correspondingly, hours of high fading were distributed throughout the day and night. Therefore, these data provide no basis for assuming a pronounced diurnal distribution of severe fades.

Histograms of fade duration for the 11.6-GHz attenuation events on

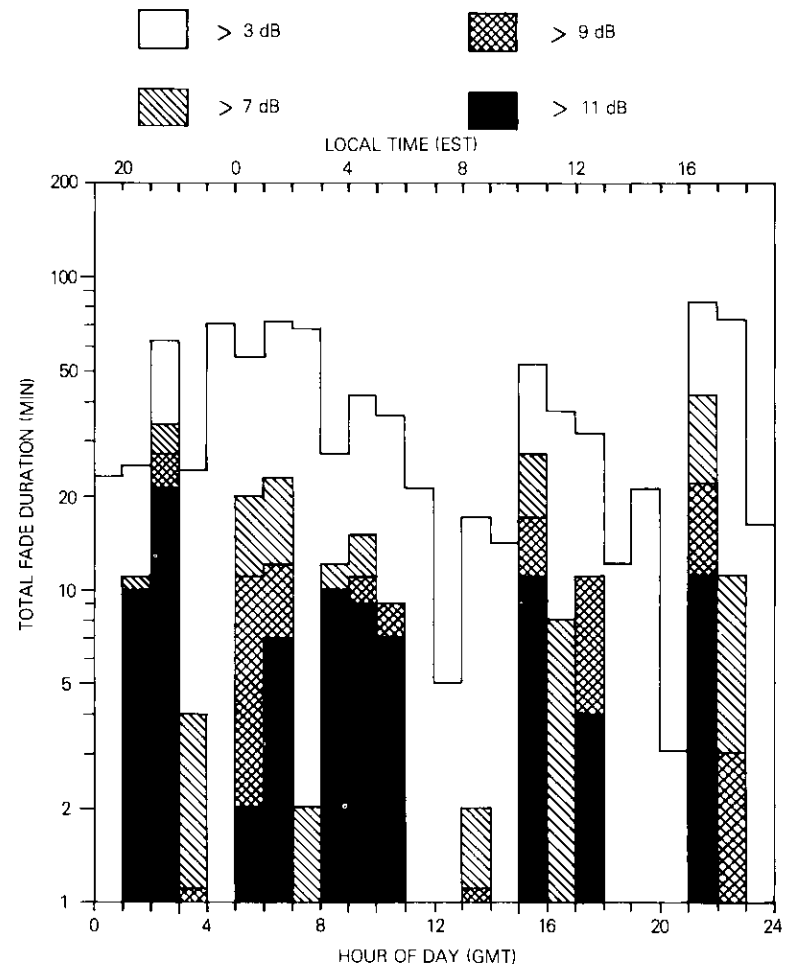


Figure 6. *Diurnal Distribution of Durations (min) that 11.6-GHz Excess Rain Attenuation was greater than the value on ordinate at Lenox, West Virginia, from October 25, 1977, to October 24, 1978*

the Etam and Lenox paths are given in Figures 7 and 8, respectively. As with the cumulative distribution of attenuation, the results demonstrate that for the measurement year fading was more severe, *i.e.*, attenuation events were more numerous and longer lasting, on the path from Etam than on that from Lenox. However, the general nature of the distributions

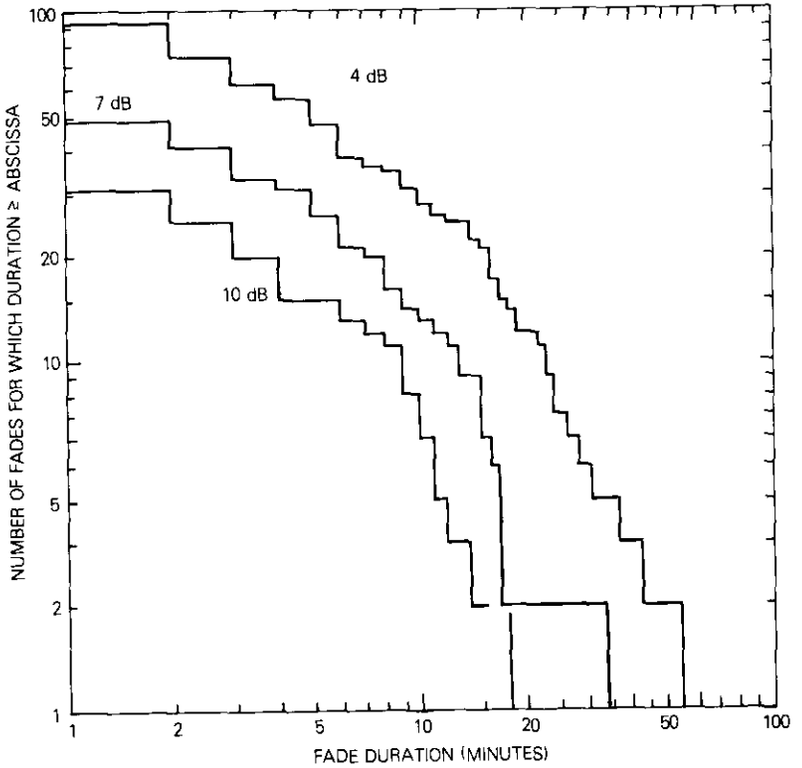


Figure 7. *Fade Duration Histogram Observed for Etam, West Virginia, from October 25, 1977, to October 24, 1978 (Histograms are shown for excess rain attenuation levels of 4, 7, and 10 dB.)*

is the same. No single event was responsible for the observed large number of minutes in long fades at either site. Etam had fifteen 10-dB fades of 5 minutes or longer and six 10-dB fades of 10 minutes or longer. There were nine 10-dB fades of 5 minutes or longer at Lenox and three of 10 minutes or longer.

Acknowledgment

The authors would like to acknowledge the personnel of the Etam earth station, in particular, Mr. R. Parsons, for collecting these data; Mrs. O. Piontek, Mrs. K. Yeh, and Mrs. J. Wilhoite for their work in data reduction and computer analysis; Messrs. F. Lee and R. Trushel for their assistance with computer software; and finally to Messrs. W. Patterson and B. Williams, without whose project support none of this work would have been undertaken.

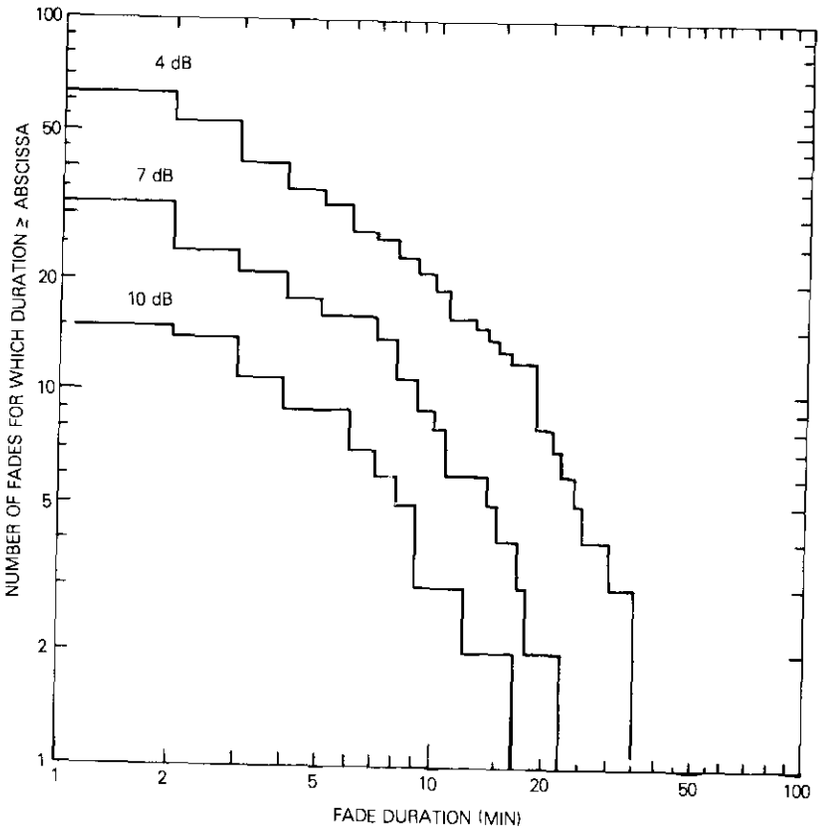


Figure 8. *Fade Duration Histogram Observed at Lenox, West Virginia, from October 25, 1977 to October 24, 1978 (Histograms are shown for excess rain attenuation levels of 4, 7, and 10 dB.)*

References

[1] H. Meyerhoff, A. Buige, and E. A. Robertson, "15.3-GHz Precipitation Attenuation Measurements Using a Transportable Earth Station at Utibe, Panama," *COMSAT Technical Review*, Vol. 4, No. 1, Spring 1974, pp. 169–187.

[2] D. B. Hodge, "An Empirical Relationship for Path Diversity Gain," *IEEE Transactions on Antennas and Propagation*, AP-24, No. 2, March 1976, pp. 250–251.
[3] P. L. Rice and N. R. Holmberg, "Cumulative Time Statistics of Surface-Point Rainfall Rates," *IEEE Transactions on Communications*, COM-21, No. 10, October 1973, pp. 1131–1136.

Coplanar waveguide FET amplifiers for satellite communications systems

R. E. STEGENS

(Manuscript received January 15, 1979)

Introduction

Since its introduction by C. P. Wen in 1969 [1], [2], coplanar waveguide (CPW) has found little application in practical MICs even though it offers the advantage of having all the conductors on the same surface. This arrangement provides a convenient ground return path for components mounted on the CPW surface, thereby obviating the need for metal ridges and separate substrates, which are required for microstrip. However, studies of simple CPW structure behavior have produced poor agreement between measured and predicted performance. Some researchers have concluded that CPW is inferior to microstrip in virtually every way [3].

Recent efforts at COMSAT Laboratories have shown that selected transmission line structures such as shunt-open or short-circuited stubs can be accurately realized in CPW form provided that certain restrictions are placed upon the circuit topology, including enforced symmetry about the propagating line and the use of bond wires to ensure equipotential ground plane surfaces. Under these conditions, the expected advantages of CPW transmission lines in the design of FET amplifiers, for example, have been realized; subsequently, in mid-1978 successful single-stage amplifiers were demonstrated.

Although the requirements for symmetry and the use of bond wires at each junction may place restrictions upon the circuit designer who is familiar with the sophisticated circuits and techniques used with fused-silica microstrip microwave integrated circuits (MICs), the CPW circuits

R. E. Stegens is Assistant Manager of the Microwave Circuits Department at COMSAT *Laboratories.*

tested at COMSAT Laboratories have demonstrated unexpected benefits. These include simplified circuit mounting techniques, easy cascading of individual circuits, reduced radiation, and the possibility of realizing circuit elements on the CPW which are not available on microstrip, such as series connected transmission lines. It has also been revealed that circuit adjustment can be performed more easily than on microstrip, provided that the designer recognizes the need to utilize CPW elements which can be easily varied; for example, short-circuited shunt transmission lines rather than open-circuited lines.

The following results indicate that CPW MIC structures can now be designed which significantly reduce the mechanical complexity and fabrication time of MIC components and subsystems utilized in earth terminals and spacecraft. Thus, this transmission line configuration has the potential for significantly reducing the MIC production costs.

Launcher and frame

A mounting system has been designed* which incorporates SMA launchers and accommodates single or cascaded CPW circuits of any length with a simple mechanical adjustment. The SMA/CPW transition is made via a 3-point pressure contact on the top surface of the MIC board. The CPW substrate is placed in the mounting structure and secured, the transition assemblies are positioned, and the pressure contact is adjusted. This fixture demonstrates that simple yet versatile mounting structures can be designed for use with CPW circuits. A similar versatile system has not been found with microstrip MICs.

CPW dielectric selection

CPW parameters were studied to select the proper dielectric material to support the CPW circuit. The dielectric must provide good mechanical strength while maintaining a well-controlled dielectric constant of a magnitude capable of producing a broad range of CPW impedances. The material should be suitable as a base for metallization layers as thin as 0.0254 mm. Based upon the characteristics in Table 1, alumina ($\epsilon_r = 10$) was selected as the best material for initial tests. A substrate thickness of 1.27 mm was used for initial test circuits; later 2.54-mm material was procured.

*Patent application has been filed.

TABLE 1. CPW PARAMETERS FOR VARIOUS DIELECTRIC MATERIALS

Material	ϵ_r	Impedance Range (Ω)	Specific Stiffness [a]	Line Resolution ($\times 10^{-3}$ cm)
"Softboard" [b]	2.5	51–180	0.1–0.4	2.5
Fused Silica	3.8	36–180	2.0 (Fragile)	2.5
Al_2O (Alumina)	10	24–118	10.7	0.5
$BaTi_4O_9$	38	14–65	≅10	0.5

[a] Modulus of elasticity in tension divided by density. Units are meters.
[b] For example, teflon-glass, polystyrene, and polyphenylene oxide.

Theoretical considerations

Figure 1 defines the CPW line parameters. It was found that, for

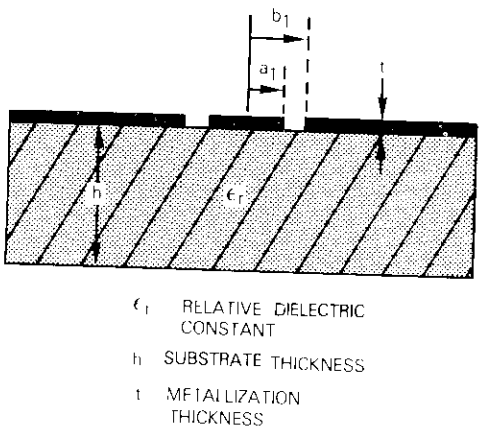


Figure 1. *CPW Transmission-line Definitions*

$t \ll (b_1 - a_1)$ and $h \gg (b_1 - a_1)$, Wen's values for impedance and phase velocity agreed to within 2 percent with data computed using a relaxation technique [4]:

$$Z_o = 1/(CV_p)\Omega$$

$$V_p = V_o\sqrt{\frac{2}{\epsilon_r + 1}} \quad m/s$$

where $$C = 2(\epsilon_r + 1)\epsilon_o \frac{K(a_1/b_1)}{K(\sqrt{1 - a_1^2/b_1^2})} \; F/m \quad .$$

In these equations, ϵ_o is the permittivity of free space, V_o is the speed of light, Z_o is the CPW characteristic impedance, V_p is the phase velocity, and K is the complete elliptic integral of the first kind. Figure 2 shows Z_o for alumina dielectric, with $\epsilon_r = 10$, negligible t, and large h. The effective dielectric constant is therefore 5.5, independent of a_1/b_1.

A study of the theoretical effect of nonzero conduction thickness [4] was performed since this effect had only partially been treated [5]. The results shown in Figure 3 indicated that an increase in impedance and a decrease in effective dielectric constant should be expected even with very thin metallization when the line dimensions are small.

SMA/CPW transition tests

To minimize the SMA/CPW discontinuity, the line width was made equal to the outer diameter of the coaxial connector by using a Cablewave 971 connector ($2b_1 = 1.27$ mm). A 5.08-cm (50Ω) CPW line was constructed on 1.27-mm alumina and measured from 2 to 18 GHz with an SMA/CPW transition on each end. The results (Figure 4) indicate that the transitions are quite good except in the 12- to 14-GHz region, where each exhibits a return loss of 13 dB.

Figure 5 shows the measured losses, which closely followed the $\sqrt{f}$ theoretical dependence from 2 to 12 GHz. The relatively high loss (0.35 dB/cm at 12 GHz) may be due to the thin (2-μm) CPW metallization.

Elementary circuit element tests

Simple circuits consisting of shunt open and short-circuited transmission lines were etched on 2.5- × 2.5-cm CPW substrates and measured for comparison with computer models. Good agreement was observed for circuit elements which were symmetrical about the primary transmission line with longitudinal bond wires used to maintain an equal potential across breaks in the ground plane.

It was determined that asymmetrical circuits could be used, provided that transverse bond wires were added to ensure an equal potential on either side of the CPW line. Figure 6 shows an example of measured results; a single shunt short-circuited CPW transmission line, designed to be 180° long at 6.0 GHz, was etched in the center of a 5.08-cm CPW line. With three pairs of bond wires used at the junction, a resonator Q of 145 was indicated at 12 GHz and the transmission characteristic corresponded to

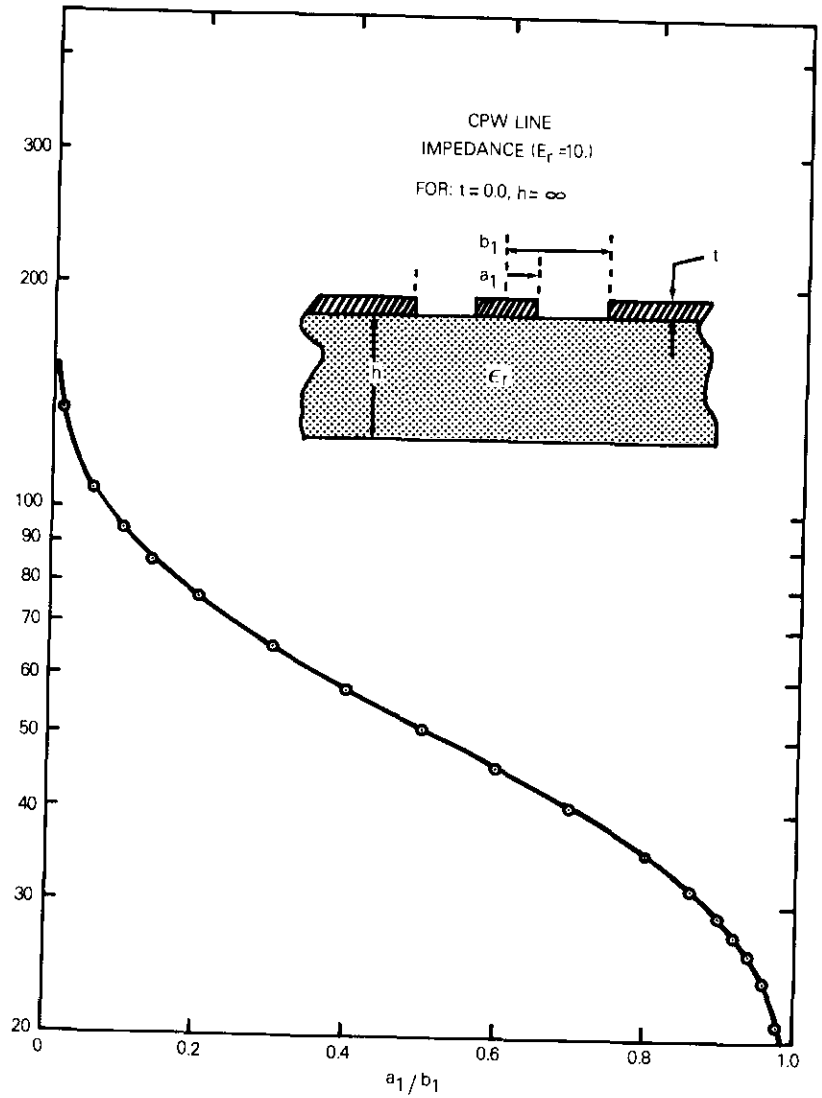


Figure 2. *CPW Line Impedance from Wen [1]*

theory up to 18 GHz. The number and length of the bond wires were found to be uncritical, and there was no dispersion. This indicates that a

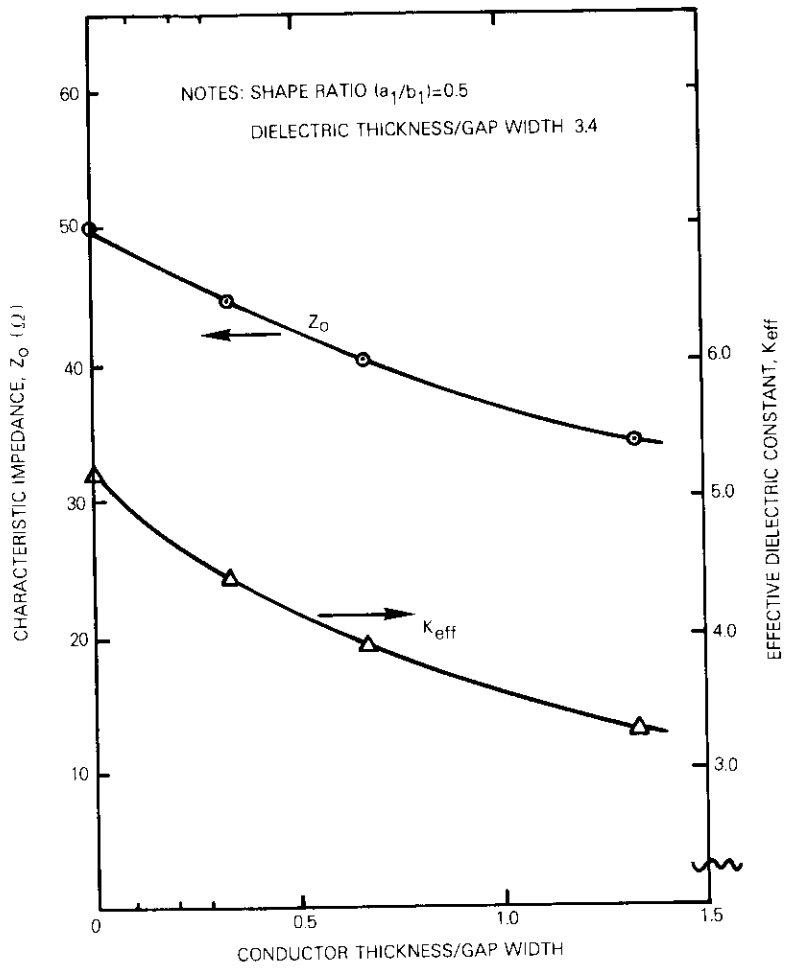


Figure 3. *Line Parameter Dependence on Conducting Strip Thickness*

CPW line with 1-mil gaps and 2.2-mil center conductors should be well behaved up to at least 108 GHz.

With the junction bond wires sequentially removed from the test circuit, measured performance did not agree with theory. Use of two identical circuit elements on either side of the transmission line eliminated the need for all except the longitudinal bond wires.

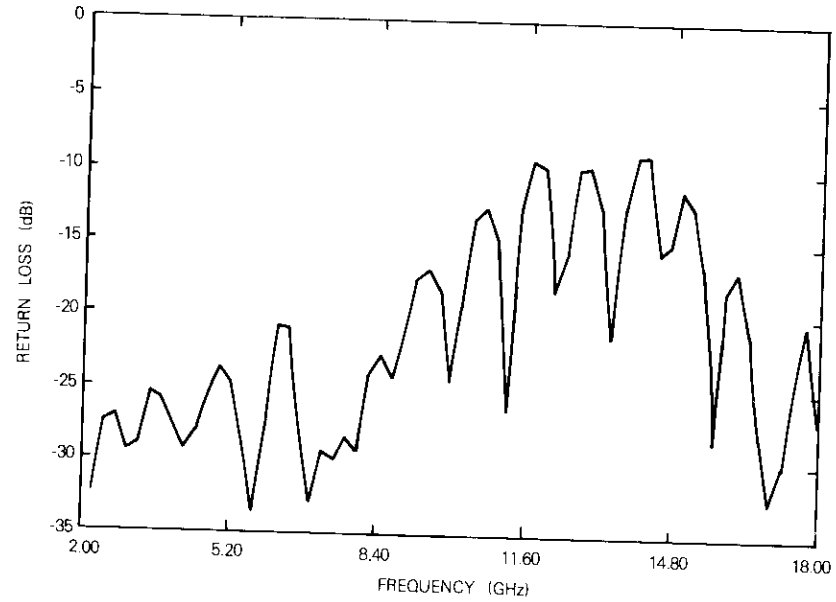


Figure 4. *Measured Return Loss of a 5.08-cm CPW Line with SMA Transitions (Termination is an SMA load)*

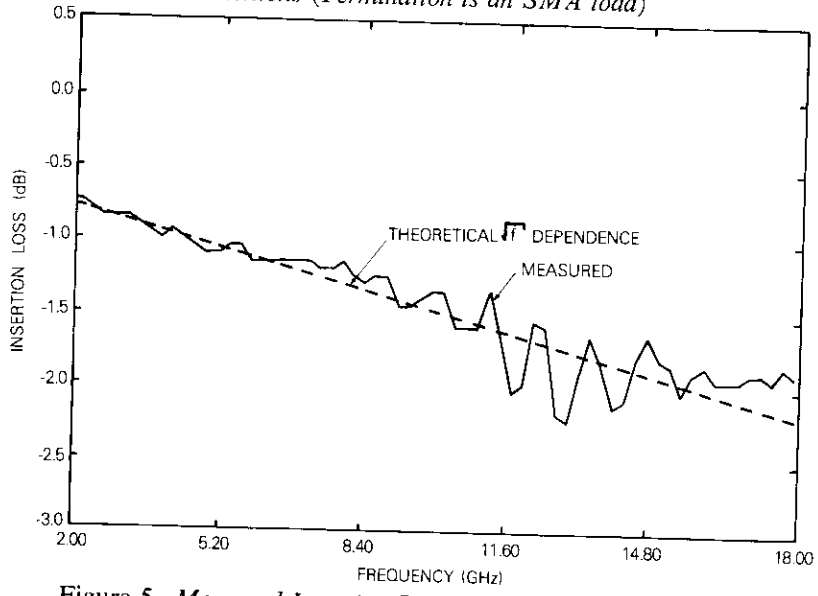


Figure 5. *Measured Insertion Loss of a 5.08-cm CPW Line*

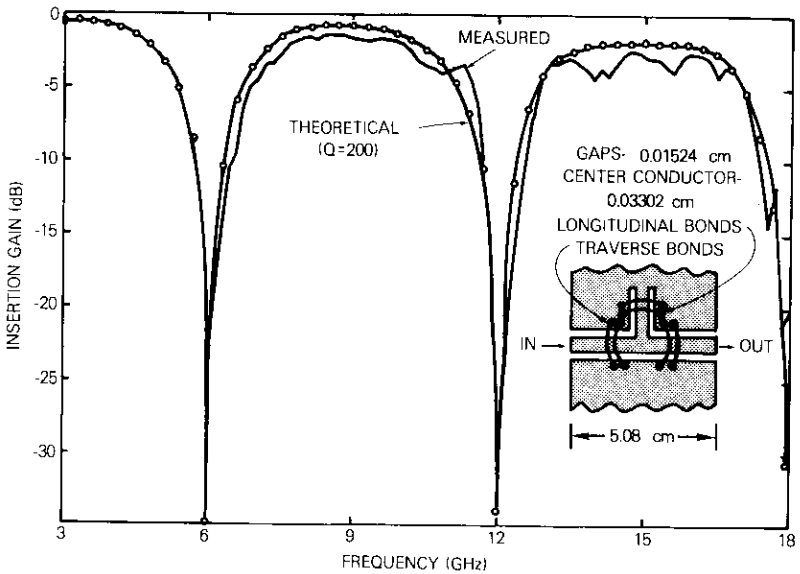


Figure 6. *Measured and Theoretical Transmission Loss for a CPW Shunt Shorted Stub on a 5.08-cm Through Line*

FET amplifier realization and tests

A single-stage FET amplifier using NEC's 388 chip was designed using only symmetrical lines, longitudinal bond wires at each junction, and impedance levels well within the range of the CPW on alumina. The amplifier was computer modeled for comparison with measured data. Figures 7 and 8 show the equivalent circuit and a photograph of the amplifier, respectively.

The results (Figure 9) show good agreement with theory provided that the effective reduction in dielectric constant caused by the metallization thickness (18×10^{-3}mm) is considered. Figure 3 indicates an effective dielectric constant of 4.8 for the open-circuited stub at the output of the amplifier. It was concluded that the relatively small discrepancy is caused by both nonzero t and unmodeled junction effects, including the longitudinal bond wires.

This single-stage amplifier demonstrated the feasibility of using CPW for future MIC designs subject to the specified constraints. Fabrication time, including etching and cutting the substrate to size and mounting and bonding all components, was about one-third of that for a comparable

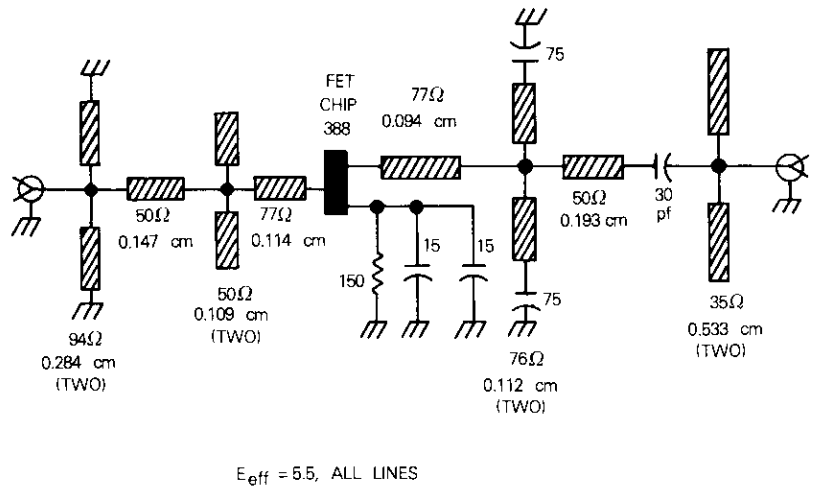


Figure 7. *Equivalent Circuit for 1-Stage CPW FET Amplifier*

microstrip 1-stage amplifier. The mounting structure was also much simpler, since no elaborate MIC carrier was necessary. The amplifier was easier to tune than microstrip versions and did not produce TE_{10} resonances when enclosed in a 2.5- × 2.5- × 2.5-cm box.

Multistage amplifiers

As a test to determine the ease with which CPW circuits can be cascaded, two of the single-stage FET amplifiers in Figure 8 were cascaded by placing the circuits end-to-end in the CPW mounting structure and repositioning the connectors. The resulting 2-stage amplifier performed well (Figure 10) and was easily tuned. The two CPW substrates were connected by bond wires across the center conductors and each ground plane. The gain of 14 dB across 11.7–12.2 dB was as expected, considering the predicted 2-dB circuit losses (Figure 5).

Figure 11 shows a developmental 3-stage FET amplifier on a single 3.8-cm CPW substrate; details of the mounting fixture are visible. The construction of the CPW amplifier may be compared with present microstrip versions which require a precisely machined INVAR carrier with three ridges, four separately sliced fused silica substrates, and a sophisticated mounting structure designed to accommodate one particular amplifier size and suppress box resonances.

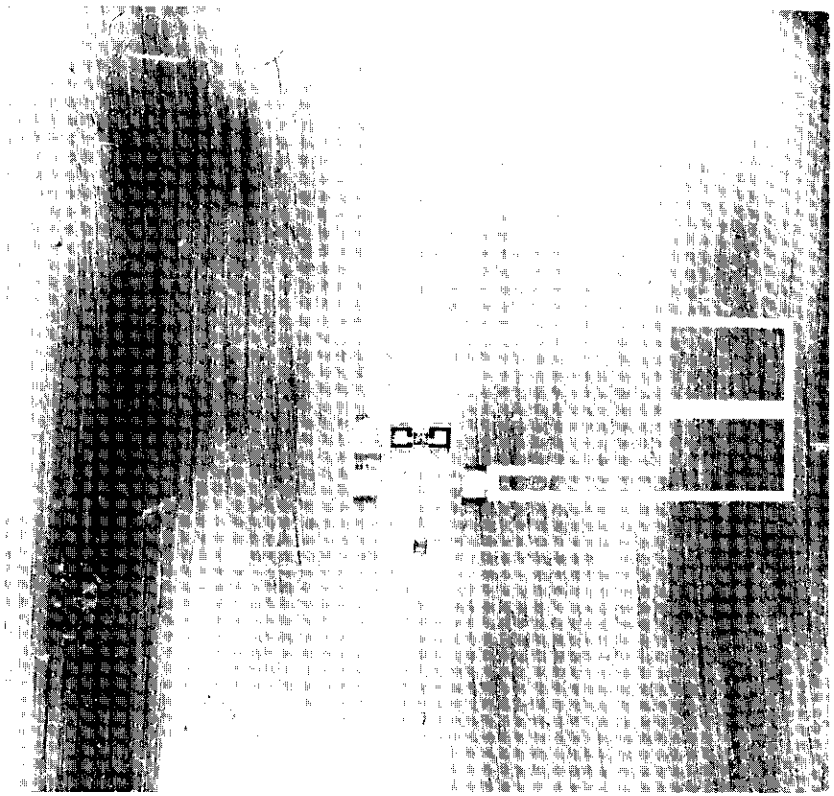


Figure 8. *Photograph of 1-Stage CPW, FET Amplifier (Substrate is 2.5-cm × 2.5-cm × 1.27-mm Al_2O_3)*

Conclusions

It has been demonstrated that MICs such as FET amplifiers operating in the 12/14-GHz frequency bands can be realized in an inexpensive form using CPW. Benefits include reduced fabrication time, less complex mounting structures, and easier cascading of individual circuits. Compared with microstrip, probable additional benefits include reduced radiation and inter-circuit coupling and higher frequency operation.

The losses measured on the CPW transmission line were greater than those on microstrip; however, some of the losses can be attributed to excessively thin CPW metallization. Short-circuited CPW transmission lines were easily adjustable in length; only open-circuited lines can be adjusted

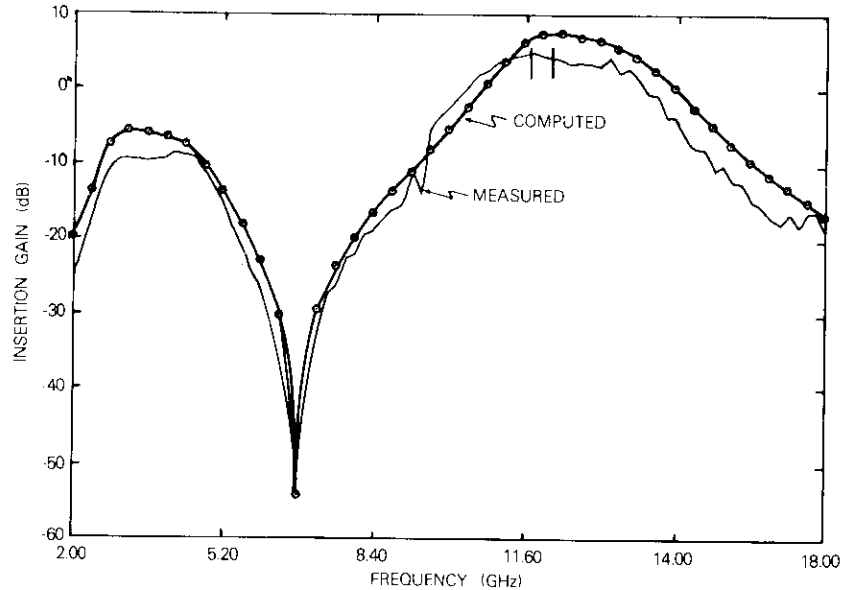


Figure 9. *Computed and Measured Response of 1-Stage CPW FET Amplifier*

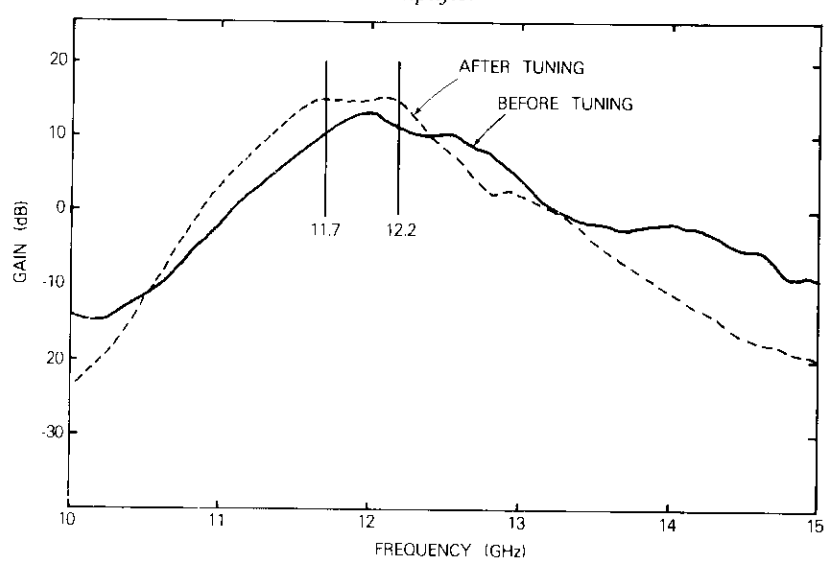


Figure 10. *Performance of Two Single-Stage CPW FET Amplifiers Cascaded in a Single Support Structure*

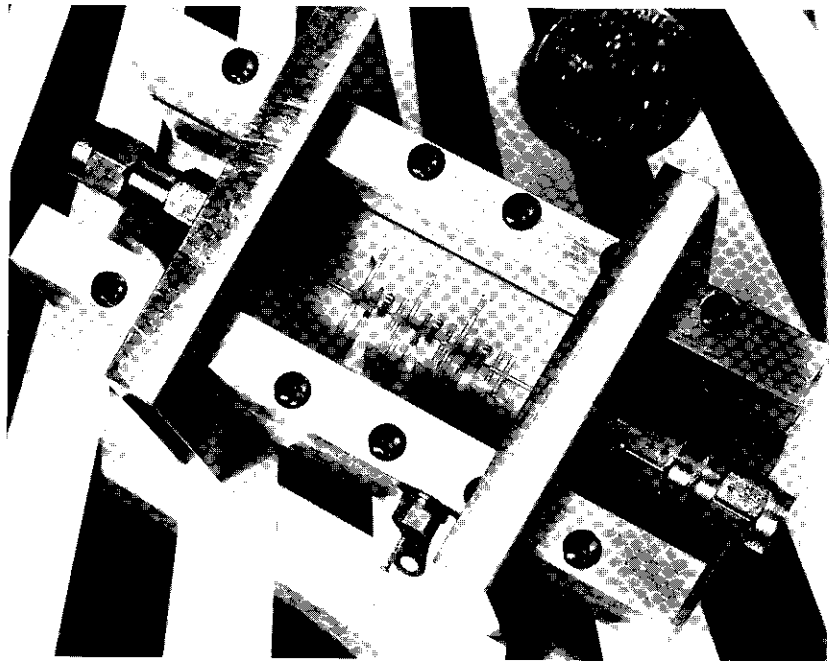


Figure 11. *Developmental 12-GHz 3-Stage CPW FET Amplifier*

on microstrip. The requirement for bond wires at every CPW junction to maintain equipotential ground planes is important.

As the understanding of CPW junctions and circuit elements advances and the fabrication of accurately realized filters and couplers becomes possible, it seems likely that this form of MIC circuit may prove to be superior to microstrip for certain applications, particularly in terms of production costs.

Acknowledgments

The author wishes to thank S. Taylor and J. Molz for the assembly and measurement of the circuits, and T. Morgan for the assembly of the FET *amplifiers. R. Barber etched the circuits, and E. Bainbridge created the rubilith masks. W. Megna generated the photography both for the circuit masks and this note.*

References

[1] C. P. Wen, "Coplanar Waveguide: A Surface Strip Transmission Line Suitable for Nonreciprocal Gyromagnetic Device Applications," *IEEE Transactions on Microwave Theory and Techniques*, MTT-17, No. 12, December 1969, pp. 1087–1090.

[2] C. P. Wen, "Coplanar Waveguide Directional Couplers," *IEEE Transactions on Microwave Theory and Techniques*, MTT-18, No. 6, June 1970, pp. 318–322.

[3] J. McDade and D. Stockman, "Microwave Integrated Circuit Techniques," Final Report AFAL-TR-73-234, May 1973.

[4] Y. S. Lee, Private Communication.

[5] T. Kitajawa et al., "A Coplanar Waveguide with Thick Metal Coating," *IEEE Transactions on Microwave Theory and Techniques*, MTT-24, No. 9, September 1976.

cisión y la velocidad de respuesta en las modalidades normal y de reserva, en la transición de una a otra, en varias submodalidades, y durante la falla de los volantes. Las pruebas fueron respaldadas con análisis y simulaciones por computadora. Cuando el sistema funcionó con un mínimo de tres volantes, los errores de actitud máximos no sobrepasaron el límite admisible, incluso cuando el par torsor del motor falló totalmente. Se demostró la importancia de inyectar impulso a un volante fallido en la matriz de distribución. Además, se corroboró que el control del impulso, en contraste con el control tradicional del par torsor, ofrece muchas ventajas de diseño y funcionamiento. Todo parece indicar que un sistema de volantes de reacción oblícuos sería interesante para mejorar la precisión y la confiabilidad de funcionamiento de los satélites de comunicaciones geoestacionarios.

Desempeño de protocolos para el control de enlaces de datos por canales de comunicaciones multiplejados por división en el tiempo

A. K. KAUL

Abstracto

En este artículo se presentan modelos analíticos para predecir el rendimiento total obtenible con protocolos de alto nivel para el control de enlaces de datos (HDLC), o protocolos similares, utilizando un método de corrección de errores por eliminación (Memoria de N bloques) en canales de comunicaciones sincronizados por multiplaje mediante división en el tiempo (STDM). Estos modelos, aplicables tanto a sistemas que transmiten una trama de datos en ráfagas múltiples como a los que transmiten varias tramas de datos por ráfaga, se pueden emplear en los canales terrestres de multiplaje por división en el tiempo (TDM) y en los canales de satélite de acceso múltiple por división en el tiempo (TDMA). Se resumen los cálculos necesarios, y mediante un sencillo sistema TDMA para satélite se ilustra la manera de utilizar los modelos y la forma en que los diversos parámetros del sistema están subordinados al rendimiento total.

Author Index, CTR 1978

INDEX OF PRESENTATIONS AND PUBLICATIONS BY COMSAT AUTHORS IN 1978

The following is a cross-referenced index of technical publications and presentations by COMSAT authors outside of *COMSAT Technical Review*. The code number at the end of an entry is the reprint code number by which copies may be ordered from Lab Records at COMSAT Laboratories, P.O. Box 115, Clarksburg, MD 20734.

ALI, Z. M., see Dill, G. D. [78CLR60].

ALI, Z. M., "A Configurable Parallel Arithmetic Structure for Recursive Digital Filtering," *IEEE PROC. Int. Symp. on Circuits & Syst.*, May 78, pp. 289–296 [78CLR29].

ALI, Z. M., "A High-Speed FFT Processor," *IEEE COM-26*, May 78, pp. 690–696 [78CLR26].

ANDERSEN, J. B.* & Zaghloul, A. I., "Wide-Angle Scan of Linear Arrays," *IEEE AP-S Int. Symp. May 78 DIGEST*, pp. 174–177 [78CLR27].

ARROYO, B.†, see Wachs, M. R. [78CLR36].

ATIA, A. E., see Zaki, K. A.* [78CLR30].

BARGELLINI, P. L., "A Review of U.S. Satellite Communications Systems Technology," Pres. at EUROSPACE-INICTEL Symp. on Comm. Sat., May 78 [UPO64CL].

BARGELLINI, P. L., "Evolution of U.S. Domestic Satellite Communications," *3rd Jerusalem Conf. on Inf. Tech. Aug. 78 PROC.*, J. Moneta, ed., North Holland Publishing Co. [78CLR46].

BARGELLINI, P. L., "Invariants in the Teaching of Electrical Science," *IEEE E-21*, Feb. 78, pp. 22–25 [78CLR01].

BARGELLINI, P. L., "Principles and Evolution of Satellite Communications," *Acta Astronautica*, Mar.–Apr. 78, pp. 135–149 [78CLR18].

BARGELLINI, P. L., "The Impact of Technology on the Capacity and Economics of Communications," Pres. at EUROSPACE VIIth U.S.-European Conf., Sept. 78 [UPO65CL].

BARRETT, M. F., see Hsing, J. C. [78CLR11].

BERMAN, A. L., see Mahle, C. E. [78CLR51].

BETZ, F. E.*, Dunlop, J. D. & Stockel, J. F., "The First Year in Orbit for the NTS-2 Nickel Hydrogen Battery," Int. Energy Conv. Eng. Conf. *PROC.*, Vol. 1, Aug. 78, pp. 67–73 [78CLR76].

BROWN, M. P., see Dicks, J. L. [78CLR32].

---

*Non-COMSAT author.

†INTELSAT assignee.

BURKITT, F. J., see Potts, J. B. [78CLR04].
BURWELL, C., see Dill, G. D. [78CLR67].
CALVIT, T. O., "MARISAT: A Progress Report," Pres. at Radio Tech. Comm. for Marine Services, Wash., D.C., Nov. 78, *SYMP. PAPERS*, Vol. 1, 11 pages [78CLR80].
CAMPANELLA, S. J., see Chakraborty, D. [78CLR62].
CAMPANELLA, S. J., see Hodson, K. [78CLR63].
CAMPANELLA, S. J., "Digital Speech Interpolation Techniques," *NTC Dec. 78 CONF. REC.*, Vol. 2, pp. 14.1–14.1.5 [78CLR86].
CAMPANELLA, S. J. & Pontano, B. A.,* "The INTELSAT TDMA Field Trial," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 299–303 [78CLR61].
CASEY, J. P., see Van Trees, H. L. [78CLR05].
CAUGHRAN, P. M., see Early, L. B. [78CLR21].
CHAKRABORTY, D., Noguchi, T.,† Campanella, S. J. & Wolejsza, C. J., "Digital Modem Design for Nonlinear Satellite Channels," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 122–130 [78CLR62].
CHIDAMBARAM, T. S., see Van Trees, H. L. [78CLR05].
CHIDAMBARAM, T. S., "A Forecasting Model for Evaluating the Potential of a New Communications Service," *ITU Telecom. J.*, Oct. 78, pp. 541–546 [78CLR75].
CHILDS, W. H., See Mahle, C. E. [78CLR51].
CHILDS, W. H., Mahle, C. E. & Potukuchi, J., "An Integrated DQPSK Demodulator for 14-GHz Satellite Communications Applications," *IEEE MTT-S Int. Microwave Symp. June 78 DIGEST*, pp. 64–66 [78CLR44].
COLLINS, G. D., see Dill, G. D. [78CLR66].
COOK, W. L., Kaiser, J., Marchese, J.,* Hodge, G.* & Popot, M.,* "An Experiment in High-Speed Computer Communications via Satellite," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78–604 [78CLR12].
COOL, R. W., see Curtin, D. J. [78CLR50] and Curtin, D. J. [77CLR72].
COOPERMAN, R. S. & Kasser, J., "A Receive-Only MARISAT Terminal," *IEEE EASCON '78 REC.*, pp. 356–361 [78CLR52].
CURTIN, D. J. & Cool, R. W., "Testing of Ultra-Thin Solar Cells," *IEEE 13th Photovoltaic Spec. Conf. June 78 PROC.*, pp. 438–443 [78CLR50].
CURTIN, D. J. & Cool, R. W., "Thermal Cycle and Charge Tests of Solar Array Modules," Pres. at Photovoltaic Solar Energy Conf. Sept. 77, Luxembourg, *PROC.*, Holland: D. Reidel Pub. Co., pp. 695–709 [77CLR72].
DAVIS, R. C., see Edelson, B. I. [78CLR24].
DAVIS, R. C., Esch, F. H., Palmer, L. & Pollack, L., "Future Trends in Communications Satellite Systems," *Acta Astronautica*, Mar.–Apr. 78, pp. 275–298 [78CLR22].
DEAL, J. H., see Watanabe, T.* [78CLR65].

*Non-COMSAT author.
†INTELSAT assignee.

DEAL, J. H., "Digital Transmission Involving Intersatellite Links," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 160–165 [78CLR64].
DEVIEUX, C., see Fang, D. J. [78CLR53].
DICKS, J. L. & Brown, M. P., "INTELSAT V Satellite Transmission Design," *IEEE ICC June 78 CONF. REC.*, pp. 2.2.1–2.2.5 [78CLR32].
DiFONZO, D. F., see Kreutel, R. W. [78CLR17].
DiFONZO, D. F., "Antennas: Key to Communications Satellite Growth," *Microwave Syst. News*, June 78, pp. 83–91 [78CLR25].
DILL, G. D. & Ali, Z. M., "Application of Transmultiplexers at Satellite Earth Stations," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 225–230 [78CLR60].
DILL, G. D., Burwell, C., Edy, R., Hodson, K.†, Umeda, Y.*, Fumino, T.*, Ito, H.* & Tachikawa, S.*, "120-Mbit/s TDMA Test Bed," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 291–298 [78CLR67].
DILL, G. D., Muratani, T.*, Tsuji, L. & Collins, G. D., "Simulated SS-TDMA System Test Results," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 175–179 [78CLR66].
DORIAN, C., "The MARISAT System," *Acta Astronautica*, Mar.–Apr. 78, pp. 243–260 [78CLR20].
DOUGHERTY, H. T.*, Estin, A. J.* & Morgan, W. L., "The Orbiting Standards Platform," Pres. at 1978 Antenna Appl. Symp., Sept. 78, *PROC.* [78CLR77].
DUNLOP, J. D., see Betz, F. E.* [78CLR76].
DUNLOP, J. D., see Stockel, J. F. [78CLR07].
EARLY, L. B., Reber, C. J. & Caughran, P. M., "Economics of Communications Satellite Systems—1976," *Acta Astronautica*, Mar.–Apr. 78, pp. 261–273 [78CLR21].
EDELSON, B. I., see Morgan, W. L. [78CLR08].
EDELSON, B. I. & Davis, R. C., "Satellite Communications in the 1980's and After," Pres. at Royal Soc. Discussion, London, Mar. 77, *Telecommunications in 1980's and After*, London: The Royal Society, 1978, pp. 159–174 [78CLR24].
EDELSTEIN, F.* & Flieger, H. W., "Satellite Battery Temperature Control," Pres. at Int. Heat Pipe Conf., May 78, AIAA Paper No. 78–448 [78CLR48].
EDY, R., see Dill, G. D. [78CLR67].
ESCH, F. H., see Davis, R. C. [78CLR22].
ESTABROOK, P.*, Krowne, C. M.*, Crescenzi, E. J.* & Stegens, R. E., "A Low Noise Single-Ended GaAs Schottky FET Amplifier for a 14 GHz Satellite Communication Application," *IEEE MTT-S Int. Microwave Symp. June 78 DIGEST*, pp. 119–131 [78CLR45].
FANG, D. J., "Modulation Transfer from TDMA On-Off Bursting to Carriers in Memoryless Nonlinear Devices, Part II—Baseband Performance," *IEEE COM-26*, Apr. 78, pp. 439–449 [78CLR16].

*Non-COMSAT author.
†INTELSAT assignee.

*Non-COMSAT author.
†INTELSAT assignee.

*Non-COMSAT author.
†INTELSAT assignee.

MORGAN, W. L., "Communications Satellites—1973 to 1983," *IEEE ICC June 78 CONF. REC.*, pp. 2.1.1–2.1.3 [78CLR31].
MORGAN, W. L., "Earth Stations for Satellite Communications," *NTC Dec. 78 CONF. REC.*, Vol. 2, pp. 32.2.1–32.2.3 [78CLR87].
MORGAN, W. L., "Satellite Characteristics Summary," *Acta Astronautica*, May–June 78, pp. 455–466 [78CLR23].
MORGAN, W. L., "Space Stations," *Sat. Comm.*, Apr. 78, pp. 32–39 [78CLR02].
MORGAN, W. L. & Edelson, B. I., "The OAF Concept Extended," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-546 [78CLR08].
MURATANI, T.*, Snyder, J. S., Saitoh, H.*, Koga, K.*, Mizuno, T.* & Yasuda, Y.*, "Application of FEC Coding to the INTELSAT TDMA System," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 108–115 [78CLR72].
NEYRET, P.*, "Polarization Properties of Feed Array for Shaped Beam Antenna," *IEEE AP-S Int. Symp. May 78 DIGEST*, pp. 361–364 [78CLR39].
NOGUCHI, T.†, see Chakraborty, D. [78CLR62].
OLSEN, R. L.*, Rogers, D. V. & Hodge, D. B.*, "The aRb Relation in the Calculation of Rain Attenuation," *IEEE AP-26*, pp. 318–329 [78CLR47].
ONUFRY, M., see Suyderhoud, H. G. [78CLR71].
OWENS, J. R.* & Morgan, W. L., "In-Orbit Operating Experience with the INTELSAT Satellites," *Acta Astronautica*, Mar.–Apr. 78, pp. 151–171 [78CLR19].
PALMER, L., see Davis, R. C. [78CLR22].
PARTHASARATHY, R.† & Kelley, T. M., "Leasing of INTELSAT Transponders for Domestic Services," *IEEE EASCON '78 REC.*, pp. 310–315 [78CLR55].
PARTHASARATHY, R.† & Lee, W., "Utilization of the INTELSAT Network," *IEEE ICC June 78 CONF. REC.*, Vol. 1, pp. 10.1.1–10.1.3 [78CLR34].
PARTHASARATHY, R.†, see Van Trees, H. L. [78CLR05].
PENTLICKI, C. J., "An Overview of Communications Satellites in the STS Era," *IEEE EASCON '78 REC.*, pp. 348–353 [78CLR57].
POLLACK, L., see Davis, R. C. [78CLR22].
POTTS, J. B. & Burkitt, F. J., "Operational Planning for the Utilization of INTELSAT V Satellites," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-529 [78CLR04].
POTUKUCHI, J., see Childs, W. H. [78CLR44].
RAMOS, A., see Hsing, J. C. [78CLR11].
REBER, C. J., see Early, L. B. [78CLR21].
REVESZ, A. G., "The So-Called Amorphous Silicon—an SiHx Polymer Film," *Thin Solid Films*, May 78, pp. L29–L33 [78CLR42].
ROGERS, D. V., see Olsen, R. L.* [78CLR47].
SINGH, K. J. & Forcina, G., "Satellite Remote Numbering Systems: General Requirements and a Proposed New Approach," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 16–22 [78CLR68].

*Non-COMSAT author.
†INTELSAT assignee.

SLACK, E. R., see Lipke, D. W. [78CLR03].
SLACK, E. R., "Financial Aspects of Lease versus Purchase for Satellites," *IEEE ICC June 78 CONF. REC.*, Vol. 2, pp. 18.1.1–18.1.5 [78CLR35].
SLACK, E. R., "Lease vs. Purchase for Satellites—Risk Factors," *IEEE EASCON '78 REC.*, pp. 306–309 [78CLR56].
SYNDER, J., see Muratani, T.* [78CLR72].
STEGENS, R. E., see Estabrook, P.* [78CLR45].
STOCKEL, J. F., see Betz, F. E. [78CLR76].
STOCKEL, J. F., Dunlop, J. D. & Betz, F.*, "NTS-2 Nickel Hydrogen Battery Performance," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-536 [78CLR07].
STUART, K. L., "Reducing Switching Stress in High Power High Voltage DC-DC Converters," National Solid-State Power Conv. Conf., May 78, *PROC.*, pp. B1-1–B1-6; *Solid State Power Conv.*, July/Aug. 78, pp. 42–45 [78CLR28].
SUYDERHOUD, H. G., Onufry, M. & Sennott, J.*, "Subjective Performance Assessment of Delta Modulation in Terms of Equivalent Noise at 16, 24, 32, and 40 kbit/s," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 243–247 [78CLR71].
SWEARINGEN, D. W., "Multiple Shore Station Interworking in MARISAT," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-550 [78CLR09].
TSENG, F. & Hyde, G., "Frequency Scaling of Attenuation Above 10 GHz," *IEEE EASCON '78 REC*, pp. 396–403 [78CLR54].
TSUJI, L., see Dill, G. D. [78CLR66].
VAN TREES, H. L., see Hoversten, E. V. [78CLR70].
VAN TREES, H. L., Casey, J. P., Chidambaram, T. S., Parthasarathy, R.† & Chasia, H.*, "Development of Traffic Scenarios for the Future INTELSAT System," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-530 [78CLR05].
VAN TREES, H. L., Fuenzalida, J. C., Mohajeri, M. & Martin, J. E., "Planning for the Post-1985 INTELSAT System," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-532 [78CLR06].
WACHS, M. R. & Arroyo, B.†, "Technical Consideration of Frequency Sharing Between Satellite Communications and Radar," *IEEE ICC June 78 CONF. REC.*, Vol. 3, pp. 35.5.1–35.5.5 [78CLR36].
WATANABE, T.*, Saitoh, H.*, Ogawa, A.* & Deal, J. H., "Space Diversity System for TDMA Satellite Links," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 319–326 [78CLR65].
WEINREICH, D. E., "Simulation of MARISAT Offshore Data Transfer," Pres. at AIAA 7th Comm. Sat. Syst. Conf., Apr. 78, AIAA Paper No. 78-555 [78CLR10].
WELTI, G. R., "Intersatellite Link for Multiple-Access Telephony," *IEEE EASCON '78 REC.*, pp. 432–440 [78CLR58].

*Non-COMSAT author.
†INTELSAT assignee.

WELTI, G. R., "TDM/PCM/FDMA Telephony Using 2P-APM," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 51–56 [78CLR73].
WOLEJSZA, C. J., see Chakraborty, D. [78CLR62].
WU, W. W., "On the Efficacy of Traffic Assignment in Satellite-Switched TDMA Systems," 4th Int. Conf. on Digital Sat. Comm., Oct. 78, *PROC.*, pp. 180–185 [78CLR74].
WU, W. W., "The Power of Desarguesian Sets," *ITC Nov. 78 PROC.*, Vol. 14, pp. 465–471 [78CLR81].
ZAGHLOUL, A. I., see Andersen, J. B. [78CLR27].
ZAGHLOUL, A. I. & MacPhie, R. H.*, "Multi-Mode Analysis of Mutual Coupling Between Rectangular Apertures," Pres. at Int. Conf. on Antennas and Propagation, London, Nov. 78, *PROC.*, Pt. 1: *Antennas*, pp. 438–442 [78CLR84].
ZAKI, K. A.* & Atia, A. E., "Sensitivity Analysis of Multi-Coupled Cavity Filters," 1978 IEEE Int. Symp. on Circuits & Systems, May 78, *PROC.*, pp. 790–793 [78CLR30].

*Non-COMSAT author.
†INTELSAT assignee.

Corrigendum

In the paper "Tabulations of raindrop induced forward and backward scattering amplitudes" by D. J. Fang and F. J. Lee, *COMSAT Technical Review*, Vol. 8, No. 2, Fall, 1978, the following revisions should be made:

1. In Table 1 (page 458), the real part of the backward scattering amplitude for horizontal polarization at 1.25 mm should be $1.3299e-5$.

2. In Table 6 (page 459), the imaginary part of the forward scattering amplitude for vertical polarization at 3.25 mm should be $-4.8458e-5$.

3. In Appendix C (page 484), equations (C-2), (C-3) and (C-4) should be:

$$A^{h,v} = 8.686 \times \mathrm{Im}(K^{h,v}) \times 10^3 \qquad \text{(C-2)}$$

$$\phi^{h,v} = \frac{180}{\pi} \times \mathrm{Re}(K^{h,v}) \times 10^3 \qquad \text{(C-3)}$$

$$N(R,a) = 16 \times 10^4 \exp\left\{-3.67 \frac{2a}{0.089R^{2.1}}\right\} . \qquad \text{(C-4)}$$