Library

DIGITAL TRANSMISSION OVER

LAND MOBILE RADIO CHANNELS

by

Dr. Robert W. Donaldson

DEPARTMENT OF ELECTRICAL ENGINEERING
FACULTY OF APPLIED SCIENCE
THE UNIVERSITY OF BRITISH COLUMBIA

February 1979

FINAL REPORT

FREQUENCY ASSIGNMENT FOR LAND MOBILE RADIO SYSTEMS IN THE 900 MHZ BAND: DIGITAL TRANSMISSION OVER LAND MOBILE CHANNELS

by

Dr. Robert W. Donaldson, Professor
Department of Electrical Engineering
Faculty of Applied Science
The University of British Columbia

for

Department of Communications
Ottawa, Ontario, CANADA

under

Department of Supply and Services Contract No: OSU78-00221
Period: September 1, 1978 - March 1, 1978

ABSTRACT

Increasing demand for digital transmission over Land Mobile radio channels requires technical data in support of guidelines relating to transmitted spectra, data rates, bit error rates and spectral utilization efficiency.

Various digital modulation schemes including PAM, SSB, PSK, DPSK and FSK are assessed. Maximum bit rates permissible under existing FM voice channel emission constraints as well as under new constraints proposed by Communications Canada are determined. For example, the new constraints permit 36 Kb/s for binary FSK, and 20 Kb/s for binary DPSK signals with 100% excess bandwidth. Corresponding rates for existing channels are 13.3 Kb/s for binary FSK and 8.1 Kb/s for binary DPSK. Bit error rates (BER) for the various modulation formats are determined for both Gaussian and Rayleigh channels which indicate BER extrema for Land Mobile channels.

Analysis of block codes for forward error correction (FEC) indicates the availability of large coding gains (40 dB for (31,16) BCH code at 10^{-7} decoded BER) on independent-error Rayleigh channels, which channels require large average power in comparison with Gaussian channels. Analysis of codes on dependent-error channels, which situation occurs when vehicles move at moderate speeds, is hampered by lack of a suitable channel model. However results based on measured error distributions show coding gains of approximately 10 dB.

Error detection and retransmission (ARQ) shows very large coding gains for independent-error Rayleigh channels (60 dB for (31,16) BCH code at 10^{-7} decoded BER). Results based on measured distributions indicate large coding gains; for example, in excess of 33 dB for (31,26) BCH code at decoded BER $\approx 10^{-5}$. For very noisy channels, repeated transmission of data blocks and bit averaging prior to FEC or ARQ is advocated.

Various ARQ protocols are considered and information throughput rates are determined. A single data block structure is proposed for transmission of control information, text, and digitized voice each of which has different speed-vs-accuracy requirements.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I INTRODUCTION

I-1 Scope and Objectives of the Study

This report presents the results of a Communications Canada Contract study whose motivation, objectives and requirements are summarized below:

MOTIVATION:

With the recent advancements in Land Mobile system development, it has become more evident that future data transmission demand over land mobile channels, for dispatch operations will represent an important part of the communication traffic loading over the allocated land mobile bands.

The need for the development of guidelines identifying transmission characteristics, requirements, and limitations of digital transmission is becoming pressingly great. These guidelines will be utilized in the development of new regulations and standards governing the technical characteristics and operation of equipment and systems employing digital means of communications transmitted over the land mobile channels.

The guidelines will cover items such as: modulation techniques, transmission bit rate, bandwidth requirements, bit rate error requirements, multipath effect reflected as required additional link margin, C/I, C/N, etc.

OBJECTIVES:

To produce technical data in support of the development of a guideline identifying transmission characteristics, requirements and limitations of digital transmission including data and coded voice over existing and proposed land mobile bands.

REQUIREMENTS:

1) Evaluation of different digital modulation techniques under the land mobile channel environment.

2) Maximum information bit rate that can be accommodated over the existing land mobile channels.

3) Forward error correction techniques that are suitable for land mobile channels and their benefits (dB's) in alleviating multipath fading.

4) Trade off between required bandwidth, spectrum utilization efficiency and equipment complexity.

5) Maximum bit rate that can be transmitted over existing FM voice channel.

At the outset it was recognized that an important task is to select the salient aspects of the requirements to produce a useful set of technical data which would facilitate formulation of the guidelines referred to under "objectives".

I-2 Executive Summary

The following detailed summary is presented here for convenience, and replaces summaries which might otherwise appear at the conclusion of each chapter.

Chapter 2 briefly characterizes land mobile communication channels, which show fluctuations in received signal level in excess of 90 dB. Various alternatives are available to combat these fluctuations including space diversity, time diversity via forward error correction (FEC) coding, and variable rate transmission via error detection and retransmission (ARQ). The need for spectral emission constraints is established for spectrum sharing based on conventional channelization.

Chapter 2 also describes the functional parts of digital communication systems as well as the restrictions implied by digital signalling. Relationships between the rate of signal level variations, vehicle speeds and data block lengths are established.

Chapter 2 identifies three types of message sources requiring transmission over land mobile channels, including voice, text, and control information. The data rate and accuracy requirements are seen to differ considerably; fortunately voice which requires a high data rate has relatively low accuracy requirements, and conversely for control information.

Chapter 3 determines transmission spectra and error rates obtainable using pulse amplitude modulation (PAM). To obviate the (adaptive) receiver threshold adjustments needed for multilevel PAM under conditions of signal level variation, signalling is restricted to binary antipodal. Modulator pulse shapes are further restricted to those which produce zero intersymbol interference (isi) following receiver matched filtering, again to obviate the need for signal-level-dependent receiver adjustments. These restrictions limit consideration to conventional rectangular baseband pulses and to pulses with raised-cosine spectra which provide a tradeoff between relative excess bandwidth α and bit rate. With maximum excess bandwidth and double sideband modulation the spectral emission constraints proposed by Communications Canada permit data rates of up to 20 Kb/s. Use of single sideband with zero excess bandwidth permits rates to 62 Kb/s. Practical considerations including phase and timing error problems would undoubtedly require some excess bandwidth which would keep the actual SSB rate well below 62 Kb/s. For conventional rectangular pulse shapes, the proposed spectral constraints limit rate to 4 Kb/s for DSB.

Existing FM voice channel constraints require at least 99 percent of the transmitted energy within ±8 KHz of the carrier frequency. This constraint limits the data rates to 32.3 Kb/s for SSB with no excess bandwidth, 16.2 Kb/s for DSB with no excess bandwidth, 8.1 Kb/s for signals having raised cosine spectra with maximum excess bandwidth, and 1.8 Kb/s for conventional rectangular pulses.

Chapter 4 considers digital PSK and FSK. Conventional PSK signals have a constant amplitude from which it immediately follows that binary PSK is identical to conventional binary PAM. However PSK need not be restricted to two levels since detection is based on the ratio of outputs from quadrature filters, and this ratio is independent of signal level. Noise rather than bandwidth limits the PSK data rate since the transmitted signal spectra is independent of the number of PSK levels.

For bandwidth-efficient PSK signalling, non-constant amplitude pulse shapes may be used; the resulting spectra are identical to those of PAM with the same pulse shapes. Use of pulses with raised-cosine spectra permit bit rates of $(\log_2 L)\ R_{PAM}$ where L is the number of PSK levels and R_{PAM} is the corresponding PAM data rate; R_{PAM}=31 Kb/s, 20Kb/s and 4 Kb/s for minimum excess bandwidth pulses, maximum excess bandwidth pulses and conventional rectangular pulses, respectively.

To obviate the need for carrier phase estimation differential PSK (DPSK) is available, with a maximum power penalty of 3 dB relative to PSK. Normally the spectra of DPSK is identical to that of PSK. For environments where bandwidth limits data rates DPSK with raised cosine signalling having a maximum excess bandwidth appears attractive, since phase estimation is unnecessary and timing error tolerance is relatively high.

FSK signal spectra are not simply related to the baseband modulating signal spectrum, and may be wider than or narrower than that of the baseband signal. The spectrum depends on the baseband signal pulse shape and the peak-to-peak frequency deviation relative to the bit rate. For a deviation ratio of 0.5 minimum shift keying (MSK) results, and exact determination of the transmitted spectra is possible. We show that for binary MSK the data rate permitted by proposed Communications Canada spectral constraints is 36 Kb/s. Existing FM voice channel constraints permit data rates of 13.3 Kb/s for binary MSK.

The relatively high data rates permitted by the proposed emission constraints occur because the transmitted spectra of appropriately designed digital signals neatly fit the defined channel whose bandwidth from zero to -26 dB below the unmodulated carrier power is ± 10 KHz; the bandwidth falls linearly (in dB) to ± 20 KHz at -60 dB. However a high bit rate R implies a low energy per bit and thus a potentially high bit error rate, in which case noise rather than bandwidth may limit R. A reduction of R and, for FSK a corresponding reduction in peak deviation does increase the signal-to-noise ratio (SNR) in proportion to 1/R. However some bandwidth is thereby unused, and more effective alternatives exist which use all of the bandwidth. These alternatives include channel encoding, or a reduced bit rate together with bandwidth expanding modulation of an FM carrier which increases SNR in proportion to R^{-3}.

Also presented for PAM, PSK, DPSK and FSK are expression for bit error probabilities p for both Gaussian and Rayleigh channels. For Gaussian channels, p decreases exponentially with SNR; for Rayleigh channels the decrease is linear in SNR. The only SNR degredation assumed is additive noise.

Chapter 5 considers forward error correction (FEC) to combat bit errors on land mobile channels. Although much of the discussion applies to both block and conventional codes, quantitative results are presented for block codes only. It is argued that memoryless Gaussian noise channel models apply to stationary vehicles, and that memoryless Rayleigh fading channel models approximate the behaviour of transmissions when the vehicle moves rapidly and traverses several wavelengths during transmission of a single data block.

Our results for block codes show large coding gains for Rayleigh channels. For example, for a decoded bit error probability $p_b=10^{-7}$ a

(31,26) single error correcting Hamming code shows a coding gain of 26 dB; for this same value of p_b a (31,16) code shows a gain of 40 dB. Coding gains for these same codes on Gaussian channels are 1.7 dB and 2 dB, respectively for $p_b=10^{-7}$. Because the transmitted power required for Rayleigh channels is ten's of dB's larger than for Gaussian channels, the large potential coding gains for Rayleigh channels are of interest.

When vehicles travel at moderate speeds and transmit short blocks of data at high rates, the resulting digital transmission channels exhibit memory. Unfortunately no digital channel model is available which suitably approximates the channel behaviour under these conditions, and use of inadequate models, not withstanding their intuitive appeal, leads to grossly inaccurate results and wrong conclusions. Some recent results based on measured error distributions for vehicles travelling at moderate speeds are available. These results together with our own suggest that at moderate vehicle speeds coding gains available may approach or even exceed 10 dB for decoded bit error probabilities below 10^{-5}.

Chapter 6 deals with the use of block codes for error detection and block retransmission (ARQ). For any given code ARQ systems show much lower decoded bit error probabilities than FEC systems. The cost of this improved performance is a reduced throughput resulting from retransmissions when the channel degrades. Coding gains available for ARQ relative to FEC are large for Rayleigh channels and moderate for Gaussian channels. For (31,26) and (31,16) codes with $p_b=10^{-7}$, gains are 17 dB and 18 dB for Rayleigh channels and 2.8 dB and 5.3 dB for Gaussian channels, and these coding gains are in addition to those available for FEC relative to uncoded transmission. Results based on measured error distributions for vehicles moving at moderate rates show a coding gain of 33 dB coupled with a 3:1 gain in bit rate for ARQ relative FEC for a (31,16) code. Thus, large ARQ coding gains do seem to be available.

Chapter 6 also considers throughput reductions and information transmission delays for various ARQ protocols including send-and-wait, continuous (go-back-N) and selective repeat. For channels which are "good" most of the time throughput reductions are small.

Also considered in Chapter 6 are data block formats suitable for mobile channels, as well as the way in which the bolck parameters, including block size (length) depend on the channel error rate, the ARQ protocol and the average message length. A block structure is proposed which includes some bits for mode control, to allow a single block format to be used to transmit the three types of messages identified in Chapter 2. Under the proposed scheme, control messages would be transmitted using stop-and-wait ARQ, voice would be transmitted with no error protection and text would be transmitted using FEC.

The Appendix deals with error controls on very noisy channels, where channel bit error probabilities are so low that FEC yields virtually no performance improvement and ARQ results in a severely reduced throughput efficiency because of multiple retransmissions. The simplest alternative (and for short isolated control blocks the only apparent alternative) is to retransmit each channel bit many times and average the bit decisions using majority voting. The averaged channel bits would then be decoded using FEC or ARQ.

II DIGITAL COMMUNICATION OVER LAND MOBILE CHANNELS

II-1 Land Mobile Radio Channels

Bandwidth and noise limit the rate at which information can be transmitted over any communication channel. In land mobile radio channels, legislation limits the bandwidth occupied by any user. Noise originates from various sources [J1, A1], including (Gaussian) receiver front-end amplifier noise, (impulsive) automobile ignition noise, co-channel interference from other users of the frequency channel, adjacent channel interference from users on other frequency channels, and random FM. Random FM occurs during vehicle motion and is caused by random phase changes associated with received signal level fluctuations described below. Because the relative geographic positions of both co-channel and adjacent-channel users change over time the associated interference levels also vary.

The level of a signal received by a mobile varies as the mobile moves. This level variation includes a fading component, illustrated in Fig. 1.1, resulting from multipath interference changes. Adjacent peaks in the received signal envelope are separated by one-half a carrier wavelength. The amplitude probability density $f_r(\alpha)$ of the received signal level r is reasonably well approximated by the Rayleigh distribution as follows: [W1, F1]:

$$f_r(\alpha) = \begin{cases} 2\alpha/r_0 \exp(-\alpha^2/r_0) & \alpha > 0 \\ 0 & \alpha \leq 0 \end{cases} \qquad (2.1)$$

In 2.1, $s = \sqrt{\pi r_0}/2$ is the local mean level of the received signal.

Also of interest is the distribution of r^2 the received signal energy [W1, F1]:

$$f_{r^2}(\alpha) = \begin{cases} (1/r_0) \exp(-\alpha/r_0) & \alpha \geq 0 \\ 0 & \alpha < 0 \end{cases} \qquad (2.2)$$

where r_0 is the average energy in the received signal.

The local mean signal level s varies slowly over several wavelengths as a result of changes in path transmission characteristics. These changes

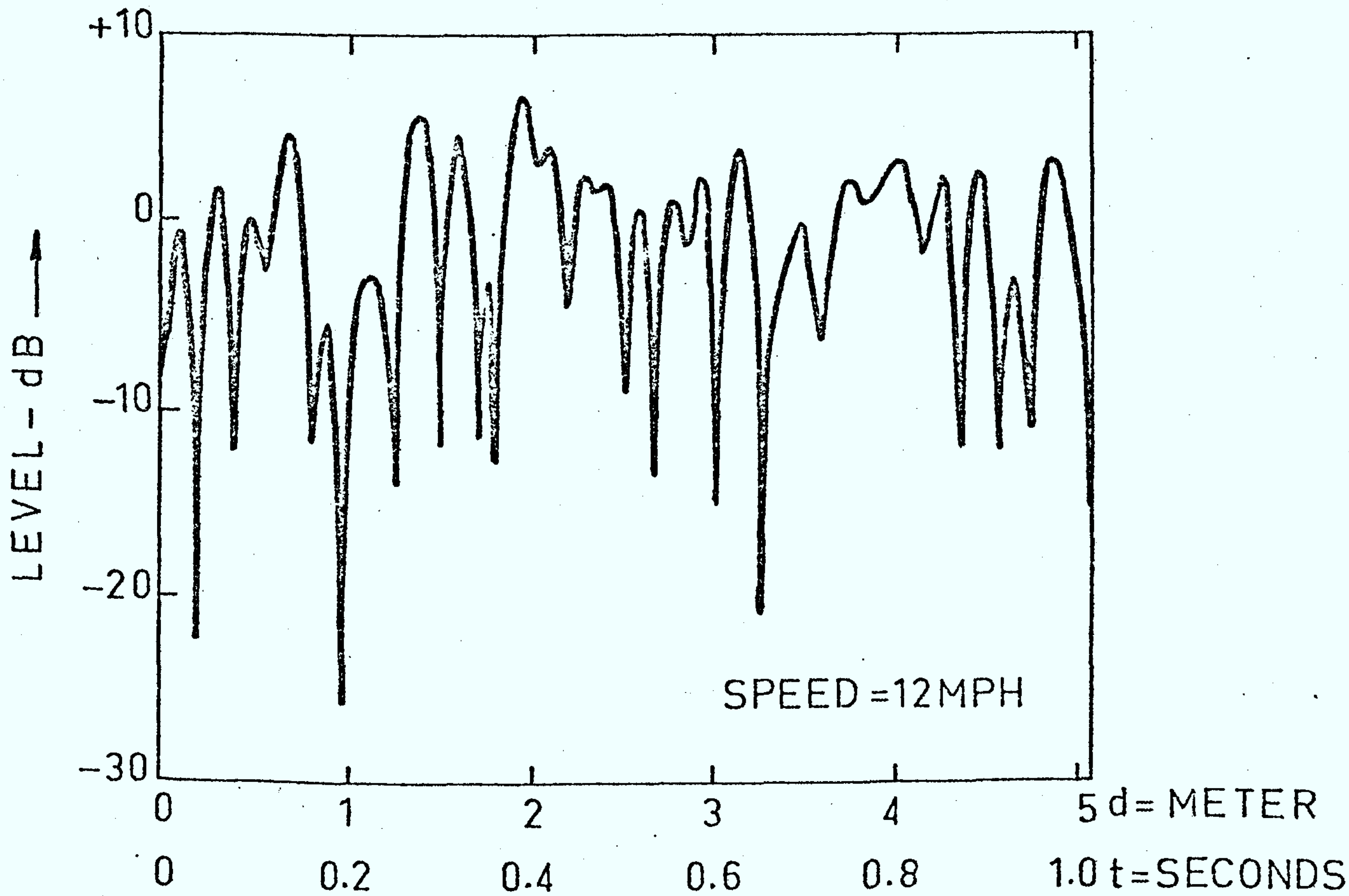

Fig. 2-1 Received signal level vs. time in an urban environment. Vehicle speed: 12 mph. carrier frequency: 850 MHz (after Arredondo and Smith [A1]).

result from variations in shadowing caused by topographical features such as street width, building height and hills. The amplitude probability density of s in dB is reasonably well approximated by the log-normal distribution, as follows [J1, F1]:

$$f_s(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(\alpha-m)^2/2\sigma^2) \tag{2.3}$$

In (2.3) m is the mean signal level in dB averaged over several wavelengths (typically 50 m), and σ is the standard deviation in dB of the local mean; σ has been quoted as 6 dB in London [F1] and as 8-12 dB in Japan and the USA [F1, J1].

Finally, the local mean m varies over the mobile coverage area as the distance D between transmitter and receiver changes. This variation is proportional to D^{-n} where $2<n<4$ [J1, O1] for $D \lesssim 40$ km (25 miles). The rate of attenuation of n tends to increase with D. A recent study shows that $n \simeq 3.7$ in urban Philadelphia [O1].

II-2 Information Sources for Land Mobile Channels.

Three types of information sources can be identified as requiring transmission over land mobile radio channels, as follows:

1. Voice information (speech) is real time information with considerable natural redundancy. In the absence of a prohibitively larger storage buffer, speech requires transmission as it is generated; i.e. real-time transmission. The actual source information rate is less than 100 bits/sec; however any viable encoding technique, such as adaptive differential PCM (ADPCM) requires a transmission rate in excess of approximately 10 Kb/s [F2, H1, C1]. Because of its natural redundancy random bit error rates as low as 10^{-2}

do not prevent intelligible transmission of speech [F2, H1, C1]. Documentation regarding burst error effects on digitally transmitted speech is scarce; however some results are available [Y1, J1].

2. <u>Text information</u> like voice contains redundancy. Shannon [S1] estimated that English text is approximately 50% redundant in that random deletion of up to half of the message does not prevent reconstruction (possibly with difficulty) of the remainder. Text symbols are normally generated at a rate sufficiently low to require neither immediate nor real-time transmission. Transmission errors are annoying in that they may cause "spelling" errors upon reconstruction of the received text. However these errors are normally correctable by the reader provided they don't occur too frequently. Video facsimile would fall into the text information category because of its redundant and non-real-time nature, although buffer storage limitations may require a scanning rate commensurate with average data transmission rate.

3. <u>Control Information</u> includes vehicle status reports, vehicle dispatch orders and street addresses. Because control information contains minimal redundancy, accuracy of its transmission is essential. Although real-time transmission of control data is seldom necessary, prompt transmission is normally required.

One could argue for information-source categories of additional to those listed above. However, these are adequate for our purposes, and establish that different information sources imply different transmission and error rate requirements. Fortunately, information sources requiring accurate transmission do not normally require high data transmission rates; conversely information requiring high speed transmission does not normally carry stringent accuracy requirements.

Communication in a mobile environment often involves transmission of two or more types of information, each with different speed-vs-accuracy priorities. It may therefore be advantageous to provide for two or more digital transmission modes. The implications of this strategy are considered in Chapter 6.

II-3 Digital Communication Systems.

Fig. 2-2 shows in block diagram a digital communication system, whose component subsystems perform the following tasks [G1, L1]:

1. Source encoder: converts the message source which may be either analog or digital into a binary digit stream which ideally has no redundancy.

2. Channel encoder: maps source digit sequences into channel digit sequences which contain controlled amounts of redundancy to combat channel transmission errors, either by means of forward error correction, or error detection and retransmission.

3. Modulator: converts binary digit sequences from the channel encoder into signals suitable for transmission

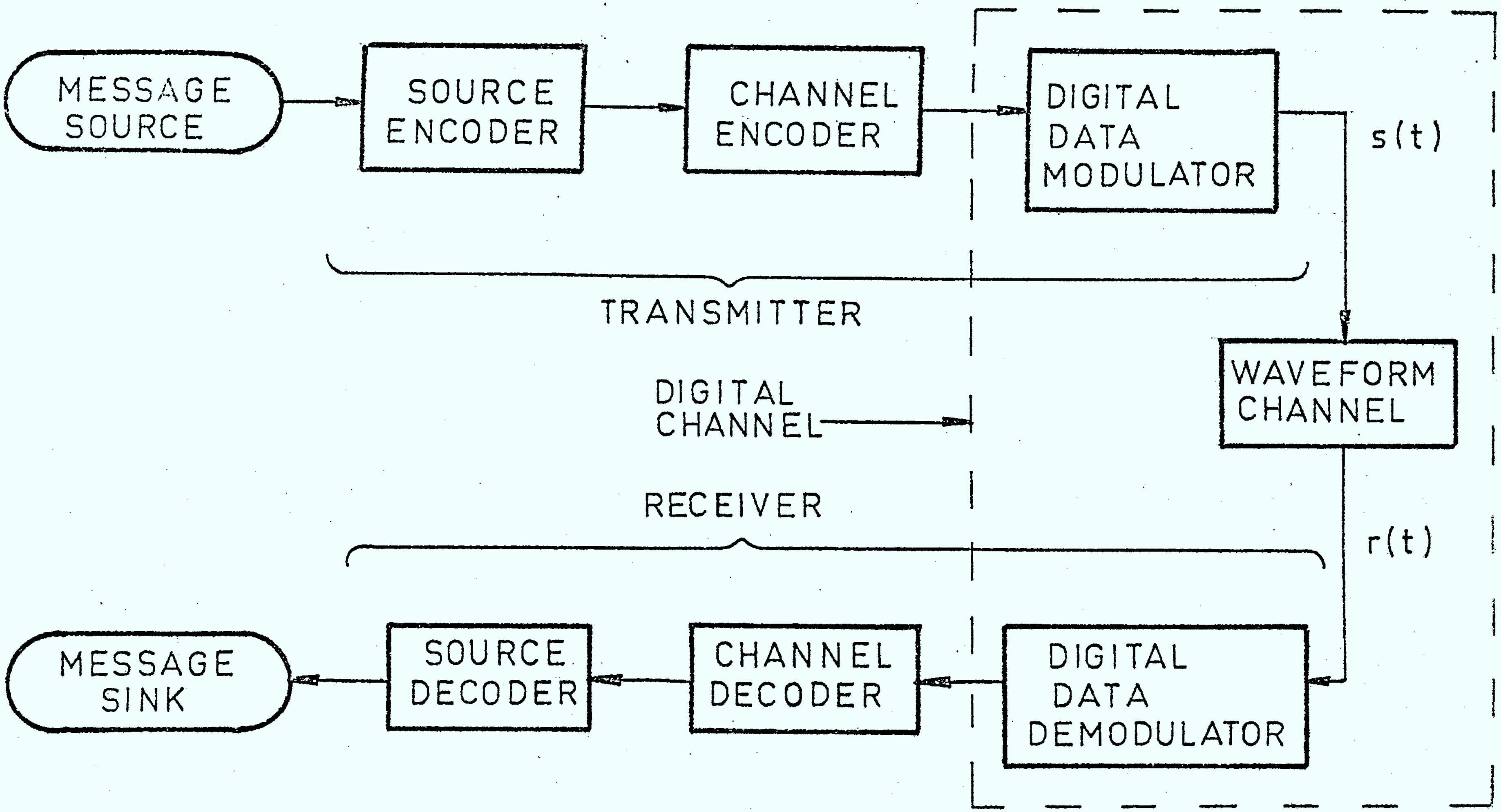

Fig. 2-2 Block diagram of a digital communication system.

over the physical waveform channel.

4. Demodulator: extracts from the received signal r(t) a replica of the channel-encoded digits.

5. Channel Decoder: maps the demodulated binary digits into source digits.

6. Source Decoder: uses the channel decoder output bits to reconstruct a replica of the original message.

The term digital communication implies a fundamental constraint on the transmitter and receiver in Fig. 2-2, namely that the transmitted signal consists of a superposition of time-translates of a few basic signals [L1]. Thus, each successive sequence of $N=\log_2 M$ binary digits from the channel encoder is mapped into a modulator pulse, one or more of whose characteristics such as amplitude, phase, frequency, position or duration depend on the binary sequence. The result is M'ary pulse amplitude modulation (PAM or digital AM), phase modulation (PM), frequency modulation (FM), pulse position modulation (PDM), or pulse duration modulation (PDM). In many cases of practical interest, M=2 and binary signalling results. In all cases of practical interest M is small, typically $M \lesssim 32$.

The restriction to digital signalling is for ease of system implementation. Even with this constraint, the bit-error probability can be much arbitrarly small by proper channel encoder/decoder design provided that the rate at which source digits is transmitted does not exceed the capacity in bits/sec of the digital channel [L1, G1, W1].

In designing the modulator and demodulator, knowledge of the waveform channel behaviour is essential. The modulator is normally constrained in in either peak or average transmitter power, and, in the case of radio channels in spectral emission characteristics. Simplicity of implementation is a primary consideration in implementation of the demodulator which is normally designed to minimize the probability of a demodulated bit error. Synchronization at both the carrier signal level and binary symbol level is an additional necessary and often difficult task of the demodulator [L2, S2]. Symbol synchronization is often an added consideration in modulator design Effective channel encoders can be implemented using linear feedback shift registers. Decoding is more difficult, and except for selected (but very useful) codes requires decoding algorithms whose implementation is often prohibitively expensive. Design and performance analysis of channel encoders and decoders requires knowledge of the bit-error statistics of the digital channel in Fig. 2-2, including bit-error dependencies.

Improved source encoding is needed for highly redundant sources such as speech and video data if significant reductions in transmitted bit rates (and therefore in required channel bandwidth) are to be realized. We do not consider in this report any of the multitude of source coding schemes which, because of the continually improving micro-circuitry are becoming ever easier to implement.

II-4 Utilizing Land Mobile Radio Channels for Digital Communication

Land mobile radio channels present problems additional to those associated with conventional point-to-point links. The signal level variations resulting from fading, local mean variations, and variations in transmitter-receiver separation D can exceed 20 dB, 10 dB and 60 dB, respectively, which together can cause a total received signal level vraiation in excess of 90 dB. To reduce adjacent channel interference, constraints are imposed on the power spectral density of the transmitted signal.

One way to combat the large signal-to-noise ratio fluctuations which accompany signal variations is to use one or more types of diversity. For example, spatial diversity in which several base station antennas are located at the edge of the radio coverage area can reduce fluctuations in base-station received signal level resulting from all these of the above-noted causes [J1]. By switching these antennas during transmission from the base station, variations in mobile receiver signal level due to shadowing and D^{-n} loss can also be reduced considerably [J1]. Space diversity systems utilize the fact that the probability of two or more transmission paths being poor at the same time is considerably less than the corresponding probability for any individual path. Forward error correcting codes involve use of time diversity.

An alternative to diversity involves transmitting information at a rate which depends either directly or indirectly on the signal-to-noise ratio (SNR) at the receiver. Transmission rates are high when SNR is high, and conversely. Variable-rate transmission normally inplies a two-way channel between transmitter and receiver. The receiver may estimate its signal-to-noise ratio and transmit this estimate to the transmitter, whose information transmission rate is adjusted accordingly [C2, C3]. Alternatively, the decoder in Fig. 2-2 may detect, rather than correct errors in transmitted bit sequences, and request retransmission of those sequences received in error. Retransmission probability is highest and information throughput is lowest during when SNR is low.

To determine the performance of various modulation and coding schemes both the mobile waveform channel and digital channel in Fig. 2-2 require characterization. Of particular interest is the rate of signal level change relative to vehicle speed. The salient facts are summarized below.

The carrier signal wavelength $\lambda=c/f$ where c is the velocity of light (3×10^8 m/sec) and f the carrier frequency. At 900 MHz, $\lambda\approx$ 1ft. Thus, fading minima in Fig. 2.1 correspond to vehicle movements of 1/2 ft. (1/6 m). At 15 mph (22'/sec) a mobile moves 44 half-wavelengths in one sec. For data rates of 10 Kb/s, 227 bits are transmitted as a vehicle moving at 15 mph moves one-half wavelength. For data rates of 1 Kb/s, 11 bits are transmitted as a vehicle moving at 30 mph moves one-half wavelength. Thus, at 10 Kb/s and 15 mph a 100-bit data block is transmitted as the vehicle moves one-quarter wavelength. At 1 Kb/s and 30 mph a 100-bit block is transmitted as the vehicle moves 5 wavelengths. It follows that the received signal level is relatively constant over one bit period, but not during the transmission of one data block. The mean signal level, however, being relatively constant over 50 m, is relatively constant over one transmitted data block.

Vehicle motion also broadens the power spectral density of the radiated signal on the order of 50 Hz at 20 mph. Because radiated spectra normally have bandwidths in excess of 10 KHz, this spectrum broading can be neglected below 1 GHz, for spectral occupancy calculations.

Conventional spectrum management policies involve division of that portion of the radio frequency spectrum assigned to the land mobile radio service into frequency channels. To prevent a user of any channel from interfering unduly with users in other channels, limitations on the spectrum of the transmitted signal s(t) (see Fig. 2-2) are imposed, along with limitations on the total transmitted power. This latter limitation is needed to avoid excessive interference with others using the same frequency channel in other geographic locations.

Fig. 2-3 shows the spectral limitations proposed in by Communications Canada [G2]. In Chapters 2 and 3 we demonstrate the determination of maximum transmission rate and the selection of modulator pulse shapes to enable the transmitted signal spectrum to tightly fit inside the proposed emission characteristic.

The division of the land mobile spectrum into frequency channels is not the only viable spectrum management option. Considerable interest exists in alternative schemes whereby each user shares the entire frequency band using spread spectrum multiple access (SSMA) [D1, C4, S3]. Definitive comparisions of the capabilities of SSMA with the more conventional channel assignment schemes are now becoming available [C4].

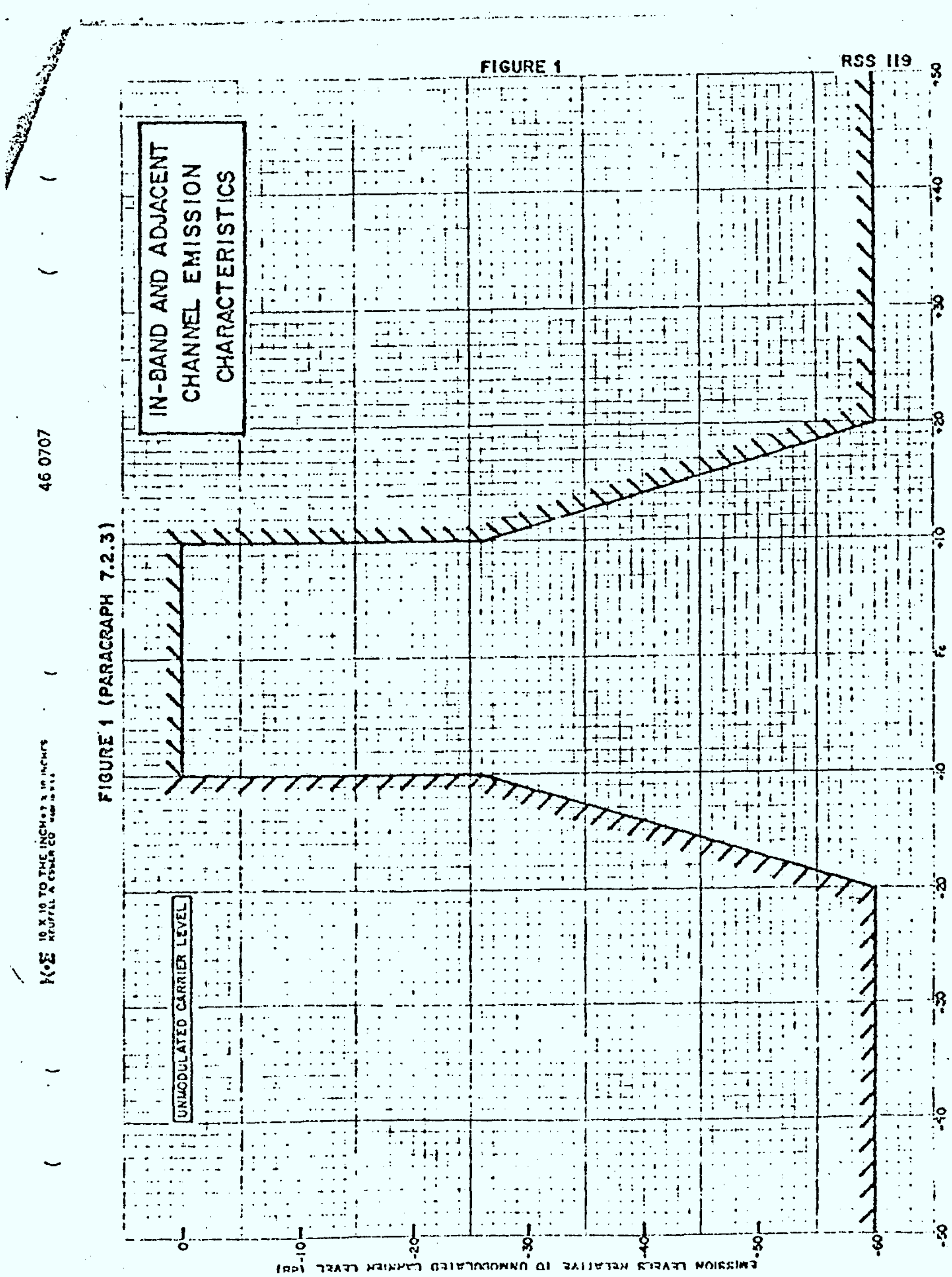

Fig. 2-3 In-band and adjacent channel emission constraints proposed by Communications Canada.

II-5 References for Chapter 2

[A1] G.A. Arrendondo and J.I. Smith, "Voice and data transmission in a mobile radio channel at 850 MHz", IEEE Trans. Veh. Technol., vol. VT-26, pp. 94-105, Feb. 1977.

[C1] D.L. Cohn and J.L. Melsa,"The residual encoder - an improved ADPCM system for speech digitization", IEEE Trans. Commun., vol. COM-23, pp. 935-941, Sept. 1975.

[C2] J.K. Cavers, "Variable rate transmission for Rayleigh fading channels", IEEE Trans. Commun., vol. COM-20, pp. 15-22, Feb. 1972.

[C3] J.K. Cavers and S.K. Lee, "A simple buffer control for variable-rate communication systems", IEEE Trans. Commun., vol. COM-24, pp. 1045-1048, Sept. 1976.

[C4] G.R. Cooper and R.W. Nettleton, "A spread-spectrum technique for high-capacity mobile communications", IEEE Trans. Veh. Technol., vol. VT-27, pp. 264-275, Nov. 1978.

[D1] R.C. Dixon, Spread Spectrum Systems. New York, N.Y.: Wiley, 1976.

[F1] R.C. French, "Error rate predictions and measurements in the mobile radio data channel", IEEE Trans. Veh. Technol., vol. VT-27, pp. 110-117, Aug. 1978.

[F2] J.L. Flanagan, Speech Analysis, Synthesis and Perception, 2nd Ed. New York, N.Y.: Springer-Verlag, 1972.

[G1] R.G. Gallayer, Information Theory and Reliable Communication. New York, N.Y.: Wiley, 1968.

[G2] Government of Canada, Dept. of Communications: Radio Standards Specification RSS-119, Issue 1, Draft 2; Proposed Effective Date: 15 Dec. 1980; Release Date: 22 July 1978.

[H1] B.A. Hanson and R.W. Donaldson, "Subjective evaluation of an adaptive differential voice encoder with oversampling and entropy coding", IEEE Trans. Commun., vol. COM-26, pp. 201-208, Feb. 1978.

[J1] W.C. Jakes, Microwave Mobile Communications. New York, N.Y.: Wiley, 1974.

[J2] N.S. Jayant, "Step-size-transmitting differential coders for mobile telephony", in Conf. Rec. Int. Conf. Commun., San Francisco, Ca., June 1975, pp. 30:6-30:10.

[L1] R.W. Lucky, J. Salz and E.J. Weldon, Jr., Principles of Communication Engineering. New York, N.Y.: McGraw-Hill, 1968.

[L2] W.C. Lindsay and M.K. Simon, Telecommunication Systems Engineering. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

[O1] G.D. Ott, "Urban path-loss characteristics at 820 MHz", IEEE Trans. Veh. Technol., vol. VT-27, pp. 189-198, Nov. 1978.

[S1] C.E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IS: University of Illinois Press, 1962; also in Bell Syst. Tech. J., July 1949 and Oct. 1949 and Scientific American, July 1949.

[S2] J.J. Stiffler, Theory of Synchronous Communications. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[S3] R.A. Scholtz, "The spread spectrum concept", IEEE Trans. Commun., vol. COM-25, pp. 748-756, Aug. 1977.

[W1] J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York, N.Y.: 1965.

[Y1] J. Yan and R.W. Donaldson, "Subjective effects of channel transmission errors on PCM and DPCM voice communication systems", IEEE Trans. Commun., vol. COM-20, pp. 281-290, June 1972.

III LINEAR MODULATION FOR DIGITAL TRANSMISSION OVER LAND MOBILE CHANNELS

III-1 Description of Pulse Amplitude Modulation (PAM)

In pulse amplitude modulation the digital data modulator in Fig. 2-2 generates a signal s(t) whose constitutive pulse amplitudes depend on the binary digit sequences impressed upon the modulator [L1]. Thus

$$s(t) = [\sum_k a_k \, g(t-kT)] \cos (\omega_c t + \phi)] \tag{3-1}$$

where a_k is the amplitude corresponding to the k^{th} binary input sequence, g(t) is the basic modulator pulse shape, ω_c is the carrier signal frequency, ϕ is the carrier phase and $T=R_B^{-1}$ where R_B is the pulse transmission rate (the band rate). In L-level PAM, a_k can assume one of L values and the bit rate $R=R_B \log_2 L$. In binary PAM L=2 and $R_B=R$.

A special case of binary PAM is conventional ON-OFF amplitude shift keying (ASK) where $a_0=0$, $a_1=A$ and

$$g(t) = \begin{cases} 1 & 0 \leq t \leq T \\ 0 & t < 0;\ t > T \end{cases} \tag{3-2}$$

Each pulse g(t) is limited in duration to [0,T] and the absence or presence of a signal conveys the information.

Conventional ASK has three disadvantages:

1. The bandwidth of the transmitted signal s(t) is excessive.
2. The decision (slicing) threshold at the receiver depends on the received signal level which can range over 90 dB.
3. The peak-power capabilities of the transmitter are excessive by 3 dB, since (for equiprobable bits) no energy is transmitted half of the time.

The last two disadvantages are obviated by using binary antipodal ASK where a_0=-A and a_1=A in which case the decision level remains at zero independent of A. The disadvantage of this "improvement" is that coherent detection is required; with ON-OFF ASK incoherent detection is feasible. The excessive bandwidth which results from rectangular-pulse baseband signalling can be reduced by shaping the pulses g(t) and removing the constraint that they be non-zero only for $0 \leq t < T$. The result is an increase in the complexity of the digital data modulator and demodulator.

Use of more than two signal levels in the presence of large and rapid signal level variations requires adaptive adjustment of decision thresholds [W1]. Signal level estimation [L2, F1] needed for such adjustment would complicate the receiver and introduce estimation errors and ensuing error rate degredations.

Large and rapid signal level variations also require use of pulse shapes which are zero at all sampling times except the zeroth to avoid intersymbol interference (isi). Existing isi-combatting receiver techniques [B1, S1, T1, T2] are not directly applicable since these assume fixed and known signal levels.

The restriction to two-level PAM also prevents use of partial response signalling [K1, K2] which requires at least two decision thresholds both of which cannot be zero.

Implementation of modulators for PAM pulse shaping is greatly facilitated by digital electronic and microprocessor technology for digital waveform generation [G1, S2]. Decreasing component costs relieve the designer of the need to rely solely on analog components to generate modulator pulse shapes.

III-2 PAM Spectral Characterisitcs

The power spectral density of a transmitted L-level PAM signal s(t) is S(f) where

$$S(f) = \frac{1}{2}\,[P(f+f_s) + P\,(f-f_s)] \qquad (3\text{-}3)$$

where P(f) is the power spectral density of the baseband signal $p(t)= \sum_k a_k\, g(t-kT)$ in (3-1). If the symbols a_k are statistically independent then [L2]

$$P(f)= \frac{1}{T^2}\sum_{j=-\infty}^{\infty} \Big|\sum_{k=1}^{L} P_k\, G(j/T)\Big|^2\, \delta(f-\frac{k}{T}) + \frac{1}{T}\sum_{k=1}^{L} P_k\, |G(f)|^2 \qquad (3\text{-}4)$$

where P_k is the probability of level a_k, $\delta(f)$ is the unit impulse, and G(f) is the Fourier transform of g(t); thus

$$G(f) = \int_{-\infty}^{\infty} g(t)\, e^{-j2\pi ft}\, dt \qquad (3\text{-}5)$$

The first term in (3-4) is a non-information bearing line spectrum which vanishes when any positive level a_k is balanced by a corresponding negative level $a_j=-a_k$ with probability $p_j=-p_k$. Thus, for binary antipodal PAM with $a_1=A$ and $a_2=-A$, the first term in (3-4) is zero. For binary ON-OFF PAM the first term is not zero. In any case this first term depends on G(f) only at discrete values of frequency f=j/T, where j is an integer.

The challenge for the designer of the digital data modulator is to specify g(t) (or equivalently G(f)) to meet existing bandwidth constraints such as those in Fig. 2-3, to maximize the bit rate R, and to provide for digital data demodulation which is easily implemented and which yields an acceptable bit-error rate.

Fig. 3-1 shows three different pulse shapes for $g(t)$, all of which show zero intersymbol interference at the sampling times $t=nT$. Fig. 3-2 shows the corresponding plots of $S(f)$; the sampling periods have been selected to enable the spectra to fit tightly inside proposed spectral emission constraint of Fig. 2-3. Pulse shape (b) is time-limited but has spectral components at all frequencies, whereas pulse shapes (a) and (c) are not time-limited but are strictly limited in bandwidth.

Determination of the maximum permissible data rate is easiest for the sin t/t - type pulse with spectral density

$$S_a(f) = \begin{cases} K_a = 1/2\, W_a & |f| < W_a \\ 0 & |f| > W_a \end{cases} \qquad (3\text{-}6)$$

Amplitude K_a is selected to ensure that the transmitted power is equal to that of the unmodulated carrier signal. The transmitted power P_T in the transmitted signal

$$s(t) = p(t)\sqrt{2P}\cos(\omega_c t + \phi) \qquad (3\text{-}7)$$

is

$$P_T = P\int_{-\infty}^{\infty} P(f)\, df \qquad (3\text{-}8)$$

from which it follows immediately that $\int_{-\infty}^{\infty} P(f)\, df = 1$, where $P(f)$ is the power spectral density of baseband signal $p(t)$.

Use of (3-6) together with

$$10\log_{10}E(f) = 8 - 3.4\,|f| \qquad 10 \leq |f| \leq 20 \qquad (3\text{-}9)$$

where $E(f)$ is the spectral emission constraint boundary in Fig. 2-3 for frequencies $10 \leq |f| \leq 20$ (f in KHz), yields the maximum value of $W_a \simeq 15.5$ KHz. This value of W_a solves the following equation:

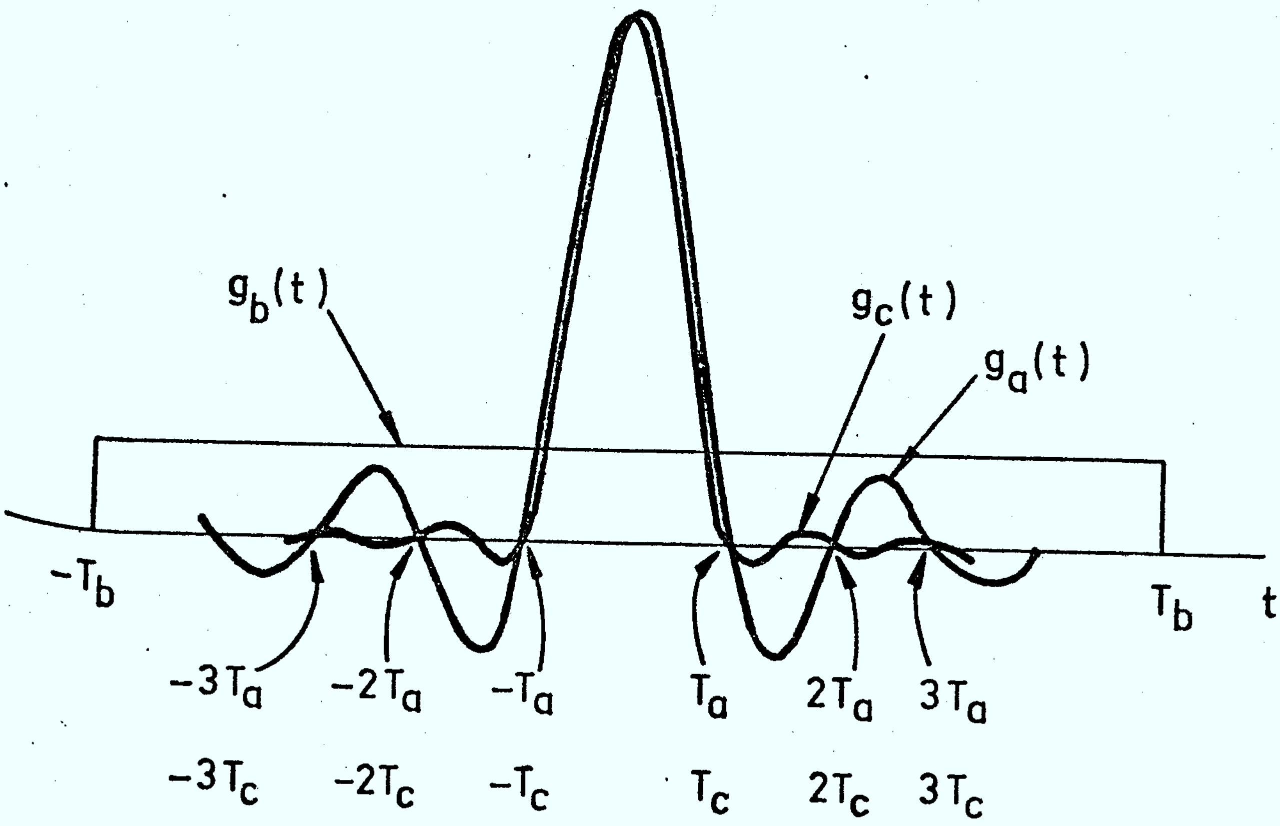

Fig. 3-1 PAM Pulse shapes. $g_c(t)$ is the pulse shape following matched filtering at the demodulator.

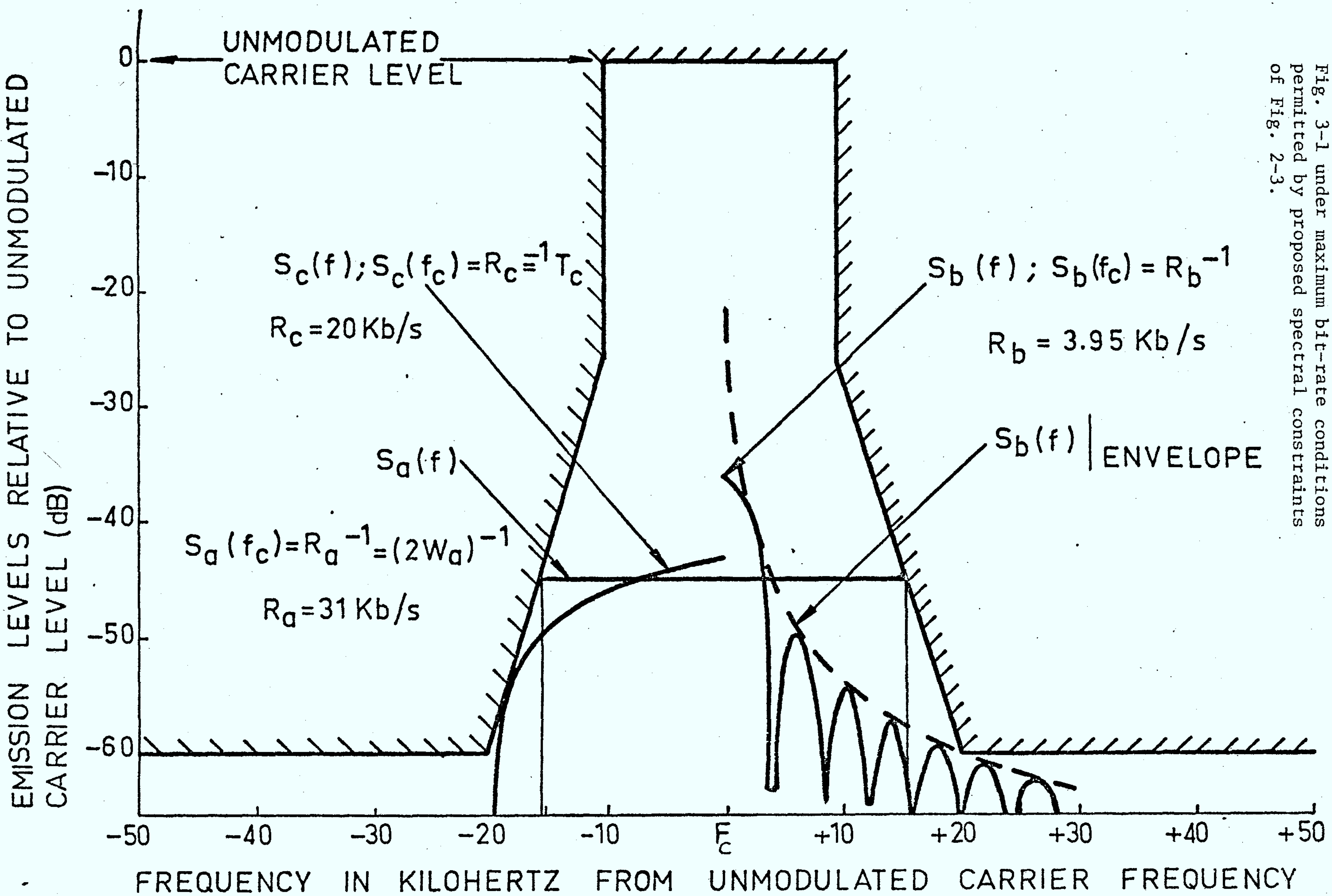

Fig. 3-2 Power spectral density for the pulse shapes in Fig. 3-1 under maximum bit-rate conditions permitted by proposed spectral constraints of Fig. 2-3.

$$10 \log_{10} (1/2\ W_a) = 8 - 3.4\ W_a \qquad (3\text{-}10)$$

It follows that the maximum bit rate is $R_a = 2\ W_a \approx 31$ Kb/s.

A similar determination of the maximum bit rate can be made for conventional binary antipodal ASK for which

$$S_b\ (f) = K_b\ T_b\ (\sin \pi f T_b / \pi f T_b)^2 \qquad (3\text{-}11a)$$

To ensure $\int_{-\infty}^{\infty} S_b(f)df = 1$, $K_b = 1$.

To estimate the maximum bit rate $R_b = T_b^{-1}$ for which $S_b(f)$ tightly fits the spectral emission constraint one can determine the minimum T_b for which the envelope $1/(\pi f)^2 T_b$ equals -60 dB for f=B=20 KHz. The resulting value of R_b is obtained from

$$R_b/(\pi\ 20{,}000)^2 = 10^{-6} \qquad (3\text{-}11b)$$

which yields $R_b \approx 3.95$ Kb/s. Thus, the maximum rate permissible for conventional rectangular pulses under the proposed emission constraint is 12.7% of that permitted for sin t/t - type pulses.

Pulses with a raised cosine spectrum having a large relative excess bandwidth α ($0 \leq \alpha \leq 1$) reduce considerably sampling-time error effects when $\alpha=0$ which corresponds to sin t/t -type pulses ($g_a(t)$ in Fig. 3-1); $\alpha=1$

corresponds to $g_c(t)$ in Fig. 3-1. To minimize the error probability [L1] the pulse actually transmitted has transform $\sqrt{G(f)}$ and the receiver filter also has transfer characteristic $\sqrt{G(f)}$ assuming an all-pass channel; thus raised cosine pulses actually appear at the demodulator output baseband filter. The power spectral density $S_c(f)$ of the transmitted raised-cosine signal is

$$S_c(f) = \begin{cases} K_c = T_c \\ \frac{T_c}{2}\left[1-\sin[(\pi T_c/\alpha)(f-\frac{1}{2T_c})]\right] & (1-\alpha)/2T_c \leq |f| \leq (1+\alpha)/2T_c \\ 0 & |f| > (1+\alpha)/2T_c \end{cases} \quad (3\text{-}12)$$

To fit $S_c(f)$ tightly inside the spectral emission characteristic one plots $10 \log_{10} S_c(f)$ for various values of T_c and thereby determines the maximum bit rate $R_c = T_c^{-1}$. The result will always exceed $[(1+\alpha)/4]R_a$ where R_a is the maximum bit rate for the case $\alpha=0$. Fig. 3-2 yields $R_c \simeq 20$ Kb/s which, because of the emission constraint roll-off from -26dB to -60dB greatly exceeds the lower bound $R_a/2 \simeq 15.5$ Kb/s.

Reduction of α would increase the bit rate as well as the timing error sensitivity. Thus, selection of excess bandwidth α permits a trade-off between these two quantities.

Selection of other pulse shapes limited to [0,T] such as triangles, half-cosine pulses and other zero isi pulses not limited to [0,T] results in bit rates which lie between $R_a \simeq 31$ Kb/s and $R_b \simeq 4$ Kb/s.

III-3 PAM Error Performance on Land Mobile Channels

For zero isi PAM pulses, including the three pulses in Fig. 3-1 and pulses with raised cosine spectra with $0\leq\alpha\leq1$ the minimum obtainable symbol error probability for L equally spaced levels is as follows [L1]:

$$P(e) = 2\left(1 - \frac{1}{L}\right) \operatorname{erfc}\left(\sqrt{\frac{3}{L^2-1} \quad \frac{2P}{R_B N_0}}\right) \tag{3-13}$$

where

$$\operatorname{erfc}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\alpha^2/2}\, d\alpha \tag{3-14}$$

In (3-13) P is the received signal power, R_B is the baud rate and $N_0/2$ the power spectral density of the white Gaussian noise. To obtain this minimum error rate requires matched filter detection followed by bit-synchronous sampling.

For binary signalling L=2, R_B=R is the bit rate, and

$$p = \operatorname{erfc}\left(\sqrt{2P/RN_0}\right) \tag{3-15}$$

where p is the bit error probability. Fig. 3-3 shows p vs $\rho_0 = 2P/RN_0$ for L=2.

In a Rayleigh fading environment, P(e) must be averaged over the received signal-to-noise ratio $2PT/N_0$, using the density function (2-2). For binary signalling,

$$p = \int_{-\infty}^{\infty} \operatorname{erfc}(\alpha)(\bar{\rho}_0)^{-1} \exp(-\alpha/\bar{\rho}_0)\, d\alpha = \frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\overline{\rho_0}^{-1}}}\right) \tag{3-16}$$

where $\overline{\rho_0}$ is the averaged received signal-to-noise ratio.

Fig. 3-3 shows p vs $\overline{\rho_0}$, and also demonstrates the difference in bit error rate behaviour for white Gaussian and Rayleigh fading channels.

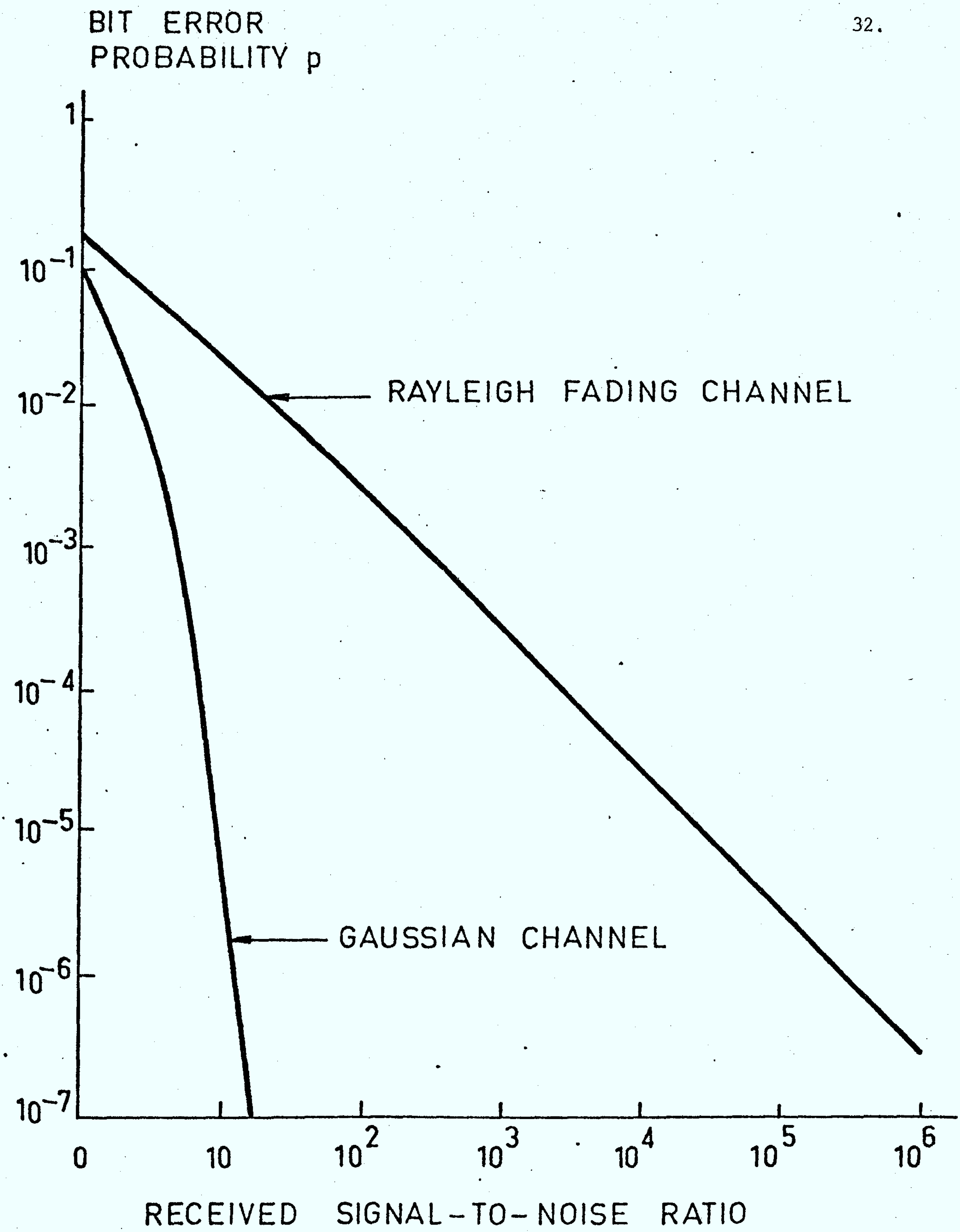

Fig. 3-3 Bit-error probability p for a white Gaussian channel and for a Rayleigh fading channel vs receive signal-to-noise ratio ρ_0 and $\overline{\rho_0}$.

For Rayleigh channels, p decreases linearly with $\overline{\rho_0}$; for Gaussian channels p decreases exponentially with ρ_0. To reduce the bit error rate from 10^{-3} to 10^{-5} requires a 3 dB increase in transmitter power for non-fading channels, and a 100 dB increase for Rayleigh channels, since both ρ_0 and $\overline{\rho_0}$ are proportional to transmitted power.

III-4 Techniques for Reducing PAM Error Rates

Recall from Section 2-4 that variations in received signal power due to Rayleigh fading can exceed 20 dB, with potential additional variations of 70 dB due to shadowing and distance attenuation. From (3-15) one sees that a SNR sufficient to yield $p \approx 10^{-7}$ at one (stationary) mobile location would yield $p \approx 1/2$ in another location where ρ_0 is reduced by 20 dB. Thus, in many cases the received SNR rather than bandwidth constraints would limit the information transmission rate.

The following approaches are available to combat noise on land mobile channels:

1. Increase the transmitter power
2. Reduce the noise and/or interference
3. Reduce the data rate
4. Employ spatial diversity
5. Reduce the radio coverage area of a region, thereby reducing loss due to transmitter-receiver separation
6. Employ forward error correction (FEC)
7. Employ some form of variable-rate information transmission.

Increasing the transmitter power is a "brute-force" approach which would be possible only if the existing transmitted power was below the lawful maximum.

Reduction of the noise implies an more expensive receiver and/or reduced co-channel and/or adjacent channel interference. Reduced adjacent channel interference implies tighter out-of-band emission standards or increased channel separation. Tightened emission standards imply narrowed transmission spectra via reduced data rates. Increased channel separation implies fewer available channels. Reduced co-channel interference implies fewer channels available to any cellular region [J1] which implies increased waiting time or blocking probability and reduced overall system information throughput.

Lowering the bit rate reduces bit error rate p in accordance with (3-15) or (3-16). However, the channel bandwidth occupied is also reduced in proportion to R (see Fig. 3-2). Utilization of this available bandwidth can result in a further SNR increase - and a corresponding further improvement in bit-error rate. For example, the binary PAM baseband waveform could frequency modulate a carrier whose frequency deviation is low enough to meet spectral emission standards. Demodulation by an FM receiver operating above threshold would yield a PAM signal with a SNR of $3\beta^2 PT/N_0$ where bandwidth expansion factor $\beta = \sqrt{W_c/W}$; W_c is the RMS channel bandwidth [W1, L3] and W the carrier signal bandwidth. Since W_c is relatively constant for small transmission spectral changes, a linear reduction in data rate R would yield a linear increase in both β and T and a SNR increase proportional to R^{-3}.

Even in the absence of vehicle motion spatial diversity increases PT/N_0, by virtue of signal averaging. Various diversity combining schemes have been analysed and shown to yield large SNR improvements even with receive noise correlations [J1, M2]. During vehicle motion spatial diversity greatly reduces SNR variations due to Rayleigh fading, shadowing and changes in transmitter-receiver separation, as noted in Section 2-4. Reduction in radio coverage area reduces received signal level variations resulting from changes in receiver-transmitter separation. The penalty is more base stations. An added benefit is fewer mobiles per base station, which increases the potential information throughput over the entire coverage area. The selection of base station sites, coverage areas, and assignment of frequency channels to coverage areas is a spectrum management problem of considerable current interest and importance [J1, K3].

Forward error correction is a form of time diversity and is considered in detail in Chapter 5. Variable rate transmission was discussed in Section 2-4; its use in the form of automatic repeat request (ARQ) transmission in conjunction with error detection is the subject of Chapter 6.

III-5 Single Sideband (SSB), Vestigital Sideband (VSB) and Quadrature Amplitude Modulation (QAM).

It is well known that the required PAM channel bandwidth can be reduced by a factor of two by single sideband transmission of the modulated signal [L1]. The resulting transmitted signal is of the form

$$s(t) = m(t)\cos(\omega_c t+\phi) + \hat{m}(t)\sin(\omega_c t+\phi) \tag{3-17}$$

where $\hat{m}(t)$ is the Hilbert transform of data pulse sequence $m(t)$. Thus

$$\hat{m}(t) = \int_{-\infty}^{\infty} \frac{1}{\pi\tau}\, m(t-\tau)\,d\tau \tag{3-18}$$

$$m(t) = \sum_k a_k \; g(t-kT) \qquad (3\text{-}19)$$

where g(t) is the basic modulator pulse shapes. SSB signal spectra are obtained from DSB spectra simply by deleting either the upper or lower side-band.

From Fig. 3-2 one sees that if sin t/t - type pulses are used, the spectral emission constraint permits a SSB data rate of 62 Kb/s. Unfortunately the resulting SSB signal has no excess-bandwidth; consequently the error probability is drastically reduced by any carrier phase detection error or by any sampling-time jitter [L1, L2, G2, H1, F2].

From (3-17) one obtains some appreciation of the effects of phase error on the received signal. For DSB $\hat{m}(t) = 0$, and if ϕ now represents the receiver phase error, then the baseband portion of the demodulated output signal following multiplication of the received signal by $\cos \omega_c t$ is $\cos \phi \; m(t)$ + noise. The received signal energy is thus reduced by $\cos^2 \phi$. Knowledge of $f_\phi(\alpha)$, the probability density of ϕ permits calculation of the error rate averaged over ϕ [M1, R1, L2]. The actual form of $f_\phi(\alpha)$ depends on the phase estimator used.

For SSB a phase error ϕ yields a demodulated baseband signal $m(t) \cos \phi + \hat{m}(t) \sin \phi$ + noise. Not only is m(t) reduced by $\cos \phi$, cross-talk from the quadrature channel is also present and increases with ϕ.

Some tolerance to phase and timing errors is gained by using vestigital sideband modulation (VSB) in which a vestige of the suppressed sideband is transmitted, together with the retained sideband [L1, H1, D1]. A VSB signal has the form of (3-17) except that

$$\hat{m}(t) = \int_{-\infty}^{\infty} h_\beta(\tau)\; m(t-\tau)\; d\tau \tag{3-20}$$

where $h_\beta(t)$ is the impulse response of the filter that shapes the quadrature component of the VSB signal. SSB is a special case of VSB in which $h(t)=1/\pi t$.

The reduced sensitivity to phase and timing errors of VSB with respect to SSB is achieved at the expense of increased transmission bandwidth. For example, a VSB signal with identical vestigital and sideband roll-offs and having the raised cosine spectral shape of (3-12) with $\alpha=1$ yields a transmitted signal whose spectral density is identical to that of a DSB signal with a raised cosine spectrum with $\alpha=\frac{1}{2}$.

The design of pulse shapes for minimizing effects of timing and phase errors has received recent attention [M1, G2, T2, F1, F2, L1]. These studies indicate that it is the excess bandwidth beyond the minimum required which determines the sensitivity to phase and timing errors.

If zero phase error occurs in (3-17) then $m(t)$ and $\hat{m}(t)$ can be uncorrelated data streams which can be recovered by multiplication of the received signal by $\cos \omega_c t$ and $\sin \omega_c t$, respectively. Quadrature amplitude modulation (QAM) results, and for the spectral characteristic in Fig. 3-2 a bit rate of up to 62 Kb/s is permitted by transmission of alternate data bits via in phase and quadrature carriers. As with SSB, absence of excess bandwidth causes phase and timing errors to have devastating effects on bit-error rates [G2, R1, H1].

The actual details of how excess bandwidth, synchronization algorithms, transmission pulse shapes, pilot tone power allocation, fading rates and fading levels, and error rates interact is rather complicated and could form the basis for a separate study.

III-6 Use of PAM on Existing FM Land Mobile Voice Channels

Spectral constraints for existing FM land mobile voice channels require that 99 percent of the transmitted spectral power lie within ± 8KHz of the carrier frequency.

If sin t/t type signals shown in Fig. 3-1 are used, then from (3-6) it follows that the maximum permissible bit-rate is R=16.162 Kb/s. If raised cosine signalling with $\alpha=1$ is used then the maximum permissible bit rate is R=8.183 Kb/s. For values of $0<\alpha<1$, the maximum possible bit rate is (approximately) $R=16.162/(1+\alpha)$ for signals with raised cosine spectra.

Use of SSB, VSB or QAM permits (in theory) a data rate of up to 32.324 Kb/s with sin t/t-type signalling, although phase and timing errors would require lower rates in a practical situation.

Reference to others' works [S3, G3] shows that the maximum data rate permissible for conventional rectangular baseband pulses to be R≈1.8 Kb/s.

The maximum data rates R above are obtained by determining the maximum value of R for which

$$\int_{-B}^{B} P(f)\,df = 0.99 \int_{-\infty}^{\infty} P(f)\,df \qquad (3\text{-}21)$$

where P(f) is the power spectral density of the baseband PAM signal, (which depends on R) and B=8 KHz.

III-7 References for Chapter 3

[B1] T. Berger and D.W. Tufts, "Optimum pulse amplitude modulation, part I: transmitter - receiver design and bounds from information theory", vol. IT-13, pp. 196-208, Apr. 1967.

[D1] R. Dogliotti and U. Mengali, "Tracking loops for carrier reconstruction in vestigital sideband (VSB) supressed-carrier data transmission", IEEE Trans. Commun., vol. COM-25, pp. 374-379, Mar. 1975.

[F1] D. Falconer and J. Salz, "Optimal reception of digital data over the Gaussian channel with unknown phase jitter", IEEE Trans. Inform. Theory, vol. IT-23, pp. 117-126, Jan. 1977.

[F2] L.E. Franks, "Further results on Nyquist's problem in pulse transmission", IEEE Trans. Commun. Technol., vol. COM-16, pp. 337-340, Apr. 1968.

[G1] P.J. van Gerwin, N.A.M. Vorhoeckx, H.A. van Essen and F.A.M. Snijders, "Microprocessor implementation of high-speed data modems ", IEEE Trans. Commun., vol. COM-25, pp. 238-250, Feb. 1977.

[G2] R.D. Gitlin and E.Y. Ho, "The performance of staggered quadrature amplitude modulation in the presence of phase jitter", IEEE Trans. Commun., vol. COM-23, pp. 348-352, March 1975.

[G3] S.A. Gronemeyer and A.L. McBride, "MSK and offset QPSK modulation", IEEE Trans. Commun., vol. COM-24, pp. 809-820, Aug. 1976.

[H1] F.S. Hill, "Optimum pulse shapes for pulse-amplitude modulation data transmission using vestigital sideband modulation" IEEE Trans. Commun. vol. COM-23, pp. 352-361.

[J1] W.C. Jakes, Microwave Mobile Communications, New York, N.Y.: Wiley, 1974.

[K1] P. Kabal and S. Pasupathy, "Partial-response signalling", IEEE Trans. Commun., vol.COM-23, pp. 921-934, Sept. 1975.

[K2] E.R. Kretzmer, "Generalization of a technique for binary data communication", IEEE Trans. Commun., vol. COM-25, pp. 738-744, July 1977.

[K3] T.J. Kahwa and N.D. Georganis, "A hybrid channel assignment scheme in large-scale, cellular-structured mobile communication systems", IEEE Trans. Commun., vol. COM-26, pp. 432-439, Apr. 1978.

[L1] R.W. Lucky, J. Salz and E.J. Weldon, Jr., Principles of Data Communication, New York: McGraw Hill, 1968.

[L2] W.C. Lindsay and M.K. Simon, Telecommunication Systems Engineering Englewood Cliffs, N.J.: 1973, Ch.1.

[L3] B.P. Lathi, An Introduction to Random Signals and Communication Theory, Scranton, Pa.; International Textbook Company, 1968.

[M1] R. Matyas, "Effect of noisy phase reference signals on coherent detection of FFSK signals", IEEE Trans. Commun., vol. COM-26, pp. 807-815, June 1978.

[M2] Y. Miyagaki, N. Morinaga, and T. Namekawa, "Error probability characteristics for CPSK signal through m-distributed fading channel", IEEE Trans. Commun., vol. COM-26, pp. 88-100, Jan. 1978.

[R1] S.A. Rhodes, "Effect of noisy phase reference on coherent detection of offset-QPSK signals", IEEE Trans. Commun., vol. COM-22, pp. 1046-1055, Aug. 1974.

[S1] J. Salz, "Optimum mean-square decision feedback equalization", Bell Syst. Tech. J., vol. 52, pp. 1341-1374, Oct. 1973.

[S2] F.A.M. Snijders, NAM Voerhoeckx, H.A. van Essen and P.J. van Gerwin, "Digital generation of linearly modulated data waveforms", IEEE Trans. Commun., vol. COM-23, pp. 1259-1270, Nov. 1975.

[S3] M.K. Simon, "A generalization of minimum-shift-keying (MSK)-type signalling based on input data symbol pulse shaping", IEEE Trans. Commun., vol. COM-24, pp. 845-856, Aug. 1976.

[T1] D.W. Tufts, "Nyquists problem - the joint optimization of transmitter and receiver in pulse amplitude modulation", Proc. IEEE, vol. 53, pp. 248-259, Mar. 1965.

[T2] D.W. Tufts and T. Berger, "Optimum pulse amplitude modulation, part II: inclusion of timing jitter", IEEE Trans. Inform. Theory, vol. IT-13, pp. 209-216, Apr. 1967.

[W1] J.M. Wozencraft and I.M. Jacobs, <u>Principles of Communication Engineering</u>. New York, McGraw-Hill, 1965, Ch. 4.

IV NON-LINEAR MODULATION (FSK, PSK) FOR DIGITAL TRANSMISSION OVER LAND MOBILE CHANNELS

IV-1 Description of Frequency Shift Keying (FSK) and Phase Shift Keying (PSK).

Non-linear modulation of the argument of a sinusoidal carrier signal generates angle modulated transmitted signals.

Define

$$x(t) = \sum_{k=-\infty}^{\infty} a_k \, g(t-kT) \qquad (4\text{-}1)$$

where $\{a_k\}$ denotes the symbol sequence, $g(t)$ a basic pulse shape and T the symbol period. A digital FM modulated carrier signal $s(t)$ with radian frequence ω_c and average power P results when $x(t)$ modulates the carrier, as follows:

$$s(t) = \sqrt{2P} \; \cos \left(\omega_c t + \omega_d \int_{-\infty}^{t} x(\tau) \, d\tau + \theta\right) \qquad (4\text{-}2)$$

In (4-2) θ is a (random) phase angle and the instantaneous frequency is $\omega_c + \omega_d \, x(t)$. When $x(t)$ is normalized to have unity peak amplitude, ω_d is the peak frequency deviation of $s(t)$. In general increasing ω_d increases the bandwidth of $s(t)$ and reduces the obtainable bit error rate.

When $g(t)$ in (4-1) is a rectangular pulse

$$g(t) = \begin{cases} 1 & 0 < t \leq T \\ 0 & t \leq 0, \quad t > T \end{cases} \qquad (4\text{-}3)$$

and conventional frequency shift keying (FSK) results. It is not yet clear how best to select $g(t)$ to give $s(t)$ desired spectral properties and maintain low error rates.

A conventional digital phase modulated signal is of the form

$$s(t) = \sqrt{2P} \sum_{k=-\infty}^{\infty} g(t-kT)(\cos \omega_c t + \phi_k) \qquad (4-4)$$

where rectangular pulse g(t) is defined in (4-3). Phase sequence $\{\phi_k\}$ varies in accordance with symbol sequence $\{a_k\}$; for example $\phi_k=2k\pi/L$ $(k=1,2,\ldots,L)$ where L is the number of symbol levels.

Equation (4-4) may be generalized in two ways. First, the amplitudes of the rectangular pulses may be modulated by arbitrary pulse shapes h(t). Second, the constant phase angles ϕ_n may be replaced by phase functions $f_{\phi_n}(t)$. For simplicity of implementation, we restrict $f_{\phi_n}(t) = \phi_n f(t)$. With these generalizations, the phase modulated signal becomes

$$s(t) = \sqrt{2P} \sum_{k=-\infty}^{\infty} h(t-kT) \cos(\omega_c t + \sum_{n=-\infty}^{\infty} \phi_n f(t-nT)) \quad 0<t\leq T \qquad (4-5)$$

Neither h(t) nor f(t) need be restricted to the interval [0,T].

Analog FM and PM spectra are similar to one another and differ only by virtue of an integration or differentiation of the baseband waveform. Digital FSK and PSK differ by virtue of the fact that phase continuity is normally preserved across FSK symbol boundaries, but not across PSK symbol boundaries.

In contrast to digital FSK, digital PSK does not permit bandwidth expansion and is similar in spectral characteristics to PAM.

IV-2 Spectra and Bit-Error Rates for PSK

Eqn. (4-4) which describes PSK can be rewritten as follows:

$$s(t) = \sqrt{2P}\left[\left(\sum_{k=-\infty}^{\infty} b_k\, g(t-kT)\right)\cos \omega_c + \left(\sum_{k=-\infty}^{\infty} c_k\, g(t-kT)\right) \sin \omega_c t\right] \quad (4\text{-}5)$$

where

$$b_k = \cos \phi_k \quad (4\text{-}6a)$$

$$c_k = \sin \phi_k \quad (4\text{-}6b)$$

$$\phi_k = 2\pi k/L \quad (4\text{-}6c)$$

Conventional PSK consists of quadrature PAM signals. For binary PSK $\phi_k = 0$ or π, in which case $b_k = \pm 1$ and $c_k = 0$. The resulting signal $s(t)$ is identical to PAM with rectangular baseband pulses. From the discussion in Chapter 3 it follows immediately that $S(f)$, the power spectral density of $s(t)$ is given by the response $S_b(f)$ in Fig. 3-2.

As indicated in Fig. 4-1, optimal detection of conventional PSK in the presence of additive white Gaussian noise involves filtering of the received signal by filters matched to $\sin \omega_c t$ and $\cos \omega_c t$, and processing of the sampled outputs by an optimal decision device [W1, L1]. Fig. 4-2 shows the decision regions for 8-level PSK assuming equiprobable symbols. For example the decision that $\vec{s}_1(\phi_1 = \pi/8)$ has occurred will result if and only if $r_1 > 0$ and $0 < \tan^{-1} r_2/r_1 < \pi/4$.

For L=2 and L=4, symbol error probability P(e) for a white Gaussian channel is obtainable in closed form [L1, W1]:

$$P(e) = \operatorname{erfc}\left(\sqrt{\rho_0}\right) \qquad (L=2) \qquad (4\text{-}7)$$

$$P(e) = 1 - \left(1 - \operatorname{erfc}\left(\sqrt{\rho_0/2}\right)\right) \qquad (L=4) \qquad (4\text{-}8)$$

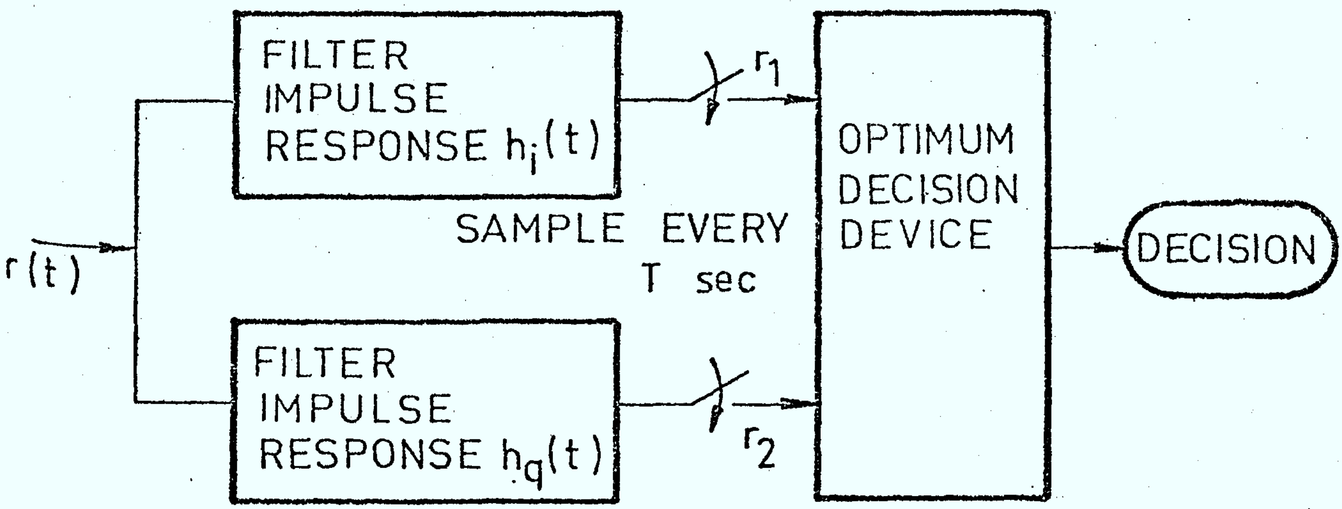

Fig. 4-1 Optimum PSK detector. $h_i(t)=[\cos \omega_c t]U(t)$. $h_q(t)=[\sin \omega_c t]U(t)$. $U(t)=1;\ 0<t\leq T$; $U(t)=0$ otherwise.

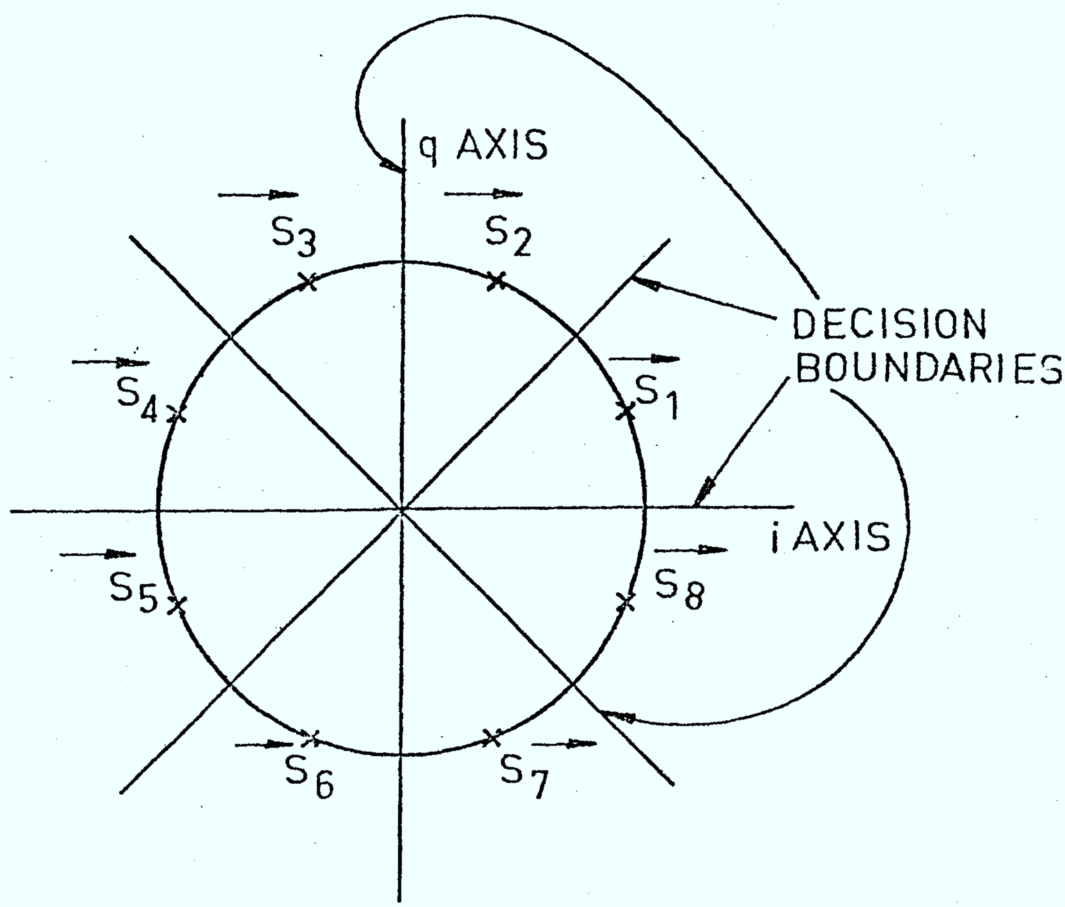

Fig. 4- 2 Decision space for 8-level PSK : $\{\vec{S_i}\}$ denote the transmitted signal vectors.

where erfc (x) is defined by (3-14) and $\rho_0 = 2PT/N_0$. For other values of L, P(e) must be evaluated numerically [L1], or bounded. A simple bound is as follows [T1]:

$$\alpha \lesssim P(e) \lesssim 2\alpha \tag{4-9a}$$

$$\alpha = \text{erfc}\ [(\sin \pi/L)\ \sqrt{\rho_0}\ ] \tag{4-9b}$$

As L increases the upper bound in (4-9) becomes increasingly tight. Since [W1]

$$\text{erfc}\ (x) \lesssim (1/2)\ \exp\ (-x^2/2) \tag{4-10}$$

it follows that for L-level PSK

$$P(e) \lesssim \exp\ [-(\sin^2\pi/L)\rho_0/2] \tag{4-11a}$$

Asymptotically for large ρ_0 [L1]:

$$p(e) \approx \exp\ (-\rho_0 \sin^2\pi/L) \tag{4-11b}$$

For Rayleigh fading channels P(e) must be averaged over the range of ρ_0 to yield $\overline{\rho_0}$. Averaging the bounds in (4-9) yields

$$\bar{b} \lesssim P(e) \lesssim 2\bar{b} \tag{4-12a}$$

$$4\bar{b} \approx (\overline{\rho_0}\ \sin^2\pi/L)^{-1} \qquad (\rho_0 >> 1) \tag{4-12b}$$

For 2-level PSK, P(e) is given by (3-16) assuming above threshold operation. For 4-level PSK

$$P(e) = 1 - [1 - \frac{1}{2}(1 - \frac{1}{\sqrt{1+2/\overline{\rho_0}}})]^2 \tag{4-13}$$

Because the optimum PSK receiver in Fig. 4-1 compares ratios of numbers based on the received signal, PSK on Rayleigh channels need not be restricted to binary. This situation is in contrast to PAM where decision levels for multilevel PAM require adjustment in accordance with received signal levels.

The PSK error exponent in (4-11) varies as $\sin^2 \pi/L$. Consequently, for a given band rate 1/T, the maximum permissibly value of L and hence the maximum bit rate would be determined by the maximum allowed bit error probability.

For fixed a value of L, the maximum bit rate $R=\log_2 L/T$ would be determined by spectral emission constraints on a relatively noise-free channel and by noise on a noisy channel. Bit rate limitations based on spectral emission constraints are summarized for 2- and 4-level PSK in Figs. 4-3 and 4-4 and follow directly from the results in Chapter 3. Also shown in Figs. 4-3 are symbol error probabilities, which also apply to Fig. 4-4.

Because 2-level PSK and binary antipodal PAM are identical, Fig. 3-3 represents binary PSK error probability vs ρ_0 and $\overline{\rho_0}$.

IV-3 Differential PSK

Use of differential PSK (DPSK) obviates the need to track the carrier signal phase. In DPSK the received symbol phase is compared with that of the previously detected symbol. The added receiver complexity resulting from differential detection is usually more than offset by the absence of carrier phase recovery equipment. Some error rate increase occurs since errors tend to occur in pairs; if the detected phase of a symbol is incorrect, the differentially a detected phase of the following symbol is also likely.

For differential detection and large signal-to-noise ratio ρ_0 [L1],

$$P(e) \simeq \exp\left(-2\, \rho_0 \sin^2(\pi/2L)\right) \qquad (4\text{-}14)$$

Modulation	Baseband Pulse Shape	Maximum Data Rate (Kb/s)	Symbol Error Probability: Gaussian Channel $\rho = 2P/N_o$	Symbol Error Probability: Rayleigh Channel $\bar{\rho} = 2\bar{P}/N_o$
2Ø-PSK(PAM)	rectangular	4		
2Ø-PSK(PAM)	sint/t	31		
2Ø-PSK(PAM)	raised cosine spectra; x=1	20	$\mathrm{erfc}(\sqrt{\rho/R}\,)$	$\simeq \frac{1}{4\,\bar{\rho}/R}$
SSB (binary)	sint/t	62		
2Ø-DPSK	rectangular	4		
2Ø-DPSK	sint/t	31	$\frac{1}{2}\exp(-\rho/R)$	$\frac{1}{2(1+\bar{\rho})}$
2Ø-DPSK	raised cosine; x=1	20		
4Ø-DPSK	rectangular	8		
4Ø-DPSK	sint/t	62	$1-[1-\frac{1}{2}\exp(-\rho/2R)]^2$	$1-[1-\frac{1}{2(1+\bar{\rho})}]^2$
4Ø-DPSK	raised cosine; x=1	40		
$\left\{\begin{array}{l}\text{Binary FSK,} \\ h = 0.5 \\ \text{Binary MSK}\end{array}\right\}$	rectangular	36	$\mathrm{erfc}(\sqrt{\rho/\ R}\,)$	$\simeq \frac{1}{4\bar{\rho}/R}$

Fig. 4-3 Summary of Maximum Bit Rates Based Solely on the Spectral Emission Constraints of Fig. 2-3

Modulation	Baseband Pulse Shape	Maximum Data Rate (Kb/s)
2Ø-PSK(PAM)	rectangular	1.8
2Ø-PSK(PAM)	sint/t	16.2
2Ø-PSK(PAM)	raised cosine spectra; α=1	8.2
SSB(binary)	sint/t	32.3
2Ø-DPSK	rectangular	1.8
2Ø-DPSK	sint/t	16.2
2Ø-DPSK	raised cosine; α=1	8.2
4Ø-DPSK	rectangular	3.6
4Ø-DPSK	sint/t	32.4
4Ø-DPSK	raised cosine; α=1	16.2
Binary FSK h=0.5 Binary MSK	rectangular	13.3

Fig. 4-4 Summary of Maximum Bit Rates Based Solely on Existing FM Voice Channel Emission Constraints

Comparison of the exponent of (4-14) with that of (4-11) shows a DPSK degradation factor

$$\partial = \sin^2\pi/L \; / \; 2 \sin^2\pi/2L \qquad (4\text{-}15)$$

which varies from $\partial=1$ (zero dB) when L=2 to a maximum of $\partial=2$ (3 dB) when $L \to \infty$.

For finite SNR values, binary DPSK shows a higher error rate than PSK. The maximum SNR degradation is 3 dB which occurs a low SNR values. For binary DPSK on Gaussian channels [L1, W1]

$$P(e) = 0.5 \exp(-\rho_0) \qquad (4\text{-}16)$$

On Rayleigh channels [E1]

$$P(e) = 1/2 \; [1 + \overline{\rho_0}] \qquad (4\text{-}17)$$

The spectra of the transmitted DPSK signal is identical to that of the transmitted PSK signal if the symbols to be transmitted are statistically independent. It follows that conventional rectangular pulse DPSK signals, like PSK signals have the spectral response show by $S_b(f)$ in Fig. 3-2.

Detection of DPSK signals involves generation of $z_1(t)=r_1(t)r_1(t-T)$ where $r_1(t)$ is the output from filter $h_i(t)$ in Fig. 4-1. Signal $z_1(t)$ is then lowpass filtered to remove double frequency carrier terms to yield a statistic y_1. A similar type of operation is performed on the output of the filter with impulse response $h_q(t)$ to yield statistic y_2. If the pulse g(t) in (4-4) has zero isi at the sampling instants, then y_1 and y_2 can be processed by an optimum decision device to detect the transmitted sequence $\{a_k\}$ [L1].

Figs. 4-3 and 4-4 summarize maximum bit rates permitted by the spectral emission constraint in Fig. 2-3.

IV-4 FSK Signal Spectra

It is well known that the power spectral density S(f) of the FSK signal s(t) in (4-2) is not easily determined [G1, T2, L1, R1, R2, S1]. In two special and limiting cases, however, spectral determination is possible, as follows:

1. The peak frequency deviation ω_d in (4-2) is large in which case wideband FM results and

$$S(f) \simeq K\, f_x(f) \tag{4-14}$$

where $f_x(\alpha)$ is the amplitude probability density of x(t) and K is a constant selected in accordance with the transmitted power.

2. The frequency deviation ω_d is small in which case narrowband for results. From (4-2),

$$s(t) \simeq \sqrt{2p}\,[\cos(\omega_c t+\theta)-\omega_d[\int_{-\infty}^{t} m(\tau)d\tau]\sin(\omega_c t+\theta)] \tag{4-15}$$

$$S(f)/P \simeq \frac{1}{2}\{\delta(f+f_c)+\delta(f-f_c)\}+\frac{\omega_d^2}{2}\left\{\frac{S_m(f+f_c)}{[2\pi(f+f_c)]^2}+\frac{S_m(f-f_c)}{[2\pi(f-f_c)]^2}\right\} \tag{4-16}$$

where $\delta(f)$ is the unit impulse.

For wideband FM the amplitude probability density function of the modulating signal determines the transmitted signal spectra. For narrowband FM, the spectral density of the modulating signal determines the transmitted signal spectra which, because of the factor $(1/2\pi f)^2$ in

(4-15b) can be narrower than that of the modulating signal itself. Thus, FM can provide either bandwidth expansion or bandwidth compression. One would expect FM spectra for intermediate values at ω_d to depend on both the amplitude probability density and the spectra of the modulating signal. For arbitrary modulating signals x(t) in (4-2) or even for conventional rectangular FSK pulses g(t) closed form expressions for S(f) are unavailable, and approximate methods must be used to find S(f). Garrison's [G1] approach involves approximation of the baseband modulating signal pulses g(t) in (4-1) by a duration-limited step-wise approximation. The method is fairly general but as Garrison states [G1]:

> ".... the casual reader should not be mislead into believing that arbitrary configurations for premodulation filtering are amenable to analysis. Machine computation time can readily become prohibitive unless the pulse approximation can be conveniently bounded in both duration and quantization number."

When the peak frequency deviation $f_d=\omega_d/2\pi=1/4T=R/4$ where R is the bit rate, minimum shift keying (MSK) results [S1, A1, G2, D1]. For binary MSK the spectrum S(f) is known exactly [G2]:

$$S_{MSK}(f) = \frac{8PT(1+\cos 4\pi fT)}{\pi^2(1-16T^2f^2)^2} \qquad (4\text{-}17)$$

where f is the frequency offset from the carrier.

Fig. 4-5 shows $S_{MSK}(f)$ plotted to fit tightly within the spectral emission constraints of Fig. 2-3. These constraints permit a bit rate R=36 Kbs.

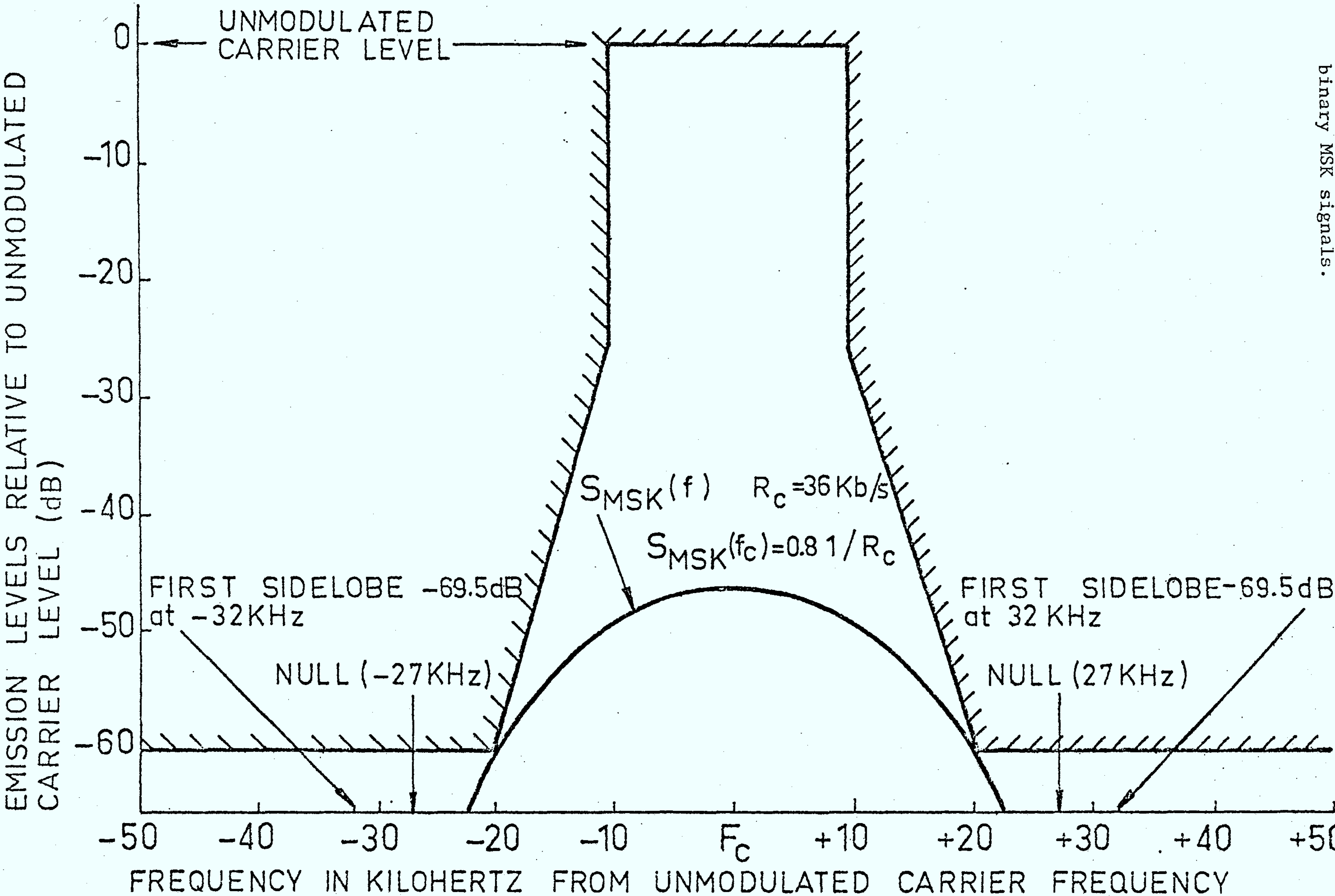

Fig. 4-5 Power spectral density for binary MSK signals.

It is possible to obtain closed form expressions for M-ary MSK spectra [S2]. The (complex) expressions are not presented here, since noise rather than bandwidth would likely limit the transmission rate of MSK on actual mobile channels.

Determination of the maximum data rate permitted under the spectral emission constraints in Fig. 2-3 for FSK signals with arbitrary deviation ratios $2f_d/R$ is a computationally tedious task. In effect, numerically calculated spectra would have to be plotted on Fig. 2-3 for various values of R. That value of R yielding a snug fit to the constraint would define the maximum R. In calculating spectra it would be necessary to determine the spectral amplitude relative to the unmodulated carrier power.

IV-5 Bit-Error Rates for FSK

A conventional FM receiver of the type normally used for (incoherent) detection of FSK signals appears in Fig. 4-6. The transfer characteristics of both the bandpass and lowpass filters affect the bit-error rates [T3]. Even if the filters' spectral shapes are fixed the bandwidths are important; if the bandwidth of either filter is too narrow, the output signal is badly distorted; if the bandwidths are too wide the noise power transmitted is excessive.

Tjhung and Whittke [T3] used computational techniques to optimize the bandwidths of both Gaussian and rectangular bandpass filters for a minimum binary FSK bit error rate. The lowpass filter was an integrate-and-dump. Various peak frequency deviations were considered, with $h=2f_d/R \approx 0.7$ and a bandpass bandwidth $B \approx R$ being optimum. White Gaussian noise was added during transmission. Over the bit-error range $p \lesssim 10^{-7}$ signals

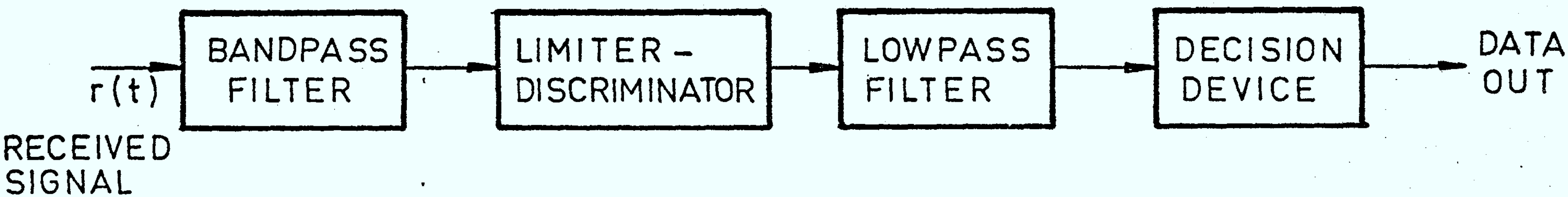

Fig. 4-6 Conventional FM receiver.

with h = 0.5 are seen [T3] to be approximately 2 dB inferior, in terms of transmitted power, to the FSK signals with h = 0.5.

Actually, MSK signals can be detected coherently, since they can be viewed as staggered QAM signals of duration 2T on each quadrature channel [S2, D1]. Since coherent detection provides a 3 dB transmitter power gain over incoherent detection, FSK signals with h = 0.5 do seem to yield the minimum error probability. Because of the lower value of f_d they would also be expected to yield a higher bit rate than FSK signals with h=0.7.

For binary MSK signals with coherent detection, the bit-error probability p is given by (4-7) for Gaussian channels and by (4-13) for Rayleigh channels. For incoherent detection of MSK signals [E1]

$$P = 0.5 \exp(-\rho_0/2) \qquad \text{(Gaussian channels)} \qquad (4\text{-}18)$$

$$P = 0.5 (1 + \overline{\rho_0}/2)^{-1} \qquad \text{(Rayleigh channels)} \qquad (4\text{-}19)$$

FSK signals with arbitrary deviation ratios show P(e) vs ρ_0 behaviour as follows [L1] where L is the number of FSK levels:

$$P(e) \simeq (1-\frac{1}{L}) \frac{\cot(\pi/2L)}{\sqrt{\cos(\pi/L)}} \exp[-2\rho\sin^2(\pi/2L)] \qquad (4\text{-}20)$$

For Rayleigh channels it follows from (3-16) and (4-20) that

$$P(e) \simeq (1-\frac{1}{L}) \frac{\cot(\pi/2L)}{\sqrt{\cos(\pi/L)}} \frac{1}{1+2\rho\sin^2(\pi/2L)} \qquad (4\text{-}21)$$

It is seen that DPSK and FSK have the same asymptotic error performance. However FSK, under the spectral constraints in Fig. 2-3 has a higher allowable data rate for binary signalling.

IV-6 Digital Transmission of PSK and FSK over Existing Land Mobile FM Voice Channels

The maximum bit rates permitted under the existing FM voice channel emission constraint which confines 99% of the transmitted energy to ±8 KHz appear in Fig. 4-4. The results for PSK and DPSK follow directly from the results in Chapter III and the fact that PAM and PSK produce identical spectra.

The results for MSK are obtained from Simon's [S2] paper, which shows MSK to have the highest bit rates for all frequency pulse shapes g(t) as defined in (4-1). The curve of Gronemeyer and McBride [G2] indicates this same maximum rate of 13.3 Kb/s.

Many conventional FM modulators employ pre-emphasis, with complimentary de-emphasis at the demodulator. Our results for FM assume that pre-emphasis and de-emphasis is either not employed, or is neutralized by complimentary baseband filters at both transmitter and receiver. Neutralization implies de-emphasis of the baseband waveform prior to impression on the pre-emphasized FM modulator, and pre-emphasis of the demodulated signal prior to data recovery.

IV-7 PSK and FSK Bandwidth Reduction Alternatives

Two approaches are available to reduce the bandwidth of conventional PSK signals. One involves the replacement of conventional rectangular pulses g(t) in (4-4) with pulse h(t) such as raised cosine pulses with desirable spectral properties. The resulting error probability expressions for PSK and DPSK in Sections 4-2 and 4-3 apply for any pulse shapes h(t) which do not cause intersymbol interference. The disadvantages of using non-rectangular pulses is that more complex pulse-shaping instrumentation is needed at the transmitter and receiver and that the transmitter does not operate continuously at peak power as for rectangular pulse shapes.

An alternative approach to improving PSK spectra is to relax the usual restriction that phase ϕ_k remain constant during the symbol period T. Although various authors [P1, G3] have determined power spectra when f(t) in (4-5) assumes a variety of pulse shapes, only Prabhu [P1] appears to have calculated the resulting symbol error probability. Prabhu's detailed study yields two conclusions. First, considerable bandwidth savings result from using phase functions f(t) which are non-zero for one or two pulse periods. Specifically binary PSK with

$$f(t) = \begin{cases} 0.5\ [1+ \cos(\pi t/T)] & |t| < T \\ 0 & |t| > T \end{cases} \qquad (4\text{-}22)$$

results in a bandwidth which is 15% of that resulting from using constant phase functions ϕ_k; bandwidth being defined in terms of 99% spectral energy containment. A 1.4 dB transmitter power increase on a white Gaussian channel restores the bit error probability to that of conventional binary PSK.

A second result of interest occurs when conventional (rectangular pulse) PSK is filtered prior to transmission by four-pole Butterworth filters to produce a spectral bandwidth equal to that of similarly filtered PSK with a phase functions given by (4-22). For a bit error probability $P=10^{-6}$, the PSK signal with the phase function of (4-22) requires 0.8 dB more than that of the constant phase PSK signal. If the definition of bandwidth is changed to imply a 99.9% spectral energy containment, the PSK with the time-varying phase is inferior by 0.5 dB to constant phase PSK. Since filtering of PSK signals prior to their transmission is equivalent to selecting non-rectangular amplitude pulse shapes g(t) in (4-4) or h(t) in (4-5), Prabhu's results suggest that appropriately selected amplitude pulse shapes with constant phase PSK is not easily improved upon by utilizing time-varying phase functions.

Improvement of the bandwidth properties of FSK by use of non-rectangular pulse shapes g(t) in (4-1) has been demonstrated for MSK signals [S1]. For example, use of a pulse yielding a sinusoidal frequency deviation results in a transmitted binary FSK spectrum which, for levels less than -30 dB below the spectrum's centre frequency falls of much faster than binary MSK. However bit error probabilities for non-conventional FSK pulses are unavailable.

In Chapter III it was suggested that excessive noise might be combatted by using the PAM signal to FM modulate a carrier. Immediately the question arises as to the selection of the PAM pulse shapes or equivalently the selection of the FM pulse shapes in (4-1). The pulse shape which for a given frequency deviation minimizes P(e) is presently unknown. It is clear that for large deviations conventional rectangular pulses are unsuitable, since they would not fill the bandwidth but, in the limit, would yield spectral impulses located either side of the carrier frequency. For large deviation ratios, pulses should be selected to give the FM modulating signal an amplitude probability density appropriate to the desired spectral shape.

IV-8 Comparison of PAM, PSK and FSK

In comparing PAM, PSK and FSK we note first that in all cases the symbol error probability P(e) decreases exponentially with a linear increase in SNR on Gaussian channels. On Rayleigh channels the error probability decrease is linear with an increase in SNR.

For binary signalling the spectral emission constraints in Fig. 2-3 permit a maximum rate of 62 Kb/s for SSB with no excess bandwidth, which rate however would necessarily be reduced considerably to facilitate synchronization and timing. Binary MSK at 36 Kb/s and 2ϕ-DPSK with excess bandwidth $\alpha=1$ at 20 Kb/s both show high bit rates which would be obtainable if the SNR would permit sufficiently low error rates. Four-phase DPSK at 40 Kb/s is also potentially attractive if the SNR is sufficient.

Existing voice bandwidth constraints reduce considerably the maximum bit rates permitted: 32.3 Kb/s for SSB, 13.3 Kb/s for MSK, 8.2 for 2ϕ-DPSK, and 16.2 Kb/s for 4ϕ-DPSK.

To determine the best choice of modulation we could use Salz's [S3] approach, suitably modified to determine for a specified symbol error probability the best type of modulation and the corresponding data rate for a given SNR. We omit this determination, however, for three reasons. First, it is necessary to precisely define the occupied bandwidth of the transmitted signal s(t), which is a difficult and tedious task for FSK. Second, practical considerations particularly for SSB would limit the actual data rate below the maximum obtainable. Finally, the maximum data rates in Figs. 4-3 and 4-4 are well above those currently in use [A2, Z1], which suggests that noise rather than bandwidth constraints may limit the rate. For this latter reason we devote the remainder of this report to considerations of coding for error control.

We note that all error probabilities shown in Fig. 4-3 are for ideal Gaussian and Rayleigh channels, and that the Rayleigh channel results exclude random FM [J1], time delay spreads [J1, A2] and also exclude any improvements available using spatial diversity [J1, M1].

IV-9 References for Chapter 4

[A1] F. Amaroso and J.A. Kivett, "Simplified MSK signalling technique", IEEE Trans. Commun., vol. COM-25, pp. 433-440, April 1977.

[A2] G.A. Arrendondo, J.C. Feggeler and J.I. Smith, "Advanced mobile telephone service: Voice and data transmission", Bell Syst. Tech. J., vol. 58, pp. 97-123, Jan., 1979.

[D1] R. Debuda, "Coherent demodulation of frequency shift keying with low deviation ratio", IEEE Trans. Commun., vol. COM-20, pp. 429-435, June 1972.

[E1] R.E. Eaves and A.H. Levesque, "Probability of block error for very slowly Rayleigh fading in Gaussian noise", IEEE Trans. Commun., vol. COM-25, March, 1977, pp. 368-374.

[G1] G.J. Garrison, "A power spectral density analysis for digital FM", IEEE Trans. Commun., vol. COM-23, pp. 1228-1243, Nov. 1975.

[G2] S.A. Gronemeyer and A.L. McBride, "MSK and offset QPSK modulation", IEEE Trans. Commun., vol. COM-24, pp. 809-820, Aug. 1976.

[G3] L.J. Greenstein, "Spectra of PSK signals with overlapping baseband pulses", IEEE Trans. Commun., vol. COM-25, pp. 523-530, May 1977.

[L1] R.W. Lucky, J. Salz and E.J. Weldon, Jr., Principles of Data Communication. New York: McGraw Hill, 1968.

[M1] Y. Miyagaki, N. Morinaga, and T. Namekana, "Error probability for CPSK signal through m-distributed fading channel", IEEE Trans. Commun., vol. COM-26, pp. 88-100, Jan. 1978.

[P1] V.K. Prabhu, "PSK-type modulation with overlapping baseband pulses", IEEE Trans. Commun. vol. COM-25, pp. 980-990, Sept. 1977.

[R1] H.E. Rowe, Signals and Noise in Communication Systems. Princeton, N.J.: Van Nostrand, 1965.

[R2] H.E. Rowe and V.K. Prabhu, "Power spectrum of digital frequency-modulation signal", Bell Syst. Tech. J., vol. 54, pp. 1095-1125, July-Aug. 1975.

[S1] O. Shimbo, "General formula for power spectra of digital FM signals", Proc. IEEE, vol. 113, pp. 1783-1789, Nov. 1965.

[S2] M.K. Simon, "A generalization of minimum-shift-keying (MSK)-type signalling based on input data symbol pulse shaping", IEEE Trans. Commun., vol. COM-24, pp. 845-856, Aug. 1976.

[S3] J. Salz, "Communications efficiency of certain digital modulation systems", IEEE Trans. Commun., vol. COM-18, pp. 97-102, April 1970.

[T1] H.L. van Trees, Detection, Estimation and Modulation Theory: Part I. New York, N.Y.: Wiley, 1978, ch. 4, pp. 381.

[T2] M. Taub and D.L. Schilling, Principles of Communications Systems. New York, N.Y.: McGraw-Hill, 1971, ch. 4.

[T3] T.J. Tjhung and P.H. Wittke, "Carrier transmission of data in a restricted band", IEEE Trans. Commun., vol. COM-18, pp. 295-304, Aug. 1970.

[W1] J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York, N.Y.: Wiley 1968, ch. 4.

[Z1] L.E. Zegers and C.B. Dekker, "A comparison of digital transmission techniques for standard FM mobile radio sets", IEEE Trans. Commun., vol. COM-25, pp. 364-368, March 1975.

V FORWARD ERROR CORRECTING USING BLOCK CODES

V-1 Introduction

Errors resulting from digital transmission of symbols are inevitable. This chapter deals with the correction of such errors. The following chapter considers other methods of error control including error detection and retransmission.

Block codes and convolutional codes are available for forward error correction (FEC) [L1, G1, P1, L2, L3]. Block coders augment k-bit source digit sequences by the addition of r = n-k check bits which depend on the source digit sequence. The r redundant check bits permit correction of up to t errors in an n-bit block where

$$t = \begin{cases} (d-1)/2 & d \text{ odd} \\ (d/2) - 1 & d \text{ even} \end{cases} \tag{5-1}$$

and d is the minimum Hamming distance between two transmitted codewords. Block codes are easily generated using feedback shift registers. Decoding of selected codes is possible using feedback shift register circuits or majority logic gates. In particular, all single-bit error correcting Hamming codes (a subclass of BCH codes) can be decoded using either approach. Selected multiple-error correcting BCH codes can also be decoded using one or both of these types of decoders.

In contrast to block codes which check each source information bit once, convolutional codes check each source bit N times where N is the constraint span of the code. Some special convolutional codes are decodable using shift register circuits. However longer more powerful codes require elaborate decoders [W1, F1, L3].

Convolutional codes are well suited to transmission of long blocks of data up to hundreds of bits in length, particularly over one-way channels whose primary limitation is noise rather than bandwidth. For transmission of short messages on either one-way or two-way channels, convolutional codes seem to offer no clear advantage over block codes [L1].

Block codes and convolutional codes are available for use on random error channels, burst error channels or on channels exhibiting random and burst error behaviour. Block codes have the advantage that their performance is easily calculated once an appropriate statistical description of the digital channel is available.

Much of the discussion in this and the following chapter applies to both types of codes. However specific performance assessments are presented for block codes only. As explained in Section V-3 the major issue at this time involves determination of those circumstances in which FEC coding is beneficial, rather than detailed comparisons between code types.

V-2 Performance of FEC Block Codes on Memoryless Channels

The most commonly used measure of block code performance is P_e, the probability of a block (or word) error. Although this measure says nothing concerning the distribution of errors, it is a simply defined, mathematically tractable and useful quantitative measure of performance.

Define $P(m,n)$ as the probability of m errors in a block of n bits. For a FEC code which corrects all errors in t or fewer bits and erroneously decodes all words with more than t errors [L2]

$$P_e = \sum_{m=t+1}^{n} P(m,n) \tag{5-2}$$

For a binary symmetric memoryless channel with bit-error probability p,

$$P(m,n) = \left(\begin{smallmatrix} m \\ n \end{smallmatrix}\right) p^m (1-p)^{n-m} \tag{5-3}$$

where

$$\left(\begin{smallmatrix} m \\ n \end{smallmatrix}\right) = \frac{m!}{n!(n-m)!} \tag{5-4}$$

Except in special circumstances, P(m,n) for land mobile channels is <u>not</u> closely approximated by (5-3).

Also of interest is p_b, the probability of error of a decoded information bit, where

$$p_b = P_w P(e) \tag{5-5}$$

In (5-5), P_w denotes the probability of an error in the k information bits, given an error in the n-bit word, and depends on the code. Errors would normally cause the closest (in the Hamming sense) codeword to the one transmitted to be decoded, in which case $P_w \simeq d/n$. For lower density codes such as orthogonal codes $P_w \simeq \frac{1}{2}$ [L2, S1].

The code parameters which determine P_e are n, k and d. For a given channel bit rate, the rate at which information is transmitted increases with k/n. The number of errors which can be corrected for a given value of n increases with d/n. Various bounds on d/n are available [L2, P1] as follows:

$$2^{n-k} \geq \sum_{i=0}^{t} \left(\begin{smallmatrix} n \\ i \end{smallmatrix}\right) \quad \text{(Hamming bound)} \tag{5-6}$$

$$n-k \geq 2(d-1) - \log_2 d \quad \text{(Plotkin bound)} \tag{5-7}$$

$$2^{n-k} < \sum_{i=0}^{d-2} \binom{n}{i} \qquad \text{(Varsharmov-Gilbert-Sacks bound)} \qquad (5\text{-}8)$$

Upper bounds on d for any code are provided by (5-6) and (5-7), while (5-8) provides a lower bound on d which forestalls use of codes with needlessly small d/n values. BCH codes are efficient in that d/n is relatively large for any given k/n value.

Fig. 5-1 illustrates code comparisons based on (5-2), (5-3), (5-4) and (5-5). Codes considered include the (31,26) single-error correcting Hamming code and the (31,16) BCH code, which is three-step orthogonalizable and can therefore be decoded using majority logic gates [L1, M2]. These codes were selected not only for there decodability, but also because they are of a length suitable for dispatch messages, because they lie between the shorter 15-bit and longer 63- or 127- bit BCH codes, and because actual field test results are available for comparison purposes (see Section IV-5). Channels considered include Gaussian noise channels, (which applies to stationary vehicles) and Rayleigh fading channels (which applies to some degree to rapidly moving vehicles). Incoherent FSK modulation was assumed, with received SNR ρ_0 and $\overline{\rho_0}$ for which channel bit error probability p is as follows:

$$p = 0.5 \exp(-\rho_0/2) \qquad \text{(Gaussian channel)} \qquad (5\text{-}9)$$

$$p = (2+\overline{\rho_0})^{-1} \qquad \text{(Rayleigh channel)} \qquad (5\text{-}10)$$

Comparisons between coded and uncoded transmission should be on the basis of constant information bit rate, rather than on the basis of constant channel bit rate. Decoded bit error probability P_b can be reduced simply by reducing the channel bit rate; moreover the associated cost is less than the cost of coding which, therefore is of interest only if the necessarily

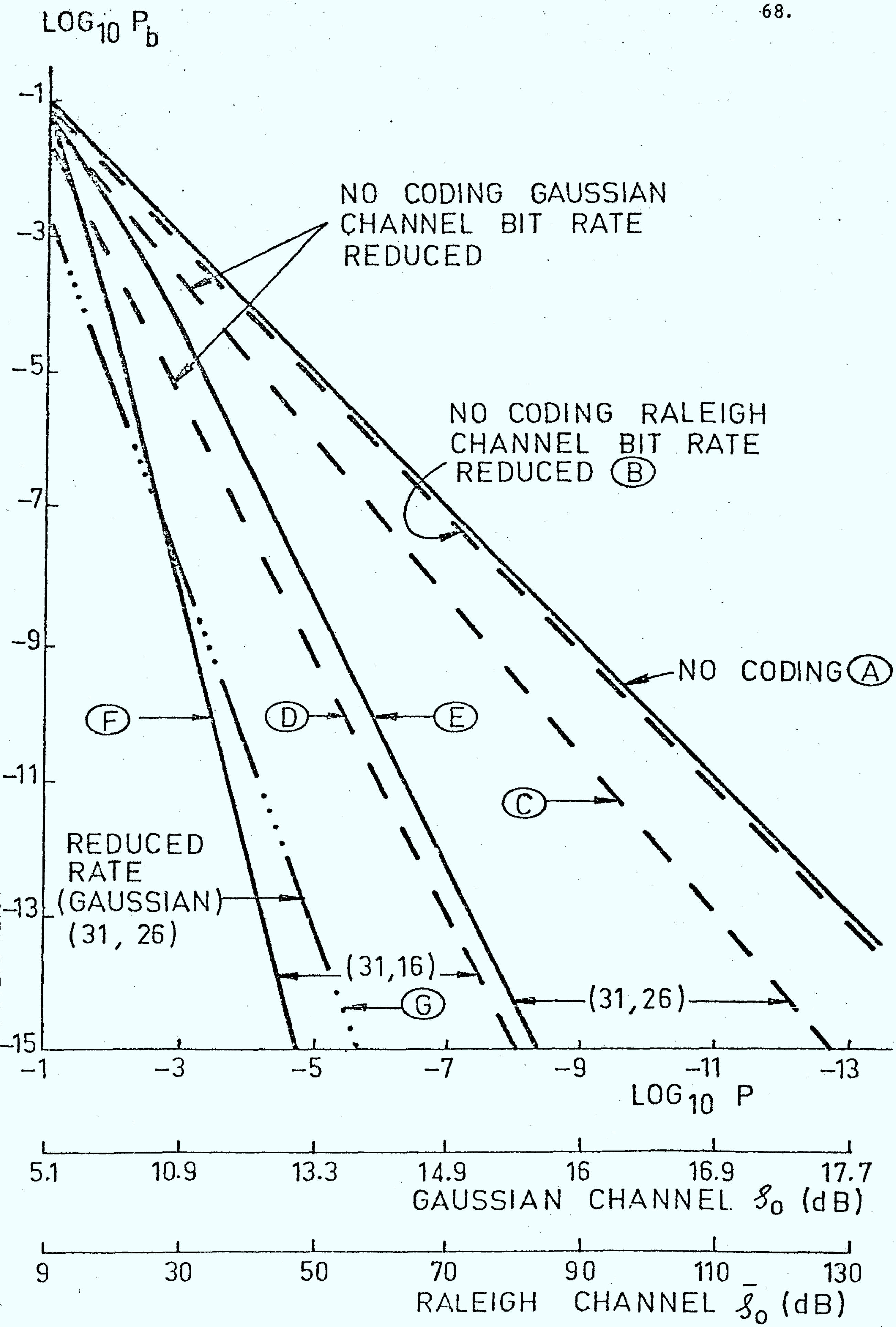

Fig.5-1 *Decoded bit error probability* P_b *for FEC using (31,26) and (31,16) block codes; p denotes channel bit error probability;* ρ_0 *and* $\bar{\rho}_0$ *denote signal-to-noise ratio for Gaussian and Rayleigh channels, respectively. The dotted lines show* P_b *with no coding and a reduced transmission rate.*

increased channel bit error probability is more than offset by the error correcting capability of the code.

Evaluation of the (31, 26) code is obtained by comparing curves C and E in Fig. 5-1, for the Gaussian channel and curves B and E for the Rayleigh channel, and not curves A and E. For Rayleigh channels large coding gains are available; 26 dB for a (31, 26) code when $p_b=10^{-7}$. The corresponding gain is 1.7 dB for the Gaussian channel.

At low decoded bit error probabilities, larger gains are available for a (31, 16) code; comparisons of curves B and F for the Rayleigh channel and curves D and F for the Gaussian channel shows gains of 40 dB and 2 dB, respectively.

Of interest is a (31, 26) code at a channel bit rate by the factor 16/26, rather than using a (31, 16) code; the information bit rate in the two cases is the same. Comparison of curves F and G shows that the (31, 16) code is better for $p_b < 10^{-7}$, but for $p_b > 10^{-7}$ the (31, 26) code with the rate reduction is better. For Rayleigh channels, a rate-reduced (31, 26) code lies 2.1 dB to the left of curve E and is not shown, since in this case the (31, 16) code is best for $p_b \lesssim 10^{-1}$; the coding gain in comparison to the rate-reduced (31, 26) code is 16 dB at $p_b=10^{-7}$.

For any code a threshold $\tilde{p}_b$ exists above which direct transmission is preferable to coding. For the (31, 26) code, $\tilde{p}_b \simeq 3 \times 10^{-2}$ in Fig. 5-1, for the (31, 16) code $\tilde{p}_b \simeq 5 \times 10^{-3}$. For sufficiently low values of p_b coding will always be beneficial on a memoryless channel, since for small p

$$P_e \simeq \frac{n!}{(t+1)!\,[n-(t+1)]!} \; p^{t+1} \qquad (5\text{-}11)$$

Use of other codes or other modulation schemes yields similar types of results. Relatively small coding gains are available for Gaussian channels. Very large gains are available on Rayleigh channels, the reason being that errors which occur during deep fades, if corrected tend to nullify the effects of fades, in which case the time-diversity effects of FEC cause the Rayleigh channel to perform like a Gaussian channel.

Fig. 5-2 presents results for a 7-error correcting (31, 6) code which is 3-step majority logic decodable [L1]. Fig. 5-3 presents results for a (15, 11) single error correcting Hamming code and a 1-step majority logic decodable (15, 7) double-error correcting BCH code.

The 47 dB coding gain at $p_b=10^{-7}$ for the (31, 6) code is in marked contrast to the -0.2 dB gain for the Gaussian channel. The code would be attractive for a mobile channel, to combat Rayleigh fading during rapid vehicle motion.

The coding gains obtainable using the (15, 11) and (15, 7) code are similar to those obtainable using a codes with n=31 and similar code rates.

Use of other modulation schemes yields results similar to those obtained here; Gaussian channels show a exponential decrease with ρ_0 and therefore small coding gains. Rayleigh channels show a linear decrease with $\overline{\rho_0}$ and large potential coding gains, which are of interest since the power required for Rayleigh channels is much larger than that for Gaussian channels, as Figs. 5-1, 5-2, and 5-3 demonstrate.

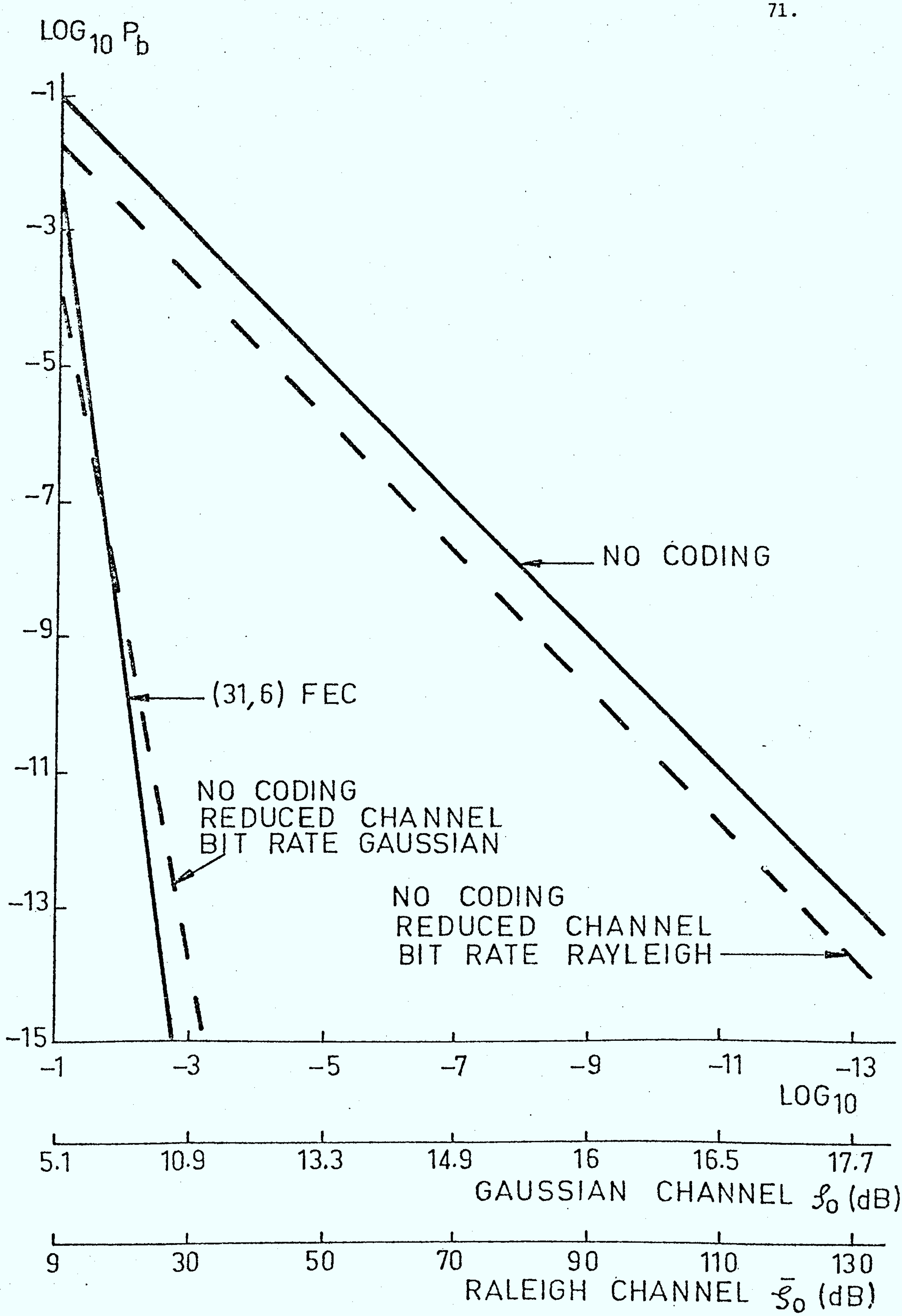

Fig. 5-2 P_b for a (31,6) 7-error correcting BCH code

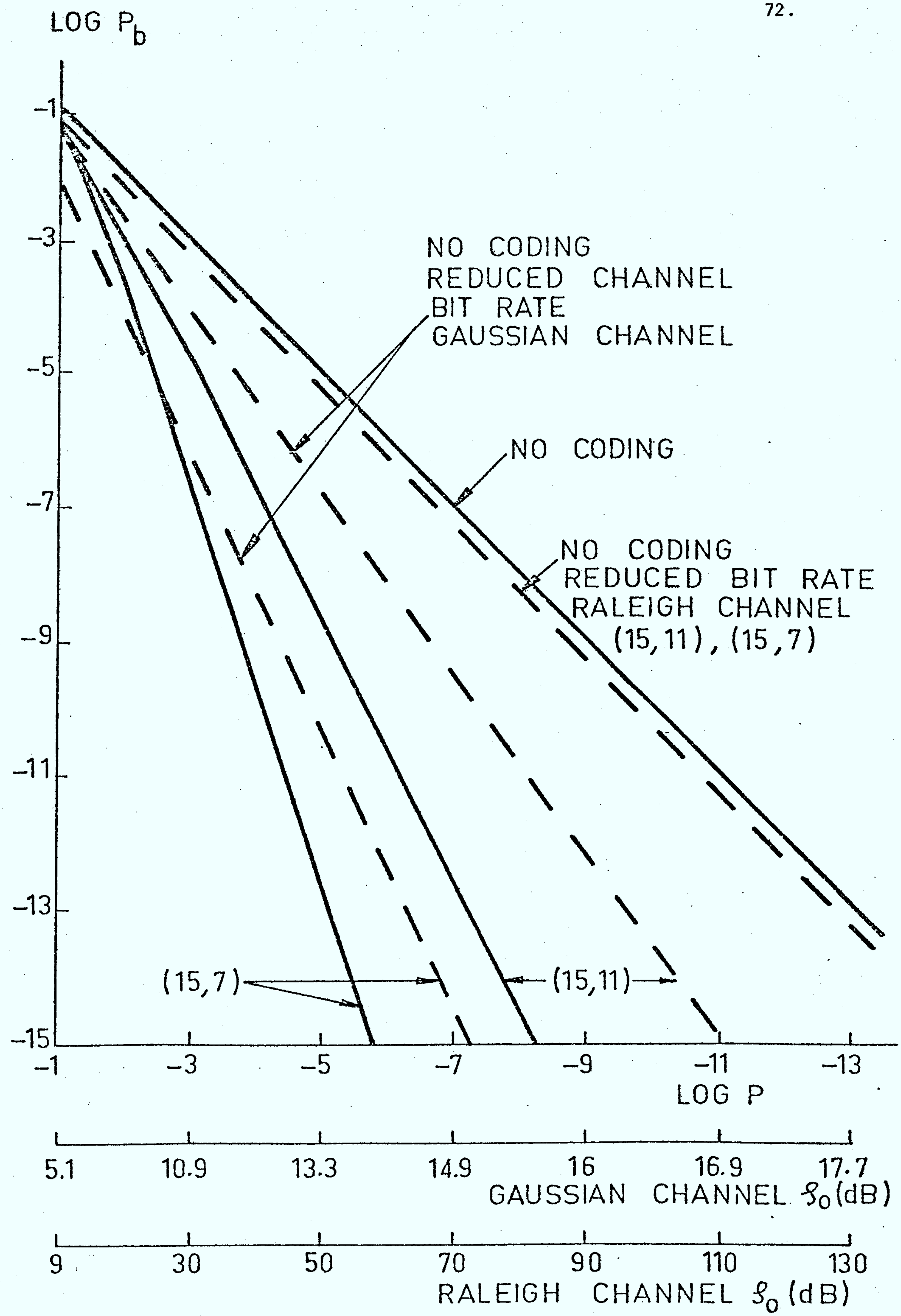

Fig. 5-3 P_b for (15,11) and (15,7) block codes

V-3 Applications of Memoryless Channel Results to Land Mobile Channels

The memoryless channel results in the previous section seem applicable to land mobile channels under the following circumstances:

1. The mobile is stationary during transmission of a block of data, in which case the Gaussian memoryless channel assumption for P(m,n) is reasonalbe.

2. The transmitted bits are interleaved [L1, C1], in which case the Rayleigh memoryless channel approximation is reasonable provided the vehicle traverses many carrier wavelengths during an interleaving period.

If a mobile traverses many wavelengths during the transmission of a data block, error effects may resemble those of a memoryless Rayleigh channel, even in the absence of interleaving. These comments together with the results of the previous section suggest FEC as being reasonably effective in combatting errors when the vehicle is stationary or when the vehicle moves many wavelengths during transmission of a single data block, provided the average signal level is high enough that positive coding gains result.

One cannot expect FEC codes to be effective under all circumstances. To protect against errors when vehicles are located in deep fade zones would require an information bit rate which would be prohibitively low under normal operating conditions.

The above conclusions are supported by Chase [C1] who showed that when FEC codes correct errors in data blocks transmitted over Rayleigh channels at a rate which causes the received SNR to remain constant over the block duration, no improvement is achievable in decoded bit error probability. The same study showed that codes which interleave bits among data blocks which themselves are independent in received SNR yield considerably improved performance. For example, at decoded bit error probabilities of 10^{-3}, coding gains in excess of 10 dB resulted, and these gains increased as decoded bit error probabilities decreased.

Chase's results support the conclusion that to combat slow Rayleigh fading, the bits in each data block must be transmitted over a time span sufficient to ensure that most of these bits will be received correctly. In a mobile radio environment, interleaving would not always be effective since vehicles may remain stationary indefinitely. The fact that many transmissions consist of short infrequent blocks of control information requiring immediate and accurate transmission also frustrates use of interleaving which in any case requires transmitter and receiver buffer storage.

The best way to ensure error control under all operating conditions is to have available for use an ARQ strategy of the type considered in Chapter 6.

V-4 Models of P(m,n) for Land Mobile Radio Channels

Precise comparison of various coded systems with each other and with uncoded systems in non-random error environments requires knowledge of P(m,n) in (5-2). For channels with memory there are three basic approaches to obtaining P(m,n):

1. Calculate bit-error dependencies by using the behaviour of the fluctions in the received signal level.

2. Develop a descriptive mathematical model for the digital channel in Fig. 2-2.

3. Measure P(m,n).

The first approach, which is the most direct seems not to have yielded useful results, even though correlation functions for the received signal envelopes have been obtained for vehicles moving at constant velocity [J1]. The fact that vehicle velocity is variable, and that vehicles often transmit and receive when stationary would only increase the difficulty of utilization the first approach.

The second approach has led to the development of progressively more complex finite- and semi-infinite state models of channels [G2, L2, K1, A1, E1, E2, C2, F2, G3] the simplest example of which is Gilbert's two-state model, depicted in Fig. 5-4. The Gilbert channel is either in the good (G) state where the bit-error probability equals zero, or in the bad state (B) where the bit-error probability is 1-h. State transition probabilities are as indicated in Fig. 5-4, from which P(m,n) is obtained recursively [G2, K1]. In order to allow for random error as well as burst error behaviour Elliot [E1] modified Gilbert's model to allow the error probability in the good state to be non-zero. The result is a more flexible model and more involved computations of P(m,n).

Expansion of the number of states in the Gilbert model with each state having non-zero error probability has been advocated [K1]. A five-parameter model was found to give results which agreed reasonably well with these observed for a Tropospheric scatter channel [C2].

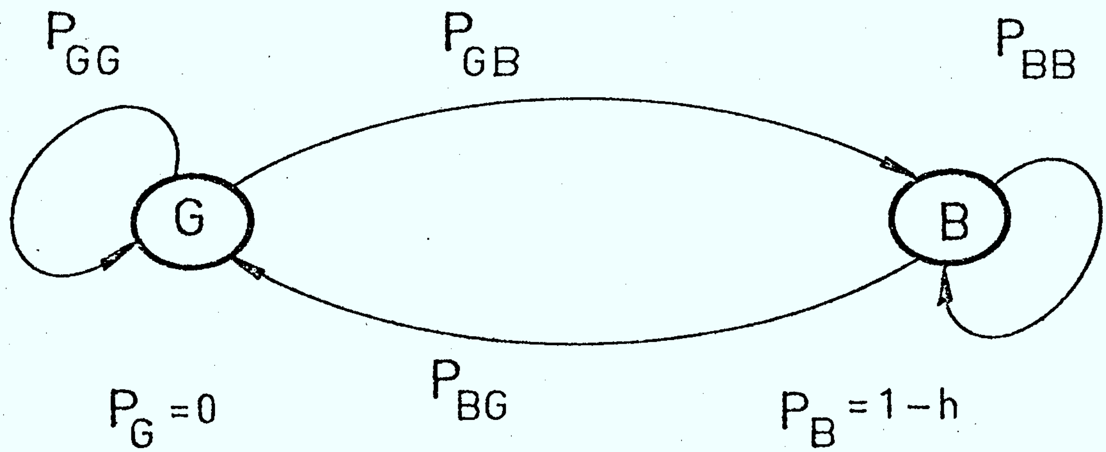

Fig. 5-4 Gilbert burst noise channel model.

The importance of avoiding models which do not suitably represent actual channel behaviour becomes apparent in comparing results from two separate studies assessing the utility of FEC codes in burst noise environments. In one study [F3], the Gilbert two state model was used to simulate a burst noise channel; it was shown that FEC did not reduce the information bit error probability for the channel at a fixed information bit transmission rate. In Mabey's [M1] study, measured values of $P(m,n)$ for an actual mobile channel were used to obtain the opposite result. Although the Gilbert two state model simulates bursty behaviour of mobile channels in a generally appealing way, it does not model the channel sufficiently well to yield reliable conclusions regarding FEC code performance.

A state model suitable for land mobile channels seems unavailable. To simulate the fading process illustrated in Fig. 2-1 it would seem necessary to include in any state model a periodic variation in state transition probabilities, with the period itself a random variable. The absence of a suitable model requires that confirmation of expected FEC code performance be obtained from either measured $P(m,n)$ data or actual field tests, particularly where vehicle velocity causes the fading rate to fall between the extremes approximated by the memoryless Gaussian and memoryless Rayleigh channel.

V-5 Results for FEC Codes Based on Measured $P(m,n)$ Distributions

Evaluations of block codes based on actual measurements of $P(\geq m,n)$, the probability that the number of errors in an n-bit block of bits exceeds m appear in a recent article by Mabey [M1]. The measured $P(\geq m,n)$ distributions were obtained at 462 MHz in London, England during transmission from a base station to a moving vehicle using data rates of 1200 b/s and 4800 b/s.

The following conclusions emerge from the Mabey's results:

1. $P(\bar{>}m,n)$ falls of more slowly as m increases than for a channel with independent errors.

2. Random error correcting codes are as good as or better than burst error correcting codes. Mabey states [M1]:

"The restricted usefulness of burst codes on the mobile radio data channel, which is generally described as bursty, deserves some explanation. Burst correcting codes perform well on the classic bursty channel [4] where error bursts are never longer than B bits and are separated by error-free guard spaces which are never shorter than G bits, if the guard spaces are long enough for most bursts to be corrected. In mobile radio, the error bursts are caused by fades in the received signal as the vehicle moves through the standing wave pattern created by multipath propagation. However the following factors contribute to a channel which is better described as "messy" rather than bursty. There is variation in fade spacing, duration, and depth; a shallow fade of even quite long duration may produce errors of only low density which may be corrected more reliably by a random error correcting code. The mean received signal level fluctuates; at a low mean, fades will tend to be long with only short gaps, whereas at a high mean, fades will be short with long gaps. Also, vehicle speed varies, and in addition errors caused by ignition interference are scattered through nonfaded and faded periods."

3. The block error probability P_e does not change much with block length n; in independent error channels, P_e decreases as n increases at data rates below channel capacity [G1].

4. For a given P_e, increased information throughout can occur if the channel bit rate is increased and the code rate is reduced.

5. Interleaving of bits prior to transmission yields considerably improved performance over that which results in the absence of interleaving. Specifically, at $P_e=10^{-3}$, coding gains for a(31,16) triple error-correcting BCH code of approximately 4.5 dB resulted

when thirty-two 31-bit blocks were interleaved. An additional 2 dB would result on a random-error channel with the same signal-to-noise ratio.

Mabey's results do not permit comparisons of the kind shown in Figs. 5-1, 5-2 and 5-3, since the channel bit rate was fixed at one of two values. However conclusion 4 above is supported by the following comparisons.

1. A (31,21) double error correcting code at 1200 b/s and a (31,16) triple error correcting code at 4800 b/s give nearly equal values of P_e at a equal SNR values over the observed range $10^{-1} \stackrel{\sim}{<} P_e \stackrel{\sim}{<} 10^{-4}$. The corresponding information throughput rates are 813 and 2477 information b/s which is a 1:3 ratio.

2. The (31,16) code at 1200 b/s and a(31,6) 7-error correcting code at 4800 b/s give nearly equal values of P_e over the above-mentioned range, at up to 8 dB lower SNR values (for $P_e \simeq 10^{-4}$) than used to obtain the results in the above paragraph. The information bit rates in this latter case were 619 and 929 b/s, a ratio of 1:1.5.

Comparison of the 813 b/s of the (31,21) code at 1200 b/s with the 929 b/s of the (31,6) code at 4800 b/s shows a coding gain in of the latter code in excess of 8 dB at $P_e \simeq 10^{-4}$. Unfortunately, Mabey does not present results for uncoded transmission. However the 8 dB coding gain noted, together with Mabey's other results as well as our results in Figs. 5-1, 5-2 and 5-3 suggest that coding gains available without interleaving during vehicle motion lie between those available for

Gaussian and Rayleigh channels, and may approach 10 dB, or more. Additional actual results or a viable channel model are needed for definitive quantative conclusions.

Finally, we note the use of a rate $\frac{1}{2}$ type B-1 convolutional mobile channels at a channel bit rate of 4800 b/s [C3]. Qualitative performance results are presented.

V-6 References for Chapter 5

[A1] J.P. Adoul, B.D. Fritchman and L.N. Kanal, "A critical statistics for channels with memory", IEEE Trans. Inform. Theory, vol. IT-18, pp. 133-141, April 1972.

[C1] D. Chase and L.J. Weng, "Multiple-burst correcting techniques for slowly fading channels", IEEE Trans. Inform. Theory, vol. IT-22, pp. 505-513, Sept. 1976.

[C2] R.T. Chien, A.H. Haddad, B. Goldberg and E. Moyes, "An analytic error model for real channels", in Conf. Rec., IEEE Int. Conf. Commun., Philadelphia, PA., 1972, pp. (15-7)-(15-12), also in [G3].

[C3] R.T. Chien, V.E. Clayton, P.E. Boudrean and R.R. Locke, "Error correction in a radio-based data communications system", IEEE Trans. Commun., vol. COM-23, pp. 458-462, April 1975.

[E1] E.O. Elliot, "Estimates of error rates for codes on burst noise channels", Bell Syst. Tech. J., vol. 42, pp. 1977-1997, Sept. 1963.

[E2] E.O. Elliot, "A model of the switched telephone network for data communications", Bell Syst. Tech. J., vol. 44, pp. 89-110, Jan. 1965.

[F1] G.D. Forney, "The Viterbi algorithm", Proc. IEEE, vol. 61, pp. 268-278, March 1973.

[F2] B.D. Frichtman, "A binary channel characterization using partitioned Markov chains", vol. IT-13, pp. 221-227, April 1967.

[F3] C. Fujiwara, M. Kasahara, K. Yamashita, and T. Namekawa , "Evaluations of error control techniques in both independent-error and dependent-error channels", IEEE Trans. Commun., vol. COM-26, pp. 785-793, June 1978.

[G1] R.G. Gallager, Information Theory and Reliable Communication. New York, N.Y.: Wiley, 1968.

[G2] E.N. Gilbert, "Capacity of burst-noise channels", Bell Syst. Tech. J., vol. 39, pp. 1253-1265, Sept. 1960.

[G3] B. Goldberg, Ed., Communication Channels: Characterization and Behaviour. New York, N.Y., IEEE Press, 1975; also in IEEE Trans. Commun., vol. COM-23, pp. 1189-1197, Nov. 1975.

[J1] W.C. Jakes, Jr., Ed., Microwave Mobile Communications. New York, N.Y.: Wiley, 1974.

[K1] L.N. Kanal and A.R.K. Sastry, "Models for channels with memory and their applications to error control", Proc. IEEE, vol. 66, pp. 724-744, July 1978.

[L1] S. Lin, An Introduction to Error-Correcting Codes. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

[L2] R.W. Lucky, J. Salz and E.J. Weldon, Principles of Data Communication. New York, N.Y.: McGraw-Hill, 1968.

[L3] W.C. Lindsay and M.K. Simon, Telecommunication Systems Engineering. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

[M1] P.J. Mabey, "Mobile radio data transmission-coding for error control", IEEE Trans. Veh. Technol., vol. VT-27, pp. 99-110, Aug. 1978.

[M2] J.L. Massey, Threshold Decoding. Cambridge, MA.: MIT Press, 1963.

[P1] W.W. Peterson and E.J. Weldon, Jr., Error-Correcting Codes, 2nd Ed. Cambridge, MA.: MIT Press, 1972.

[S1] J.J. Stiffler, Theory of Synchronous Communication. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[T1] S. Tsai, "Markov characterization of the HF channel", IEEE Trans. Commun. Technol., vol. COM-17, pp. 24-32, Feb. 1969.

[W1] J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York, N.Y.: Wiley, 1965.

VI ERROR CONTROL USING ERROR DETECTION AND RETRANSMISSION

VI-1 Error Detection using Block Codes

Block codes may be used for error detection instead of or as well as for error correction. When error detection is employed, the probability of a detectable block error P_d is of interest, along with the probability of a block decoding error P_e and the probability P_c that the block is decoded correctly.

In the absence of any error correction [L1]:

$$P_c = P(m=0,\ n) \tag{6-1}$$

$$P_e \simeq 2^{-(n-k)} \sum_{m=d}^{n} P(m,n) \tag{6-2}$$

$$P_d = 1-P_c-P_e \tag{6-3}$$

where P(m,n), n, k, and d are defined in Chapter 5.

A block code which is used for error detection only detects all errors except those which change the transmitted codeword into another codeword. Since the ratio of the number of codewords to the total number of words is $2^k/2^n$ most errors are detected. The factor 2^{n-k} in (6-2) together with the lower limit d rather than $\simeq d/2$ for FEC makes P_e much lower for error detection than for FEC. However, the converse is true of P_c since one or more errors in a block results in retransmission rather than decoding. As the channel degrades the number of retransmissions increases, and the actual information throughput adjusts automatically to match channel quality.

Fig. 6-1 shows the decoded bit error probability P_b for (31,26), (31,16) and (31,6) BCH codes. Also shown for comparison purposes are the

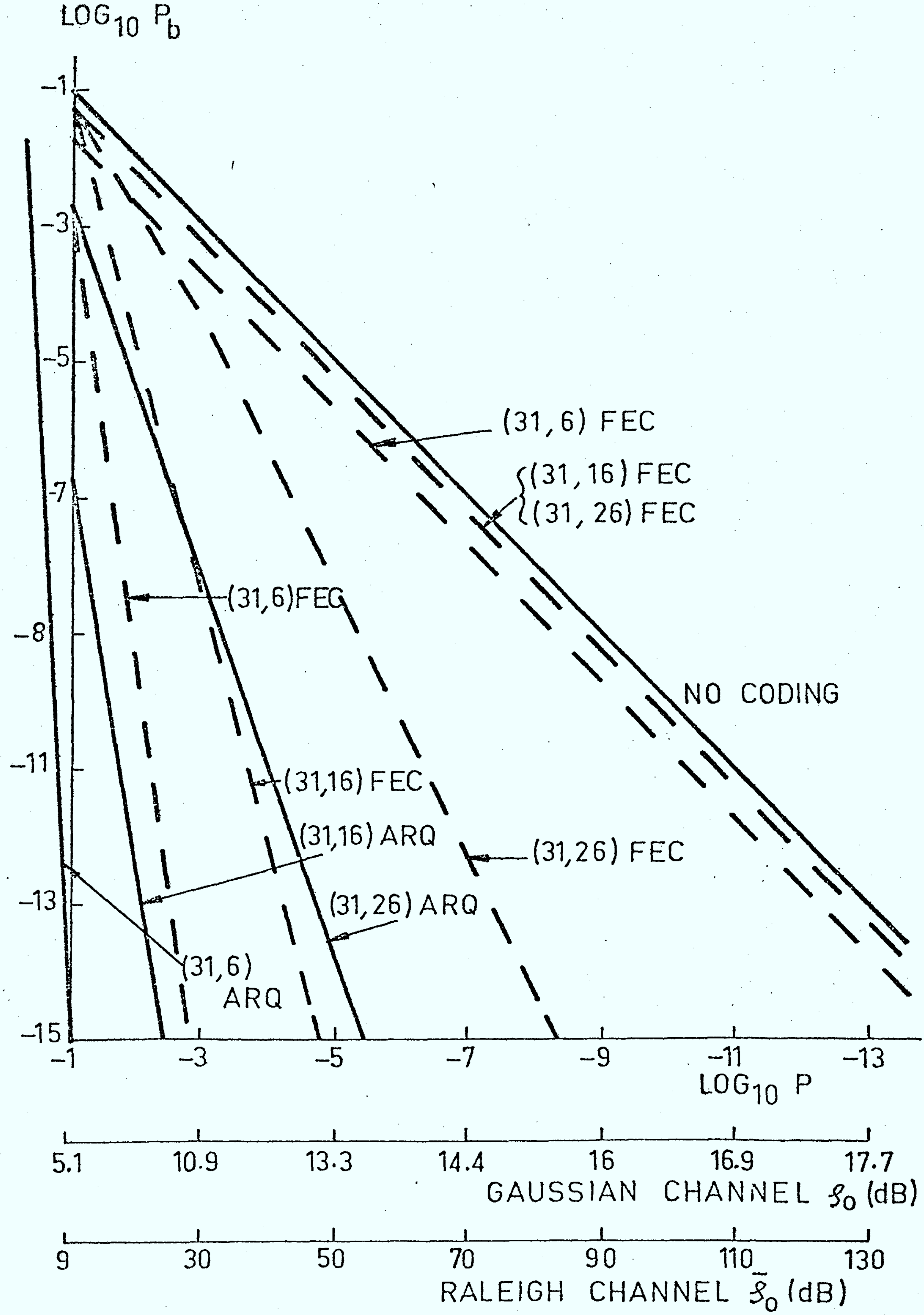

Fig. 6-1 *Decoded bit error probability* P_b *vs. channel bit error probability p for ARQ. Also shown (dotted lines) is* P_b *for FEC.*

the decoded bit error probabilities obtained in Chapter 5 when these same codes are used for forward error correction. Figs. 6-2 and 6-3 show coding gains obtainable for FEC and ARQ. On Gaussian channels at $p_b=10^{-7}$, for example, a (31,26) code used for ARQ rather than FEC yields a gain of 2.8 dB, and a (31,16) code used with ARQ rather than FEC yields a gain of 5.3 dB. These gains are in addition to the gains of 1.7 dB and 2 dB, respectively for FEC relative to direct transmission. A (31,6) code at $p_b=10^{-7}$ show a 0.2 dB loss for FEC and a 4 dB gain for ARQ.

Large coding gains are available for Rayleigh channels, particularly at low values of p_b. For example, at $p_b=10^{-7}$ coding gains for ARQ relative to FEC for (31,26), (31,16) and (31,6) codes equal 17 dB, 18 dB and 10 dB, respectively, and these gains are in addition to the gains of 26 dB, 40 dB, and 47 dB for FEC relative to direct transmission with no coding.

It is clear from Figs. 6-1, 6-2 and 6-3 that coding gains available depend on the specific code used, the channel type and the value of p_b. For a Rayleigh channel Fig. 6-3 shows that for $p_b \gtrsim 10^{-8}$ a $(31,16)_{ARQ}$ format yields the highest coding gain; for $p_b \lesssim 10^{-8}$ a $(31,6)_{ARQ}$ format is best. For a Gaussian channel Fig. 6-2 shows a $(31,16)_{ARQ}$ format as best for $10^{-14} \lesssim p_b \lesssim 10^{-3}$.

Use of Figs. 6-1, 6-2 and 6-3 to assess the benefits of using error detection and retransmission on mobile channels is similar to the procedure proposed in Chapter 5. For stationary vehicles the Gaussian curves are applicable, while the Rayleigh curves apply to moving vehicles when the errors are independent.

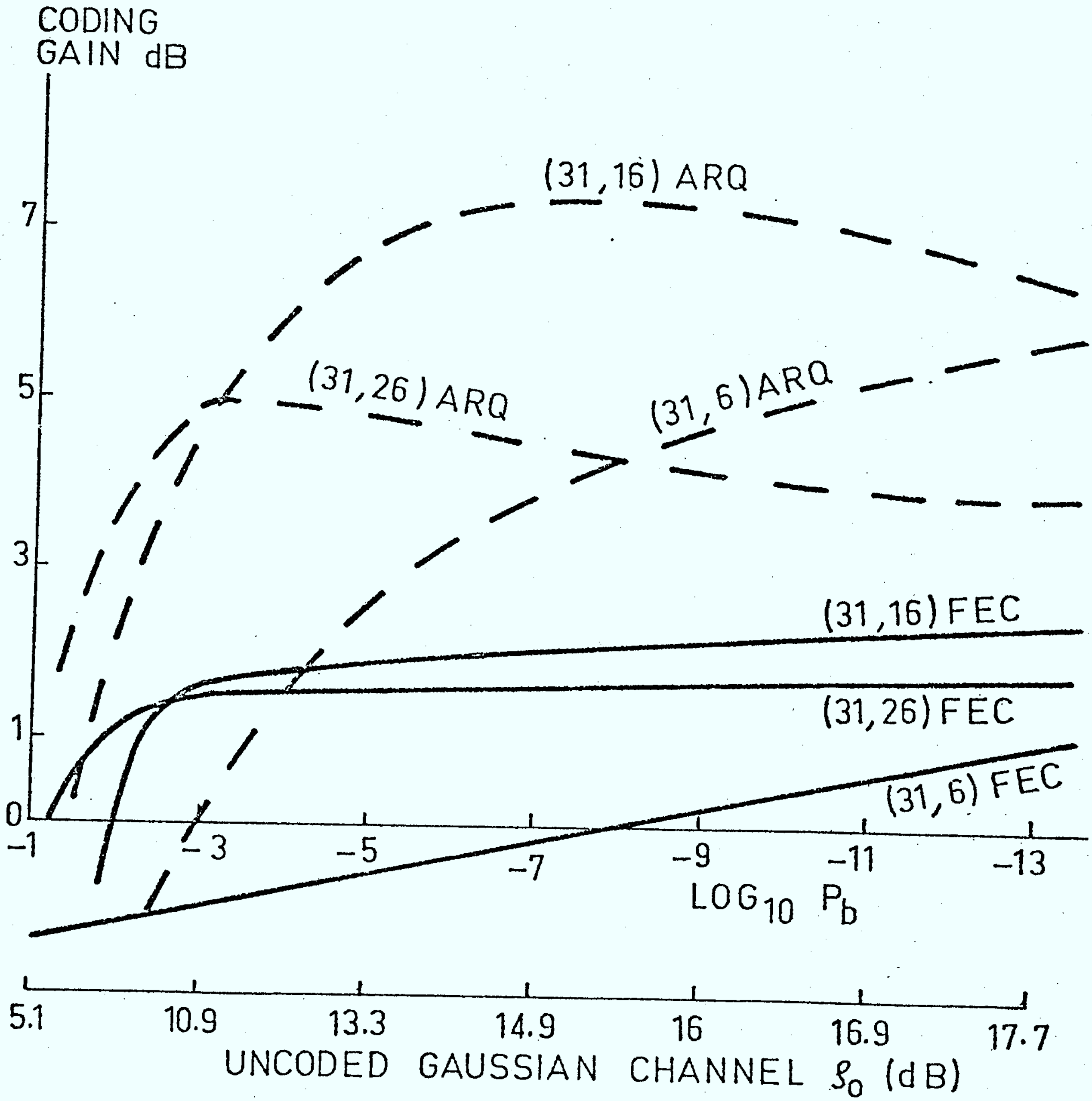

Fig. 6-2 Coding gains for Gaussian Channels

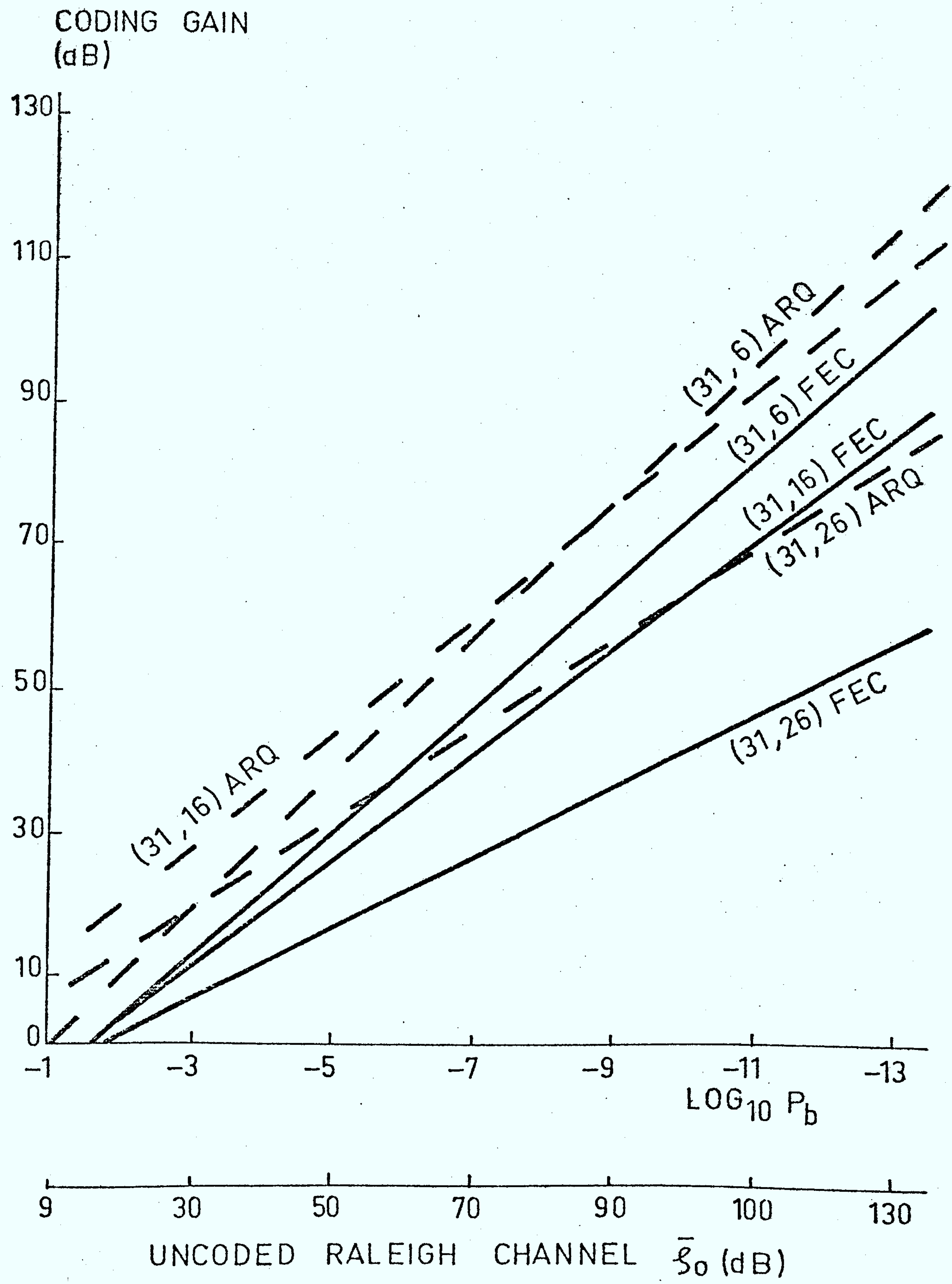

Fig. 6-3 Coding gains for Rayleigh channels

Results based on field measurements were obtained by Mabey [M1] under conditions described in Section 5-5. His results confirm the availability of large additional coding gains for ARQ relative to FEC. Specifically use of a (15,7) BCH code which can convert all one and two-bit errors in the FEC mode or detect all errors up to and including four bits in the ARQ mode at 1.2 Kb/s uses 20 dB less transmitted power at $P_e=10^{-4}$ with ARQ than with FEC. With error detection at $P_e=10^{-5}$ a (31,16) BCH code at 4800 b/s shows a 33 dB coding gain for ARQ relative to FEC. As noted in Section 5-5, the (31,16) BCH code with FEC at 4800 b/s yielded P_e values virtually identical to the (31,21) BCH code with FEC at 1200 b/s. It follows that if the (31,16) code is used in the ARQ mode instead of the (31,21) in the FEC mode at $P_e \approx 10^{-5}$ a coding gain of 33 dB results, and a 3-fold improvement in bit rate also occurs.

At $P_e=10^{-5}$ Mabey's [M1] results show $P_d \approx 10^{-4}$ for the $(31,16)_{ARQ}$ format at 4800 b/s, which implies virtually no loss in throughput efficiency relative to FEC (see section VI-4).

Comparing the coding gain of more than 33 dB (coding gain normally implies equal information transmission rates) with the 44 dB shown for the $(31,16)_{ARQ}$ format in Fig. 6-3 suggests that the gains associated with independent bit errors on Rayleigh channels may be indicative of those obtainable when the vehicle is moving.

The considerably higher coding gains for ARQ relative to FEC tend to support the statement by Burton and Sullivan [B1] which applies to mobile channels as well as conventional telephone channels:

"A significant portion of this paper has been spent justifying what is essentially a *fait accompli*, namely, the widespread use of ARQ techniques in computer communications. While we have seen that ARQ systems currently in use may not be satisfactory in future applications, the solution lies not in FEC systems, but rather in a better ARQ technique. The conclusion drawn in this paper may leave coding theorists somewhat upset, and indeed the authors have been frequently confronted with arguments which have the following general tone: Shannon's capacity theorem states that for any channel there is a quantity called the channel capacity. As long as the information rate through the channel does not exceed this capacity, it is theoretically possible, with the use of forward error-correcting codes alone, to achieve an error probability as low as one desires. Therefore, there should, in principle, be no need for ARQ systems.

Our answer lies in the difference between principle and practice. For certain idealized models of telephone channels, the channel capacity has indeed been calculated. To our knowledge, however, no one knows how to determine the capacity of a real telephone channel, warts and all. Furthermore, while the FEC techniques discussed in this paper appear promising, there is a very *ad hoc* tone about them; they are based more on coarse intuitive notions of what the channel is like than on any fundamental understanding. In short, then, to a coding theorist, the telephone channel is a very formidable problem, and there is little hope that forward-error-correcting codes capable of performing to the satisfaction of most computer-communications users can be developed in the near future. In the meantime, ARQ systems are likely to continue filling the need simply and efficiently."

VI-2 Combined Error Detection and Correction Using Block Codes

Block codes with minimum distance $d \geq 5$ can be used to correct errors in t or fewer bits where t is bounded as in (5-1) and to detect all errors involving δ bits where $t<\delta\leq d-1-2t$. In this case [L1]

$$P_c = \sum_{m=0}^{t} P(m,n) \qquad (6\text{-}4)$$

$$P_e \simeq \frac{\sum_{i=0}^{t} \binom{n}{i}}{2^{n-k}} \sum_{m=d-t}^{n} P(m,n) \qquad (6\text{-}5)$$

$$P_d = 1 - P_c - P_e \qquad (6\text{-}6)$$

Combined error correction and detection is particularly useful on satellite channels where the transmission time from source to receiver and back again exceeds 500 ms [S1, S2]. Inclusion of some error correction improves throughput efficiency over that which results from detection only situations requiring frequent retransmissions. Retention of some error detection capability provides for higher decoded bit error probabilities than for FEC only.

Selection of t to optimize throughput efficiency has been considered [F1] for both random and idealized Gilbert burst error channels. As explained in Chapter 5, the results are not directly applicable to mobile channels because of the inadequacy of the channel models used in the study [F1].

VI-3 Alternative ARQ Protocols

When a received data block contains detectable errors, retransmission of the block ensues. Various retransmission protocols are available, as follows [B2, B3, S1, F1]:

1. Send-and-wait ARQ: Following transmission of a block, the sending terminal waits for either a positive acknowledgement (ACK) or a negative achknowledge (NACK) from the receiving terminal before sending the next block or retransmitting the same block.

2. Continuous (Go-Back-N) ARQ: The transmitter continues to send blocks until it receives a NACK, at which point it retransmits

the current block as well as the previously transmitted N-1 blocks. A transmitter buffer is needed to save N blocks for possible retransmission. Following detection of one or more errors in a block, the receiver ignores all subsequent blocks prior to receiving the retransmitted block. Transmission delays and receiver decoding delays require $N \geq 2$.

3. Selective ARQ: The transmitter operates continuously, but retransmits only those blocks in which errors have been detected.

Stop-and-wait ARQ is advantageous because of its implementation simplicity and is particularly suited to half-duplex channels [B1]. However conventional channel sharing is based on line switching [K1, S3], and considerable idle time occurs as explained in Section VI-4 with the result that overall system information throughput is lower than that of the other ARQ protocols. Selective repeat systems have the highest overall system throughput but would normally require numbering of the data blocks to facilitate identification for retransmission. The go-back-N scheme would also require block numbering for round-trip transmission times with high variability as for queueing, or contention-accessing of broadcast channels [T1, K1, A1]. Variations of each of the above schemes are possible [K1, S3, S4]. ACK's only may delivered to the transmitter, with retransmission occurring after a time-out period. This variation, which may be used with any of the three schemes above, avoids the problems which arise in incorrect decoding of NACK's. The disadvantages are that the time-out period must be long enough to accommodate round-trip delays, otherwise unnecessary retransmissions may occur. Long time-out periods imply reduced overall system information throughput for the stop-and-wait protocol and increased transmitter buffer storage for the other two protocols.

Another variation results if NACK's only are sent to the transmitter. The advantage here is that acknowledgement traffic is considerably reduced, since NACK's would normally occur much less frequently than ACK's. In a situation in which data traffic flows regularly in each direction, ACK's or NACK's can be embedded in data blocks. However, when data flow is predominately one-way, acknowledgement information would have to be sent in its own block which would contain the usual overheads. There is an advantage to sending acknowledgement information separately, however; when queueing delays occur acknowledgement blocks can be given priority over regular data blocks, thereby reducing retransmission delays [S3, P1].

VI-4 Throughput and Delay for ARQ Protocols

The throughput efficiency η for a digital communication link may be defined as the number of the information bits decoded relative to the total number of bits transmitted. Thus, for an (n,k) FEC code $\eta = k/n$. For an (n,k) code used in ARQ situations [S1]

$$\eta = \frac{k}{n} \frac{1}{1+N[P_d/(1-P_d)]} \qquad (6\text{-}7)$$

where N denotes the value applicable to the go-back-N protocol.

Equation (6-7) also applies to both stop-and-wait and selective ARQ in which cases N=1 and

$$\eta = (k/n)(1-P_d) \qquad (6\text{-}8a)$$

Normally $P_e << P_d$, $1-P_d \simeq P_c$ and, for N=1

$$\eta \simeq P_c\ k/n \qquad (6\text{-}8b)$$

The degradation in throughput effeciency of ARQ relative to FEC is obtained directly from (6-1). For N=1, Fig. 6-4 shows this degredation $\tilde{\eta}$ for a block code of length n=31 and indicates that for $p \sim 10^{-3}$, $P_c \gtrsim 0.9$ in which case the following $(\tilde{\eta},N)$ pairs of values result: (0.9, 1); (0.82, 2); (0.69, 4). The effect of N in reducing η is relatively large for low values of P_d.

The delay D from the time that a message is presented to the transmitter until it is processed by the receiver depends on various component delays including those due to queueing for channel access [S3, K1], transmission of bits over the channel, propagation, decoding, and modem start-up or turn-around [D1]. Let T_1 and T_2 denote the one-way and round-trip delays, respectively (often $T_2 \simeq 2T_1$); then

$$D = T_1 + T_2 \ (P_d + 2P_d^2 + 3P_b^3 + \ldots. + iP_d^i + \ldots.)$$

$$D/T_1 = 1 + (T_2/T_1) \ [P_d/(1-P_d)^2] \qquad (6\text{-}9)$$

For $1-P_d \simeq P_c$ the increase in D relative to T_1 is $(T_2/T_1)(P_d/P_c{}^2)$ which is normally small. For $T_2/T_1=2$ and $P_d=10^{-2}$ the relative delay increase is 2%.

If the delays T_1 and T_2 are variable, then T_1, T_2 and D in (6-9) are average delays.

It is seen that for channels which are "good" most of the time, retransmissions are infrequent and η and D are not reduced significantly below their FEC values. When the channel degrades increased retransmissions and delays together with reduced values of η is the price paid for low decoded bit error probabilities.

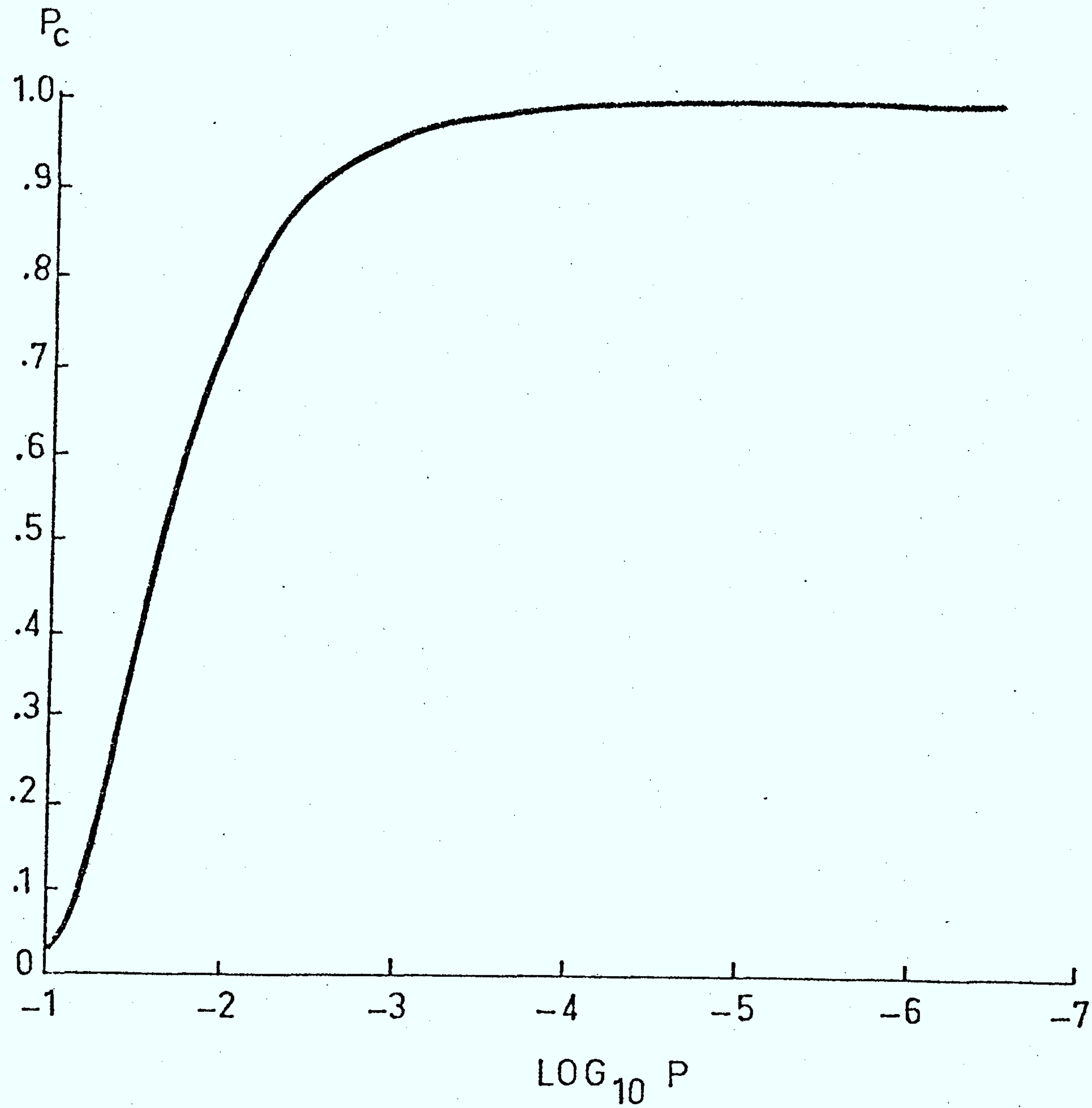

Fig. 6-4 Probability of correct decoding P_c for ARQ vs. channel bit error probability p; block length n=31

For send-and-wait ARQ, η gives a somewhat optimistic indication of overall system information throughput when D is large. In this case the channel is idle much of the time, and conventional channel sharing schemes do not permit other users access to the channel during these idle periods. For go-back-N protocols, the lower limit on N is determined by D [S1].

Curves like those in Fig. 6-1 together with (6-7) and (6-9) can be used to assess both FEC and ARQ codes in terms of decoded bit error probability, throughput efficiency and delay D. For example, a (31,26) code with stop-and-wait ARQ on a random error channel with bit error probability $p=10^{-2}$ yields $p_b=1.4 \times 10^{-5}$, $\eta=0.72$ (26/31)=0.60 and for $T_2/T_1=2$ a relative delay increase of 108%. Use of a (31,16) FEC code on the same channel yields $p_b=7 \times 10^{-5}$ and $\eta=0.52$. Because of its zero relative delay increase the FEC code may be the preferred choice in this situation, not withstanding its slightly higher p_b and lower η.

For some types of messages including control messages some form of ARQ seems essential to enable to transmitter to know whether or not its message was received.

VI-5 Data Block Overheads

In addition to k information bits a data block may also contain r=n-k check bits, v address bits, s block-synchronization bits c mode control bits and/or g acknowledgement bits (see Fig. 6-5).

What are the optimum values for k, r, v, s, c and g? What is the optimum total block length? These important questions are not easily answered.

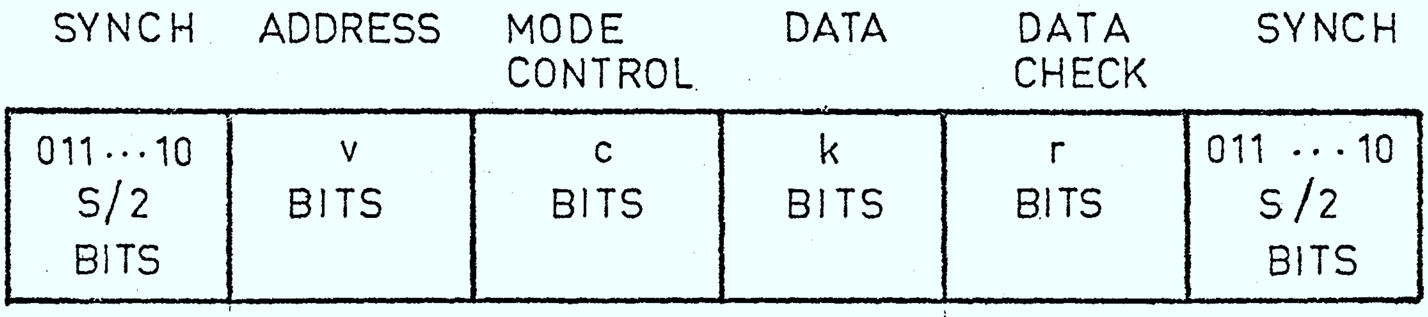

Fig. 6-5 Proposed message block structure

It is desirable to use blocks long enough to accommodate most messages, in order to reduce overhead. However long blocks contain wasted fill data when short messages occur, unless the block length is variable. This latter alternative reduces data block overhead but adds to system complexity since the encoder and decoder must adapt to varying block lengths. Long blocks may be advantageous on land mobile channels; if a block spans several fades the effect may be not unlike that which occurs when interleaving is used, in which case the effectiveness of FEC would be enhanced. On the other hand use of error detection only, without FEC favours short blocks if retransmissions are to be minimized, since the probability of a block error increases with block length.

The ARQ protocol used influences block lengths. Send-and-wait ARQ favours long blocks particularly when acknowledgement delays are large, in order to reduce the idle channel time referred to in Section VI-4. Continuous ARQ favours block lengths which decrease with as acknowledgement time increase, since some blocks previously received correctly may nonetheless require retransmission. Block length for selective ARQ would be independent of acknowledgement delay, although buffer storage as well as transmission and decoding delay would increase with bolck length. These statements are confirmed in a study by Chu [C1] who determined the optimum block length in terms of bit error rate for both random error and burst error channels, acknowledgement delay, and the average length of geometrically distributed messages.

Another issue is the optimum amount of forward error correction relative to error detection. Some FEC capability would reduce the number of message retransmissions at the expense of increased decoder complexity and increased error probability. Also of importance is the appropriate amount of FEC for acknowledgement information, whose accuracy is more critical than that of actual messages. Fujiwara et al [F1] considered minimization of the decoded bit error probability by selecting block code parameters to optimally balance error detection, error correction and protection of acknowledgement information. Independent error channels and (idealized) Gilbert burst noise channels were considered. For example, consider an overall block length of b=63 bits including k information bits, r check bits, and g acknowledgement bits. It was found that for a decoded bit error probability of 10^{-4} and throughput efficiency $k/(k+r+g)=0.7$ with continuous go-back-3 ARQ on an independent error channel, 16 check bits should be used to provide for correction of all single-bit errors and detection of all two- and three-bit errors, and that three bits should be used to code NACK's and ACK's.

Block synchronization is another important issue. One approach is to place the synch. sequence 011 .... 10, consisting of u ones between beginning and ending zeros, at the beginning and also possibly at the end of each block. To prevent this synchronizing sequence from appearing elsewhere in the data block, a zero is inserted after every sequence of u-1 consecutive ones prior to transmission and removed following decoding. It can be shown [D2] that $u=1+\sqrt{N}$ minimizes the synchronization overhead when a synchronization sequence preceeds a data block containing N bits, excluding synch. bits, and that $u = 1 + \sqrt{N/2}$ is optimum when the sequence also terminates the data block. For this latter case with N=100 and 900, respectively, optimum values for u are 8 and 22 and the ratio of synchronizing bits to total block length (including both synch. bits and the maximum number of stuffed bits) is 0.259 and 0.092, respectively. For the case of a single synch. sequence at the start, u=11 and u=31 is optimum for the two cases considered, and the overhead ratio is 0.187 and 0.065, respectively.

The use of prefixes and suffixes is but one of several ways to provide for block synchronization [S5].

VI-6 Data Block Structures for Land Mobile Channels

What data block structure is best for land mobile channels? A definitive answer to this important question seems unavailable.

The discussions in the preceeding section together with those in Section II-2 concerning the speed and accuracy requirements of various message types suggests several alternatives, one of which is illustrated in Fig. 6-5. The mode control bits indicate the message type and are coded using a low rate FEC code. The address bits would also be FEC coded; it would be feasible to check both the address and mode control information bits by a single parity bit sequence.

When control messages constitute the k data bits, the r check bits would be used for error detection only. Acknowledgements could be sent in special short blocks, possibly over a separate channel to avoid delays. Either an ACK or NACK could be sent following reception of each control message. Failure to receive an acknowledgement during a time-out period would result in a retransmission. A simple stop-and-wait protocol avoids the need for control message numbering, and seems adequate provided the block length is sufficient to include most control messages. The messages themselves could be restricted to a list designed to fit into a single data block. If s bits are used to code ACK's and NACK's, the probability of confusing the two acknowledgement words under majority decoding is

$$P_{ACK} = \sum_{i=(g+1)/2}^{g} \binom{n}{i} p^{i}(1-p)^{g-i} \qquad (g \text{ odd}) \qquad (6\text{-}10)$$

where p is the channel bit error probability. Fig. 6-6 shows P_{ACK} vs p for various values of g.

Transmission of real-time information such a voice would require neither FEC nor ARQ, since error control could be exercised by speaker and listener. Thus, the r check bits in Fig. 6-5 could be used together with the k data bits to provide n=k+r bits for voice transmission.

Transmission of text messages could be accomplished by use of the control message format in Fig. 6-5, provided the ensuing time delays could be tolerated.

Alternatively the check bits could be used for FEC rather than for error detection. In this latter case the absence of acknowledgements would permit continuous transmission.

Mode control provides for throughput-delay-reliability trade-off requirements consistent with the various message types, while permitting the use of a single encoder and decoder. Operators could select the mode to suit their requirements.

Other message block formats are possible [S4, K1]. The design of the block format together with the channel sharing scheme has an important bearing on overall system throughput, and deserves further consideration.

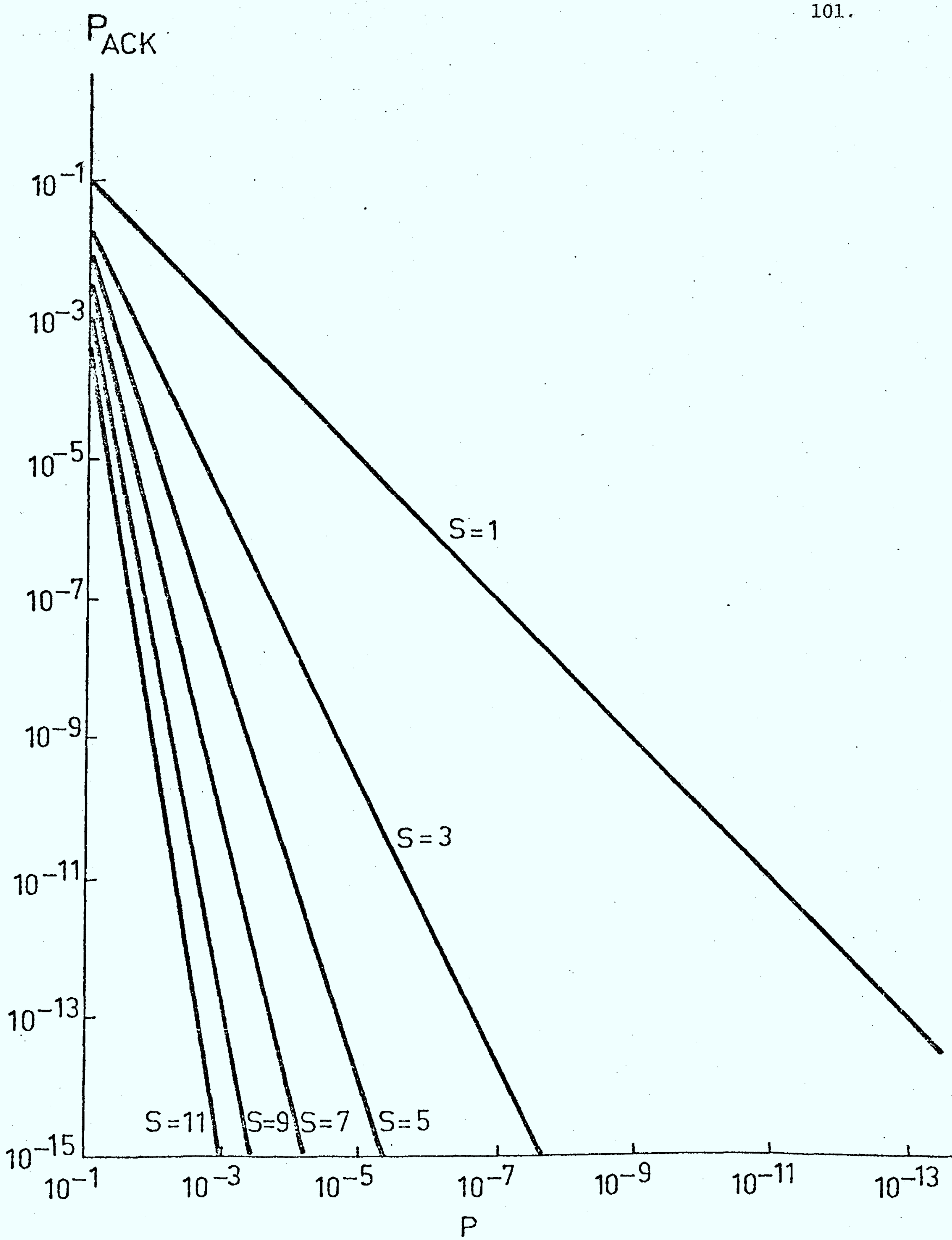

Fig. 6-6 Error probability P_{ACK} for acknowledgement information vs. channel bit error probability p and number of acknowledge bits s

Finally, the bits additional to the n=k+r bits in Fig. 6-5 add to overhead and reduce throughput efficiency below that determined in Section VI-3 where the only overhead bits considered were the check bits.

VI-7 References for Chapter 6

[A1] N.Abramson, "The throughput of packet broadcasting channels", IEEE Trans. Commun., vol. COM-25, pp. 117-128, Jan. 1977.

[B1] H.O. Burton and D.D. Sullivan, "Errors and error control", Proc. IEEE, vol. 60, pp. 1293-1301, Nov. 1972.

[B2] R.J. Bernice and A.H. Frey, Jr., "An analysis of retransmission systems", IEEE Trans. Commun., vol. COM-12, pp. 135-145, Dec. 1964.

[B3] R.J. Bernice and A.H. Frey, Jr., "Comparison of error control techniques", IEEE Trans. Commun., vol. COM-12, pp. 146-154, Dec. 1964.

[C1] W.W. Chu, "Optimal message block size for computer communications with error detection and retransmission strategies", IEEE Trans. Commun., vol. COM-22, pp. 1516-1525, Oct. 1974.

[D1] J.R. Davey, "Modems", Proc. IEEE, vol. 60, pp. 1284-1292, Nov. 1972.

[D2] R.W. Donaldson, Design of Distributed Computer-Communication Networks, Report to the Department of Communications, Ottawa, Ontario, Canada, July 1977.

[F1] C. Fujiwara, M. Kasahara, K. Yamashita and T. Namekawa, "Evaluations of error control techniques in both independent error and dependent-error channels", IEEE Trans. Commun., vol. COM-26, pp. 785-794, June 1978.

[K1] L. Kleinrock, Queueing Systems, Vol. 2: Computer Applications. New York: J. Wiley, 1976, Ch. 5 and 6.

[L1] R.W. Lucky, J. Salz and E.J. Weldon, Principles of Data Communication. New York: McGraw-Hill, 1968.

[M1] P.J. Mabey, "Mobile radio data transmission-coding for error control", IEEE Trans. Veh. Technol., vol. VT-27, pp. 99-110, Aug. 1978.

[P1] R.L. Pickholtz and C. McCoy, Jr., "Effects of a priority discipline in routing for packet-switched networks", IEEE Trans. Commun., vol. COM-24, pp. 506-516, May 1976.

[S1] A.R.K. Sastry, "Performance of hybrid error control schemes on satellite channels", IEEE Trans. Commun., vol. COM-23, pp. 689-695, July 1975.

[S2] A.R.K. Sastry and L. Kanal, "Hybrid error control using retransmission and generalized burst-trapping codes", IEEE Trans. Commun., vol. COM-24, pp. 385-393, April 1976.

[S3] M. Schwartz, Computer-Communication Network Design and Analysis. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[S4] C.A. Sunshine, "Effeciency of interprocess communication protocols for computer networks", IEEE Trans. Commun., vol. COM-25, pp. 287-293, Feb. 1977.

[S5] J.J. Stiffler, Theory of Synchronous Communication. Englewood-Cliffs, N.J.: Prentice-Hall, 1971.

[T1] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part III-Polling and (dynamic) split channel reservation multiple access", IEEE Trans. Commun., vol. COM-24, pp. 832-844, Aug. 1976.

APPENDIX-ERROR CONTROLS ON VERY NOISY CHANNELS

When the digital channel in Fig. 2-2 is very noisy, FEC will be of limited use since a relatively large fraction of the data blocks will have more errors than can be corrected. If an ARQ system is used, many retransmissions will result and throughput efficiency will fall, as explained in Section VI-4.

One simple way to combat the harmful effects of very noisy channels is to retransmit each bit N times, and use a majority decision at the receiver to arrive at averaged bit decisions. This approach has been used by the author [D1] to greatly improve the reliability of vehicle location and identification systems, by Sindhu [S1] on high bit-error satellite channels and more recently by Arredondo, Feggeler and Smith [A1] in the Bell System's advanced mobile telephone service (AMPS) test facility. The reduction in (averaged) bit error probability $\bar{p}$ is as indicated in Fig.6-4; for example if N=5 and $p=10^{-2}$, $\bar{p}=10^{-5}$. The implementation of bit averaging is relatively simple. The disadvantages include reduced throughput efficiency over that available using interleaving [C1, L1], transmitter buffer storage required to save data blocks for retransmission, and receiver hardware for storage and averaging.

Although efficient and effective interleaving is not always feasible on mobile channels. To be effective, interleaved bits must span many fades. From the discussion in Section II-4 it follows that a 10 Kb/s transmission rate of blocks 100 bits long would require interleaving of 50 blocks to span 25 fades at vehicle speeds of 15 mph. In many situations, including transmission of isolated control message blocks, interleaving would not really be feasible, and bit averaging would seem to be the only viable way to reduce the averaged bit error rate to where FEC or ARQ could be effective.

The current version of AMPS [A1] relies solely on bit averaging and FEC. In accordance with the discussions of Chapter VI such a system seems inefficient since retransmissions would often occur needlessly, for example when a vehicle is packed in high SNR zone. The system would aslo seem to be unsatisfactory for vehicles parked in a deep fade, in which case the absence of an acknowledgement protocol would prevent transmitters from realizing that their messages had not been received.

References

[A1] G.A. Arredondo, J.C. Feggeler and J.I. Smith, "Advanced mobile telephone service: Voice and data transmission", Bell Syst. Tech. J., vol. 58, pp. 97-123, Jan. 1979.

[C1] D.N. Chase and L.-J. Weng, "Multiple-burst techniques for slowly fading channels", IEEE Trans. Inform. Theory, vol. IT-22, pp. 505-511, Sept. 1976.

[D1] R.W. Donaldson, "Design of transponder signals and receivers for automatic vehicle location and identification systems", IEEE Trans. on Commun., vol. COM-23, pp. 489-500, May 1975.

[L1] S. Lin, An Introduction to Error-Correcting Codes. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

[S1] P.S. Sindhu, "Retransmission error control with memory", IEEE Trans. Commun., vol. COM-25, pp. 473-479, May 1977.