FREQUENCY ASSIGNMENT FOR LAND MOBILE RADIO SYSTEM IN THE 900 MHz BAND: DIGITAL TRANSMISSION OVER LAND MOBILE CHANNELS-SPECTRUM USAGE CONSIDERATIONS

by

Dr. Robert W. Donaldson

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

February, 1981

FINAL REPORT

FREQUENCY ASSIGNMENT FOR LAND MOBILE RADIO SYSTEM IN THE 900 MHz BAND: DIGITAL TRANSMISSION OVER LAND MOBILE CHANNELS-SPECTRUM USAGE CONSIDERATIONS

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 Serial No: OST80-00020
Period: September 1, 1980 - March 31, 1981

ABSTRACT

Quantitative criteria are defined for design and evaluation of land mobile radio systems. These criteria include delay or blocking probability, reliability as measured by signal-to-noise-plus-interference ratio SNR, and throughput, and are expressed in terms of the important system variables. Cost, which changes rapidly with technology, is not expressed precisely, but is related in a semi-quantative way to system variables.

The criteria are used to determine performance obtainable with cellular systems using conventional frequency division multiple access, and fixed channel assignments. Particular emphasis is placed on the limitations resulting from co-channel and adjacent-channel interference, since this ultimately limits information throughput. Maximum R/Δ values for various modulation formats and channel assignment schemes are determined (where R is the bit-rate and Δ the channel spacing), for fixed values of SNR, defined at a particular cell radius. Throughput (spectrum efficiency) in bits/sec/Hz is determined for various SNR values, modulation formats, assignment schemes and channel codes, for various values of delay or blocking probability and number of channels per mobile. The way in which channel encoding can be used to enhance spectrum efficiency and provide additional system design flexibility is illustrated for both forward error-correction and error-detection/retransmission modes.

Specific results are obtained for systems of up to 19 cells, with and without reuse of frequency channels. Such systems are typical of those in use or being proposed for immediate use. For given values of delay or blocking probability, highest throughputs are obtained when frequency reuse is employed, together with channel encoding. The best code rate depends on the reuse plan. The appropriate error control protocol depends on the type of message to be transmitted; forward error correction (FEC) is best for real-time speech, FEC or ARQ are appropriate for text, and ARQ is best for control messages where accuracy requirements are paramount.

The results bear directly on policy issues, including the selection of Δ, spectral emission constraints, questions as to whether or not channels should be used in an unrestricted way for both analog and digital transmission, and system designs which should be encouraged to improve spectrum efficiency.

Issues requiring further study are identified, including the mutual interference between analog FM speech and data, accurate determination of bit-error probabilities over a radio coverage area, error control protocols and data block formats for improved efficiency, low-bit rate digital speech encoders, and other schemes, including spread spectrum for sharing of land mobile radio channels.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

I INTRODUCTION

I-1 Scope and Objectives of the Study

This report presents the results of a Communications Canada Contract study whose motivation, purpose and general approach are summarized below.

MOTIVATION

Recent advances in land mobile system development has provided increasing evidence that digital transmission of dispatch data as well as other digitally encoded information will constitute an important part of land mobile communication traffic. The need for the development of guidelines identifying transmission characteristics, requirements, and limitations of digital transmission is becoming increasingly urgent. These guidelines would lead to the development of new regulations and standards governing the technical characteristics and operation of equipment and systems employing digital transmission of information over land mobile communication channels.

The following items are of specific interest:

1) To study tradeoffs between channel bandwidth, spectrum efficiency, and equipment complexity when utilizing digital modulation techniques on Land Mobile channels.
2) Multi-user sharing for data transmission - to evaluate different sharing strategies, taking into consideration blocking probability, queueing delay, data rate and channel bandwidth.
3) Data reflecting relations between channel bandwidth and spectrum efficiency (information bits) as well as the relations with equipment complexity.
4) Brief description of each different strategy. For each scheme

produce data relating the number of users to blocking probability, queueing delay and channel occupancy.

PURPOSE

To produce technical data in support of the development of guidelines for the use and operation of Land Mobile radio equipment and systems in the 900 MHz band for communication over Land Mobile channels utilizing digital modulation techniques.

GENERAL APPROACH

The work involves fundamental questions concerning the sharing of radio broadcast spectrum over time, frequency and space among a population of users.

In order to compare alternative sharing schemes, it is necessary to choose performance criteria, and to relate these to system variables and management protocols. The variables and protocols would then be selected to give good performance.

Many communication applications involve two-way transmission of information, often in an interactive conversational environment, where the following performance criteria are meaningful:

1) Either average delay between the time when a message is ready for transmission and the time of its receipt at the destination, or probability that the message is blocked from transmission because all facilities are in use (for some management protocols both criteria, which are inter-related, are appropriate).
2) Message transmission reliability.
3) System cost.

4) Information throughput.

These criteria are not usually independent. For example, an increase in information throughput or reliability normally implies an increase in system delay, if the system cost remains unchanged.

System variables include number of users, average message generation rates, average message lengths, parameters associated with various strategies for sharing communication resources, digital modulation format, data rate, error detection/correction code, message retransmission protocol, and switching scheme. Switching alternatives involve conventional line switching, and packet switching. Sharing (multiplexing) strategies include the conventional division of the frequency spectrum into channels with particular channels being assigned to different geographic regions, spread spectrum multiple accessing, and aloha-types of contention multiplexing. Packet switching variables include packet length and routing algorithms. Conventional spectrum channelization parameters include channel bandwidth and rules for assignment of channels to geographic regions.

It is desirable that analysis, evaluation and design procedures be based as much as possible on analytical models of various system configurations. Such models facilitate calculation of performance criteria and optimization of these criteria. In those cases where quantative determination of performance is not feasible, approximation techniques or simulations would be required.

A thorough and detailed consideration of all aspects related to the proposed research is surely not feasible under the constraints of the present contract. Thus, it was recognized at the outset that an important part of the study would involve selection from among various alternatives an approach which would provide useful technical data consistent with the study's motivation and purpose.

RESULTS

This report emphasizes cellular systems employing spectrum sharing via frequency division multiple access (FDMA). Particular emphasis is given to performance limitations resulting from co-channel and adjacent-channel interference. Cellular systems are of current interest, are in current use, and represent a natural and readily implemented extension of conventional FM systems for analog voice transmission.

The report deals extensively with all issues raised under "motivation". Curves and tables are provided to facilitate determination of throughput and spectrum efficiency for a given delay or blocking probability, at a specified signal-to-noise-plus-interference ratio. A detailed executive summary follows.

I-2 Executive Summary

This report presents a quantitative analysis of the performance of cellular systems with fixed channel assignments. The analysis procedures developed are applicable to other spectrum sharing schemes.

Chapter 2 identifies three types of message sources, including speech, text and control data. The data rates and accuracy requirements of these differs considerably. Speech has a high data rate but low accuracy requirements, while the converse applies to control information.

Chapter 2 also examines the functional tasks of digital communication systems, and reviews digital modulation formats and implementations particularly suited to land mobile radio systems. These formats include binary PAM, PSK (and DPSK) and FSK. Of particular importance are transmission bandwidths, spectral roll-off rates which affect adjacent-channel interference, and susceptibility to signal level fluctuations. These fluctuations result from Rayleigh fading, shadowing and mobile-base separation changes, and can exceed 90 dB. Interference effects are described, and the effects of co-channel and adjacent-channel interference are described quantitatively for digital signals, by an interference function $C(\Delta/R)$, where Δ and R denote the channel separation and bit rate, respectively. Bit-error probability expressions are summarized for the various modulation formats considered. Alternative spectrum sharing alternatives are summarized. It is noted that conventional frequency division multiplexing is an easily implemented extension of existing analog FM systems, and that fixed channel assignments are easily implemented and seem best for heavily loaded channels.

Chapter 3 sets out system performance criteria, and expresses these in terms of system variables and message traffic parameters. Performance criteria selected include delay D or blocking probability P_B, system throughput (spectrum efficiency) Q in bits/sec/Hz, reliability, and cost. Both D and P_B are expressed in terms of the number of channels/mobile m, and the channel utilization factor. Reliability is defined in terms of the signal-to-noise-plus-interference ratio SNR, defined as the wanted signal at distance $r/\sqrt{2}$ from the transmitter where r is the cell radius; interference is assumed located at cell centre. Cost is determined indirectly by observing values for m and the number of bases N, and by noting the complexity of the modulation format, channel assignment scheme and channel code. System comparisons involve finding the minimum Δ/R values which will meet a given SNR requirement, assuming co-channel and adjacent-channel interference to be the major sources of disturbance. Once Δ/R has been obtained, P_B, D and Q are determined, in terms of m.

Use of both P_B and D as performance criteria creates system design conflicts. Minimal delay favours minimal values for m and Δ/R, and large R values. Low P_B values favour small Δ/R and large m, with R equal to (and not in excess of) the rate required for real-time transmission. Both criteria are of interest, since a channel may carry both real-time speech and non-real-time data traffic. Also included in Chapter 3 is a discussion of channel occupancy, and graphs for determining performance criteria of various system configurations.

Chapter 4 presents a thorough analysis of single cell systems. Effects of Δ, R, m, system bandwidth B and modulation format on SNR, P_B and D and Q are considered in detail. For SNR $\lesssim$ 30 dB, MSK is seen to yield higher spectrum efficiencies than PSK, PAM and SFSK; the reason for this result is the relatively fast spectral roll-off rate of MSK for this SNR range. Above 40 dB, SFSK is most efficient.

Chapter 5 considers multi-cell systems which do not reuse frequency channels, and provides detailed results for systems having three, seven and nineteen cells. Channels are assigned to minimize adjacent-channel interference. Maximum throughputs for the three types of systems differ considerably. For MSK with SNR = 20 dB, for example, maximum throughputs for systems containing 1, 3, 7 and 19 cells are, respectively, .95, 1.58, 1.58 and 4.6; for PSK these maximum values equal 0.30, 0.65, 1.2 and 3.6. Except for certain special situations, MSK permits greater spectrum efficiency than PSK. Also indicated in Chapter 4 is the range of values of Q and m for which various cellular plans are best for given P_B values. Generally, as Q increases, larger numbers of cells become advantageous.

Chapter 6 considers cellular systems which reuse frequency channels, to reduce the minimum number of channels per mobile and to improve spectrum efficiency. Reuse causes co-channel interference which limits the obtainable

SNR value. Maximum R/Δ values are determined, subject to SNR limitations. As the number of groups of channels increase, co-channel interference decreases, and adjacent-channel interference also decreases slightly. For 3-cell, 4-cell, 7-cell and 9-cell reuse plans, maximum throughputs for MSK (PSK) at SNR = 20 dB are: 5.0 (3.3), 6.0 (4.0), 4.5 (4.0), 5.3 (9), as compared with 4.6 (3.8) for a 19-cell system with no reuse. With P_B constrained at 0.2, spectrum efficiencies for the above systems are: 4.0 (2.6), 4.6 (3.1), 3.0 (2.6), 3.2 (5.4) vs. 1.8 (1.5) for a 19-cell system with no frequency reuse.

The importance of accurately specifying the propagation factor n in determining spectrum efficiency is also illustrated in Chapter 6. Our results are obtained with n = 3.5.

Chapter 7 examines the use of codes for forward error correction (FEC) as well as for error detection and retransmission (ARQ). High interference levels and high channel bit rates may result in unacceptably high raw bit-error probabilities. Channel encoding combats these problems, as illustrated by specific examples, and permits greater flexibility and spectrum efficiency through increased reuse than otherwise possible. For example, use of MSK with 40-channel transceivers, $P_B = 0.2$, and decoded information bit-error probability $p_b = 10^{-4}$ requires a 12-cell reuse pattern for 19-cell systems, with throughput Q = 0.5 bits/sec/Hz. Use of a (15, 7) double-error correcting codes permits a 4-cell reuse pattern, with Q = 2.8 bits/sec/Hz. The p_b value is a nominal one at a distance $1/\sqrt{2}$ time the cell radius. More work is needed to determine its value at various cell locations.

Chapter 7 also proposes coding to accommodate the different speed-accuracy requirements of the different types of messages identified in Chapter 2. The accommodation is via mode control bits in data blocks, whose structure is examined in relation to coding and spectrum efficiency.

Chapter 8 summarizes the major results of this study, indicates how they relate to important spectrum management policy issues and identifies issues needing additional study. The work has a direct bearing on Δ and R, whose selection depends to some extent on whether or not separate frequency bands are to be used for digital transmission. If channels are to be used in an unrestricted way for both digital and conventional analog FM voice transmission, additional work is needed to determine nutual adjacent-channel interference between analog and digital signals. Earlier work shows this mutual interference to be very dependent on the pre-modulation filtering of the speech signal. This filtering affects the spectral roll-off rates of the transmitted signal, and it is this roll-off rate which ultimately determines adjacent-channel interference levels. For this reason, roll-off rates may be a useful addition to regulatory spectral emission constraints.

More detailed calculations of bit-error probabilities at various cell locations are required. Both averages and worst case error rates would be useful. These are needed to more accurately determine best values for Δ and R, and most favourable channel assignments and reuse plans.

Finally, additional study is needed to make performance comparisons, as illustrated in this study, between cellular systems and other spectrum sharing schemes, including spread spectrum multiple access.

II MODULATION, CODING AND SPECTRUM SHARING FOR LAND MOBILE RADIO CHANNELS

II-1 Introduction

In order to assess the performance of any communication system, it is first necessary to fully understand system operation, and in particular to isolate key system parameters and operational policies. This task is accomplished in the present chapter.

It is also important to clearly articulate performance criteria, and to express these criteria in terms of the system parameters and operational policies. Chapter III accomplishes this task.

The present chapter identifies the important functional tasks performed by a digital communication transmitter and receiver, delineates types of message sources, summarizes land mobile channel physical behaviour, presents for various modulation formats transmitted spectra, interference effects and bit-error probabilities, summarizes the effects of using various error control protocols, and briefly examines various spectrum sharing alternatives.

II-2 Information Sources for Land Mobile Radio 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 [F1, H1, C1]. Because of its natural redundancy, random bit error rates as low as 10^{-2} do not prevent intelligible transmission of speech [F1, H1, C1]. Documentation regarding burst error effects on digitally transmitted speech is scarce; however, some results are available [Y1, J1].

2. Text information 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. Control information 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.

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 all three types of information, each with different speed-vs-accuracy priorities. It may therefore be advantageous to provide for two or more digital transmission modes, one of which would favour speed while another would favour accuracy. Implementation of this strategy is considered in Section 7-5.

II-3 Functional Tasks of Digital Communication Transceivers

Fig. 2-1 shows in block diagram a digital communication link, 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, or both.

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

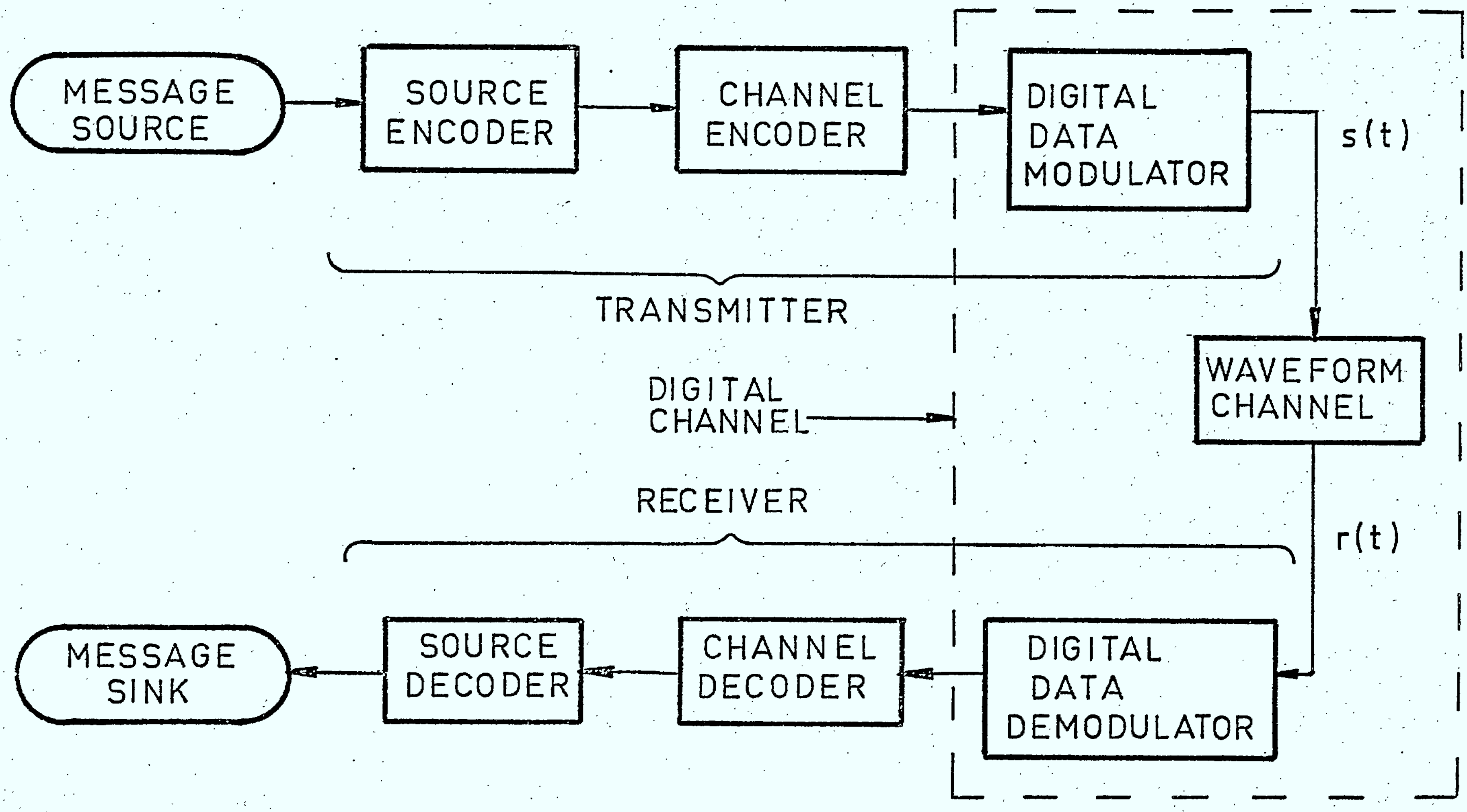

Fig. 2-1 Block diagram of a digital communication link.

The term digital communication implies a fundamental constraint on the transmitter and receiver in Fig. 2-1, 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 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 made 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 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 [G1, L3, P1]. 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. 2-1, 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 Physical Behaviour of Land Mobile Radio Channels

Land mobile radio channels are dispersive in time, frequency, and space [J2, C2, K1]. Relative motion between the transmitter and receiver causes the level of a received narrow-band signal to fluctuate rapidly, as indicated in Fig. 2.2. These level changes are due to scattering and multipath, and vary with the receiver's spatial location and with the frequency band occupied by the signal. The rate of signal level change depends on the velocities of the mobile transceiver, which velocity is itself variable. Adjacent peaks in the signal envelope correspond to changes in the relative separation between transmitter and receiver of one-half the (narrow-band) signal's 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, F2]:

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

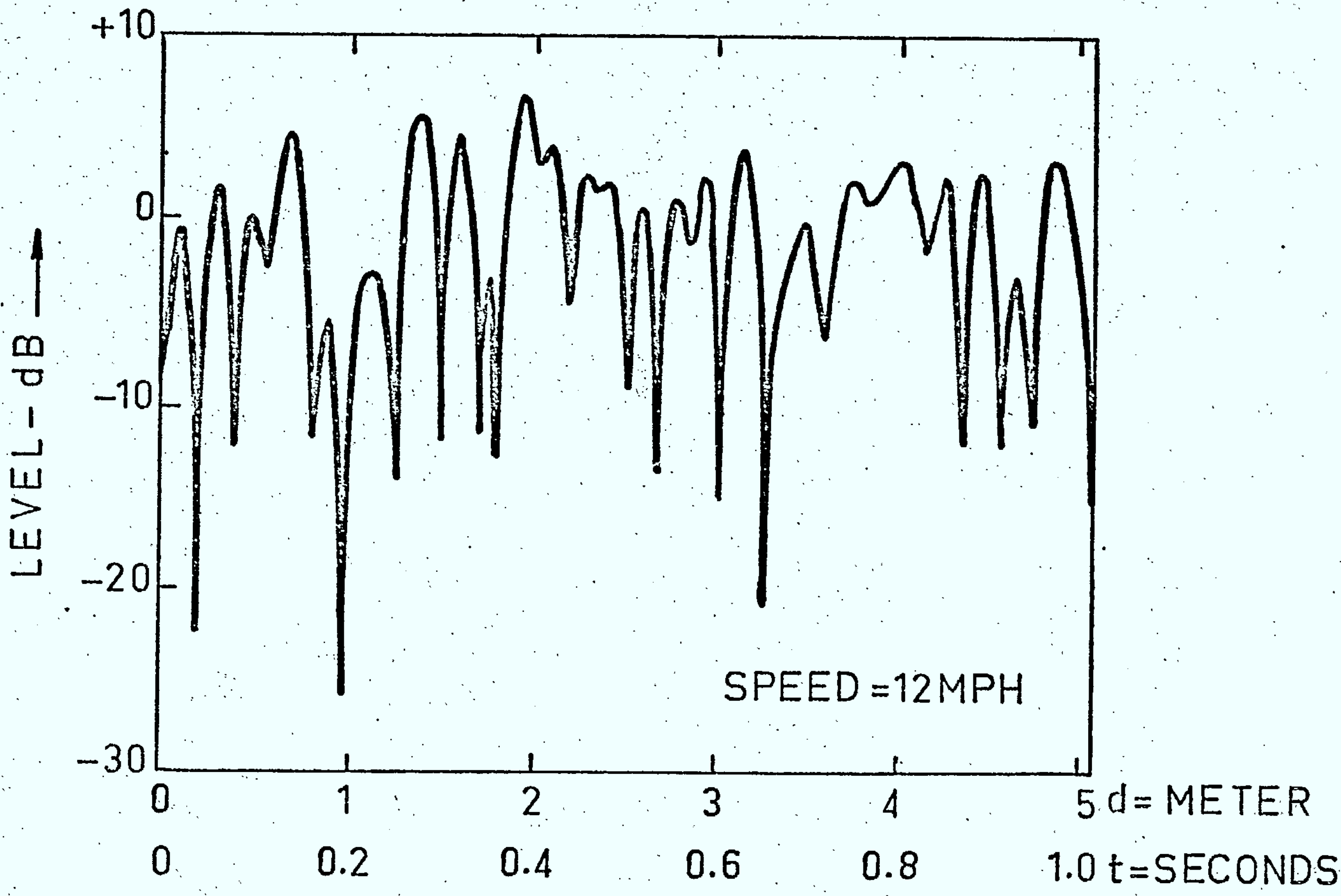

Fig. 2-2 Received signal level vs. time in an urban environment. Vehicle speed: 12 mph. Carrier frequency: 850 MHz. (after Arredondo and Smith [A1]).

where $s = \sqrt{\pi r_0/2}$ is the local mean level of the received signal and r_0 is the average energy in the received signal. The distribution of r^2 the received signal energy is [W1, F1]:

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

The local mean signal level s varies slowly over several wavelengths as a result of gradual changes in path transmission characteristics. These changes 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 [J2, F2]:

$$f_s(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(\alpha-m)^2/2\sigma^2) \tag{2-3}$$

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 quated as 6 dB in London [F2] and as 8-12 dB in Japan and the U.S.A. [F2, J2].

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 $2<n<4$ [J2, 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].

The rate of signal level variation relative to vehicle speed affects the performance of various digital transmission schemes. 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, $\lambda \simeq 1$ft.. Thus, fading minima in Fig. 2-2 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 necessarily during the transmission of one data block when using conventional 30 kHz channels. The mean signal level, however, being relatively constant over 50 m, is relatively constant over one transmitted data block.

The random doppler shifts in signal frequency caused by 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.

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 variation in excess of 90 dB. One way to combat these large fluctuations is to use one or more types of diversity, based on the fact that these variations are not uniform over time, space or frequency. For example, spatial diversity utilizes several base station antennas located at the edge of the radio coverage area to reduce fluctuations in base-station received signal level resulting from all these of the above-noted causes [J2]. 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 [J2]. 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 [L1, G1] codes involve use of time diversity. Frequency diversity is an inherent part of recently proposed spread spectrum systems, in which each transceiver makes full use of the available spectrum.

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 implies 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 [C3, C4]. Alternatively, the decoder in the receiver 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 when SNR is low; thus, the actual information rate adjusts automatically to the state of the channel.

Again, it should be emphasized that Fig. 2.2 corresponds to a signal whose bandwidth is sufficiently narrow that the level of all frequency components vary in unison (flat fading). Conventional 30 kHz land mobile channels meet this narrow-band criterion; as indicated earlier, spread spectrum systems do not.

II-5 Modulation Alternatives for Land Mobile Radio Channels

Digital modulation formats for conventional FM voice channels (≈30 kHz bandwidth) have been considered in earlier reports [D1]. The methods deemed suitable include binary antipodal PAM, PSK (and differential PSK), and FSK.

In pulse amplitude modulation (PAM) the digital data modulator in Fig. 2-1 generates a signal s(t) whose constitutive pulse amplitudes depend on the symbol sequences $\{a_k\}$ impressed upon the modulator [L1]. Thus

$$s(t) = [\sum_k a_k\, g(t-kT)]\, \sqrt{2P}\, \cos\, (\omega_c t + \theta) \tag{2-3}$$

where $a_k = \pm 1$, g(t) is the basic modulator pulse shape, ω_c is the (radian) carrier signal frequency, θ is the carrier phase, $T=R^{-1}$ where R is the bit rate and P is the power of the unmodulated carrier signal.

PAM is best restricted to binary antipodal to avoid the need for decision level adjustment in accordance with signal level changes (the decision level remains at zero). Pulses g(t) with zero intersymbol interference (isi) at sampling times are also required, since isi receivers depend on the signal level.

Fig. 2-3 shows three different pulse shapes, all of which show zero intersymbol interference at the sampling times $t=nT$, $n \neq 0$. The two non-rectangular pulses are derived from the family of pulse shapes

$$f(t) = K\left(\frac{\sin \Pi t/T}{\Pi t/T}\right)\left(\frac{\cos\, \alpha\, \Pi t/T}{1-(2\alpha t/T)^2}\right) \tag{2-4}$$

where α $(0 \leq \alpha \leq 1)$ controls the pulse "excess" bandwidth [L1] and K the pulse power. For $\alpha=0$, $f_a(t)$ in Fig. 2-3 results, whereas $f_c(t)$ occurs for $\alpha=1$. The larger is α, the less is the performance degradation caused by sample-timing errors [L1]. The actual modulator pulse g(t) is derived from the corresponding modulator pulse shape f(t), as follows: for the rectangular pulse, $g_b(t)=f_b(t)$

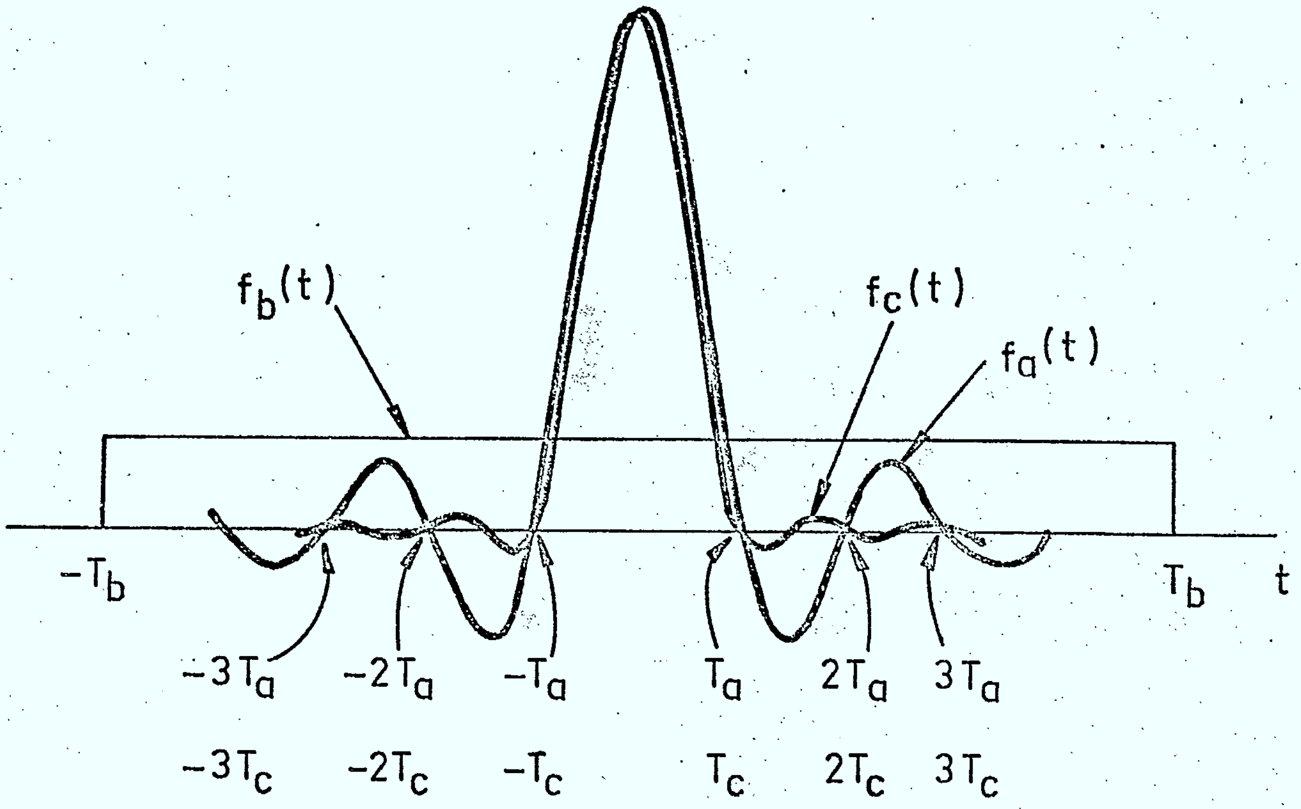

Fig. 2-3 PAM pulse shapes where f(t) is the pulse shape following matched filtering at the demodulator.

while for the family specified by (2-4) g(t) is the inverse Fourier transform of $\sqrt{F(f)}$ where F(f) is the Fourier transform of f(t). These comments follow from the fact that minimization of bit error probability on a white Gaussian noise all-pass channel requires that the modulator pulse shape and receiver baseband filter impulse response be identical [L1].

The power spectral density S(f) of the transmitted PAM signal s(t) is

$$S(f) = \frac{1}{2}[P(f+f_c) + P(f-f_c)] \tag{2-5}$$

where P(f) is the power spectral density of the baseband signal p(t) = $\Sigma\, a_k\, g(t-kT)$. Fig. 2-4 shows the corresponding plots of S(f); the sampling periods have been selected to enable the spectra to fit tightly inside a "top-hat" spectral emission constraint. 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 for binary signalling is easiest for the sin t/t - type pulse with spectral density

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

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

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

is

$$P_T = P \int_{-\infty}^{\infty} P(f)\, df \tag{2-8}$$

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).

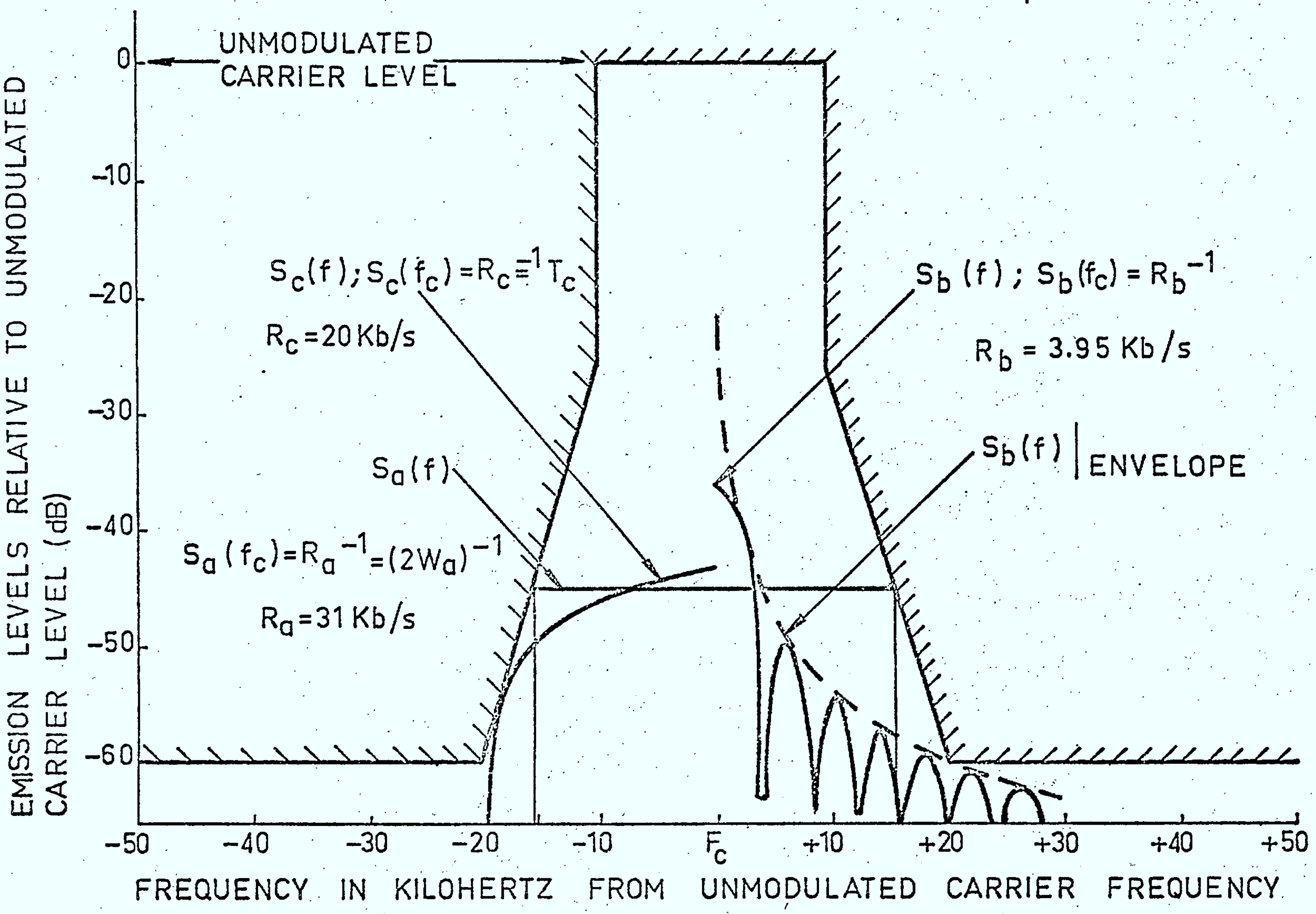

Fig. 2-4 Power spectral density for the pulse shapes in Fig. 2-3 under maximum bit-rate conditions.

Use of (2-6) together with

$$10 \log_{10} E(f) = 8 - 3.4 \; |f| \qquad 10 \leq |f| \leq 20 \tag{2-9}$$

where E(f) is the spectral emission constraint boundary in Fig. 2-4 for frequencies $10 \leq |f| \leq 20$ (f in kHz), yields the maximum value of $W_a \approx 15.5$ kHz. This value of W_a solves the following equation:

$$10 \log_{10} (1/2 \; W_a) = 8 - 3.4 \; W_a \tag{2-10}$$

It follows that the maximum bit rate is $R_a = 2 \; W_a \approx 31$ Kb/s. It should be noted that the total absence of spectral energy outside the frequency band $f_c \pm 15.5$ kHz in Fig. 2-4 implies zero interference with adjacent channels whose separation from f_c exceeds 31 kHz.

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

$$S_b(f) = K_b \; T_b \; (\sin \pi f T_b / \pi f T_b)^2 \tag{2-11a}$$

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=20 kHz. The resulting value of R_b is obtained from

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

which yields $R_b \approx 3.95$ Kb/s. Thus, the maximum bit rate permissible for conventional rectangular pulses under the "top-hat" constraint is 12.7% of that permitted for sin t/t - type pulses. In addition some interference is generated for adjacent channels at arbitrarily large frequencies from f_c.

The power spectral density $S_c(f)$ of the transmitted signal with $\alpha=1$ is

$$S_c(f) = \begin{cases} K_c = T_c & |f| < (1-\alpha)/2T_c \\ \frac{T_c}{2} \left[1 - \sin[(\pi T_c/\alpha)(f - \frac{1}{2T_c})] \right] & (1-\alpha)/2T_c \leq |f| \leq (1+\alpha)/2T_c \\ 0 & |f| > (1+\alpha)/2T_c \end{cases} \tag{2-11}$$

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 $(1 - \frac{\alpha}{2}) R_a$ where R_a is the maximum bit rate for the case $\alpha=0$. Fig. 2-4 yields $R_c \approx 20$ Kb/s which, because of the emission constraint roll-off from -26dB to -60dB, exceeds the lower bound $R_a/2 \approx 15.5$ Kb/s. To avoid adjacent-channel interference, frequency channels must be separated by at least 40 kHz.

Reduction of α would increase the maximum allowable bit rate as well as the timing error sensitivity. Thus, selection of excess bandwidth α permits a trade-off between these 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 $R_a \approx 31$ Kb/s and $R_b \approx 4$ Kb/s.

Optimum digital data demodulators for PAM in white Gaussian noise are as shown in Fig. 2-5. In the absence of isi, the filter impulse response h(t) is matched to the modulator pulse shape g(t).

For FSK, the transmitted signal s(t) in Fig. 2-1 is of the form

$$s(t) = \sqrt{2P} \cos \left(\omega_c t + \omega_d \int_{-\infty}^{t} x(\tau)\, d\tau + \theta\right) \qquad (2\text{-}12)$$

where

$$x(t) = \sum_{k=-\infty}^{\infty} a_k\, g(t-kT) \qquad (2\text{-}13)$$

and $\{a_k\}$ denotes the sequence of symbols to be transmitted, θ is a random phase angle, P is the received power, ω_c is the carrier frequency, $\omega_c + \omega_d\, x(t)$ is the instantaneous frequency, and ω_d is the peak frequency deviation when x(t) is normalized to have unity peak amplitude. When g(t) in (2-13) is a rectangular pulse

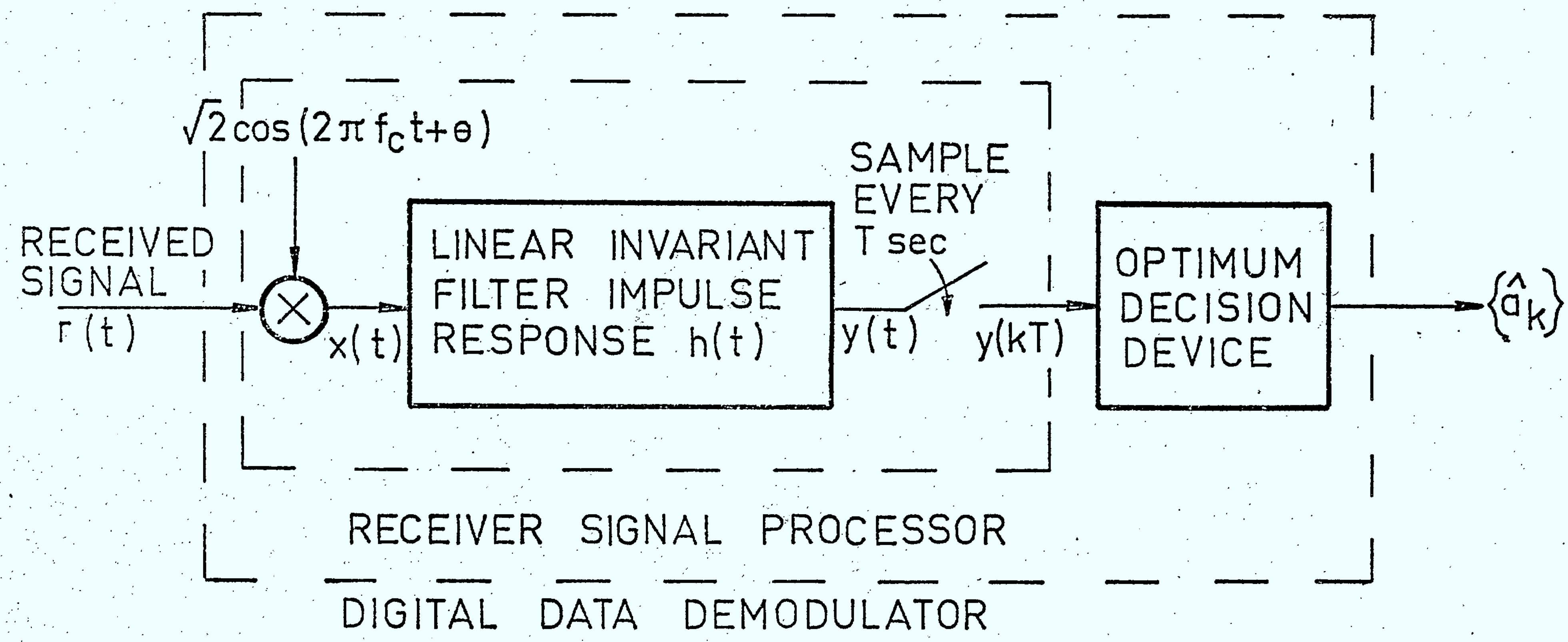

Fig. 2-5 Receiver for coherent demodulation of PAM signals.

$$g(t) = \begin{cases} 1 & 0<t\leq T \\ 0 & t\leq 0,\ t>T \end{cases} \qquad (2\text{-}14)$$

and conventional FSK results. The selection of $g(t)$ to give $s(t)$ the most desirable spectral shape while maintaining low error rates remains an unsolved problem, in general.

Receivers for FSK signals are of the form shown in Fig. 2-5, except that for optimum reception, the receiver signal processor consists of a bank of filters, with each filter matched to one of the N possible FSK pulses [L1,W1] (N=2 for binary signalling). Such receivers perform coherent detection. Optimum incoherent detection involves matched filtering followed by envelope detection; the result is a simpler receiver, since the phase θ in (2-12) need not be tracked by the receiver. Another (suboptimum) receiver consists of a bandpass filter, followed by a limiter-discriminator, followed by a lowpass filter as the receiver signal processor [T1,J2].

When the data rate R for conventional binary FSK is such that $T = 4f_d = 2\omega_d/\pi$, minimum shift keying (MSK) results. MSK signals can be viewed as being composed of quadrature carrier PAM signals modulated by alternative data bits [K2,A2,G2]. Viewed in this way, MSK signals can be detected coherently using the receiver in Fig. 2-6. The result is a 3dB improvement in received power utilization over that which results when MSK is detected (coherently) as binary orthogonal FSK. The basic pulse shapes $g(t)$ for each quadrature data stream are as follows:

$$g(t) = \begin{cases} \frac{1}{\sqrt{T}} \cos \frac{\pi t}{2T} & |t|\leq T \\ 0 & |t|>T \end{cases} \qquad (2\text{-}15)$$

The transmitted MSK signal is of the form

$$s(t) = \sqrt{P}\Big[\sum_{i\ \mathrm{even}} a_i g(t-iT)\cos(2\pi f_c t+\theta) + \sum_{i\ \mathrm{odd}} a_i g(t-iT)\sin(2\pi f_c t+\theta)\Big] \qquad (2\text{-}16)$$

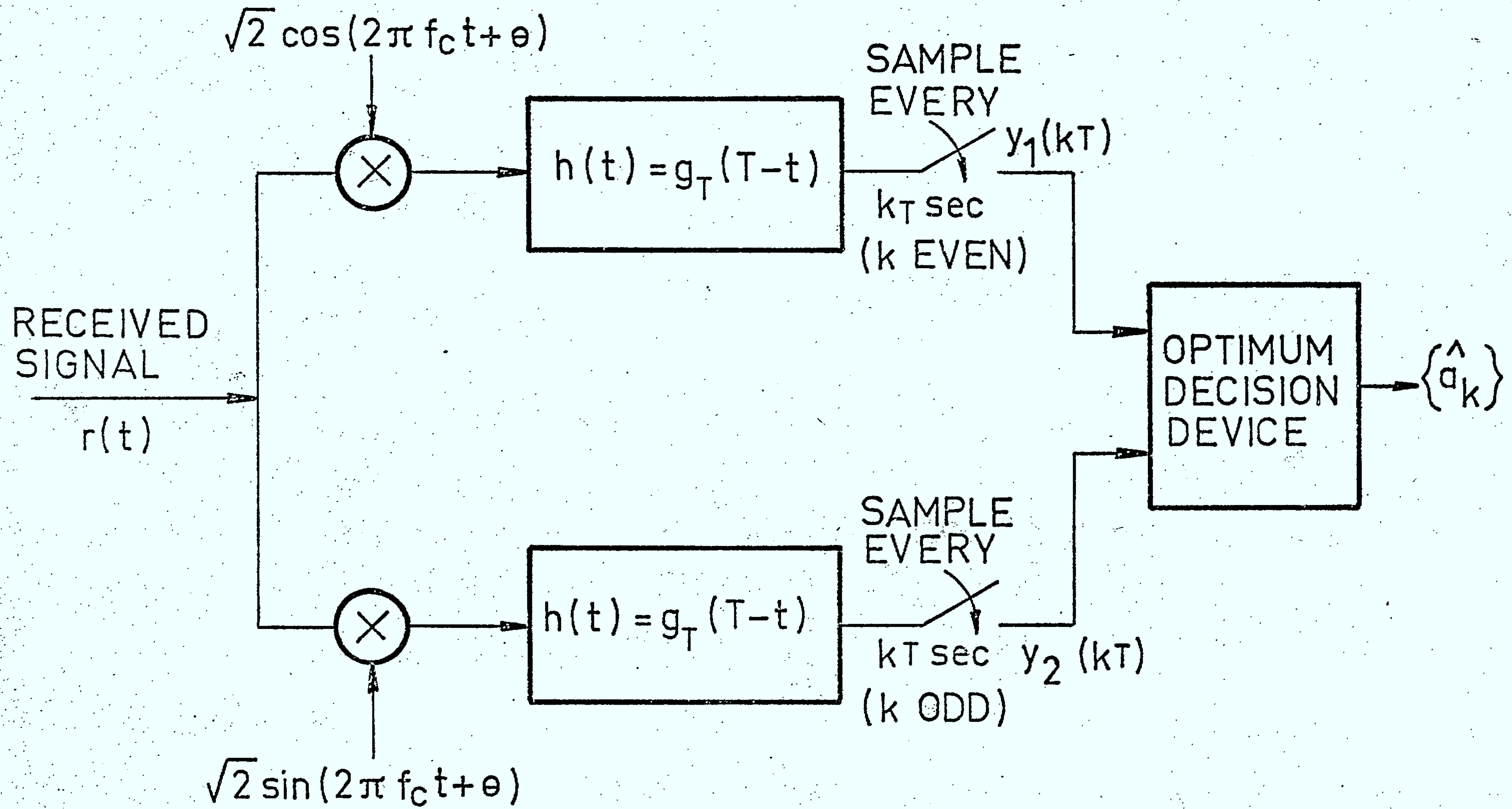

Fig. 2-6 Optimum binary MSK receiver.

where P is the transmitted power, $\{a_i\}$ are the binary symbols, and θ a random phase angle.

In general, the determination of the power spectral density of digital FM signals is difficult, and requires approximate techniques [G3,S3,T1] However, for binary MSK the power spectral density $S_{MSK}(f)$ is as follows:

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

where f is the frequency offset from the carrier. Fig. 2-7 shows $S_{MSK}(f)$ fitted tightly within the "top-hat" characteristic referred to earlier, for which the maximum permitted bit-rate is R=36 Kbs.

The compact MSK spectrum results because of phase continuity between bits. Extension of phase continuity to the first and higher derivatives leads to MSK-type signals, such as sinusoidal FSK (SFSK) [A2], where the transmitted signal consists of two quadrature data streams of the form of (2-16) with

$$g(t) = \begin{cases} \frac{1}{\sqrt{T}} \cos[\frac{\pi t}{2T} - U \sin \frac{2\pi t}{T}] & |t|<T \\ 0 & |t|\geq T \end{cases} \tag{2-18}$$

where U=1/4. Analysis shows that SFSK signals have spectra which fall off as f^{-8}, as compared with f^{-4} for MSK and f^{-2} for conventional PSK. Fig. 2-8 shows spectra for these three types of signals. Note that SFSK is more difficult to implement than MSK.

A conventional digital phase modulated signal is of the form

$$s(t) = \sqrt{2P} \sum_{k=-\infty}^{\infty} g(t-kT)(\cos \omega_c t + \phi_k) \tag{2-19a}$$

where rectangular pulse g(t) is defined in (2-14). Phase sequency $\{\phi_k\}$ varies in accordance with symbol sequence $\{a_k\}$; for example $\phi_k=2k\pi/L$ $(k=1,2,...,L)$ where L is the number of symbol levels.

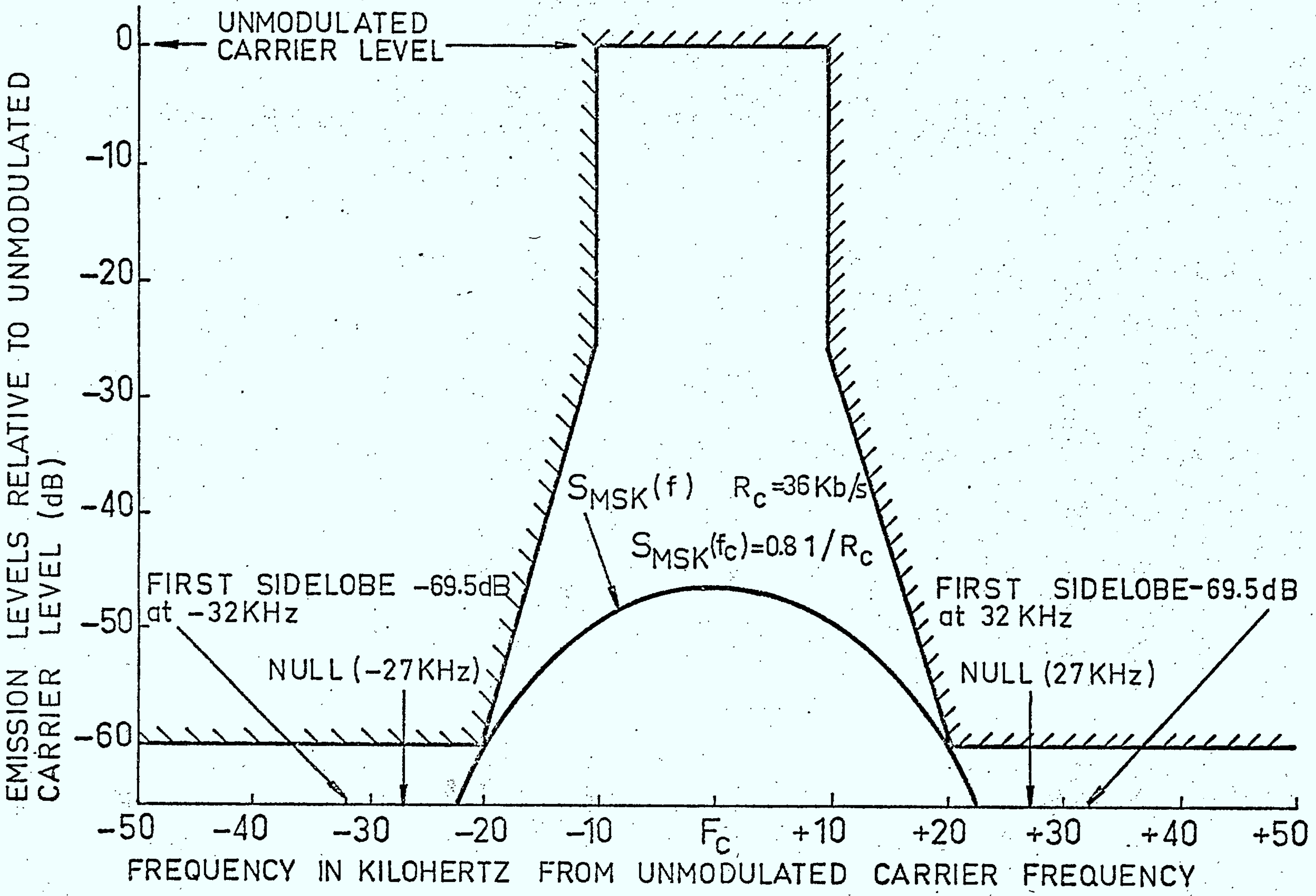

Fig. 2-7 Power spectral density for binary MSK signals at maximum bit rate.

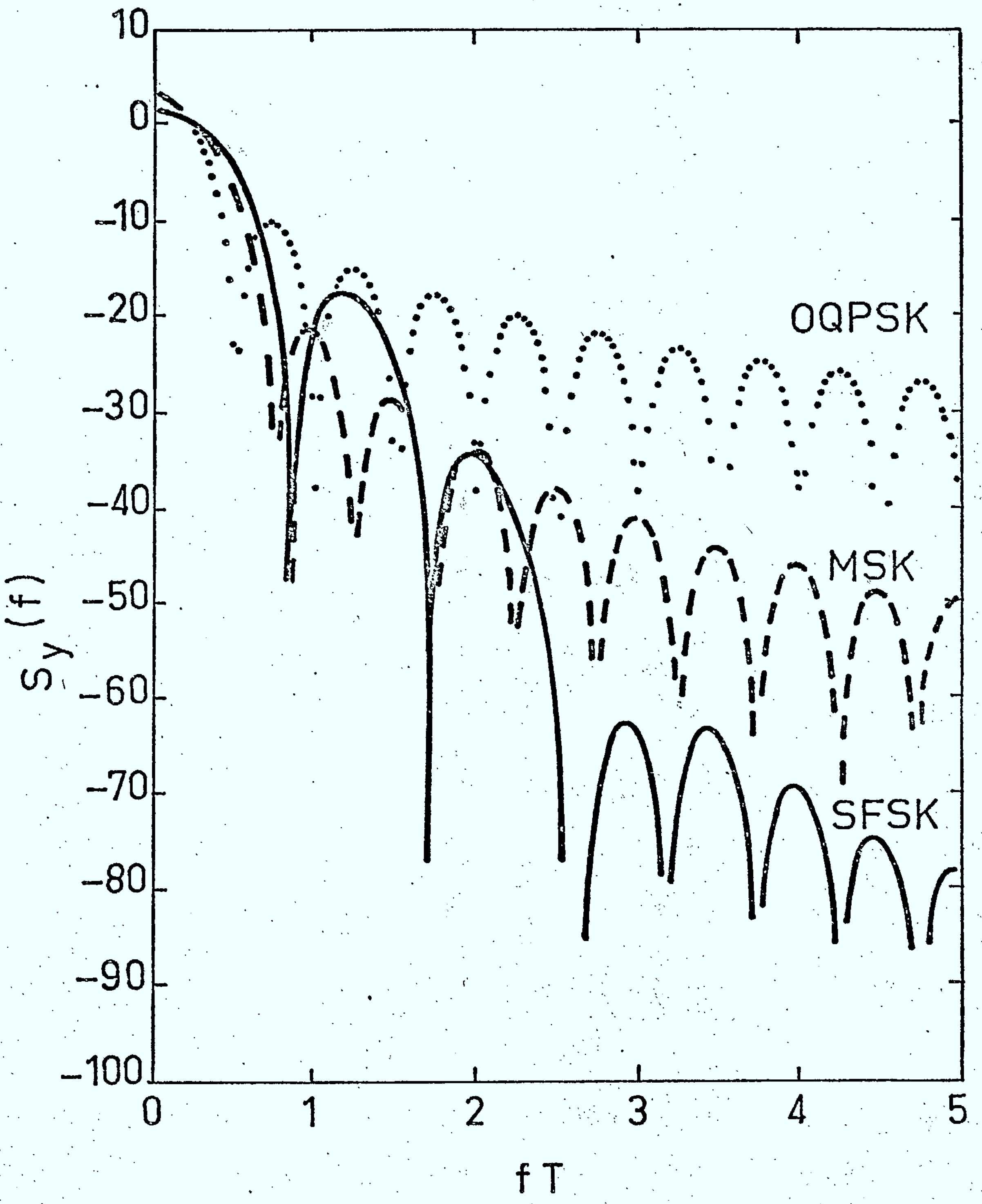

Fig. 2-8 FM spectra for MSK, OQPSK and SFSK signals (after Simon [S4]).

Eqn. (2-19a) may be rewritten as follows:

$$s(t) = \sqrt{2P}\{[\sum_{k=-\infty}^{\infty} b_k\, g(t-kT)]\cos \omega_c t + [\sum_{k=-\infty}^{\infty} c_k\, g(t-kT)]\sin \omega_c t\} \qquad (2\text{-}19b)$$

where

$$b_k = \cos \phi_k \qquad (2\text{-}19c)$$

$$-c_k = \sin \phi_k \qquad (2\text{-}19d)$$

$$\phi_k = 2\pi k/L \qquad (2\text{-}19e)$$

Thus, conventional PSK consists of quadrature PAM signals. For binary PSK ϕ_k=0 or π, in which case $b_k=\pm1$ and $c_k=0$. The resulting signal s(t) is identical to PAM with rectangular baseband pulses, and the power spectral density of s(t) in (2-19) is given by the response $S_b(f)$ in Fig. 2-3.

Optimal detection of conventional PSK on Gaussian white noise channels involves synchronous demodulation of the quadrature carriers sin $\omega_c t$ and cos $\omega_c t$, processing of each of these demodulated signals by filters matched to g(t), sampling of the filtered outputs and optimal decision making [L1,W1]. Because the optimum decision device compares ratios of numbers from the demodulated quadrature carriers, PSK on Rayleigh fading channels need not be restricted to binary.

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 [L1,W1]; if the detected phase of a symbol is incorrect, the (differentially) detected phase of the following symbol is also likely to be in error. A transmitter power increase of not more than 3dB reduces the DPSK error rate to that obtainable using PSK. DPSK signal spectra are like PSK spectra.

II-6 Interference Effects

Our earlier report [D1] derives in detail the signal-to-noise plus interference ratio at the output of the filters in Fig. 2-5 and 2-6.

Define

$$y(t) = x(t) + n(t) \tag{2-20}$$

where x(t) is the output due to the desired data signal, and n(t) is due to noise plus both co-channel and/or adjacent-channel interference. The output signal-to-noise-plus-interference ratio, SNR is as follows:

$$\begin{aligned} SNR &= E[x\ ]/E[n\ ] \\ &= [(N/S) + \sum_v (I_v/S) + \sum_d (I_d/S)]^{-1} \end{aligned} \tag{2-21}$$

In (2-21) S, N, I_v and I_d represent energy resulting from the transmitted data signal, Gaussian noise of power spectral density $N_o/2$, FM voice signal interference, and interference from other data signals. In this work, we assume that all receiver filter characteristics H(f) are matched to the transmitted modulator pulse shapes $G_T(f)$, (i.e. $H(f)=G_T^*(f)$) where we specifically include bit period T; note that $G_T(f)$ is the Fourier transform of modulator pulse $g_T(t)=g(t)$. We also assume no FM voice interference. These restrictions are not essential, although matched filtering maximizes SNR in white noise, and accurate results for voice interference require a reliable model for the transmitted FM voice signal spectrum [D1]. Thus,

$$N/S = N_o/2P_sT \tag{2-22}$$

$$I_d/S = \sum_d (P_d/P_s) C_d(\Delta T, g) \tag{2-23}$$

where P_s and P_d denote the received power in the desired and interfering data signals, respectively, and C_d is the interference function which depends on modulator pulse shape g(t) and the product ΔT, where Δ is the separation between adjacent channels. (Typically, Δ=30 kHz.) Function C_d is as follows:

$$C_d(\Delta T,g) = (2T)^{-1}\int_{-\infty}^{\infty}|G_T(f-\Delta)|^2|G_T(f)|^2 df \tag{2-24a}$$

$$= (1/2T^2)\int_{-\infty}^{\infty}|G_T(\lambda-\Delta T)|^2|G_T(\lambda)|^2 d\lambda \tag{2-24b}$$

$$= \frac{1}{T}\int_0^{\infty}\rho_g{}^2(\tau)\ \cos\ 2\pi\Delta\tau d\tau \tag{2-24c}$$

$$= \int_0^{\infty}\rho_g{}^2(x)\ \cos\ 2\pi\Delta T x dx \tag{2-24d}$$

and

$$\rho_g(\tau) = \int_{-\infty}^{\infty} g_T(t)\, g_T(t-\tau) dt \tag{2-25}$$

These results apply to binary PAM and PSK signals, and to MSK signals as well if the bit period T is replaced by 2T, and if C_d is multiplied by the factor two [D1,K2]. For MSK P_s and P_d includes the total received power, as for PAM and PSK.

Summarized below are functions $C_d(\Delta T)$ for binary PAM, PSK, MSK and SFSK signals [D1]:

For raised cosine PAM signalling with α=0,

$$|G_T(\lambda)|^2 = \begin{cases} 0 & |\lambda| \geq 1/2 \\ T & |\lambda| < 1/2 \end{cases} \tag{2-26}$$

Thus, C_d=0 for $|\Delta T| \geq 1$, and for $|\Delta T| < 1$ (without loss in generality $\Delta > 0$)

$$\begin{aligned} C_d &= (1/2T^2)\int_{\Delta T-\frac{1}{2}}^{\frac{1}{2}} T^2 d\lambda \\ &= (1-\Delta T)/2 \end{aligned} \tag{2-27}$$

For raised cosine signalling with α=1,

$$|G_T(\lambda)|^2 = \begin{cases} 0 & |\lambda| \geq 1 \\ T\left[\frac{1+\ \cos\ \pi\lambda}{2}\right] & |\lambda| < 1 \end{cases} \tag{2-28}$$

For $|\Delta T| \overline{>} 2$, $C_d=0$. For $\Delta T<2$

$$C_d = \frac{1}{8} \int_{\Delta T-1}^{1} (1+ \text{cox } \pi\lambda)(1+ \cos \pi(\lambda-\Delta T))d\lambda$$

$$= \frac{1}{8} (1- \frac{\Delta T}{2})(2+ \cos \pi\Delta T)+ \frac{3}{16\pi} \sin \pi\Delta T \tag{2-29}$$

For rectangular-pulses (conventional PSK)

$$g_T(t) = \begin{cases} 1/\sqrt{T} & |t| \overline{<} T/2 \\ 0 & |t|>T/2 \end{cases} \tag{2-30}$$

and

$$\rho_g(\tau) = \begin{cases} 1- \frac{|\tau|}{T} & |\tau| \overline{<} T \\ 0 & |\tau|>T \end{cases} \tag{2-31}$$

$$C_d = \int_0^1 (1-x)^2 \cos (2\pi\Delta Tx)dx$$

$$= \frac{1}{2(\pi\Delta T)^2} [1- \frac{\sin 2\pi\Delta T}{2\pi\Delta T}] \tag{2-32}$$

For $\Delta T=0$, $C_d=1/3$.

With MSK regarded as staggered quadrature PAM signals,

$$g_{2T}(t) = \begin{cases} \frac{1}{\sqrt{T}} \cos \pi t/2T & |t|<T \\ 0 & |t|>T \end{cases} \tag{2-33}$$

and

$$\rho_g(\tau) = \begin{cases} \frac{1}{\pi} \sin \frac{\pi|\tau|}{2T} + (1- \frac{|\tau|}{2T}) \cos \frac{\pi|\tau|}{2T} & |\tau|<2T \\ 0 & |\tau| \overline{>} 2T \end{cases} \tag{2-34}$$

For SFSK $g_{2T}(t)=g(t)$ in (2-18) with $U=1/4$ and

$$\rho_g(\tau) = (1- \frac{|\tau|}{2T}) \cos \frac{\pi|\tau|}{2T} + \frac{1}{T} \int_0^{T-\frac{|\tau|}{2}} \cos (\frac{\pi t}{T} - 2U \sin \frac{2\pi t}{T} \cos \frac{\pi|\tau|}{T})dt \tag{2-35}$$

For MSK and SFSK, C_d is most easily determined numerically.

Figs. 2-9 and 2-10 show $C_d(\Delta T)$ for the various signalling schemes. These results are for optimum receivers. Use of suboptimum receivers results in increased values for C_d. For example, the dotted curve in Fig. 2-10 is for MSK detection using a more easily implemented single matched filter, sampled every T sec., as in Fig. 2-5, where [K3]

$$h(t) = \begin{cases} \cos(\pi t/T) & |t| \leq T/2 \\ 0 & |t| > T/2 \end{cases} \qquad (2\text{-}36)$$

The parameter U in (2-37) can be varied to reduce C_d for small values of Δ at the expense of an increase in C_d for larger Δ[E1].

Interference which results when the signal and interfering signals have different modulator formats can be determined [D1]. Crosstalk C_d decreases as ΔT increases at a rate determined by that data format with the <u>slowest</u> spectral roll-off rate [R1], and the mutual interference effects of two data signals with different modulator pulse shapes is identical [D1]; thus, if g(t) and h(t) are the modulator pulse shapes of the two data signals then

$$C_d(\Delta T) = \frac{1}{T}\int_0^\infty \rho_g(x)\,\rho_h(x)\cos 2\pi\Delta T x\,dx \Big/ \left[\int_0^\infty \rho_g(x)\,\rho_h(-x)\,dx\right]^2 \qquad (2\text{-}37)$$

Design a data signal with spectra to fit within emission characteristics is not altogether satisfactory in terms of reducing adjacent-channel interference. Rather, what is required are signals with fast spectral roll-off rates, which lead to rapidly decreasing values of C_d.

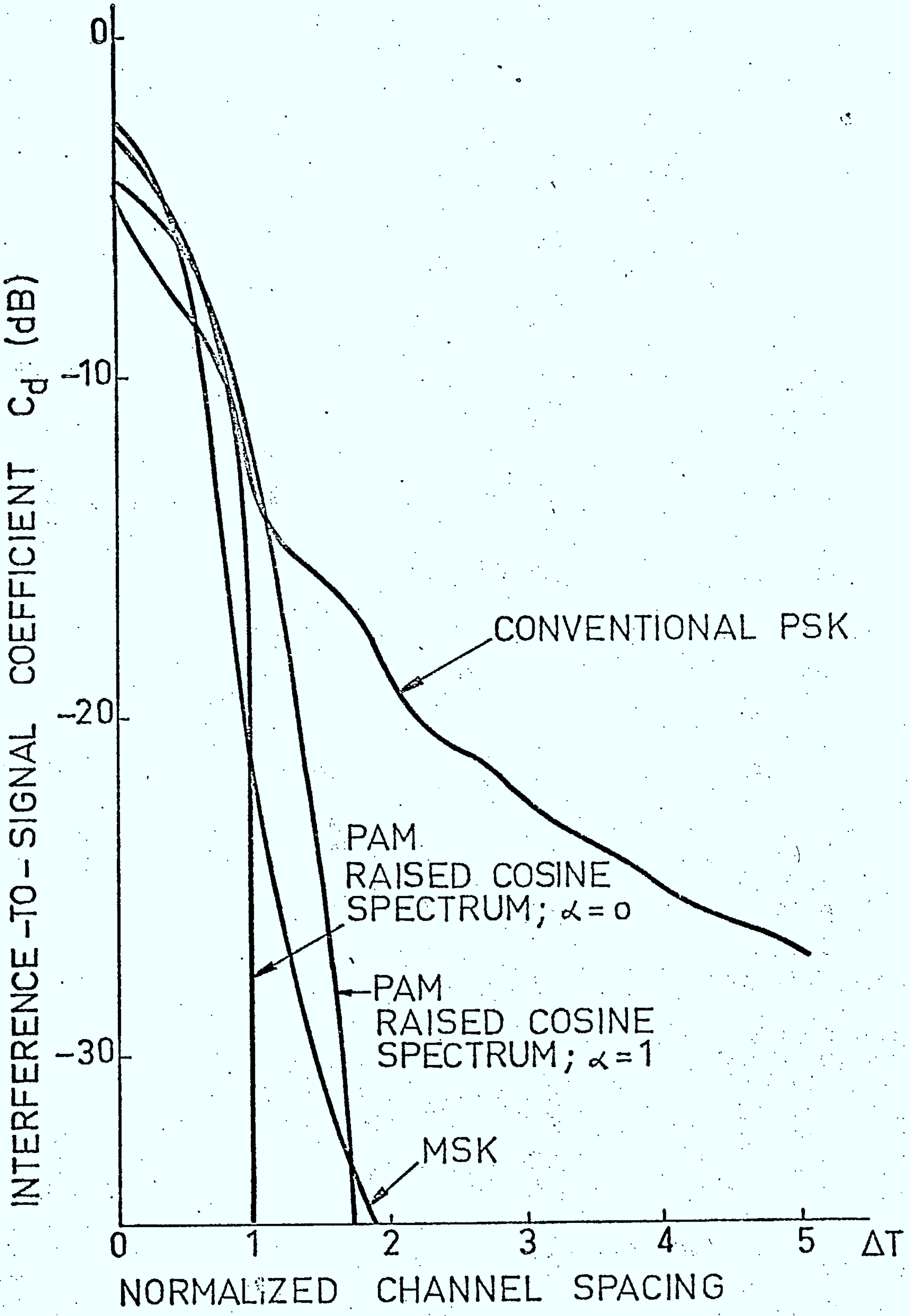

Fig. 2-9 Crosstalk function C_d vs. $\Delta/R = \Delta T$ for various data signals.

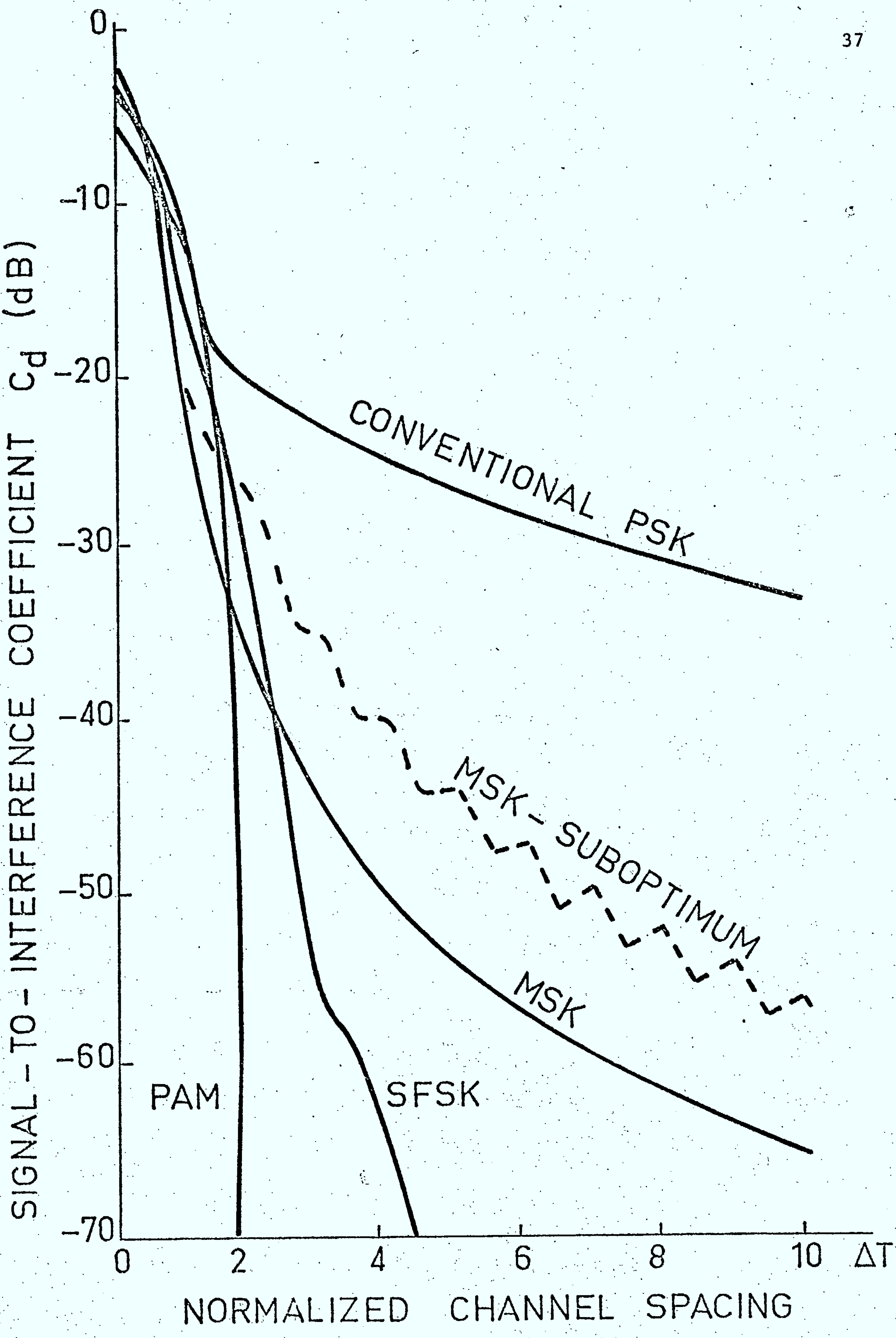

Fig. 2-10: C_d for MSM, conventional PSK, PAM, and SFSK.

II-7 Bit-Error Probability Determination

For binary antipodal signalling on white Gaussian noise channels, the bit-error probability p with optimum receivers is

$$p = \operatorname{erfc}(\sqrt{SNR}) \tag{2-38}$$

where

$$SNR = 2P/RN_o \tag{2-39}$$

$$\operatorname{erfc}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-y^2/2)dy \tag{2-40}$$

and P, R and $N_o/2$ denote, respectively, the received signal power, noise power spectral density and bit rate. Binary PAM and binary PSK are examples of binary antipodal signalling, as are binary MSK and 4-phase PSK when viewed as quadrature streams of binary PAM data, provided P is the power in each stream.

When the signal level is subject to Rayleigh fading, p must be average over these levels. The result is

$$p = (2 + \overline{SNR})^{-1} \tag{2-41}$$

where the overbar, omitted in what follows, represents an average.

Figure 2-11 shows p as given by (2-38) and (2-41) vs. SNR. The effects of Rayleigh fading are clear. Even though deep fades are relatively rare, when they occur the bit-error probability increases severely, and this increase dominates the average.

Non-white Gaussian noise, multi-level signalling or suboptimum receivers change the above values for p. However, in all cases, p decreases exponentially with SNR for Gaussian channels, and linearly for Rayleigh channels.

In land mobile channels, there are several additional factors affecting bit-error probability. In many cases, interfering signals with Rayleigh varying levels constitute the primary disturbance. When there are

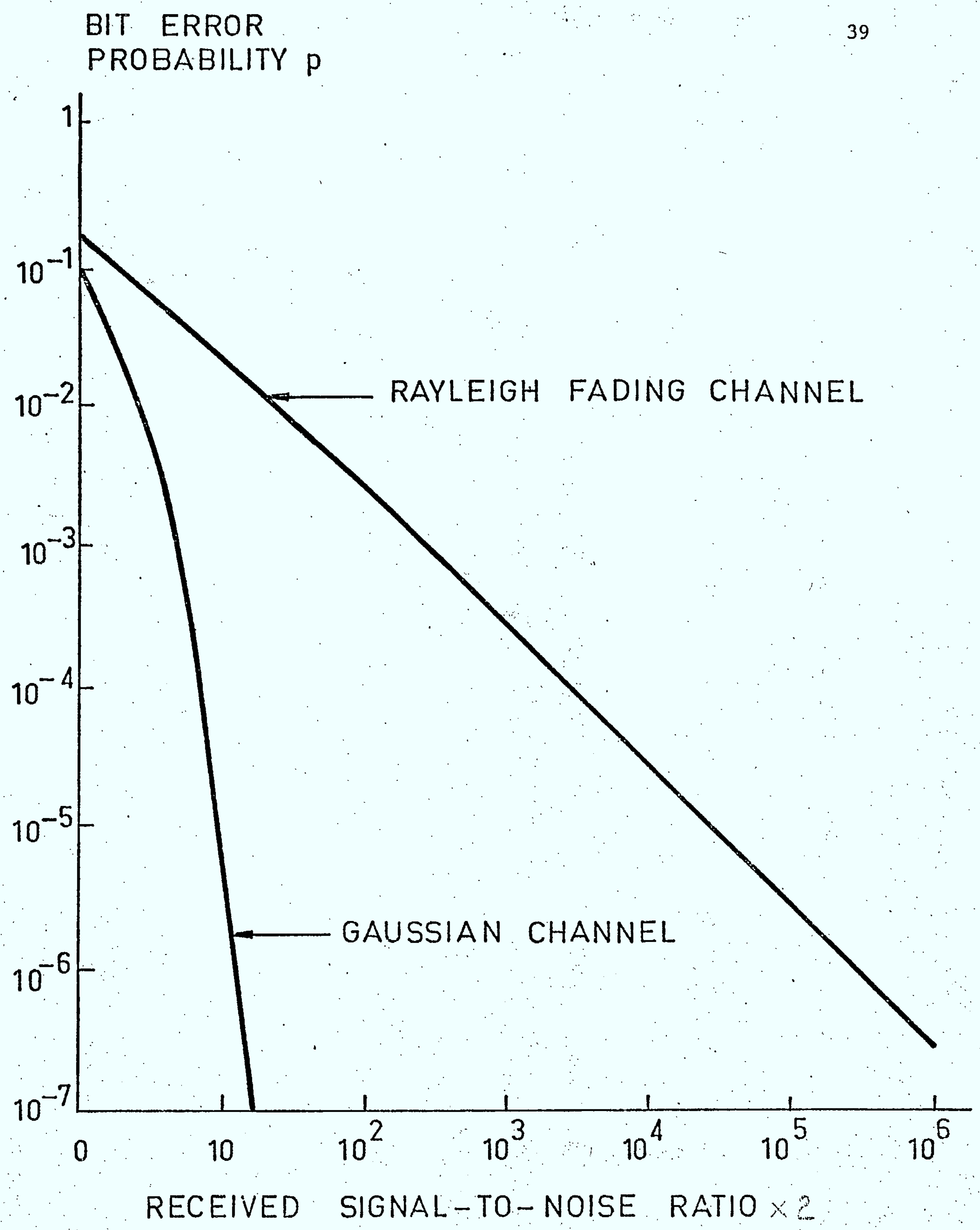

Fig. 2-11: Bit-error probability curves for Gaussian and Rayleigh channels; binary antipodal signalling.

many interferers of nearly equal effect, the central limit theorem may be used to approximate interference as (non-white) Gaussian noise. When there are few interferers, p must be averaged over both signal and interference levels. Also, vehicle movement causes changes in the mean signal level as a result of shadowing and distance separation. These latter changes would normally result in block error probabilities being averaged [E2]. In block averages, bit-error variations resulting from vehicle movement and Rayleigh fading cause burst errors [L1,P1,L3]. The effects of these bursts can be reduced by interleaving [L1].

II-8 Channel Sharing Alternatives

Various approaches exist for sharing the scarce radio spectrum. The usual approach involves its division into channels of 25 KHz or 30 KHz. Communication between a mobile and base can proceed when a channel to which both are tuned becomes free. Base stations can access all channels, whereas mobiles are normally able to access only a few. Mobile costs increase with the number of accessible channels.

Division of a radio coverage area into cells enables reuse of frequency channels [J2]. Channel reuse causes co-channel interference whose level depends on the distance between centres of cells assigned the same channels. In the cellular plan of Fig. 2-12, seven groups of channels are assigned (permanently) as shown. Assignment schemes also affect adjacent-channel interference levels [M1].

Channels may be assigned permanently to cells, or may be assigned dynamically in accordance with message traffic demands. Dynamic assignments require much more computer software to ensure that reuse separations are

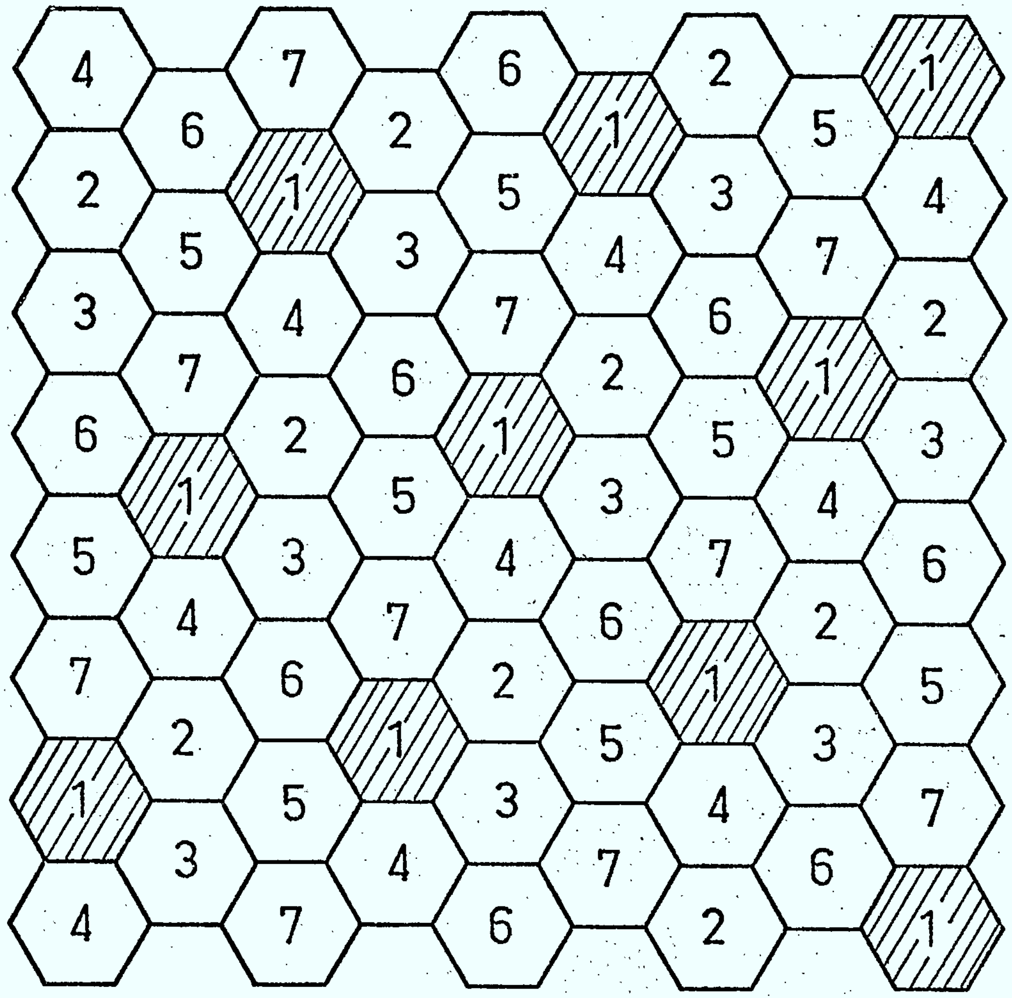

Fig. 2-12: Cellular System Plan.

maintained, but can improve spectrum efficiency, particularly at low traffic levels [J2].

An alternative to conventional frequency division multiplexing is spread spectrum (SSMA) [C5]. All users are assigned a unique code sequence which uses the entire frequency spectrum. This code provides for user identification, addressing, and interference suppression. Frequency diversity is also provided, since fading over a wideband channel of several hundred kHz is selective, and all frequencies are not subject to fading at the same time. The code sequence itself consists of narrowband pulse sequences, with the various pulses having different centre frequencies. Receivers for SSMA are currently more expensive than for conventional FDM systems, but mass production would decrease costs. At present it is not clear whether or not SSMA provides for better spectrum efficiency than does FDM. More analysis is required to make accurate performance comparisons for the two types of systems.

Various switching alternatives exist for land mobile radio systems. Line (or circuit) switching is the older, conventional strategy; a connection between transmitter and receiver is established and held for the duration of the call. A newer method, packet switching [K4,S5], involves segmentation of messages into fixed length packets, which are transmitted over the first available channel. Packet switching may enable more efficient utilization of the radio spectrum. However, new problems arise, including routing of packets, out-of-order packet arrivals, and overflow of reassembly buffers. Some comparisons of the costs and efficiency of line and packet switched systems exist but conclusive comparisons are not yet available.

II-9 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.

[A2] F. Amoroso, "Pulse and spectrum manipulation in the minimum (frequency) shift keying (MSK) format", IEEE Trans. Commun., vol. COM-24, pp. 381-384, March 1976.

[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] G.R. Cooper and R.W. Nettleton, "Spectral efficiency in cellular land-mobile communications: A spread-spectrum approach", School of Elec. Eng. Purdue Univ. Tech. Rpt. TR-EE 18-44, 1978.

[C3] J.K. Cavers, "Variable rate transmission for Rayleigh fading channels", IEEE Trans. Commun., vol. COM-20, pp. 15-22, Feb. 1972.

[C4] 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.

[C5] 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.W. Donaldson, "Frequency assignment for land mobile radio system: Digital transmission over land mobile channels - interference considerations", Report to Dept. of Communications, Ottawa, Canada, Jan. 1980.

[E1] R.E. Eaves and S.M. Wheatly, "Optimization of quadrature - carrier modulation for low crosstalk communication and close packing of users", IEEE Trans. Commun., vol. COM-27, pp. 176-185, Jan. 1979.

[E2] R.E. Eaves and A.H. Levesque, "Probability of block error for very slow Rayleigh fading in Gaussian noise", IEEE Trans. Commun., vol. COM-25, pp. 368-374, March 1977.

[F1] J.L. Flanagan, Speech Analysis, Synthesis and Perception, 2nd Ed., New York, N.Y.: Springer-Verlag, 1972.

[F2] 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.

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

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

[G3] G.J. Garrison, "A power spectral density analysis for digital FM", IEEE Trans. Commun., vol. COM-23, pp. 1228-1243, Nov. 1975.

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

[J1] 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.

[J2] W.C. Jakes, Microwave Mobile Communications. New York, N.Y.: Wiley, 1974.

[K1] R.S. Kennedy, Fading Dispersive Communication Channels. New York: Wiley, 1969.

[K2] I. Kalet, "A look at crosstalk in quadrature-carrier modulation systems", IEEE Trans. Commun., vol. COM-25, pp. 884-892, Sept. 1977.

[K3] I. Kalet and E. White, "Suboptimal continuous shift keyed (CSK) demodulation for efficient implementation of low crosstalk data communication", IEEE Trans. Commun., vol. COM-25- pp. 1037-1041, Sept. 1977.

[K4] L. Kleinroch, Queuing Theory, vol. 2: Computer Applications. New York: Wiley, 1976.

[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.

[L3] S. Lin, An Introduction to Error-Correcting Codes. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

[M1] J.J. Mikulski, "A system plan for a 900-MHz portable radio telephone", IEEE Trans. Veh. Technol., vol. VT-26, pp. 76-81, Feb. 1977.

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

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

[R1] B. Reiffen and B.E. White, "On low crosstalk data communication and its realization by continuous-frequency modulation schemes", IEEE Trans. Commun., vol. COM-26, pp. 131-135, Jan. 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] O. Shimbo, "General formula for power spectra of digital FM signals", Proc. IEEE, vol. 113, pp. 1783-1789, Nov. 1965.

[S4] 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.

[S5] M. Schwartz, Computer Communication Network Analysis and Design.

Englewood Cliffs, N.J.: Prentice-Hall, 1977.

[T1] T.J. Tjhung and P.H. Whitke, "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.: 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 SYSTEM PERFORMANCE CRITERIA

III-1 Desired Attributes of System Performance Criteria

Analysis, assessment and design of systems is ultimately based on how well a system accomplishes its tasks for a given cost. Objective evaluations require quantitative performance measures whose selection is an important aspect of system design and evaluation. Performance criteria should be accurate enough to adequately describe system behaviour and yet simple enough to facilitate analysis, design, assessment and optimization. Compromises between accuracy and simplicity are nearly always necessary.

Useful criteria for digital mobile radio communication systems include information throughput, cost, system reliability and either time delay or blocking probability. Bit-error probability or output signal-to-noise ratio is often used as a measure of reliability. These four are not independent but are related as described in this and subsequent chapters.

Other performance criteria which are of interest include system sensitivity to changes in various parameters, and channel occupancy which is discussed in Section III-4.

III-2 Delay and Blocking Probability

Consider a base station serving a region where messages are generated independently at an average rate λ per sec; message lengths are exponentially distributed with mean length μ^{-1} bits/message, and are served by one of m channels, and each channel transmits at rate R bits/sec as in Fig. 3-1. Calls which cannot be handled immediately are delayed and queued for transmission, normally on a first-come first-serve (FCFS) basis. This is the Erlang C call-

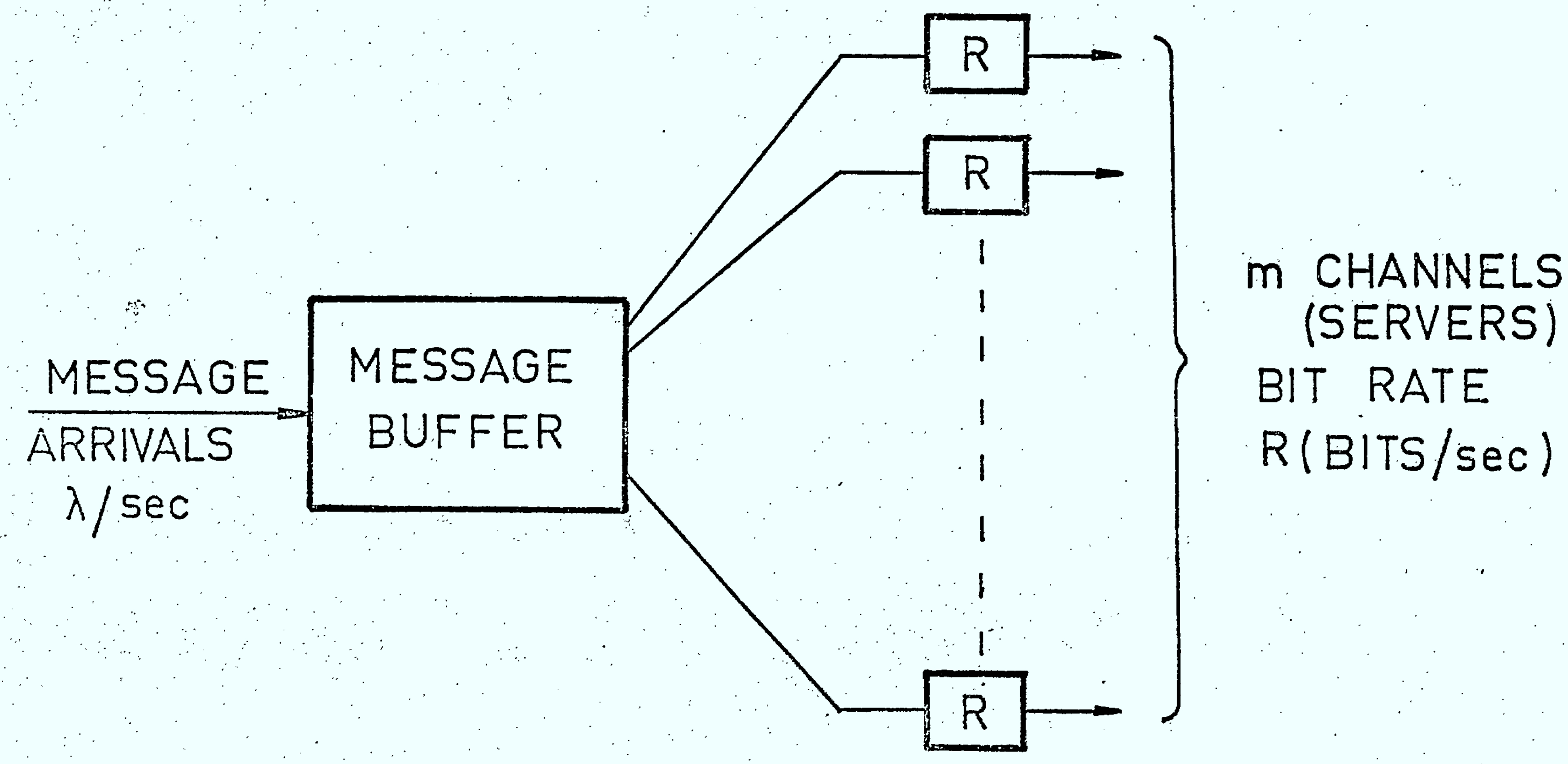

Fig. 3-1 An m-server system.

handling discipline, and approximates what would occur in non-real-time transmission of text or control information of the type described in Chapter 2.

Analysis is based on c_k, the probability of k different messages being in the system, either being served or waiting in a queue, where [K1,S1]:

$$c_k = \begin{cases} c_0\rho^k/k! & k\leq m \\ c_0(\rho/m)^k m^m/m! & k\geq m \end{cases} \tag{3-1}$$

$$\rho = \lambda/\mu R \tag{3-2}$$

and

$$c_0 = \{\sum_{k=0}^{m-1}(\rho^k/k!) + (\rho^m/m![1-(\rho/m)])\}^{-1} \tag{3-3}$$

The delay D between generation of a message and completion of transmission is

$$D = (1/\mu R) + (P_m/\mu mR[1-(\rho/m)]) \tag{3-4}$$

where P_m is the probability that the system contains m or more messages;

$$P_m = \{c_0/[1-(\rho/m)]\}\rho^m/m! \tag{3-5a}$$

Note that

$$c_0/[1-(\rho/m)] = \{\rho^m/m! + [1-(\rho/m)]\sum_{k=0}^{m-1}\rho^k/k!\}^{-1} \tag{3-5b}$$

In (3-4), $1/\mu R$ is the average message transmission time. The second term is the product of the average waiting time and the probability P_m that m or more messages reside in the system, in which case all channels are busy. The results actually apply to any service discipline not dependent on message lengths [K1].

Often, one considers the normalized delay d, which describes D relative to the total service capacity μmR; thus

$$d = (\mu m R) D$$

$$= m + P_m/[1-(\rho/m)] \tag{3-6}$$

Fig. 3-2 shows d in (3-6) vs $\rho/m = \lambda/\mu m R$ which is the traffic carried λ relative to the system service capacity. As m increases for any given value of ρ/m, d increases. Clearly, relatively few high capacity servers (channels) are preferred to a larger number of low capacity servers, when ρ/m is fixed.

The effects of scaling both λ and R by the same factor a are readily determined from (3-4) and (3-5). For m fixed, ρ and therefore P_m depend only on the ratio $\lambda/\mu R$; thus

$$D(m, a\lambda, aR) = D(m, \lambda, R)/a \tag{3-7}$$

This is a significant result, which says that uniform scaling of both λ and R actually reduces D by the scale factor, for any number of servers m.

An alternative call-handling discipline is Erlang B[K1,J1], where calls arriving when all channels are busy are discarded (in effect, the buffer in Fig. 3-1 has zero capacity). Voice traffic is said to obey such behaviour [J1]. The usual assumption is that blocked calls never return.

Under Erlang B,

$$b_k = b_0 \rho^k / k! \qquad k \leq m \tag{3-8}$$

where

$$b_0 = \left[\sum_{k=0}^{m} \rho^k / k! \right]^{-1} \tag{3-9}$$

The blocking probability P_B is

$$P_B = b_m \tag{3-10}$$

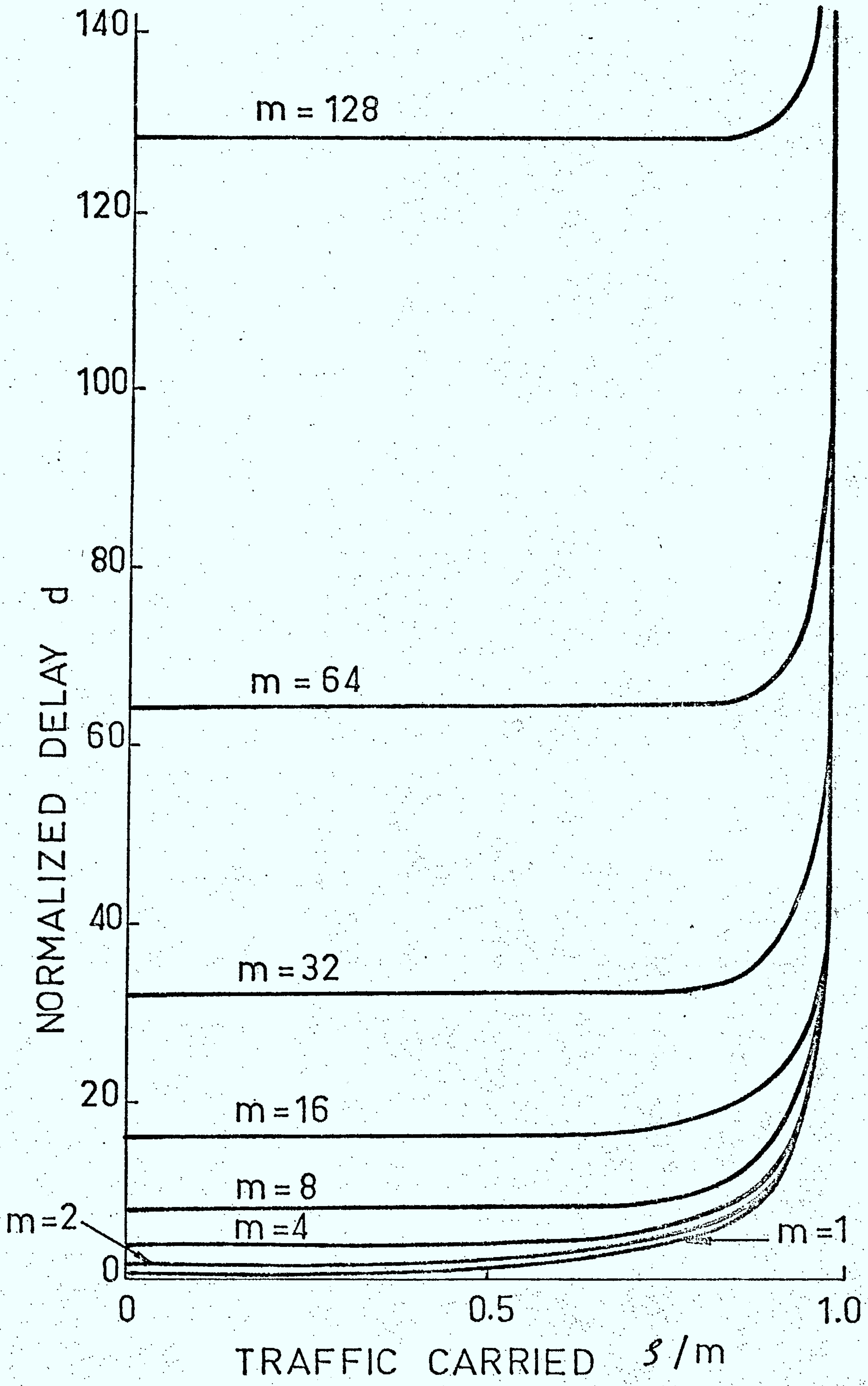

Fig. 3-2a Queuing delay (normalized) for an m-server system vs. traffic carried ρ/m. a) d

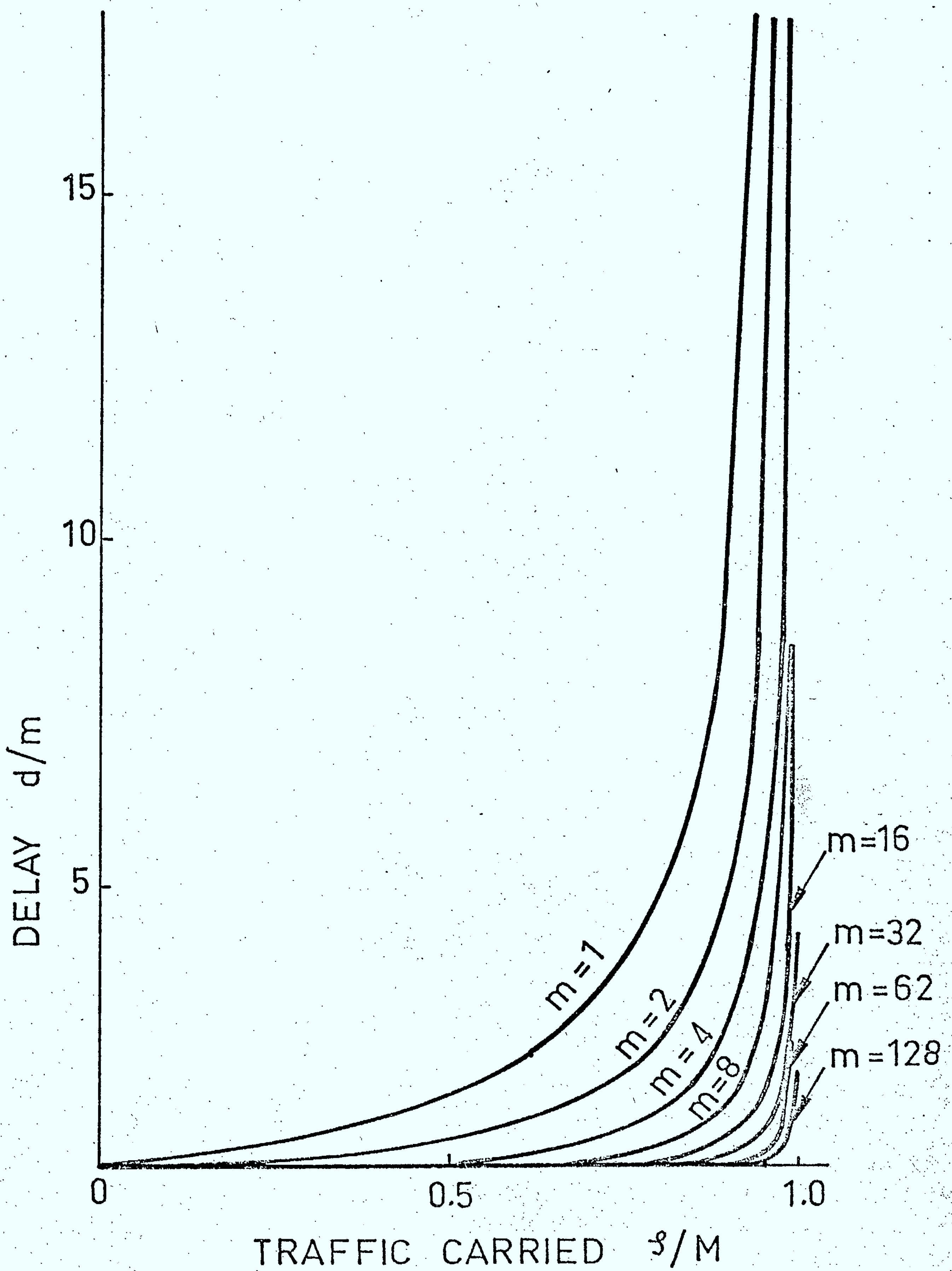

Fig. 3-2b Queuing delay (normalized) for an m-server system vs. traffic carried ρ/m. b) d/m

Fig. 3-3 shows P_B as given by (3-10) vs $(\rho/m)(1-P_B)$, which is the traffic carried $\lambda(1-P_B)$ relative to the system service capacity μmR. It is clear that for a fixed level of traffic carried, many low capacity channels are preferable to a few high capacity ones, when P_B is the performance criterion.

Other call-handling disciplines have been proposed to model system behaviour, including those in a recent study [N1] where calls either remained in the queue (with probability θ) or left immediately (with probability $1-\theta$). Parameter θ provides for a range of disciplines; however, the study supports the general utility of the Erlang B and C disciplines, corresponding to $\theta=0$ and $\theta=1$, respectively.

III-3 System Throughput

System throughput is the total number of bits successfully transmitted, often normalized with respect to time, total system bandwidth B, and/or coverage area A. When normalization is with respect to B, throughput is often referred to as spectrum efficiency (in bits/sec/Hz). We define throughput Q as follows:

$$\begin{aligned} Q &= [\Lambda A/\mu B][1-P_B] \\ &= \beta[1-P_B] \end{aligned} \qquad (3\text{-}11)$$

where Λ is the message arrival rate per unit of coverage area, and μ the average message length.

The number of mobiles is reflected only in the value of Λ which equals the product of mean call attempt rate per mobile λ_v and the number of mobiles.

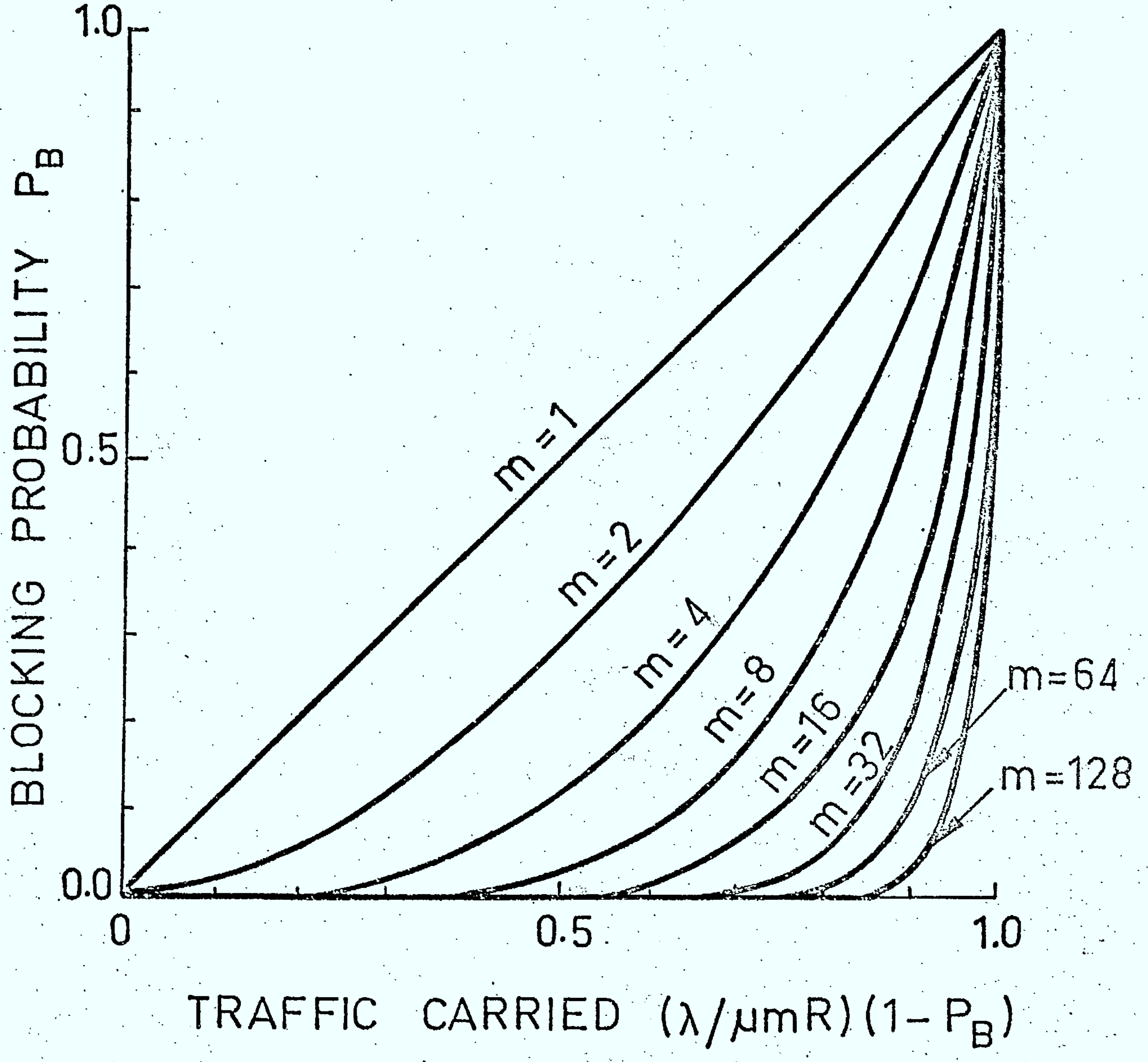

Fig. 3-3 Blocking probability P_B for an m-server system vs. traffic carried, $\rho(1-P_B)/m$.

When mobiles are distributed unevenly over the coverage area, then

$$\Lambda A = \lambda_v \int_A n(x,y)\,dxdy \tag{3-12}$$

where n(x,y) represents the density of mobiles in the neighbourhood of x,y.

In an Erlang B environment, P_B is given by (3-10). For Erlang C, $P_B=0$. Figure 3.4 shows Q vs. m for various P_B and d values.

III-4 Channel Occupancy

Channel occupancy may be defined either as the average number of channels occupied, per cell, or as the probability that exactly k of the m available channels are occupied.

If the latter definition is employed, then channel occupancy for each value of k is given by c_k in (3-1) for Erlang C call-handling or by b_k in (3-8) for Erlang B call-handling. Figure 3-5 shows p_k vs. ρ/m and k for Erlang B call-handling, for m=2 and m=8. When k=m, $p_m=P_B$. It is seen that as the traffic offered ρ/m increases, the probability is that an increasing number of channels will be occupied.

Figure 3-6 shows P_B vs. ρ/m, which allows Fig. 3-5 or (3-8) to be used to determine p_k vs. traffic carried $\rho(1-P_B)/m$.

Comparison of (3-1) and (3-8) shows that for k<m, occupancies p_k for Erlang C equal those for Erlang B times the scale factor c_0/b_0. For k=m, the scale factor is $c_0/b_0\ (1-\frac{\rho}{M})$. Figure 3-7 shows these scale factors vs. ρ/m, for m=2 and m=8. One sees that as ρ/m increases, Erlang C call-handling has an increasingly higher value of p_m and increasingly lower values for p_k (k<m) than does Erlang B. The reason for this behaviour is that Erlang C queues calls, while Erlang B does not.

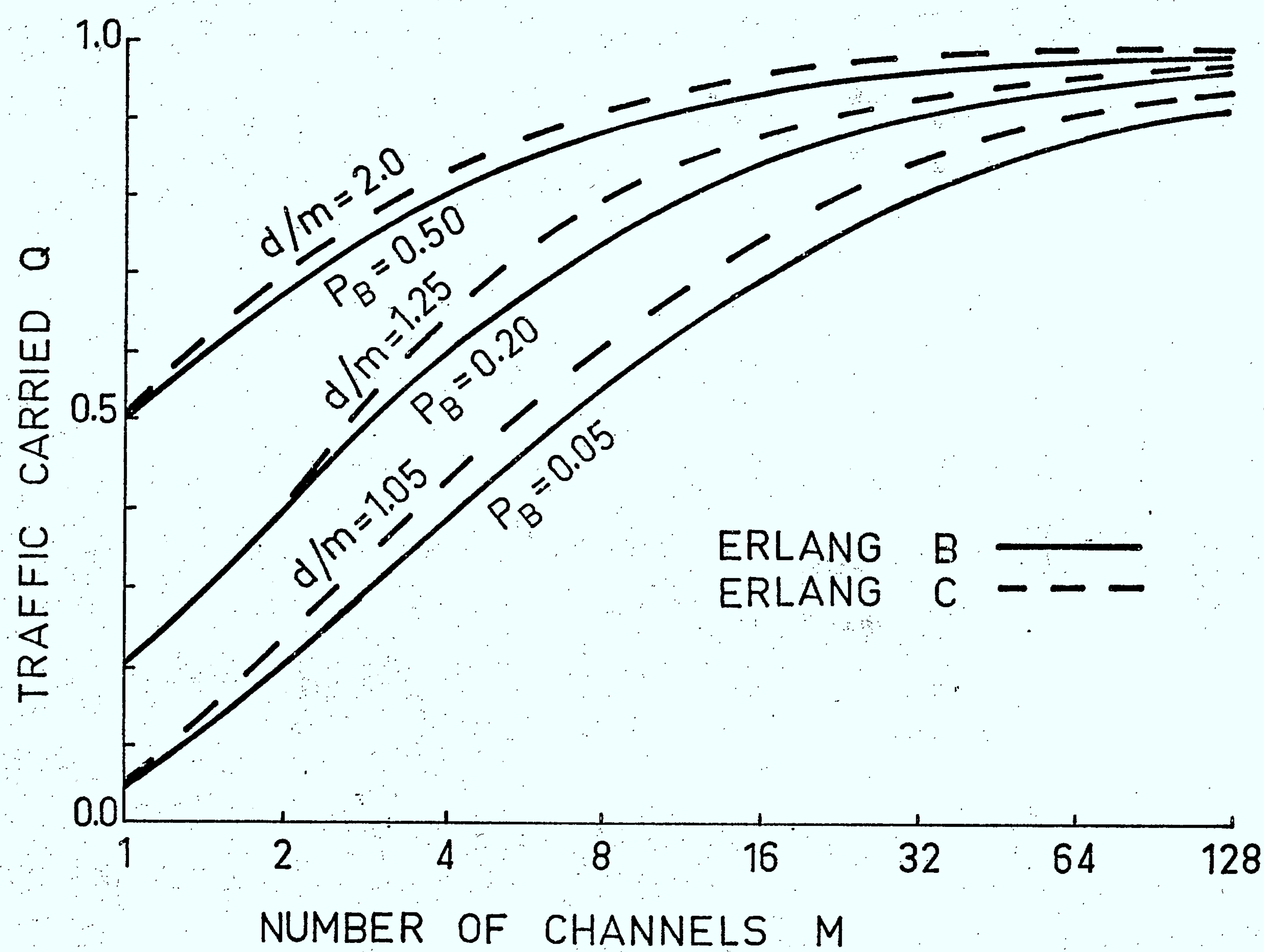

Fig. 3-4 Traffic carried (per channel) vs. number of channels m.

Fig. 3-5 Channel occupancy P_K vs. ρ/m for Erlang B call handling.

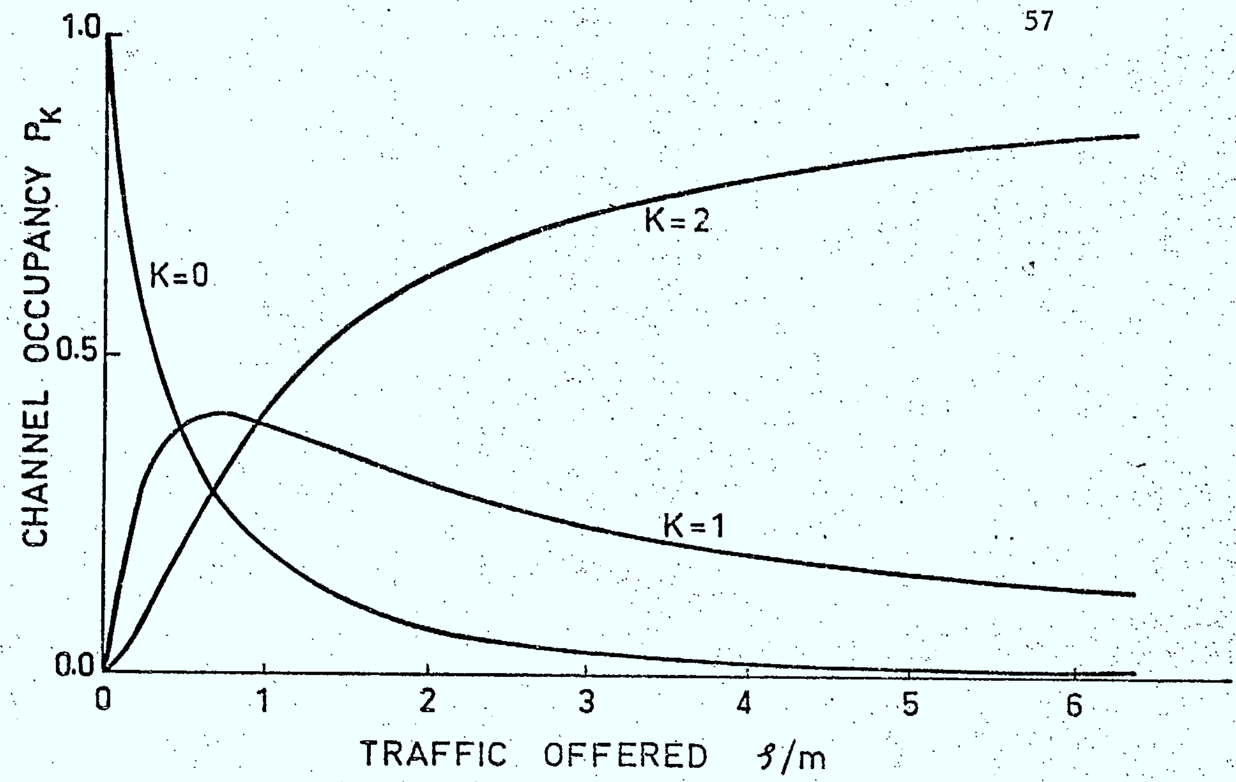

Fig. 3-5a m=2

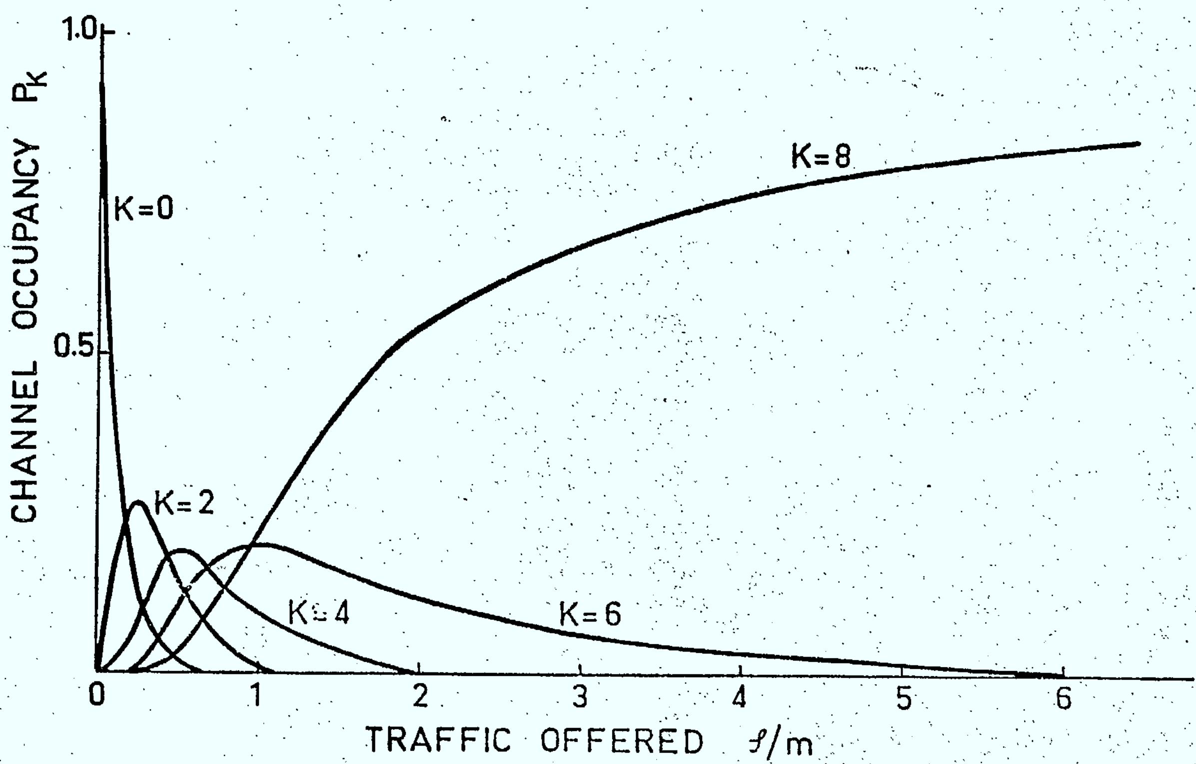

Fig. 3-5b m=8

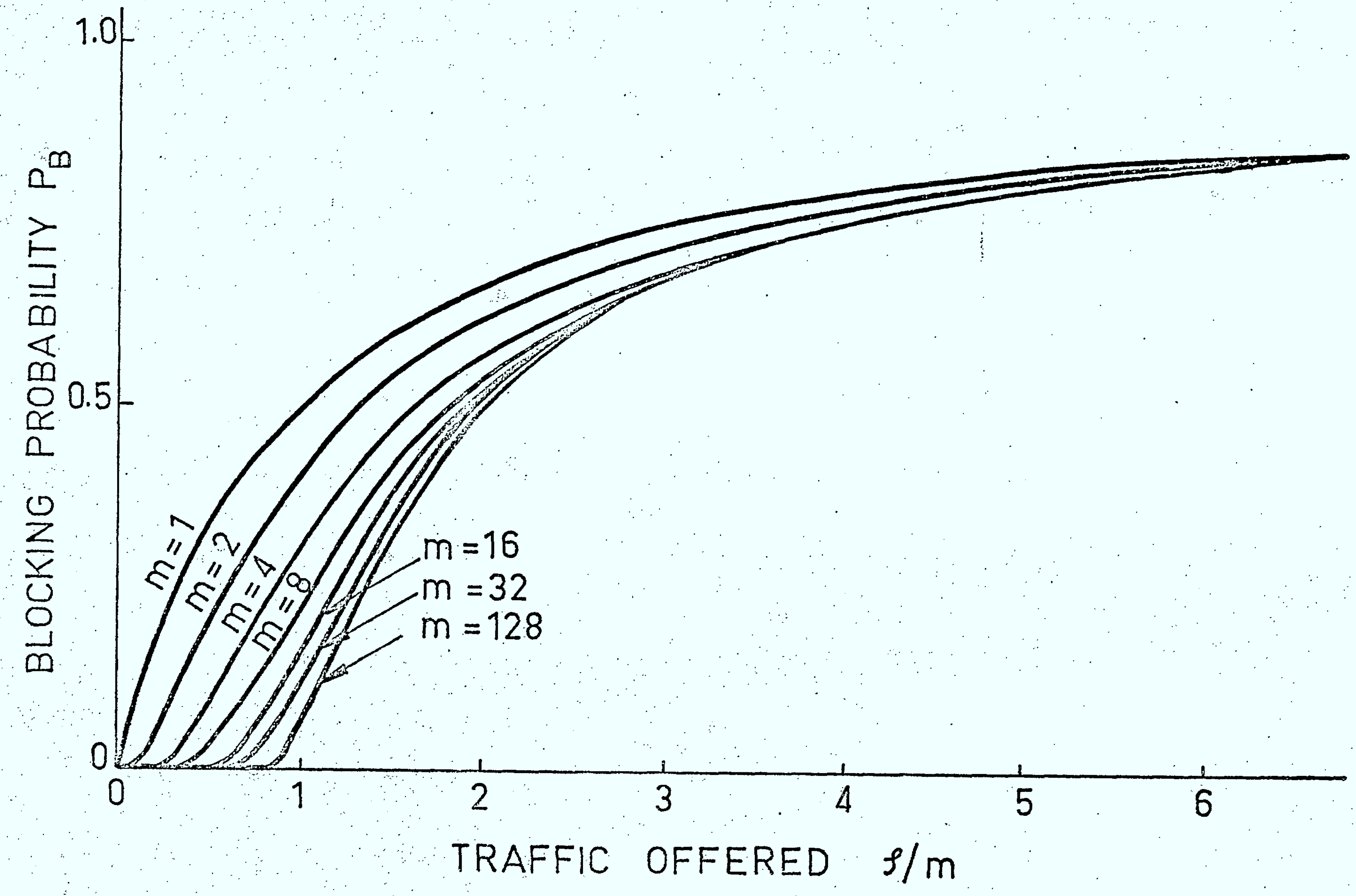

Fig. 3-6 Blocking probability P_B vs. traffic offered, ρ/m.

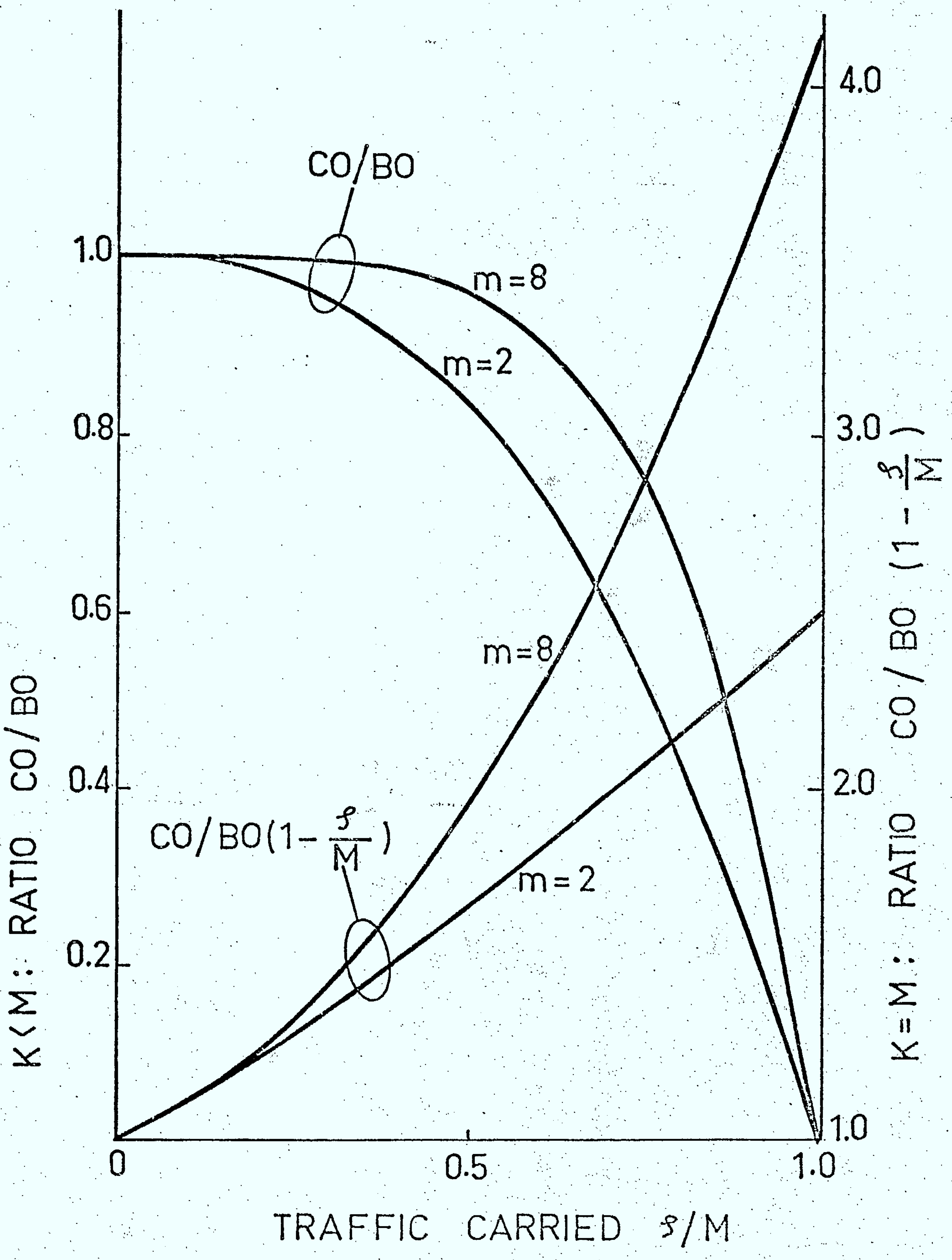

Fig. 3-7 Ratio of $\{p_k\}$ vs. ρ/m, Erlang C / Erlang B; $m=2$ and $m=8$.

If occupancy ϕ is defined to be the average number of channels occupied at any given moment, then it is shown in Appendix B that ϕ=Q, the system throughput. Thus, Figs. 3-2 and 3-3 can be viewed as displaying ϕ vs. either P_B (for Erlang B) or d (for Erlang C). Figure 3-8 shows ϕ for Erlang B vs. traffic offered. One sees that as d or P_B increase, ϕ also increases, as expected. This definition for ϕ also equates ϕ to the probability that any given channel will be occupied, assuming that all channels are equally loaded.

Again, the reminder that these results apply to a system where m is fixed, or at least changes sufficiently slowly that steady-state queueing theory is applicable [K1].

III-5 Reliability

Precise definitions of system reliability are elusive, not only in the present context but in other system design contexts as well [D1,W1,F1]. We define reliability as the output signal-to-noise-plus-interference ratio SNR, calculated as in the following convenient way. The desired signal, as well as any interfering signals from the same radio coverage cell are assumed to be located at a median radius $r/\sqrt{2}$, where r is the cell radius. (Note that approximately half of the area lies between $r/\sqrt{2}$ and r and that the remainder lies between 0 and $r/\sqrt{2}$.) The receiver is assumed to be at the cell's centre. Interfering signals are assumed to originate from the centre of any other cell. Thus, for desired and interfering signals from the same cell, the ratio $P/P_k=1$, where P_k and P denote interfering and desired signal powers, respectively. For an interfering signal from an immediately adjacent cell, $P/P_k=(\sqrt{6})^n$, where n is the propagation factor (nominally, $n \approx 3 \cdot 5$). If inter-

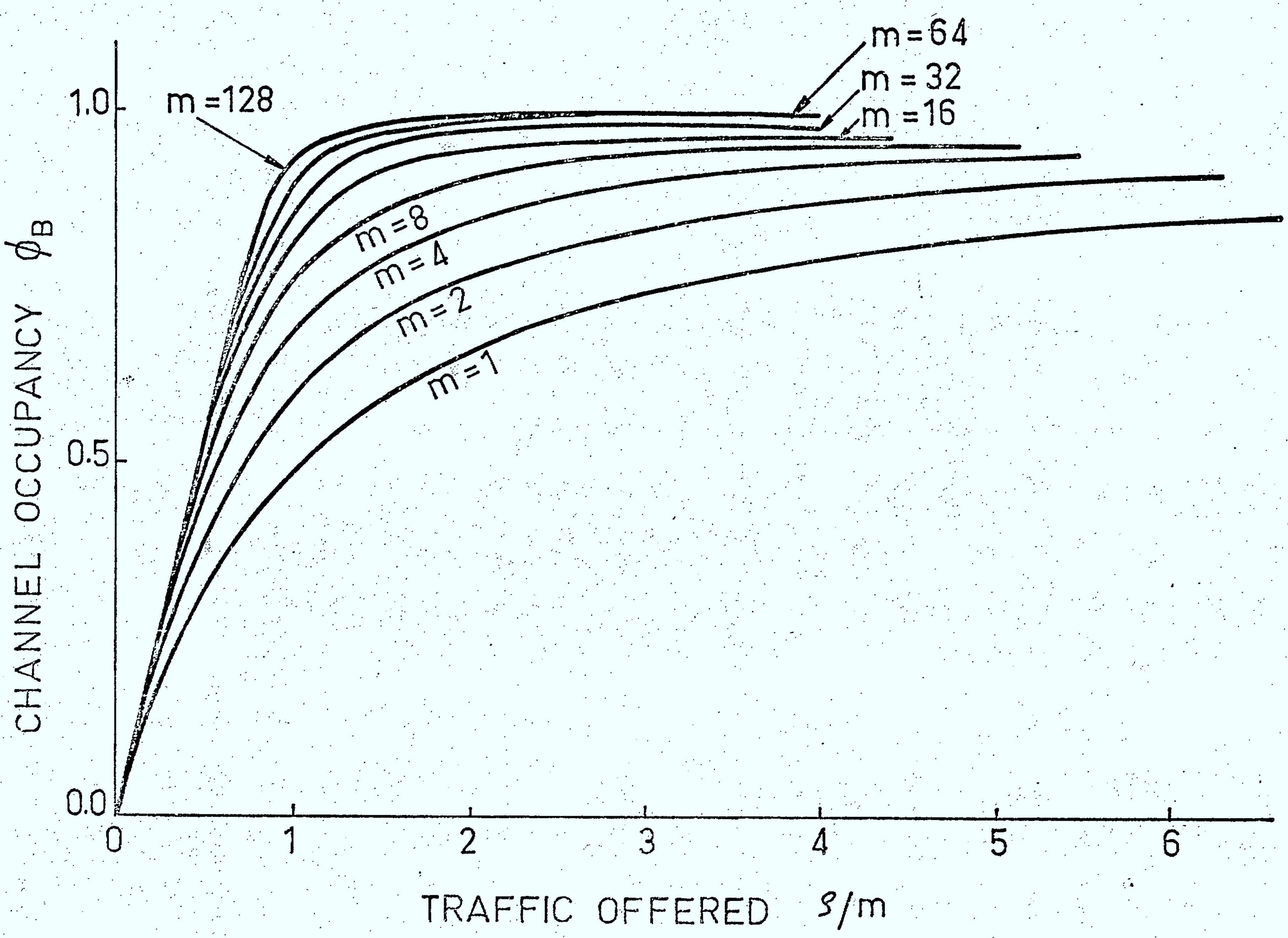

Fig. 3-8 Channel occupancy ϕ_B vs. traffic offered ρ/m; Erlang B call-handling.

ference is the predominant source of disturbance, which we assume, then

$$\text{SNR} = \left[\sum_k \phi_k \, (P_k/P)^n \, C(\Delta/R)\right]^{-1} \tag{3-13}$$

where (P/P_k) depends only on the relative cell locations of the signal and interferers. Channel occupancy ϕ_k is for channel k.

In (3-13) all interferers have the same modulation format, and Δ_k is the channel spacing between the signal and interferer; normally Δ_k is some integer multiple of a basic spacing Δ. If some interferers have different modulation formats, then function C will be different for these interferers.

The above approach, in effect, averages the effects of Rayleigh and lognormal fading, but explicitly includes spatial separation and channel occupancy effects, as well as the effects of the propagation factor n, modulation format, channel separation and data rate. Also implicitly included (in ϕ) are the number of channels m and the traffic level ρ.

Determination of the actual (uncoded) bit error probability p would require averaging of p with respect to Rayleigh and lognormal fading of both the signal and interferers, as explained in Section 2-7.

III-6 Cost Considerations

System cost is closely related to complexity. Costs can be expressed as a sum of mobile costs and base station costs.

In the case of base facilities, cost increases with the number of base stations. An increase from single to multiple base systems would normally involve addition of facilities for channel scanning, as well as for vehicle location in order that the base nearest a mobile could be assigned thereto. In fact, the addition of facilities to locate mobiles may be a useful one in any case, since mobile location is all that is required in many instances. With location equipment in place the actual data transmissions for location only could be minimized. The accuracy required of location equipment increases as the coverage area decreases, but the signal-to-noise requirements for data transmission become less. In cellular systems where frequency channels are not reused, location need not be very accurate, and could probably be based on received signal strength. Various authors [J1,O1,R1] present good discussions of the vehicle location problem.

The algorithms used to assign channels to cells affects cost (or complexity) of base facilities. Dynamic assignment schemes involve considerable hardware and software to store current assignments, and to provide co-ordination and calculations for assigning available channels. Fixed assignment schemes, where each channel is permanently assigned to a specific cell, cost less to implement, and for heavier traffic, actually seem more effective.

Mobile costs include a fixed cost, plus a variable cost which increases with the number of channels per mobile. A mobile with more than one channel would often require channel scanning equipment, which would add to its cost.

Typical costs (in 1979 dollars, Canadian) are $10,000.00 for bases, excluding location equipment and land lines. A single-channel mobile costs, typically, $1,500.00. Clearly, bases cost more than mobiles. However, the expected useful life of bases is often longer, and their cost is spread over the number of mobiles served. Thus, schemes which increase the number of channels on which each base can communicate may actually be cost effective overall, since the number of channels per mobile may be reduced for any given throughput level.

Costs of both mobile and base transmitters and receivers would tend to increase with the overall system bandwidth B. Transceiver costs would be about the same for the various modulation schemes discussed in Chapter 2, although some cost savings would result if optimum MSK coherent detection were replaced by suboptimum incoherent detection. However, the SNR penalty of approximately 6dB would be regained only at some additional cost. The SFSK receiver electronics would involve some cost increases above those for the other modulation schemes. However, the cost per channel or per bit transmitted could actually be less, because of reduced sensitivity to adjacent-channel interference, which would permit higher throughputs.

Use of channel codes for error correction and/or detection would increase receiver costs, but would also very likely improve throughput. Thus, the cost per bit transmitted may decrease.

Costs are seen from the point of view of those who are to pay them. If one entity provides both base and mobile facilities, overall costs would tend to be minimized. If bases and mobiles were supplied by different entities, each would tend to minimize its own costs, possibly at the others' expense. However, costs might still fall because of increased competition.

Technology changes come rapidly, particularly in digital hardware and software. The result is a continued change in cost and implementation complexity. Benefit/cost ratios also change substantially over time. Ever-increasing numbers of vehicles and increasing fuel costs make the use of mobile communication facilities increasingly attractive. Thus, channel control schemes or transceiver implementations which are presently uneconomical, may be economical in the future.

A detailed study of mobile system costs is beyond our scope. In this report costs are addressed, indirectly, by noting cost-determing parameter values and operational features. Much literature is available which deals with communication system costs [H1,L1,S1,Y1].

III-7 Overall Approach to System Analysis and Design

Our objective at this point is to devise a rational and viable approach to system analysis and design which will include all performance criteria and important system parameters.

The following approach is adopted:

1. The signal-to-noise-plus-interference ratio SNR is constrained to a specified value. This constraint will define a lower limit on Δ/R, which limit will depend on the SNR value, the modulation format, the number of cells N, the propagation factor n and the way in which channels are assigned to cells.
2. Both P_B and D as determined in Section 3-2 are determined in terms of the various system variables, including the throughput factor β in (3-11).
3. The system throughput Q in bits/sec/Hz is determined in (3-11).
4. P_B and/or D is expressed vs. Q for various system parameter values. Particular emphasis is given to selecting these parameters to maximize Q for P_B or D fixed.
5. Cost is determined indirectly by observing the values required for B, the total number of channels $M=B/\Delta$, the number of channels per mobile m, the number of bases N, and by noting the complexity of the channel assignment scheme, modulation format and channel code.

The above approach permits the total number of system parameters to be reduced to a manageable few; for example, the number of mobiles, average message length, radio coverage area, system bandwidth, and call attempt rate per mobile are conbined into a single variable.

The use of both d and P_B does involve system design conflicts; a low D value favours minimal values for m and Δ/R, and large R values. Low values for P_B favour minimum Δ/R values but large values for m, with R equal to (and not in excess of) the rate required for real-time traffic. Implications of these conflicts are considered in Chapter 4. Both criteria are useful, since systems originally designed for one type of digital traffic could, in time, carry all three types of traffic described in Section 2-2; in this latter case both criteria would be needed.

How should the SNR constraint value be chosen? The obvious answer is that the choice should yield an acceptable bit-error probability p. Unfortunately, p is not available as an average over coverage area, lognormal and Rayleigh fading, and interference levels, although recent useful work by French [F2] relates SNR to required geographic separation for combatting co-channel interference. The absence of a precise relationship between p and SNR does not invalidate our approach, since p is a monotone decreasing function of SNR. However, precise knowledge of p is of interest and is useful for evaluating the performance of channel codes. Coding can substantially improve throughput on Rayleigh channels, which are similar in many ways to mobile channels.

In the following three chapters, throughput relates to all bits transmitted. In Chapter 7, we discuss the application of channel coding, and show how throughput is easily modified to include its effects.

The sensitivity of the performance criteria to changes in system parameters is examined, primarily by actually varying the parameters, and examining the effects on performance.

III-8 References for Chapter 3

D1 J. DeMercado, N. Spyratos and B.A. Bowers, "A method for calculation of network reliability", IEEE Trans. Rel., vol. R-25, pp. 71-76, June 1976.

F1 H. Frank and I.T. Frisch, Communication, Transmission and Transportation Networks. Addison-Wesley, Reading, Mass., 1971.

F2 R.C. French, "The effect of fading and shadowing on channel reuse in mobile radio", IEEE Trans. Veh. Technol., vol. VT-28, pp. 171-182, Aug. 1979.

H1 A.D. Hall, III, "An overview of economics of scale in existing communication systems", IEEE Trans. Syst., Man and Cybern., vol. SMC-5, pp. 2-14, Jan. 1975.

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

K1 L. Kleinrock, Queueing Systems, vol. 2: Computer Applications. New York: Wiley, 1976.

L1 R.N. Lane, "Spectral and economic efficiencies of land mobile radio systems", IEEE Trans. Veh. Technol., vol. VT-22, pp. 93-104, Nov. 1973.

N1 M. Nesenbergs, "A hybrid of Erlang B and C formulas and its applications", IEEE Trans. on Commun., vol. COM-27, pp. 59-68, Jan. 1979.

O1 G.D. Ott,"Vehicle location in cellular mobile radio systems", IEEE Trans. Veh. Technol., vol. VT-26, pp. 43-46, Feb. 1977.

R1 S. Riter and J. McCoy, "Automatic vehicle location - an overview", IEEE Trans. Veh. Technol., vol. VT-26, pp. 7-11, Feb. 1977.

S1 L. Schiff, "Traffic capacity of three types of common user mobile radio communication systems", IEEE Trans. Commun. Technol., vol. COM-18, pp. 12-21, Feb. 1970.

W1 R.W. Wilkov, "Analysis and design of reliable computer networks", IEEE Trans. Commun., vol. COM-20, pp. 666-678, June 1972.

Y1 B. Yaged, Jr., "Economics of scale, networks, and network cost elasticity", IEEE Trans. Syst., Man and Cybern., vol. SMC-5, pp. 30-40, Jan. 1975.

IV ANALYSIS AND DESIGN OF SINGLE-CELL SYSTEMS

IV-I Single-Cell Systems

In many existing systems all mobiles are served by a single base station, with $M = B/\Delta$ two-way channels transmitting R bits/sec in each direction. Single-cell systems do not require co-ordination among different base stations and are therefore relatively inexpensive.

For single-cell systems, the signal and all interfering powers P and P_k in (3-13) are equal, in which case

$$SNR^{-1} = \sum_k \phi_k \, C(k\Delta/R) \tag{4-1}$$

where k ranges over positive and negative integer values.

All but the two channels at the edges of the frequency spectrum have channels on either side of themselves. When all channels are loaded equally, as will be the case in what follows, $\phi_k = \phi$, and

$$SNR^{-1} \simeq 2\phi \sum_{+k} C(k\Delta/R) \tag{4-2}$$

In actually determining Δ/R to meet the SNR requirement all except immediately adjacent-channels have negligible effect for MSK, SFSK and PAM. For PSK all but the two closest (in frequency) channels have negligible effect.

With m as the number of channels per mobile, Λ as the message traffic per unit of coverage area, $M = B/\Delta$ as the total number of channels, and λ as the traffic per group of m channels, $\lambda = \Lambda A m/M$. The utilization factor for m channels is

$$\begin{aligned}\rho &= \lambda/\mu R \\ &= (\lambda/\mu B)(B/\Delta)(\Delta/R) \\ &= (\Lambda A/\mu B)m(\Delta/R)\end{aligned}$$

Thus,

$$\rho/m = \beta\Delta/R \tag{4-3}$$

Equation (4-3) assumes that all channels carry equal traffic and that each channel has binary data rate R.

It follows from (4-3) that Figures 3-3 to 3-7, inclusive display P_B and channel occupancies vs $Q\Delta/R$.

The delay D normalized relative to the total system capacity μMR is τ, where

$$\tau = (M\mu R)D \tag{4-4a}$$

$$= (m\mu R)D(M/m)$$

$$= (M/m)d \tag{4-4b}$$

and where $d = (m\mu R)D$. Thus, the delay vs throughput curves in Figure 2-2 can be used to determine τ, provided the resulting delay d is multiplied by (M/m).

IV-2 Effects of Bit Rate and Channel Spacing on System Performance

Two questions of immediate concern are selection of bit rate R and channel spacing Δ. We begin by assuming that R is fixed, and that Δ is increased from the minimum value permitted by the SNR contraint in (4-2). In order for ρ/m in (4-3) to remain unchanged β must decrease to keep $\beta\Delta$ constant. There results a corresponding decrease in $M = B/\Delta$, and in the total system service capacity μMR. If m remains unchanged then so do P_B and ϕ_B. If m decreases with M, however, then P_B would increase even if $\beta\Delta$ remained constant.

Maintaining m fixed while $\beta\Delta$ remains constant leaves d and ϕ_d unchanged. Although τ decreases, since M/m decreases, actual delay D remains unchanged. If m decreases to maintain M/m constant, d and τ both decrease, but D again remains unchanged.

It follows that an increase in Δ from its minimum allowed value while maintaining R fixed requires a corresponding decrease in throughput to prevent an assured increase in both blocking probability and delay. If m decreases with β, P_B will increase even with this decrease in β. In any case Δ/R should be minimized.

The selection of the actual values for R and Δ is more difficult. For real-time voice traffic, R would be the minimum value required to transmit at a rate required for adequate speech quality, and Δ/R would then be minimized, as explained above. In addition m would be made as large as possible, subject to cost constraints. The choice $m = M$ minimizes P_B.

For non-real-time data, maximizing R with Δ/R fixed minimizes M while keeping total service capacity μMR constant. Minimizing M minimizes $\tau = (M/m)d$ for any fixed value of m, and therefore minimizes D. The actual value of m is not very important; an increase in m reduces M/m but increases d by an almost equal amount. When channels are divided into groups of m channels each, with each mobile assigned one of these groups, M/m equals the number of such groups. An increase in m is offset by the decrease in the number of groups, usually.

The maximum value for R will be governed by Gaussian noise; as R increases the energy per bit decreases, with the result that Gaussian noise will eventually provide the SNR constraint limitation. The maximum R at which this limitation occurs depends on the type of receiver, radiated power, and cell size. Unless this limit on R fails to exceed that value required for real-time traffic, a conflict arises when both real-time and non-real-time traffic are present. The latter favours wide-band high-bit-rate channels, while the former favours minimum rates required for real-time transmission.

IV-3 System Performance for Various Modulation Formats

For now we assume a required SNR level of 20 dB. For MSK with approximately 50% channel occupancy it follows from (4-2) and Figure 2-9, that $\Delta/R = 0.95$ is the minimum value permitted for Δ/R. For $\Delta = 30$ kHz, which is conventional channel spacing, $R \simeq 31.6$ kHz. This value of R fits the MSK spectrum well within the "top-hat" constraint in Figure 2-7, and is adequate for differential PCM transmission of voice. If the channel were Gaussian, then bit error probability $p \simeq 10^{-22}$, in accordance with Section 2-7. For Rayleigh channels with an average signal-to-Gaussian-noise ratio of 20 dB, $p = 0.02$ (See Figure 2-11). Determination of a truly suitable value for SNR requires that p be related to SNR, as discussed in Section 3-4.

For binary PSK

$$SNR^{-1} \simeq 2\phi[C(\Delta/R) + C(2\Delta/R)] \tag{4-5}$$

With SNR = 20 dB and $2\phi \simeq 1$, $C(\Delta/R) \simeq -21$ dB and $\Delta/R = 2.5$. If $\Delta = 30$ kHz then $R = 12$ kHz. (This value greatly exceeds the 4 kHz maximum bit rate permitted by the "top-hat" constraint in Figure 2-4.)

When compared with the MSK results, (4-3) shows that the same values for P_B, d and channel occupancies will result for PSK only if throughput factor β is reduced by $\Delta/R \simeq 2.6$. Thus, the curves in Figures 3-2 to 3-7 inclusive apply to single-cell systems with PSK modulation, provided the actual throughput is 0.40 times that shown on the traffic carried axes. The curves apply for MSK provided the actual throughput is 1.05 times that shown on the traffic carried axes.

For binary PAM with no excess bandwidth ($\alpha = 0$), $C(\Delta/R) \simeq -20$ dB for $\Delta/R = 1$, which is therefore the minimum value allowed, assuming 50% channel occupancy. For $\alpha = 1$, $\Delta/R = 1.3$ is the minimum allowed value. For SFSK, $\Delta/R = 1.4$ is the minimum allowed value (See Figure 2-10).

These results are summarized in Table 4-1, which shows the minimum Δ/R values for SNR = 20 dB. These Δ/R values also represent the factor by which throughput Q must be divided in order to use Figures 3-2 to 3-7 inclusive, to determine d, τ, and ϕ. Also shown in Table 4-1 are the maximum bit rates permitted for Δ = 30 kHz as well as implied bit error probabilities for Gaussian and Rayleigh channels, as presented in Section 2-7. For Gaussian channels p is approximated by

$$p \simeq (1/2) \exp(-SNR/2) \tag{4-6}$$

where SNR is in actual units (not dB's).

IV-4 Effects of Different SNR Constraints

The fact that the constraint SNR = 20 dB permits a PSK bit rate which greatly exceeds that permitted by the "top-hat" characteristic in Chapter 2 may imply an unacceptably large average bit error probability.

Table 4-1 shows the effects of using different SNR values. In particular the minimum Δ/R values permitted are shown, assuming a 50% channel occupancy. As before, these Δ/R values represent the factor by which actual throughput must be divided in using Figures 3-2 to 3-7 to determine P_B, τ and ϕ.

The rate at which Δ/R must decrease as SNR (and $[C(\Delta/R)]^{-1}$) increases depends ultimately on the roll-off rates of the transmitted spectra. In examining Table 4-1, one sees the clear inferiority of PSK relative to MSK, even in moving from 20 dB to 30 dB. The MSK rate falls by a factor of 1.8, while the PSK rate falls by a factor of four. For 10 dB $\lesssim$ SNR $\lesssim$ 33 dB, MSK has the highest R/Δ ratio. For SNR $\gtrsim$ 40 dB, SFSK lies between PAM and MSK, in terms of the maximum allowable R/Δ value. The minimum Δ/R values are plotted in Fig. 4-1.

SNR Constraint	Implied Bit Error Probabilities		Minimum Δ/R Values				Maximum R(kHz) for Δ = 30 kHz			
	Gaussian	Rayleigh	MSK	PSK	PAM(α=1)	SFSK	MSK	PSK	PAM(α=1)	SFSK
10 dB	$10^{-2.2}$	2×10^{-1}	.65	.8	.87	.8	46.2	37.5	34.5	37.5
20 dB	10^{-22}	2×10^{-2}	.95	2.5	1.3	1.4	31.6	12.0	23.1	21.4
30 dB	10^{-220}	2×10^{-3}	1.4	8	1.7	1.9	21.4	3.5	17.7	15.8
40 dB	$<10^{-1000}$	2×10^{-4}	2.4	25	1.9	2.4	12.5	1.20	15.8	12.5
50 dB	$<10^{-1000}$	2×10^{-5}	4.0	82	1.95	2.8	7.5	.37	15.4	10.7
60 dB	$<10^{-1000}$	2×10^{-6}	7.5	256	2	3.7	4.0	.12	15	8.1

Table 4-1: Illustrating the effects of various modulation formats and SNR constraints on minimum Δ/R ratios. (50% channel occupancy)

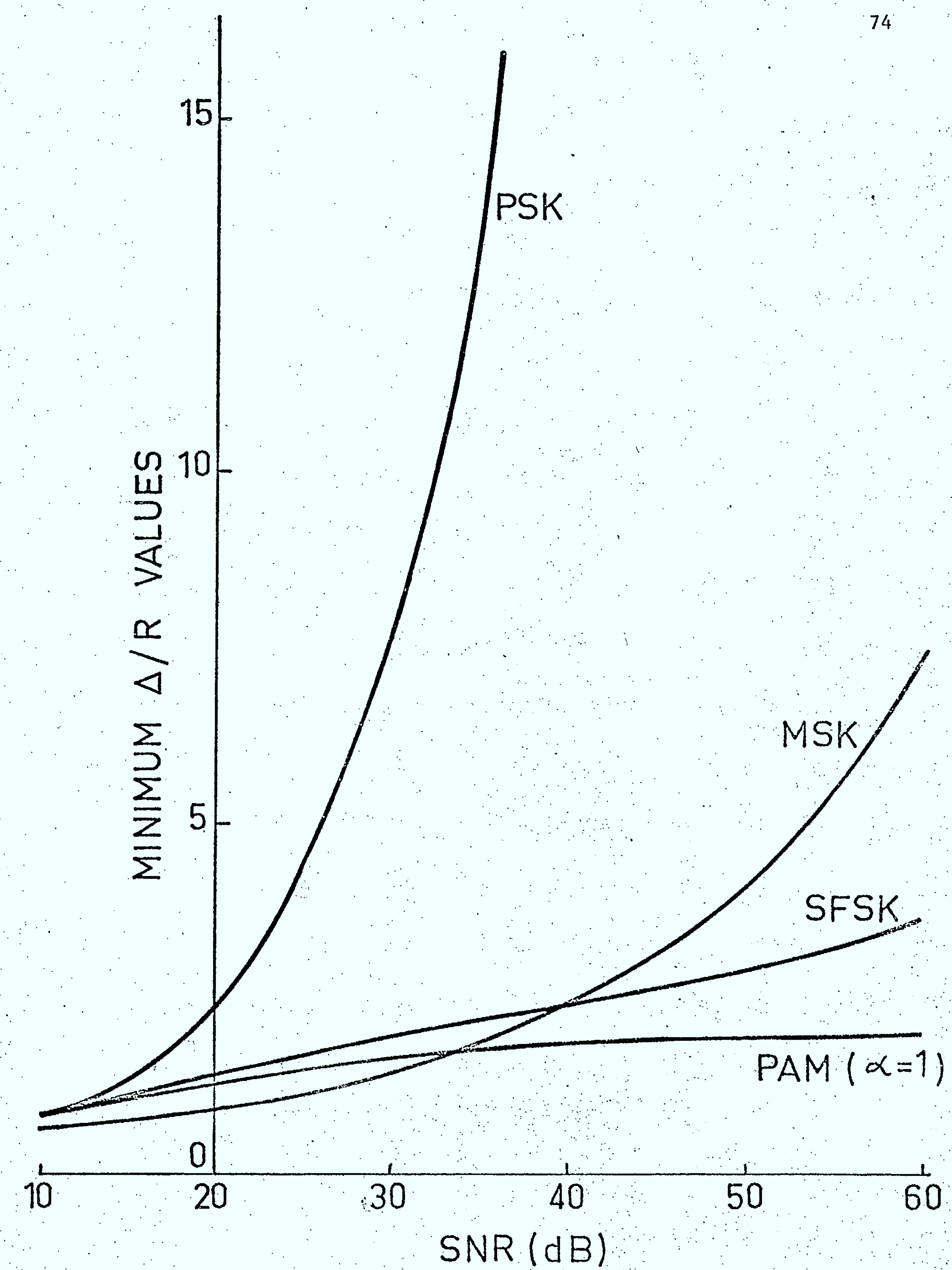

Fig. 4-1: Minimum Δ/R values for various SNR levels.

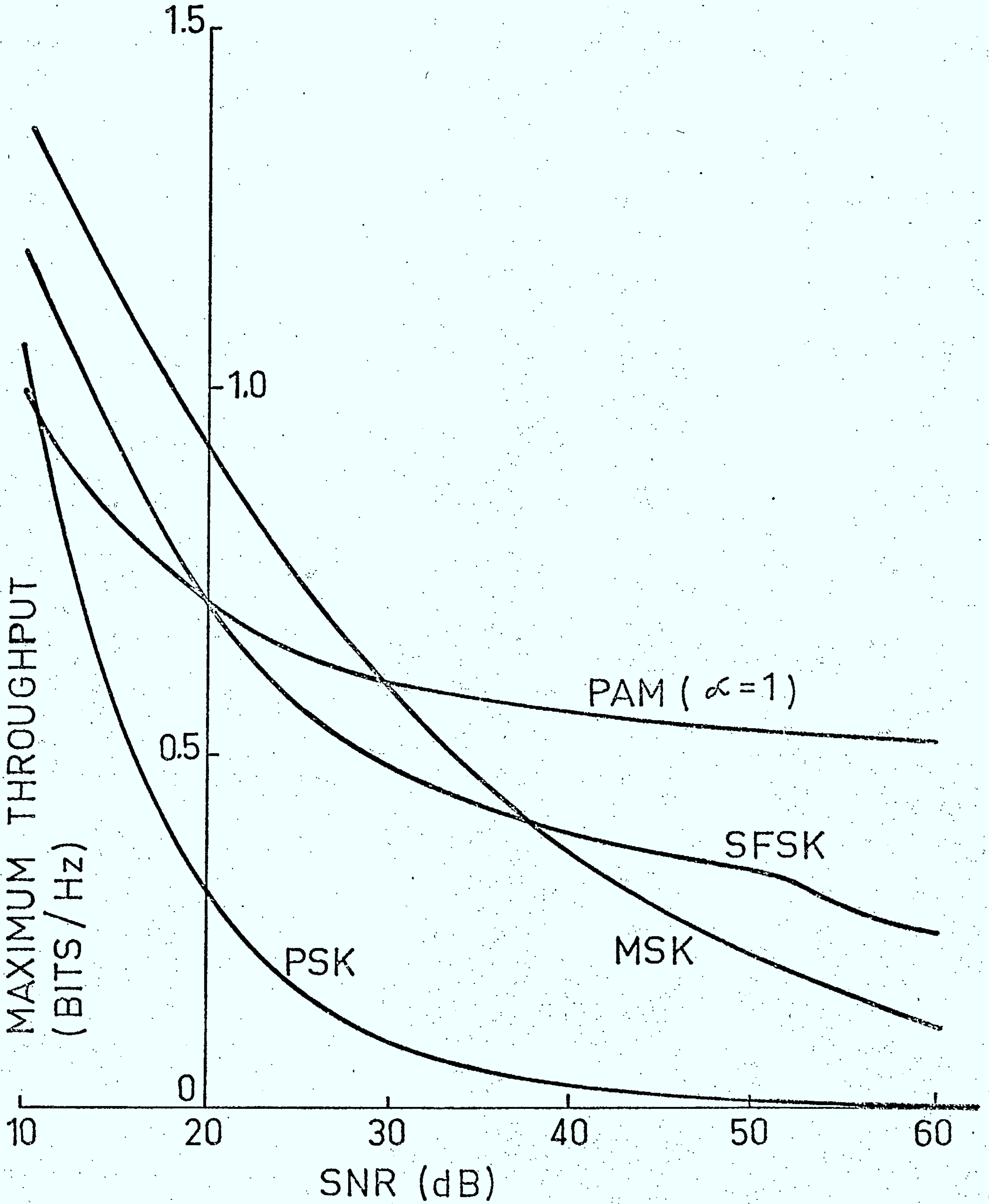

Fig. 4-2: Maximum throughput vs. SNR constraint for various modulation formats.

If the channel occupancy is below 50% as would often be the case, the Δ/R value could be reduced slightly from the values shown. Since channel occupancy fluctuates with traffic level, a fixed value of 50% is safe and reasonable in calculating SNR.

Figure 4-2 shows the maximum R/Δ values vs. the SNR constraint, obtained by assuming a channel occupancy $\phi=1$. These R/Δ values represent the maximum obtainable spectrum efficiency. They are slightly lower than the R/Δ values in Table 4-1, since doubling the occupancy from 50% to 100% in effect increases the SNR constraint by 3dB, and reduces the maximum values of R/Δ.

IV-5 System Analysis and Design Example

The use of Table 4-1 and (4-3) together with the curves in Chapter 3 is straightforward. For example, with SNR=20dB, the minimum Δ/R value is 0.95 for MSK, from Table 4-1. For the curves in Chapter 3, traffic carried is 0.95Q for this minimum Δ/R value, and Q=1.05 is the maximum value permitted.

With Q=0.45, $R=\Delta$, and m=2, traffic carried equals 0.45 in Chapter 3, $P_B=0.25$ and $d\simeq2.5$. If R=30 kHz and $\mu^{-1}=1000$ bits/message, the actual delay $D=d/\mu mR=41.7$ msec. If R and Δ are both halved to R=15 kHz, P_B remains unaltered but D doubles to 83.3 msec. However, if R alone decreases to 15 kHz with Δ remaining fixed at 30 kHz, then the value on the traffic carried axis in Chapter 3 becomes 2.0 Q; with Q=0.45 (2Q=0.90) P_B increases to 0.8, d increases to $d\simeq10.6$ and the actual delay D to 353 msec. A further halving of R to 7.5 kHz with Δ constant would require a corresponding halving of Q just to maintain $P_B=0.8$ and D=353, msec., and a factor of four reduction in Q to return P_B and D to their values at R=30 kHz.

If the SNR requirement for an MSK system with m=2 increases to

40dB, the minimum Δ/R increases to 2.4; in which case the traffic carried on the axes in Chapter 3 is 2.4Q. With m=2, P_B and d remain at 0.25 and 2.5, respectively, provided Q decreases to 0.45/2.4=18.8. The maximum allowed Q=0.417.

If PSK rather than MSK is used with SNR=40dB, Table 4-1 shows the maximum $Q=1/25\simeq 0.04$. Provided Q is reduced to 0.45/25=.0018, P_B and d remain at 0.25 and 2.5 respectively, when m=2.

If we reconsider MSK with SNR=20dB and let m=4, then with Q=0.45, P_B=0.1, $d\simeq 4.23$ and D decreases from 41.7 msec. when m=2 to 35.3 msec., assuming Δ=R=30 kHz. If P_B=0.25 is acceptable, then with m=4 the traffic carried on Fig. 3-3 and the MSK throughput can increase to 0.63. However, d then increases to 4.9 and D to 40.8 msec. Thus, in this particular case, doubling the number of channels per mobile has resulted in a 40% increase in throughput with no change in P_B and virtually no change in actual delay D. However, there is a cost associated with doubling the number of mobiles per channel.

V MULTIPLE-CELL SYSTEMS WITH NO REUSE OF FREQUENCIES

V-1 Multiple-Cell Systems

Division of a radio coverage area into cellular regions involves several considerations additional to those for single-cell systems. Assignment of channels to cells, cell size, and assumptions regarding the propagation factor n all affect interference levels in neighbouring cells.

In what follows, we assume uniform traffic over the coverage area. Unless otherwise stated, all N cells are of the same size and shape. We also assume that the same bandwidth B_k and the same number of channels m_k are assigned to each cell. In this case throughput factor $\beta_k = \Lambda A_k/\mu B$ is the same for all cells, and is equal to $\beta = \Lambda A/\mu B$ for a single-cell system. Also, the throughput Q in bits/Hz/sec. is equal to that for each cell, since

$$QB = \sum_{k=1}^{N} Q_k B_k \tag{5-1}$$

$$B_k = B/N \qquad (k = 1,2,\cdots,N) \tag{5-2}$$

$$Q_k = Q \qquad (k = 1,2,\cdots,N) \tag{5-3}$$

In this chapter we consider systems with 3, 7 and 19 cells. Such a progression would be quite natural in a region experiencing growth accompanied by increasing traffic levels.

V-2 Three-Cell System

Figure 5-1 shows the area covered by a three-cell system. Use of three circles of equal size for coverage of a circular region results in gaps and overlap, which could be reduced in practice by use of directional

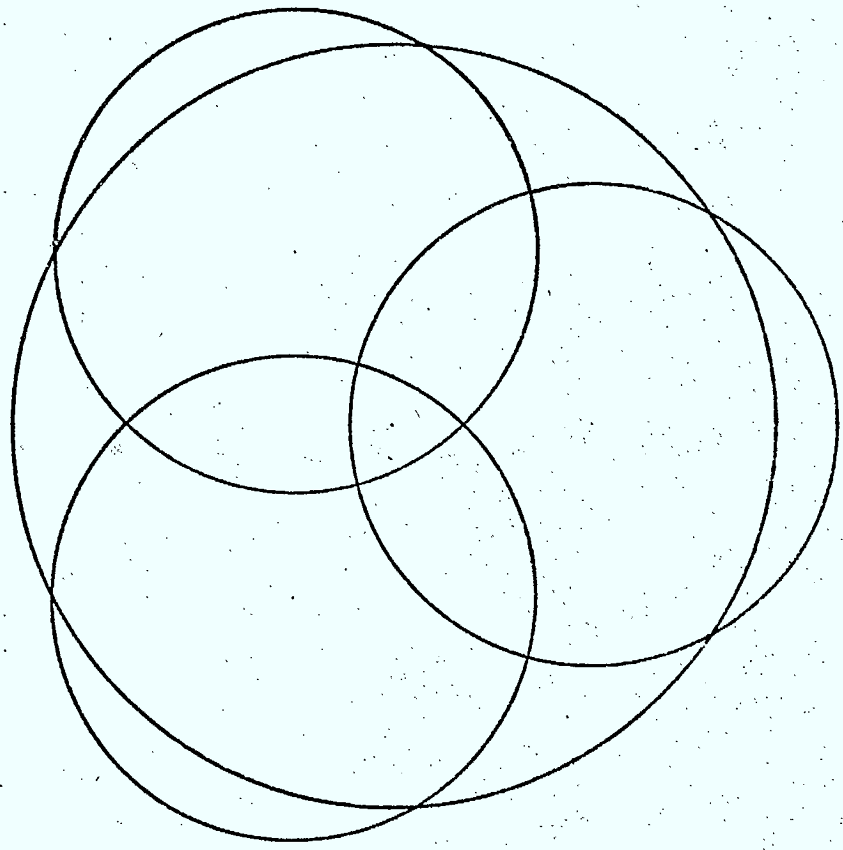

Fig. 5-1: Coverage for a three-cell system.

antennae. We ignore these effects in our analysis.

As noted in Section 3-5,

$$P/P_k = (\sqrt{6})^n \tag{5-4}$$

independent of the radii of the three coverage circles. If adjacent frequency channels are assigned to adjacent cells, then channels 1, 4, 7, 10, ••• will be assigned to one cell, channels 2, 5, 8, 11, ••• will be assigned to a second cell, with channels 3, 6, 9, 12, ••• assigned to the remaining cell. For each channel,

$$\begin{aligned} SNR^{-1} &\simeq 2\phi[(\sqrt{6})^{-n}\{C(\Delta/R) + C(2\Delta/R)\} + C(3\Delta/R)] \\ &\simeq 2\phi[0{\cdot}043\ C(\Delta/R) + C(3\Delta/R)] \end{aligned} \tag{5-5}$$

where we have used $n = 3{\cdot}5$, and have assumed $C(2\Delta/R) << C(\Delta/R)$.

Ratio Δ/R must now be selected to ensure that the SNR constraint in (5-5) is satisfied. For example with SNR = 20 dB one sees from the fact that $(\sqrt{6})^{3\cdot5}$ corresponds to 13.6 dB, and from Fig. 2-9 that for MSK, the minimum value of $\Delta/R \simeq 0.5$ for 50% channel occupancy ($2\phi=1$). For 100 percent ($2\phi=2$) occupancy $\Delta/R \simeq 0.65$, and the maximum throughput is, therefore, 1.5. This throughput is approximately 50% higher than for single cell systems.

Figure 5-2 shows minimum Δ/R values for various modulation formats vs. SNR, assuming 50% occupancy. Figure 5-3 shows the maximum throughput R/Δ, in which case ϕ=100%.

For SNR $\lesssim$ 15 dB, minimum Δ/R values are governed by interference from the same cell as the signal , and separated 3Δ Hz in frequency. For MSK, SFSK and PAM, adjacent-cell interference at separation Δ Hz governs the minimum Δ/R for SNR $\gtrsim$ 15 dB. For PSK, same-cell interference determines the minimum Δ/R for SNR $\gtrsim$ 15 dB.

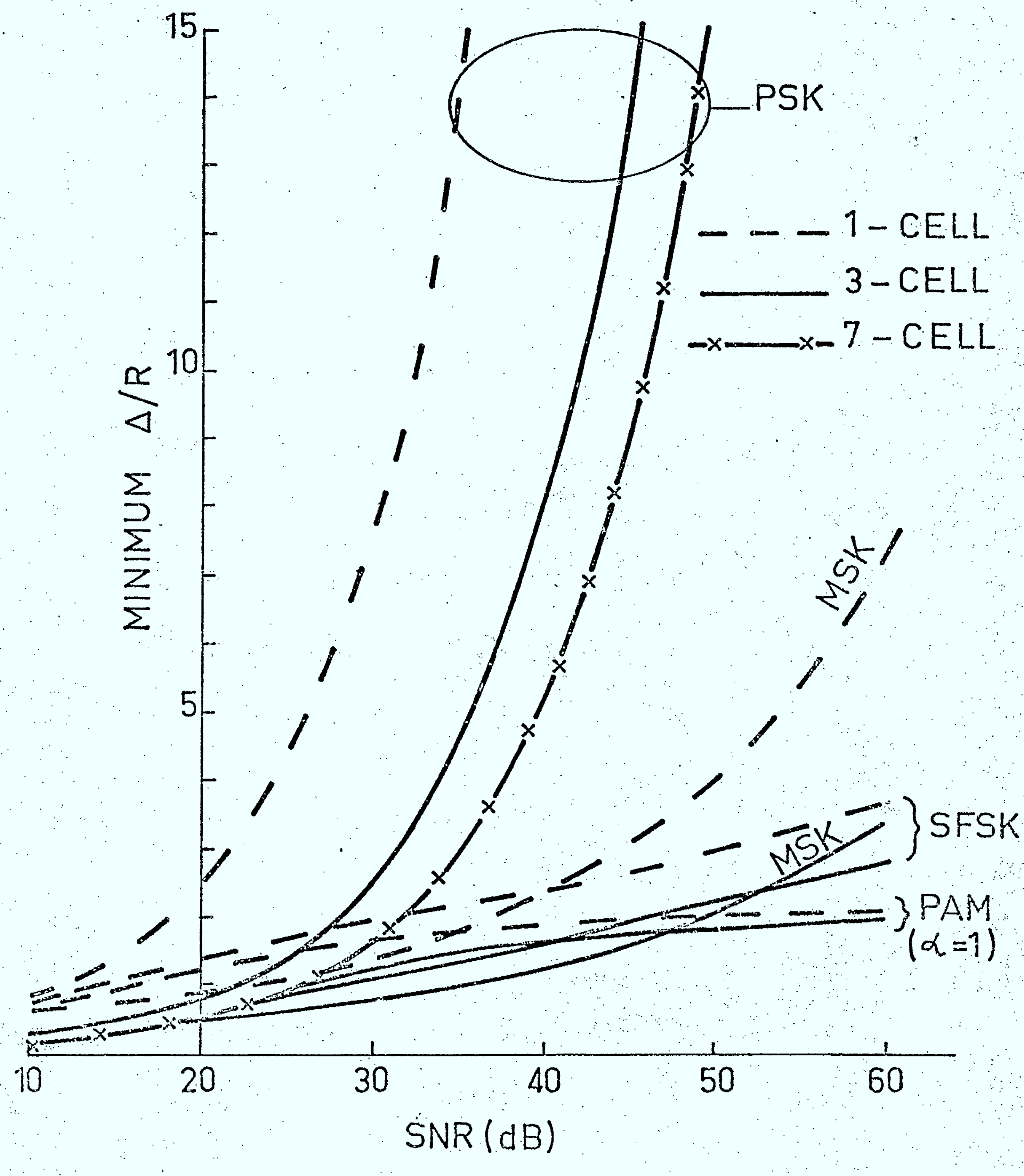

Fig. 5-2: Minimum Δ/R values vs. SNR for three-cell systems.

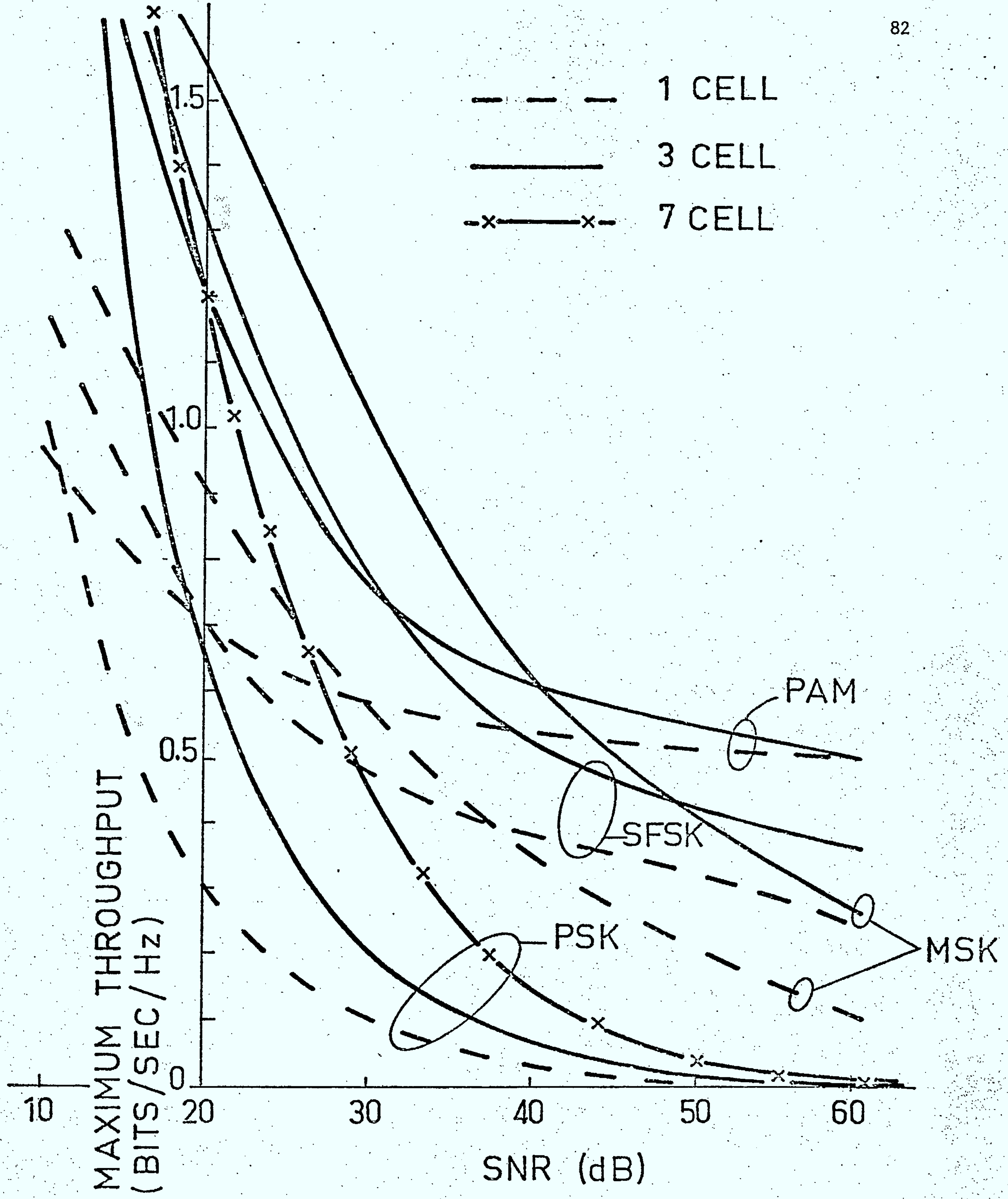

Fig. 5-3: Maximum throughput (bits/sec/Hz) vs. SNR for three-cell systems.

Performance comparisons between one-cell and three-cell systems are readily made using the curves in Chapter 3, Fig. 5-2 and Fig. 5-3. For a fixed number of channels per mobile, the number of channels per mobile per cell for a single-cell system is three times that for a three-cell system. From these facts and Fig. 3-3, the curves in Fig. 5-4 are obtained showing system throughput Q vs. number of channels per mobile, for various P_B values for MSK with SNR = 20 dB, with Δ/R selected for 100% occupancy.

Figure 5-4 shows that for throughputs above the equal P_B locus, three-cell systems have lower P_B values, and that the converse is true for lower throughputs Q. The number of channels per mobile needed to make three-cell systems advantageous is seen to increase as P_B decreases. For example, with three channels per mobile and $P_B \lesssim 0.5$, three-cell systems are never advantageous. However, with 19 or more channels per mobile and $P_B \gtrsim 0.05$, three-cell systems are always better. With 40 channels per mobile, for example, and $P_B = 0.2$, a single-cell system has a throughput of approximately 0.92, compared with the 37% increase to 1.25 for a three-cell system.

Similar comparisons can be made for delay D, by using Fig. 3-2. If D_1 and D_3 denote delays for the one- and three-cell systems, respectively, then

$$D_3/D_1 = (d_3/d_1)(m_1/m_3)(R_1/R_3) \tag{5-6}$$

where d_i, m_i and R_i denote, respectively, the normalized delay, number of channels/mobile/cell and the bit rate for one- and three-cell systems. One sees immediately that if Δ/R is reduced by increasing R in moving from a one-cell to a three-cell configuration, D_3/D_1 is smaller than if Δ is reduced, with R unchanged.

In using Fig. 3-2 to determine d_1 and d_3, the traffic carried axis must be scaled by Δ/R. Thus, $(d_3/d_1)(m_1/m_3)$ is often less than unity, and is much less than unity when the single-cell throughput $Q \simeq 1$. Again, with

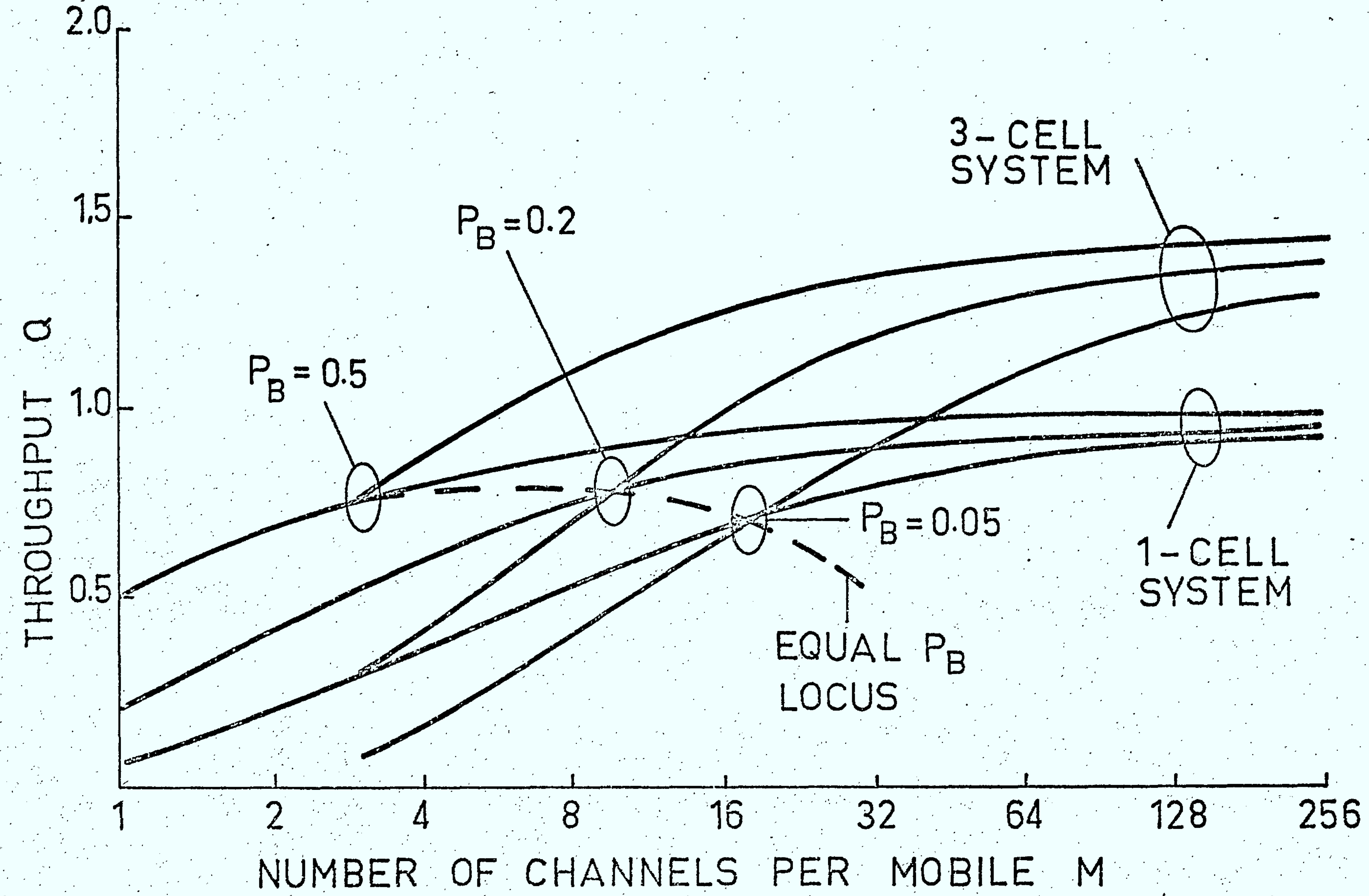

Fig. 5-4: Throughput comparisons for one- and three-cell systems vs. number of channels per mobile; MSK, SNR = 20 dB.

SNR = 20 dB and MSK modulation, consider Q = 0.75 for a single-cell system. With m_1 = 15 and m_3 = 5, Fig. 3-2 yields $d_1 \simeq 16$; $Q_3 = 0.5$, and $d_3 \simeq 5$. Thus, if $R_1 = R_3$, $D_3/D_1 \simeq 0.94$. However, if $R_3 = 1.5\ R_1$ in which case the channel separation is unchanged, $D_3/D_1 \simeq 0.63$. Thus, the delay of the single-cell system is 50% higher than that of the three-cell system. In an Erlang B environment, $P_B = 0.08$, as compared with $P_B = 0.12$ for the single-cell case. Again, there is a 50% improvement in performance in using a three-cell system.

The same type of comparisons can be made for other modulation formats and SNR constraints. The value of Q where three-cell systems become beneficial decreases as the maximum throughtput for three-cell systems relative to that for one-cell systems increases.

V-3 Seven-Cell System

In seven-cell systems, channels are placed into one of seven groups such that each cell contains channels separated by at least 7Δ Hz. Figure 5-5 shows three possible assignments of groups channels to cells. All other (fixed) channel assignments where contiguous outer cells do not contain immediately adjacent frequency channels are equivalent to these, in terms of interference. To see this, note that without loss in generality groups 1 and 2 can be assigned, respectively, to the centre and any outer cell. Groups 7 can then be assigned in one of three different ways. The remaining channels must then be assigned in one of the three ways shown.

The centre cell is subject to more interference than any of the outer cells, and the interference effects are not the same for each outer cell. Unless the centre cell's coverage area is reduced to accommodate increased retransmissions caused by the higher interference level, this cell limits perform-

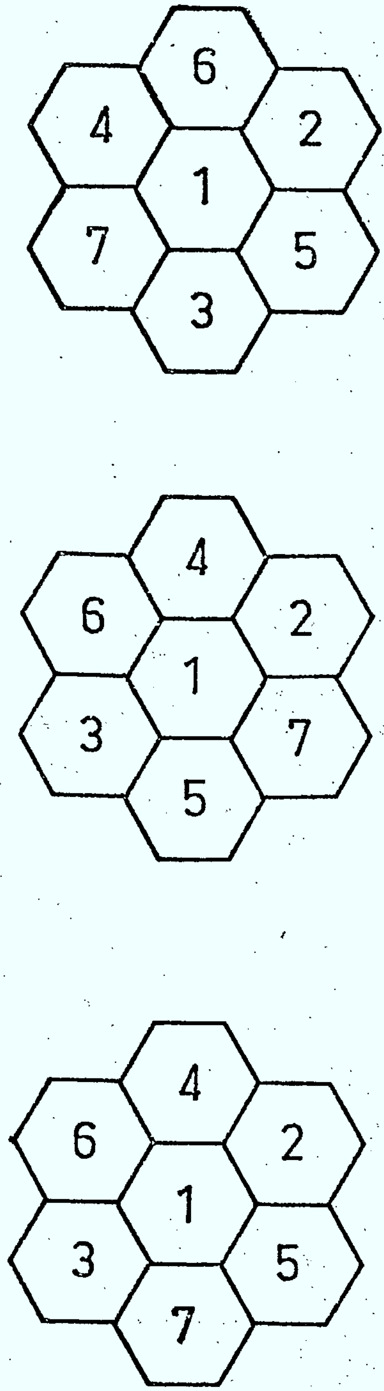

Fig. 5-5: Seven-cell system channel assignments.

ance. If the centre cell area is reduced, the outer cell centres more closer to each other, and interference levels in these cells increases. Also, considerable overlap in coverage results, unless non-isotropic antennae are used. While such a plan permits some increases in throughput, increases for MSK, SFSK and PAM relative to that for three-cell systems is small for SNR $\tilde{>}$ 15 dB, since adjacent-channel interference limits SNR.

For PSK throughput increases are possible. For the centre cell

$$SNR^{-1} \simeq 2\phi[C(7\Delta/R) + (\sqrt{6})^{-n}\sum_{k=1}^{3} C(k\Delta/R)] \qquad (5\text{-}7)$$

Since $C(3\Delta/R)$ in (5-6) limits SNR for PSK, the minimum Δ/R values for PSK for 50% channel occupancy and maximum throughput R/Δ (100% occupancy) are much more favourable than the corresponding values obtained for three-cell systems in the previous section. The PSK results are included on Figs. 5-2 and 5-3. Actually, adjacent-cell interference limits performance in this case for SNR $\tilde{>}$ 15 dB.

For the other modulation schemes, same-cell interference limits performance for SNR $\tilde{<}$ 15 dB, and in this case improvements in Δ/R over three-cell systems are by factors of 3/7 in accordance with (5-6) and (5-7).

V-4 Nineteen-Cell System

Figure 5-6 shows a 19-cell system, with groups of channels arranged in such a way that immediately adjacent channels are assigned to cells centred $2\sqrt{3}$ r apart, where r is the length along any hexagon side. Channels separated by 2Δ Hz are assigned to cells centred 3r apart. Thus, contiguous cells have channels separated at least 3Δ Hz in frequency. The SNR varies from cell-to-

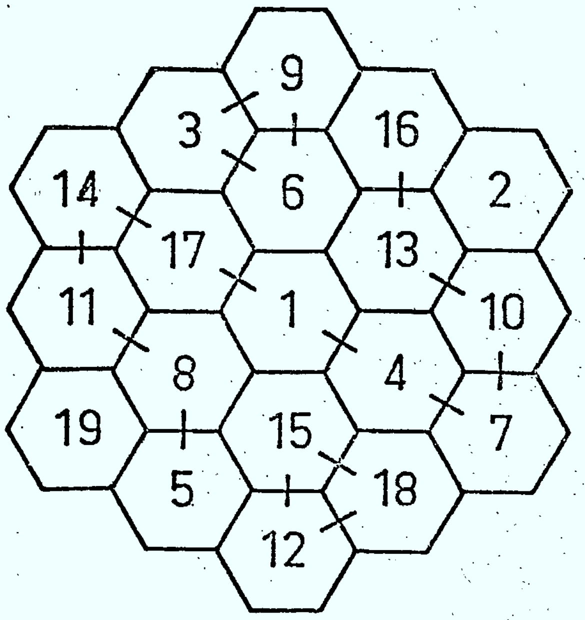

Fig. 5-6: Nineteen-cell system channel assignments.

cell, but in the worst case is approximated as follows:

$$SNR^{-1} \simeq 2\phi[(2\sqrt{6})^{-n}C(\Delta/R)+(3\sqrt{2})^{-n}C(2\Delta/R)+(\sqrt{6})^{-n}C(3\Delta/R)+C(19\Delta/R)] \qquad (5\text{-}8)$$

For n=3.5, the terms $(2\sqrt{6})^n$, $(3\sqrt{2})^n$ and $(\sqrt{6})^n$ correspond, respectively, to 24.3 dB, 22 dB, and 13.6 dB.

In the 19-cell system, both the adjacent-cell and same-cell interference is much smaller than that for the 3-cell and 7-cell system. One would therefore expect increased throughput R/Δ for the various modulation formats, in the 19-cell system.

Figure 5-7 shows minimum Δ/R values for PSK, MSK and SFSK assuming 50% channel occupancy. Figure 5-8 shows maximum throughput R/Δ (100% occupancy). Shown for comparison purposes are results for 3-cell systems. In both figures, interference from immediately adjacent cells limits Δ/R for SNR $\gtrsim$ 15 dB. Same-cell interference limits Δ/R for SNR $\lesssim$ 15 dB. The increased throughputs possible for 19-cell systems, compared with 3-cell (and 7-cell) systems are substantial; for example, with SNR $\simeq$ 20 dB, the former has maximum throughputs three times as large as the latter.

Throughput vs. number of channels per mobile m for various P_B values is shown in Fig. 5-9, for MSK with SNR = 20 dB. Curves for 3-cell systems are shown for comparison. Also shown are loci where 19-cell systems outperform 3-cell systems, and where 3-cell systems outperform single-cell systems. One sees that 19-cell systems always outperform 3-cell (and 7-cell) systems for $P_B \gtrsim 0.2$, and also for $P_B \gtrsim 0.05$ provided $m \gtrsim 22$. For a 38-channel transceiver (2 channels/cell) and $P_B \simeq 0.20$, 19-cell throughput is 2.2 vs. 1.2 for 3-cell systems. Again, Δ/R is selected to give SNR = 20 dB for ϕ = 100%.

Similar comparisons can be made for delay. Consider a 38-channel system employing MSK with SNR $\simeq$ 20 dB. With throughputs Q_3 = 1.13 and

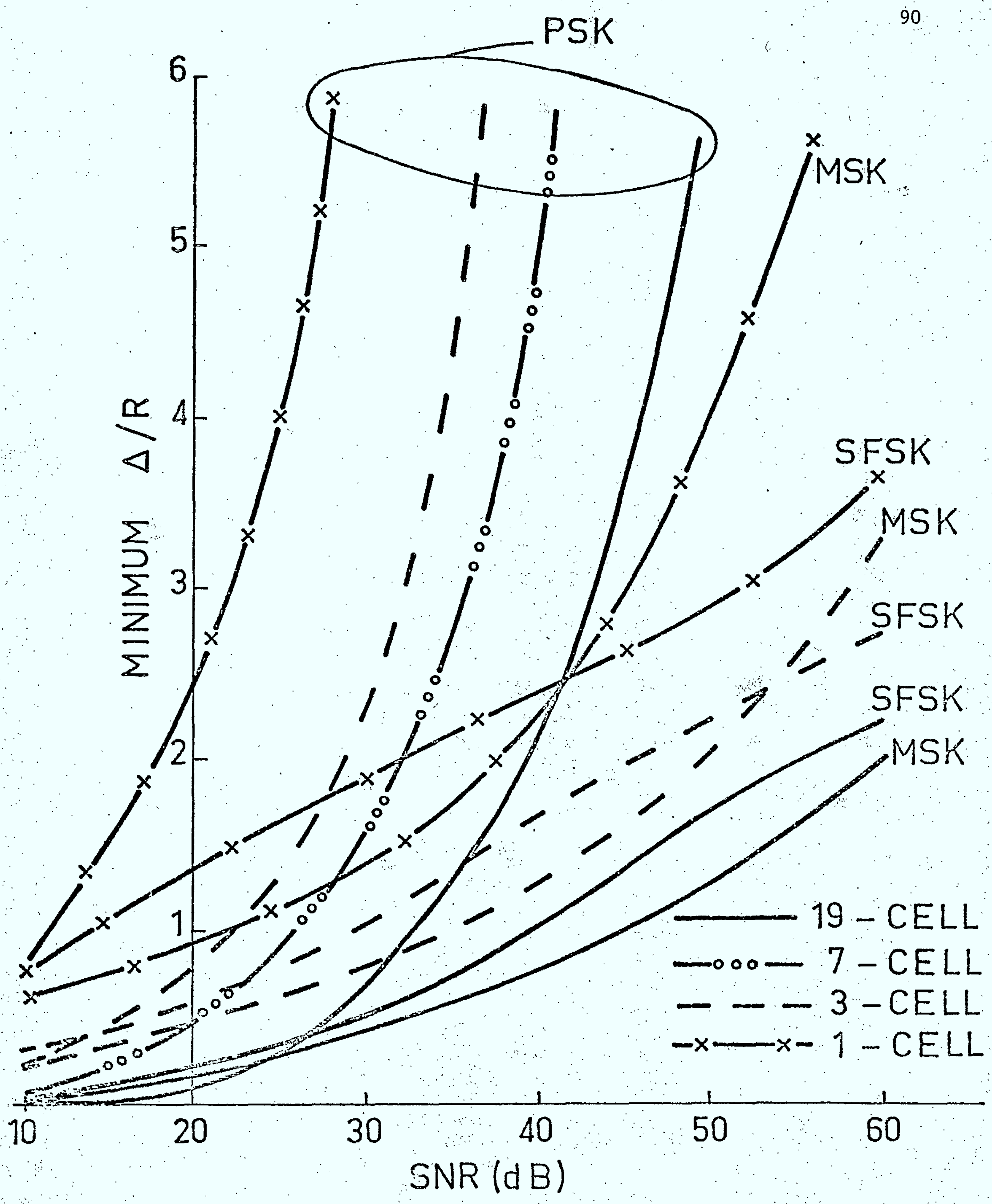

Fig. 5-7: Minimum Δ/R values for 19-cell systems.

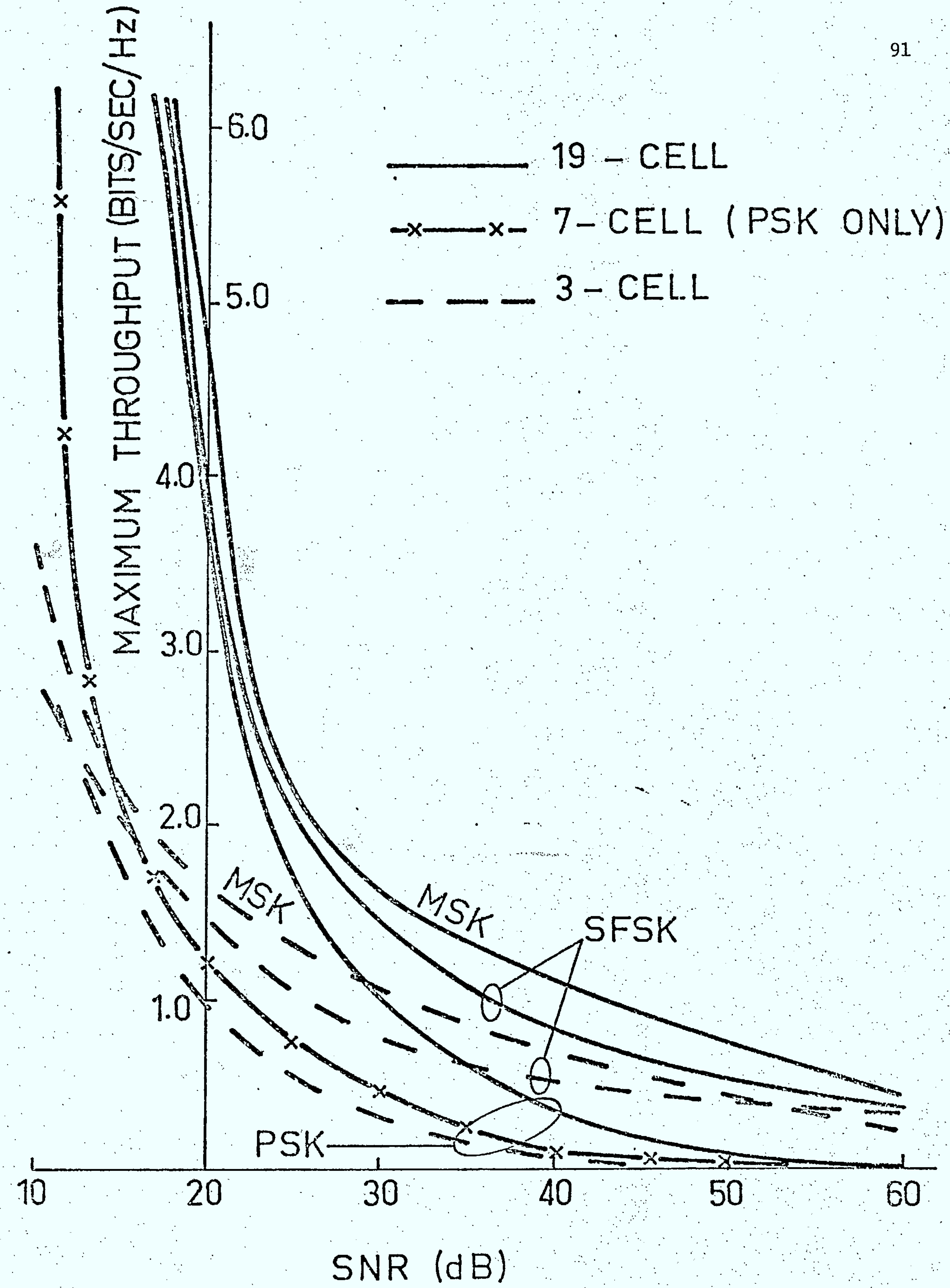

Fig. 5-8: Maximum throughput for 19-cell systems.

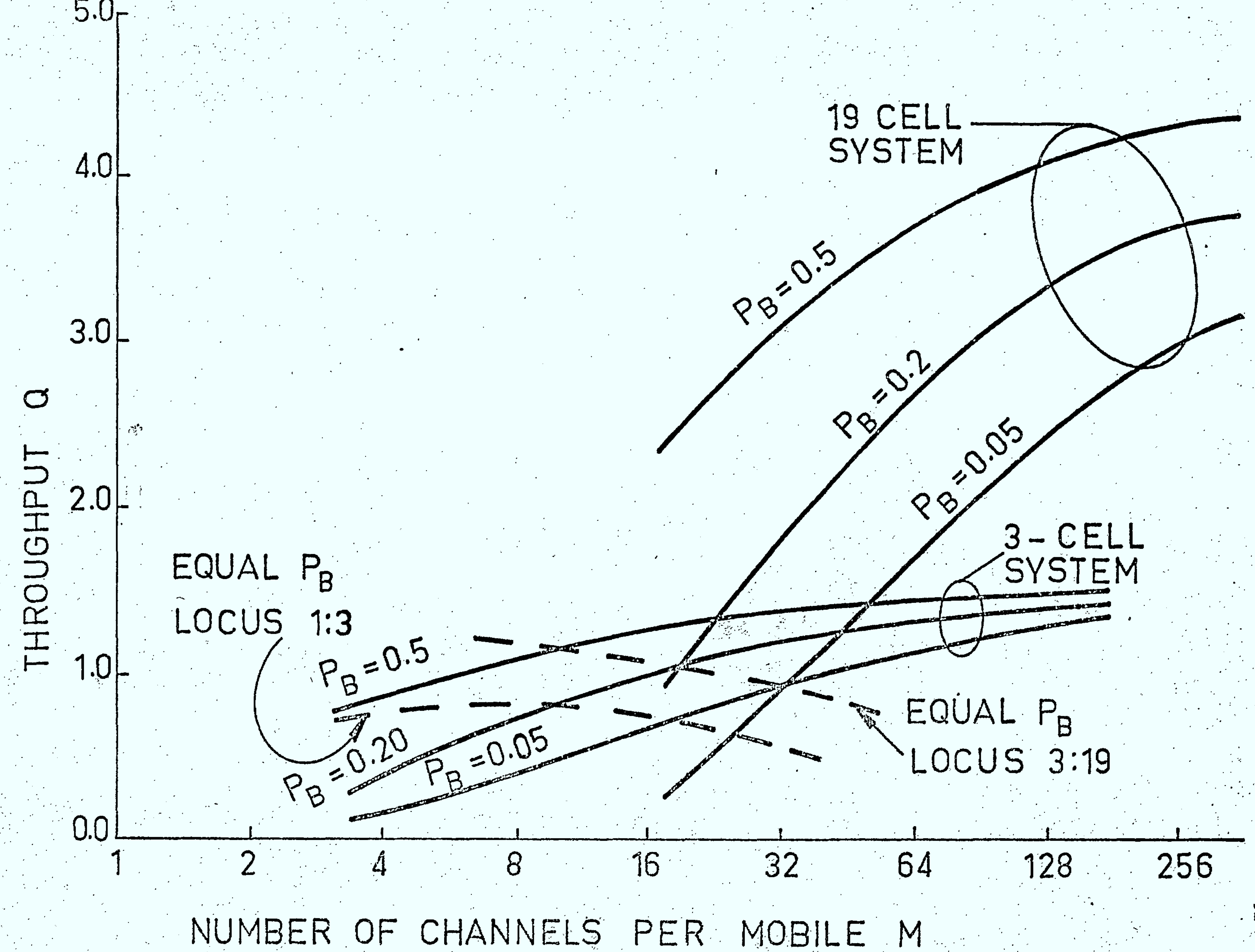

Fig. 5-9: Throughput for 19-cell system vs. number of channels per mobile; MSK, SNR = 20 dB.

$Q_{19} = 3.3$, $d_3 \simeq 13$, $d_{19} \simeq 2$; $m_3 = 2$, $m_{19} \simeq 13$; and $R_3/R_{19} \simeq 3$ if increased throughput results from increased bit rates. Thus, the throughput triples and the delay falls to one-third. However, P_B increases from 0.15 for a 3-cell system, to 0.45. If Δ rather than R is changed in moving from a 3-cell to a 19-cell system $D_3 \simeq D_{19}$, and P_B increases from 0.15 to 0.24, provided m equals 6, which implies a 114 channel transceiver.

A natural question at this point is whether some (fixed) assignment of channels to cells, other than that shown in Fig. 5-6, would yield a larger maximum throughput. For the centre cell, no substantial reduction in interference is possible, since $2\sqrt{3}$ r is the maximum separation possible from cells containing adjacent frequency channels. The size of this centre cell could be reduced; however, other cells would then be closer to each other, and their interference levels would increase. As well, considerable coverage overlap would occur. Further, we have not succeeded in finding an assignment such that all cells except the centre cell have adjacent-channel separations of at least 4Δ, although we have not exhaustively examined all possible assignments. We conclude, therefore, that fixed channel assignment schemes with throughputs in excess of those shown in Fig. 5-8 do not exist.

VI MULTIPLE-CELL SYSTEMS EMPLOYING CHANNEL REUSE

VI-1 Multiple-Cell Systems with Channel Reuse

Reuse of frequency channels in cells which are sufficiently far apart obviates the need to continually increase the number of channels per mobile as the number of cells increases. However, co-channel interference results, and is independent of the bit-rate. Typically, co-channel coefficient $C(0) \simeq -3$ dB (see Fig. 2-9) unless PSK is used, in which case $C(0) \simeq -5$ dB. This co-channel interference eventually limits the SNR obtainable at any given channel occupancy ϕ; the only way to increase SNR further is to reduce ϕ, which results in a proportional reduction in throughput.

The number of groups of channels G in a cellular plan is not arbitrary, but must be selected as follows, where j and k are non-negative integers:

$$G = (j + k)^2 - jk \tag{6-1}$$

Table 6-1 shows the values of G permitted, together with the distance D between the centres of cells with identical frequency channels. Also shown for later reference is the ratio P/P_k for the wanted signal and co-channel interferer, for n = 3, 3.5 and 4; i.e.

$$(P/P_k) = (D\sqrt{2})^n \tag{6-2}$$

in accordance with Section 3-5.

In what follows we assume that all cells have equal area A_k, number of channels/mobile m_k and bandwidth B_k; $A_k = A/N$ and $B_k = B/G$ where N and G denote, respectively, the number of cells and groups of channels. System throughput Q and factor β are related to cell throughput Q_k and factor β_k as follows, where $\beta = \Lambda A/\mu B$:

G	D/r	P/P$_k$			P/P$_k$ (dB)		
		n = 3.0	3.5	4.0	n=3.0	3.5	4.0
3	3	76.4	157	324	18.8	22.0	25.0
4	3.46	117	259	573	20.7	24.1	27.6
7	4.58	271	692	1760	24.3	28.4	32.5
9	5.2	397	1080	2930	26.0	30.3	34.7
12	6	611	1780	5180	27.9	32.5	37.2
13	6.24	687	2040	6070	28.4	33.1	37.8
16	6.93	941	295	9230	29.7	34.7	39.7
19	7.55	1220	3980	13000	30.9	36.0	41.1
21	7.91	1400	4680	15700	31.5	36.7	42.0
25	8.66	1840	6430	22500	32.6	38.1	43.5
27	9	2060	7360	26200	33.1	38.7	44.2
28	9.17	2180	7850	28300	33.4	39.0	44.5
31	9.6	2500	9220	34000	34.0	39.7	45.3
36	10.4	3180	12200	46800	35.0	40.9	46.7

TABLE 6-1 Permitted numbers of channel groups G, normalized reuse distance D/r, and signal-to-co-channel interference ratio P/P_k.

$$\beta_k = \Lambda A_k/\mu B_k$$

$$= \beta G/N \qquad (6\text{-}3)$$

$$Q_k = \beta_k(1-P_B)$$

$$= QG/N \qquad (6\text{-}4)$$

Note that (6-4) reduces to (5-3) when G = N (no channel reuse). From (6-3) and (6-4) the advantages of channel reuse become clear; the system throughput Q exceeds the cell throughput Q_k by the factor N/G.

VI-2 Nineteen-Cell System with Seven-Cell Reuse Pattern

Consider a 19-cell system with seven groups of channels assigned as shown in Fig. 6-1. The geographic separation between cells with the same channel group is 4.58 r. All but the centre cell is subject to co-channel interference from two other cells. Thus, the average number of co-channel interferers is 38/18 ≈ 1.9.

Immediately adjacent channels in contiguous cells occur as follows: two channels of interference for the seven inner cells, one channel for ten of the outer cells, and no channels for the remaining two outer cells. Thus, on the average there are 24/19 ≈ 1.25 such channels per cell, at distance separation $\sqrt{3}$ r. Similar analysis shows 24/19 ≈ 1.26 channels per cell at separation 3 r. From these arguments it follows that

$$SNR^{-1} \simeq \phi \{(1.9\ C(0)/(4.58\sqrt{2})^n) + [(1.25/(\sqrt{6})^n) + (1.26/(3\sqrt{2})^n)]\ C(\Delta/R)\} \qquad (6\text{-}5)$$

For n = 3.5, (6-5) is as follows, where ϕ is the channel occupancy:

$$SNR^{-1} \simeq \phi\ (0.0027\ C(0) + 0.0627\ C(\Delta/R)) \qquad (6\text{-}6)$$

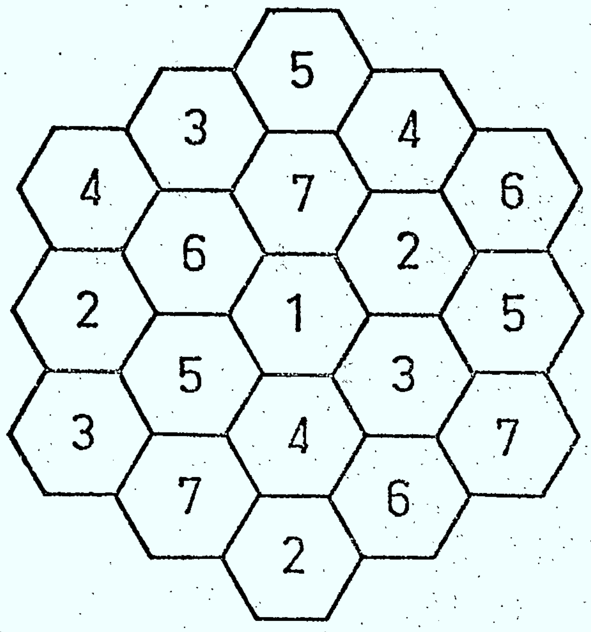

Fig. 6-1: Nineteen-cell system with seven-cell reuse pattern.

Constants 0.0027 and 0.0627 correspond, respectively, to -25.6 dB and -12.0 dB. Since $C(0) \simeq -3$ dB for MSK, SFSK and PAM and $\simeq -5$ dB for PSK, it follows that SNR $\lesssim$ 28.6 dB for the first three modulations, and $\lesssim$ 30.6 dB for PSK, assuming ϕ = 100%. For lower SNR values, adjacent-channel interference dominates, but as Δ/R increases, co-channel interference eventually limits SNR, unless ϕ is reduced. With adjacent-channel and co-channel interference equal, SNR $\simeq$ 25.6 dB for all but PSK for which SNR $\simeq$ 27.6 dB. The corresponding values of $C(\Delta/R)$ are (25.6-12) = 13.6 dB and (27.6-12) $\simeq$ 15.6 dB, respectively, for which $\Delta/R = 0.75$ for MSK and $\Delta/R = 1.4$ for PSK.

For the cellular scheme in Fig. 6-1, the throughput multiplier $N/G \simeq 2.7$.

Figure 6-2 shows maximum throughputs for 19-cell systems with and without channel reuse. To avoid congestion, SFSK and PAM curves are not shown; however, their behaviour is generally between that for PSK and MSK. The circles on the curves in Fig. 6-2 indicate the point at which channel occupancy ϕ must be reduced from unity for further SNR increases.

One sees from Fig. 6-2 that there is no one modulation scheme and channel assignment scheme which is best over all SNR levels. For 21 $\lesssim$ SNR $\lesssim$ 27 dB, MSK with channel reuse permits the highest throughput. At 25 dB, for example, throughput Q = 3.4 for MSK with reuse vs. 2.2 without reuse, and vs. 1.6 for PSK without reuse. For SNR $\lesssim$ 20 dB or SNR $\gtrsim$ 28 dB, absence of reuse yields higher throughputs for MSK. For PSK, differences between reuse and no-reuse are small for SNR $\lesssim$ 29 dB.

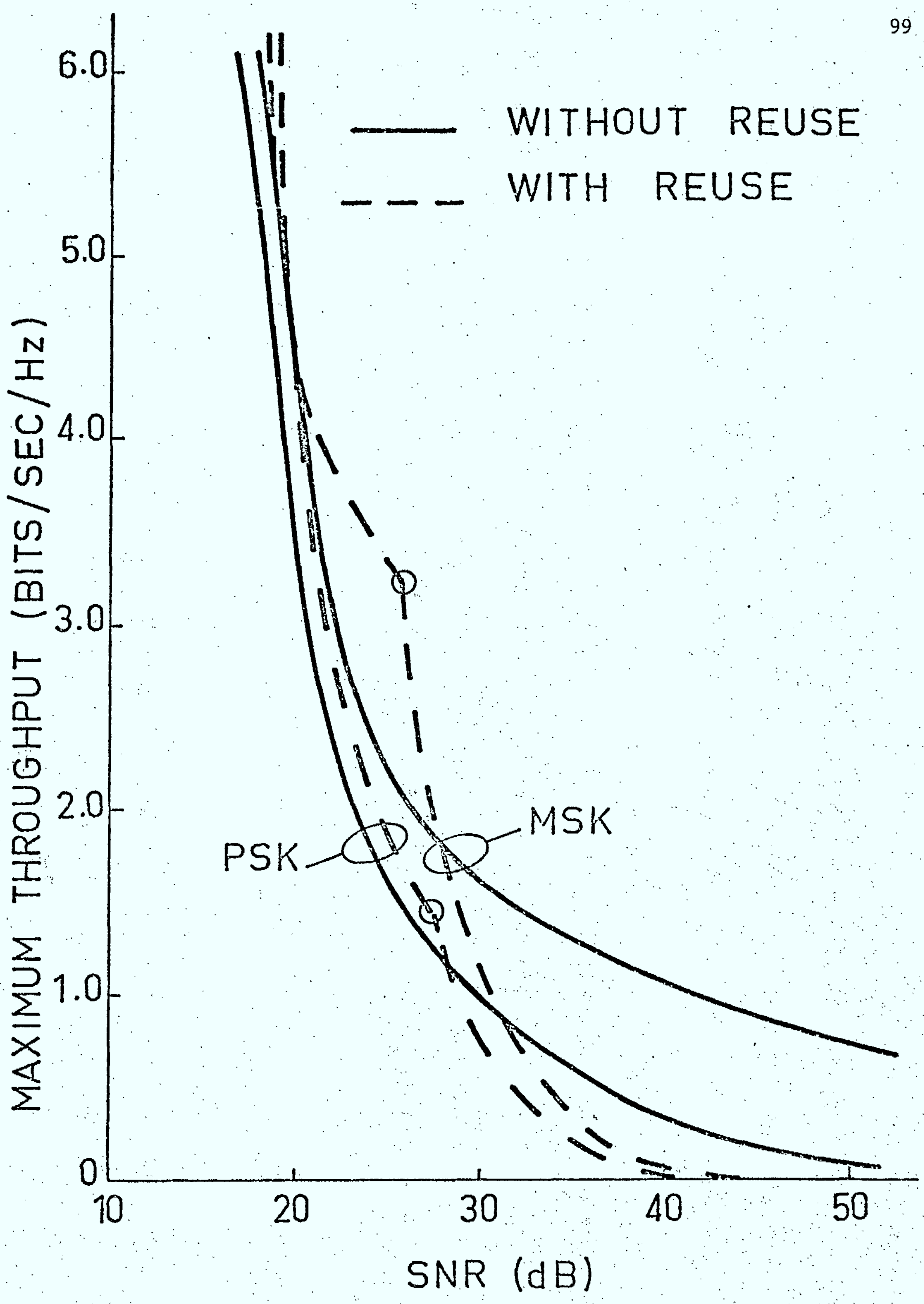

Fig. 6-2: Maximum throughputs for nineteen-cell systems with and without channel reuse; PSK and MSK.

VI-3 Nineteen-Cell System with Three-Cell Reuse Pattern

Figure 6-3 shows 19-cell systems having G groups of channels, where G = 3, 4 and 9.

In the 3-cell pattern, all seven interior cells are bordered by six cells, each of which may generate interference from an immediately adjacent frequency channel. Six of the twelve outer cells are bordered by four cells offering interference from immediately adjacent frequency channels; the other six are bordered by three neighbouring cells with channels separated by Δ Hz. All cells are subject to interference from immediately adjacent channels in more distant cells, and from channels 2Δ Hz or more from the wanted channel; however, the additional interference from these sources is negligible. Thus, for a channel Δ Hz away in a contiguous cell $P_k/P \simeq (\sqrt{6})^{3.5} \simeq 23.0$ (13 dB), while a channel in a cell whose centre is twice the distance has a P_k/P value $2^{-3.5} \simeq 0.089$ times 23.0. Thus interference from the more distant cell adds another 0.37 dB (1.089), which is negligible considering that the value of n is itself not accurately determined.

Co-channel interference for the centre cell is from six cells 3r units away. Each cell in the inner ring of six is subject to co-channel interference from four cells 3r units away and from one cell $3\sqrt{3}$ r units away; this latter contribution is safely ignored. Co-channel interference for outer cells is from either two or three cells 3r units away, two cells 6r units away and one cell $3\sqrt{3}$ r units away.

If we average the co-channel and adjacent-channel interference levels, we obtain co-channel interference from an average of 3.2 cells 3r units away, and immediately-adjacent-channel interference from 4.4 contiguous cells. Thus,

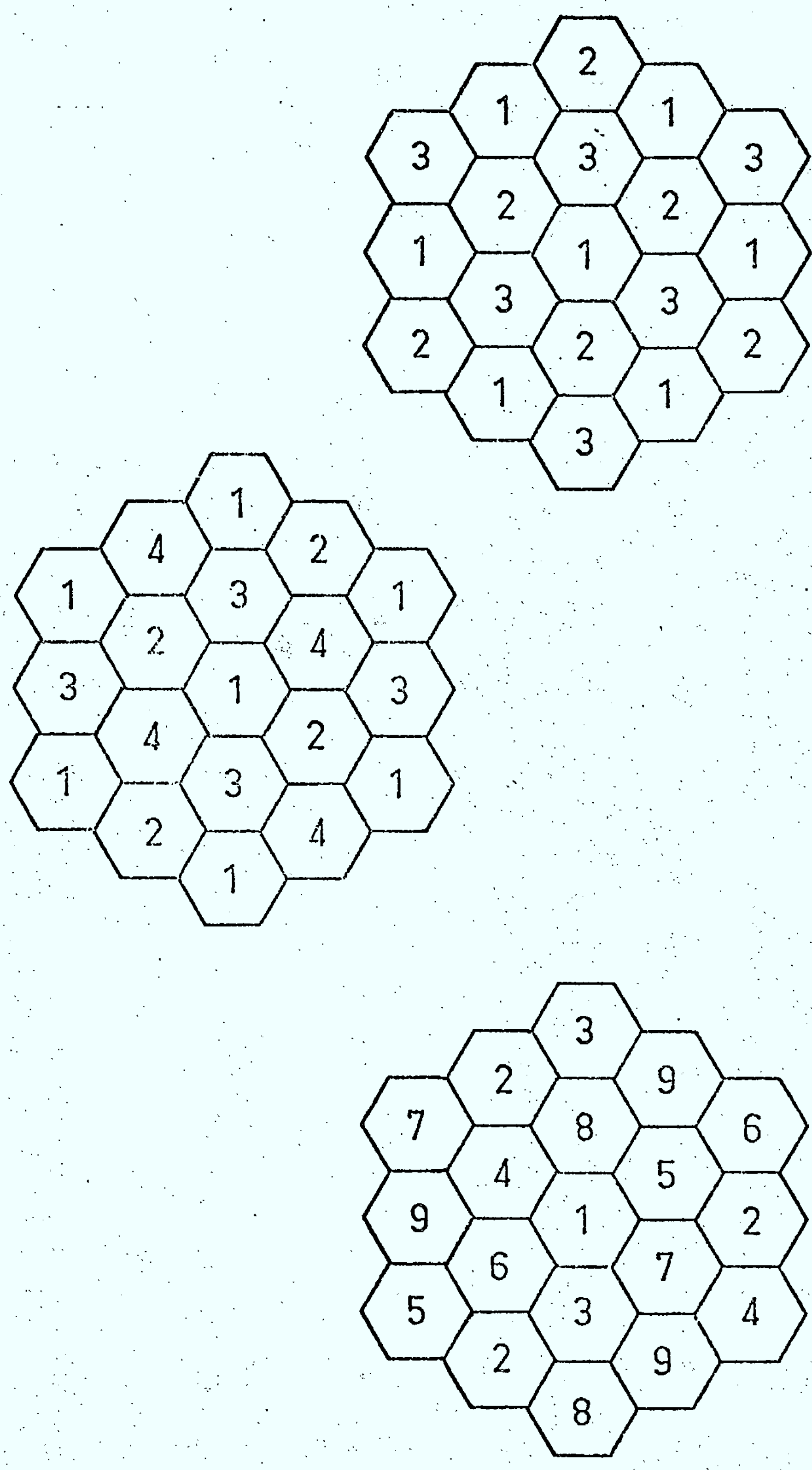

Fig. 6-3: Nineteen-cell system plans with three-, four- and nine-cell reuse.

$$SNR^{-1} \simeq \phi[3.2\ C(0)(3\sqrt{2})^{-n} + 4.4\ C(\Delta/R)(\sqrt{6})^{-n}] \tag{6-7}$$

$$\simeq \phi[0.020\ C(0) + 0.19\ C(\Delta/R)] \tag{6-8}$$

The coefficients 0.020 and 0.19 correspond to -17.0 dB and -7.20 dB, respectively. Since $C(0) \simeq -3$ dB for MSK, SFSK and PAM, and $C(0) \simeq -5$ dB for PSK, the maximum obtainable SNR values for $\phi = 1$ are 20 and 22 dB, respectively. When Δ/R is such that adjacent-channel and co-channel interference are equal, SNR is reduced by 3 dB from these values to 17 dB and 19 dB, respectively. In this case, adjacent-channel interference levels are $-20 + 7.2 = -12.8$ dB for MSK and -14.8 dB for PSK. For these levels, $\Delta/R = 0.74$ for MSK and 1.2 for PSK. For SNR $\lesssim$ 18 dB adjacent-channel interference dominates. However, as (Δ/R) increases to improve SNR, co-channel interference becomes dominant and eventually limits SNR unless ϕ is reduced.

Figures 6-4 and 6-5 show maximum throughputs for MSK and PSK, respectively, for various 19-cell systems, including the 3-cell reuse plan. The throughputs obtained are $19/3 = 6.33$ times the R/Δ values which result for various SNR levels. These R/Δ values are obtained from Fig. 2-9, and are not tabulated here, but are easily obtained by dividing ordinate values on Figs. 6-4 and 6-5 by 6.33.

VI-4 Nineteen-Cell System with Four-Cell Reuse Pattern

The discussion for the 4-cell pattern in Fig. 6-3 closely parallels that for the 3-cell pattern.

If co-channel and adjacent-channel interference are averaged,

$$SNR^{-1} \simeq \phi[2.84\ C(0)(2\sqrt{6})^{-n} + 2.95\ C(\Delta/R)(\sqrt{6})^{-n}] \tag{6-9}$$

$$\simeq \phi[0.011\ C(0) + 0.128\ C(\Delta/R)] \tag{6-10}$$

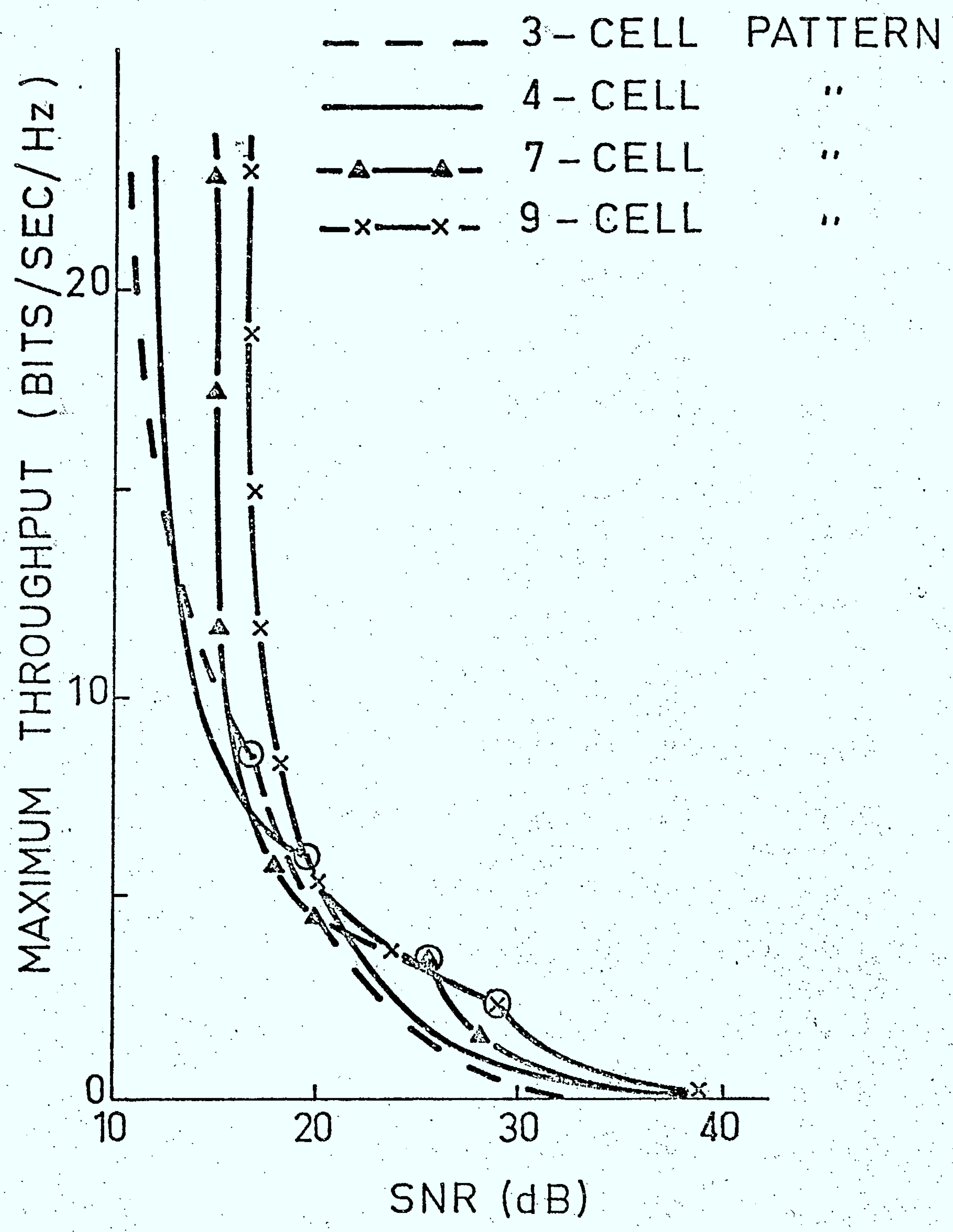

Fig. 6-4: Maximum throughput for 19-cell system with three-, four-, seven- and nine-cell reuse, MSK.

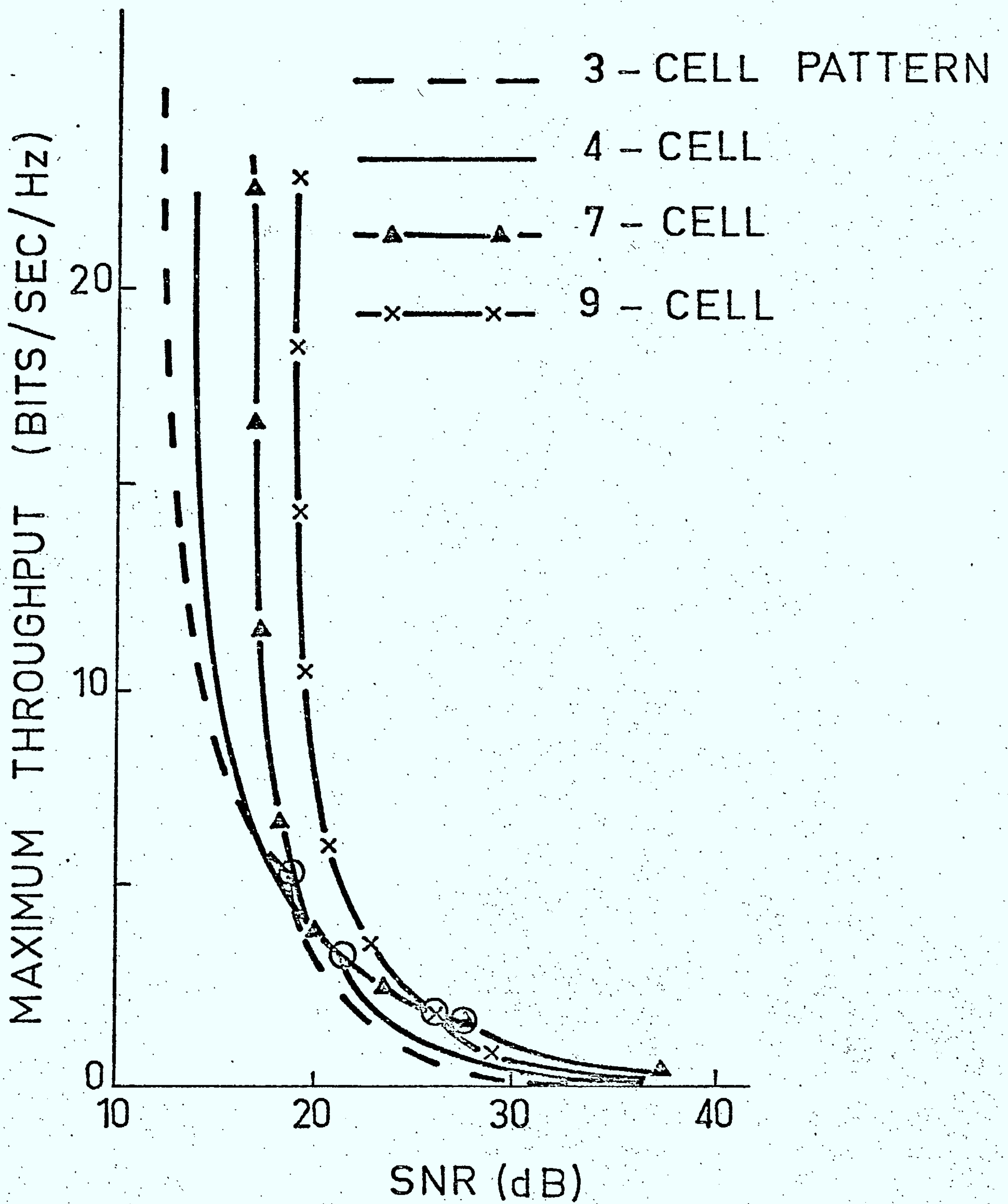

Fig. 6-5: Maximum throughput for 19-cell system with three-, four-, seven- and nine-cell reuse; PSK.

Coefficients 0.011 and 0.128 correspond to -19.6 dB and -8.9 dB, respectively. For MSK, SFSK and PAM, SNR cannot exceed 22.6 dB; for PSK the limit is 24.6 dB. When adjacent-channel interference equals co-channel interference, then SNR levels drop to 19.6 and 21.6 dB, respectively; $C(\Delta/R)$ = -13.7 dB for MSK and -15.7 dB for PSK, and $\Delta/R = 0.75$ for MSK and 1.45 for PSK.

Maximum throughputs, shown in Figs. 6-4 and 6-5, are obtained by multiplying R/Δ values by $19/4 = 4.75$.

VI-5 Nineteen-Cell System with Nine-Cell Reuse Pattern

The discussion for the 9-cell reuse pattern in Fig. 6-3 parallels that for the 3- and 4-cell cases. Averaging the co-channel and adjacent-channel interference yields

$$SNR^{-1} \simeq \phi[1.26\ C(0)(3\sqrt{6})^{-n} + 0.77\ C(\Delta/R)(\sqrt{6})^{-n}] \tag{6-11}$$

$$[0.0012\ C(0) + 0.034\ C(\Delta/R)] \tag{6-12}$$

In averaging the adjacent-channel interference, interference from contiguous cells, as well as from cells separated by 3r, has been included, since interference in cells 3r away accounts for almost 1/3 of the total.

Coefficients 0.0012 and 0.034 correspond to -29.2 dB and -14.1 dB, respectively. For MSK, PAM and SFSK, SNR cannot exceed 32.2 dB; for PSK the SNR limit is 34.2 dB. When adjacent-channel interference equals co-channel interference, these SNR levels decrease to 29.2 and 31.2 dB, respectively.

Figures 6-4 and 6-5 show the maximum throughputs vs. SNR, and are obtained by multiplying R/Δ values by $19/9 = 2.11$.

VI-6 Capacities of Larger Cellular Systems

The preceeding analysis extends easily to systems having more than nineteen cells. The numerical values of the quantities multiplying the co-channel term C(0) and the adjacent-channel terms C(Δ/R), C(2Δ/R), ... are determined as before. In most situations the C(Δ/R) term will dominate all other adjacent-channel interference terms. The numerical quantities referred to above may be averages determined as in the previous sections of this chapter, or may be worst case values; we favour averages as being more meaningful.

For example, a 49-cell system with a seven-channel reuse pattern, has an average co-channel interference at 100% occupancy of approximately $[3.5/(4.58\sqrt{2})^n]$ C(0), and an average adjacent-channel interference of $[2/(\sqrt{6})^n]$ C(Δ/R). For n = 3.5, these numerical coefficients equal 0.0051 (-23 dB) and 0.087 (-10.6 dB), respectively. Therefore, the maximum SNR obtainable is 23 dB for MSK, PAM and SFSK, and 25 dB for PSK. The maximum is seven times the R/Δ value at the specified SNR level, determined as illustrated previously. The terms multiplying C(0) and C(Δ/R) for the 19-cell system with a seven-cell reuse pattern equal -25.6 dB and -12.1 dB, respectively. Thus the curves in Fig. 6-2, if moved horizontally to the left by 1.5 dB and then scaled vertically by 49/19 = 2.58 represent maximum throughputs for the 49-cell system over the 100% occupancy portion of the curve.

The portions where ϕ is reduced are obtained as before. For MSK at SNR = 20 dB, the maximum throughput equals 10.0, as compared with 4.0 for the 19-cell system. Thus, a throughput increase of 250% requires a 258% increase in the number of cells.

Similarly, a 133 cell system with a 19-cell reuse pattern has a C(0) multiplier of $3.5/(7.55\sqrt{2})^{3.5} \approx -30.6$ dB. Thus, the maximum obtainable SNR for 100% occupancy is 30.6 dB unless PSK is used, in which case this maximum is 32.6 dB. For 50% occupancy, these values are 3 dB higher.

VI-7 Effects of Channel Reuse on Delay and Blocking Probability

Delay and blocking probability were obtained in Chapter 5 by determining minimum values for Δ/R, and by using (4-3) and the curves in Chapter 3.

Figures 6-4 and 6-5 can be used to find the maximum R/Δ values with ϕ = 50% by moving each curve 3 dB horizontally to the right, and by noting that the resulting ordinates equal $(N/G)(R/\Delta)$. Since for each cell,

$$\rho/m = \beta_k \, \Delta/R$$

$$= \beta(G/N)(\Delta/R) \qquad (6\text{-}15)$$

the resulting ordinate values yield ρ/m for any given SNR value and throughput factor β. For 100% occupancy, shifting of the curves is not needed.

For example, consider MSK with SNR = 20 dB, and a 19-cell system with a 4-channel reuse pattern. From Fig. 6-4 one obtains $(N/G)(R/\Delta) \approx 6$ at ϕ = 100%. To obtain the curves in Fig. 6-6, which shows throughput Q vs. number of channels/mobile for various P_B values, it is necessary only to move each point on the single-cell system curves in Fig. 5-4 four units horizontally to the right, and scale the ordinate values by 6.

Figure 6-6 clearly indicates the value of reusing channels. Not only is the maximum throughput of 6 with reuse greater than the 4.6 value with no reuse, the maximum throughput is approached much more rapidly, as the number of channels per mobile increases from unity. For a 40-channel system, for example, the difference in throughputs for P_B = 0.2 is 4.6 vs. 1.8; the former throughput is 2.56 times the latter. Even though reuse of channels introduces more adjacent-channel interference which results in a lowering of R/Δ, the increased channel capacity resulting from reuse usually more than compensates.

Table 6-2 shows throughput comparisons for PSK as well as MSK, for various channel assignment plans. Except for the 9-cell reuse plan in a

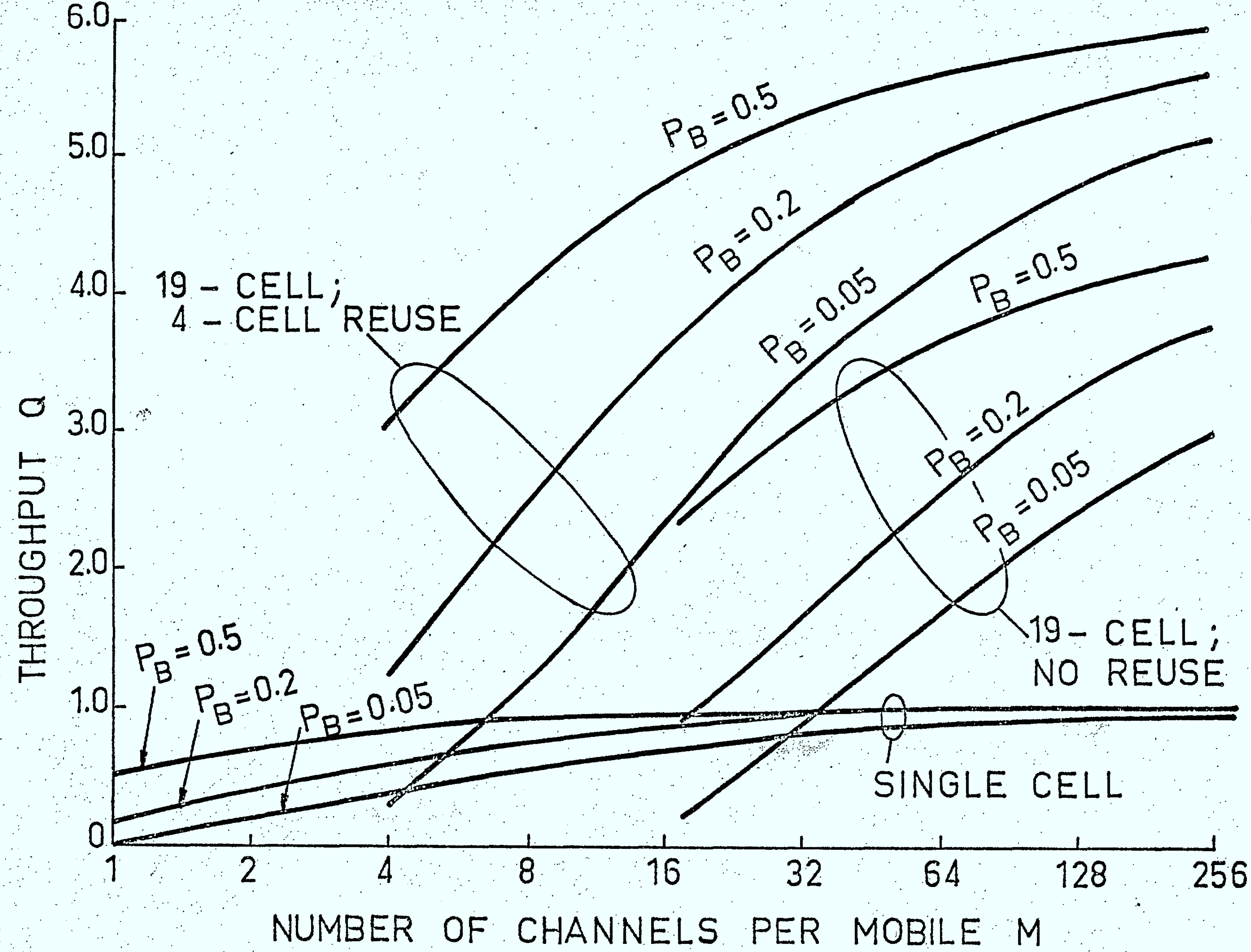

Fig. 6-6: Throughput for 19-cell system with 4-cell reuse pattern, P_B fixed; MSK, SNR = 20 dB.

TYPE OF SYSTEM	(N/G)(R/Δ)		THROUGHPUT Q			
			P_B = 0.2		μRD = 1.5	
	MSK	PSK	MSK	PSK	MSK	PSK
Single base	0.95	0.30	0.86	0.27	0.92	0.29
3 cells	1.5	0.95	1.2	0.76	1.4	0.86
7 cells	1.5	1.2	1.0	0.78	1.2	0.97
19 cells; no reuse	4.6	3.8	1.8	1.5	2.9	2.3
19 cells; 3-cell reuse	5.0	3.3	4.0	2.6	4.5	3.0
19 cells; 4-cell reuse	6.0	4.0	4.6	3.1	5.2	3.5
19 cells; 7-cell reuse	4.5	4.0	3.0	2.6	3.7	3.2
19 cells; 9-cell reuse	5.3	9	3.2	5.4	4.1	6.9

TABLE 6-2 Throughput (spectrum efficiency in bits/sec./Hz) for MSK and PSK for various systems. SNR = 20 dB; 40 channels/mobile; R/Δ based on 100% channel occupancy.

19-cell system, MSK is better than PSK; the anomaly cited is due to the lower value of $C(\Delta/R)$ for PSK for Δ/R near zero.

Similar arguments apply to delay considerations, as can be seen by comparing Figs. 3-3 with Fig. 3-2(b) which shows $d/m = \mu RD$ vs. ρ/m. Figure 6-7 shows Q for MSK with SNR = 20 dB, assuming 100% occupancy, with $\mu RD = 1.5$. Table 6-2 shows spectrum effeciencies of various systems, for SNR = 20 dB and $\mu RD = 1.5$, with 40 channels/mobile, and R/Δ based on 100% occupancy. The actual delay depends on μR, and in particular on any variation in R to accommodate various values permitted for R/Δ, as explained in Chapters 4 and 5.

VI-8 Effects of Propagation Factor n

Both the adjacent-channel and co-channel interference levels depend on the power ratio $P/P_k = (D\sqrt{2})^n$, where D is the distance between the centres of local and interfering cells, and n is the propagation factor. One sees that the interference in dB varies directly in proportion to the value assumed for n. Thus, if one assumes n = 3 and, in fact, n = 4, SNR in dB is 33% better than calculated, which implies a higher allowed value for R/Δ, with a direct increase in spectrum efficiency. The actual value of n depends in a complex way on various factors including base antenna height, topography, and distance separation between transmitter and receiver.

We do not attempt to summarize factors affecting n, but merely illustrate in Fig. 6-8 its effect on the maximum throughput for a 19-cell system with a 4-cell reuse pattern using MSK. Our value of 3.5 is probably safe; the value 4.0 is often used in analytical work.

Table 6-1 shows the variation of P/P_k for various reuse patterns for n = 3, 3.5 and 4.

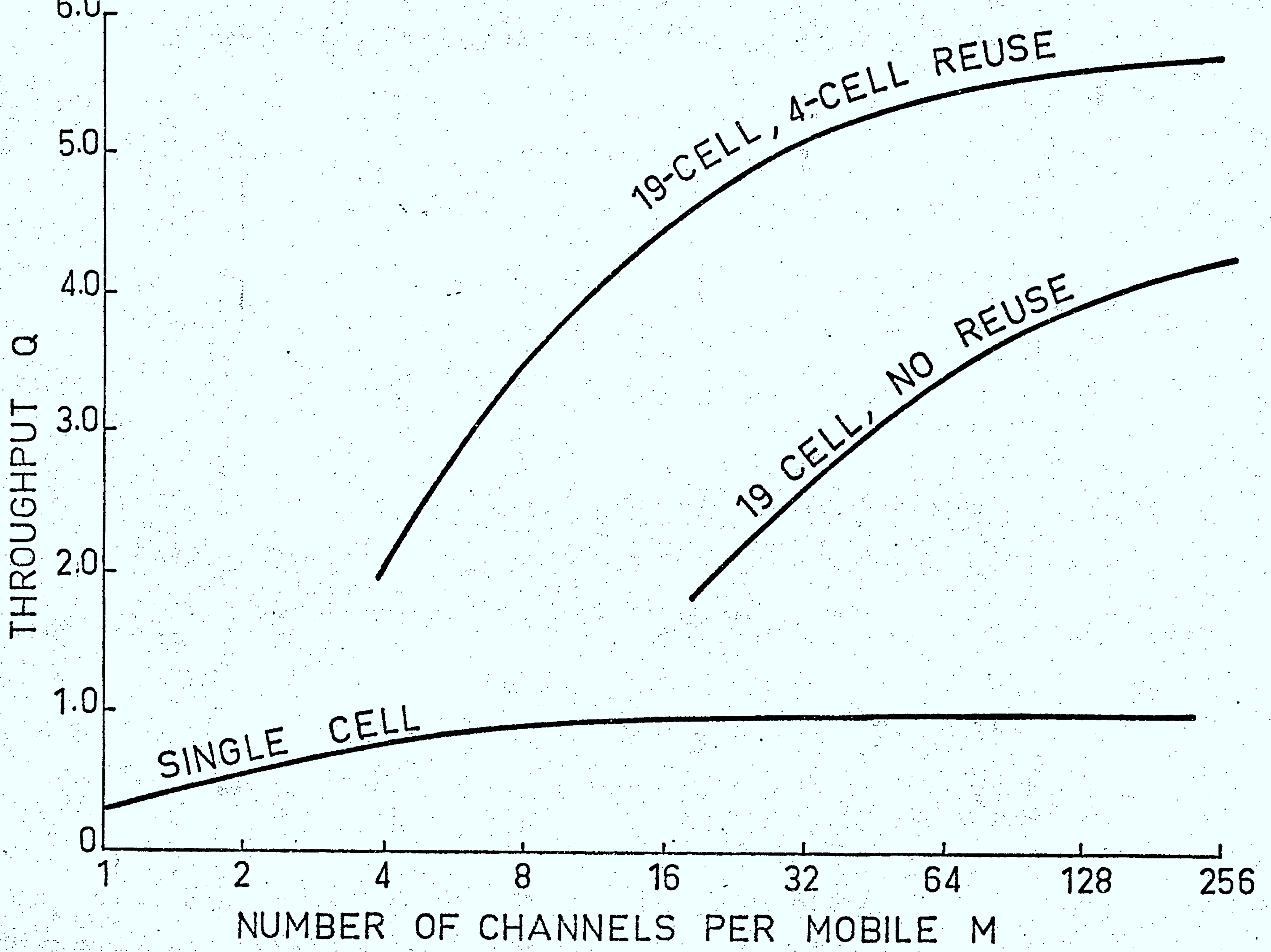

Fig. 6-7: Throughput for 19-cell system with 4-cell reuse pattern; μRD = 20, MSK, SNR = 20 dB.

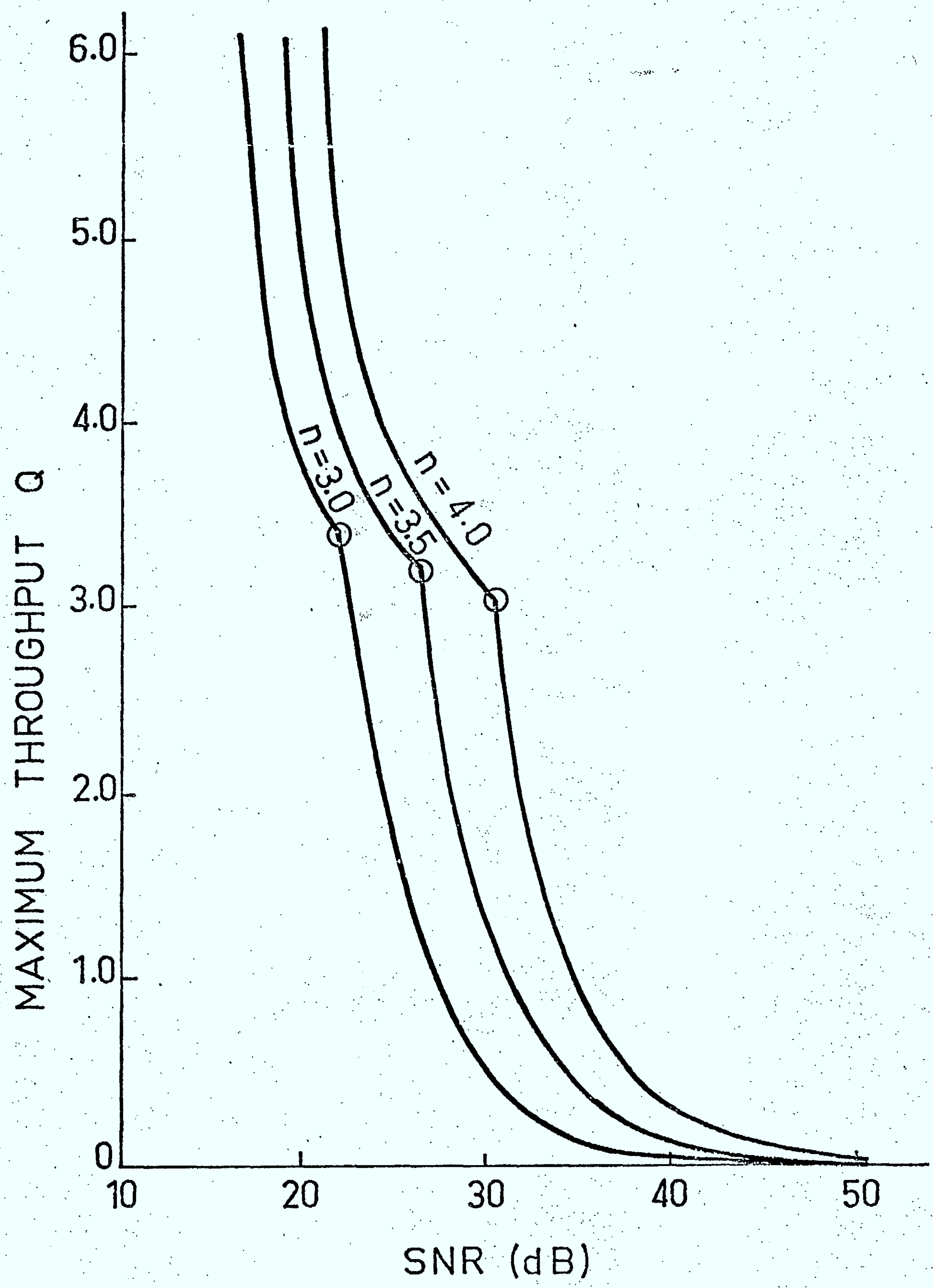

Fig. 6-8: Maximum throughput for 19-cell system with 7-cell reuse, for n = 3.0, 3.5 and 4.0; MSK.

VI-9 Effects of Suboptimum Receivers and Channel Occupancies on Throughput

The SNR equations used to determine throughputs in this and earlier chapters are based on optimum receivers.

As noted in Chapter 2, use of suboptimum receivers reduces SNR by up to 3 dB if DPSK is used in place of PSK, and up to 6 dB if incoherent detection of MSK replaces optimum (coherent) reception. The effect of suboptimum reception is to shift the throughput vs. SNR curves horizontally to the left by 3 or 6 dB.

Channel occupancy also has an effect on throughput. Reduction in ϕ from 100% to 50% increases SNR by 3 dB, and each further reduction of 50% involves an additional 3 dB SNR increase. In addition, reductions in ϕ cause corresponding reductions in throughput. Thus, the throughput vs. SNR curves are shifted horizontally to the right in accordance with the SNR increase, and the resulting ordinate values are then reduced by the percent reduction in ϕ.

Use of suboptimum reception and/or reductions in ϕ affect all curves in the same general way. Thus, comparisons between various system plans do not change for any given modulation format. However, the fact that DPSK involves a 3 dB SNR reduction as opposed to a 6 dB reduction for incoherent MSK improves the performance of suboptimum PSK relative to that of suboptimum MSK.

VII CODING, ERROR CONTROL AND BIT-ERROR PROBABILITY

VII-1 Coding for Error Correction and Error Detection

As noted in Chapter 2, the very large fluctions in signal-to-interference-plus-noise ratio cause correspondingly large fluctuations in raw bit-error probability. No matter how good the average SNR, bit-errors are unavoidable during deep fades. The only way to combat these fades is by some form of diversity. Channel coding is a form of time diversity which is highly effective on fading channels, either in a forward error correction (FEC) or error detection and retransmission (ARQ) mode.

Block codes and convolutional codes are available for FEC [L1, G1, P1, L2]. Block coders augment k-bit source-digit sequences. The r redundant check bits permit correction of up to t errors in an n-bit block where

$$t = \begin{cases} (d-1)/2 & d \text{ odd} \\ (d/2)-1 & d \text{ even} \end{cases} \tag{7-1}$$

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 types of decoders.

In contrast to block codes which check each source bit once, convolutional codes check each 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].

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

The most commonly used measure of block code performance is P_e, the probability of a block (or word) error. 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 [L2].

$$P_e = \sum_{m=t+1}^{n} P(m,n) \tag{7-2}$$

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

$$P(m,n) = \binom{m}{n} p^m (1-p)^{n-m} \tag{7-3}$$

where

$$\binom{m}{n} = \frac{m!}{n!(n-m)!} \tag{7-4}$$

As discussed later, (7-3) must be used with care in estimating error performance for land mobile channels.

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

$$p_b = P_w P_e \tag{7-5}$$

In (7-5), 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].

When error detection is employed using block codes, 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 [L2]:

$$P_c = P(m=0, n) \tag{7-6}$$

$$P_e \simeq 2^{-(n-k)} \sum_{m=d}^{n} P(m,n) \tag{7-7}$$

$$P_d = 1-P_c-P_e \tag{7-8}$$

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 (7-3) 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.

Decoder logic for error detection is simpler than for error correction. However, some additional memory storage is required at the transmitter, to store messages for possible retransmission, under various retransmission protocols.

VII-2 Effects of Coding on Throughput and Delay

Under FEC coding, the actual information throughput is k/n times the total number of bits transmitted. The delay increases by any additional time required to transmit parity bits, and for decoding.

When ARQ is used, additional delay as well as throughput reductions can occur as a result of retransmissions, depending on the type of ARQ protocol used. Basic ARQ protocols are summarized below [B1, B2, B3, F2, S2]:

1. Send-and-wait ARQ: Following transmission of a block, the sending terminal waits for either a positive acknowledgement (ACK) or a negative acknowledgement (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, in conventional line-switched systems considerable idle time may occur and reduce information throughput. Selective repeat systems have the highest overall system throughput but require numbering of the data blocks to facilitate identification for retransmission.

Variations of each of the above schemes are possible [K1, M1, S3, S4]. For example, stop-and-wait ARQ may involve transmission of a sequence of relatively short blocks (which are numbered), and only those in which errors are detected are included for retransmission in the next sequence. This option appears particularly suitable on land mobile channels, where bursts of noise may cause many errors in some blocks while others remain error-free. 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, P2].

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 [S2]

$$\eta = \frac{k}{n} \frac{1}{1+N[P_d/(1-P_d)]} \tag{7-9}$$

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

Equation (7-9) also applies to both stop-and-wait and selective ARQ in which cases N=1 and

$$\eta = (k/n)(1-P_d) \tag{7-10}$$

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

$$\eta \simeq P_c \; k/n \tag{7-11}$$

The degradation in throughput efficiency of ARQ relative to FEC is obtained directly from (7-9). For N=1, Fig. 7-1 shows this degradation $\tilde{\eta}$ for a block codes of length n = 15, 31,..,1023. For n = 31, $p \gtrsim 10^{-3}$ implies $P_c \gtrsim 0.9$ 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. Notwithstanding, for some types of messages, including control messages as described in Chapter 2, some form of ARQ seems essential to enable the transmitter to know whether or not its message was received.

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

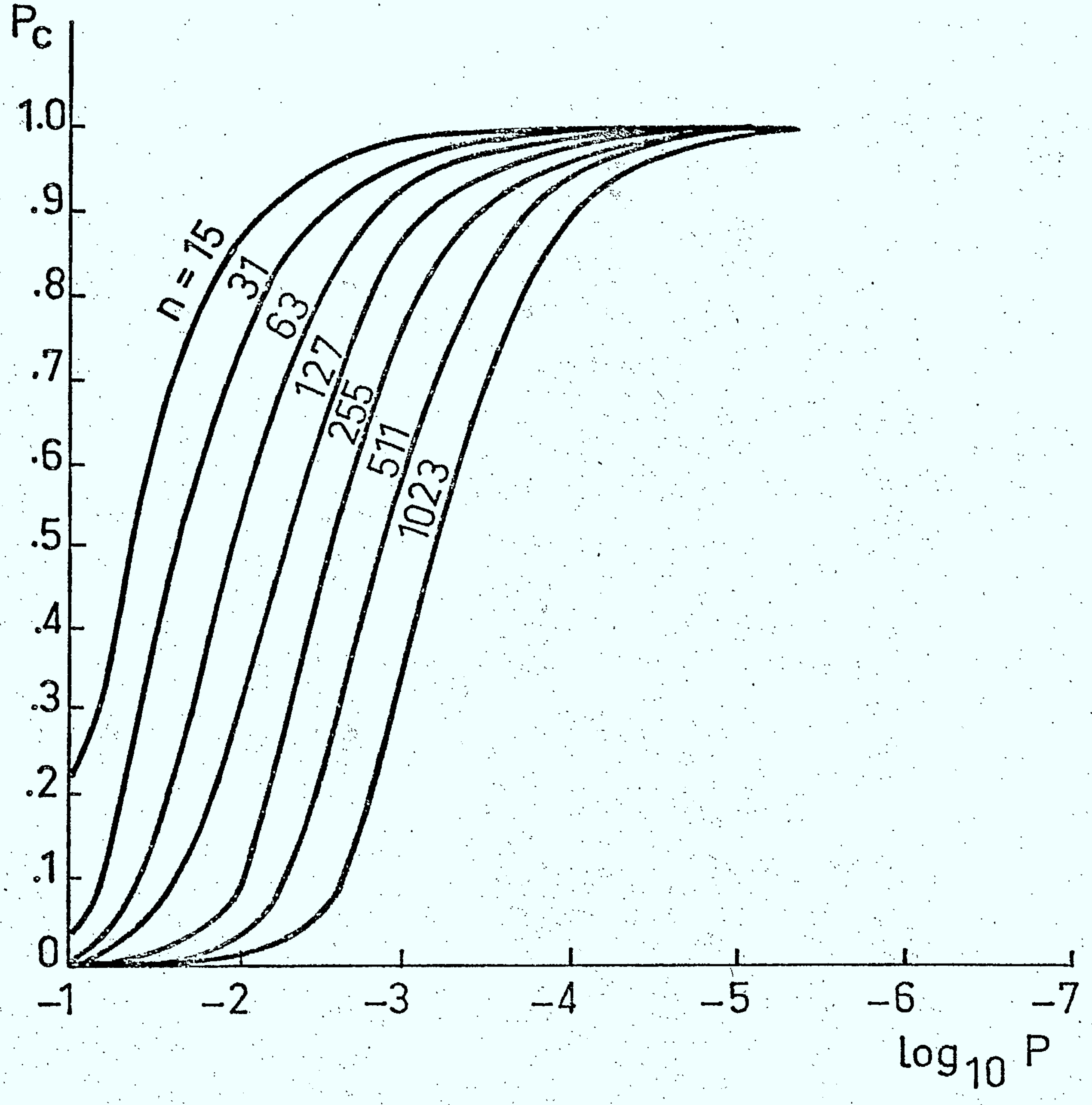

Fig. 7-1: Probabilities of correct decoding for ARQ for various block lengths n.

delays including those due to queueing for channel access, transmission of bits over the channel, propagation, decoding, and modem start-up or turn-around. Let T_1 and T_2 denote the one-way and round-trip delays, respectively (often $T_2 \simeq 2T_1$); then

$$D = T_1 + T_2 \left(P_d + 2P_d^{\,2} + 3P_b^{\,3} + \ldots. + iP_d^{\,i} + \ldots.\right)$$

$$D/T_1 = 1 + (T_2/T_1)\,[P_d/(1-P_d)^2] \quad (7\text{-}12)$$

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%.

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.

VII-3 Effects of Coding on Spectrum Effeciency

To demonstrate the use of channel encoding in land mobile radio systems, we begin by considering alternative means of obtaining bit-error probabilities p_b after decoding in the range 10^{-3} to 10^{-4}. This value of p_b would be suitable for real-time digital voice transmission. A Rayleigh channel model is used, which model would be valid for short block lengths with interleaving (see Sections 2-4 and 7-4) and many interferers whose combined interference level would be relatively constant. For this model, it follows from Section 2-7 that

$$p \simeq (SNR)^{-1} \quad (7\text{-}13)$$

We assume SNR to be calculated as in earlier chapters, which implies that interference dominates noise, and that p as given by (7-13) is for radius $r/\sqrt{2}$ from the base. Obviously, p will vary with respect to distance between mobile and base, as discussed in Chapter 2 and in Section 7-4.

From the curves in Chapter 6 one sees that a reuse plan must include at least G=9 groups of channels for SNR ≃ 30 dB ($p \simeq 10^{-3}$) unless throughput is to be reduced below $\phi=1$. (As explained earlier, an actual ϕ value well below unity is often compensated for, in SNR, by suboptimum receivers.) Analysis procedures used in Chapter 6, together with Table 6-1 shows that G=12 is needed for SNR ≃ 35 dB with $\phi=1$, and G=16 is needed for SNR ≐ 40 dB ($p \doteq 10^{-4}$). For a 19-cell system with no reuse and SNR = 35 dB ($p \simeq 3 \times 10^{-4}$), the maximum MSK throughput Q ≃ 1.35 (See Fig. 5-8), if direct transmission of bits occurs.

If a 4-cell reuse pattern is employed, SNR ≃ 20 dB is the maximum value permitted with $\phi=1$, in which case Q≃6. If a double-error correcting (15, 7) FEC code is used, $p_b = 4.6 \times 10^{-4}$, and the maximum information throughput is (7/15)x6 = 2.8. Use of a triple-error correcting (31, 16) code yields $p_b = 3.2 \times 10^{-4}$ and Q = 3.1. Both maximum throughput levels are more than double that for the 19-cell system with no reuse of channels. More important, information throughput achieved for given P_B or μRD values and a given value of m favour the 4-cell reuse plan even more strongly. For example, with 40-channel mobile transceivers and $P_B = 0.20$, Q = (16/31) x 4.6 = 2.4 for a 4-cell reuse plan with the (31, 16) code vs. Q = 1.35(0.4) = 0.5 for a 19-cell system with no reuse. Clearly, FEC with small G values can greatly affect the spectrum efficiency.

Similar comparisons are easily made for 3-, 7- and 9-cell systems. For example, for a 7-cell system with $\phi=1$, the maximum SNR = 26 dB

($p \approx 2.5 \times 10^{-3}$), in which case Q = 3.4. Use of a shortened (11, 7) code yields $p_b = 3.4 \times 10^{-4}$ and Q = (7/11)x3.4 = 2.2. (The (11, 7) single-error correcting code is obtained by shortening the (15, 11) code [L1]. Shortening permits construction of simple codes of virtually any length and code rate.) Similarly, for a 3-cell reuse plan, SNR = 17.5 dB ($p = 1.8 \times 10^{-2}$) is the largest value possible with $\phi=1$, in which case Q = 8.5. Use of a (15, 5) triple-error correcting code results in $p_b = 1.4 \times 10^{-4}$ and maximum Q = 2.8, a value equal to that for the 4-cell reuse plan with a (15, 7) double-error correcting code. However, for a fixed number of channels per mobile, and fixed P_B or μRD value, the 3-cell pattern has a higher throughput, because of the increased number of channels per cell. However, as discussed in the following section, the close proximity of the co-channel interferers may cause too much fluctuation in p_b as the mobile location varies.

Consider next codes from those above (all of which have simple hardware decoders) for error detection. For the (15, 7) code in the 4-cell reuse plan with $p = 10^{-2}$, $p_b = 1.2 \times 10^{-9}$, and $P_c = 0.86$. Use of a (31, 16) code in the same situation yields $p_b = 2.4 \times 10^{-14}$ with $P_c = 0.72$. The latter value of p_b may be needlessly small, and may not justify a throughput reduction to 72% of the FEC value, vs. 84% for the (15, 11) code. In some applications, $p_b = 10^{-9}$ may be more than adequate, in which case a (15, 11) code could be used with $p_b = 2.8 \times 10^{-5}$; the corresponding maximum throughput, including the 0.86 factor would be (11/15) x 6.0 x 0.86 = 3.8.

Our purpose here has not been an exhaustive study of code performance, but rather to illustrate the greatly increased flexibility in system design offered by channel encoding. The implications of this flexibility are discussed in Section 7-5.

VII-4 Bit-Error Probability Variations

In an actual mobile radio system, raw bit-error probability p varies as mobiles move, because of shadowing and changes in base-mobile distance separations. Interference levels also vary, for the same reasons. As p varies, so does p_b.

One way to account for this variation is to average p_b over the coverage area, and over the variations in mean signal level. Unless there are many interfering signals, effects of interference level variations should also be averaged, as explained in an earlier report [D1]. In averaging p_b, it is reasonable to assume that both shadowing levels and distance separation remain constant during transmission of one block of data, in accordance with the discussions in Section 2-4. It is also important that shadowing correlations between the wanted and interfering signals be included [F3].

Perhaps the greatest difficulty in accurately averaging p_b is to estimate P(m,n), on which p_b depends. When a vehicle is stationary, P(m,n) is given by (7-3) with p calculated as for a Gaussian channel, provided, in the case of mobile-to-base interference, interference levels are stationary. When a vehicle is moving fast enough to encounter several fades during transmission of blocks of interleaved bits, the Rayleigh approximation for p is valid. In the absence of interleaving, or during stop-and-go vehicle movement, P(m,n) exhibits memory, and the channel appears to include both random and burst-error behaviour [M2]. Coding gains observed on real channels are considerable [M2], and are not inconsistent with those based on Rayleigh models with independent errors [D1] in random noise. Bit error probabilities calculated for independent-error Rayleigh channels may provide reasonable estimates even when channel memory is involved, since bursts of errors can provide for higher p_b values than random errors, with the same raw bit error rate p. The reason is that under burst errors, only selected blocks are damaged, while random errors are not confined to selected blocks. In any event, rigorous calculation of averages for p and p_b would be difficult.

Whether or not an average for p_b is calculated, it is clear that p_b will improve, in general, as the mobile moves in close proximity to the base. At small base-mobile separations, both p_b and p will be very good. The actual best values will depend on the minimum mobile-base separation.

Conversely, on the cells' outer edges, p and p_b will be at their worst. An alternative (or complementary) approach to averaging p_b is to compare these worst values with nominal values obtained assuming the mobile-base separation of $r/\sqrt{2}$. Consider a 4-cell reuse plan, with MSK; the worst case occurs with a mobile at one of the six apexes of the centre cell. In this case, $P/P_k = 1$ for the adjacent-channel interference in one of the contiguous cells, which can be reduced, if necessary by reducing R/Δ. For example, reduction from the maximum value of 1.25 to 1 reduces $C(\Delta/R)$ from -15 dB to -21 dB.

For co-channel interference, with n = 3.5 and 4-cell reuse

$$P/P_k \simeq ((1.5\sqrt{3})^n/2)/C(0)$$

$$\simeq 14.5 \text{ dB} \qquad (7\text{-}14)$$

This level of interference implies $p = 10^{-1.45} = 3.6 \times 10^{-2}$. For double-error correction using a (15, 7) code, $p_b = 2 \times 10^{-2}$, in comparison with the nominal value 4.6×10^{-4} obtained earlier. Use of the same code for error-detection yields $p_b = 6.8 \times 10^{-7}$ vs. the nominal 1.2×10^{-9}. The throughput reduction factor is $P_c = 0.63$, vs. the nominal 0.84. The increase in delay D in (7-12) for $T_2/T_1 = 2$ is $P_d/P_c^2 = (1-P_c)/P_c^2 = 0.93$ (an increase of 93%) vs. the nominal 0.25 (25%).

The effects of shadowing on the mean signal and interference levels depend heavily on shadowing correlations. For 100% correlation, there is no effect. For no correlation, the effect can be very large; typically the mean level variations have a 3 dB standard (power) deviation about the mean. If two standard deviations occur in SNR, the result is a factor of $\pm 10^{0.6} = 4$ variation in p, and for a t-error correcting code, the variation factor approximates 4^t. For a 2t error-detecting code the variation factor is 4^{2t}. Thus, for t = 2, p_b could vary by factors of ±16 and ±256, respectively, for FEC and ARQ, with corresponding variations in throughput and delay reduction factors.

A similar analysis is possible for all other channel assignment schemes. The variation in co-channel interference decreases with an increase in G, and this fact may favour somewhat larger G values than would otherwise be indicated.

VII-5 Data Block Structures for Land Mobile Radio Systems

An important determinant of spectrum efficiency is the data block structure, including its relationship to FEC or ARQ codes.

Figure 7-2 shows a structure suitable for land mobile radio channels. The data block includes k information bits, r = n-k check bits, v address bits, s synchronization bits and c mode control bits. Included in the n information plus check bits may be g bits for acknowledgement of messages transmitted in the other direction. The n bits may consist of short sub-blocks which are decoded separately. The bits in any sub-block may be spread out (interleaved) over the longer n-bit sequence. Interleaving is probably helpful if FEC is used [L1, C1], but may cause too many retransmis-

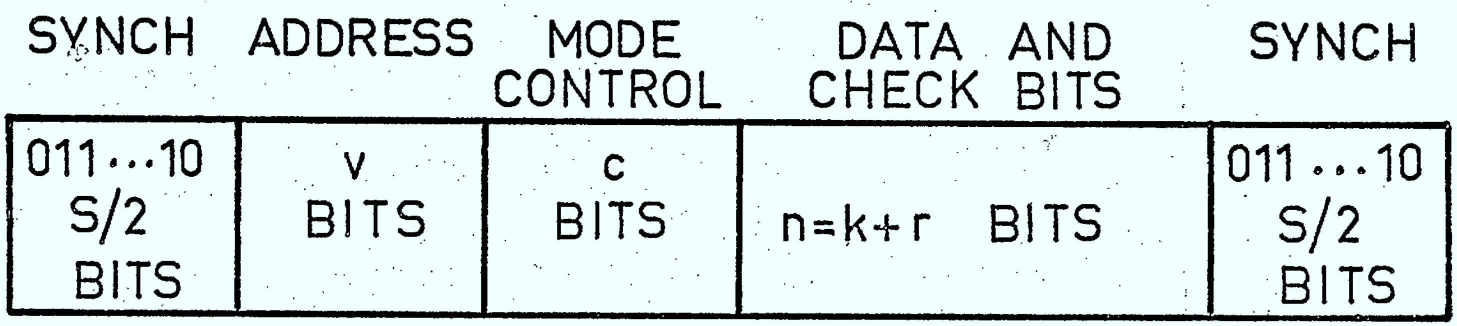

Fig. 7-2: Message block structure.

sions if ARQ is employed. The mode control bit(s) indicate to the decoder whether FEC or error-detection, or a combination of these is to be performed.

FEC or direct transmission with no coding would be used for digital transmission of speech, ARQ for control messages, and either a combination of FEC and ARQ, or FEC alone for text. In situations where all three types of messages can occur, some form of mode control seems essential - address and mode control bits could be spread out over the block, and made very reliable by many check bits.

In an ARQ environment, retransmitted sub-blocks can be combined with those received earlier for improved FEC decoding; it has been proposed, for example, that subsequent retransmissions consist solely of additional bits to check the earlier information bits, creating a 2n, k sub-block. Numerous possibilities exist with ARQ to combat the large fluctuations in raw bit error probability p. Obviously, tradeoffs exist between coder/decoder hardware complexity and spectrum efficiency. For FEC, the only alternative is to reduce the information bit rate when p degrades, for example near cell edges; in real-time transmission, such reduction implies a change in the source encoder, either in the sampling rate or number of quantization levels, or both. Such on-line changes would be possible for adaptive differential PCM [H1, C3] but would result in some increase in equipment complexity.

Values for k, r, v, s, c and g as defined above, should be selected to maximize information throughput, which is equal the number of bits/sec transmitted times the factor

$$\tilde{\eta} = \frac{k}{n+v+s+c+g} = (k/n)(n/[n+v+s+c+g]) \tag{7-15}$$

For FEC, g = 0. For ARQ the above factor $\tilde{\eta}$ is multiplied by P_c, and the spectrum efficiency is $\tilde{\eta}\cdot P_c\cdot Q$. Optimization of $\tilde{\eta}\cdot P_c\cdot Q$ is complicated by the fact that P_c and Q depend on k, n, v, s, c and g. The nature of the dependence on k and n was illustrated in the previous section; while (7-15) may suggest values of $k/n \simeq 1$ as best, smaller k/n values actually yield larger values of Q, by permitting greater reuse of channels.

The ARQ protocol used influences the choice of block lengths. Send-and-wait ARQ favours long blocks, particularly when acknowledgement delays are large, in order to reduce the idle channel time. Selective ARQ favours short sub-blocks. Chu [C2] 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. He did not, however, include consideration of sub-blocks.

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. [F2] 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 prior to transmission after every u-1 bits and removed following decoding. It can be shown that $u=1 + \sqrt{\alpha}$ minimizes the synchronization overhead when a synchronization sequence preceeds a data block containing α bits, excluding synch. bits, and that $u=1 + \sqrt{\alpha/2}$ is optimum when the sequence also terminates the data block. For this latter case with α = 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 [S1].

VII-6 References for Chapter 7

[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] 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] 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.

[C3] D.C. 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.

[D1] R.W. Donaldson, "Frequency assignment for land mobile radio system: Digital transmission over land mobile channels - interference considerations," Report to Dept. of Communications, Ottawa, Jan. 1980.

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

[F2] 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.

[F3] R.C. French, "The effect of shadowing on channel reuse in mobile radio", IEEE Trans. Veh. Technol., vol. VT-28, pp. 171-181, Aug. 1979.

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

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

[K1] L. Kleinrock, <u>Queueing System, Vol. 2: Computer Applications</u>. New York: J. Wiley, 1976, Ch. 5 and 6.

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

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

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

[M1] J.M. Morris, "Optimal block lengths for ARQ error control schemes", IEEE Trans. Commun., vol. COM-27, pp. 488-493, Feb. 1979.

[M2] P.J. Mabey, "Mobile radio data transmission-coding for error control", IEEE Trans. Veh. Technol., vol. VT-27, pp. 99-110, Aug. 1978.

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

[P2] 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] J.J. Stiffler, Theory of Synchronous Communication. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[S2] A.R.K. Sastry, "Performance of hybrid error control schemes on satellite channels", IEEE Trans. Commun., vol. COM-23, pp. 689-695, July 1975.

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

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

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

VIII DISCUSSION AND CONCLUSIONS

VIII-1 Summary and Implications of Results

This report defines quantitative criteria for designing and evaluating land mobile radio systems. These criteria include delay or blocking probability, reliability and throughput, and are expressed in terms of the important system variables. Cost, which changes rapidly with technology, is not expressed precisely, but is related in a semi-quantative way to system variables.

The criteria are used to determine performance obtainable with cellular systems using conventional frequency division multiple access, and fixed channel assignments. Particular emphasis is placed on the limitations resulting from co-channel and adjacent-channel interference, since this ultimately limits information throughput. Maximum R/Δ values for various modulation formats and channel assignment schemes are determined, for fixed values of SNR, defined at a particular cell radius. Throughput (spectrum efficiency) in bits/sec/Hz is determined for various SNR values, modulation formats, assignment schemes and channel codes, for various values of delay or blocking probability and number of channels per mobile. The way in which channel encoding can be used to enhance spectrum efficiency and provide additional system design flexibility is illustrated for both forward error-correction and error-detection/retransmission modes.

Specific results are obtained for systems of up to 19 cells, with and without reuse of frequency channels. Such systems are typical of those in use or being proposed for immediate use. For given values of delay or blocking probability, highest throughputs are obtained when frequency reuse is employed, together with channel encoding. The best code rate depends on the reuse plan. The appropriate error control protocol depends on the type of message to be transmitted; forward error correction (FEC) is best for real-time speech, FEC or ARQ are appropriate for text, and ARQ is best for control messages where accuracy requirements are paramount.

VIII-2 Policy Questions

The results of this study bear directly on the following spectrum management policy questions:

1. What value should be selected for channel spacing Δ?
2. What spectral emission constraints, if any, should be specified?
3. Should all channels be used in an unrestricted way for both analog and digital transmission, or should separate channels be used for each?
4. What system designs should be encouraged, in the interests of spectrum efficiency?

Selection of Δ relates to channel bit rate R, since adjacent-channel interference depends on R/Δ. In the absence of interference, receiver front-end noise limits R; however, this limit will vary with coverage area, transmitter power, and receiver quality, including diversity capability. Perhaps some nominal SNR limit should be agreed upon, for design purposes.

When interference is present, with MSK modulation and SNR = 20 dB, the maximum R/Δ value is in the order of 1.0 for single-cell systems, 1.5 for systems having between three and seven cells with no reuse of channels, and 1.25 for systems with 4- and 7-cell reuse patterns. For higher SNR values, or for PSK modulation, R/Δ is less than these values.

If blocks of channels are reserved solely for digital transmission, the range 10 KHz $\lesssim$ R $\lesssim$ 25 KHz seems reasonable. This choice for Δ implies a similar range for R. The range 15 KHz $\lesssim$ R $\lesssim$ 20 KHz would permit digital transmission of speech using adaptive differential PCM, or even non-adaptive PCM at lower quality, with some forward error correction. Data could also be transmitted at the same R value, with channel encoding used to select the rate and accuracy of the decoded information bits. At R $\approx$ 15 KHz, for example, a rate 1/3 code would yield an information bit rate compatible with a 4800 bit/sec telephone channel, and would provide for considerable error protection. The code rate could be selected in accordance with interference levels, by means of mode control bits in the data blocks, as explained in Section 7-5.

For ease of spectrum management, as well as for fullest utilization of all channels, it is desirable to use all channels for both analog FM voice and data transmission. The upper range of Δ (25 KHz) specified above is compatible with existing analog FM systems; the lower range (10 KHz) is close to the narrower channel bandwidths being considered by some agencies. What is presently lacking are sufficiently reliable estimates for adjacent-channel interference between voice and data, as a function of data modulation format, data rate, analog signal "bandwidth" and pre-modulation processing of voice. Of particular interest is the spectral roll-off rates of FM voice signals, since these rates ultimately determine interference levels.

Existing spectral emission constraints issued by various agencies are either in terms of spectral bounds like those in the "top-hat" characteristics in Chapter 2, or in terms of energy collected by specified bandpass filters centred on immediately adjacent channels. These criteria are helpful and reasonably easily administered. However, they ignore spectral roll-off rates, which ultimately determine maximum information throughput. Simply stated, fast roll-off rates imply higher R/Δ values for a given level of interference, and better spectrum efficiency.

When conventional FDM systems are used solely for digital transmission, reuse of relatively few groups channels maximizes information throughput, for given values of blocking probability or delay and numbers of channels per mobile. As few as four groups may be best. A more thorough analysis of bit-error probabilities is needed to provide more accurate estimates of spectrum efficiency for various reuse plans. When channels are permitted unrestricted use by both analog and digital transmissions, additional work on mutual interference is needed to determine the best reuse patterns.

At this point, conclusive comparisions between conventional FDM cellular systems, and newer systems such as SSMA seem to be unavailable. However, the performance criteria established in this work could be applied, in order to make such comparisons.

VIII-3 Issues Requiring Further Study

From the discussions in the previous section, follow suggestions for further work:

1. There is a need to have a better understanding of the mutual interference effects, both co-channel and adjacent-channel, between analog voice transmissions and digital transmissions. Some mutual interference calculations are in an earlier report by the author, and are based on Gaussian speech models. More work is needed with other models, and some measurements would eventually be required, for confirmation. Of

particular interest is the selection of pre-modulation filters to provide good quality speech and reduced interference levels via fast spectral roll-off rates.

2. More effort is required to determine both raw and decoded bit-error probabilities for digital transmission. It is not clear whether averages over a coverage area or some other measures of bit-error probability, such as nominal and worst case values would be most appropriate. The answer will depend in part on the availability of channel statistics such as P(m,n), and on required computational efforts. Eventually, field test results should be compared with calculations.

3. Other schemes for sharing the radio spectrum should be evaluatedin terms of the criteria used in this study. Alternatives include SSMA, packet radio transmission, and dynamic channel assignments in FDM cellular systems. Interference considerations are of particular importance.

4. More work is needed on the evaluation of various error control protocols for ARQ systems, in conjunction with data block format design. Of particular importance are the implications on implementation cost, delay and spectrum efficiency.

5. Evaluations of digital speech encoders for land mobile radio applications are needed. Comparisons between costs, bandwidth requirements, and speech quality comparisons between analog and digital systems are needed. Present technology suggests that adaptive differential encoders may be more spectrally efficient than analog systems.

APPENDIX A - LIST OF SYMBOLS

A Performance Criteria

D - queuing delay

d - queuing delay relative to μmR

P_B - blocking probability

Q - system throughput (spectrum efficiency)

ϕ - probability of a channel being in use (channel occupancy)

SNR - data receiver output signal-to-noise-plus-interference ratio

p - raw bit-error probability

B Message Source Parameters

λ_v - call attempt rate per vehicle (messages/sec)

Λ - call attempt rate per unit area (messages/sec/m)

μ^{-1} - average message length (bits/message)

$n(x,y)$ - area density of mobiles

C System Parameters

N - number of base stations

B - total system bandwidth (Hz)

Δ - channel separation (Hz)

R - bit rate (bits/sec)

m - number of channels per mobile

P_s - received power of data signal (watts)

P_d - received power of interfering signal (watts), also probability of a detected block error

n - propagation factor; also channel code block length

$N_o/2$ - power spectral density of noise (watts/Hz)

T - bit period ($T = R^{-1}$) (sec.)

C_d – interference coefficient

M – total number of channels in system ($M = B/\Delta$)

r – cell radius (m); also $r = (n-k)$ – number of check bits

λ – message arrived rate per group of m channels (messages/sec)

ρ – utilization factor for m-channel group ($\rho = \lambda/\mu R$)

k – number of information bits per block

P_c – probability of a correctly decoded block

P_e – probability of error in a block of bits

η – throughput efficiency for a channel code

$\tilde{\eta}$ – throughput efficiency for a data block

P_b – decoded information bit-error probability

G – number of channel groups

D – reuse distance between cell centres

APPENDIX B - PROOF THAT ϕ=Q

The average number of channels occupied in an m-server system is

$$Z = \sum_{k=0}^{m} k\, p_k + m \sum_{k=m+1}^{\infty} p_k \qquad \text{(B-1)}$$

For Erlang B systems, the second term is zero. Thus occupancy

$$\phi_B = Z/m$$

$$= (b_0/m)\rho \sum_{k=0}^{m-1} \rho^k/k!$$

$$= (\rho/m)b_0[\sum_{k=0}^{m} \rho^k/k! - \rho^m/m!]$$

$$= (\rho/m)[1-P_B]$$

$$= Q$$

For Erlang C systems, the second term in (B-1) is $m\, P_{m+1}$. Thus

$$\phi_c = (c_0/m)\rho \sum_{k=0}^{m-1} \rho^k/k! + P_{m+1}$$

$$P_m = [c_0/(1-\frac{\rho}{m})]\rho^m/m!$$

Thus

$$\phi_c = (\rho/m)[1-P_m] + P_{m+1}$$

Since

$$P_{m+1} = P_m - p_m$$

$$\phi_c = (\rho/m) + (1-\frac{\rho}{m})\, P_m - p_m$$

Since

$$p_m = c_0\, \rho^m/m!$$

it follows that

$$\phi_c = \rho/m$$

$$= Q$$