54

# CODE-DIVISION MULTIPLEXING

by
Ceilidh Hoffmann

submitted to
the Department of Electrical Engineering & Computer Science
as partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004



Author's Signature: _____

Department of Electrical Engineering and Computer Science

Certified by: _____

Kai-Yeung Siu
Thesis Supervisor

Accepted by: _____

Arthur C. Smith
Chair, Department Committee on Graduate Students

# CODE-DIVISION MULTIPLEXING

by
Ceilidh Hoffmann

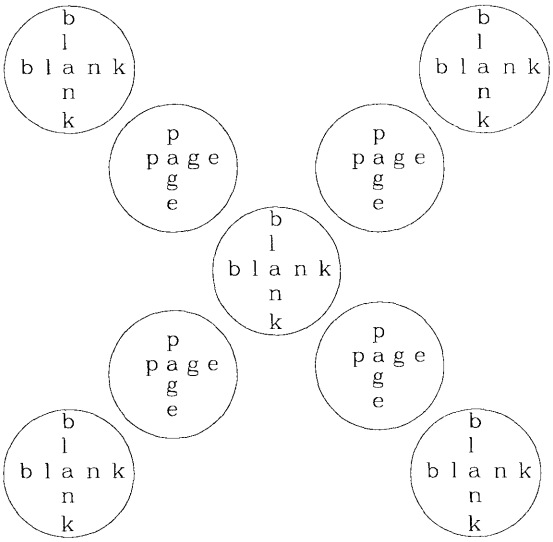
Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2004,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

## ABSTRACT

We study forward link performance of a multi-user cellular wireless network. In our proposed cellular broadcast model, the receiver population is partitioned into smaller mutually exclusive subsets called cells. In each cell an autonomous transmitter with average transmit power constraint communicates to all receivers in its cell by broadcasting. The broadcast signal is a multiplex of independent information from many remotely located sources. Each receiver extracts its desired information from the composite signal, which consists of a distorted version of the desired signal, interference from neighboring cells and additive white Gaussian noise. Waveform distortion is caused by time and frequency selective linear time-variant channel that exists between every transmitter-receiver pair.

Under such system and design constraints, and a fixed bandwidth for the entire network, we show that the most efficient resource allocation policy for each transmitter based on information theoretic measures such as channel capacity, simultaneously achievable rate regions and sum-rate is superposition coding with successive interference cancellation. The optimal policy dominates over its sub-optimal alternatives at the boundaries of the capacity region. By taking into account practical constraints such as finite constellation sets, frequency translation via carrier modulation, pulse shaping and real-time signal processing and decoding of finite-length waveforms and fairness in rate distribution, we argue that sub-optimal orthogonal policies are preferred. For intra-cell multiplexing, all orthogonal schemes based on frequency, time and code division are equivalent. For inter-cell multiplexing, non-orthogonal code-division has a larger capacity than its orthogonal counterpart. Among intra-cell orthogonal schemes, we show that the most efficient broadcast signal is a linear superposition of many binary orthogonal waveforms. The information set is also binary. Each orthogonal waveform is generated by modulating a periodic stream of finite-length chip pulses with a receiver-specific signature code that is derived from a special class of binary antipodal, superimposed recursive orthogonal code sequences. With the imposition of practical pulse shapes for carrier modulation, we show that multi-carrier format using cosine functions has higher bandwidth efficiency than the single-carrier format, even in an ideal Gaussian channel model. Each pulse is shaped via a prototype baseband filter such that when the demodulated signal is detected through a baseband matched filter, the resulting output samples satisfy the Generalized Nyquist criterion. Specifically, we propose finite-length, time overlapping orthogonal pulse shapes that are g-Nyquist. They are derived from extended and modulated lapped transforms by proving the equivalence between Perfect Reconstruction and Generalized Nyquist criteria. Using binary data modulation format, we measure and analyze the accuracy of various Gaussian approximation methods for spread-spectrum modulated (SSM) signalling. We show that both high rate techniques —parallel channel, single gain and single channel, reduced gain— are equivalent under the Gaussian model with or without multipath fading. For seamless multiplexing of SSM channels of various rates, we propose a flexible scheduling policy that removes code blocking and affords statistical multiplexing by dynamically reassigning signature codes. The algorithm is able to support both bursty and constant-bit rate connections without code tree partitioning.

Thesis Supervisor: Kai-Yeung Siu


blank
page
blank
page
blank
page
blank
page
blank
blank
blank

# ACKNOWLEDGMENTS

During my stay for the past several years at MIT, I was fortunate to have met and become acquainted with the following wonderful and amazing group of intellectuals. I take this opportunity to thank each and every one of them: My thesis committee members Profs. Dave H. Staelin and Franz Kaertner, and my thesis advisor Dr. Kai-Yeung (Sunny) Siu. Sunny was instrumental in getting me into the research project I was most interested in. My knowledge and interest in networking and higher layer communication protocols were instilled by him. He is my advisor, a good friend and a close confidant. Prof. Staelin has truly been a model professor in giving me advice and encouragement to keep marching ahead. Prof. Kaertner is the sweetest and kindest professor I met at MIT. My sincere gratitude goes to my technical examination committee members Profs. G. Dave Forney, Vahid V. Tarokh, Jeff H. Lang, Vincent V. Chan and Jeffrey H. Shapiro. I am very grateful to Prof. Forney for the time he spent helping me prepare my exam presentation. He clearly showed me the difference between good and great. I also wish to thank my graduate counsellor Prof. Dimitri Bertsekas for his guidance over the years, NTT DoCoMo of Japan for providing the grant that made it possible to conduct research in a wide variety of topics that subsequently lay the seeds for my thesis, and Prof. Gilbert Strang for being a model teacher and whose course in wavelets and filter banks sparked my interest in that subject. My attendance and knowledge gained in his course invariably changed the direction and focus of my thesis. I am indebted to my colleagues Drs. Mingxi Fan and Anthony Kam for their collaborative work in Chapters 6 and 7. I thoroughly appreciate and respect our diverse backgrounds and the synergy we manage to create in 1-107. Above all, I would not have made it through all the ups and downs without the support and motherly concern of Ms. Marilyn Pierce of EECS, especially at times when I was on the road performing my Hudinic acts.

On a personal note, the following people have profoundly impacted, and ultimately guided me —one way or the other— in my making key decisions that would forever alter my life: All my family members including Grandma, Hans, and Bob, Drs. Howard Heller, Peter Kassel, Stephan Kurtz and Patient Advocate Anne Boppe, and lastly Mr. Ravi Agarwal. I don't need to spell out how much you mean to me.

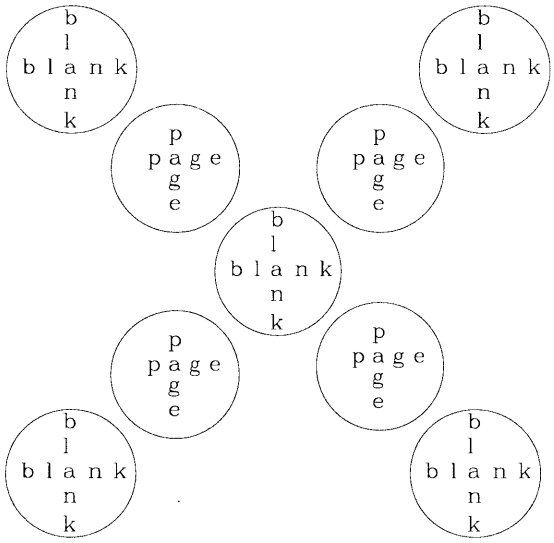
Finally, my faintest whispering thank you to Duo for all the love, sadness and memories. I once read that it takes more than a good memory to have good memories. I will cherish each and every moment I had with you. You are my everything, my Immortal....

C. H.
cth@alum.mit.edu
Boston, MA

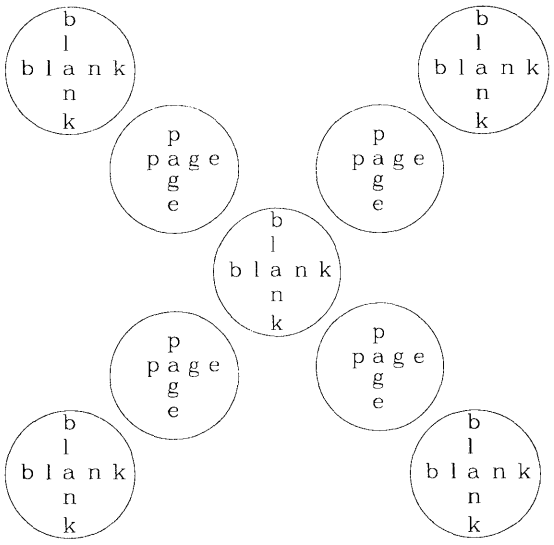
This document is typeset in Charter font using WinEdt (v.5) text editing software. The final LaTeX file is compiled with MikTeX (v.2.4) text formatting package in PC Windows environment. It is equivalent to the standard LaTeX $2_\varepsilon$ package. All drawings are sketched in AutoCAD LT and then converted to Encapsulated Postscript (.eps) files. All relevant computer files (.tex, .dwg and .eps) used in the generation of printer-friendly files —.pdf and .ps files— are contained in a CD-ROM that is submitted along with this paper copy.

To Nick,

*for making me see that*
*once in life,*
it *can happen twice.*

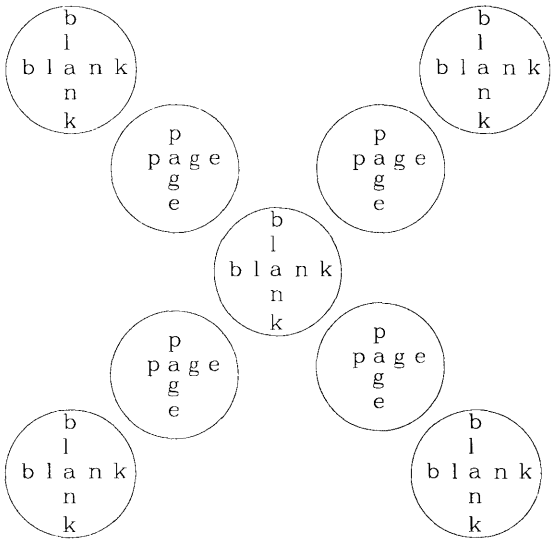

blank
page
blank
page
blank
page
page
blank
blank

# CONTENTS


blank
page
blank
page
blank
page
page
blank
blank

# FIGURES

# TABLES


blank
page
blank
page
blank
page
page
blank
blank

# 1

# INTRODUCTION

In wireless communication, the radio frequency (RF) bandwidth is a premium. The only effective means of maximizing bandwidth utilization is by reusing it over and over again at distant locations. Any conceivable wireless communication system —for example, broadcast radio and television channels, walkie-talkie, Citizen Band radio, Personal Mobile Radio, air traffic control, wireless local area networks etc.— adopts this strategy. In the literature, the precise technical jargon is known as *frequency re-use*; it is measured in terms of a single parameter called *the re-use number* $N_T$. Reuse of bandwidth is also the main culprit of interference in a cellular wireless network. Interference due to frequency re-use is synonymous to bacteria: Without it, the entire cellular network is very inefficient in bandwidth utilization; With it, a carefully designed control mechanism is required to maintain the network in equilibrium. Two mutually exclusive techniques that mitigate signal degradation due to interference are propagation loss and coding. The former is mostly a by-product of geometry while the latter is an application of algebra. Since interference originates from distant transmitters, it can be reduced by increasing the distance of separation —but at a cost of lower bandwidth efficiency per transmitter. (This remark is explained in detail in Chapter 3.) Determination of an optimal re-use distance is a difficult task for cellular radio system designers. It is a complex problem involving many interdependent factors such as resource allocation policy, *information-dependent* coding, data modulation format, characteristics of the propagation medium and the design and complexity of mobile transceiver. The final choices of these factors made by European GSM[1] and American IS-95[2] system designers are compared in Table 1.1.

[1]GSM —Global System for Mobile Communications— is a pan-European cellular mobile radio standard that has been adopted in all regions around the world with a global market share of over 80% as of year 2003.

[2]U.S. IS (Interim Standard) 95 is a cellular mobile radio standard pioneered by Qualcomm Inc. that uses direct-sequence spread-spectrum technology and is incompatible —and in direct competition with— TDMA based GSM.

**Table** 1.1: System Parameters of GSM and IS-95 Cellular Networks

| Parameter | GSM | IS-95 |
|---|---|---|
| Channel Multiplexing<br>Multi-User Accessing | FDM/TDM (downlink)<br>FDMA/TDMA (uplink)<br>with time advance | superimposed O-CDM (downlink)<br>PN spread N-CDMA (uplink) |
| Duplex mode | frequency division | frequency division |
| Re-use number $N_r$ | 4 or 7 | 1 |
| Carrier Modulation s(t) | single-carrier | single-carrier |
| Data Modulation g(t) | MSK (downlink)<br>MSK (uplink) | QPSK (downlink)<br>O-QPSK (uplink) |
| Pulse-shaping f(t) | Gaussian (BT = 0.3) | Root raised cosine ($\tau = 0.22$) |
| Coding<br>(Information-independent) | None | Pseudo-Noise Coding<br>SF = 2 (dl), SF = 42.67 (ul) |
| Convolutional Coding<br>(Information-dependent) | $r_c = 1/2$, K = 5 (both links) | $r_c = 1/2$, K = 9 (downlink)<br>$r_c = 1/3$, K = 9 (uplink) |
| Modulation spectral efficiency<br>(bits/symbol) | 0.8 | 1.6 |
| Bandwidth efficiency $N_{eff}$<br>(Effective reuse number) | 4 | 1.5 |

Please refer to page 381 for acronyms.

## 1.1 Historical Perspective and Motivation

An alternative to propagation loss for interference suppression is *information-independent* coding —commonly known as *spreading*. Its sole purpose is to reduce interference from distant transmitters, *not* background thermal noise. In contrast, *information-dependent coding* —which is more commonly known as single-user error control or channel coding— is optimized for detection of a signal perturbed with samples of an additive white Gaussian noise process —and to a lesser extent, the suppression of interference. Both error control coding and spreading techniques achieve their objectives by expanding the signal transmission bandwidth —relative to the uncoded case— by $1/r_c$ and $k_c$ respectively. (The parameter $r_c$ is the *code rate* of an error-control code and $k_c$ is the *excess spreading factor* of a signature code sequence.) Claude Shannon formulated the Channel Coding Theorem by proving that in a single-user AWGN channel, reliable communication —with arbitrarily low error probability— is achievable only through coding. This result is also applicable to a multi-user communication system if the code is optimized for both AWGN and interference. The construction of such codes and their associated decoding methods remain as the Holy Grail in communication system design. In practice system

designers rely on a cascade of two independent operations: single-user error-control coding to combat Gaussian noise followed by channelization coding for interference suppression. Time-, frequency- and code-division multiplexing schemes are examples of channelization coding in time, frequency and code domains, respectively. When the reuse number $N_r$, the coding rate $r_c$ and the excess spreading factor $k_c$ are taken into account, the transmission bandwidth relative to the uncoded case is increased by

$$N_{eff} = \frac{N_r k_c}{r_c} \tag{1.1}$$

where $N_{eff}$ is defined as the *effective reuse number*. If $W_{tot}$ is the total allocated RF bandwidth for a cellular network, the usable bandwidth per cell (transmitter) is

$$W_c = W_{tot}/N_{eff} \tag{1.2}$$

The optimal set of numbers that should be assigned to these three variables in eqn. (1.1) for cellular mobile radio is often a contentious matter debated fiercely in the academic world as well as in the industry. Numerous published articles and texts have compared capacities of GSM and IS-95 networks. Other researchers have analyzed the capacities of generic TDMA- and CDMA-based networks. These comparative results may be biased, unfair or incomplete for the following reasons:

- *GSM vs. IS-95*: The comparison of GSM versus IS-95 is not always equivalent to TDMA versus CDMA because key network design parameters such as data modulation format, channel coding rate and constraint length, voice compression algorithm and receiver architecture skew the overall network performance. The critical role played by the underlying channel multiplexing or multi-accessing method (TDMA or CDMA) is no longer apparent. For example, IS-95 affords a much more powerful convolutional code with a longer constraint length than GSM due to its continuous transmission capability. In contrast, GSM bursts in TDM/TDMA mode do not allow excessive delay for voice packets. For non-voice applications, it is conceivable that a TDMA system can also use a convolutional code of longer constraint length. In terms of data modulation, GSM designers selected minimum-phase shift keying due to its constant envelope property and equivalence to FM/FSK modulation. The former condition allows the use of non-linear RF amplifiers. The latter condition allows the use of an FM demodulator, either in non-coherent, differential or coherent detection mode. In IS-95, both links adopt spectrally more efficient I/Q modulation formats. Since the power efficiency of RF amplifiers is less critical at the base site, IS-95 designers opted for QPSK in the downlink. For the uplink, its variant offset QPSK —which has a smaller peak-to-average amplitude ratio than QPSK— was chosen. In terms of pulse shapes, GSM adopts a non-Nyquist Gaussian pulse for its smooth spectral roll-off. Its goal is to improve bandwidth efficiency with controlled ISI. In IS-95, root raised-cosine Nyquist pulses are used.

- *TDMA vs. CDMA*: Analytical results obtained from comparative study of generic TDMA and CDMA systems may not be directly applicable to practical networks since most theoretical models assume unconstrained signal design, very long codewords and complex decoding algorithms that may not be realizable in practice. For example, in the one-to-many[3] broadcast channel, both TDM (as adopted in GSM downlink[4]) and orthogonal CDM (as adopted in IS-95 downlink) have the same achievable rate regions. It is true, however, that the optimal channel multiplexing protocol —with a larger capacity region than TDM— is a non-orthogonal CDM with successive interference cancellation. However, this is not the chosen protocol for IS-95. In the many-to-one multiple access channel, optimal CDMA with maximum-likelihood joint detection has a larger rate region than TDMA. On the other hand, naive CDMA (as adopted in IS-95 uplink) has a smaller rate region than TDMA.

- *Downlink vs. Uplink*: Cellular networks —like most other communication systems— operate in duplex mode; i.e., information is exchanged from a stationary base transceiver to mobile terminals and vice versa. The preferred duplexing mode for cellular networks is frequency-division[5] where two disjoint RF bands are assigned for base-to-mobile (downlink) and mobile-to-base (uplink) transmissions, respectively. Thus, any comparative analysis of TDMA versus CDMA is incomplete unless the study is conducted for both downlink and uplink. The GSM standard has the same TDMA architecture, modulation and coding formats in both links. In contrast, the IS-95 standard applies different channel multiplexing, modulation and coding formats in each link.

- *Single (fixed) vs. Multiple (variable) Rates*: Most comparative studies focus on TDMA and CDMA network architectures that support a single class of constant-bit-rate channels only. This model is well suited for voice and low-rate data applications. It obviously fails to address the feasibility and complexity issues involved in the support of variable data rates in a single channel over the duration of a connection or among different channels. Most often the added complexity in multi-rate communication is at the receiver. For a base site, transmission at a higher information rate is simply a matter of rate-splitting among several parallel low-rate channels. In TDM/TDMA, multi-rate reception involves signal detection over a longer time window. In contrast, the complexity of a CDM receiver may not scale easily for higher information rates.

Our work herein is another attempt —one among many from various researchers and practitoners— that measures and quantifies the contribution of key parameters of a cellular

[3]The one-to-many "broadcast" channel is not the same as "downlink." See Chapter 3 for details.

[4]downlink = forward link

[5]The other option is time-division duplexing where a single RF band is used. The time frame is divided into two sub-frames, with each sub-frame dedicated to base-to-mobile or mobile-to-base transmission.

network architecture design. We limit our scope to communication in the downlink only. We do not in any capacity profess that our models and analytical results are unbiased, fair and complete. Nevertheless, we have taken a broad approach in which important design parameters —resource allocation policy and its associated channel multiplexing scheme, carrier and data modulation/demodulation techniques, pulse shaping for spectral containment, peak-to-average transmission ratio and resulting latency bound, receiver complexity and scalability— and their intertwined relationships are all taken into account. Before we begin our quest for the Holy Grail, we review some preliminaries:

## 1.2 Definitions and Network Models

A communication network consists of at least two entities: a *source* and a *sink*. Information is exchanged *from* a source *to* a sink. The term "user" is synonymous to "sink." An M-user communication network consists of M source-sink pairs. A source feeds information to its designated *transmitter*. Communication is established between a transmitter and a *receiver* through a medium called the *channel*. The receiver then delivers the message to its associated sink[6]. The transmitters and receivers play the role of middlemen who help facilitate the transaction —in this case, reliable and efficient communication from a source to its sink.

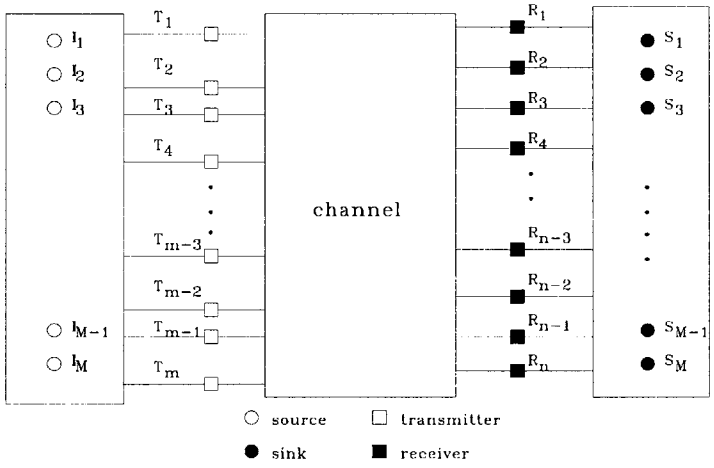



**Figure** 1.1: A generic communication system with M source-sink pairs

In general, a communication channel is a two-way street. An example of a two-way channel is the twisted-pair duplex telephone line in which information is sent and received by both entities simultaneously. In the work presented herein, information exchange is simplex —i.e.,

[6]It is assumed that information exchange between a source and its transmitter —and similarly between a receiver and its sink— is instantaneous (zero delay) and error-free.

in one direction only. Therefore every entity is either a source or a sink, but not both.

With stated stipulations, several communication channel models can be envisaged. At a minimum when there are only two entities —a source and a sink— we observe the well-studied single-user channel model. When the source-sink population is much larger, three different types of multi-user channel models —broadcast, multiple-access and interference— can be defined. Their distinction is quantified by a single determinant known as *cooperation*. "Autonomy" is antonymous to "cooperation." Cooperation is equivalent to pooling of resources; a resource is either a transmitter or a receiver. To facilitate our classification of multi-user channel models, we depict a block form of a generic network with M source-sink pairs in Fig. 1.1. The channel is effectively an input-output (I/O) device with $T = m$ transmitters as inputs and $R = n$ receivers as outputs. Cooperation may exist at the input side among sources or at the output side among sinks. In one extreme all sources and sinks are in full cooperation; this is the degenerate single-user model. In the other extreme with no cooperation at both ends, we obtain the *interference channel* in which $T = R = M$. The two remaining channel models are constructed when there is full cooperation at one (input or output) side only. When there is full output cooperation ($R = 1$) and autonomous transmitters ($T = M$), a *multiple-access* channel is observed. Here, the emphasis is on the receiver since it must —on behalf of all sinks— gather or *access* information from *multiple* sources. With full input cooperation ($T = 1$) and a separate receiver for each sink ($R = M$), the *broadcast channel* results. All sources pool a single common transmitter. In return the transmitter *broadcasts* a compound signal to all users (sinks). All three multi-user channel models are illustrated in Fig. 1.2. A dashed-line block is enclosed among a group of sources or sinks that are cooperating. The compound signal —say, $s(t)$— is a linear superposition of M time-synchronized, independent signals,

$$s(t) = \sum_{i=1}^{M} s_i(t)$$

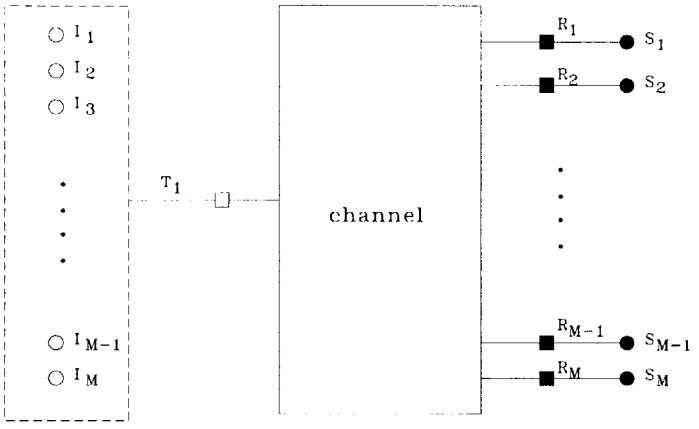
where each elemental signal $s_i(t)$ is representative of message $m_i$ from source $I_i$. *Channel multiplexing* is a transmitter's task of converting independent messages to a broadcast signal:

$$\{m_1, m_2, \ldots, m_M\} \longrightarrow \{s_1(t), s_2(t), \ldots, s_M(t)\}$$

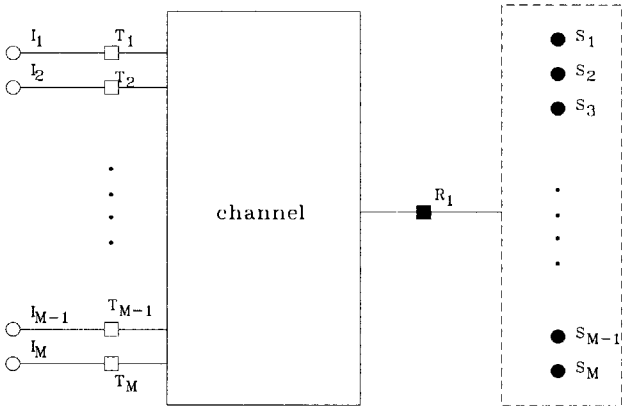
such that the same broadcast signal is most suitable for transmission to various remotely located receivers through their respective channels. If an intermediate step where messages are modulated by *signature code sequences* is included,

$$\{m_1, m_2, \ldots, m_M\} \longrightarrow \{b_1[j], b_2[j], \ldots, b_M[j]\} \longrightarrow s(t)$$

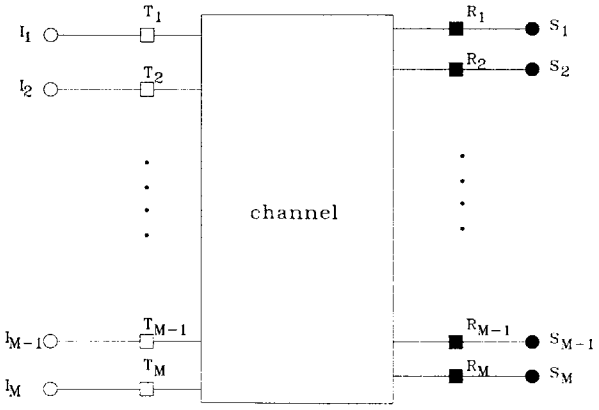
the broadcast signal is said to be *code-division multiplexed*.




(a) broadcast model with full input cooperation




(b) multiple-access model with full output cooperation

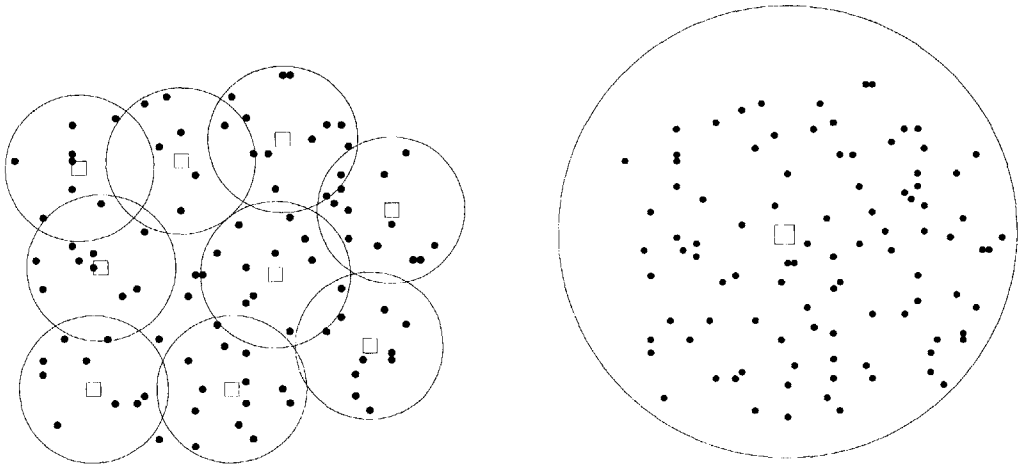



(c) interference model with no cooperation

**Figure** 1.2: Multi-user communication channel models

## 1.3 Problem Statement

When cooperation is *partial*, two variations of the interference channel can be deduced. They are labelled collectively as *composite channels*. We are only interested in the *broadcast-interference composite model*[7] with partial input cooperation ($T < M$) and autonomous receivers. It mimics the base-to-mobile (downlink) cellular radio propagation channel. Each *mobile user* or *radio* receives broadcast signal from its target antenna site as well as interference signals from nearby transmitters. We coin such model the *cellular broadcast channel*.



(a) (b)

**Figure** 1.3: An aerial view of a densely populated geographical area

Fig.1.3(a) depicts an aerial view of several fixed transmitters (□) and many scattered sinks (•) in some densely populated geographical area. If we further assume that an omni-directional antenna is mounted at each transmitter site and neglect physical objects in the propagation medium, the signal strength is the same along the perimeter of a circle of arbitrary radius — similar to circular (barometric) contours. This "imaginary" coverage area is commonly known as a "cell." As an alternative, it is possible to design a communication network consisting of a single transmitter only with a huge circular footprint (see Fig. 1.3(b)) to accommodate every scattered receiver. This in fact is the paradigm for terrestrial and satellite broadcast services such as radio and television. In both applications, the number of sources (radio or television channels) is much smaller than the number of receivers. This violates our stipulation that a source exists for each sink. If each sink (a listener or a viewer) demands a different TV or radio

[7]The other composite channel is multiple-access with interference. A good example of such channel is the cellular uplink. Each base receiver must jointly decode information from its scattered target mobiles in the presence of interfering signals from out-of-cell mobiles.

broadcast channel[8], the channel "pipeline" would be very big; i.e., a prohibitively large amount of bandwidth is required. To circumvent the hunger for bandwidth, the region is partitioned into smaller coverage areas. The allocated bandwidth is shared and reused among transmitters. In a nutshell, this is the cellular re-use concept.

*Problem Statement:*
*Consider a large mobile radio communication network depicted in Fig 1.4. Assume the total coverage area is unbounded, and it is partitioned along imaginary lines by concatenation of hexagonal cells of equal size. If the available (radio frequency) bandwidth is $W_{tot}$ Hz and the omni-directionally radiated power from each transmitter is limited to* P *watts, what is the optimal resource allocation policy for each transmitter?*

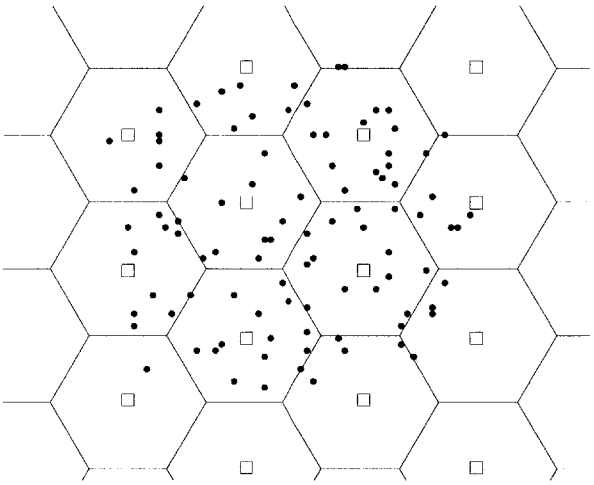


**Figure** 1.4: An aerial view of a cellular communication network

### 1.3.1 Optimality Criteria

This begs the question: "How is optimality defined, and what other parameters should be taken into account in selecting the 'best' resource allocation policy for each transmitter?" If optimality is measured in terms of the sum-rate (i.e., the sum of information rates to all users within a cell), then the best strategy is to communicate only with the receiver that maintains

[8]The term "channel" is used in different contexts throughout the monograph. Here it refers to a TV or radio station that is tuned in. It also applies to the radio propagation medium through which a signal is transmitted, or a connection —a logical channel such as a frequency band, a time-slot or a code sequence— established between each transmitter-receiver pair.

the best channel response. This policy is evidently unfair to other receivers with poorer respective signal strengths. At the other extreme a "socialist" policy supports the same rate to every receiver regardless of their respective demands. Current cellular systems adopt such a strategy where each circuit-switched channel carries a fixed information rate. This policy is too stringent to support various multimedia applications that require variable information rates. Rate adaptation applies not only to the partitioning of the total sum-rate among receivers, but also to dynamic variation of information rate to each receiver throughout the lifespan of a connection.

### Achievable Rate Distribution

We may investigate this problem from a different angle. Suppose there is a set of requested rates by receivers:

$$\mathbf{R}^* = \left(R_1^*, R_2^*, \ldots, R_m^*\right)$$

We can select an "optimal" resource allocation policy

$$\mathbf{R}^o = \left(R_1^o, R_2^o, \ldots, R_m^o\right)$$

that best matches the requested rates with

- the minimum mean squared error:

$$\min \sum_{i=1}^{m} \left(R_i^* - R_i^o\right)^2$$

- the minimum error in rate-sum:

$$\min \sum_{i=1}^{m} \left(R_i^* - R_i^o\right)$$

- the largest percentage of user population whose requested rates are met:

$$\max \sum_{i=1}^{m} \mathbf{1}_i$$

  where the indicator function $\mathbf{1}_i = 1$ when $\left(R_i^* \geq R_i^o\right)$ is true, and zero otherwise.

- the largest achievable rate per connection:

$$\max\left\{R_1^o, R_2^o, \ldots, R_m^o\right\}$$

- the maximization of the minimum rate in all connections:

$$\max\left\{\min\left\{R_1^o, R_2^o, \ldots, R_m^o\right\}\right\}$$

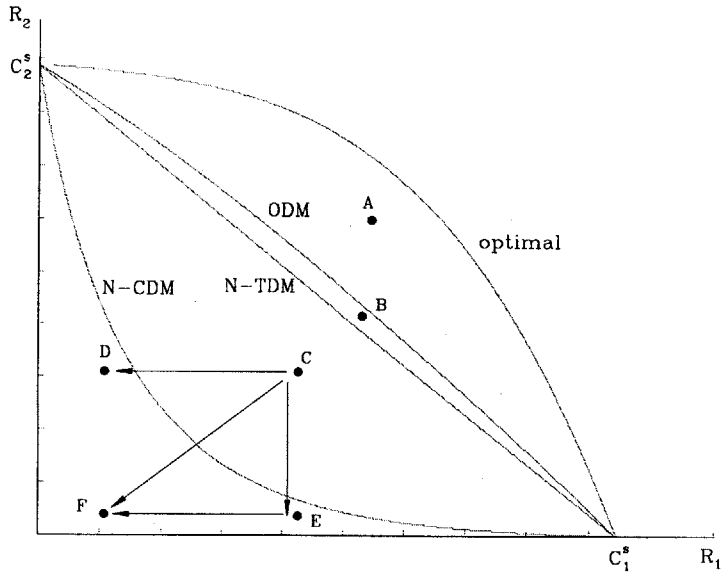



**Figure** 1.5: Achievable rate pair $(R_1, R_2)$ for a two-user Gaussian broadcast channel. Any rate pair enclosed by the outermost bound (including point A) is achieved by the optimal scheme. Any rate pair, including point B, is achievable by any of the orthogonal schemes.

Depending on the requested rate vector $\mathbf{R}^*$, there may be an outright winner or several winners with ties. Consider the capacity-region plot shown in Fig. 1.5. This is the broadcast capacity of a two-receiver system in an ideal bandlimited channel perturbed by additive white Gaussian noise processes with power spectral densities $\sigma_1^2$ and $\sigma_2^2$, respectively. The optimal policy is superposition coding with successive multi-stage decoding. (Details are given in Chapter 3.) The x- and y- intercepts are single-user capacities $C_1^s$ and $C_2^s$ of users 1 and 2, respectively. Any rate pair $(x_i, y_i)$ within the region bounded by the "optimal" curve and the two perpendicular axes can be supported by the transmitter. The figure also gives achievable rate regions (ARR) for several sub-optimal multiplexing schemes. If the target rate pair is point A, then the optimal scheme must be used to meet the demand of all receivers. If the target rate pair is point B, then we have a choice between the optimal and various sub-optimal orthogonal-division multiplexing (ODM) schemes. A similar design option is available for point C where all but naive code-division multiplexing (N-CDM) are good candidates. Moreover, the gap in ARR of the optimal policy over ODM —or likewise the larger rate region of ODM over naive time-division multiplexing (N-TDM)— is proportional to the difference in noise levels $\sigma_1^2 - \sigma_2^2$. In fact when $\sigma_1^2 = \sigma_2^2$ the optimal and orthogonal-division multiplexing schemes collapse onto the straight-line rate region of N-TDM. It becomes apparent that measuring the optimality of a resource policy based only a set of simultaneously achievable set of rates is insufficient. We must consider other factors such as:

### Rate Adaptation

During the life of a connection, from the initial handshake to subsequent channel release, the information rate of a user may vary from zero to some upper limit. Referring to Fig. 1.5, at any time instant the supported rate pair may slide from point C to D (or E or F). At another instant it may slide again from D to F. The adaptation of rates between the transmitter and its users must be coordinated dynamically and seamlessly. The break-before-make process —first releasing an existing connection and subsequent acquisition of a new connection— for variable rate support is not efficient utilization of network resources. If carried out, the total network capacity is reduced due to increased control signal overhead, and above all, the exchange of information is not seamless. With dynamic rate adaptation, a user's information rate is adjusted block-by-block. In time-division multiplexing, a block is a time slot —which is a fraction of time frame allocated to a user. Likewise, in frequency-division multiplexing (FDM) a block is a fraction of the total frequency band. A control flag is appended in each block sent from the transmitter to each user informing the size of the block. Hence we must consider the ease and flexibility of arranging such a procedure in every channel multiplexing scheme.

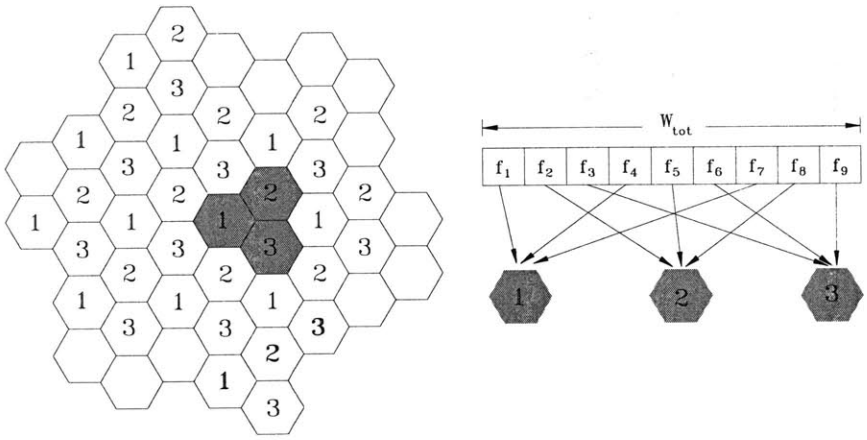



**Figure** 1.6: Assignment of frequency blocks in FDM

### Receiver Complexity

For successful coordination of dynamic rate adaptation, a user's receiver must have the capability of accepting —i.e., demodulating, detecting and decoding— a block of variable size. The pertinent issue here is scalability of receiver design for increased information rate. In TDM this is a trivial matter since a larger block implies a longer time window of reception. In a FDM-based cellular network as shown in Fig. 1.6, assigned frequency bands are interleaved such that spectral overlap (i.e., adjacent channel interference) is reduced among frequency bands

allocated to a cell. This strategy precludes the use of a wideband contiguous frequency band for higher transmission rates. Therefore, an FDM receiver may require a tunable filter with adjustable (rubber) bandwidth or a set of fixed-bandwidth filters. It is obvious that the latter option does not scale efficiently.

In the optimal multiplexing policy, the receiver complexity issue is somewhat unique: it is unrelated to the information rate. Rather it is directly proportional to the number of users and their relative signal-to-noise ratio (SNR). The user with the lowest noise level has the highest complexity since it must first decode signals of other users with higher noise levels. The real-time implementation of such a procedure may not be feasible in a fading environment. We elaborate on these remarks in Chapter 3. As presented therein, the receiver complexity issue takes a different flavor when combined with the frequency re-use concept of cellular networks.

### Dynamic Resource Allocation

In rate adaptation the aim is to provide rate guarantees on a per-user basis. It is mainly concerned with the physical layer (e.g., receiver hardware complexity) and the medium-access layer (e.g., control channel flag) issues. It is a local optimization policy; its interest lies in the support of rate $R_{max}/n$ for an arbitrary $n \geq 1$. It does not address —or is even concerned with— the issue of maximizing the overall network information capacity, which is of course a network layer global optimization problem. This subject of dynamic resource allocation is broken down into network operations such as call admission control, data channel scheduling and queuing and prioritization of data packets. The goal is to support a certain pre-defined level of quality-of-service (QoS) —loosely measured in terms of maximum information rate, the ratio of maximum to average information rate, call admission and dropping rates— to every user while maximizing the time-averaged capacity of the entire network. It is well known that TDM with its bursty transmission property generally has a high ratio of maximum-to-average information rates. The opposite is true for FDM. In orthogonal code-division multiplexing (O-CDM), as demonstrated in Chapter 7, the total network capacity also depends on the type of signature code sequences.

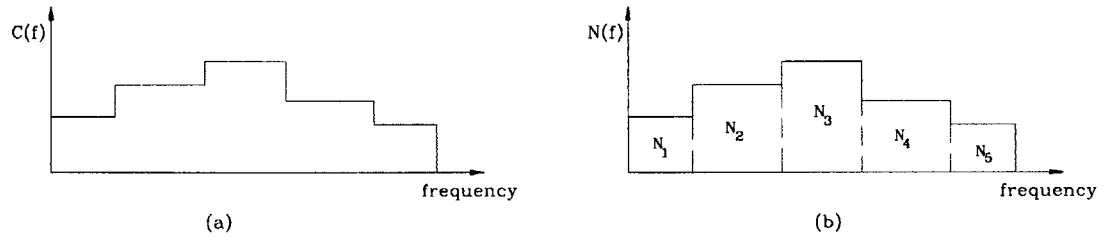



**Figure 1.7:** Additive Gaussian channel models
(a) frequency selective channel (b) non-uniform Gaussian PSD

### Carrier Modulation and Pulse Shaping

The channel capacity of a power-limited, discrete-time AWGN channel is a fixed value. The type of carrier modulation format (single or multiple) for information transport is irrelevant in the discrete model. When the channel model is modified for continuous time, it is well known that both single- and multi-carrier modulated signals achieve the same capacity as long as optimal brick-wall pulse shapes are used. The higher spectral efficiency of multi-carrier modulation format have been touted in the Gaussian channel model when the channel transfer function is frequency selective (time dispersive), or the noise density is non-uniform across the transmission bandwidth. In that case the total spectrum is sliced into many parallel channels such that ideal flat frequency response is maintained in each sub-channel. This is illustrated in Fig. 1.7. This is the water-filling policy in the frequency-domain. The transmit power density is optimally allocated according to some criterion such that the spectral efficiency is maximized over the entire frequency band. Our argument goes one step further: Even in a frequency-flat AWGN channel, multi-carrier modulation achieves a larger capacity than its single-carrier counterpart when modulated signals are generated using practical pulse shapes. The capacity gain is mainly due to smaller excess bandwidth (beyond the minimum Nyquist bandwidth) and controlled overlap of sub-carriers in the multi-carrier format. The excess bandwidth is usually measured in terms of the fraction of mainlobe bandwidth that overshoots the Nyquist bandwidth. We show that the excess bandwidth is larger for a wideband single-carrier modulated signal than its multi-carrier counterpart. We portray this condition graphically in Fig. 1.8. Our argument is made precise in Chapter 5.

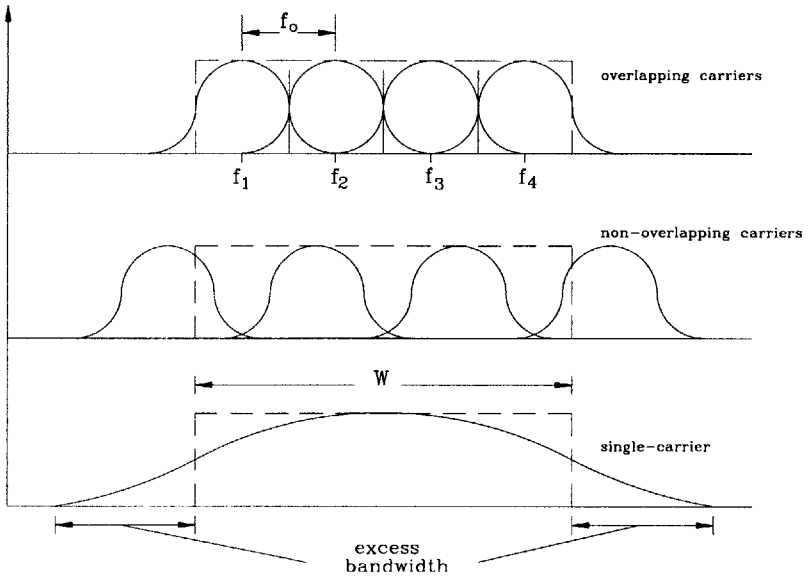



**Figure** 1.8: Excess bandwidth of transmit signal with pulse shaping

## 1.4 System Model

Due to symmetry, the design and analysis of the entire cellular network reduces to the design and analysis of communication techniques in any arbitrary hexagonal cell. By convention we denote the target cell as cell number 0 and the desired target receiver in cell 0 as user 0. The system consists of the transmitter at the center, many receivers scattered within the target cell plus interfering transmitters surrounding the target cell. The model is depicted in Fig. 1.9.

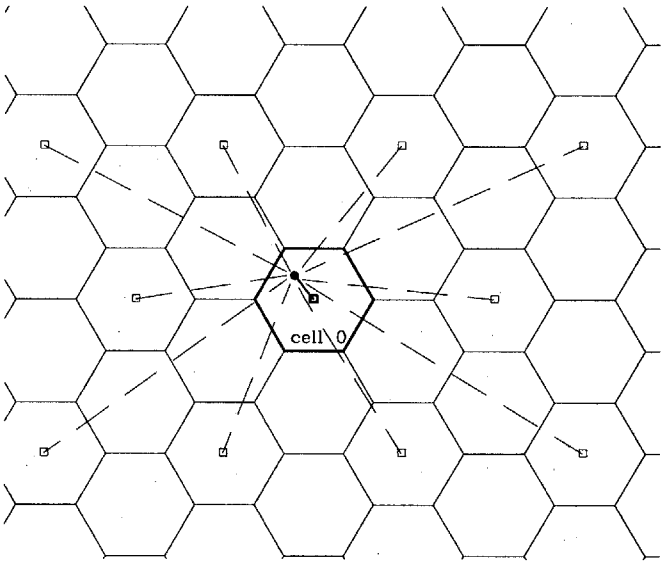



**Figure** 1.9: System model of a symmetric cellular broadcast network

A block diagram outlining key components of the cellular broadcast communication system is illustrated in Fig. 1.10. The messages from M independent information sources are pooled and converted into code symbols:

$$\{m_1, m_2, \ldots, m_M\} \longrightarrow \{X_1[j], X_2[j], \ldots, X_M[j]\}$$

Throughout we assume code symbols are binary-valued. The M parallel code sequences are then code-division multiplexed into a composite broadcast code sequence b[n]:

$$\{X_1[j], X_2[j], \ldots, X_M[j]\} \longrightarrow \{b_1[n], b_2[j], \ldots, b_M[j]\}$$

$$b[n] = \sum_{i=1}^{M} b_i[n]$$

where it is understood that each channel symbol $X_k[j]$ is channelization coded or modulated

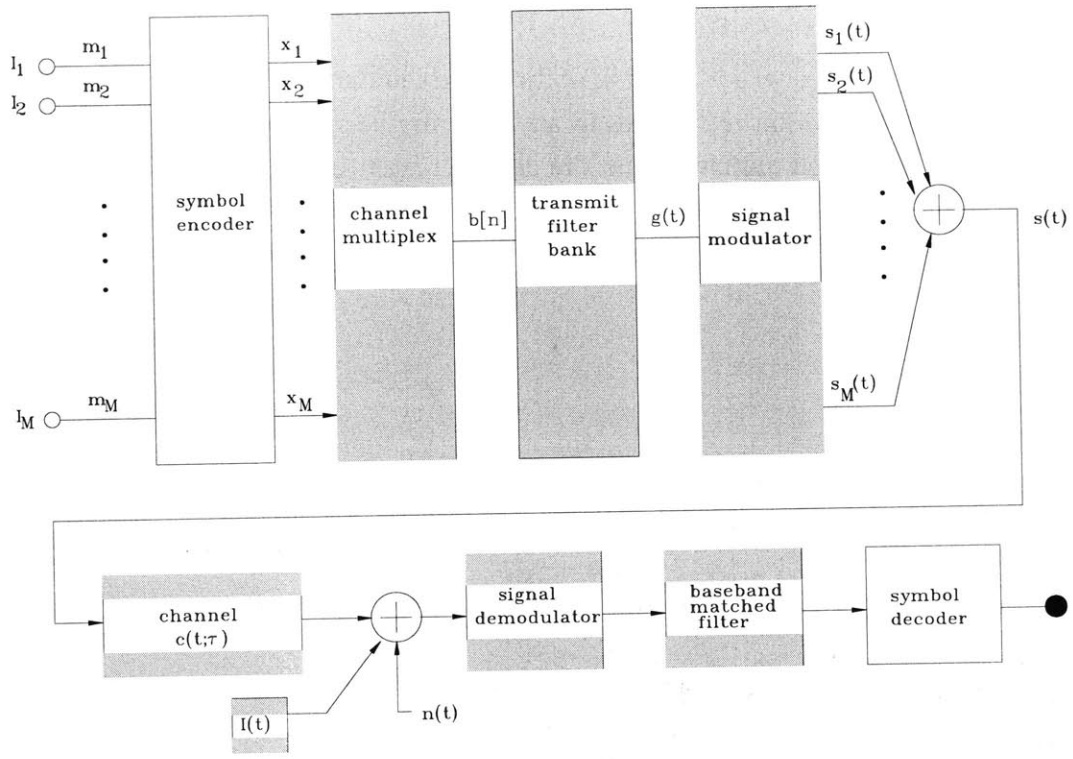



**Figure** 1.10: Block diagram outlining major components of a cellular communication system

with its respective modulating sequence $c_k[n]$:

$$\text{Level-Shift:} \quad X_k[j] \longrightarrow x_k[j]$$
$$\text{Channelization:} \quad b_k[n] = x_k[j] \cdot c_k[n]$$

Finally, the code sequence is converted to a continuous-time RF waveform through a two-step process of discrete-to-analog (D/A) conversion and carrier modulation:

$$\text{D/A Conversion:} \quad b[n] \longrightarrow g(t)$$
$$\text{Carrier Modulation:} \quad g(t) \longrightarrow s(t)$$

The transmit broadcast signal $s(t)$ is perturbed by linear channel distortions consisting of time variation and multipath propagation. The received signal $r(t)$ of target user 0 consists of channel-distorted desired signal, interfering signals from neighboring transmitters plus receiver front-end thermal noise that is modelled as white Gaussian noise process. As can be seen from the figure, we are mostly interested in synthesis and analysis of shaded functional blocks such as transmit and receive filter banks, carrier modulator and demodulator, and above all, the channel multiplexing operation and design of channelization codes. As performance measure,

we wish to evaluate the capacity and achievable rate, as well as uncoded bit-error rate and channel blocking probability of target user 0. As we shall see, all three benchmarks are highly dependent on aforementioned functional blocks and the time-variant impulse response $c(t;\tau)$ of the channel, and to a lesser extent, on the type and complexity of the target receiver.

## 1.5 Thesis Summary

By taking into account all the factors listed in previous sections, we show that a resource policy based on orthogonal code-division multiplexing is the preferred option. This conclusion is reached in several stages through the method of elimination. We show that:

- The optimal policy based on multi-resolution coding may not be feasible in practice. This is due to limited knowledge in the design of such codes and the enormous complexity imposed on all receivers. We next consider three orthogonal multiplexing schemes that divide bandwidth into time-slots, frequency bands or orthogonal codes. On a single-cell basis, they are all equivalent. In a multiple-cell environment, a policy based on code-division multiplexing is preferred over time- and frequency-division multiplexing. The preference is not necessarily due to a larger sum-rate for CDM but due to reduced complexity and flexibility at both transmitting and receiving ends.

- The support of variable rates to a target receiver in CDM is achieved by assigning a single binary code, parallel binary codes or a single multi-level code. We show that their performances, measured in terms of bit-error rate in zero-mean AWGN channel with or without multipath fading, are comparable. A receiver designed for a single binary code channel has the least complexity. Thus, CDM with single binary code channels is the preferred mode of multi-rate transmission.

- In the previous step, we show that higher-bit rate schemes based on single and parallel binary code channels are comparable in terms of bit-error rate. This compares their performance on a single-user basis. At a system-wide level, CDM with single binary code channels may have a lower throughput due to "code blocking" constraint that the system imposes when assigning recursive binary orthogonal codes. This condition does not arise in CDM with parallel binary orthogonal code channels. We then propose several methods and algorithms that remove code blocking. As an added bonus we show that CDM with single binary code channels is the preferred mode not only in terms of multi-rate support but also for statistical multiplexing of code channels.

- As seen from the system model depicted in Fig. 1.10, channel multiplexing is a discrete-time operation involving input data sequences and signature codes. We have yet to de-

clare the type of pulse shaping and carrier modulation that maximum the overall spectral efficiency of the cellular system. The transmit signal $s(t)$ can be one of the following two: a single-carrier-modulated wideband format or a series of contiguous multiple-carrier-modulated narrowband formats. We show that a multiple-carrier-modulated pulse has a narrower (i.e., more bandwidth efficient) power spectral density when practical Nyquist waveforms only are allowed for pulse shaping. In fact, we describe a special class of finite-length time-overlapped pulses that meets the generalized Nyquist criterion for zero intersymbol interference.

## 1.6 Organization

When analyzing a complex communication system with a large number of entities —with each entity belonging to one of several different classes, it is always a challenge to make the material accessible and easy to follow. Most often it is not because the content and its underlying theories are difficult to comprehend, but simply because the symbols and notations used in the analysis become unnecessarily cumbersome and confusing. We also run the risk of exhausting all possible Greek and Roman characters. We have therefore taken the liberty —and much headache— in preparing a list of symbols and notations used throughout this document. For consistency and homogeneity with references listed herein, we retain, whenever possible, the most commonly used symbols and notations as they appear in the literature. They are listed in page 381 under various sub-categories.

Every chapter begins with a *Summary* section. It quickly reviews related material from previous chapters, and then summarizes the main results of the chapter. The purpose is brevity, not absolute clarity or completeness. It is a synopsis of the overall content to a reader who is well-versed in the subject matter.

Chapters 2–7 end with a section titled *Notes and References*. In it we cite background material and references pertinent to material presented in each chapter. It also highlights certain methodologies, procedures and relevant topics that are not covered or those that we were not able to resolve. Appendices are annexed at the end of several chapters. They contain background material that serve as the vehicle in facilitating the description and understanding of results presented in the main body of the chapter.

Since the topics covered in various chapters are related and recycled, it is highly recommended that the entire document be read in the logical order presented. The main topics of discussion in the remaining chapters are as follows:

Ch. 2: The physical attributes of various fading phenomena and their associated statistical models are reviewed. An exhaustive list of statistical distribution functions that model the fading amplitude of the fading signal envelope and their first- and second-order moments

are derived and compared. The contents serve as background material for topics that follow in the remaining chapters.

Ch. 3: Key information-theoretical results pertaining to the channel capacity and simultaneously achievable rate regions of various multi-user channel models are reviewed and derived. In particular, the analytical results are extended to a composite channel that mimics the downlink of a cellular propagation model. It is then concluded that information-independent coding is preferred over propagation loss due to distance for inter-cell interference suppression. Thus CDM/TDM with spreading is a better candidate than FDM/TDM without spreading.

Ch. 4: The design and analysis of signature waveforms and their associated code sequences are detailed. Emphasis is placed on a class of recursive orthogonal binary antipodal code sequences. Its auto- and cross-correlation properties are examined. Alternative interpretation of these orthogonal codes in terms of a balanced tree structure is described. The recursive property enables flexibility in rate adaption and reduced hardware complexity at the receiver.

Ch. 5: The design of practical analog pulse shapes is treated. Related topics such as transmission bandwidth, power spectral density, bandwidth efficiency, Nyquist criterion for zero inter-symbol interference are discussed. The connection between the Nyquist criterion in the analog domain and the Perfect Reconstruction criterion in the discrete domain are highlighted. The significance of practical Nyquist time-limited pulse design methods are explored. It is concluded that multi-carrier modulation with time-overlapped pulses is more bandwidth efficient than single-carrier modulation with raised-cosine Nyquist pulses.

Ch. 6: The performance of various data modulation-demodulation schemes are listed. In particular, we focus on two competing techniques of single-code with variable spreading and multiple codes with single spreading. The performance measure is in terms of the average bit-error probability. The validity of the Gaussianity of the interference plus noise term is studied.

Ch. 7: The code blocking issue that arises in recursive orthogonal CDM scheme is explained. As a follow-up, two network scheduling schemes that reassign signature codes horizontally and vertically inside a code tree are presented. Another code reassignment scheme that exploits statistical multiplexing of bursty connections is also considered.

## 1.7 A Note on Nomenclature

In order to avoid any confusion in terms of interpreting concepts and terminology, we —in this section— list and define several important keywords that appear over and over in other chapters. We have tried to conform our definitions and interpretations as close as possible to those that already exist or well accepted in the literature. Unfortunately, universal acceptance or unambiguity of key parameters and concepts is not possible due to historical and legacy precedence as well as different interpretation by scholars and practitioners in many branches of information sciences and engineering. For the sake of clarity and consistency, we define the following:

### Multiplexing vs. Multiple-Access

According to our own interpretation on page 28, channel multiplexing is a task performed by a broadcaster (transmitter) on behalf of many disparate and independent information sources for the collective benefit of all target receivers. It is assumed that the transmitter also has some knowledge of the channel and receiver states as well as the authority to construct a single compound broadcast signal that is efficient and reliable for transmission. In contrast, multiple-accessing assumes joint detection and decoding of signals from several autonomous transmitters. It is a task reserved for a receiver that is linked to more than one sink. In the cellular broadcast model that we study, each receiver is connected to a single sink. Reception from multiple transmitters —for example, for the purpose of transmit signal diversity— is not allowed. When discussing commercial cellular and other wireless networks, it has become the norm to use terms such as TDMA, FDMA and CDMA even though all three are signalling protocols for the uplink. Whenever we discuss a particular network such as GSM or Qualcomm IS-95, we will explicitly state whether we are referring to uplink or downlink or both.

### Channel Multiplexing vs. Channel Modulation

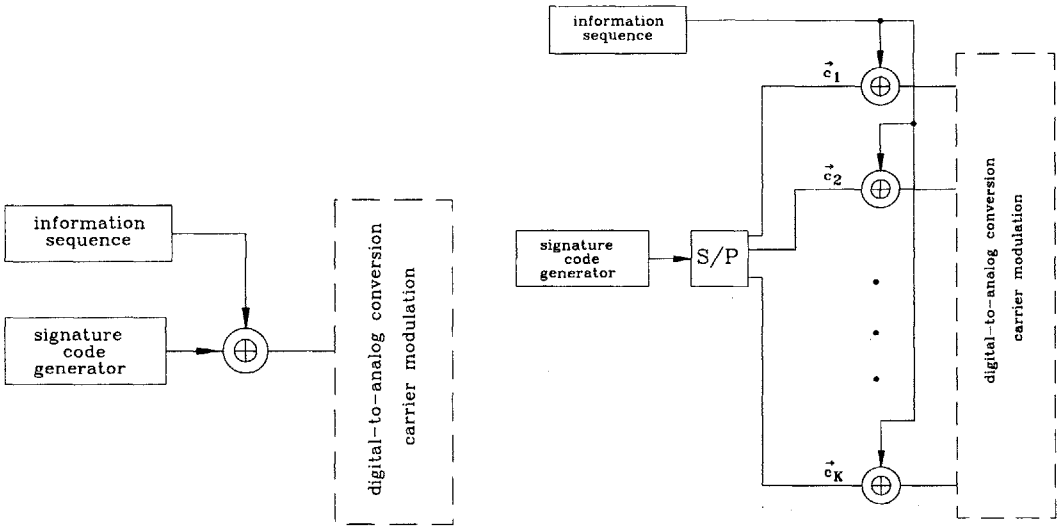
As depicted in Fig. 1.10, channel multiplexing is a discrete-time operation that is carried out before discrete-to-analog conversion and pulse shaping. Since we are studying digital communication only, all code symbols belong to a finite (PAM-type) set. Analog waveforms are required only for actual passband transmission via a radio propagation channel. Therefore, channel/carrier modulation refers to frequency translation from baseband to RF. There are two options in channel modulation: single versus multi-carrier. As far as we are concerned, the type of carrier modulation —single or multiple— is independent from the choice for channel multiplexing.

### Multi-Carrier Modulation vs. OFDM

Multi-carrier modulation is an analog IF or RF modulation technique in which the transmit signal is a linear sum of multiple (sub-)carrier modulated signals. Valid sub-carriers are cosine and sine continuous-time functions. No strict requirement is imposed on the relationship between any two sub-carriers in terms of spacing between center frequencies, time synchronization and carrier phase coherence. On the other hand, *Orthogonal* frequency-division multiplexing (OFDM) is a special kind of multi-carrier modulation technique implemented in the discrete-time domain. For historical reasons the name is closely related to frequency-division multiplexing of analog telephone channels. By applying (inverse) discrete-time Fourier transformation (DFT) of a block of discrete data sequence, a multi-carrier modulated signal —actually a discrete-time sequence— can be generated. (The sequence is D/A converted to analog waveform by interpolation and lowpass filtering.) The Fourier exponential sub-carriers are mutually orthogonal, resulting in "orthogonal" channel multiplexing. There are strict requirements for such orthogonality in terms of carrier spacing and time alignment since frequency responses of sub-carrier modulated signals overlap —unlike channel orthogonality in conventional FDM where it is assumed that contiguous channels have non-overlapping frequency responses. Note that orthogonality no longer holds if the signals in sub-carriers are not time synchronized. Furthermore, orthogonality is maintained at the receiver only if all subcarrier-modulated signals are detected altogether. If a particular sub-carrier modulated signal is selectively bandpass filtered, the filter output is distorted by overlapping signals from neighboring sub-carriers. In short, OFDM is treated as a carrier *modulation* format, and not as a channel multiplexing scheme. To avoid confusion, we refrain from using the term "OFDM" when describing multi-carrier modulation.

### FDM vs. OFDM

As stated above, OFDM is a multi-carrier modulation format that employs a fast Fourier transformer (FFT) in the discrete-time domain as a replacement for a bank of analog frequency oscillators. We know that there is no connection between OFDM and FDMA; however, there is a potential for confusion between OFDM and FDM. In frequency-division multiplexing (FDM), no time alignment is assumed among various carriers. Moreover, it is implicitly assumed that each receiver selectively bandpass filters only one out of many carrier-modulated waveforms for information retrieval. Thus, little frequency overlap is assumed among neighboring carrier-modulated signals. *In FDM, channels are disjoint. In OFDM, channels overlap.* Furthermore, FDM is a channel multiplexing scheme whereas OFDM —which is equivalent to multi-carrier modulation in discrete-time domain— is not.




(a)direct sequencing (b) time or frequency hopping

**Figure** 1.11: Direct and indirect sequencing of a discrete-time information sequence

### Various Code-Division Multiplexing Schemes

As stated earlier, any channel multiplexing scheme including code-division is viewed as an encoding technique in the discrete-time domain. If a binary information sequence of clock rate $R_b$ is *directly* encoded (i.e., binary modulo-2 added) with a signature sequence of clock rate $R_c$, the resulting scheme is called *direct-sequence* code-division multiplexing (DS-CDM). If, however, the information sequence is indirectly encoded by first serial-to-parallel converting the signature sequence, it can be either *frequency-hopped* code-division multiplexing (FH-CDM) or *time-hopped* code-division multiplexing (TH-CDM), depending on post-processing in continuous-time analog domain. Direct and indirect sequencing are illustrated in Fig. 1.11.

If the serial-to-parallel converter is followed by a single-carrier modulator, the resulting signal resembles a traffic channel in TH-CDM. Likewise, multi-carrier modulation after S/P conversion is equivalent to a FH-CDM channel. Hence, we may view time hopping as indirect sequence CDM in the time domain, and frequency-hopping as indirect sequence CDM in the frequency domain. It is apparent that if the signature sequence is a periodic train of impulses, time hopped CDM is equivalent to time-division multiplexing (TDM), and FH-CDM becomes FDM. Unless stated otherwise, CDM always refers to DS-CDM. Note that in strict information theoretic sense, spread-spectrum coding is not a necessary step in channelization coding of a CDM signal. A good example is superposition coding proposed by Bergmans [7], [8]. Otherwise, in almost all communication system models, it is assumed that spread-spectrum coding is an integral part of code-division multiplexing. Therefore, spread-spectrum multiplexing (SSM)

—a term commonly used in communications literature— is synonymous to CDM. When CDM is combined with single- or multiple-carrier analog modulation, it is referred to respectively as single-carrier CDM (SC-CDM) or multi-carrier CDM (MC-CDM).

### Code-Division Multiplexing and Spectrum Spreading

Spectrum spreading is commonly understood as a modulation process where the data modulated signal is multiplied (modulated) with a random-like signature analog waveform such that the resulting signal has a power spectral density much larger than the Nyquist bandwidth. The despreading operation —i.e., the opposite of spread spectrum modulation— is carried out by demodulating the received signal with the same signature waveform with proper time alignment. Thus spread spectrum modulation is a post-processing operation that takes place after data modulation, and despreading occurs before data demodulation. In CDM, the operation of channel multiplexing is carried out in discrete-time domain before digital-to-analog conversion. Furthermore, the compound signal before spreading may have a power spectral density extent equal to the Nyquist bandwidth. This is the case when the signature codes are orthogonal. In all of our analyses we reserve the term "spreading" for an operation that transforms a data modulated signal into into a wideband transmit signal with PSD width much larger than the Nyquist bandwidth.

### Coding vs. Modulation

Historically, spread spectrum modulation is viewed as signal processing in analog domain. It reflects the use of analog multipliers for SS modulation before the digital age. An analog spread-spectrum modulator is depicted in Fig. 1.12.

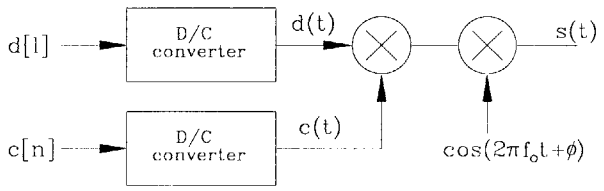



**Figure** 1.12: Conventional description of spread-spectrum modulation

It is still common —including the most recent research articles— to express a spread-spectrum modulated signal as:

$$s(t) = c(t)d(t)\cos(2\pi f_o t + \phi)$$

where $d(t)$ the data modulated baseband signal and $c(t)$ is the spread-spectrum modulating

signal. Next, the analog waveforms are expressed in terms of embedded discrete-sequences:

$$d(t) = \sum_{n=-\infty}^{\infty} d[n]\,\Pi_{T_b}(t - nT_b)$$

$$c(t) = \sum_{n=-\infty}^{\infty} c[n]\,\Pi_{T_c}(t - nT_c)$$

where $\Pi_T(t)$ is a rectangular pulse of duration $T$ and unit amplitude. $T_b$ and $T_c$ are bit and chip durations, respectively. We, however, assume both orthogonal (Walsh) channelization and pseudo-noise spreading as *information-independent* coding schemes since both functions are carried out in discrete-time. We express a SSM signal as

$$s(t) = g(t)\,\cos(2\pi f_o t + \phi)$$

by emphasizing that spread spectrum modulation is a baseband process with

$$g(t) = b[n] \sum_{-\infty}^{\infty} \psi(t - nT_c)$$

where $\psi(t)$ is a unit amplitude, possibly rectangular common chip pulse of duration $T_c$; it is also the impulse response of the baseband transmit filter. The data and spread-spectrum discrete sequence is

$$b[n] = d[l] \cdot c[n]$$

Different indices $l$ and $n$ are used to emphasize different clock rates —$R_b$ for data $d[l]$ and $R_c$ for spread-spectrum code $c[n]$. With discrete sequences there is no longer concern for maintaining time epoch synchronization and phase coherence between data $d(t)$ and SS modulating signal $c(t)$. If spread-spectrum modulation is a two-step process involving Walsh coding, then

$$b[n] = d[l] \cdot p[n] \cdot w_k[m]$$

where $w_k[m]$ is the $k^{th}$-row vector of a Walsh matrix. Its clock rate need not be the same the pseudo-random code $p[n]$. Based on our interpretation, the generation of a spread-spectrum modulated signal can be compartmentalized into discrete and analog sections as illustrated in Fig. 1.13. Nevertheless, we sometimes refer to Walsh and PN coding as Walsh and SS modulation, respectively. In general the term "modulation" is valid for

1. *Data Modulation*: bit-to-symbol mapping, even though it is a baseband operation typically carried out in discrete-time.

2. *Sequence Modulation*: Channelization coding (Walsh modulation) and PN coding (spread-spectrum modulation). Both are carried out in discrete-time domain.

3. *Carrier Modulation*: Analog multiplication for frequency translation.

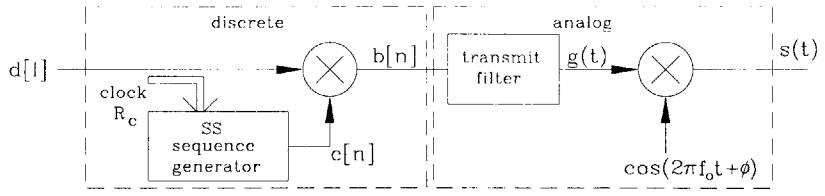



**Figure** 1.13: Preferred description of spread-spectrum modulation

### Processing Gain vs. Spreading Factor

Following the suggestion by Dr. M. K. Simon, the processing gain $P_G$ is defined as the ratio of the transmission bandwidth (the extent of the power spectral density) of the signal and its raw uncoded information rate $R_b$. Its unit is Hertz per bit per second. It is independent of the various signal processing techniques such as information dependent channel coding, orthogonal channelization and PN spreading. The spreading factor SF —or equivalently, the spreading gain SG— is a ratio of chip rate $R_c$ and (possibly channel encoded) bit rate R. Its unit is chips per bit. Unlike the processing gain, the spreading factor is always integer-valued.

Historically, the processing gain is the ratio of the signal-to-noise ratios after and before matched filtering:

$$P_G = \frac{\text{output signal-to-noise ratio}}{\text{input signal-to-noise ratio}} = \frac{SNR_o}{SNR_i} = \frac{W_{ss}}{W}$$

Matched filtering causes the desired signal to "collapse" from its transmit wideband SS signal (of bandwidth $W_{ss}$) to the original narrowband signal of Nyquist bandwidth $W$ while expanding the bandwidth of narrowband interference to $W_{ss}$. Hence, after "processing" the receiver output has a power gain of $W_{ss}/W$. The historical $P_G$ is equivalent to our accepted definition only when $W = R_b$.

### Signal-to-Noise Ratio vs. Bit Energy-to-Noise Density Ratio

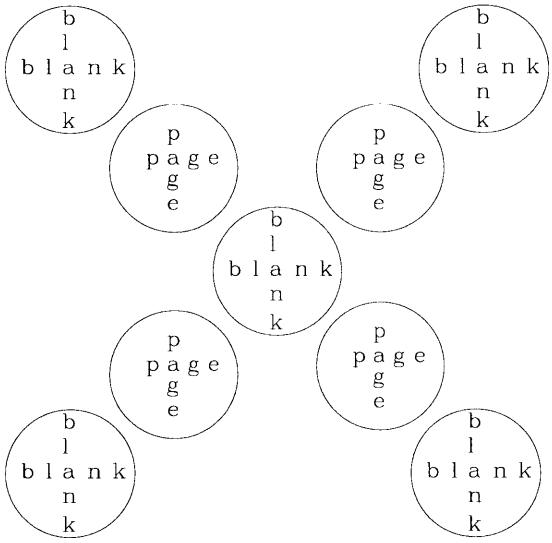
The bit energy $\mathcal{E}_b$ is a fictitious quantity related to symbol energy $\mathcal{E}_s$:

$$\mathcal{E}_b = \frac{\mathcal{E}_s}{\log_2 M}$$

where M is the number of bits of information contained in a symbol. The signal-to-noise *power* ratio SNR is defined as

$$\begin{aligned} SNR &= \frac{\text{signal power}}{\text{noise power}} = \frac{P}{N} = \frac{P}{N_o W} \\ &= \frac{\mathcal{E}_b R_b}{N_o W} = \frac{\mathcal{E}_b}{N_o} \cdot \frac{1}{P_G} \end{aligned}$$

where $N_o$ is the (flat) noise power spectrum and $W$ is signal bandwidth.


blank
page
blank
page
blank
page
blank
page
blank
blank

# 2

# FADING PHENOMENA AND MODELS

## Summary

There is a large collection of excellent texts and journal articles that cover many aspects of signal fading for radio wave propagation. Our results presented herein are not new; in fact, the entire content is merely careful note-taking and re-interpretation of material from various references listed in the bibliography. However, most communication texts treat fading as an abstract mathematical construct with statistical modelling, and fail to account for its accuracy as well as deviation from actual observed fading phenomenon and its associated empirical data obtained from field trials. Closed-form analytical results such as channel capacity and bit-error rates are only as good as the underlying mathematical fading models they are based on. When studying fading channel models, two important benchmarks in system performance are:

- The validity of a particular statistical model and its associated distribution function in predicting the strength and time variation of received signal. This mainly involves curve-fitting of statistical samples to empirical data samples.
- The severity of a fade measured in terms of various parameters that define a statistical model and its associated distribution function. This step is critical in comparative analysis of a mobile communication system perturbed by different types of fading. For example, it is incomplete to state the loss in capacity when a channel suffers from flat fading (compared to a non-fading AWGN case) without identifying the type of fading (i.e., its associated pdf) and its fading parameter (i.e., severity of fade).

It is our view that the inclusion of a chapter on channel fading is not redundant; Rather, it gives a unique synopsis linking experimental data and analytical fading channel models. The contents

presented herein also serve as background material for related topics covered in Chapters 3 and 6:

- In Chapter 3 the capacity and achievable rate regions of broadcast and compound channels in a multipath fading environment are computed using various probability density functions (pdf) that model the statistical behavior of a fading received signal envelope. These statistical distribution functions are presented in Appendix 2A.
- In Chapter 6, the baseband fading channel model (described in Sec. 2.2) is used to compute the bit-error probability of several multi-rate transmission schemes.

## 2.1 Fading Channel Characteristics

Fading in radio propagation refers to *random* fluctuation and gradual degradation in the amplitude (or power) of received signal. It is well known from the law of propagation that in *free space* the received power level of a signal is inversely proportional to the square of distance between the transmitter and receiver. The signal *attenuates* as it propagates through space. This type of signal deterioration is not considered fading since the power loss is a function of a *deterministic* variable: distance. Amplitude fluctuation (without attenuation) is caused by constructive and destructive combining of the RF carrier phases from multiple replicas of the same transmit waveform. The *statistical* nature of a faded signal is the result of two causes:

- Random scattering and diffraction of a transmitted signal from large and small objects in the propagation medium —the radio channel.
- Time variation due to mobility in one or more of the following entities: the transmitter, the receiver or the physical medium itself.

Any change in the angle of signal propagation due to reflection and refraction from large objects is called *scattering*[1]. When the objects are small — in the vicinity of several wavelengths, *diffraction* occurs as a travelling wave bounces off the edge or corner of objects. The transmitted signal can still be measured in the shadow region behind the obstacle (see Fig. 2.1)[2]. The reception of multiple copies of a transmitted waveform due to scattering and diffraction results in *multipath interference*. Signal fading is not observed without multipath reception. However, the reception of multiple copies of the transmit signal alone does not generate a faded signal unless time variation of the amplitude due to random signal fluctuation is also realized[3]. There is a cautionary note on the definition of a radio channel; it is time-variant if:

[1] A single primary wave incident on a rough surface is said to create many *scattered* secondary waves.

[2] Huygen's Diffraction Principle explains this phenomenon [99].

[3] Otherwise, we are merely modelling a time-invariant channel with echoes such as the intersymbol-interference (ISI) channel.

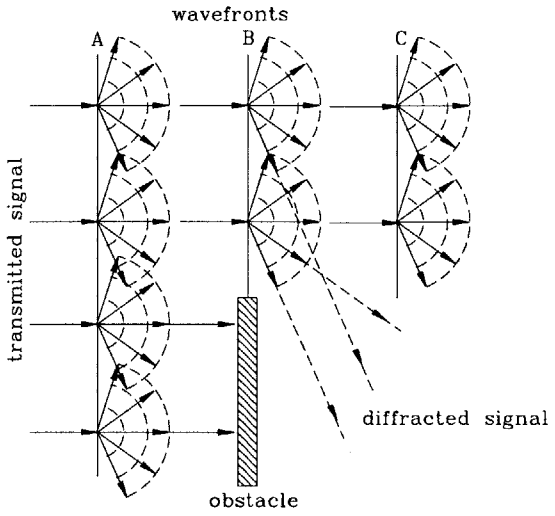



**Figure** 2.1: Penetration of diffracted signals behind obstructed objects.

- the receiver is in motion and the propagation medium is static (time-invariant); i.e., all physical objects and structures between the transmitter-receiver pair are immobile. Of course, the transmitter can also be in motion, but we assume base sites are stationary.

- the medium itself is time-varying, due to the movement of physical objects such as cars, pedestrians etc.

As an example, consider a mobile receiver moving at a walking pace of 2 mi/hr. A radio signal at 900 MHz transitions from its peak level to a minimum over a short distance of 3 inches. Hence, the received signal fluctuates about its median level 5-6 times every second. We define this type of fading as *localized* or *short-term fading*. The received sample mean signal strength (averaged over several tens of wavelengths) remains constant as the instantaneous level fluctuates around this mean value. In the literature this type of fading is known as "fast" fading. The name stems from the fact that a moving receiver observes rapid fluctuation in the instantaneous level of received signal. We however reserve the term fast fading for a different kind of channel disturbance. As a receiver moves further away from the transmitter, the mean value of the received signal decreases. The main culprit is propagation (path) loss due to increased distance in separation. Once the physical medium consisting of hilly terrain, buildings, trees and other large objects are taken into account, the mean value itself is a random parameter. This type of fading where the mean signal slowly fluctuates as it attenuates over large distances is called *wide area* or *long-term fading*. More common terms are "slow" fading and "shadowing" of large objects.

It is generally accepted that for terrestrial communication in the VHF, UHF and SHF regions,

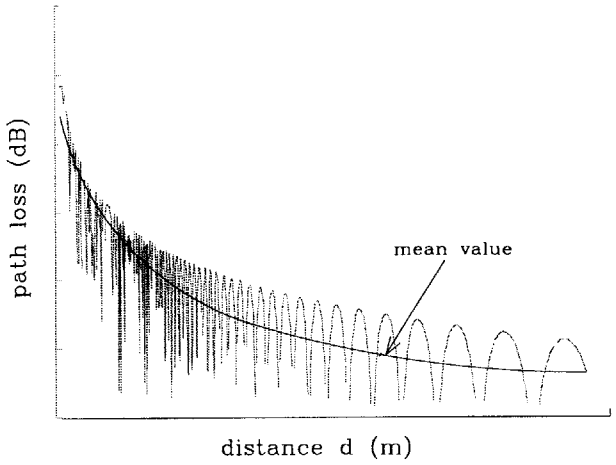



**Figure 2.2:** Plot of path loss (in dB scale) vs. distance (in meters) of a received signal direct line-of-sight path and a delayed path from a single reflecting surface.

the radio propagation channel has a fairly linear frequency response. By modelling it as a time-variant linear filter, its impulse response can be expressed as

$$c_p(t;\tau) = \sqrt{L_p} \sum_{n=1}^{L(t)} \alpha_n(t)\, \delta\Big(\tau - \tau_n(t)\Big) \tag{2.1}$$

The variables in the RHS can be broken down into the following:

- The *propagation* or *path loss* is the ratio of received power $P_r$ and transmit power $P_t$ levels (in watts):

$$L_p = \frac{P_r}{P_t} = A\frac{L_f L_{ta} L_{ra}}{L_d} \tag{2.2}$$

  A is the proportionality constant, $L_f$, $L_{ta}$, $L_{ra}$ and $L_d$ are path losses due to carrier frequency $f_c$, transmit antenna height $h_t$, receive antenna $h_r$ and distance d, respectively. A simplified formula based on geometrical models is

$$L_p = \frac{P_r}{P_t} = \Big(d\frac{\lambda}{2\pi d}\Big)^2 G_{ta}\, G_{ra} \sin^2\Big(\frac{2\pi h_t\, h_r}{\lambda d}\Big) \tag{2.3}$$

  where $\lambda$ is the signal wavelength, $G_{ta}$ and $G_{ra}$ are transmit and receive antenna gains, respectively[4]. A plot of path loss $L_p$ as a function of distance d is shown in Fig. 2.2. Note the variation of signal strength around its mean value and the high fluctuation rate at short distances. For large distances where $d \gg h_t$, $h_r$, $\sin x \approx x$ and the equation

[4] An isotropic antenna has unity gain.

simplifies to:

$$L_p = G_{ta}\, G_{ra} \frac{(h_t\, h_r)^2}{d^\gamma} \tag{2.4}$$

The dependence of path loss on carrier frequency is removed. The propagation exponent $\gamma$ ranges between 2 to 8. In free space, $\gamma = 2$. When there is reflection from a single flat plane, it can be shown that $\gamma = 4$. Over the years, numerous field measurements have shown that the path loss formula of eqn. (2.4) is unsatisfactory in predicting signal strength over various types of terrain and frequency bands. Hata's linear regression formula is much more accurate. It is based on graphs of empirical path loss data measured by Okumura. For quasi-smooth terrain in an urban area, the various path loss terms of eqn. (2.2) in logarithmic (dB) scale are:

$$\begin{aligned}
A &= 69.55 \\
L_f &= 26.16 \log f_c \text{ (in MHz)} && 150 \le f_c \le 2000 \\
L_{ta} &= -13.82 \log h_t \text{ (in m)} && 30 \le h_t \le 200 \\
L_{ra} &= (1.56 \log f_c - 0.8) - (1.1 \log f_c - 0.7) h_r \text{ (in m)} && 1 \le h_r \le 10 \\
L_d &= (6.55 \log h_t - 44.9) \log d \text{ (in km)} && 1 \le h_r \le 20
\end{aligned}$$

Minor variations of above formula are used for suburban and rural open space coverage areas. No accurate statistical model exists for path loss.

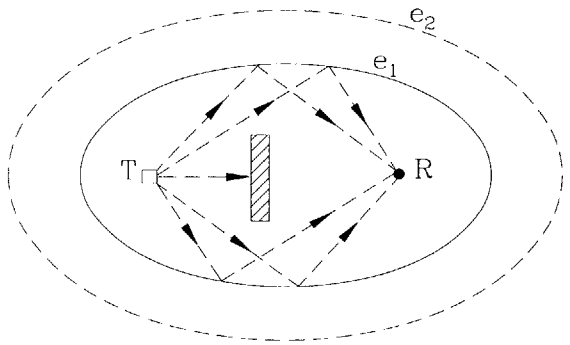



**Figure** 2.3: Single scattering model with multiple unresolvable paths and direct no line of sight (LOS) path.

- L(t) is the number of *resolvable* signal paths[5]. Two paths are resolvable if the difference in their arrival times of a carrier-modulated pulse is as large as the inverse of pulse rate $1/R_p$. (It does not make much sense to define "resolvable paths" when the transmit signal

[5]See Sec. 2.3.4 for a concise definition of *resolvable* multipaths.

is a pure carrier wave.) Consider Fig. 2.3. The transmitter and receiver are located at the foci of confocal ellipses. All signal paths that scatter and diffract at the perimeter of the inner ellipse $e_1$ travel the same distance. They are not resolvable. An example where the two paths may be resolvable is shown in Fig. 2.4 in which the scattering and diffraction occur at the perimeters of confocal ellipses $e_1$ and $e_2$. In general, the number of resolvable paths varies over time; however, it is commonly treated as a constant over several pulse transitions if the fading is slow. (Slow fading is defined in Sec. 2.3.3.)

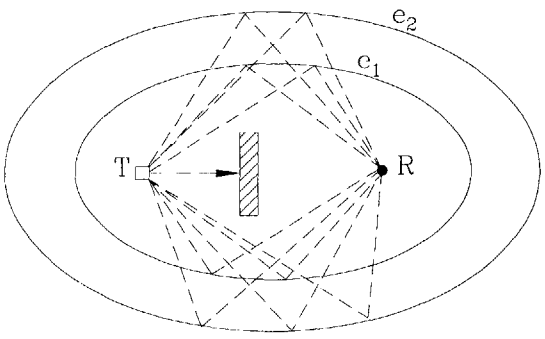



**Figure 2.4:** Multiple scattering model of two resolvable paths (scattered from two concentric ellipses) with no line of sight (LOS) path.

- For each realization, the *channel gain* $\alpha_n(t)$ in the $n^{th}$ path is a sample function of a random process. The relative path delays are $\tau_1 = 0 < \tau_2 < \ldots < \tau_L$ with an added constraint $\tau_{n+1} - \tau_n \geq T_p = 1/R_p$. From Fig. 2.3, we know that the $n^{th}$ path may actually be a superposition of a large number of unresolvable paths $L_u$ such that:

$$\alpha_n(t) = \sum_{l=1}^{L_u} \alpha_{nl}(t) \tag{2.5}$$

If each signal replica traverses through a different physical path in reaching the receiver, it is plausible to assume that the unresolvable gain processes $\alpha_{nl}(t)$ are independent. On the other hand, using the multiple scattering model of Fig. 2.5, the channel gain $\alpha_n(t)$ in the $n^{th}$ path can be expressed as a product of signal gains from multiple scattering objects:

$$\alpha_n(t) = \prod_k \alpha_{nk}(t)$$

Since scattering occurs from different objects, it is assumed that the gain factors $\alpha_{in}(t)$ are statistically independent. In logarithmic [dB] scale,

$$\alpha_n(t)\ [\text{dB}] = \sum_k \alpha_{nk}(t) \tag{2.6}$$

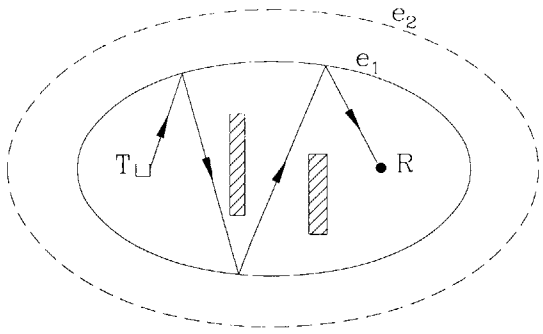



**Figure 2.5:** Multiple scattering model of a single path with no line of sight. Each scatterer reduces the signal strength.

In eqn. (2.1) it is assumed that the number of delayed paths L is finite at all time. This assumption is shown to be quite accurate for terrestrial radio communications in VHF/UHF frequency bands and where no over-the-horizon propagation takes place. The key results of this section are summarized below:

- The prediction of a received signal strength at a distance d from a transmitter is based on three sets of measures: path loss, local mean and instantaneous fluctuation about the mean. Path loss is a deterministic measure based on a simple mathematical formula of eqn. (2.2). The variation of local mean over long distance and fluctuation of instantaneous level over short distances is modelled statistically.

- Fading is due to time variation of the channel and reception of multiple replicas of transmit signal. Multipath interference is a result of scattering and diffraction of a transmit signal from various physical objects encountered in the transmission path. Two multipath replicas are resolvable if their arrival times are at least $T_p$ sec. apart.

- Each received signal replica from a resolvable path is most likely a superposition of signals with the same delay — originating from a large number of scatterers. Furthermore, each replica from an elemental path may have bounced off many objects before arriving at the receiver.

## 2.2 Equivalent Baseband Representation

The passband impulse response of a fading channel expressed in eqn. (2.1) is interpreted as the response of a linear filter at time t due to an input impulse at time $t - \tau$. It is a function of both time delay $\tau$ and position t. For a linear time-invariant (LTI) channel, the impulse

response $c(\tau)$ is a function of time delay only. An equivalent representation of an LTI channel — to within a scale factor— is its frequency response $C(f)$, where $C(f) = \mathcal{F}\{c(\tau)\}$. By defining the Fourier transform and its inverse as

$$\mathcal{F}[f(t)] = \int f(\tau)\, e^{j2\pi f\tau}\, d\tau$$

$$\mathcal{F}^{-}[f(t)] = \int f(\tau) e^{-j2\pi f\tau}\, d\tau$$

we can define three types of frequency responses for a linear time-variant (LTV) filter $c(t;\tau)$:

$$C(t;f) = \mathcal{F}_{\tau}\{c(t;\tau)\}$$

$$C(\nu;\tau) = \mathcal{F}_{t}\{c(t;\tau)\}$$

$$C(\nu;f) = \mathcal{F}\mathcal{F}\{c(t;\tau)\}$$

The middle term, the Fourier transform in the t variable —also called the *Delay-Doppler Spread function*— is the most useful since it is a measure of the channel response due to multipath delay $\tau$ and Doppler frequency shift $\nu$. In practice, two kinds of sounding experiments are conducted in measuring the response of an LTV channel. The first involves transmission and reception of a single-tone, radio frequency carrier wave. Due to its high frequency resolution and constant amplitude over time, it is used to measure the Doppler shift, the channel gain and relative phase delay. The second experiment uses a stream of short pulses with very low duty cycle. Its high time resolution and wide power spectrum are suitable for measuring resolvable multipath replicas. Ultimately, the goal is to understand the effect of fading once the received passband signal is down-converted and filtered at the baseband level. We next describe the baseband channel response when the input is a single-tone unmodulated carrier or a stream of impulses.

### 2.2.1 Frequency Resolution

Consider a single-tone, continuous-wave (CW) signal

$$A_c \cos 2\pi f_c t$$

propagating through a single-path, linear time-invariant channel. The received signal is:

$$\begin{aligned} r_p(t) &= \alpha A_c \cos 2\pi f_c (t-\tau) \\ &= \alpha A_c \cos(2\pi f_c t + \phi) \qquad \phi = -2\pi f_c \tau \\ &= A_c\left[X \cos 2\pi f_c t + Y \sin 2\pi f_c t\right] \end{aligned} \tag{2.7}$$

where the in-phase and quadrature components are:

$$\begin{aligned} X &= \alpha \cos\phi \\ Y &= -\alpha \sin\phi \qquad\qquad 0 \le \alpha < 1 \text{ and } 0 \le \phi < 2\pi. \end{aligned}$$

If the channel is time-invariant with multiple paths, two different scenarios are possible.

- In the first model, it is assumed that there exists only a single scattering object between the transmitter and the receiver[6]. If both the transmitter and receiver are located at the focal points of an ellipse, then all signal replicas that are scattered from objects located at the perimeter of the ellipse $e_1$ will arrive at the receiver with the same delay $\tau_1$ (see Fig. 2.3). For $\tau_1 = \tau$, and following eqn. (2.7), the received signal is:

$$\begin{aligned} r_p(t) &= \Big(\sum_l \alpha_l\Big) A_c \cos 2\pi f_c (t-\tau) \\ &= A_c\Big[\Big(\sum_l X_l\Big)\cos 2\pi f_c t + \Big(\sum_l Y_l\Big)\sin 2\pi f_c t\Big] \end{aligned} \tag{2.8}$$

where $l$ is the number of scattering multiple propagation paths.

$$\begin{aligned} X &= \sum_l X_l = \Big(\sum_l \alpha_l\Big)\cos\phi = \alpha\cos\phi \\ Y &= \sum_l Y_l = -\Big(\sum_l \alpha_l\Big)\sin\phi = -\alpha\sin\phi \end{aligned} \tag{2.9}$$

$$\text{Note,} \qquad \alpha^2 = \sum_l \alpha_l^2 \le 1 \qquad \text{due to energy conservation.} \tag{2.10}$$

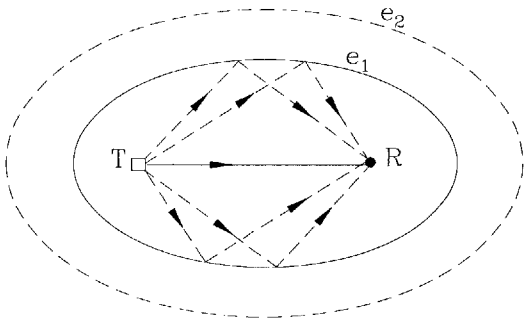



**Figure 2.6:** Single scattering model with multiple unresolvable paths including line of sight

- In the second model, there exists a *direct line-of-sight* (LOS) path from the transmitter to the receiver. If the transmitter is secondary, this path is called *specular*. Typically, the LOS path without any scattering or diffraction has the largest energy content and the shortest

[6]Of course, the transmitter itself may be *secondary*, i.e. it is a reconstruction of scattered and diffracted signals from a distant transmitter.

path, and it is considered deterministic. Modifying eqn. (2.8), where $\beta$ is the attenuation in LOS path,

$$r_p(t) = \Big(\sum_l \alpha_l\Big) A_c \cos 2\pi f_c(t-\tau) + \beta A \cos 2\pi f_c t \tag{2.11}$$
$$= A_c\Big[(\beta + \sum_l X_l)\cos 2\pi f_c t + (\sum_l Y_l)\sin 2\pi f_c t\Big]$$

$$X = \beta + \sum_l X_l = \beta + \Big(\sum_l \alpha_l\Big)\cos\phi \tag{2.12}$$
$$= \beta + \alpha\cos\phi$$

$$Y = \sum_l Y_l = -\Big(\sum_l \alpha_l\Big)\sin\phi \tag{2.13}$$
$$= -\alpha\sin\phi$$

If the channel is time-varying:

$$X(t) = \alpha(t)\cos\phi(t)$$
$$Y(t) = -\alpha(t)\sin\phi(t)$$

The real in-phase and quadrature random processes of a passband channel can be represented in the baseband by a single complex function $Z(t) = \alpha(t)e^{\phi(t)}$:

$$|Z(t)| = \sqrt{X^2(t) + Y^2(t)} = \alpha(t) \tag{2.14}$$

$$\phi(t) = \tan^{-1}\left(\frac{Y(t)}{X(t)}\right) \tag{2.15}$$

The effect of carrier frequency on fading is observed only in the phase offset $\phi(t) = -2\pi f_c\tau(t)$.

### 2.2.2 Time Resolution

Consider a linearly modulated passband signal

$$s_p(t) = \Re\{s(t)e^{j(2\pi f_c t+\phi)}\}$$

where $s(t)$ is the corresponding baseband information signal:

$$s(t) = \sum_i A_i e^{j\theta_i} f(t - iT_p)$$

where $R_p = 1/T_p$ is the information (pulse) rate and $\{A_i e^{j\theta_i}\}$ is the finite signal constellation set. The finite-energy common pulse shape is denoted by $p(t)$. Ideally, $p(t)$ is an impulse. Considering only a one-shot transmission with $A_i = A_c$, the transmit passband signal is:

$$s_p(t) = A_c f(t)\Big[\cos\theta\,\cos 2\pi f_c t - \sin\theta\,\sin 2\pi f_c t\Big]$$

The received signal is:

$$\begin{aligned} r_p(t) &= A_c \alpha f(t-\tau)\Big[\cos\theta\,\cos 2\pi f_c(t-\tau) - \sin\theta\,\sin 2\pi f_c(t-\tau)\Big] \\ &= A_c \alpha f(t-\tau)\Big[\cos\theta\,\cos(2\pi f_c t + \phi) + \sin\theta\,\sin(2\pi f_c t + \phi)\Big] \\ &= \Re\Big[A_c e^{j\theta} f(t-\tau)\;\alpha e^{j\Phi} e^{j2\pi f_c t}\Big] \end{aligned}$$

The equivalent complex baseband expression is:

$$r(t) = \alpha\, e^{j\,\Phi}\, s\big(t-\tau\big)$$

If there are more than one resolvable path ($\tau_{n+1} - \tau_n \geq T_p$) and that the total number of resolvable paths is L,

$$r(\tau) = \sum_{n=1}^{L} \alpha_n\, e^{j\,\phi_n}\, s\big(\tau - \tau_n\big)$$

Note that the indeterminate variable for the argument of $r(\cdot)$ is switched from $t$ to $\tau$. By including time variation of channel parameters, the final form of a baseband received signal is:

$$r(t;\tau) = \sum_{n=1}^{L(t)} \alpha_n(t)\, e^{-j\,\phi_n(t)}\, s\Big(\tau - \tau_n(t)\Big) \tag{2.16}$$

The equivalent baseband multipath channel impulse response is:

$$c(t;\tau) = \sum_{n=1}^{L(t)} \alpha_n(t)\, e^{-j\,\phi_n(t)}\, \delta\Big(\tau - \tau_n(t)\Big) \tag{2.17}$$

The relative carrier phase offset $\phi_n(t) = \phi_o + 2\pi f_c \tau_n(t)$ is the sum of a (fixed) modulator carrier phase $\phi_o$ plus the offset due to delay $\tau_n(t)$. Since the receiver has no knowledge of $\phi_o$, its value must be estimated through a phase-locked-loop mechanism. At the receiving end it is treated as a uniformly distributed random variable $\in [0, 2\pi)$. Since $\phi_o$ and $\tau_n(t)$ are independent, $\phi_n(t)$ and $\tau_n(t)$ are also independent. We then safely assume for each path $n$, the random variables $\phi_n$, $\alpha_n$ and $\tau_n$ are mutually independent.

So far the channel response $c(t;\tau)$ is treated as a deterministic time function. In fact, it is a function of several random parameters L, $\alpha_n$, $\phi_n$ and $\tau_n$ and must be treated as a stochastic process in the $t$ variable. It is already established that for each resolvable path $n$, the corresponding variables $\alpha_n$, $\phi_n$ and $\tau_n$ are mutually independent. Not much is said about the relationship between these parameters for different $n$, or for a particular variable along different paths. In general, their joint probability functions are required to completely characterize the statistical properties of $c(t;\tau)$, leading to a very complex description of a fading channel. Certain assumptions can simplify this model. They are considered next.

### 2.2.3 Wide-Sense Stationarity

It is safely assumed that in a localized area, the fading statistics of a channel do not vary appreciably. When the dynamics of the propagation medium or the mobile receiver is slow compared to the duration of several data symbols, it may be possible to accurately estimate the gain and delay response of a channel. The estimation process must be carried out on a periodic basis. Such a channel where the channel remains time-invariant for a block of data symbols is said to be *quasi-wide-sense stationary*. In the duration of each data block, it is wide-sense stationary (WSS). Thus its mean is time-independent, and its autocorrelation function is time-invariant:

$$\begin{aligned} R_c(t_1, t_2; \tau_1, \tau_2) &= \frac{1}{2} E\left[c(t_1; \tau_1)\, c^*(t_2; \tau_2)\right] \\ &= \frac{1}{2} E\left[c(0; \tau_1)\, c^*(t_2 - t_1; \tau_2)\right] \\ &= R_c(t_2 - t_1; \tau_1, \tau_2) \\ &= R_c(\Delta t; \tau_1, \tau_2) \qquad \text{where } \Delta t = t_2 - t_1 \end{aligned} \tag{2.18}$$

In general, the autocorrelation $R_c(t_1, t_2; \tau_1, \tau_2)$ is a function of two time variables $t_1$ and $t_2$. Its characterization in frequency domain, the power spectral density $\psi(f_1, f_2; \tau_1, \tau_2)$, is obtained by taking double Fourier transforms. It can be shown that for a wide-sense stationary process,

$$\delta(f_2 - f_1)\, \psi(f_2; \tau_1, \tau_2) = \mathcal{F}^- \mathcal{F}\left[R_c(\Delta t; \tau_1, \tau_2)\right]$$

where the *Delay-Doppler cross power spectral density* function is defined as:

$$\psi(f_2; \tau_1, \tau_2) = \int R_c(t_1, t_2; \tau_1, \tau_2) \exp\left(-j2\pi f_2 \Delta t\right) d(\Delta t) \tag{2.19}$$

$f_1$ and $f_2$ are the corresponding frequency variables of $t_1$ and $t_2$ respectively. Note that the power spectral density function is *not* frequency-invariant; i.e., it is not wide-sense stationary in the frequency domain. However, it is interpreted as follows: If two scattering components of the transmit signal of different or same delays $\tau_1$ and $\tau_2$ experience different amounts of Doppler shifts ($f_1 \neq f_2$), then the two paths are uncorrelated. The implication is that two paths with different angles of arrival (regardless of their arrival times) are uncorrelated since their Doppler shifts are different[7] Similarly, two paths with the same angle of arrival, but with different Doppler shifts, are also uncorrelated. For single path or a channel with a single resolvable path,

[7]It is known that a vehicle travelling at a speed of $v$ in an $x - y$ plane at an angle $\zeta$ from the x-axis experiences a frequency (Doppler) shift of a transmitted signal by an amount equal to $(v/\lambda)\cos(\zeta - \alpha)\cos\beta$ where $\alpha$ is the angle from the x-axis and $\beta$ is the projection angle onto the $x - y$ plane of the incoming wave.

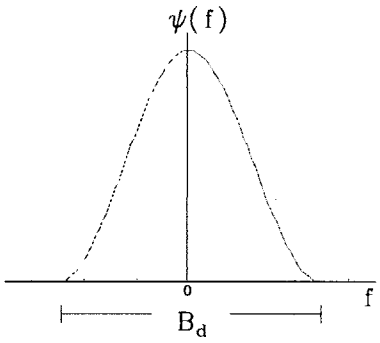



**Figure 2.7:** Doppler Spread function. The end-to-end extent of its support is called the Doppler Spread $B_d$.

$\tau_1 = \tau_2$ and setting it equal to zero, we obtain the *Doppler Spread* function:

$$\begin{aligned}\psi(f) &= \psi(f;\, 0, 0) \\ &= \int R_c(\Delta t;\, 0, 0) \exp\left(-j2\pi f \Delta t\right) d(\Delta t) \\ &= \mathcal{F}^{-}\left[R_c(\Delta t)\right]\end{aligned} \tag{2.20}$$

An example of a doppler spread function is illustrated in Fig. 2.7. The maximum frequency where the value of the Doppler spread function $\psi(f)$ is essentially zero is called the *Doppler Spread* $B_d$. Its reciprocal

$$(\Delta t)_c = (t_2 - t_1)_c = \frac{1}{B_d} \tag{2.21}$$

is called the *coherence time*. Thus a channel with a small Doppler spread has a large coherence time, implying that the channel dynamics vary very slowly. The channel induces *slow* fading. If $(\Delta t)_c$ is a fraction of $T_p$ (the inverse of pulse rate), the received signal suffers from *fast* or *time-selective* fading. In system design and analysis, the autocorrelation and its Doppler spread functions are rarely computed; instead, the performance measure is based on the worst-case scenario where the maximum Doppler shift is used as a benchmark. The maximum Doppler shift experienced by a mobile receiver travelling with a speed of $v$ is given by

$$B_d(\max) = \frac{v f_c}{c} \tag{2.22}$$

where c is the speed of electromagnetic waves in free space.

### 2.2.4 Uncorrelated Scattering

Recall that $c(t;\tau)$ is a random process in the t variable. By taking a pair of time delays $\tau_1$ and $\tau_2$, its correlation function, analogous to eqn. (2.18), can be computed. If we assume the

channel gains $\alpha_n$ or phase shifts $\phi_n$ or both are uncorrelated for paths with different delays, the autocorrelation function simplifies to:

$$R_c(t_1, t_2; \tau_1, \tau_2) = R_c(t_1, t_2; \tau_2)\, \delta(\tau_2 - \tau_1)$$

This is called the *Uncorrelated Scattering* (US) channel model. By setting $t_1 = t_2 = 0$, the *Delay Spread* function is expressed as:

$$R_c(\tau_1, \tau_2) = R_c(\tau_2)\, \delta(\tau_2 - \tau_1)$$

The delay spread function is also commonly known as the *multipath delay profile* function. Since there is a discrete set of resolvable multipaths, the plot of $R_c(\tau)$ tends to have several "spikes." This is illustrated in Fig. 2.8. (Here, it is assumed that the delays are relative to the specular or shortest path with $\tau_1 = 0$.) Several probabilistic models for multipath delay profile are given in Appendix 2B. Analogous to its frequency domain counterpart of eqn. (2.21), we can define

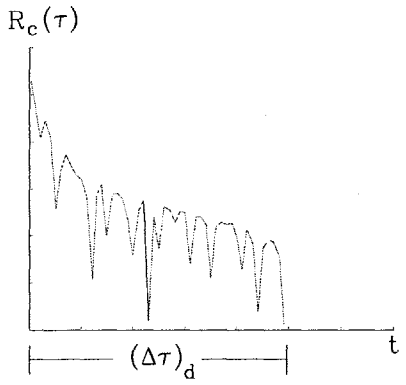



**Figure 2.8:** Delay Spread function. Unlike the Doppler spread function, it only has a finite positive support. Its extent is called the Delay Spread $(\Delta\tau)_d$.

the *Delay Spread* of a uncorrelated scattering multipath channel as the maximum value of time delay where the correlation function is essentially zero:

$$(\Delta\tau)_d = (\tau_2 - \tau_1)_d = \frac{1}{B_c} \tag{2.23}$$

Its reciprocal $B_c$ is the *coherence bandwidth*. Two unmodulated tones whose frequency separation is larger than $B_c$ are perturbed by different channel gains and phase shifts. Thus, a bandwidth of a baseband data modulated signal is wider than $B_c$, it is said to suffer from *frequency selective* fading.

### 2.2.5 WSSUS Channel

The simplest non-degenerate fading channel model combines both wide-sense stationarity (WSS) and uncorrelated scattering (US) properties. It is known as the wide-sense stationary, uncorrelated scattering (WSSUS) channel. It has been shown through many experimental measurements that a mobile radio propagation medium can accurately be modelled as a WSSUS channel in a local setting. The autocorrelation function simplifies to:

$$R_c(\Delta t;\Delta\tau) = R_c(t_2 - t_1;\tau_2 - \tau_1)$$

Its Fourier transform in the $\Delta t$ variable gives the *Scattering* function:

$$S(\Delta f,\Delta\tau) = \mathcal{F}_{\Delta t}\Big[R_c(\Delta t;\Delta\tau)\Big]$$

It is a two-dimensional *deterministic* function that succinctly and completely defines the time and frequency characteristics —the delay and Doppler spreads— of a WSSUS channel.

## 2.3 Classification of Fading

In the remaining sections, we ignore signal attenuation due to path loss. Fading is regarded as the large-scale variation of the mean and small-scale fluctuation of the instantaneous level about the mean. Next we distinguish fading as either short- or long-term, fast or slow, frequency flat or selective. However, we point out that such classifications are only for purposes of system design and analysis; there is no clear line that separates one type of fading from its opposite. There is however a clear distinction between fast and short-term fading, as well as between slow and flat fading.

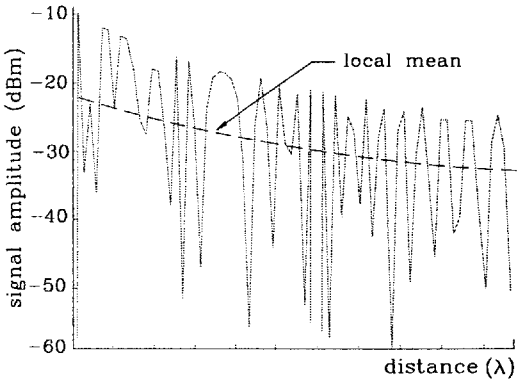



**Figure** 2.9: Localized short-term fading

### 2.3.1 Short-Term vs. Long-Term Fading

In a multipath environment, the signal strength of a mobile receiver fluctuates due to constructive and destructive combining of the phases of replicas. The random nature of amplitude or power variation is usually plotted in a decibel scale over a distance measured in wavelengths. Here we are interested in signal strength as a function of *distance*, not *time*. Fading of this kind over several hundred wavelengths is considered "short-term" or "localized" fading, since the average signal level over such a distance does not vary considerably. This is illustrated in Fig. 2.9 where the distance is measured in wavelengths. Over large distances, the mean signal strength decreases due to path loss. However, the drop in strength is not monotonic. It also fluctuates. This is categorized as "long-term" or "large-area" fading. See Fig. 2.2 for an illustration. The only difference between short- and long-term fading is the distance traversed. Each plays a different role in system analysis. For data detection and decoding, it is the instantaneous fluctuation of short-term fades that are most detrimental. In contrast, the mean signal strength is critical in analyzing the coverage area of a cell. For example, in an urban high-density area, we wish to know the diameter of a cell (with an omni-directional antenna) such as the local mean signal strength is guaranteed to 95% of the receiver population at the cell edge. We can then determine the number and location of base sites to support a cellular wireless service with a defined level of quality-of service (QoS). This is the basis for network planning.

For purposes of system design and analysis, the channel gain $\alpha_n(t)$ along path $n$ is considered as a product of both short- and long-term fading effects; i.e., it has a short-term fading component superimposed on a long-term random local mean value:

$$\alpha_n(t) = R_n(t) \cdot m(t)$$

where $R(t)$ and $m(t)$ are short- and long-term channel gains, respectively. For short-term analysis, the median level is normalized and treated as a constant; $m(t) = m = 1$. If fading is *slow*, the path gain is treated as a random variable $\alpha_n = R_n$ with a certain probabilistic distribution. Commonly used probabilistic density functions (pdf) for short-term fading gain are Rayleigh, Rice, Log-normal, Nakagami, M and One-sided Gaussian. All are described in detail in Appendix 2A. Their parameterized shapes are shown in Fig. 2.10. All fading pdfs have a positive infinite range $[0, \infty]$. Table 2.1 also summarizes their main characteristics. A brief overview is given below.

- Among the six pdfs, Rayleigh and One-sided Gaussian have a single free parameter. There are two free parameters for the remaining pdfs. Thus the former two are the least flexible in curve-fitting to empirical data. For Rayleigh distribution, the shape of its pdf cannot be varied. The only free parameter $\sigma$ shrinks or stretches the shape horizontally. In one-sided Gaussian pdf, its shape is one-half of bell curve. The free parameter $\sigma$ controls the spread. In both cases, the free parameter is a *scaling* metric.

**Table** 2.1: Characterization of Fading PDFs

| Name | Alternative name | free parameters | Comments |
|---|---|---|---|
| Rayleigh | Chi | $\sigma^2$ | two-dimensional chi density |
| Rice | Nakagami-n | K and $\sigma^2$ | $K = n^2$ |
| Nakagami | Nakagami-q | q and $\Omega$ | also known as Hoyt pdf |
| M | Nakagami-m | m and $\Omega$ | central chi-density with<br>2m degrees of freedom<br>most flexible pdf in severity range |
| Log-normal | | $\mu$ and $\sigma^2$ | the logarithm of a Log-normal RV<br>is Gaussian |
| One-sided Gaussian | Chi | $\sigma^2$ | one-dimensional chi density<br>Worst type of fading |

- In Rice pdf, the mean $\mu$ is a *translation* metric while K controls the spread and the skewness of the function. Small K gives rise to a more skewed shape. Thus, K is effectively a *scaling* and *shaping* metric. Rice is the only pdf with a free parameter for translation, making it the best candidate when a strong specular path is present. In the limit $K = 0$, the Rice pdf degenerates to the Rayleigh pdf.

- In Nakagami pdf, the free parameters q and $\Omega$ serve as *shaping* and *scaling* metrics, respectively. In contrast, both free parameters m and $\Omega$ of the m-distribution act simultaneously as *scaling* and *shaping* metrics.

- For the same local mean, the one-sided Gaussian pdf represents the most severe fading statistic. From the pdf plot, it is obvious that all amplitude values less than the mean have higher probability. For the same reason, any skewed fading distribution, where the skew is towards the origin such that the mean is greater than the mode[8] its detrimental fading effect is more severe that a another RV with symmetric pdf under equal mean condition. For one-sided Gaussian, the mode is 0. At the opposite extreme, a delta function at the local mean represents a non-fading case.

- In terms of the severity of a fade, the m-distribution has the widest range. As depicted in

[8]The mode of a RV is its most probable value; i.e., the location along the support where the pdf plot has its highest peak. The median point divides the area under the pdf plot in half. For a symmetric pdf, the mode, the mean and the median all coincide. For a skewed pdf that is heavy towards the origin, then $R_{mod} < R_{mean}$. The opposite holds when the the pdf skews away from the origin. If a pdf has a heavy tail, then it is likely that $R_{med} < R_{mean}$ since the mean is (approximately) the sum of weighted products $\sum_x x\,p(x + \Delta x)$; thus long tails with large values of x result in heavier weights. Depending on the shape of the pdf, the mode can lie anywhere along its support. Therefore, based on the shape of the pdf, each average quantity has its strength as well as weakness.

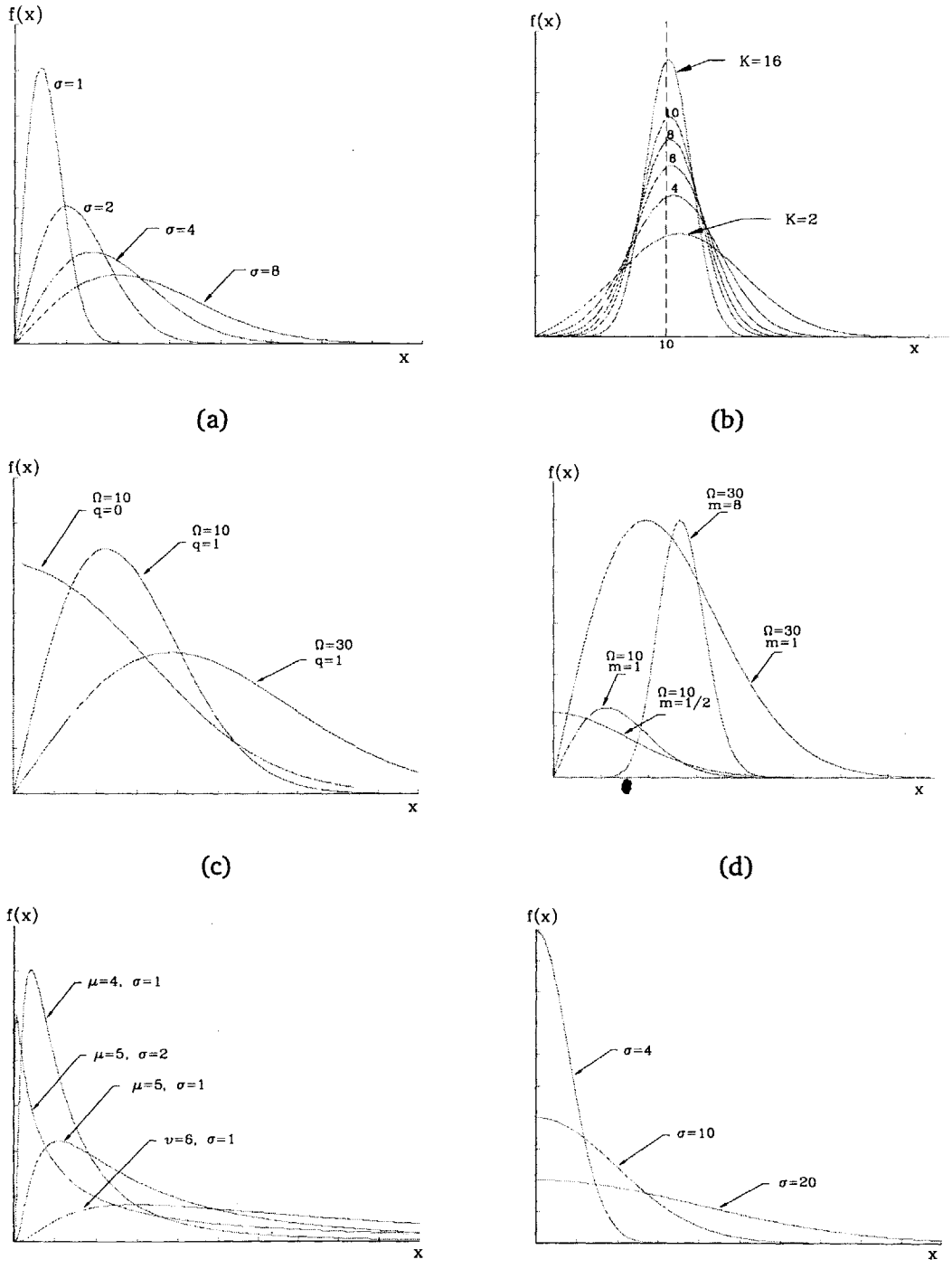



**Figure** 2.10: Fading pdfs: (a) Rayleigh (b) Rice (c) Nakagami (d) M (e) Log-normal and (f) One-sided Gaussian

Fig. 2.12, for $m = \infty$ it is equivalent to a delta function (no fade); for $m = 1/2$, it becomes the one-sided Gaussian pdf (worst-case fading). If our interest is curve-fitting for empirical data, we see from Fig. 2.10 that the Log-normal has the most flexible shape, ranging from a decaying type to a skewed shape.

For unresolvable multipath case with no line-of-sight (see Fig. 2.3), the Rayleigh density function suffices. This approximation is an application of the central limit theorem which states that the sum of independent, identically distributed (i.i.d.) random variables approaches a Gaussian distribution. Note however that the path gain is finite and bounded (see eqn. 2.10). Figure 2.11 depicts curve-fitting of a Rayleigh pdf over a histogram based on empirical data. If a strong dominant (specular) path is present, the Rayleigh model becomes inadequate. This

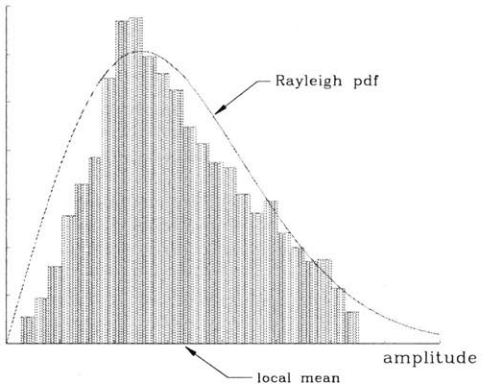



**Figure** 2.11: Curve fitting of Rayleigh pdf over empirical data. Note that due to finite transmit power, the measured signal amplitude has finite support.

is mostly because the Rayleigh pdf is a single-parameter function, and varying its mean also affects the variance. Several two-parameter pdfs such as Rice, Nakagami, M and log-normal are more suitable for fading with a specular path. Among them, the Rice pdf is the best fit since the envelope of non-zero mean Gaussian RVs with the same variance is Ricean distributed. It is confirmed through field trials that for long distances the log-normal pdf gives the best fit. A plausible explanation is based on the multiple-scatterer, single path model depicted in Fig. 2.5. Since the resultant channel gain is the product of independent gains from different scattering objects, its logarithmic form is the sum of many independent and identically distributed gains.

### 2.3.2 Severity of Fading

As we have seen, all fading pdfs are one-sided with infinite support, that is, they are defined only for non-negative values of the argument $[0, \infty)$. The severity of a fading distribution is

measured in terms of its roll-off or spread from its peak (mode) value. As an example, the delta (non-fading) pdf has zero spread. In Fig. 2.24, we show several values of the fading figure m for the Nakagami-m pdf. For $m = 1/2$, the fading is the most severe since the one-sided Gaussian density has the heaviest tail. Quantitatively, the spread is measured in terms of the variance. However, in certain fading distributions, the peak value of the pdf is not at the origin, and as a result, the amount of spread is "biased." This bias term can be removed by normalizing both the mean (or mode) and the variance. One method that has been used extensively in the literature and in practice was first proposed by U. Charash [18].

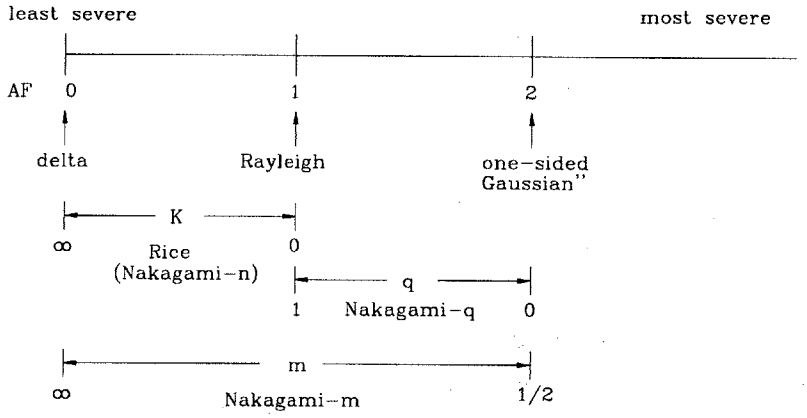



**Figure** 2.12: The range of amount of fading (AF) for various distributions

### Amount of Fading

U. Charash proposed a unified measure of the severity of fading, called the "amount of fading" (AF) by normalizing the variance of signal power by the square of its mean:

$$\begin{aligned} AF &= \frac{Var(P)}{E^2(P)} \\ &= \frac{E(P^2)}{E^2(P)} - 1 \end{aligned} \tag{2.24}$$

where the RV P denotes power, $P = \alpha^2$ and $\alpha$ is the amplitude RV. In the literature, it is common to define the instantaneous signal-to-noise power level per symbol:

$$\gamma = \alpha^2 E_s/N_o$$

where $E_s$ is the transmitted symbol energy and $N_o$ is the energy spectral density of Gaussian noise. Rewriting 2.24:

$$AF = \frac{Var(\gamma)}{E^2(\gamma)} \tag{2.25}$$

The AF values of all fading models introduced in this chapter are listed below. Each function represents the pdf of fading amplitude $\alpha$.

$$\text{One-sided Gaussian}: \quad \text{AF} = 2 \tag{2.26}$$

$$\text{Rayleigh}: \quad \text{AF} = 1 \tag{2.27}$$

$$\text{Delta}: \quad \text{AF} = 0 \tag{2.28}$$

$$\text{Rice}: \quad \text{AF}(K) = \frac{1+2K}{(1+K)^2} \qquad K \geq 0 \tag{2.29}$$

$$\text{Nakagami-q}: \quad \text{AF}(q) = \frac{2(1+q^4)}{(1+q^2)^2} \qquad 0 \leq q \leq 1 \tag{2.30}$$

$$\text{Nakagami-m}: \quad \text{AF}(m) = \frac{1}{m} \qquad m \geq \frac{1}{2} \tag{2.31}$$

$$\text{Log-normal}: \quad \text{AF}(\sigma) = \exp\left(\frac{\sigma^2}{4.34}\right) - 1 \qquad \sigma^2[\text{dB}] = \text{Var}(\alpha) \tag{2.32}$$

The amount of fading (AF) spans mostly from 0 to 2. (It is possible for $\text{AF}(\sigma) > 2$, particularly for log-normal pdf; however, as we shall see in Sec. 2.3.1, the log-normal fading qualifies more accurately as a *long-term* shadowing or attenuation due to large distance propagation. Other distributions under consideration have localized or *short-term* fading phenomena.) AF=0 is no fading and AF=2 is the worst-case one-sided Gaussian fading. At the mid-point, AF=1 corresponds to Rayleigh fading (no line-of-sight component). For Rice fading, depending on the strength K of the LOS path, AF varies from 1 (Rayleigh) to 0 (no fading). For a more severe fading spanning $1 \leq \text{AF} \leq 2$, the representative fading model is Nakagami-q. The only model that spans the entire range of AF is the Nakagami-m pdf. Thus, by varying the value of m, the Nakagami-m distribution encompasses all types of fading phenomena, from the worst-case to the non-fading AWGN model. The range of AF values for various distributions is shown in Fig. 2.12.

### 2.3.3 Slow vs. Fast Fading

In the previous subsection, we describe several probability distributions whose statistics mimic the random nature of the instantaneous amplitude level in a local area setting. We should emphasize that the radio system designer has little control over the severity of short-term fading statistics. It is simply a physical characteristic of the radio propagation medium that must be dealt with. We can only construct probabilistic models to accurately predict and analyze system degradation under such fading conditions. In contrast, the designer has tools —e.g., upper limit on receiver speed and information rate— to regulate the fading rate (slow or fast) and the contour (flat or frequency-selective) of the frequency response of the channel.

### Measure of Fading Rate and Depth

In Sec. 2.3.1 on short-term fading, it is implicitly assumed that the local mean value does not change substantially. If the motion of the mobile receiver is radial, it is highly likely that the mean value will change considerably unless its speed is low; i.e., the signal is perturbed by *slow* fading. On the other hand, if the receiver moves tangentially, then its (angular) velocity is irrelevant. Hence in short-term localized fading case, we must also determine the fading rate. As depicted in Fig. 2.9 —over the same distance, a high velocity receiver experiences faster signal fluctuation. Quantitatively, the rapidity of fades is measured in terms of the *level-crossing rate* (LCR), which is defined as the expected number of times—in a positive or negative slope—the signal strength (amplitude) crosses a predefined threshold level over some time window. The unit for the length of the time window is usually in seconds. Obviously, the higher the speed of the receiver, the larger the level-crossing rate; i.e., LCR is a function of receiver speed $v$. Besides the propagation medium, the other parameter that affects LCR is the radio carrier frequency $f_c$. The expected duration where the signal remains below this threshold over the same time window is defined as the *average fade duration* (AFD). Both are depicted in Fig. 2.13. Note the significance of the threshold. By varying its level, the fade depth (and its duration below a threshold) can be measured in terms of LCR and AFD. This information is very useful for a system designer in selecting information pulse rate, error-control code and its rate, word length, interleaver depth etc. Furthermore, LCR and AFD provide a minimum rate at which the received signal must be sampled in order to measure the severity (depth) of fade. This information is typically relayed back through a feedback channel to the transmitter to adaptively adjust its transmit power level. If the level crossing rate is low such that on

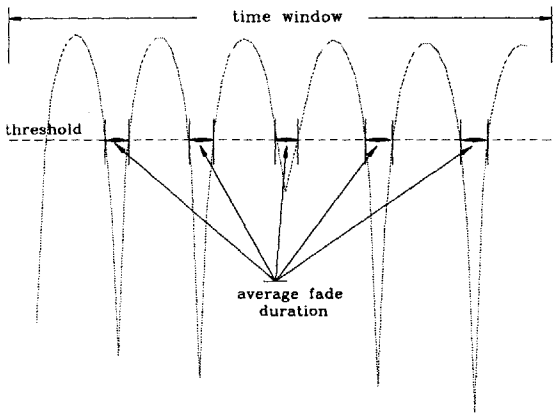



**Figure** 2.13: Level Crossing Rate (LCR) and Average Fade Duration (AFD)

average the signal amplitude along a particular path $n$ remains above the threshold over a few symbol intervals, then the carrier phase offset $\phi_n$ can be tracked accurately. The duration over

which the channel state is time-invariant is defined as the *coherence* time $\Delta t_c$. If $\Delta t_c \gg T_p$, the channel induces *slow* fading. An equivalent interpretation of slow fading is based on spectral broadening due to Doppler spread. This topic has already been addressed in Sec. 2.2.3 under wide-sense stationarity. If the bandwidth of a modulated waveform of rate $R_p$ pulses per second is $W$ Hz (e.g., $R_p \approx W$), an equivalent definition of slow fading is:

$$W \gg B_d \qquad \text{or} \qquad B_d T_p \ll 1 \tag{2.33}$$

In essence when the fading is slow, the receiver is able to track the carrier phase offset, implying that system design using phase modulation and coherent demodulation are feasible —at least theoretically.

### 2.3.4 Frequency Flat vs. Selective Fading

In short-term fading, if the relative delays $\tau_i - \tau_j$, $i > j$ of paths from local scattering and diffraction are small compared to $T_p$, then at the receiver the transmit signal has only minor distortion in its shape. Following eqn. (2.16), the received signal can be expressed as:

$$r(t) = \sum_{n=1}^{L_r} \alpha_n \, e^{-j\phi_n} \, s(t - \tau_n)$$

Substituting the expression for a stream of linearly modulated signal in place of $s(t)$, the equation becomes:

$$r(t) = \sum_{n=1}^{L_r} \alpha_n \, e^{-j\phi_n} \sum_i A_i e^{j\theta_i} p(t - iT_p - \tau_n)$$

By setting $\tau_1 = 0$, if $\tau_n \ll T_p$ for $n > 1$, then $p(t - iT_p - \tau_n) \approx p(t - iT_p)$, and $r(t)$ simplifies to:

$$\begin{aligned} r(t) &\approx \sum_{n=1}^{L_r} \alpha_n \, e^{-j\phi_n} \sum_i A_i e^{j\theta_i} p(t - iT_p) \\ &= \alpha \, e^{-j\phi} s(t) \end{aligned} \tag{2.34}$$

The corresponding channel impulse response is:

$$c(t;\tau) = \alpha(t) \, e^{-j\phi(t)} \, \delta(\tau) \tag{2.35}$$

This is the standard form of a classical *flat-fading* channel response. The amplitude response of the channel is flat over the entire transmit bandwidth. The multipath components are *unresolvable*, and the *effective* channel gain $\tau$ and phase-shift $\phi$ are the sum of i.i.d. fading RVs, respectively. For a WSSUS channel, its scattering function (see eqn. eqn:scattering function) for a flat-fading model is:

$$\begin{aligned} \mathcal{S}(\Delta f, \Delta\tau) &= \mathcal{F}_{\Delta t}\Big[R_c(\Delta t; \Delta\tau)\Big] \\ &= \delta(\Delta\tau)\psi(f) \end{aligned}$$

where $\psi(f)$ is the Doppler spread function.

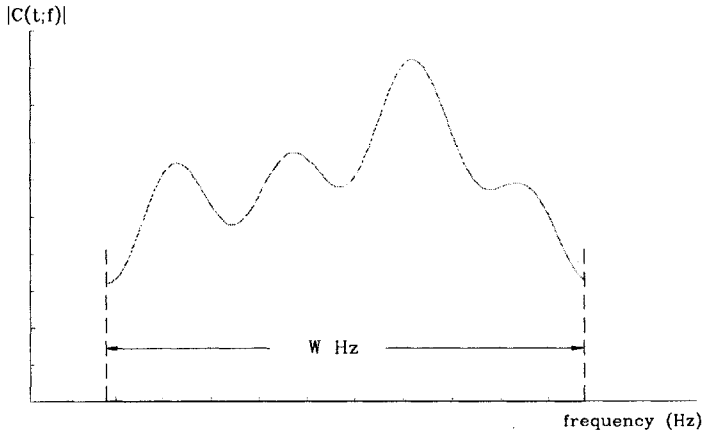



**Figure** 2.14: Frequency selective amplitude response for $L = 4$ resolvable paths

### Resolvable Multipaths

If the difference in arrival times of any two multipath signals is larger than $T_p$ such that $\tau_{n+1} - \tau_n > T_p$, the received signal is:

$$r(t) = \sum_i A_i e^{j\theta_i} \sum_{n=1}^{L} \alpha_n \, e^{-j\,\phi_n} \, p\Big[t - (i + \Delta_n)T_p - \epsilon_n\Big] \tag{2.36}$$

where

$$\Delta_n = \left\lfloor \frac{\tau_n}{T_p} \right\rfloor \quad \text{and} \quad \epsilon_n = \tau_n - \Delta_n$$

In this case, $p(t)$ is highly distorted by its attenuated and phase-shifted echoes. This is similar to reception through an ISI channel, except that in the fading case, the gains and phase-shifts are random and time-varying. The relation between delay spread and L is:

$$L = \left\lfloor \frac{(\Delta\tau)_d}{T_p} \right\rfloor + 1 \tag{2.37}$$

For simplicity, let $\epsilon_n = 0$ for $\forall n$. Then $\Delta_n = n - 1$. Rewriting eqn. (2.36):

$$\begin{aligned} r(t) &= \sum_i A_i e^{j\theta_i} \sum_{n=1}^{L} \Big(\alpha_n \, e^{-j\,\phi_n}\Big) \, p\Big[t - (i + n - 1)T_p\Big] \\ &= \sum_{n=0}^{L-1} \Big[\alpha_{n+1} \, e^{j\phi_{n+1}}\Big] s(t - nT_p) \end{aligned}$$

The L paths are *resolvable* because each replica of the transmit signal $s(t)$ is affected by an independent fading statistic $\alpha_n e^{j\phi_n}$. The term arises from statistic detection theory where it is known that the signal combining gain is maximum when all statistical components of the signal are mutually independent. The equivalent baseband impulse response for this channel is:

$$\begin{aligned} c(t;\tau) &= \sum_{n=1}^{L(t)} \alpha_n(t)\, e^{-j\,\phi_n(t)}\, \delta\Big[\tau - (n-1)T_p\Big] \\ &= \alpha_1\, e^{-j\,\phi_1}\, \delta(\tau) + \alpha_2\, e^{-j\,\phi_2}\, \delta(\tau - T_p) \ldots + \alpha_L\, e^{-j\,\phi_L}\, \delta\big(\tau - (L-1)T_p\big) \end{aligned} \tag{2.38}$$

Its frequency response for $|f| \leq W$ is:

$$\begin{aligned} C(t;f) &= \mathcal{F}\Big\{c(t;\tau)\Big\} \\ &= \mathcal{F}\Big\{\alpha_1\, e^{-j\,\phi_1}\, \delta(\tau) + \alpha_2\, e^{-j\,\phi_2}\, \delta(\tau - T_p) \ldots + \alpha_L\, e^{-j\,\phi_L}\, \delta\big(\tau - (L-1)T_p\big)\Big\} \\ &= \alpha_1\, e^{-j\,\phi_1} + \alpha_2\, e^{-j(2\pi f T_p + \phi_2)} + \alpha_3\, e^{-j(4\pi f T_p + \phi_3)} \ldots + \alpha_L\, e^{-j(2(L-1)\pi f T_p + \phi_L)} \\ &= \sum_{n=1}^{L(t)} \alpha_n(t)\, \exp\Big[-j\Big\{2\pi(n-1)fT_p + \phi_n(t)\Big\}\Big] \end{aligned} \tag{2.39}$$

It is apparent that the frequency response $C(t;f)$ is a finite series of sinusoids with random amplitudes and phases. A plot of $|C(t;f)|$ is shown in Fig. 2.14 for $L = 4$ with $T_p = 1/W$. Hence this fading of fading is *frequency-selective*.

## Notes and References

For short-term rapid fading, two widely used distributions are Rayleigh and Ricean. Both were first proposed by Steve Rice [115, 116]. The Rayleigh distribution is in honor of Lord Rayleigh who studied scattering radio wave propagation. We show that both distributions can be derived from the Gamma probability density function (pdf). For long-term fading, the mean signal strength is modelled as a Gaussian random variable in logarithmic scale. This model, first proposed by D. O. Reudink [62], is based on curve-fitting empirical data measured around New Jersey and New York City. There are several pdfs that model the combined effect of both short- and long-term fading. The most notable is the "m"-distribution proposed by M. Nakagami [92]. He also proposed other variants, the "n" and "q" distributions. By a change of variables, the "n" and Ricean distributions are the same. The "q" distribution is also known as Hoyt's distribution since it was first proposed by R. S. Hoyt [56]. For combined fading, H. Suzuki [137] a two-tier distribution function where the conditional density of the instantaneous signal amplitude is Rayleigh, conditioned on the mean signal strength that is log-normal distributed. Later, F. Hansen and F. I. Meno [49] coined this compound distribution as "Suzuki." A more generalized compound distribution where the Rayleigh pdf is replaced by the Gamma pdf is analyzed by Ho and Stuber [52]. It is shown that for a fixed mean-squared value, the one-sided Gaussian distribution is the worst type of fading.

# Appendix 2A

# Statistical Distributions for Fading Envelope

## Gamma Function

The gamma function $\Gamma(\alpha)$ is defined as:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \tag{2.40}$$

The above integral is known to exist for $\alpha > 0$ and $\Gamma(\alpha) > 0$. It is straightforward to see that

$$\text{For } \alpha = 1 \qquad \Gamma(1) = \int_0^\infty e^{-y} dy = 1 \tag{2.41}$$

$$\text{For } \alpha = 1/2 \qquad \Gamma(1/2) = \int_0^\infty e^{-y} dy = \sqrt{\pi} \tag{2.42}$$

For $\alpha > 1$, using integration by parts:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \tag{2.43}$$

For example, $\Gamma(3/2) = \frac{1}{2}\Gamma(1/2) = \sqrt{\pi}/2$. For the special case of $\alpha \in \mathbb{Z}_2$,

$$\Gamma(\alpha) = (\alpha - 1)(\alpha - 2)\ldots(3)(2)(1)\Gamma(1) = (\alpha - 1)! \tag{2.44}$$

To generalize $\Gamma(\alpha) = (\alpha - 1)!$ for all $\alpha \in \mathbb{Z}_+$, including $\alpha = 1$, set $\Gamma(1) = 0! = 1$.

## Gamma Probability Distribution

If we set $y = \frac{x}{\beta}$ and $\beta > 0$ in (2.40):

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)\left(\frac{1}{\beta}\right) dx \tag{2.45}$$

$$\text{or} \qquad 1 = \int_0^\infty \frac{x^{\alpha-1}}{\Gamma(\alpha)\,\beta^\alpha} \exp\left(-\frac{x}{\beta}\right) dx \tag{2.46}$$

Since $\alpha > 0$, $\beta > 0$ and $\Gamma(\alpha) > 0$, it is easily verified that for $0 < x < \infty$, the integrand in (2.46) qualifies as a pdf:

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\,\beta^\alpha} \exp\left(-\frac{x}{\beta}\right) U(x) \tag{2.47}$$

The random variable X is said to have a "Gamma" pdf $f(x)$ with two free parameters $\alpha$ and $\beta$. We denote a Gamma RV X by $X \sim \Gamma(\alpha, \beta)$. Several plots of the Gamma pdf for various values of $\alpha$ and $\beta$ are shown in Fig. 2.15. It can be seen that $\alpha$ is a *shaping* parameter since it controls

the amount of skewness of the pdf. For large $\alpha$, the pdf appears more symmetric, i.e. "bell-shaped." In contrast, $\beta$ is the *scaling* or *spreading* parameter since it adjusts the "spread" of the pdf. In Fig. 2.16, the Gamma pdf is plotted again with the ordinate normalized by its peak value. Careful observation reveals that the shaping parameter $\alpha$ both shapes and *translates* the density function. It is easily shown that the mean and variance of RV X are:

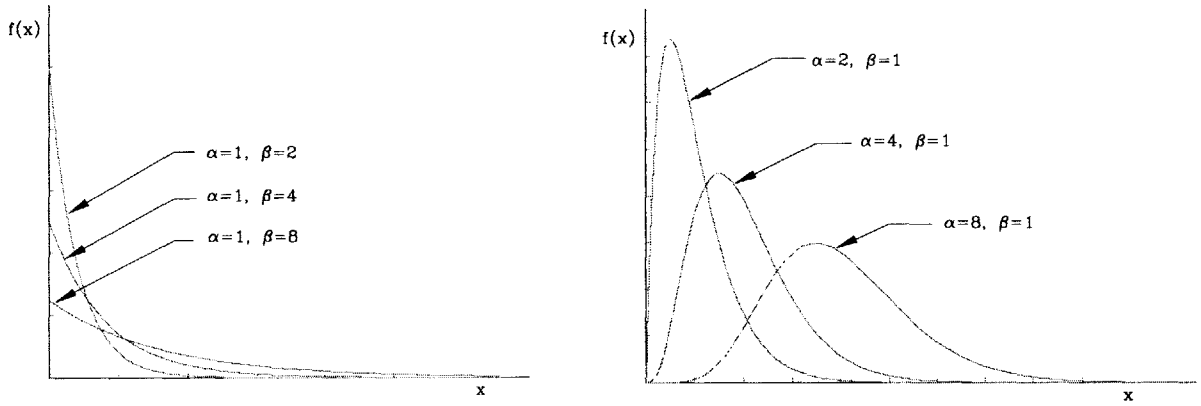



**Figure** 2.15: Gamma probability density function with various values for $\alpha$ and $\beta$.

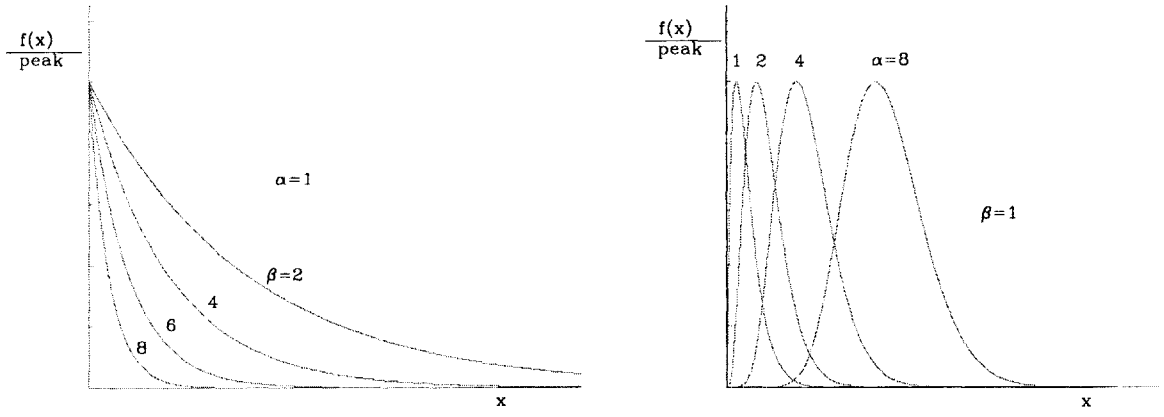



**Figure** 2.16: Peak-normalized Gamma pdf with various values for $\alpha$ and $\beta$.

$$E(X) = \alpha\beta \qquad Var(X) = \alpha\beta^2 \tag{2.48}$$

It is important to note that since the gamma pdf has two free parameters $\alpha$ and $\beta$, there is quite a flexibility in fitting the gamma pdf into one of many empirical multipath fading channel

models. From eqn. (2.48), we also note that due to two free parameters, E(X) and Var(X) are independent; i.e., we cannot deduce one value from the other. Another two-free-parameter pdf that possesses the same property is the Gaussian pdf. However, for the Gaussian RV X, E(X) is the *translation* parameter, and Var(X) is the *scaling* parameter.

## Special Cases of Gamma PDF

### Erlang PDF

For $\alpha \in \mathbb{Z}_+$, $\beta = 1/\lambda$:

$$f(w) = \frac{w^{\alpha-1}e^{-\lambda}\lambda^{\alpha}}{(\alpha-1)!}\,U(w) \tag{2.49}$$

$$E(W) = \alpha\beta = \alpha/\lambda \qquad Var(W) = \alpha\beta^2 = \alpha/\lambda^2 \tag{2.50}$$

The Erlang pdf, commonly applied in tele-traffic modelling, is usually expressed as:

$$f(w) = \frac{(\lambda w)^{\alpha-1}}{(\alpha-1)!}\lambda e^{-\lambda w}\,U(w) \tag{2.51}$$

### Exponential PDF

For the special case of Erlang pdf with $\alpha = 1$, we obtain the exponential pdf:

$$f(w) = \lambda e^{-\lambda w}\,U(x) \tag{2.52}$$

$$E(W) = \beta = 1/\lambda \qquad Var(W) = \beta^2 = 1/\lambda^2 \tag{2.53}$$

Note that unlike the Erlang or Gamma pdfs, the exponential pdf has a single free parameter $\lambda$.

### Standard Chi-Square PDF

For $\alpha = r/2$, $\beta = 2$ where $r \in \mathbb{Z}_+$, the resulting distribution is known as "standard" chi-square pdf:

$$f(x) = \frac{x^{\frac{r}{2}-1}}{\Gamma(\frac{r}{2})\,2^{r/2}}\,\exp\left(-\frac{x}{2}\right)U(x) \tag{2.54}$$

$$E(X) = \alpha\beta = (r/2)\cdot 2 = r \qquad Var(X) = \alpha\beta^2 = (r/2)2^2 = 2r \tag{2.55}$$

The standard chi-square distribution is characterized by a single parameter $r$ known as the *degree of freedom*. The reason for this terminology becomes clear in the next section.

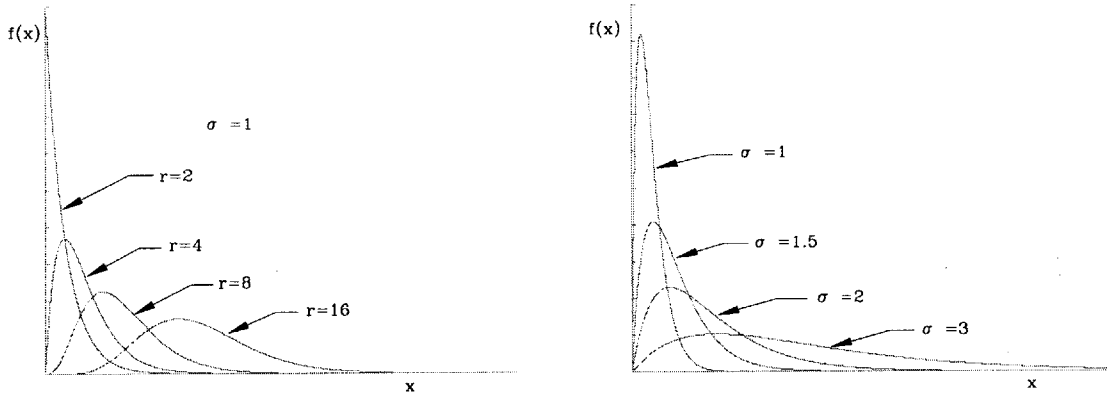
## Relation between Gamma and Other PDFs

### Gaussian PDF

Let $X_i$ be a standard Gaussian RV, i.e. $X_i \sim \mathcal{N}(0,1)$. If $\{X_i\}_{i=1}^{r}$ is a set of i.i.d. Gaussian RVs, then

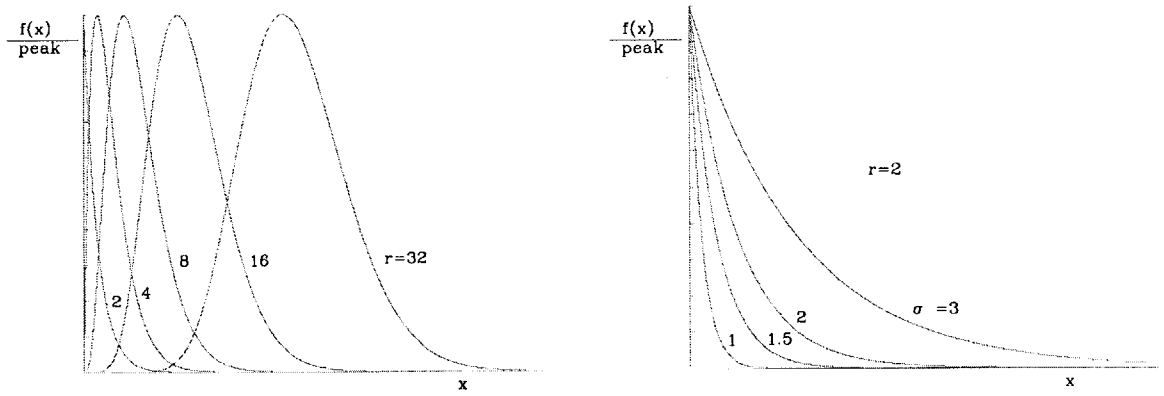
$$Y = \sum_{i=1}^{r} X_i^2$$

has a "standard" chi-square pdf with $r$ degrees of freedom. $E(Y) = r$ and $Var(Y) = 2r$. The number of independent (Gaussian) RVs is a measure of the degrees of freedom for Y.




(a) (b)

**Figure** 2.17: Central chi-square pdf with various values of $r$ and $\sigma$.




(a) (b)

**Figure** 2.18: Peak-normalized central chi-square pdf with various values of $\mu$ and $\sigma$. A larger $r$ translates, and $\sigma$ shapes the density function.

**Central Chi-Square PDF**

If $X_i \sim \mathcal{N}(0, \sigma^2)$ and $\{X_i\}_{i=1}^r$ is a set of i.i.d. Gaussian RVs, then

$$Y = \sum_{i=1}^{r} X_i^2$$

has a "central" chi-square pdf with $r$ degrees of freedom, and a *centrality parameter* $\sigma^2$.

$$E(Y) = r\sigma^2 \qquad \text{Var}(Y) = 2r\sigma^4 \tag{2.56}$$

Its pdf can be written as:

$$f(y) = \frac{y^{\frac{r}{2}-1}}{\Gamma(\frac{r}{2})\,(2\sigma^2)^{r/2}} \exp\Big(-\frac{y}{2\sigma^2}\Big)\, U(y) \tag{2.57}$$

Note that it is a two-parameter distribution function, and $E(Y)$ and $\text{Var}(Y)$ are independent unless $r = 1$.

**Non-Central Chi-Square PDF**

If $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ and $\{X_i\}_{i=1}^r$ is a set of independent Gaussian RVs, then

$$Y = \sum_{i=1}^{r} X_i^2$$

has a "non-central" chi-square pdf with $r$ degrees of freedom, a *centrality parameter* $\sigma^2$ plus a *non-centrality parameter* $\mu^2$:

$$\mu^2 = \sum_{i=1}^{r} \mu_i^2$$

Its pdf can be expressed as:

$$f(y) = \frac{1}{2\sigma^2}\Big(\frac{y}{\mu^2}\Big)^{\frac{(r-2)}{4}} \exp\Big(-\frac{y+\mu^2}{2\sigma^2}\Big)\, I_{(\frac{r}{2}-1)}\Big(\sqrt{y}\cdot\frac{\mu}{\sigma^2}\Big)\, U(y) \tag{2.58}$$

where $I_\alpha(x)$ is an $\alpha^{\text{th}}$-order modified Bessel function of the first kind:

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{\big(\frac{x}{2}\big)^{\alpha+2k}}{k!\,\Gamma(\alpha+k+1)}\, U(x) \tag{2.59}$$

It is easily shown that

$$E(Y) = r\sigma^2 + \mu^2 \qquad \text{Var}(Y) = 2r\sigma^4 + 4\sigma^2\mu^2 \tag{2.60}$$

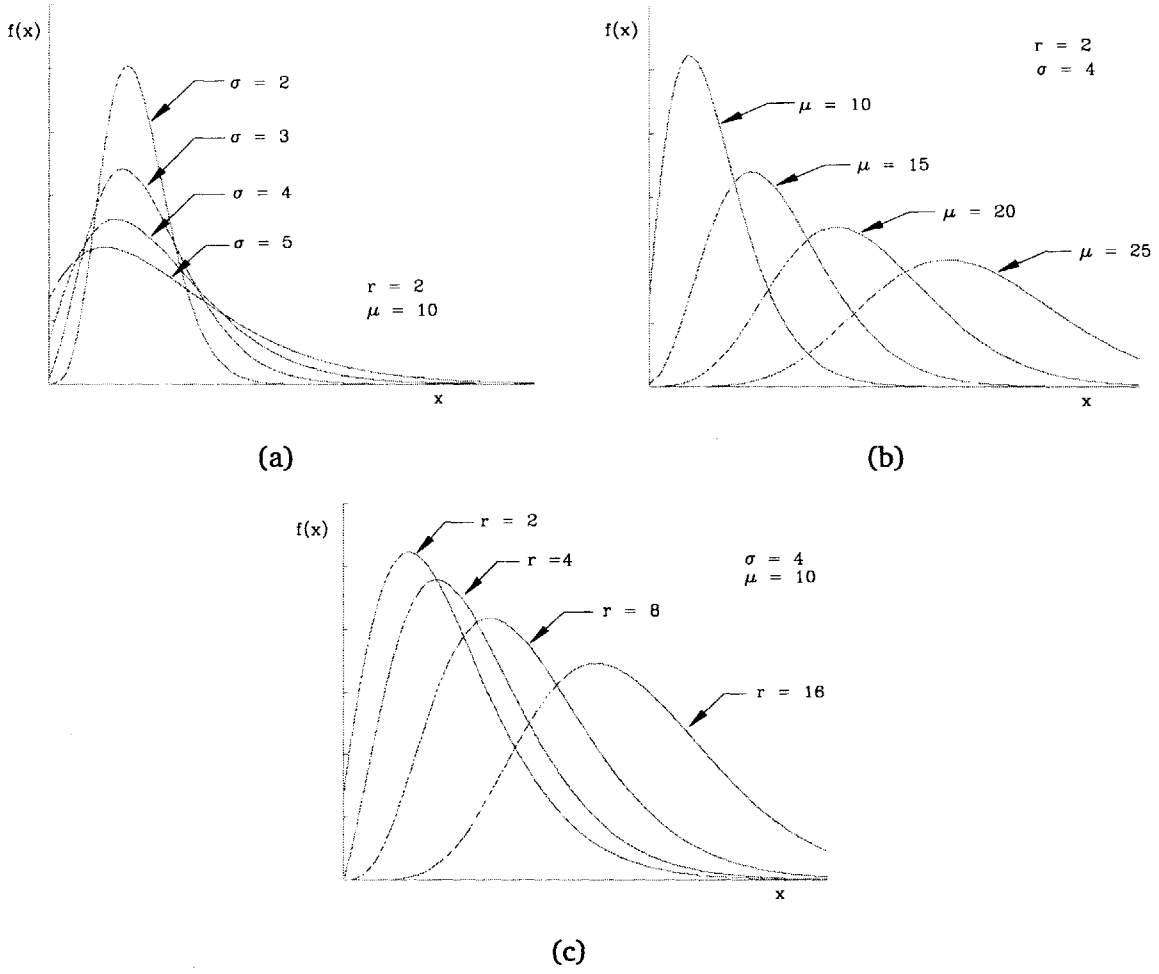



**Figure 2.19:** Non-central chi-square pdf with various values of r, $\mu$ and $\sigma$.

The non-central chi-square pdf is a function of three parameters r, $\sigma^2$ and $\mu^2$. In terms of notation, all three types of chi-square pdfs can be distinguished by the number of parameters in the argument:

$$\text{Standard}: \quad Y \sim \chi^2(r) \tag{2.61}$$

$$\text{Central}: \quad Y \sim \chi^2(r, \sigma^2) \tag{2.62}$$

$$\text{Non-Central}: \quad Y \sim \chi^2(r, \sigma^2, \mu^2) \tag{2.63}$$

### Chi PDF

If Y is a standard, central or non-central chi-square distributed RV, then $R = \sqrt{Y}$ is said to have a standard, central or non-central chi pdf, respectively. The pdf of a central chi distributed

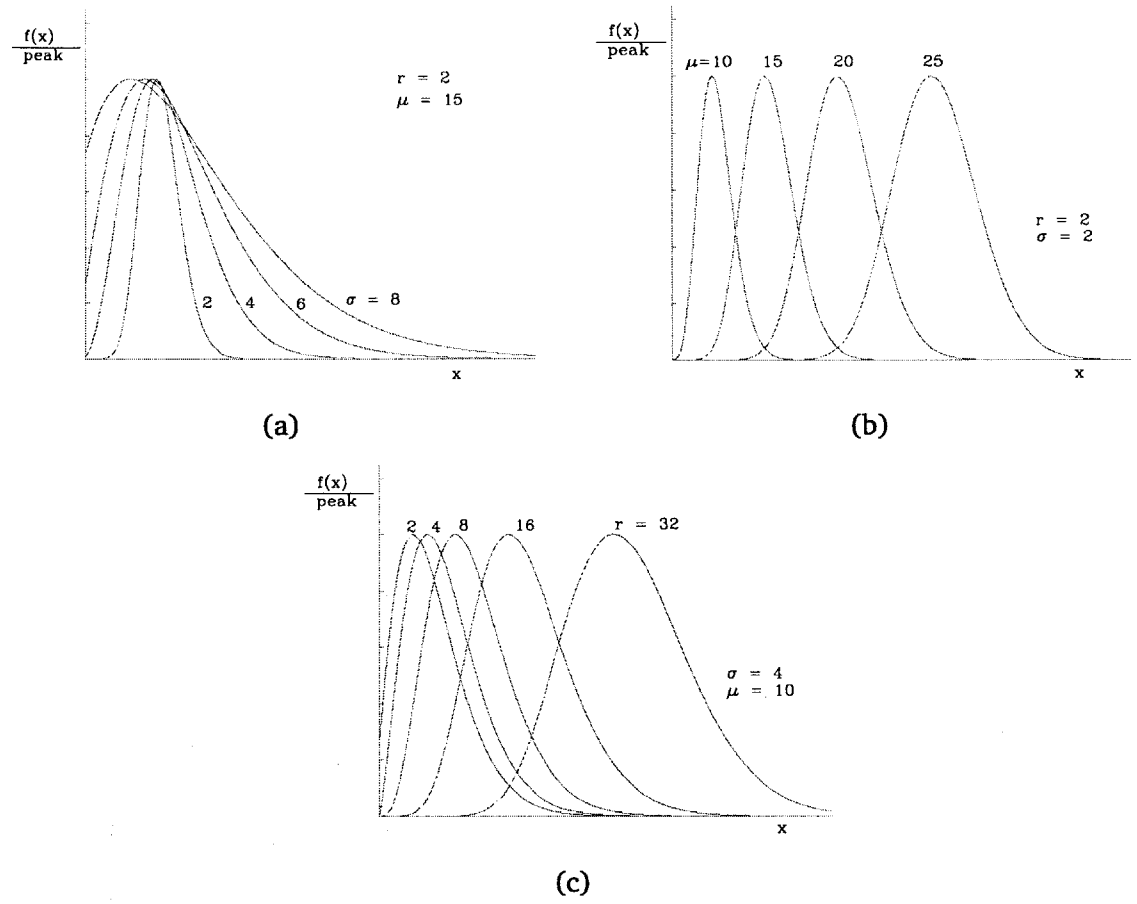



**Figure** 2.20: Peak-normalized non-central chi-square pdf with various values of r, μ and σ. A larger μ translates, and σ shapes the density function.

RV $R \sim \chi(n, \sigma^2)$ with $n$ degrees of freedom is:

$$f(r) = \frac{r^{n-1}}{2^{(n-2)/2}\,\sigma^n\,\Gamma(n/2)}\,\exp\left(-\frac{r^2}{2\sigma^2}\right)U(r) \tag{2.64}$$

Similarly, the pdf of a non-central chi distributed RV $R \sim \chi(n, \sigma^2, \mu^2)$ with $n$ degrees of freedom is:

$$f(r) = \frac{r^{n/2}}{\mu^{(n-2)/2}\,\sigma^2}\,\exp\left(-\frac{r^2+\mu^2}{2\sigma^2}\right)I_{(\frac{n}{2}-1)}\left(\frac{r\mu}{\sigma^2}\right)U(r) \tag{2.65}$$

Similar to the chi-square distributions, all three types of chi pdfs can be distinguished by the number of parameters in the argument:

$$\text{Standard}: \quad R \sim \chi(n) \tag{2.66}$$

$$\text{Central}: \quad R \sim \chi(n,\ \sigma^2) \tag{2.67}$$

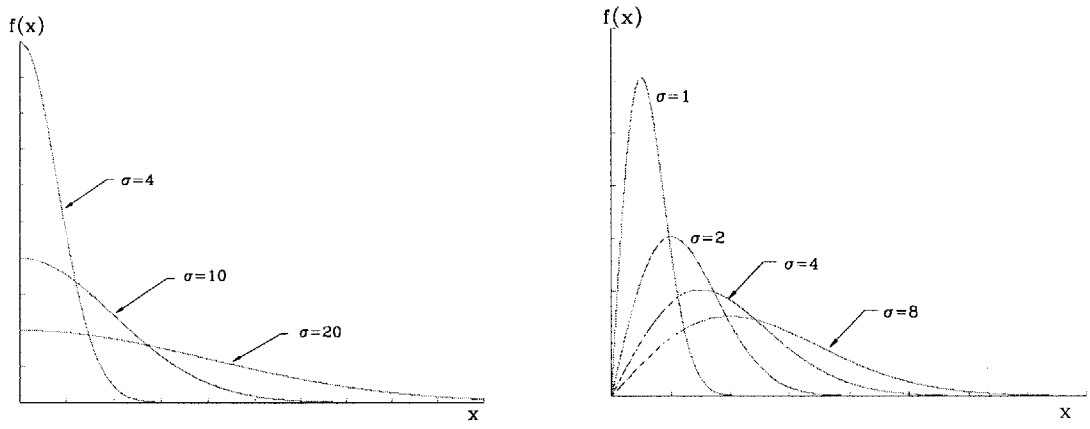
$$\text{Non-Central}: \quad R \sim \chi(n,\ \sigma^2, \mu^2) \tag{2.68}$$

### Special Cases of Chi Density

If $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ are independent Gaussian RVs, then

$$R = \sqrt{X_1^2 + X_2^2 + \cdots + X_n^2}$$

is a chi-distributed RV with $n$ degrees of freedom. Depending on the values of $\mu_i$ and $\sigma^2$, it is either standard, central or non-central type. We are particularly interested in the following cases:




(a) (b)

**Figure 2.21:** (a) One-sided Gaussian and (b) Rayleigh distributions. Both are chi $\chi$ densities with one and two dimensions, respectively.

### One-sided Gaussian PDF

If $Y \sim \chi^2(1, \sigma^2)$, or equivalently if $X \sim \mathcal{N}(0, \sigma^2)$, then $R = \sqrt{X^2}$ has a one-sided Gaussian pdf denoted by $R \sim \chi(1, \sigma^2)$:

$$f(r) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{r^2}{2\sigma^2}\right) U(r) \tag{2.69}$$

Unlike standard Gaussian pdf, the shape of the pdf of one-sided Gaussian RV is not symmetric and the mean is non-zero. It can be shown that

$$E(R) = \sqrt{\frac{2}{\pi}}\sigma \qquad E(R^2) = \sigma^2 \qquad Var(R) = \left(1 - \frac{2}{\pi}\right)\sigma^2$$

Several pdf plots of a one-sided Gaussian RV, parameterized by $\sigma$, are shown in Fig. 2.21(a). Since its mode value is always zero, this pdf represents the worst type of fading for fixed mean and variance.

**Rayleigh PDF**

If $Y \sim \chi^2(2, \sigma^2)$, or equivalently if $X_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, 2$ then $R = \sqrt{X_1^2 + X_2^2}$ has a Rayleigh pdf denoted by $R \sim \chi(2, \sigma^2)$:

$$f(r) = \left(\frac{r}{\sigma^2}\right) \exp\left(-\frac{r^2}{2\sigma^2}\right) U(r) \tag{2.70}$$

This is the conventional form of Rayleigh pdf when derived from Gaussian RVs. Several different pdf plots of a Rayleigh RV, parameterized by $\sigma$, are shown in Fig. 2.21(b). A more commonly used pdf form in mobile radio engineering is expressed in terms of the mean-squared value $\Omega = E(R^2)$:

$$f(r) = \left(\frac{2r}{\Omega}\right) \exp\left(-\frac{r^2}{\Omega}\right) U(r) \tag{2.71}$$

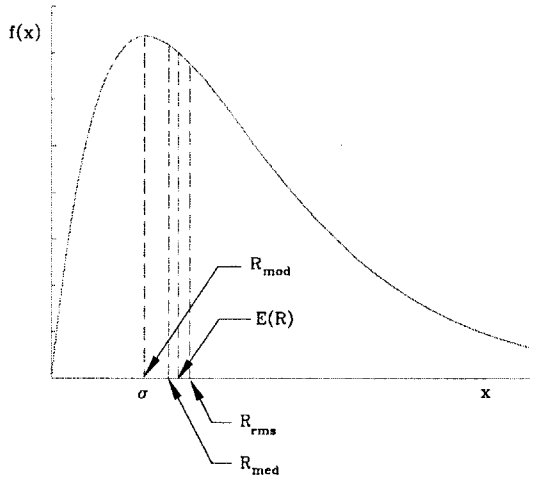



**Figure 2.22:** Rayleigh probability density function. $R_{mod} = \sigma$, $E(R) = \sqrt{\frac{\pi}{2}}\,\sigma$, $R_{med} = \sqrt{2\ln 2}\,\sigma$, $R_{rms} = \sqrt{\frac{4-\pi}{2}}\,\sigma$

We can deduce the following:

$$\text{Moment}: \quad E(R^n) = 2^{n/2}\,\Gamma\left(1 + \frac{n}{2}\right) \tag{2.72}$$

$$\text{Mean}: \quad E(R) = \sqrt{\frac{\pi}{2}}\,\sigma \tag{2.73}$$

$$\text{Mean-Square}: \quad E(R^2) = 2\,\sigma^2 \tag{2.74}$$

$$\text{Variance}: \quad \text{Var}(R) = \left(\frac{4-\pi}{2}\right)\sigma^2 \tag{2.75}$$

$$\text{Median}: \quad R_{med} = \sqrt{2\ln 2}\,\sigma \tag{2.76}$$

$$\text{Mode}: \quad R_{mod} = \sigma \tag{2.77}$$

The median $R_{med}$ corresponds to the point where the area of the pdf to its left and right are equal:

$$\int_0^{R_{med}} f(r)\, dr = \frac{1}{2}$$

The mode $R_{mod}$ is the most probable value:

$$R_{mod} = \arg\max \Big[ f(r) \Big]$$

**Rice PDF**

If $Y \sim \chi^2(2, \sigma^2, \mu^2)$, or equivalently if $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ and $X_2 \sim \mathcal{N}(0, \sigma^2)$, then $R = \sqrt{X_1^2 + X_2^2}$ has a Rice pdf denoted by $R \sim \chi(2, \sigma^2, \mu^2)$:

$$f(r) = \left(\frac{r}{\sigma^2}\right) \exp\left( - \frac{r^2 + \mu^2}{2\sigma^2} \right) I_0\left(\frac{r\mu}{\sigma^2}\right) U(r) \tag{2.78}$$

The more common form in mobile radio engineering is:

$$f(r) = \left(\frac{2r}{S}\right) \exp\left( - \frac{r^2 + \mu^2}{S} \right) I_0\left(\frac{2r\mu}{S}\right) U(r) \tag{2.79}$$

where $S = 2\sigma^2$. The moments and the peak value of a Ricean RV R are:

$$\text{Moment:} \quad E(R^n) = S^{n/2}\, \Gamma\left(1 + \frac{n}{2}\right) F(n) \tag{2.80}$$

$$\text{Mean:} \quad E(R) = \sqrt{\frac{\pi}{2}}\, \sigma\, F(1) \tag{2.81}$$

$$\text{Mean-Square:} \quad E(R^2) = S + \mu^2 \tag{2.82}$$

$$\text{Variance:} \quad \text{Var}(R) = \mu^2 + S\left(1 - \frac{\pi}{4} F^2(1)\right) \tag{2.83}$$

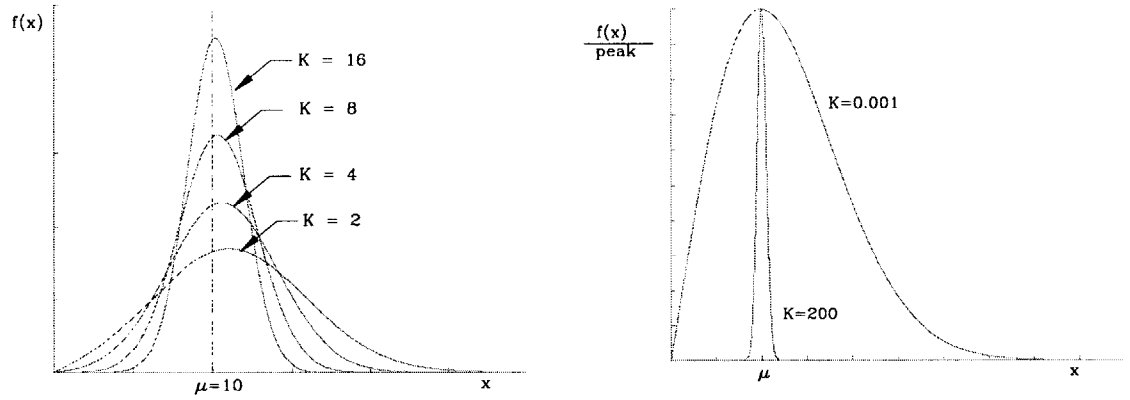
where ${}_1F_1(x, y; z) = \Phi(x, y; z)$ is the confluent hypergeometric function [47], pp. 1012, and

$$F(n) = {}_1F_1\left( - \frac{n}{2}, 1; -\frac{\mu^2}{S} \right)$$

Although the Rice pdf has two independent parameters, the severity of Ricean fade is measured in terms of a single fading parameter K:

$$K = \frac{\mu^2}{2\sigma^2} = \frac{\mu^2}{S} \tag{2.84}$$

K is the energy (or power) ratio of the line-of-sight (specular) path and non-line-of-sight paths. When the specular path is strong, K is large. As a special case, when $K = 0$, the Rice density reduces to the Rayleigh pdf. For $K \gg 1$, it approaches the Gaussian density with mean $\mu$. As $K \to \infty$, $\sigma^2 \to 0$, and the distribution approaches a delta function, i.e., the non-fading case.




(a) (b)

**Figure** 2.23: Rice probability density function with parameter K ($\mu = 10$)

## Other Important Fading Distributions

### Nakagami-n PDF

For a Nakagami-n pdf, n is the fading parameter, $0 \leq n < \infty$.

$$f(r) = \frac{2r(1+n^2)\,e^{-n^2}}{\Omega} \exp\left(-\frac{(1+n^2)r^2}{\Omega}\right) I_o\left(2nr\sqrt{\frac{1+n^2}{\Omega}}\right) U(r) \tag{2.85}$$

It can be shown that by using the following substitutions:

$$n = \frac{\mu}{\sqrt{2}\,\sigma} \qquad 2\sigma^2 = \Omega - \mu^2$$

and after some manipulations, the Nakagami-n and Rice pdfs are equivalent with $K = n^2$. As a special case, when $K = n^2 = 0$, it equals the Rayleigh pdf. As $n \to \infty$, it approaches a delta function, i.e., no fading. The moments and the peak value of a Nakagami-n RV R are:

$$\text{Moment}: \quad E(R^n) = \frac{\Gamma\left(m+\frac{n}{2}\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{n/2} \tag{2.86}$$

$$\text{Mean}: \quad E(R) = \frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\sqrt{\frac{\Omega}{m}} \tag{2.87}$$

$$\text{Mean-Square}: \quad E(R^2) = \Omega \tag{2.88}$$

$$\text{Variance}: \quad \text{Var}(R) = \Omega\left[1 - \left(\frac{\Gamma(m+\frac{1}{2})}{\sqrt{m}\,\Gamma(m)}\right)^2\right] \tag{2.89}$$

$$\text{Mode}: \quad R_{\text{mod}} = \sqrt{\Omega} \tag{2.90}$$

$$\text{Peak Value}: \quad f(R_{\text{mod}}) = \frac{2m^m}{\Gamma(m)\sqrt{\Omega}} \tag{2.91}$$

**Nakagami-m PDF**

The m-distribution, first proposed by Nakagami has the following pdf:

$$f(r) = \frac{2m^m\, r^{2m-1}}{\Omega^m\, \Gamma(m)} \exp\left(-\frac{mr^2}{\Omega}\right) U(r) \tag{2.92}$$

We denote a Nakagami-m distributed RV as $R \sim m(m, \Omega)$. It is a two-parameter distribution with

$$\Omega = E(R^2) \qquad m = \frac{\Omega^2}{(R^2 - \Omega)^2} \geq \frac{1}{2} \tag{2.93}$$

The parameter m can be considered as the inverse of the normalized variance of $R^2$. As we will see in Sec. 2.3.2, the normalized variance of a random variable is used as the single unified benchmark to compare the severity of various fading distributions. The moments and the peak value of a Nakagami-m RV R are:

$$\text{Moment}: \quad E(R^n) = \frac{\Gamma\left(m + \frac{n}{2}\right)}{\Gamma(m)} \left(\frac{\Omega}{m}\right)^{n/2} \tag{2.94}$$

$$\text{Mean}: \quad E(R) = \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma(m)} \sqrt{\frac{\Omega}{m}} \tag{2.95}$$

$$\text{Mean-Square}: \quad E(R^2) = \Omega \tag{2.96}$$

$$\text{Variance}: \quad \text{Var}(R) = \Omega\left[1 - \left(\frac{\Gamma(m + \frac{1}{2})}{\sqrt{m}\,\Gamma(m)}\right)^2\right] \tag{2.97}$$

$$\text{Mode}: \quad R_{\text{mod}} = \sqrt{\Omega} \tag{2.98}$$

$$\text{Peak Value}: \quad f(R_{\text{mod}}) = \frac{2m^m}{\Gamma(m)\sqrt{\Omega}} \tag{2.99}$$

It can be shown by using the following transformations:

$$n = 2m \qquad 2m\sigma^2 = \Omega$$

that the Nakagami-m (eqn. (2.92)) and the central chi density with n degrees of freedom (eqn. (2.64)) are equivalent. However, note the differences: the central chi pdf $\chi(n, \sigma^2)$ is defined only for $n \in \mathbb{Z}_+$, whereas the Nakagami fading parameter is valid for $\frac{1}{2} \leq m < \infty$. As special cases, $m = \frac{1}{2}$ gives the one-sided Gaussian pdf, and $m = 1$ results in Rayleigh pdf. As $m \to \infty$, it approaches the non-fading delta function.

**Nakagami-q PDF**

The Nakagami-q pdf, also known as the Hoyt distribution has the following pdf:

$$f(r) = \frac{(1 + q^2)\, r}{q\, \Omega} \exp\left(-\frac{(1 + q^2)^2\, r^2}{4\, q^2\, \Omega}\right) I_0\left(\frac{(1 - q^4)\, r^2}{4\, q^2\, \Omega}\right) U(r) \tag{2.100}$$

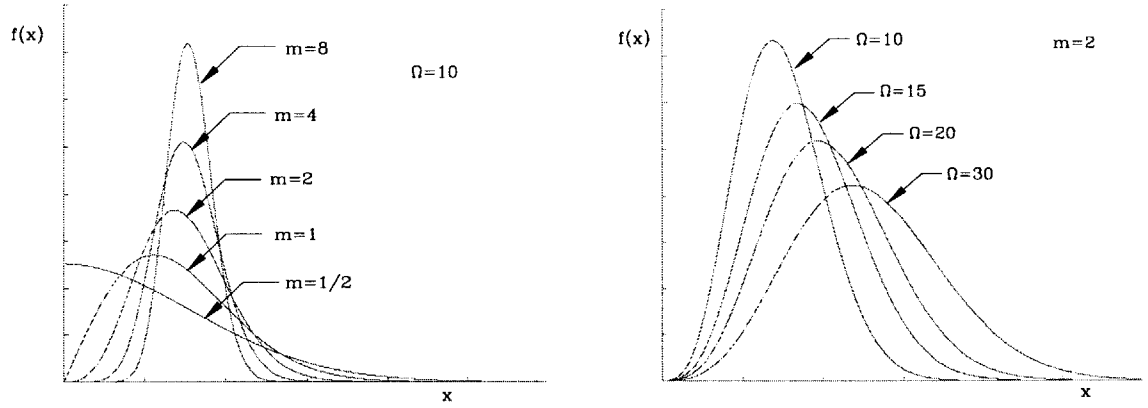



**Figure** 2.24: Nakagami-m pdf with parameters m and Ω

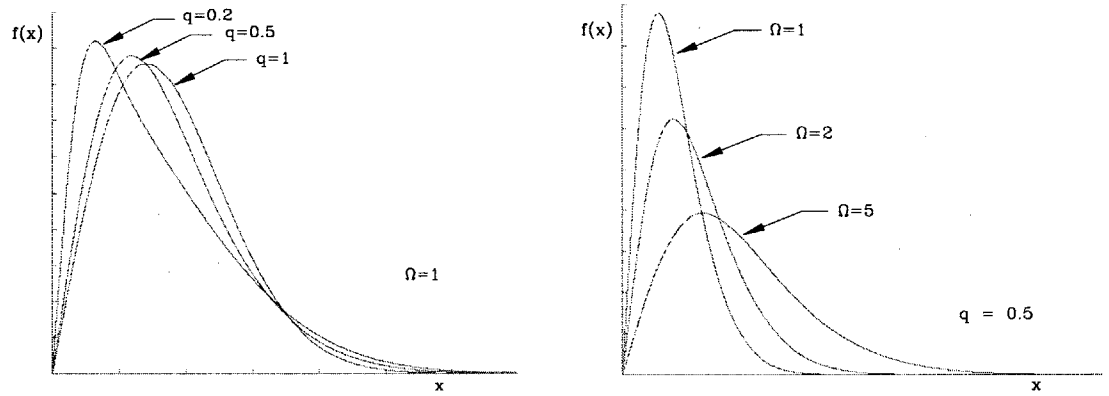



**Figure** 2.25: Nakagami-q pdf with parameters q and Ω

The fading parameter is $0 \leq q \leq 1$. The original "symmetric" form proposed by Nakagami (ironically without the parameter q) is:

$$f(r) = \frac{2r}{\sqrt{\alpha\beta}} \exp\left[-\frac{r^2}{2}\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)\right] I_0\left[\frac{r^2}{2}\left(\frac{1}{\beta}-\frac{1}{\alpha}\right)\right] U(r) \qquad (2.101)$$

It can be shown that by applying the following transformations:

$$\beta = \alpha q^2 \qquad \alpha = \frac{2\Omega}{1+q^2}$$

eqns. (2.100) and (2.101) are equivalent. It is a function of two free parameters. A q-distributed RV is denoted by $R \sim q(q, \Omega)$. Because of the $r^2$ term in the argument of the Bessel function, the q density cannot be expressed as non-central chi-square or chi density. As

special cases, for $q = 0$ it becomes the one-sided Gaussian pdf, and for $q = 1$ it is equal to the Rayleigh pdf.

**Log-Normal PDF**

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = e^X$, then Y has a log-normal distribution with pdf:

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) U(y) \tag{2.102}$$

$$E(Y) = \exp\left(\frac{4\mu + 1}{8\sigma^2}\right) \qquad E(Y^2) = \exp\left(\frac{2\mu + 1}{2\sigma^2}\right)$$

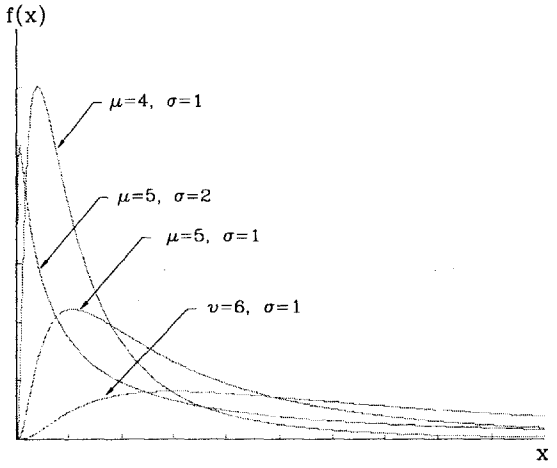



**Figure** 2.26: Log-normal distribution

**Suzuki PDF**

If a RV S has a conditional Rayleigh density, conditioned on its mode (or mean) that has a log-normal distribution, then S has a Suzuki density. Its unconditional density is:

$$f(s) = f(s|\sigma) \frac{M}{\sigma\sqrt{2\pi\alpha^2}} \exp\left[-\frac{(\ln\sigma - \mu)^2}{2\alpha^2}\right] d\sigma \tag{2.103}$$

where

$$f(s|\sigma) = \int_0^\infty \left(\frac{s}{\sigma^2}\right) \exp\left(-\frac{s^2}{2\sigma^2}\right)$$

The conditional Rayleigh pdf $f(s|\sigma)$ can be generalized to a Gamma or chi pdf.

# Appendix 2B

# Derived Distributions

### Powers of a Random Variable

Note a peculiar property of a log-normal RV: If X is a normal (Gaussian) RV, i.e., $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$Y = e^X \tag{2.104}$$

is a log-normal RV. Similarly, any power of Y:

$$Y^n = e^{nX} \tag{2.105}$$

also has a log-normal pdf since scaling a normal RV is still a normal RV. In practice, we are interested in $n \in \mathbb{Z}_+$. In particular, if the amplitude of a signal $V = e^X$ is log-normal distributed, its value $\mathcal{V}$ in decibel scale,

$$\mathcal{V}\,[\text{dB}] = 20\log(V) = 46\,X \tag{2.106}$$

is Gaussian distributed. Similarly, its power $P = V^2$ (in watts) and $\mathcal{P}$ [dB] are log-normal and Gaussian RVs, respectively. Since signal power is the proportional to the square of its envelope, we are mostly interested in the pdfs of envelopes used in short-term fading. Similar to Table 2.1, we list the density functions of the power for six commonly used short-term fading distributions.

**Table** 2.2: Relation between Power and Envelope Fading PDFs

| X | $Y = X^2$ | Comments |
|---|---|---|
| Rayleigh | central chi-square | two-dimensional<br>equivalent to exponential function |
| Rice | non-central chi-square | two-dimensional |
| Nakagami-q | no standard form | |
| Nakagami-m | central chi-square | 2m degrees of freedom |
| Log-normal | Log-normal | Here, both RVs are in log scale unit |
| One-sided Gaussian | central chi-square | one-dimensional |

### Product of Random Variables

If $\alpha_n$, $n \in \mathbb{Z}^N$, are mutually independent RVs, their product

$$\alpha = \prod_{n=1}^{N} \alpha_n \tag{2.107}$$

when expressed in decibel scale:

$$\alpha\,[\mathrm{dB}] = \sum_{n=1}^{N} \alpha_n \tag{2.108}$$

approaches the Gaussian distribution for large N, by invoking the Central Limit Theorem (CLT) [34]. This implies that $\alpha$ is (approximately) log-normal distributed.

### Sum of Random Variables

We already know that if $\{X_i\}_{i=1}^{r}$ is a set of independent Gaussian RVs, then

$$Y = \sum_{i=1}^{r} X_i^2$$

is a chi-square RV with r degrees of freedom. If $\{Y_i\}_{i=1}^{r}$ is a set of independent chi-square RVs, where each RV $Y_i$ has $n_i$ degrees of freedom, then

$$N = \sum_{i=1}^{r} Y_i^2 \tag{2.109}$$

is chi-square distributed with $n = (n_1 + n_2 + \ldots + n_r)$ degrees of freedom. In an unresolvable multipath fading model, the received signal

$$r(t) = \sum_{i=1}^{r} \alpha_i\, s(t) \tag{2.110}$$

is the sum of attenuated replicas of the transmitted signal $s(t)$. The attenuation factors $\alpha_i$ are complex RVs. The instantaneous received signal energy at time t is:

$$\begin{aligned} E_r &= E\Big(r(t)r^*(t)\Big) \\ &= E_s \sum_{i=1}^{r} \sum_{j=1}^{r} E\big(\alpha_i\, \alpha_j^*\big) \end{aligned} \tag{2.111}$$

If the RVs $\alpha_i$ are zero-mean,

$$E_r = E_s \sum_{i=1}^{r} E\Big(|\alpha_i|^2\Big) \tag{2.112}$$

If the RVs $\alpha_i$ are zero-mean and Gaussian, each $|\alpha_i|$ is a one-sided Gaussian RV. Equivalently, their sum

$$A = \sum_{i=1}^{r} \left(|\alpha_i|^2\right)$$

is chi-square distributed with $r$ degrees of freedom. Thus, the received signal energy $E_r$ is the product of the mean of a chi-square RV $A$ and the transmit signal energy $E_s$. If there are $L$ resolvable multiple fading paths, the received signal energy becomes

$$E_r = E_s \cdot \sum_{l=1}^{L} E(A_l) \tag{2.113}$$

where

$$A_l = \sum_{i=1}^{r_l} \left(|\alpha_{il}|^2\right)$$

is a chi-square distributed RV with $r_l$ unresolvable multipaths. $\{|\alpha_{il}|\}$ is the set of independent one-sided Gaussian RVs. Thus,

$$A = A_1 + A_2 + \ldots + A_L$$

is a chi-square RV with $\sum_{l=1}^{L} r_l$ degrees of freedom.

**Linear Function of a Random Variable**

When dealing with binary-valued sequences, the two possible states are $\{0, 1\} = \mathbb{F}_2$ or $\{+1, -1\} = \mathbb{D}_2$. Let $X \in \mathbb{F}_2$ be a Bernoulli RV, i.e. $X \sim B(p)$, with $P(X = 0)$=p, $P(X = 1)$=1 − p. Then,

$$Y = 2X - 1$$

is said to have a *bipolar* Bernoulli distribution, denoted $Y \sim bB(p)$. Note that $(Y) \in \mathbb{D}_2$ with

$$P(Y = -1) = p \qquad \text{and} \qquad P(Y = 1) = 1 - p$$

$Y$ is a *level-shifted* version of $X$ (see Fig. 4.11). Similarly, if $U$ is a binomial RV, $U \sim \text{Bin}(n, p)$ with probability mass function (pmf):

$$P(U = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

its corresponding *bipolar* Binomial RV $V \sim \text{bBin}(n, p)$, where $V = 2U - n$, has the following pmf:

$$P(V = m) = \binom{n}{k'} p^{k'} (1-p)^{n-k'} \qquad k' = \frac{m+n}{2} \tag{2.114}$$

Both conventional and bipolar Bernoulli and Binomial functions are plotted in Fig. 2.27. We denote $n$ as its *length* and $k$ as its *outcome*.

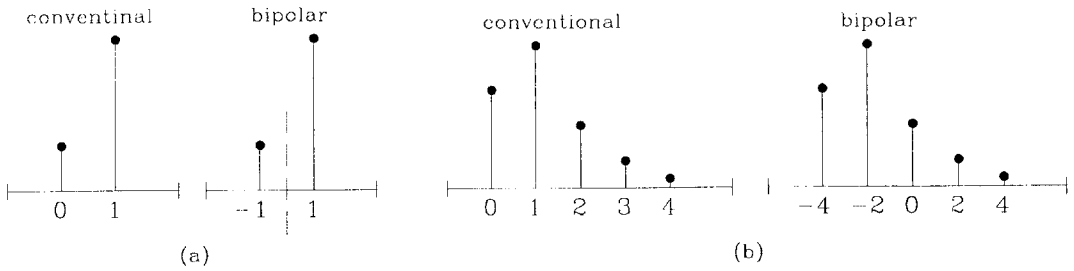



**Figure** 2.27: Conventional and bipolar functions: (a) Bernoulli (b) Binomial

A Bernoulli RV is said to be *symmetric* if its elements are equally likely, i.e.,

$$P(X = 0) = P(X = 1) = 1/2$$

A Binomial RV is symmetric if $p = 1/2$. The pmf of a symmetric bipolar Binomial RV simplifies to:

$$P(V = m) = \begin{cases} \frac{1}{2^n}\binom{n}{\frac{m+n}{2}} & n, m \quad \text{both odd or even} \\ 0 & \text{otherwise} \end{cases} \tag{2.115}$$

Symmetric bipolar Bernoulli and Binomial functions are plotted in Fig. 2.28

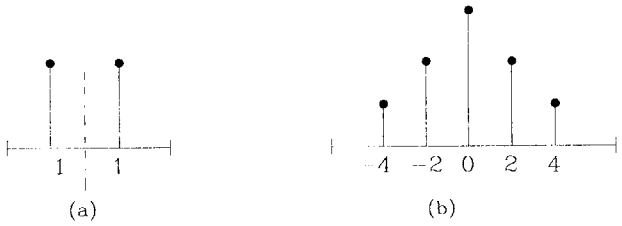



**Figure** 2.28: Symmetric bipolar density functions: (a) Bernoulli (b) Binomial

Note that for a conventional symmetric Binomial RV, if it is a sum of $n$ independent Bernoulli RVs, its length is also $n$. If $n$ is even, it has a single mode. On the contrary, when $n$ is odd there are two modes. In both cases the integer support is $k \in [0, n]$. For a bipolar symmetric Binomial RV, if it is a sum of $n$ independent bipolar Bernoulli RVs, the *absolute* value of its length is $n$. When $n$ is even, it has a single mode equal to zero. Its support consists of even integers only; i.e., $m \in \pm 2\mathbb{N}^{n/2}$. When $n$ is odd, it has two modes: $-1$ and $+1$. Its support consists of odd integers only; i.e., $m \in \mathbb{D}_{n+1}$. All four cases are illustrated in Fig. 2.29 for $n = 4$ and 5.

We already know that a Binomial RV is a sum of iid Bernoulli RVs. Therefore, the sum of a Binomial RV of length $n$ and a mutually independent (equally distributed) Bernoulli RV is another Binomial RV of length $(n+1)$. We are now interested in the distribution of the resulting

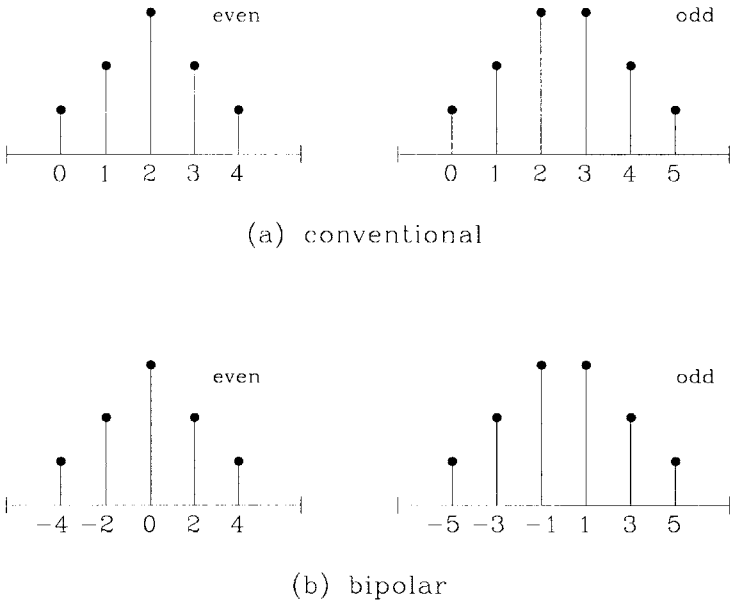



**Figure** 2.29: Symmetric Binomial PDFs of odd and even lengths

RV if the Bernoulli RV is *not* independent. Let $V$ be a symmetric, bipolar Binomial RV

$$V = X_0 + X_1 + \ldots + X_{N-2}$$

where $\{X_i\}$ is a set of iid symmetric, bipolar Bernoulli RVs. Next, let $W = V + X_0$. Also let

$$V^* = X_1 + \ldots + X_{N-2}$$

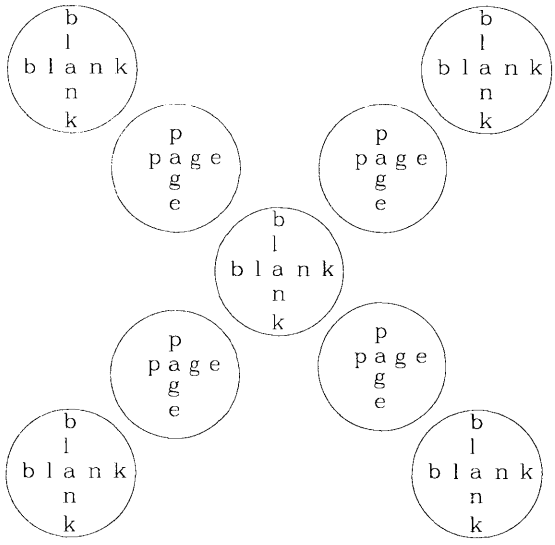
be a Binomial RV of length $(N-2)$. Then

$$\begin{aligned} P(W = m) &= P\Big[\Big(V^* = m-2 \text{ and } X_0 = 1\Big) \text{ or } \Big(V^* = m+2 \text{ and } X_0 = -1\Big)\Big] \\ &= p\,P\Big(V^* = m-2\Big) + (1-p)\,P\Big(V^* = m+2\Big) \end{aligned} \tag{2.116}$$

Since the RVs are symmetric

$$P(W = m) = \frac{1}{2}\Big[P\Big(V^* = m-2\Big) + P\Big(V^* = m+2\Big)\Big]$$

Thus, it is the sum of two Binomial RVs, both of length $(N-2)$ weighed by $1/2$. They, however, are not independent.


blank
page
blank
page
blank
page
blank
page
blank
blank

# 3

# CELLULAR BROADCAST CHANNEL

## Summary

We present some fundamental limits in terms of simultaneously achievable information rates from a single transmitter to many receivers. When there is no interference, the channel model under study is the broadcast channel. Its capacity and achievable rate regions are reviewed. When interference from other transmitters is taken into account, the transmission medium is a composite channel consisting of both broadcast and interference components. When the locations of transmitters is defined according to a certain regular geometrical structure, the resulting composite channel is called a cellular broadcast channel. Achievable rate regions of intra-cell superposition coding and three orthogonal multiplexing schemes based on time-, frequency- and code-division in a cellular broadcast channel are analyzed. Channel models without any fading, with flat and frequency-selective fading are analyzed. (The necessary background material for fading phenomenon and its representative channel models are detailed in Chapter 2.) A tutorial on hexagonal cellular geometry and some of its unique properties are presented in Appendix 3A. The information therein aids in the modelling of a cellular broadcast channel.

The main highlights of this chapter are:

- The optimal data multiplexing scheme in a broadcast model requires the use of superposition coding at the transmitter and a demultiplexing scheme based on successive interference cancellation at all receivers. This type of receiver requires "double feedback" side information on a dynamic basis. In practice, such realization may not be possible or prohibitively complex.

- Various sub-optimal orthogonal multiplexing schemes are equivalent in terms of their achievable rate regions. In a multiplexing scheme based on orthogonal codes, it may be necessary to generate a much larger number of codes than the number of receivers to support arbitrary rates.
- The most efficient protocol is based on superimposed orthogonal code-division multiplexing with possible time-sharing. This conclusion is reached by comparing not only the achievable rate regions but also other design and system constraints such as multi-rate support (measured in terms of peak-to-average rate ratio), latency, receiver complexity and network planning issues.

## 3.1 Classification of Multi-User Channels

A general communication network is an enormous web consisting of a large population of terminals and their associated communication links. For our analysis we impose the following restrictions:

1. Each terminal is exclusively a transmitter or a receiver, but not both. Bi-directional two-way links are not allowed. This restriction applies only to information not related to channel states. Control channel signalling from a receiver to its transmitter through a feedback channel is allowed.
2. There is an associated information source for each transmitter. Similarly, each receiver relays its decoded message to an information sink. There are equal number of information sources and sinks in the network. It is implicitly assumed that the channel linking a source and its transmitter —likewise, a receiver and its sink— is error- and delay-free.
3. A direct channel (without any intermediate hops) exists between every transmitter-receiver pair. No relay channel is allowed.
4. Each transmitter is autonomous and independent from other transmitters in the network. They are autonomous in the sense that no cooperation among transmitters is permitted in designing their code books. The messages from their respective sources are statistically independent. Simulcasting of the same information from more than one transmitter is not allowed. Equivalently, our model does not support multiple transmit antenna based diversity combining systems.
5. Every receiver accepts information from its intended transmitter as "useful." All other received signals are treated as interference. Multi-accessing is forbidden. Diversity reception is allowed at a receiver in multipath fading channel models.

### 3.1.1 Input-Output Device

Let T, R, I and S denote the number of transmitters, receivers, information sources and sinks, respectively. A subscripted letter will be used to identify a unit in each category. For example, $T_i$ is the $i^{th}$ transmitter in the network. For an M-terminal network, $I = S = M$ and $T, R \leq M$. Based on the above set of restrictions, a communication channel is an input-output device with T inputs and R outputs. A block form is depicted in Fig. 3.1 for $T = m$ and $R = n$.

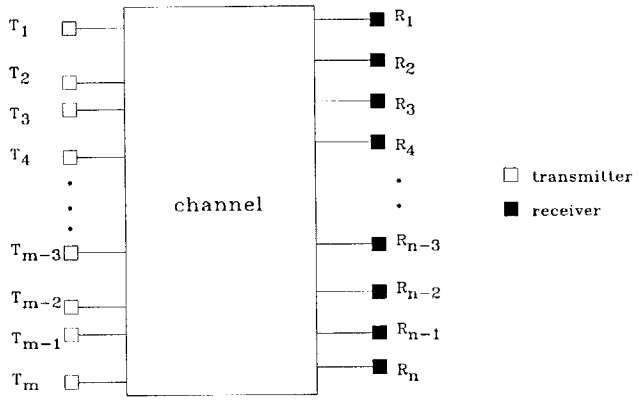



**Figure** 3.1: Channel as an I/O device with $T = m$ inputs and $R = n$ outputs

Depending on the relationship between T and R, three types of multi-terminal channel models can be envisaged.

**1 Definition** *When* $T > R = 1$ *it is called the multi-access channel. At the other extreme, when* $R > T = 1$ *it is known as the broadcast channel. When* $T = R > 1$ *an interference channel results.*

These models and their respective definitions are accepted universally in the literature.

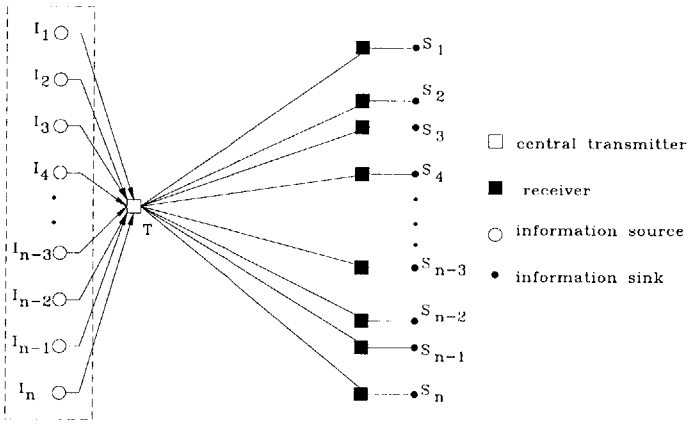
**2 Definition** *When* $T > R > 1$, *it is a composite channel consisting of both multi-access and interference components. Similarly, when* $R > T > 1$, *it becomes a broadcast-interference composite channel.*

Based on above definitions we have exhausted all possible cases relating T and R. Of course, when $T = R = 1$ we have the degenerate case of a single-user communication channel.

### 3.1.2 Multi-Terminal Cooperation

Let us now consider the relationship between the transmitter-receiver pair $(T, R)$ and its associated information source-sink pair $(I, S)$. Note that $I \geq T$ since a number of source outputs can be *multiplexed*, and the resulting multiplexed signal is transmitted through a single

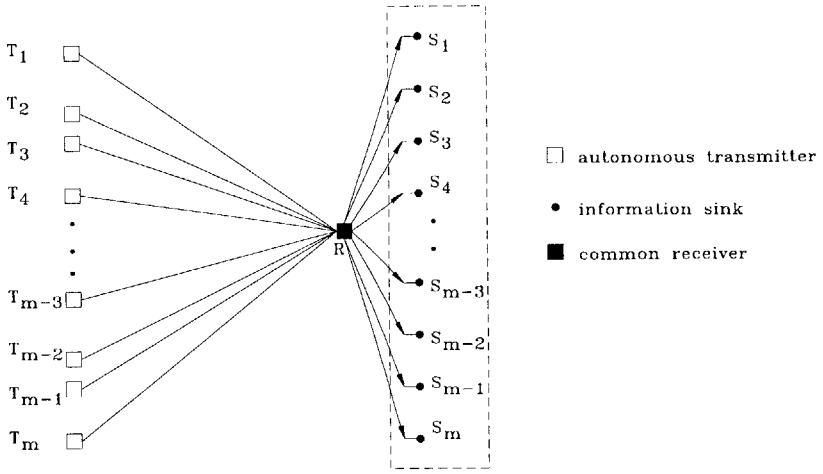
transmitter. Similarly, at the receiving end, $R \leq S$ since a separate receiver is assigned to handle the task of decoding and *demultiplexing* a mutually exclusive subset of the total received composite signal. At all times an active communication link is established between a source $I_i$ and its intended sink $S_i$. If all sources "cooperate" by conveying their independent information through a single transmitter $T = 1$, the resulting I/O device is the broadcast channel. It is depicted in Fig. 3.2. If cooperation among all sources is not available and $T > 1$, we obtain the broadcast-interference composite channel. Whether a channel is purely broadcast or composite depends only on the relation between I and T, i.e., the input side of the channel. Their difference $I - T \geq 0$ is a measure of the amount of cooperation; when $I - T = 0$, there is no cooperation among information sources. For $T = I$ the I/O device is either an interference or multiple-access, or a composite of both channels.




Broadcast channel model with n receivers.

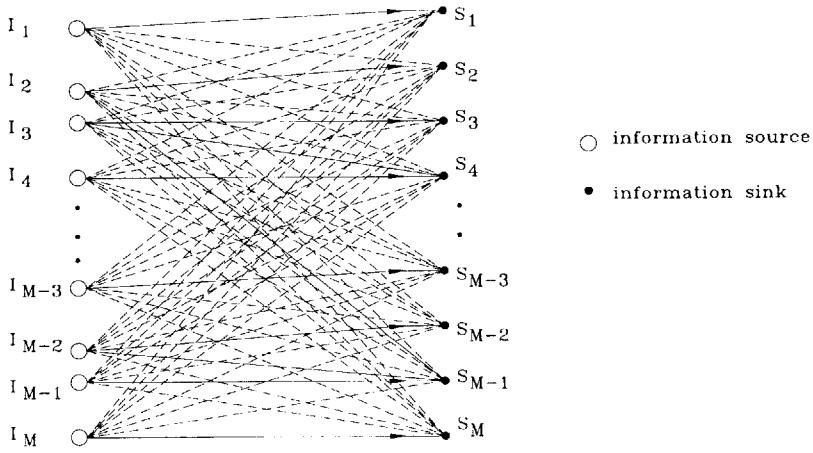
**Figure** 3.2: Each information source $I_k$ communicates with its intended receiver $R_k$ through the central transmitter T.

At the output side, if there is a single receiver ($R = 1$) for all information sinks, it is a multiaccess channel depicted in Fig. 3.3. When $R > 1$ it is a composite channel with some cooperation. The amount of cooperation is measured in terms of $S - R \geq 0$. An interference channel results when there is no cooperation at both input and output (see Fig. 3.4). In general, cooperation at either end (the input or outside side) may not be possible due to physical constraints such as spatial and/or temporal separation or design constraints such as autonomy (non-cooperativeness) and computational complexity. When there is full cooperation at both ends, the model simplifies to a single-user channel. Cooperation in broadcasting takes on such forms as synchronization of component channels in their time epochs, carrier phase coherence, non-colliding assignment of time-slots and/or frequency bands, etc. See page 171 for a classification of various time synchronism options.




Multiple-access channel model with m transmitters.

**Figure** 3.3: Each transmitter $T_k$ communicates with its intended destination sink $S_k$ through the common receiver R.




Interference channel model with M transmitter-receiver pairs.

**Figure** 3.4: Since each transmitter communicates with its intended receiver only, no cooperation exists at both transmit and receive ends.

### 3.1.3 Transition Probability Function

Consider a single-user communication channel with $T = R = I = S = 1$. Information exchange from a source to a sink takes place methodically through the following steps:

- *Message Selection*: The source has a message book $\mathcal{M}$ that contains M messages:

$$\mathcal{M} = \{1, 2, \ldots, M\}$$

For each transmission, it selects an index, say $m = j \in \mathcal{M}$. The underlying message can be something else —for example, if the message is a letter from the English alphabetic set; we set "a" as $m = 1$, "b" as $m = 2$ and so on.

- *Channel Encoding*: The transmitter devises an encoding scheme whose code sequence is most suitable for transmission via the channel. It constructs a code book containing M codewords such that there is a bijective mapping between a message j and its corresponding codeword $\mathbf{X}(j)$[1]:

$$\mathbf{X} : \{1, 2, \ldots, M\} \longrightarrow \mathcal{X}^n$$

where each codeword $\mathbf{X}[j]$ has length n,

$$\mathbf{X}[j] = [X_1[j]\ X_2[j]\ \ldots\ X_n[j]]$$

and each *symbol* $X_k[j]$ of the codeword is a letter from the alphabet set $\mathcal{X}$. Its cardinality is denoted by $|\mathcal{X}|$. Equivalently, $\mathbf{X} \in \mathcal{X}^n$; i.e., the code book consists of a small set of M carefully chosen vectors of length n from a complete set of $|\mathcal{X}|^n$. Note that unlike the message set $\mathcal{M}$, the code symbols from $\mathcal{X}$ are not necessarily positive integers. Each symbol can be any real number that is most suitable for a particular channel and receiver.

- *Information Rate*: The rate R of this $(M, n)$ code is:

$$R = \frac{1}{n} \log_2 M \quad \text{bits per transmission.} \tag{3.1}$$

Some equivalent notations:

$$(M, n) = (2^{nR}, n)$$
$$\mathbf{X} = X^n$$
$$x^k = (x_1\ x_2\ \ldots\ x_k)$$

[1]The time index of discrete-time sequences, the elements and entries of vectors, matrices and sets are expressed by enclosing any variable (except t and $\tau$) in squared brackets; for example, $X[n]$, $s_j[k]$, $H[ij]$. Any other function defined on a real linear scale is expressed with the index in parentheses; for example, $s(t)$, $F(\alpha)$, $c(t, \tau)$.

- *Transition Probability Function*: For each channel input symbol $x_k$, there exists a channel output symbol $y_k \in \mathcal{Y}$ whose relation is defined by a transitional probability function $p(y_k|x_k)$. The output alphabet $\mathcal{Y}$ need not the same as $\mathcal{X}$ in terms of symbols or cardinality. For a codeword $\mathbf{X}$, the transition function is:

$$p\left(\mathbf{Y}\middle|\mathbf{X}\right) = p\left(Y^n|X^n\right)$$

The channel is *memoryless* if the output symbol $y_k$ depends only on the current input symbol $x_k$, and not on past input and output symbols $x^{k-1}$, $y^{k-1}$:

$$p\left(y_k\middle|\, x_k, x^{k-1}, y^{k-1}\right) = p(y_k|x_k)$$

Furthermore, if no *feedback* is allowed such that the current input symbol $x_k$ is independent from past output symbols $y^{k-1}$, i.e.,

$$p\left(x_k\middle|\, x^{k-1}, y^{k-1}\right) = p(x_k|x^{k-1})$$

the transition probability function becomes:

$$p\left(y^n|x^n\right) = \prod_{k=1}^{n} p(y_k|x_k)$$

- *Channel Decoding*: Finally, the receiver constructs a decoding scheme where it estimates the sent message from its received vector $\mathbf{Y}$ based on some optimality criterion:

$$\hat{m} : \mathbf{Y} \longrightarrow \{1, 2, \ldots, M\}$$

An error occurs if $\hat{m}(\mathbf{Y}) \neq j$ when $\mathbf{X}(j)$ is sent. The conditional probability of error when the message $j$ is sent is:

$$\lambda_j = P\left(\hat{m}(Y^n) \neq j \middle| m = j, X^n(j)\right)$$

The average probability of error is defined as:

$$P_e = \sum_{j=1}^{M} \lambda_j P_j$$

where $P_j$ is the a priori probability of choosing message $j$ and $\sum_{j=1}^{M} P_j = 1$. If they are all equally likely,

$$P_e = \frac{1}{M} \sum_{j=1}^{M} \lambda_j$$

Furthermore, if the channel is symmetrical such that $\lambda_j$ is identical for all $j \in \mathcal{M}$,

$$P_e = P\left(\hat{m}(Y^n) \neq j,\, X^n(j)\right)$$

**3 Definition** *A single-user discrete communication channel is defined compactly as*

$$\left(\mathbf{X}, \mathbf{P}, \mathbf{Y}, R\right)$$

*where* **P** *is the channel transition probability matrix whose entries are*

$$p(y_j|x_i),\ j \in |\mathcal{Y}|,\ i \in |\mathcal{X}|$$

*Information rate* R *is said to be* achievable *if there exists a* $(2^{nR}, n)$ *code such that* $P_e \to 0$ *for an arbitrarily large* $n$. *The supremum of all achievable rates is the* channel capacity.

### 3.1.4 Broadcast Channel

The above definitions and notations of a single-user channel can readily be adapted to multi-terminal channel models. For $T = 1$, $R = N > 1$ and $I = S = N$, we obtain a broadcast channel. The joint transition probability of a broadcast channel is:

$$p(y_1, y_2, \ldots, y_N|v)$$

where $v \in \mathcal{V}$ is a composite symbol transmitted by the broadcaster representing independent messages for N sources, and $y_i \in \mathcal{Y}_i$ is the received symbol of receiver $R_i$. It is assumed that in a purely broadcast setting, there is no cooperation among receivers. Conditioned on the transmitted symbol $v$, the received symbols $y_i$ and $y_j$ $(i \neq j)$ are independent. The joint transition probability function is then broken into marginal transition probabilities:

$$p(y_1|v),\ p(y_2|v),\ \ldots,\ p(y_N|v)$$

In block form the broadcast channel, as illustrated in Fig. 3.5, is a single-input, multiple output device with several embedded blocks of "sub-channels," each with its associated transition probability function. To avoid cumbersome notations, we study the simplest case by setting $N = 2$.

At the input side, two information sources select their independent messages from their respective message books $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\mathcal{M}_1 = \{1, 2, \ldots, M_1\}$$
$$\mathcal{M}_1 = \{1, 2, \ldots, M_2\}$$

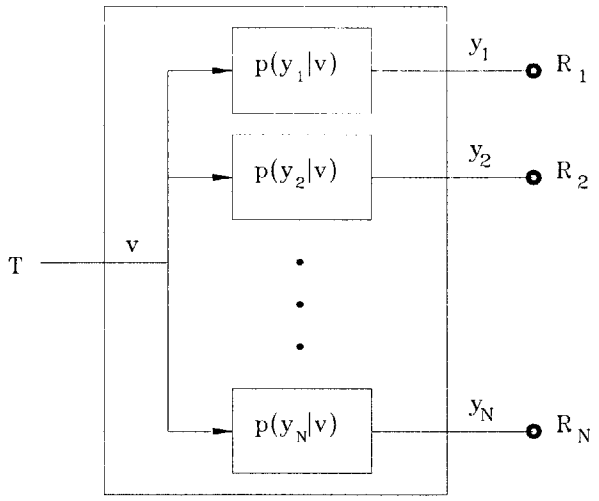



**Figure** 3.5: Broadcast channel as a single-input, multiple-output device with N receivers

In general, $M_1 \neq M_2$. The transmitter devises the following encoding scheme:

$$\begin{aligned}
\mathbf{X}_1 &: \{1,2,\ldots,M_1\} \longrightarrow \mathcal{X}_1^n \\
\mathbf{X}_2 &: \{1,2,\ldots,M_2\} \longrightarrow \mathcal{X}_2^n \\
\mathbf{V} &: \{1,2,\ldots,M_1\} \times \{1,2,\ldots,M_2\} \longrightarrow \mathcal{V}^n \\
\mathbf{V} &= \mathbf{X}_1 + \mathbf{X}_2 \\
\mathbf{V}(i,j) &= \mathbf{X}_1(i) + \mathbf{X}_2(j) \\
&= [v_1(i,j)\, v_2(i,j)\, \ldots\, v_n(i,j)] \\
v_k(i,j) &= x_{1k}(i) + x_{2k}(j)
\end{aligned}$$

The rate for this $(2^{nR_1}, 2^{nR_2}, n)$ code is:

$$R = \frac{1}{n}\log_2(M_1 M_2)$$

with the understanding that for each receiver,

$$R_1 = \frac{1}{n}\log_2 M_1 \qquad R_2 = \frac{1}{n}\log_2 M_2$$
$$\mathbf{R} = (R_1, R_2)$$

The transition probability function is:

$$\begin{aligned}
p\left(\mathbf{Y}_1\mathbf{Y}_2 \,\middle|\, \mathbf{V}\right) &= p(Y_1^n Y_2^n | V^n) \\
&= p(Y_1^n | V^n)\, p(Y_2^n | V^n)
\end{aligned}$$

If the channel is memoryless with no feedback:

$$p\left(\mathbf{Y}_1\mathbf{Y}_2 \middle| \mathbf{V}\right) = \prod_{k=1}^{n} p(y_{1k}|v_k)p(y_{2k}|v_k)$$

The decoding functions are:

$$\hat{m}_1 : \mathbf{Y}_1 \longrightarrow \{1, 2, \ldots, M_1\}$$
$$\hat{m}_2 : \mathbf{Y}_2 \longrightarrow \{1, 2, \ldots, M_2\}$$

An error occurs when a message pair $(i,j)$ is sent and $\hat{m}_1 \neq i$ or $\hat{m}_2 \neq j$. The probability of error is defined as:

$$P_e = P\left(\hat{m}_1 \neq i \text{ or } \hat{m}_2 \neq j, \mathbf{V}(i,j)\right)$$

where it is assumed that any message pair $(i,j)$ is uniformly distributed over $M_1 \times M_2$ and the channel is symmetric. A two-user channel is compactly denoted by $\left(\mathbf{V}, \mathbf{P}, \mathbf{Y}_1, \mathbf{Y}_1, \mathbf{R}\right)$.

**4 Definition** *For a discrete broadcast channel, a rate pair* $(R_1, R_2)$ *is said to be* achievable *if there exists a code of rate* $R = (R_1+R_2)$*, denoted by* $\left(2^{nR_1}, 2^{nR_2}, n\right)$ *such that* $P_e \to 0$ *for an arbitrarily large* $n$*. The* capacity region *is the closure of the set of all achievable rates.*

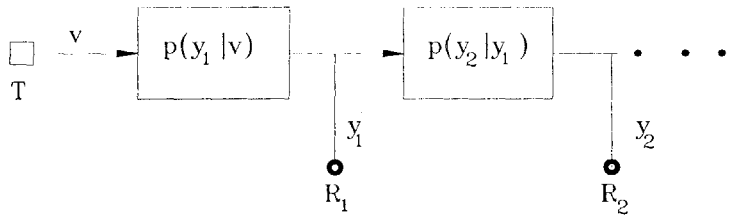



**Figure** 3.6: Physically degraded broadcast channel model with two receivers

### 3.1.5 Degraded Broadcast Channel

In certain broadcast channels such as terrestrial mobile radio, a receiver $R_1$ that is closer to the transmitter has a better reception than another receiver $R_2$ that is located further "downstream." It is said that $R_2$ receives a "physically degraded" (p-degraded) version of a signal received by $R_1$. In terms of transition probability functions:

$$p(y_1, y_2|v) = p(y_1|v)p(y_2|y_1)$$
$$p(y_2|v) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1|v)p(y_2|y_1)$$

Since $p(y_1, y_2|v) = p(y_1|v)\,p(y_2|v)$, with p-degradation $p(y_2|v) = p(y_2|y_1)$. In a general broadcast model with N receivers, if the p-degraded condition holds for all receivers

$$p(y_j|v) = p(y_j|y_{j-1}) \tag{3.2}$$

where $j \in \mathbb{Z}_2^N$, then it is called a "cascaded" physically degraded broadcast channel. This channel model is illustrated in Fig. 3.7. If eqn. (3.2) does not hold for all receivers, other kinds of degraded broadcast channel models can be constructed. An example is shown in Fig. 3.8 where each channel block as a cascaded p-degraded broadcast channel. This type of channel model is encountered when a cellular coverage is broken into partitioned zones. See the section on re-use partitioning in Appendix 3A for additional details. As a weaker condition, a broadcast

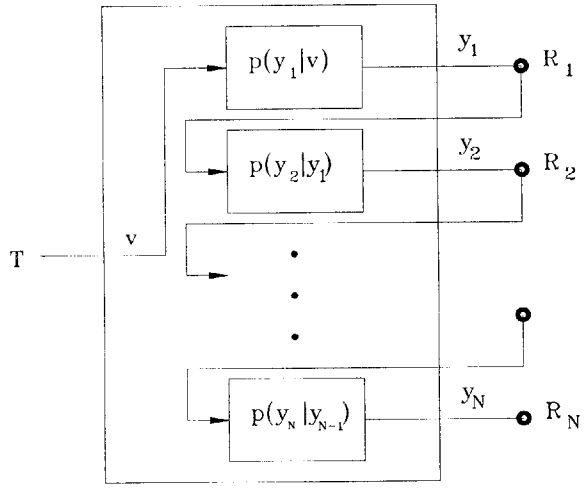



**Figure** 3.7: Cascaded, physically degraded broadcast channel with downstream degradation

channel $p(y_2, y_1|v)$ is said to be *degraded* if there exists another distribution $\tilde{p}(y_2|y_1)$ (possibly different from $p(y_2|y_1)$) such that

$$p(y_2|v) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1|v)\tilde{p}(y_2|y_1)$$

This is the less restricted version of broadcast degradation. Unless noted otherwise, a degraded broadcast channel has a transition probability function of this kind.

### 3.1.6 Interference Channel

For the unique case of T = R = I = S where each transmitter wishes to convey information to a single receiver, the *interference* channel results (see Fig. 3.4). Each receiver decodes only

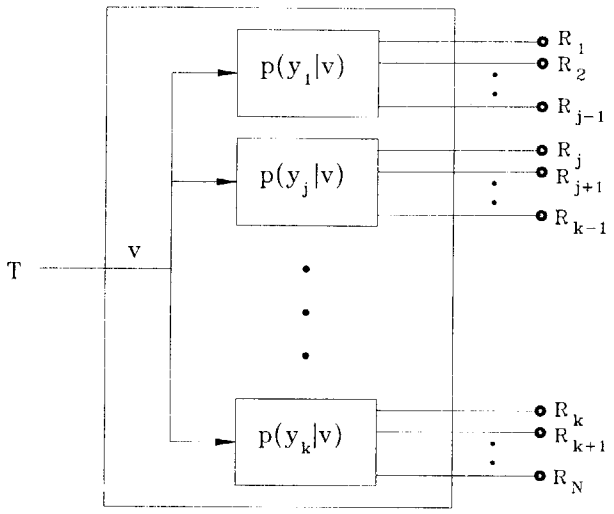



**Figure** 3.8: Sub-channel blocks where each is a cascaded p-degraded broadcast channel

the code symbol representing its intended message and treat other transmit symbols as interference. In Fig. 3.4, to differentiate reception of information from interference, the solid-line channel that connects a transmitter $T_i$ and its target receiver $R_i$ is called the *principal link*. All other dotted links that result in interference to other receivers are called *interference links*. For a broadcast channel, every link is a principal link. In an interference channel model, all transmitters are autonomous; that is, their codewords are independent. However, it is implicitly assumed that they are time-synchronized such that the $j^{th}$ code symbols from all N transmitters $y_{1k}[j], y_{2k}[j], \ldots, y_{Nk}[j]$ are observed at the same time in each receiver k.

For the two-user $N = 2$ case,
*Encoding of information:*

$$\mathcal{M}_1 = \{1, 2, \ldots, M_1\}$$
$$\mathcal{M}_1 = \{1, 2, \ldots, M_2\}$$

$$\mathbf{X}_1 : \{1, 2, \ldots, M_1\} \longrightarrow \mathcal{X}_1^n$$
$$\mathbf{X}_2 : \{1, 2, \ldots, M_2\} \longrightarrow \mathcal{X}_2^n$$

The rate for this $(2^{nR_1}, 2^{nR_2}, n)$ code is:

$$R = \frac{1}{n} \log_2 (M_1 M_2)$$
$$\mathbf{R} = (R_1, R_2) \qquad R_1 = \frac{1}{n} \log_2 M_1 \qquad R_2 = \frac{1}{n} \log_2 M_2$$

*Transition Probabilities:*

$$\begin{aligned} p\left(\mathbf{Y}_1\mathbf{Y}_2 \middle| \mathbf{X}_1, \mathbf{X}_2\right) &= p\left(Y_1^n Y_2^n | X_1^n X_2^n\right) \\ &= p\left(Y_1^n | X_1^n X_2^n\right) p\left(Y_2^n | X_1^n X_2^n\right) \end{aligned}$$

If the channel is memoryless with no feedback:

$$p\left(\mathbf{Y}_1\mathbf{Y}_2 \middle| \mathbf{X}_1, \mathbf{X}_2\right) = \prod_{k=1}^{n} p\left(y_{1k} | x_{1k}, x_{2k}\right) p\left(y_{2k} | x_{1k}, x_{2k}\right)$$

*Decoding of information:*

$$\begin{aligned} \hat{m}_1 &: \mathbf{Y}_1 \longrightarrow \{1, 2, \ldots, M_1\} \\ \hat{m}_2 &: \mathbf{Y}_2 \longrightarrow \{1, 2, \ldots, M_2\} \\ P_e &= P\left(\hat{m}_1 \neq i \text{ or } \hat{m}_2 \neq j; \mathbf{X}_1(i), \mathbf{X}_2(j)\right) \end{aligned}$$

where it is assumed that both messages $i$ and $j$ are uniformly distributed over $M_1$ and $M_2$, respectively and the channel is symmetric. A two-user interference channel is compactly denoted by

$$\left((\mathbf{X}_1, \mathbf{X}_2), \mathbf{P}, (\mathbf{Y}_1, \mathbf{Y}_2), \mathbf{R}\right)$$

**5 Definition** *For a discrete interference channel, a rate pair* $(R_1, R_2)$ *is said to be* achievable *if there exists two independent codes denoted by* $\left(2^{nR_1}, 2^{nR_2}, n\right)$ *of rates* $R_1$ *and* $R_2$ *respectively such that* $P_e \to 0$ *for an arbitrarily large* $n$. *The* capacity region *is the closure of the set of all achievable rates.*

### 3.1.7 Composite Channel

When $T = M > 1$, $R = N > 1$ and $M < N$, the model is a combination of both broadcast and interference channels, since there is at least one information source that communicates with two or more receivers. Analogously, when $T = M > 1$, $R = N > 1$ and $M > N$, the model is a combination of both multiple-access and interference channels, since there is at least one receiver that decodes information from two or more sources. Both types of compound channel are illustrated in Fig. 3.9.

**6 Definition** *In our study, a* composite *channel is defined as a compound channel that consists of* broadcast *and* interference *sub-channels only.*

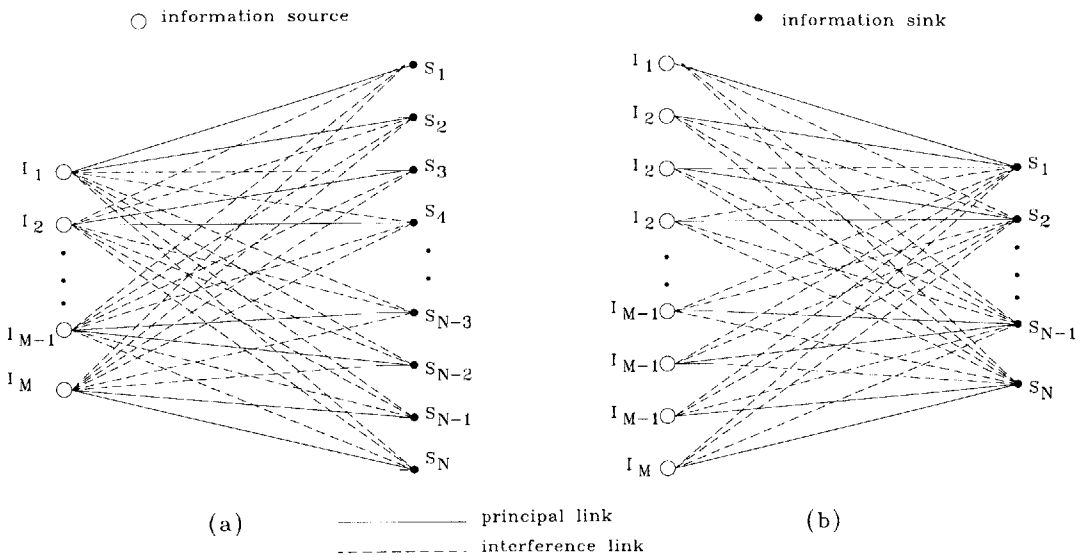
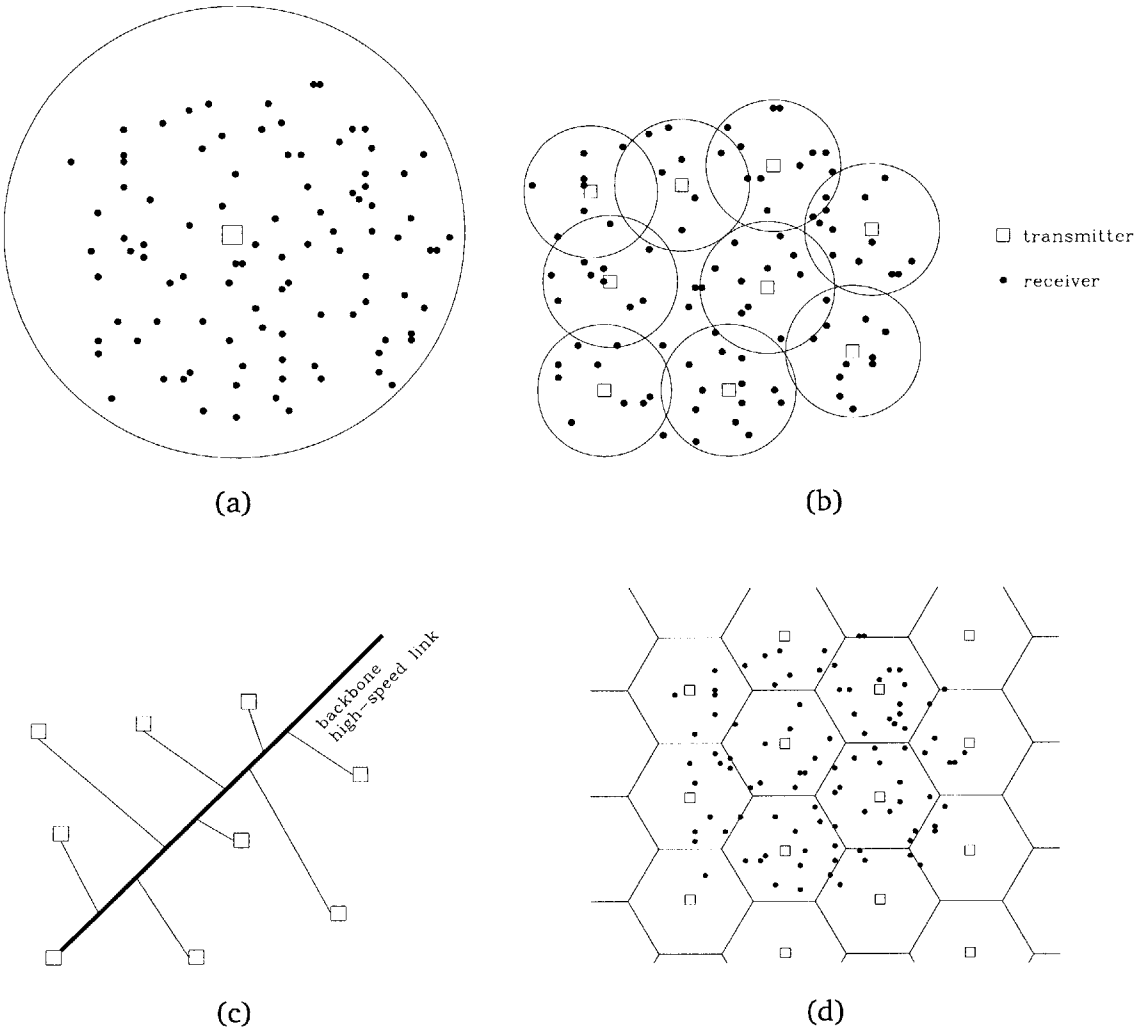



**Figure 3.9:** Two types of composite channel models. When $M > N$, it has multiple-access and interference components. When $M < N$, it is a combination of both broadcast and interference channels.

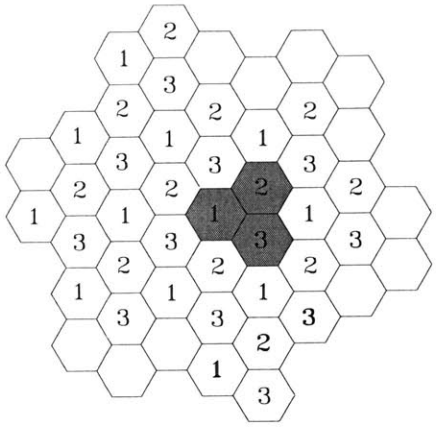
### 3.1.8 Cellular Broadcast Channel

In a complex communication network such as mobile cellular radio with a large terminal population, cooperative broadcasting or multiple-accessing is not always possible or practically sound. Cooperation is synonymous to *centralization*. Consider the aerial view (Fig. 3.10(a)) of a geographical area with a large set of scattered terminals (radios) —each terminal depicted by a • symbol. First, it would be wasteful in transmit power to cover the entire area from a single transmitter (denoted by a □ symbol in the figure). Terminals farthest from the transmitter receive a severely degraded version of the transmit broadcast signal. In practice a upper limit on the (time average) power limit per transmitter is imposed. Second, the total bandwidth allocated to the entire network must be shared among all active radios. Therefore, the information rate per radio is inversely proportional to the number of active users. This is the bandwidth constraint. If there is also a limitation on the amount of power the channel can carry, it is said to be *power-limited*. Mobile radio channels are both power- and bandwidth limited. Cellular architecture resolves both power and bandwidth constraints by setting up a distributed network. The entire coverage area is partitioned along imaginary boundary lines into many autonomous smaller coverage areas. This is illustrated in Fig. 3.10 (b). Each small coverage area is called a *cell*. By convention, an autonomous "immobile" transmitter —also known as a base station with omni-directional signal radiation— is situated at the center of a cell. Each transmitter sets

up principle links with radios or *mobiles* within its cell. Since power leaks beyond a cell edge, it causes interference to radios in neighboring cells. If there are $M_t$ transmitters and $N_t (\gg M_t)$ receivers, the entire network is equivalent to a communication system with $M_t$ inputs, $N_t$ outputs, connected via a broadcast-interference composite channel. As depicted in Fig. 3.10(c) it is understood that information from sources to their corresponding transmitters takes place along high-speed links that are error- and delay-free. The power-constraint condition is now limited to each transmitter over a local coverage area.




(a) Centralized communication network with a single transmitter
(b) Distributed network with many localized transmitters
**Figure** 3.10: (c) Information transfer from sources to transmitters via high-speed backbone links
(d) Distributed cellular network with regular grid pattern and infinite number of cells. All cell sizes are equal.



Aerial view of a planar cellular network with hexagonal cells.

**Figure** 3.11: The reuse number is $N_r = 3$. A region of $N_r$ cells (shaded gray) is called a cluster.

Due to propagation loss (see Ch. 2, Sec.2.1), the interference power level is drastically reduced in cells farther away from a transmitter. In such cells, the same bandwidth can be reused. This is the *cellular-reuse* concept and it is intricately tied to distributed network architecture. An example of cellular re-use is shown in Fig. 3.11. In all cells labelled **1**, the same frequency band is reused simultaneously. The same reuse concept applies to other cells labelled **2** and **3**. In this example the *reuse number* $N_r$ is 3. All cells that are assigned the same frequency band (bandwidth) are called *co-channel* cells. Under the cellular-reuse setting, we define this type of multi-terminal composite communication channel as the *cellular broadcast channel*. In general, the entire cellular network is blanketed with autonomous cells, each with transmit power P and bandwidth $W/N_r$. The bandwidth $W_m$ per radio is:

$$W_m = \frac{W}{N_t} \cdot \frac{M_t}{N_r} \tag{3.3}$$

where $N_t$ and $M_t$ are the total number of radios and cells (transmitters) in the network. Compare $B_m$ with the bandwidth per radio of a centralized single-transmitter network: $W/N_t$. As long as the entire coverage area is partitioned such that there are more cells than the reuse number, i.e., $M_t > N_r$, the cellular distributed approach is more *bandwidth efficient* than its alternative based on centralized architecture. More details about hexagons, notations, definitions and its geometrical properties are presented in Appendix 3A.

Eqn. (3.3) is correct only if the mobile receivers are uniformly distributed —i.e., the same number of mobiles per cell. An alternative measure of bandwidth efficiency is in terms of bandwidth per cell $W_c = W/N_r$. As long as the number of mobiles per cell $K_c$ is less than $N_t/N_r$

$$W_m = \frac{W_c}{K_c} = \frac{W}{N_r}\frac{1}{K_c} > \frac{W}{N_r}\frac{N_r}{N_t} = \frac{W}{N_t}$$

the cell-based approach is more efficient in mobile bandwidth. Note that this fact is independent of the number of transmitters $M_t$. However, in general we have no control over the distribution or the concentration of mobile population within the network. $N_t$ is in fact a random process. The only design parameters are $M_t$ and $N_r$. By assigning a smaller re-use number the cell bandwidth $W_c$ as well as mobile bandwidth $W_m$ is increased. However, this remark is misleading. By reducing the reuse $N_r$, the co-channel cells become closer and interference power is increased; as a result, the information rate per mobile $R_m$ is reduced. The *spectral efficiency*[2] (in this case, of a mobile) is measured in terms of its achievable information rate with an arbitrarily small probability error $P_e$.

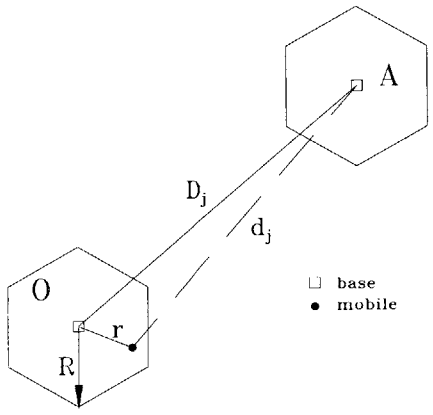



**Figure** 3.12: Interference from co-channel cell A at distance $D_j$. When $D_j \gg R$, we accept the approximation $D_j = d_j$.

### Equivalent Composite Channel Model

When the number of mobiles $N_t$ is fixed, the bandwidth efficiency of a cell is a function of several variables: its size, the number of cells in the network and the re-use number. The entire network is a broadcast-interference composite channel with $M_t$ inputs and $N_t$ outputs. By modifying this model such that the coverage area is infinite, and by tiling same-size (hexagon) cells contiguously (as shown in Fig. 3.10), the cell bandwidth efficiency becomes a function of a single parameter: re-use number $N_r$. In fact, there is no need to analyze the entire network; any cell is representative of the entire network —assuming the number (and location) of mobiles in every cell N are independent and identically distributed random variables. Every receiver

[2]In communication texts, bandwidth and spectral efficiency terms are used interchangeably. We, however, differentiate the two. The former is derived from cellular radio engineering design practices, while the latter is a strictly information-theoretic quantity. In Chapter we also use the term "bandwidth" efficiency as a measure of the power spectrum extent of a transmit signal compared to its Nyquist bandwidth.

in a cell "sees" a broadcast-interference channel since the received signal consists of broadcast signal from its base as well as interfering signals from other neighboring base stations.

It is shown in Appendix 3A that for every target cell, its neighboring cells can be partitioned into mutually exclusive sets $\mathcal{G}_c[j]$, $j \in \mathbb{Z}^{M-1}$ such that all cells in a particular set $\mathcal{G}_c[j]$ are equally distant from the target cell. Consider the layout in Fig. 3.12 where a neighboring cell **A** (i.e., its base station) from set $\mathcal{G}_c[j]$ is located at distance $D_j$ from the target cell **0**. R is the cell radius. A mobile at a distance r from the target is $d_j$ away from its neighboring base. Note that besides cell **A** the mobile receives interference from other cells in $\mathcal{G}_c[j]$. Each elemental interfering cell in a set $\mathcal{G}_c[j]$ is significant only in a collective sense since the most important determinant is the aggregate interference power from cells located the same distance from the target cell. Hence, they are all lumped into a "single" interferer at distance $d_j$. In terms of transition probabilities, the following derivation relates the channel gains of the principle link (from the target cell) and interfering links.

By applying the simplified path loss formula of eqn. (2.4), the mobile received power is:

$$P_r = G \cdot P_t \left( \frac{1}{r^\gamma} + \frac{k_1}{d_1^\gamma} + \frac{k_2}{d_2^\gamma} + \cdots + \frac{k_{M-1}}{d_{M-1}^\gamma} \right)$$

where

$$k_j = \left| \mathcal{G}_c[j] \right| \qquad \text{and} \qquad G = G_{ta} G_{ra} \left( h_t h_r \right)^2$$

Using the re-use number relation of eqn. (3.33) and the approximation:

$$\frac{d_j}{r} \approx \frac{D_j}{r} \qquad D_j \gg R$$

The path loss $L_p$ is:

$$\begin{aligned} \frac{P_r(r)}{P_t} &= \left( 1 + \frac{k_1}{(d_1/r)^\gamma} + \frac{k_2}{(d_2/r)^\gamma} + \cdots + \frac{k_{M-1}}{(d_{M-1}/r)^\gamma} \right) \frac{G}{r^\gamma} \\ &= \left( 1 + \frac{k_1}{\left[ 3N_r[1;r] \right]^{\gamma/2}} + \frac{k_2}{\left[ 3N_r[2;r] \right]^{\gamma/2}} + \cdots + \frac{k_{M-1}}{\left[ 3N_r[M-1;r] \right]^{\gamma/2}} \right) \frac{G}{r^\gamma} \end{aligned}$$

where $N_r[j;r]$ is the smallest re-use number such that

$$\frac{d_j}{r} \le \sqrt{3N_r[j;r]}$$

For mobiles located at different distances within a cell, we can partition a total of $K_c$ of them into mutually exclusive sets $\mathcal{G}_m[n]$, $n \in \mathbb{Z}^N$ of mobiles where each set consists of mobiles that are located the same distance $r_n$ from the target base and $0 \le r_1 < r_2 < \cdots < r_N \le R$. By

defining the *normalized received power* $\mathcal{P}_r[n]$ for a set of mobiles located at distance $r_n$:

$$\mathcal{P}_r[n] = \frac{P_t G}{r_n^{\gamma}} \sum_{j=0}^{M-1} \beta^{(j)}[n]$$

where the power gains are:

$$\beta^{(0)}[n] = 1$$
$$\beta^{(j)}[n] \approx \frac{k_j}{\left[3N_r[j;n]\right]^{\gamma/2}}$$

where The power gain $\beta^{(n)}$ is the square of the channel (amplitude) gain:

$$\alpha^{(n)} = \sqrt{\beta^{(n)}}$$

By default, the channel gain of the principal link $\alpha^{(0)}[1] = 1$ for mobiles in $\mathcal{G}_m[1]$. To complete the picture, we remark that this model is equivalent to a composite channel with M inputs (a single broadcaster and $M-1$ interfering neighbors) and N outputs, where N is the (random) number of mobiles in a cell. This channel model and its properties have already been introduced in Sec. 3.1.7. A modified version of the composite channel in Fig. 3.9(a) is depicted in Fig. 3.13(a) for $M = 2$ and $N = 4$. Note the significance of principal and interference links: In Fig. 3.13(b) a portion of the composite channel with two inputs and outputs is shown. This, however, is not an interference channel since both receivers treat the signal from $T_1$ as interference.

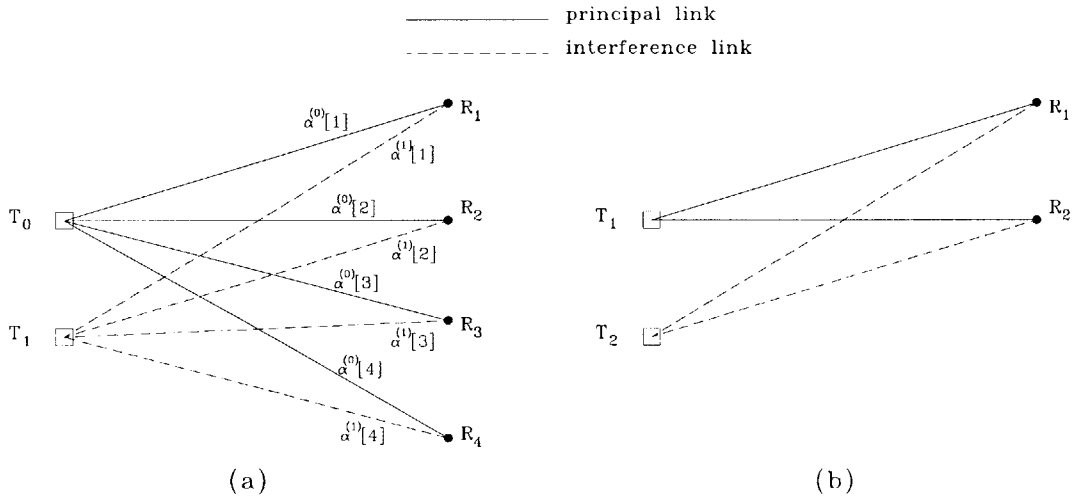



**Figure** 3.13: Composite channel model with link gains

### 3.1.9 Continuous-time Bandlimited Channel

When information is conveyed from a base transmitter to mobile receivers through a radio propagation medium, the transmit signals take on the form of *continuous-time, analog* electromagnetic waves; i.e., both inputs and outputs of the I/O device are continuous-time functions. All multi-user channels described so far in the previous sections are *discrete-time* models where each code symbol is sent at a rate of one per *transmission* or *channel use* —not once per $T_s$ seconds. The discrete-time, finite-alphabet channel model can be generalized by including the functions of a modulator and demodulator set. An equivalent block form of a discrete-time channel model with embedded continuous-time modulator-demodulator (modem) is depicted in Fig. 3.14.

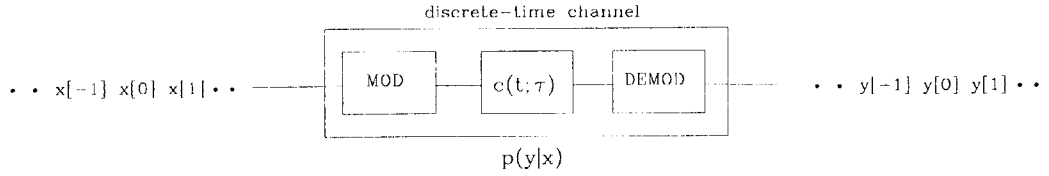



**Figure** 3.14: Discrete-time channel with embedded modem

The two-parameter time function $c(t;\tau)$ is the linear time-variant (LTV) impulse response (IR) of a radio propagation channel. Its properties are covered in Ch. 2, Sec. 2.2. The conversion of code symbols to waveforms is facilitated by the inclusion of a symbol or *data modulator* after the encoder. As shown in Fig. 3.15, the modulator consists of a periodic impulse modulating sequence followed by a pulse-shaping filter with impulse response $f(t)$. If code symbols

$$\ldots x[-1],\, x[0],\, x[1] \ldots \qquad x[j] \in \mathcal{X}$$

are complex-valued, each symbol

$$x[j] = \Big(x_I[j],\, x_Q[j]\Big)$$

can be split into in-phase and quadrature components; then each component is modulated by an *intermediate frequency* (IF) cosine or sine carrier. The modulation process is a bijective mapping of encoded symbols from the alphabet $\mathcal{X}$ to a finite set of *real* waveforms

$$\mathbf{S}(t) = \big\{s_1(t), s_2(t), \ldots, s_M(t)\big\}$$

where $|\mathcal{X}| = |\mathbf{S}(t)| = M$, and

$$x \in \mathcal{X} \longrightarrow s(t) \in \mathbf{S}(t) \tag{3.4}$$

The mapping can be expressed concisely as follows:

$$s(t) = g_I(t)\cos(2\pi f_o t) - g_Q(t)\sin(2\pi f_o t) \qquad t \in \big[jT_s, (j+1)T_s\big)$$

where $g_I(t)$ and $g_Q(t)$ are filter outputs:

$$g_s(t) = \int f(\tau) x_s(t-\tau)\, d\tau \qquad s = I, Q \qquad t \in [jT_s, (j+1)T_s) \tag{3.5}$$

The filter input $x_s(t)$ is the continuous-time equivalent of the code symbol sequence:

$$\begin{aligned} x_s(t) &= \delta_{T_s}(t) \cdot \sum_j x_s[j] \\ &= \sum_j x_s[j]\, \delta(t - jT_s) \end{aligned} \tag{3.6}$$

After backward substitution:

$$\begin{aligned} s_{T_s}(t) &= \sum_j s(t - jT_s) \\ &= \sum_j f(t - jT_s)\Big(x_I[j]\cos(2\pi f_o t) - x_Q[j]\sin(2\pi f_o t)\Big) \end{aligned}$$

Since symbols are transmitted at a rate of one every $T_s$, we denote $R_s$ as the *symbol rate* and

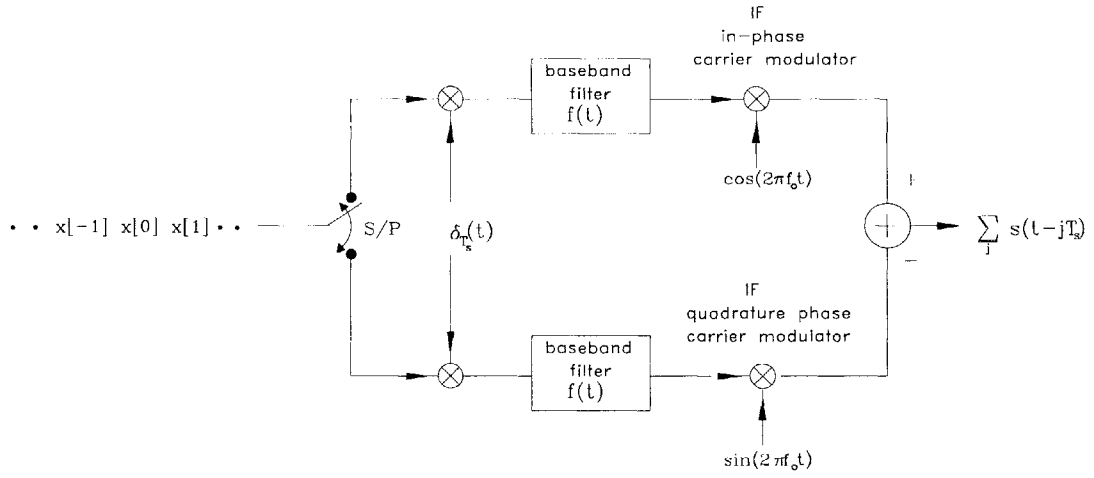



**Figure 3.15:** I/Q data modulator of complex-valued code symbols $x[j] = (x_I[j], x_Q[j])$ with pulse shaping filter impulse response $f(t)$

$T_s$ as the *symbol duration*. Note that the length of each pulse $f(t)$ —which we call the *pulse duration* $T_p$— is not constrained to $[0, T_s)$. We postpone any further discussion on this topic until Chapter 5. At this point, we simply assume that the power signal $s_{T_s}(t)$ has power spectral density contained in $W$ Hz, and that the passband channel bandwidth is also limited to $W$ Hz. It is known from the *Dimensionality Theorem* that reliable transmission at a rate of D waveforms

per second is achievable through a strictly $W$ Hz bandlimited channel if the symbol duration $T_s$ satisfies the following equality :

$$D = \lfloor 2WT_s \rfloor$$

Hence, transmission of a code symbol per transmission in the discrete case is equivalent to transmission of $2W$ waveforms every $T_s$ seconds. If the received signal is perturbed by additive zero-mean Gaussian noise process $n(t)$ at the front end:

$$r(t) = c_p(t;\tau) \star s_{T_s}(t) + n(t)$$

After down-converting in the in-phase channel:

$$= \left[c_p(t;\tau) \star s_{T_s}(t)\right] \cos(2\pi f_o t) + n(t)\cos(2\pi f_o t)$$

After baseband filtering with impulse response $h(t)$:

$$= \left\{ \left[c_p(t;\tau) \star s_{T_s}(t)\right] \cos(2\pi f_o t) + n(t)\cos(2\pi f_o t) \right\} \star h(t)$$

Assuming $f_o \gg W$, the double-frequency terms are removed by baseband filtering, and after some massaging, the baseband received signal is:

$$y_I(t) = \sum_j x_I[j]\, f(t - jT_s) \star c(t;\tau) \star h(t) + n_I(t)$$

where the baseband noise in the in-phase channel is:

$$n_I(t) = \left\{ n(t)\cos(2\pi f_o t) \right\} \star h(t)$$

Analogously, the quadrature noise can be written as:

$$n_Q(t) = \left\{ n(t)\sin(2\pi f_o t) \right\} \star h(t)$$

Since the I and Q carriers are orthogonal, $n_I(t)$ and $n_Q(t)$ are uncorrelated. Since they are Gaussian distributed, they are also independent. After sampling $y(t)$ at the symbol rate:

$$y_I(jT_s) = y_I[j] = x_I[j]\, c[j] + n_I[j]$$

where the discrete-time channel gain is:

$$c[j] = f(t - jT_s) \star c(t;\tau) \star h(t)\Big|_{t=jT_s}$$

Similar derivation for the quadrature component leads to an expression:

$$y_Q[j] = x_Q[j]\, c[j] + n_Q[j]$$

In practice, the value of $y[j] = (y_I[j], y_Q[j])$ may be quantized such that its alphabet $\mathcal{Y}$ is a discrete-valued finite set. To complete the picture, we illustrate a block diagram of an I/Q data demodulator in Fig. 3.16.

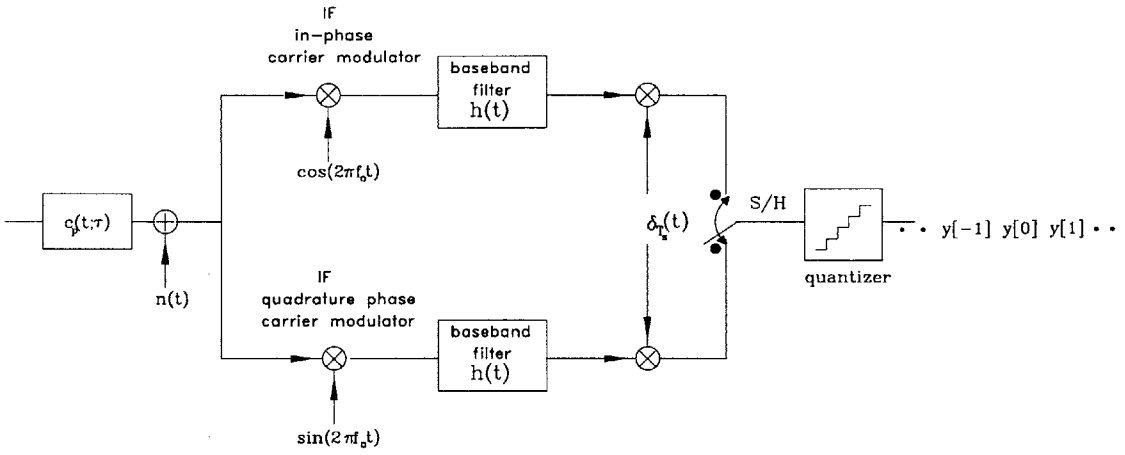



**Figure 3.16:** I/Q data demodulator of complex-valued code symbols $y[j] = (y_I[j], y_Q[j])$ with baseband filter impulse response $h(t)$

## 3.2 Single-Cell Channel Capacity

For a two-user Gaussian broadcast model,

$$y_1' = c_1 v + n_1'$$
$$y_2' = c_2 v + n_2'$$

where $c_1$ and $c_2$ are channel (amplitude) gains. Assume $c_1 > c_2$; i.e., user 1 has higher SNR. To simplify notations we drop the brackets. The noise variables $n_1'$ and $n_2'$ are zero-mean i.i.d. Gaussian RVs with equal variance $\sigma^2$. This assumption is accurate since it can be shown — based on equations from the previous section— that at sample time instant $jT_s$, the in-phase and quadrature noise components $n_I(jT_s)$ and $n_I(jT_s)$ are zero-mean, equal variance Gaussian RVs. After scaling,

$$y_1^* = \frac{y_1'}{c_1} = v + \frac{n_1'}{c_1}$$
$$y_2^* = \frac{y_2'}{c_2} = v + \frac{n_2'}{c_2}$$

Since the channel gains are compensated at the receiver, the standard-form of a Gaussian broadcast model can be written as:

$$\begin{aligned} y_1 &= v + n_1 \\ y_2 &= v + n_2 \end{aligned} \tag{3.7}$$

The variances of the noise variables $n_1$ and $n_2$ are $\sigma_1^2 = \sigma^2/c_1^2$ and $\sigma_2^2 = \sigma^2/c_2^2$, respectively. In general, $c_1 \neq c_2$ and $\sigma_1^2 \neq \sigma_2^2$. Assume $\sigma_1^2 < \sigma_2^2$. Rewriting the above equations:

$$y_1 = v + n_1$$
$$y_2 = v + n_1 + (n_2 - n_1)$$

We see that $y_2$ is a degraded version of $y_1$, consistent with the fact that user 2 is further "downstream" since $c_1 > c_2$. Thus, the additive Gaussian broadcast channel qualifies as a degraded broadcast channel.

For a two-user Gaussian interference model,

$$y_1' = c_{11} v_1 + c_{21} v_2 + n_1'$$
$$y_2' = c_{12} v_1 + c_{22} v_2 + n_2'$$

After scaling,

$$y_1^* = \frac{y_1'}{c_{11}} = v_1 + \frac{c_{21}}{c_{11}} v_2 + \frac{n_1'}{c_{11}}$$
$$y_2^* = \frac{y_2'}{c_{22}} = \frac{c_{12}}{c_{22}} v_1 + v_2 + \frac{n_2'}{c_{22}}$$

The standard form[3] is:

$$\begin{aligned} y_1 &= v_1 + a_1 v_2 + n_1 \\ y_2 &= a_2 v_1 + v_2 + n_2 \\ a_1 &= \left(\frac{c_{21}}{c_{11}}\right), \qquad a_2 = \left(\frac{c_{12}}{c_{22}}\right) \end{aligned} \tag{3.8}$$

The channels compose an *interference* channel if and only if $a_1, a_2 > 0$. The interference channel is *lossy* when $a_1, a_2 < 1$; for the opposite, the channel is *amplified* when $a_1, a_2 > 1$. It is *lossless* when $a_1 = a_2 = 1$. If the channel is symmetric such that $c_{11} = c_{22}$ and $c_{12} = c_{21}$, then

$$y_1 = v_1 + a v_2 + n$$
$$y_2 = a v_1 + v_2 + n$$

where $a = c_{21}/c_{11} = c_{12}/c_{22}$. For a Gaussian interference channel with $c_{11} = c_{22}$ (i.e., $n_1 = n_2 = n$), a degraded version of a received signal does not exist. Likewise, no degradation exists for a Gaussian multiple-access channel:

$$y = c_1 x_1 + c_2 x_2 + n$$

As we shall see, the fact that the Gaussian broadcast model has unequal noise terms results in a capacity region that is different in shape from the capacity regions of interference and multiple-access Gaussian models. The block forms of Gaussian broadcast and interference channel models are depicted in Fig. 3.17.

[3]Unlike the standard form defined in [12], we assume the noise PSDs are unequal and $N_1 < N_2$.

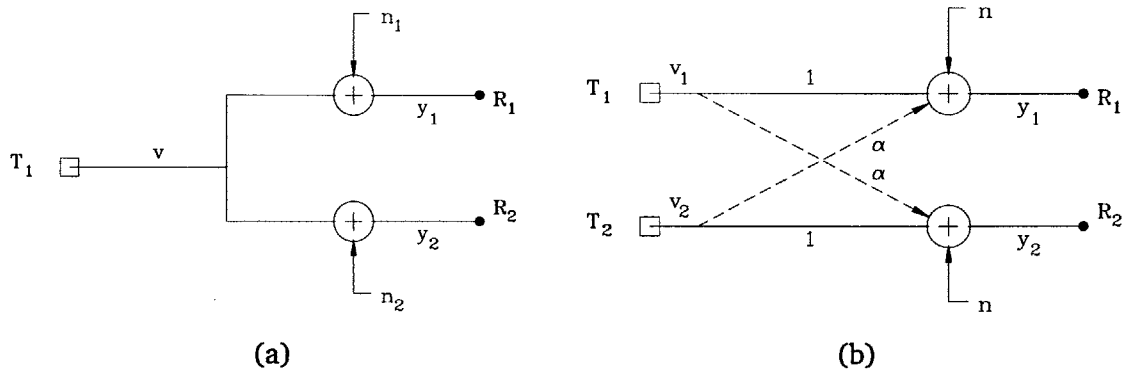



**Figure** 3.17: (a) Gaussian broadcast channel (b) Gaussian interference channel

For a two-user Gaussian composite model,

$$y_1' = c_{11}v_1 + c_{21}v_1 + n_1'$$
$$y_2' = c_{12}v_1 + c_{22}v_1 + n_2'$$

After scaling,

$$y_1^* = \frac{y_1'}{c_{11}} = v_1 + \frac{c_{21}}{c_{11}}v_2 + \frac{n_1'}{c_{11}}$$
$$y_2^* = \frac{y_2'}{c_{12}} = v_1 + \frac{c_{22}}{c_{12}}v_2 + \frac{n_2'}{c_{12}}$$

The standard form is:

$$y_1 = v_1 + a_1 v_2 + n_1$$
$$y_2 = v_1 + a_2 v_2 + n_3 \tag{3.9}$$

where $n_3 = n_2'/c_{12}$ and

$$a_1 = \left(\frac{c_{21}}{c_{11}}\right) \qquad a_2 = \left(\frac{c_{22}}{c_{12}}\right)$$

The noise terms $n_1 \neq n_3$ since one broadcast link is a degraded version of another: $c_{11} > c_{12}$. This model is depicted in Fig. 3.18. The channel is symmetric ($a_1 = a_2$) when $c_{21}c_{12} = c_{22}c_{11}$.

### 3.2.1 AWGN Only Channel

According to Shannon's capacity formula, the information rate R (in bps) through an ideal band-limited AWGN channel can be expressed as[4]:

$$R^s \leq C^s = W\log(1 + \text{SNR}) \tag{3.10}$$

[4]When expressing all Shannon capacity formulae and its variations, it is assumed that the logarithmic function is of base 2; i.e., the channel capacity is measured in bits per second or bit per channel use.

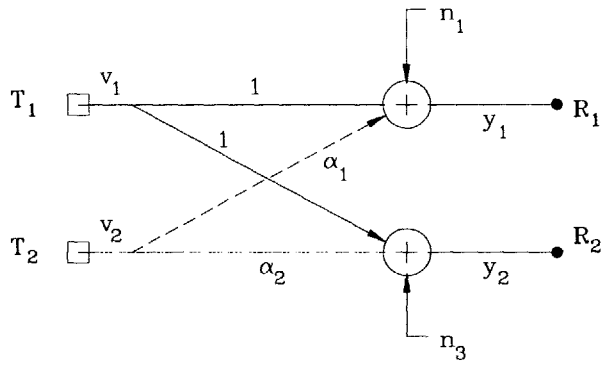



**Figure** 3.18: Gaussian composite channel

where the signal-to-noise power ratio SNR is:

$$\text{SNR} = \frac{P}{N_o W} = \frac{\mathcal{E}}{N_o T W} = \frac{2\mathcal{E}_D}{N_o} \tag{3.11}$$

The superscript in $R^s$ and $C^s$ denotes "single-user." $W$ is the channel bandwidth (in Hz), $N_o$ is the one-sided power spectral density of the Gaussian noise process, $\mathcal{E}$ is the waveform energy, and the symbol rate is $1/T$. The signal dimension is $D = 2WT$ and $\mathcal{E}_D = \mathcal{E}/D$ is the signal energy per dimension. Equality holds in eqn.(3.10) when the waveform is an infinite-length Nyquist sinc ($\sin x/x$) pulse, and its amplitude is Gaussian distributed. If pulses are non-Nyquist sinc or if the amplitudes are from a finite set, or both, the achievable rate $R^s$ is much less than the single-user capacity $C^s$. From eqn.(3.11), it can be seen that the information (bit) rate can be increased by using a higher signal energy $\mathcal{E}$ or a larger bandwidth $W$ or both. However the bandwidth $W$ is a fixed quantity for the entire network, and the signal energy $\mathcal{E}$, or equivalently, the transmit average power P is constrained such that

$$\frac{1}{n} E\Big[\sum_{i=1}^{n} v_i^2\Big] \le P$$

Each transmitted (modulated) signal can be expressed as:

$$s(t) = \sum_{k=1}^{D} A_k \varphi_k(t)$$

where $\{\varphi_k(t)\}_1^D$ is a set of orthonormal basis functions. For $D = 1$, $\varphi_1(t)$ is a sinc pulse and $s_j(t) = x_j \Pi_W(t)$ is an element of an M-ary Pulse Amplitude Modulated (M–PAM) signal set with symmetric amplitudes:

$$\mathcal{X} = \Big\{-(M-1), -(M-3), \ldots, (M-3), (M-1)\Big\} \tag{3.12}$$

**Broadcast Channel**

By relabelling the noise power densities $\sigma_1^2 = N_1$, $\sigma_2^2 = N_2$ and assuming $N_1 < N_2$, we know that the two end-points ($x$ and $y$ intercepts) of the capacity region are single-user capacities:

$$\begin{aligned} R_1 \leq C_1^s &= W \log\left(1 + \frac{P}{N_1 W}\right) \\ R_2 \leq C_2^s &= W \log\left(1 + \frac{P}{N_2 W}\right) \end{aligned} \tag{3.13}$$

By simple time-sharing (naive TDM or N-TDM), the ARR is

$$\mathcal{R}^{NT} = \left(R_1^{NT},\ R_2^{NT}\right) \leq \left(\alpha\, C_1^s,\ (1-\alpha)\, C_2^s\right)$$

The rate region is a right triangle. This policy is naive in the sense that the transmit power to each receiver is proportional to its allocated fractional time:

$$P_1 = \alpha P \qquad P_2 = (1-\alpha) P$$

For notational convenience, let $\alpha_k P$ denote the power partition of user $k$ in a K-user channel such that

$$\sum_{k=1}^{K} \alpha_k = 1$$

By using Shannon's capacity formula, it is straightforward to show that the rate region of frequency-division multiplex (FDM) is

$$\mathcal{R}^F = \left(R_1^F,\ R_2^F\right)$$

where

$$\begin{aligned} R_1^F &= \alpha_1 W \log\left(1 + \frac{P_1^F}{N_1\, \alpha_1 W}\right) \\ R_2^F &= \alpha_2 W \log\left(1 + \frac{P_2^F}{N_2\, \alpha_2 W}\right) \end{aligned} \tag{3.14}$$

Similarly for time-division multiplex (TDM)

$$\mathcal{R}^T = \left(R_1^T,\ R_2^T\right)$$

$$\begin{aligned} R_1^T &= \alpha_1 W \log\left(1 + \frac{P_1^T}{N_1 W}\right) \\ R_2^T &= \alpha_2 W \log\left(1 + \frac{P_2^T}{N_2 W}\right) \end{aligned} \tag{3.15}$$

It is easily seen that for the following power distributions

$$P_1^T = \frac{P_1^F}{\alpha_1} \qquad P_2^T = \frac{P_2^F}{\alpha_2} \tag{3.16}$$

TDM and FDM are equal in rate regions. Furthermore, if

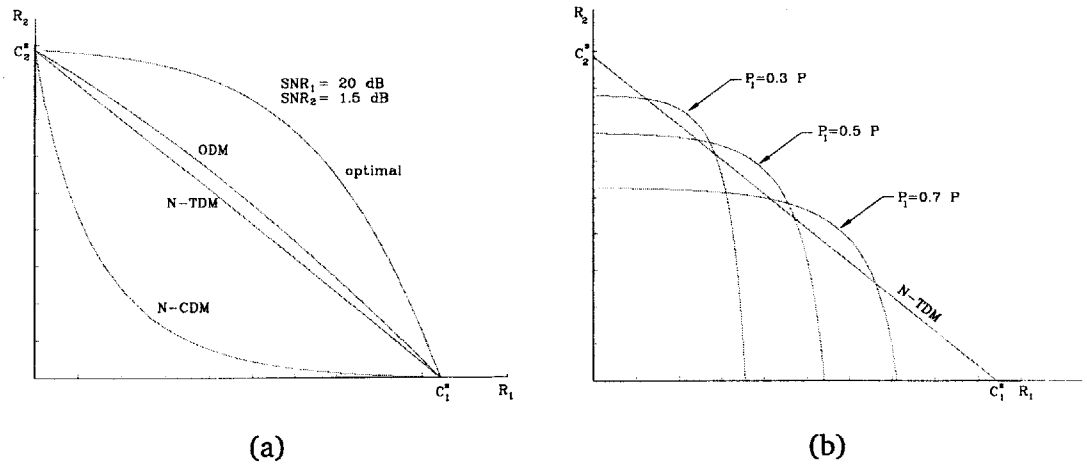
$$P_1^F = \alpha_1 P \qquad P_2^F = \alpha_2 P \tag{3.17}$$

then the rate regions of N-TDM, TDM and FDM are identical. If the transmitter has additional side information —i.e., feedback from receivers informing their noise levels— it can adjust the partitioning of its total transmit power P to maximize the rate region. By using calculus of variations, it can be shown that the optimal power levels in FDM for a given partition $\alpha$ are

$$\begin{aligned} P_1^* &= \alpha_1 P - \epsilon \\ P_2^* &= \alpha_2 P + \epsilon \end{aligned} \tag{3.18}$$

$$\epsilon = \alpha_1 \alpha_2 W(N_2 - N_1) \tag{3.19}$$

Since $N_1 < N_2$, the compensation power $\epsilon$ is positive. Intuitively, the above results state that the power level of a channel with better SNR should be reduced slightly by $\epsilon$ compared to naive power distribution. The rate regions for N-TDM, FDM and TDM are plotted in Fig. 3.19(a). Since FDM and TDM have the same rate region when the optimal power distribution is chosen, we call them collectively as *orthogonal-division multiplex* (ODM) schemes. A simple proof of the dominance of ODM over N-TDM is given in Appendix 3B. It is worth noting that when ODM deviates from its optimal power distribution —e.g., due to error in feedback information— its achievable rate-pair is smaller than that of N-TDM in rate regions where power partition is sub-optimal. This is illustrated in 3.19(b). Note that with sub-optimal power partition, its end points are much less than single-user capacities.




(a) Two-user Gaussian broadcast channel rate region

**Figure** 3.19: (b) rate regions for orthogonal channel multiplexing without transmit power adaptation

In the optimal policy there is no partitioning of time frame or frequency band. The transmitter broadcasts a superimposed signal. Note, however, that the total power is partitioned —in our two-user example, into $\alpha_1 P$ and $\alpha_2 P$. The receiver with the worst SNR (user 2 in our example) treats other users' information signal as noise. Its capacity is

$$C_2 = W \log\left(1 + \frac{\alpha_2 P}{N_2 W + \alpha_1 P}\right)$$

The receiver (user 1 in our example) with the best SNR decodes all other users' information. For example, it is able to decode user 2's signal (by treating its own signal as noise) with better SNR since $N_1 < N_2$:

$$D_2 \triangleq W \log\left(1 + \frac{\alpha_2 P}{N_1 W + \alpha_1 P}\right)$$

It then re-encodes, remodulates and subtracts other users' signals from the original superimposed signal. What remains is the receiver's own information signal plus additive Gaussian noise. Its capacity is

$$C_1 = W \log\left(1 + \frac{\alpha_1 P}{N_1 W}\right)$$

The capacity-pair of the optimal policy can be written as

$$\mathcal{C}^o = (C_1, C_2) \tag{3.20}$$

For a general K-user channel, assuming $N_k < N_{k+1}$ for all k, the capacity of the $k^{th}$ user is

$$C_k = W \log\left(1 + \frac{\alpha_k P}{N_k W + \sum_{l<k} \alpha_l P}\right) \tag{3.21}$$

It can be proven that this optimal policy dominates over ODM schemes[5].

If each receiver treats other users' signals as noise, we obtain the naive code-division multiplexing (N-CDM) scheme. Its rate pair is

$$\begin{aligned} R_1^{NC} &= W \log\left(1 + \frac{\alpha_1 P}{N_1 W + \alpha_2 P}\right) \\ R_2^{NC} &= W \log\left(1 + \frac{\alpha_2 P}{N_2 W + \alpha_1 P}\right) \end{aligned} \tag{3.22}$$

The above ARR still applies to superposition coding *without* successive interference cancellation. Therefore, superposition coding alone at the transmitting end is not sufficient for optimality. We now state several important requirements for optimal encoding and decoding in AWGN broadcast channel:

[5]Please refer to Reference [7] for a proof.

- The optimal scheme —commonly known in the literature as code-division multiplex— is actually a combination of non-orthogonal superposition code-division multiplexing and successive interference cancellation (SIC) of degraded signals. If the channel is not degraded, then SIC is not possible, and the resulting encoding-decoding scheme is no longer optimal.

- For error-free SIC, every receiver must have complete knowledge of superior (those with lower noise) and degraded (those with higher noise) signals relative to its own. Since peer-to-peer control signalling is not allowed, the noise level $N_k$ (or equivalently, the propagation power loss in channel k) must first be relayed to the base transmitter from each receiver[6], and then relayed again to all receivers. We coin this mechanism "double-feedback" information. The accuracy and periodic updating of double-feedback information becomes questionable when the channel suffers from fading or when the receivers are non-stationary or both. It can be shown that the ARR shrinks considerably away from capacity when decoding errors are made in SIC due to incorrect sorting of noise levels.

- In N-TDM power partitioning is straightforward: If requested rates are $(\alpha_1 C_1^s, \alpha_2 C_2^s)$, the power distribution is $(\alpha_1 P, \alpha_2 P)$. In TDM and FDM, this power distribution is corrected with a compensation power $\epsilon$ such that the optimal power distribution is

$$P_1^* = \alpha_1 P - \epsilon \qquad P_2^* = \alpha_2 P + \epsilon$$

  In the optimal scheme power partitioning is not straightforward. From eqn. 3.21 we can immediately deduce power allocation for user k requesting rate $C_k$ as

$$\alpha_k P = \left(2^{C_k/W} - 1\right)\left(N_k W + \sum_{l<k} \alpha_l P\right)$$

  We first compute power allocation for the channel with the best SNR, then the channel with the second best SNR and so on in a successive sequential order.

In short we stress that error-free implementation of the optimal multiplexing scheme is rather difficult. Moreover, to the best of our knowledge, a practical superposition coding scheme has yet to be discovered[7].

[6]It is generally assumed that only the mobile can accurately measure the channel gain (propagation path loss) in the forward link. Channel state information measured at the base site based on uplink signal is not reliable for base-to-mobile link quality since most cellular networks use separate frequency bands for uplink and downlink (i.e., frequency-division duplexing).

[7]Recently, multi-resolution coding has been proposed for high-definition broadcast signalling and imaging. This is a technique where satellite channels carry extra information for increased definition such that channel with higher SNR receive higher-resolution video images. This is similar, but not equivalent, to superimposition coding where all channels carry independent information for different sinks.

S. Verdu showed in [151] that in a time-synchronous M-user discrete-time Gaussian multiple-access channel, the CDMA sum-capacity is

$$\frac{1}{2}\log\left[\det\left(\mathbf{I}_M + \sigma^{-2}\mathbf{H}_M\right)\right] \tag{3.23}$$

where $\mathbf{I}_M$ is a $(M \times M)$ identity matrix, $\sigma^2$ is the two-sided power spectral density of the additive white Gaussian noise, and det[·] denotes the determinant of a matrix. $\mathbf{H}_M$ is the $(M \times M)$ cross-correlation matrix, with its entries defined as

$$H_{ij} = \int s_i(t)s_j(t)\,dt$$

where $s_i(t)$ and $s_j(t)$ are the signature waveforms assigned to users i and j, respectively. (See Chapter 4 for further details.) It is known that L mutually orthogonal waveforms can be generated at the expense of expanding the transmission bandwidth by L[8]. Assume we use M out of L orthogonal waveforms, all with equal energy $\mathcal{E}$, then

$$H_{ij} = \begin{cases} \mathcal{E}_D & i = j \\ 0 & i \neq j \end{cases}$$

Note that since thee waveforms are orthogonal, the signal energy $\mathcal{E}$ in each channel is equal to the energy per dimension $\mathcal{E}_D$. Substituting above result in eqn. (3.23), the sum-rate of orthogonal CDM in a Gaussian broadcast channel is

$$\frac{M}{2}\log\left(1 + \frac{\mathcal{E}_D}{\sigma^2}\right)$$

If the channel is continuous-time with bandwidth $W$ Hz, using the relation in eqn. (3.11):

$$MW\log\left(1 + \frac{P}{N_o W}\right)$$

where P is the power in each orthogonal channel. If the transmitter broadcasts L parallel orthogonal channels, and user 1 decodes M out of L orthogonal channels jointly, then its achievable rate is

$$R_1^C = \frac{MW}{L}\log\left(1 + \frac{P_1^C}{N_1 W}\right) \tag{3.24}$$

Note the normalization by L since orthogonal signalling expands the bandwidth by L. Similarly, if user 2 decodes $(L - M)$ out of L orthogonal channels jointly, then its achievable rate is

$$R_2^C = \frac{L-M}{L}W\log\left(1 + \frac{P_2^C}{N_2 W}\right) \tag{3.25}$$

[8]We are assuming transmit pulses are sinc waveforms with an ideal brick-wall frequency response. It is then obvious that L orthogonal waveforms can be generated by frequency translation of $L - 1$ sinc pulses.

By setting

$$\alpha_1 = \frac{M}{L} \qquad \alpha_2 = \frac{L-M}{L}$$
$$P_1^C = P_1^T = \frac{P_1^F}{\alpha_1} \qquad P_2^C = P_2^T = \frac{P_2^F}{\alpha_2}$$

Compared to eqn. (3.14) wee see that orthogonal CDM has the same achievable rate region as TDM and FDM[9]. In words the above equation states that to support rate $R_i^C$ to user i in O-CDM, the transmitter must send L parallel orthogonal, equal energy channels such that $\alpha_i L$ of them carry information desired by user i. The value of L depends on the set $\{\alpha_i\}$. To support any rate pair $(R_1, R_2)$ the length L of orthogonal codes (i.e., the dimension of an orthogonal matrix) must be large enough that $\alpha_i L$ is an integer for every i. Therefore, every for the two-user case, the number of parallel orthogonal channels L may be very large. A fraction of these L channels are assigned to user 1, with the remaining assigned to user 2. As we shall introduce in Chapter 6, this is called parallel channel, single gain (i.e., a single spreading gain for all channels) CDM, or PC-SG CDM. It is also possible to interpret $R_i^C$ as an achievable rate in O-CDM by sending a single code (orthogonal to codes assigned to other receivers) of power $P_i^C$ and bandwidth expansion factor $1/\alpha_i$. In this case, every receiver has a different spreading gain. This is called single-channel, reduced-gain CDM, or SC-RG CDM. Both variable-rate O-CDM schemes are depicted in Figure 3.20 for $M = 2$ and $L = 4$.

### Interference Channel

The capacity of interference channel in AWGN is not known. We next summarize achievable rate regions that have been reported in the literature. For a two-user system, the x– and y–intercepts of the capacity region are equal to single-user capacities:

$$R_1 \le C_1^s \qquad R_2 \le C_2^s$$

We also know that the ARR includes any rate pair bounded by the time-sharing[10] line that connects $C_1^s$ and $C_2^s$:

$$(R_1, R_2) \le \left(\alpha\, C_1^s,\ (1-\alpha)\, C_2^s\right) = \mathcal{R}^{NT}$$

By optimizing power partitions, it can be shown that the ARR of orthogonal multiplexing (such as FDM or TDM) dominates over the ARR of naive-TDM —analogous to the broadcast channel

[9]Other researchers reach the same conclusion by arguing that time, frequency and code dimensions are transferable; i.e., given the same amount of bandwidth, equal number of orthogonal codes can be generated as time-slots or frequency bands.

[10]Since the transmitters are physically separated in an interference channel, it may not be possible to establish two-way time-division multiplexed channels without overlap. We can, however, maintain non-colliding outdoing and incoming time frames by adding guard intervals. In FDM, orthogonality between outgoing and incoming frequency bands is easily preserved as long as the Doppler shift is negligible.

model. Therefore, the capacity region also includes $\mathcal{R}^F$. We next derive several achievable rate-pairs based on the standard form of eqn.(3.8). Let $N_1$ and $N_2$ denote respective PSDs of Gaussian noise samples $n_1$ and $n_2$, and assume $N_1 < N_2$.

$$y_1 = v_1 + a_1 v_2 + n_1$$
$$y_2 = a_2 v_1 + v_2 + n_2$$

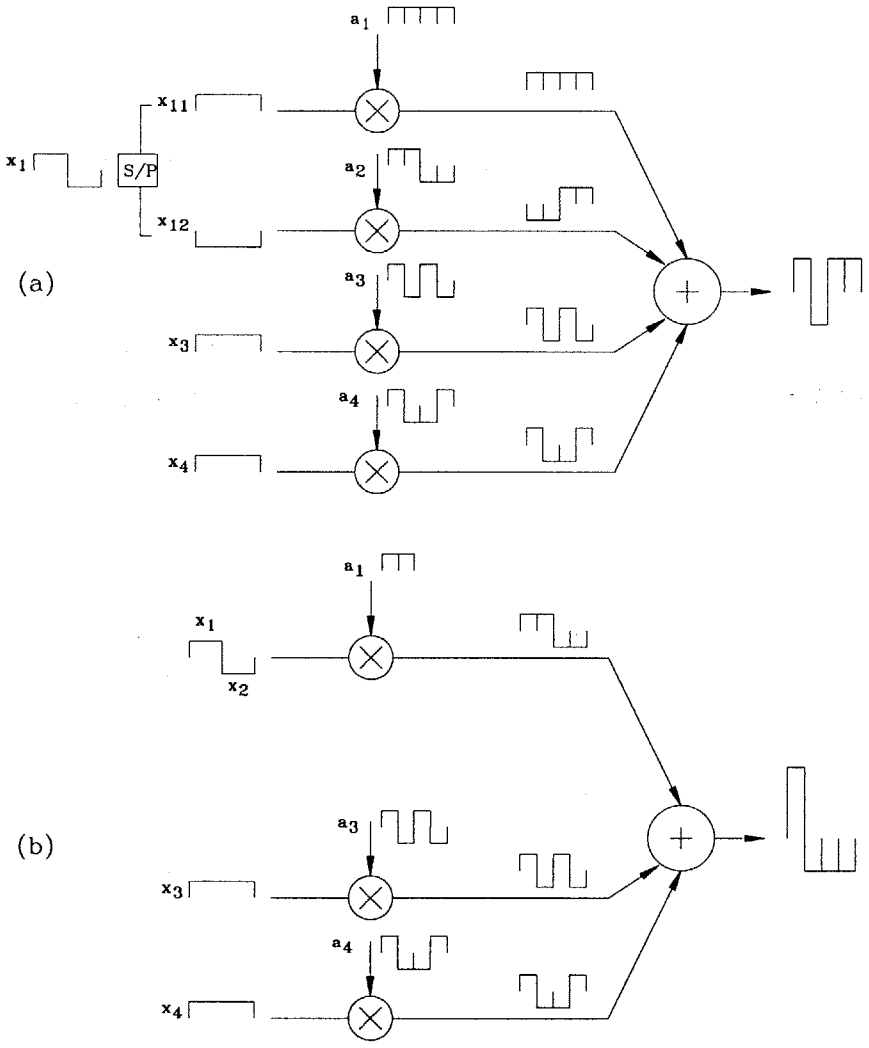



**Figure 3.20:** (a) Parallel-Channel, Single-Gain Transmission in O-CDM for 2 of 4 code channels. The input data is split between the top two code channels.
(b) Single-Channel, Reduced-Gain Transmission in O-CDM where a reduced gain channel replaces the top 2 of 4 code channels.

We first define the following rates:

$$R_1^{NC} \triangleq W\log\left(1+\frac{P_1}{N_1W+a_1^2P_2}\right)$$
$$R_2^{NC} \triangleq W\log\left(1+\frac{P_2}{N_2W+a_2^2P_1}\right)$$
$$D_1 \triangleq W\log\left(1+\frac{a_2^2P_1}{N_2W+P_2}\right)$$
$$D_2 \triangleq W\log\left(1+\frac{a_1^2P_2}{N_1W+P_1}\right)$$

We can now state —without constructive proofs based on information-theoretic coding and decoding— that the following rate pairs are contained in the capacity region:

- *Interference as Noise*: By treating interference as undesirable noise, the rate region is equal to that of naive code-division multiplexing $\mathcal{R}^{NC}$:

$$\mathcal{R}^{NC} = (R_1^{NC}, R_2^{NC})$$

- *Channel Swapping*: If each receiver reverses the roles of principal and interference channels, each can decode other user's signal while treating its own signal as interference. This rate region is equal to $\mathcal{D}$:

$$\mathcal{D} = (D_1, D_2)$$

  Of course, each receiver can do better. If each can decode other user's data bits, these bits can be re-encoded, remodulated and then subtracted out from received signal. The resulting signal consists only of desirable signal plus Gaussian noise. We are already familiar with this detection strategy; it is successive interference cancellation (SIC).

- *Successive Interference Cancellation*: If user 2's information rate is $D_2$, then receiver 1 can apply SIC to remove user 2's signal from the compound signal and decode user 1's signal perturbed only by Gaussian noise $n_1$. Its maximum AR is then equal to $C_1^s$. At the same time, receiver 2 can decode information at rate $R_2^{NC}$ by treating user 1's signal as noise. If $R_2^{NC} > D_2$, transmitter 2 must still convey information at rate $D_2$ to its receiver (although it can decode reliably at the higher rate of $R_2^{NC}$) such that the resulting interference can be removed by receiver 1. If $R_2^{NC} < D_2$, then receiver 2 can only decode at maximum rate $R_2^{NC}$ without applying SIC. Similarly, If user 1's information rate is $D_1$, the maximum AR of user 2 is $C_2^s$ and user 1's AR is the lesser of $R_1^{NC}$ and $D_1$. Hence, the ARR —by applying

superposition coding at transmitters[11] and SIC at each receiver site— is:

$$\mathcal{R}_1^o = \left(C_1^s, T_2\right) \qquad T_2 = \min\{R_2^{NC}, D_2\}$$
$$\mathcal{R}_2^o = \left(T_1, C_2^s\right) \qquad T_1 = \min\{R_1^{NC}, D_1\}$$

Note the subtlety in this proof. In general if given power $P_2$, transmitter 2 can send information to its receiver at maximum rate, say $R_2$. On the other hand, user 1 can only decode a fraction of that information (of rate $D_2$) due to propagation loss $a_1^2$ and its noise level $N_1$. By applying SIC, it can only remove that portion of user 2's interference from its signal, resulting in maximum achievable rate equal to $(C_1^s - D_2)$. Here, we are proposing a different strategy. Transmitter 2 sends its information of rate $D_2$ at power $P_2$ (much more than the required minimum for arbitrarily small $P_e$) such that *all* its information can be decoded by receiver 1, resulting in its achievable rate equal to $C_1^s$. We follow this second approach since we are interested in boundary ARR involving $C_1^s$ and $C_2^s$.

- *Noisy Interference*: Thus far we have assumed that interference channels exist (i.e., $a_1 > 0$, $a_2 > 0$) without taking into account their respective magnitudes. We also have not considered the effects of unequal noise levels ($N_1 < N_2$). We know by channel swapping the achievable rate pair $(R_1, R_2)$ is

$$R_1 = D_2 \triangleq W \log\left(1 + \frac{a_2^2 P_1}{N_2 W + P_2}\right)$$
$$R_2 = D_1 \triangleq W \log\left(1 + \frac{a_1^2 P_2}{N_1 W + P_1}\right)$$

If the interference channels are very noisy such that

$$a_2^2 \geq \frac{N_2 W + P_2}{N_1 W} \qquad \text{and} \qquad a_1^2 \geq \frac{N_1 W + P_1}{N_2 W}$$

we immediately see that (by rearranging)

$$\frac{a_2^2 P_1}{N_2 W + P_2} \geq \frac{P_1}{N_1 W} \qquad \text{and} \qquad \frac{a_1^2 P_2}{N_1 W + P_1} \geq \frac{P_2}{N_2 W}$$

Therefore, both receivers can decode their information at the maximum achievable rates $\mathcal{C}^s = (C_1^s, C_2^s)$. This odd situation where interference does not reduce single-user capacities does not occur in broadcast or multiple-access channel models. If only one of above two conditions holds, the ARR is no longer equal to $\mathcal{C}^s$. For example, if only the first condition holds

$$\frac{a_2^2 P_1}{N_2 W + P_2} \geq \frac{P_1}{N_1 W}$$

[11] In reality, superposition coding may not be feasible in the interference model since transmitters are geographically separated and joint encoding of independent messages is necessary for the generation of superposition codes.

it is obvious that

$$\frac{a_2^2 P_1}{N_2 W + P_2} > \frac{P_1}{N_1 W + P_2}$$

thus resulting in an achievable rate pair of only $(R_1^{NC}, C_2^s)$. We also notice another peculiarity: If the channel gain $a_2$ is amplified such that

$$a_2^2 \geq \frac{N_2 W + P_2}{N_1 W} > 1$$

then we see immediately that $D_1 \geq C_1^s$. Similarly, $D_2 \geq C_2^s$ when

$$a_1^2 \geq \frac{N_1 W + P_1}{N_2 W}$$

In this case it is not even necessary that the channel gain $a_1$ be amplified. However, we know that interference cannot increase single-user capacities since information contents are mutually independent; thus the ARR is

$$(S_1, S_2) \qquad S_x = \min\{C_x^s, D_x\}$$

- *Amplified Interference*: As defined earlier, an interference channel is lossless if $a_1 = a_2 = 1$. It has amplified amplitude gains if $a_1 = a_2 > 1$. When comparing naive CDM versus channel swapping, the following holds depending on the values of $a_1$ and $a_2$:

$$R_1^{NC} \lesseqgtr D_1 \qquad\qquad R_2^{NC} \lesseqgtr D_2$$

For lossy channels ($0 < a_1, a_2 < 1$), $R_1^{NC} > D_1$ since $N_1 < N_2$. This inequality still holds when the channel is lossless or amplified over a certain range. Similar but different conclusions can be drawn for $R_2^{NC}$ and $D_2$: When the channel is not lossy ($a_1, a_2 \geq 1$), $R_2^{NC} < D_2$.

All five possible shapes of the ARR of a two-user Gaussian interference channel are plotted in Fig. 3.21. Regardless of the channel gains, x– and y– intercepts are single user capacities $C_1^s$ and $C_2^s$. From these intercepting points, there emerges two straight-line AR boundary regions connecting coordinates $(C_1^s, 0)$ with $\mathcal{R}_1^o$ —and similarly, $(0, C_2^s)$ with $\mathcal{R}_2^o$. Based on this frame, we next elaborate the resulting shapes:

1. The simplest shape is the rectangle as depicted in Fig. 3.21(a), a very noisy interference model where each component channel maintains single-user capacities with the aid of superposition coding and SIC. For comparison the ARR of FDM with optimal power partitioning is also included in the plot. Any orthogonal scheme such as FDM has an ARR that is independent of channel gains. Thus, its shape and size remains the same in all plots.

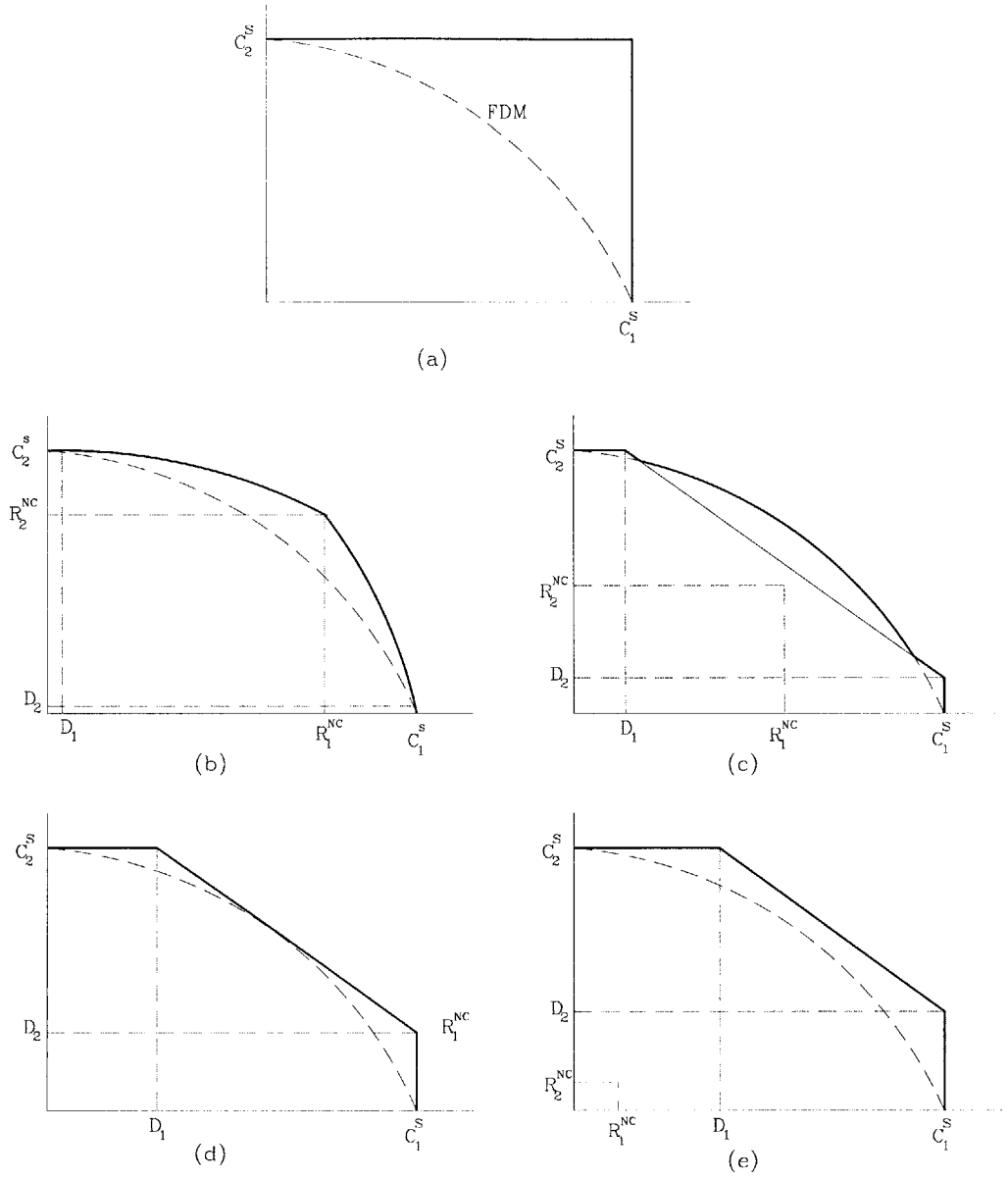



**Figure 3.21:** Three different shapes for the Gaussian interference channel ARR.
(a) very noisy interference channel where $D_1 = C_1^s$, $D_2 = C_2^s$
(b) very lossy interference channel where $D_1 \ll R_1^{NC}$, $D_2 \ll R_2^{NC}$
(c) moderately lossy interference channel where $D_1 < R_1^{NC}$, $D_2 \gtrless R_2^{NC}$
(d) lossless interference channel where $a_1 = a_2 = 1$
(e) amplified interference channel where $a_1 > 1$, $a_2 > 1$

2. For very low lossy channels ($a_1, a_2 \ll 1$), $D_x \ll R_x^{NC}$, $x = 1, 2$. Therefore, the straight-line boundary is very small. The low loss channel model is depicted in Fig. 3.21(b). In this case it is readily shown that the rate pair $\mathcal{R}^{NC}$ of naive CDM lies outside the ARR of FDM. By optimal power partitioning, the region enclosed by the curve connecting $\mathcal{R}_1^o$ and $\mathcal{R}^{NC}$ —and similarly, the curve connecting $\mathcal{R}_1^o$ and $\mathcal{R}^{NC}$— can be constructed. We note that this rate region dominates over the ARR of ODM for all possible rate pairs. Compared to the broadcast channel model, this result is quite interesting since it shows that N-CDM has a larger ARR than ODM in a Gaussian interference channel.

3. For moderately lossy channel ($a_1, a_2 < 1$), $D_1 < R_1^{NC}$ and $D_2 \lesseqgtr R_2^{NC}$. Since $\mathcal{D}$ is further away from the origin, the straight-line boundaries are more pronounced. This situation is shown in Fig. 3.21(c). It can be shown that rate pair $\mathcal{R}^{NC}$ lies inside the ARR of FDM. Hence, its ARR is smaller than that achieved by time-sharing between end points $\mathcal{R}_1^o$ and $\mathcal{R}_2^o$. Based on these observations, an irregular ARR results when channel loss is moderate. It is seen from the plot that FDM dominates over N-TDM and N-CDM in regions where $R_1 \approxeq R_2$. At boundary regions where $R_1 \gg R_2$ or $R_1 \ll R_2$, N-TDM combined with SIC dominates.

4. For a lossless channel ($a_1 = a_2 = 1$), both receivers detect the same information signal despite different noise levels. (In the standard form as detailed in [12], Gaussian noise levels are normalized by adjusting transmit power constraint such that $N_1 = N_2 = 1$.) Therefore, we conclude that receiver 2 is redundant. By applying N-CDM with SIC and time sharing, this model is equivalent to optimal detection in multiple-access channel. It is no surprise that the resulting shape of the ARR —as depicted in 3.21(d)— is a pentagon, the shape of multi-access channel capacity.

5. In a lossless channel, the pentagon ARR of N-CDM with SIC and time sharing dominates over ODM for all rate pairs except at the bisecting point of the time sharing line where the ARR of ODM touches the pentagon. Once the channels are amplified ($a_1, a_2 > 1$), the rate pair $\mathcal{D}$ becomes significant and the time-sharing curve moves away from the ODM curve. This is illustrated in 3.21(e).

To summarize, the capacity region of a Gaussian interference channel is not known. For a lossy interference channel, neither N-CDM or ODM dominates over the other for all values of channel gains. The straight-line boundary region achieved by superposition coding and SIC offers little gain when the channel is very lossy. Without SIC, naive CDM dominates over ODM only when the channel is very lossy; otherwise, its ARR is much smaller than that of ODM. When studying cellular radio models where channel gains are lossy due to path loss, it is reasonable to assume that we will not encounter lossless and amplified channels.

### Composite Channel

From eqn. (3.9) the standard form of a broadcast-plus-interference composite channel model can be expressed as

$$y_1 = v_1 + a_1 v_2 + n_1$$
$$y_2 = v_1 + a_2 v_2 + n_2$$

The outputs $y_1$ and $y_2$ represent received signals at two different locations detecting the same broadcast signal $v_1 = V_{11} + v_{12}$, whereas the remaining broadcast signal $v_2$ is interference. For each output both interference channel gains and noise variances are normalized such that the principal channel gains are unity. Applying multiplexing techniques used in broadcast and interference models, we can deduce achievable rate pairs. For example, single-user capacities $C_1^s$ and $C_2^s$ can be achieved if the interfering cell transmits at channel swapping rates

$$D_2 \triangleq W \log\left(1 + \frac{a_1^2 P}{N_1 W + P}\right)$$
$$D_1 \triangleq W \log\left(1 + \frac{a_2^2 P}{N_2 W + P_1}\right)$$

then via SIC, the target receiver can remove interference terms. Hence, the following rate pairs are achievable:

$$(C_1^s, 0) \qquad (0, C_2^s)$$

when the rates for interference channel are $D_1$ and $D_2$. By time sharing with optimal partitioning, we can achievable rate sets equivalent to the ARR of ODM. However, such rate distribution violates the "symmetry" assumption of the cellular broadcast model where each cell with equal bandwidth and power transmits at the same sum-rate. It also violates our assumption of single-user detection since the target receiver applies SIC. In the above example we are considering an extreme case where a neighboring cell reduces its rate pair to $\mathcal{D}$ to support maximum AR pair at the target cell. In the following we derive rate pairs that are achievable autonomously; i.e., without affecting the AR pair of neighboring interference site.

- By simple time sharing, the following rate pair is achievable:

$$R_1^{NT} = \alpha_1 W \log\left(1 + \frac{P}{N_1 W + a_1^2 P}\right)$$
$$R_2^{NT} = \alpha_2 W \log\left(1 + \frac{P}{N_2 W + a_2^2 P}\right)$$

- By applying optimal power partitioning, the following rate pair is achievable:

$$R_1^F = \alpha_1 W \log\left(1 + \frac{P_1}{N_1 \alpha_1 W + \alpha_1 a_1^2 P}\right)$$
$$R_2^F = \alpha_2 W \log\left(1 + \frac{P_2}{N_2 \alpha_2 W + \alpha_2 a_2^2 P}\right)$$

- By treating other user's signal as noise, the following rate pair is achievable:

$$R_1^{NC} = W \log\left(1 + \frac{P_1}{N_1 W + P_2 + a_1^2 P}\right)$$
$$R_2^{NC} = W \log\left(1 + \frac{P_2}{N_2 W + P_1 + a_2^2 P}\right)$$

- Similar to superposition coding and SIC, by removing other user's signal after detection and re-modulation, the following rate pair is achievable:

$$R_1^o = W \log\left(1 + \frac{P_1}{N_1 W + a_1^2 P}\right)$$
$$R_2^o = W \log\left(1 + \frac{P_2}{N_2 W + P_1 + a_2^2 P}\right)$$

The ARR of a AWGN composite channel is plotted in Fig. 3.22.

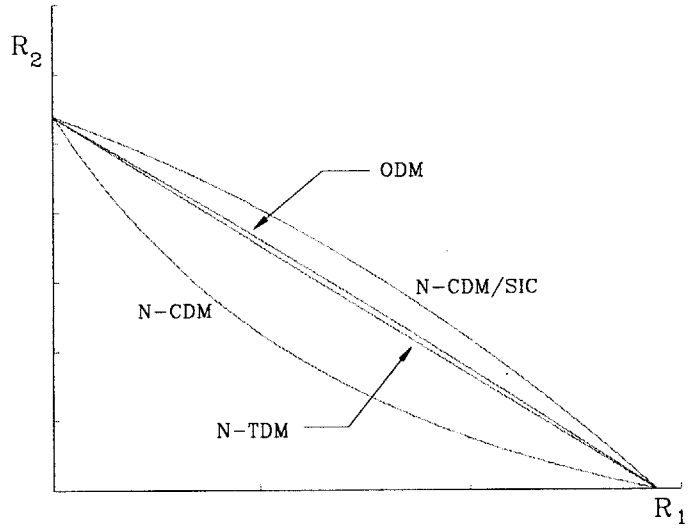



**Figure** 3.22: Achievable rate region of Gaussian composite channel $a_1^2 = 0.3$, $a_2^2 = 0.7$

We notice that various ARR of the composite channel are very similar to those of Gaussian broadcast channel. Therefore, we conclude that for intra-cell channel multiplexing, orthogonal multiplexing is the preferred option.

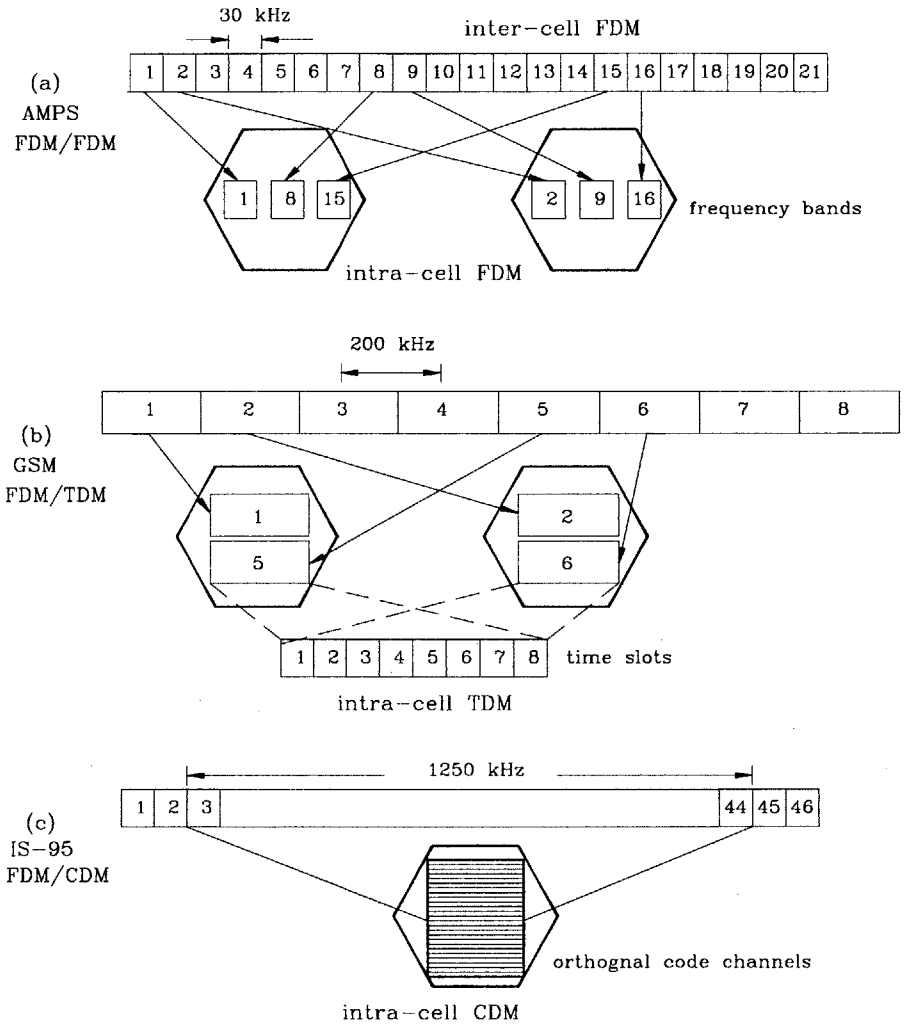
## 3.3 Channel Multiplexing in Cellular Environment

In the previous sections we showed that superposition coding with successive interference cancellation is the optimal policy in a broadcast setting. Due to its inherent encoding/decoding complexity —plus the necessity for double feedback information— sub-optimal policies that maintain mutual orthogonality among parallel user channels are preferred. All three orthogonal schemes based on time-, frequency- and code-division were shown to have equal simultaneously achievable rate regions. Naive CDM is the least preferred scheme due to its convex ARR. In the interference model, the optimal policy is not known. *Our results give no preference in the selection of a channel multiplexing scheme in the AWGN interference model.* However, we notice that the ARR of an orthogonal multiplexing scheme is close to other ARR based on superposition coding with SIC, N-CDM and N-TDM except for the first case of very noisy interference channel. In cellular radio environment where interference from neighboring co-channel[12] cells are lossy due to propagation loss and fading, it is plausible to assume that such noisy condition is not very likely. At this point is is unclear what the most efficient resource allocation policy is for the cellular broadcast channel. As we recall, this is a special type of composite channel consisting of both broadcast and interference components. In a multi-cell environment, our choice of resource allocation policy must be efficient not only within a cell but also among cells for the entire network. When selecting a resource allocation policy for a cellular network, we must explicitly state respective channel multiplexing schemes for both inside a cell (among users) and inside the network (among cells). A multiplexing scheme inside a cell is known as the *intra-cell* assignment while multiplexing among cells in a cluster is called *inter-cell* assignment. Some examples are FDM/TDM (FDM among cells and TDM within a cell), FDM/FDM and FDM/CDM. The European GSM standard —which is commonly known as a TDMA-based cellular network— overlays TDM frames over frequency-division multiplexed bands. The very first U.S. analog cellular standard AMPS uses FDM/FDM and Qualcomm's IS-95 is a hybrid FDM/CDM based network, which during its infancy replaced several inter-cell FDM channels for a single CDM band supporting many parallel orthogonal code channels. All three hybrid channel multiplexing schemes are illustrated in Fig. 3.23. Other possible hybrid schemes not shown in the figure are

- TDM schemes (TDM/FDM, TDM/TDM and TDM/CDM)
- hybrid CDM[13] schemes (CDM/CDM, CDM/TDM and CDM/FDM)

[12]For definitions and in-depth discussions on co-channel cells, reuse number, cluster and other cellular network related terms, please refer to Appendix 3A.

[13]As explained in the introductory chapter, CDM refers to direct-sequence or phase-coded CDM where time epoch synchronization at the transmitter and coherent detection at the receiver are mandatory in maintaining mutual channel orthogonality. Other non-coherent spread-spectrum modulation schemes such as frequency hopping are not considered as CDM.




Hybrid (inter-cell/intra-cell) channel multiplexing
**Figure 3.23:** schemes in commercial cellular networks;
(a) FDM/FDM (b) FDM/TDM (c) FDM/CDM

In a single-cell scenario, time synchronization among parallel transmit channels affords mutual orthogonality in the forward link. However, maintaining orthogonality among any two channels originating from different base sites (i.e., inter-cell orthogonality) is much more difficult, if not impossible. Among the three orthogonal division multiplexing schemes, FDM is the easiest since its orthogonality constraint is based on the assignment of non-overlapping frequency bands as channels. In fact, in an inter-/intra-cell channel multiplexing scheme based on FDM/FDM, the total network bandwidth $W_{tot}$ is partitioned equally into $N_r$ subsets, where each subset (of bandwidth $W_c = W_{tot}/N_r$) consists of a group of frequency bands that is assigned to a co-channel cell in a cluster. In FDM/FDM inter-cell orthogonality is disrupted only

if the channel induces spectral broadening due to Doppler shift. From eqn. (2.22) we know that for an RF signal at 2 GHz and mobile speed of 100 km/hr, the maximum Doppler shift is 185 Hz. By inserting a guard band around each FDM channel to compensate for maximum frequency drift, inter-cell orthogonality can still be maintained. In AMPS standard where each channel bandwidth is 30 kHz, a guard band of 370 Hz accounts for 1.2 % overhead.

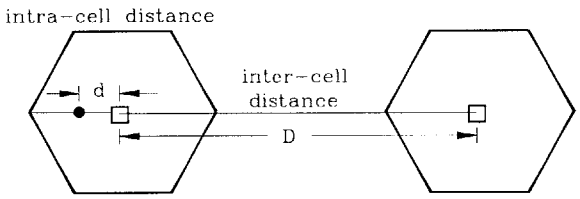



**Figure** 3.24: Maximum relative propagation delay between target and interference signals recorded at the target mobile in a cellular network

In inter-cell TDM or O-CDM, the guard times must be wide enough to compensate for maximum relative propagation delay. From Fig. 3.24 we see that the relative delay $\tau$ is maximum when it equals D; i.e., when the mobile is located anywhere along the line linking the cell center to the furthest vertex from an interfering cell. (D is the distance of propagation between the target and interference cell sites; d is the distance between target mobile and cell site.) In CDM where reuse distance is small due to spectrum spreading, $\tau_{max}$ is relatively small compared to FDM/TDM where larger reuse numbers are required for high SIR. For example, in CDM with $N_r = 1$ (i.e., co-channels are located in immediately neighboring first-tier cells) $\tau_{max} = \sqrt{3}R$. Likewise, in FDM/TDM with $N_r = 4$, $\tau_{max} = 2\sqrt{3}R$. Numerically, for a cell radius of 1 km, $\tau_{max} = 5.77\ \mu$ sec for $N_r = 1$ and $\tau_{max} = 11.5\ \mu$ sec for $N_r = 4$. In GSM (with $N_r = 4$) where each time slot equals 577 $\mu$ sec, a guard time of 23 $\mu$ sec accounts for 4 % overhead. When $N_r = 4$ and the cell radius is 5 km, the GSM guard time increases to 152 $\mu$ sec, a whopping overhead of 26% to maintain inter-cell orthogonality. In IS-95 and 3G UMTS O-CDM systems, the chip durations are 0.814 and 0.244 $\mu$ sec, respectively. Inter-cell code synchronism[14] is impossible to maintain unless the cell radius is 35 m or less. Based on these observations, we conclude that O-CDM and TDM are not viable candidates for inter-cell channel multiplexing. It is, however, possible to use N-CDM as an inter-cell channel multiplexing policy since the cross-correlation between any two co-channel signals is measured in terms of cross-correlation properties of embedded signature sequences, and not on time epoch alignment of signature waveforms. Therefore, if the network design paradigm is to maintain mutual orthogonality among inter-cell co-channels, then FDM is the only viable option, whereas for non-orthogonal inter-cell multiplexing, N-CDM can be considered.

[14]See page 171 for definition and classification of various time alignment options.

## 3.4 Cellular Radio Capacity

In an orthogonal inter-cell multiplexing policy such as FDM/ODM, co-channel interference (CCI) is suppressed via propagation loss, whereas in non-orthogonal inter-cell multiplexing policy based on N-CDM, CCI is suppressed by low normalized cross-correlation of signature waveforms for non-zero delay offsets. The capacities of FDM/ODM and N-CDM/O-CDM based cellular systems can be measured and compared in two different ways. In the first based on cellular radio engineering methods, we derive *cellular radio capacity* using cellular hexagonal geometry, average SIR and propagation path loss $L_p$ as relevant and sufficient parameters. In the second based on information-theoretic rules, we compute the simultaneously achievable rate regions —which we coin as *cellular channel capacity*— of FDM/CDM and N-CDM/O-CDM cellular broadcast networks by assuming ideal pulse shapes and ideal AWGN channel response. We study cellular capacity first.

### 3.4.1 Average Carrier-to-Interference Ratio

In cellular radio engineering we wish to measure the strength of average received power of the desired RF signal against interference power from neighboring cell sites and receiver front-end thermal noise. This benchmark is measured in terms of carrier-to-interference plus noise power ratio:

$$C/(I+N) = CINR = \frac{\text{carrier power of desired RF signal}}{\text{interference power plus noise power}} \tag{3.26}$$

In this simplified model the type of carrier or data modulation is insignificant; it is generally assumed that the desired signal as well as interference signals are unmodulated carrier tones. This assumption simplifies the thorny issue of (non-ideal) frequency response of power spectral density of wideband RF signal when it is data and carrier modulated, and the channel is frequency selective. For practical reasons, CINR is approximated as carrier-to-interference ratio C/I (CIR) by ignoring Gaussian noise power in an interference-limited topological environment. We will also use this approximation. Based on multi-cell hexagonal network model derived in Appendix 3A, the carrier-to-interference ratio is expressed as

$$CIR(r) = \frac{1}{6}\frac{\frac{P_t}{r^\gamma}}{\frac{P_t}{d_1^\gamma}+\frac{P_t}{d_2^\gamma}+\cdots} = \frac{1}{6}\frac{r^{-\gamma}}{\sum_{i=1}^{\infty} d_i^{-\gamma}} \tag{3.27}$$

where $r$ is the intra-cell distance between the base site and its target receiver, and $d_i$ is the distance between target site and $i^{th}$ tier co-channel interfering site. Here we assume all sites transmit at the same power level of $P_t$ watts. All channels from sites to the target receiver

undergo the same propagation path loss with exponent $\gamma$. Alternatively, CIR(r) can be written as

$$\mathrm{CIR}(r) = \frac{\mathcal{E}_s R}{I_o W} = \frac{\mathcal{E}_b}{I_o} \cdot \frac{R_b}{W} = \frac{\mathcal{E}_b}{I_o} \cdot \frac{1}{P_G}$$

where $P_G$ is the processing gain. (Difference between $P_G$ and spreading gain $N_c$ is discussed on page 47.) $I_o$ is the interference PSD. For non-spread spectrum modulated signalling, $P_G \cong 1$. For spread-spectrum modulated signalling, $P_G \cong N_c \gg 1$ if data modulation is binary. Link quality (e.g., bit-error rate) is commonly measured as a function of $\mathcal{E}_b/I_o$ for an interference-dominated traffic channel. Otherwise, if the channel is perturbed only by thermal noise, the benchmark is in terms of $\mathcal{E}_b/N_o$. If we aim for completeness, the link quality of a cellular forward link is measured in terms of $\mathcal{E}_b/(N_o + I_o)$. Rewriting eqn.(3.27):

$$\mathcal{E}_b/I_o(r) = \frac{\mathcal{E}_b}{I_o} = \frac{\sqrt{3\gamma}\, P_G}{6} \left(\frac{r}{R}\right)^{-\gamma} \left[\sum_i N_r(i)^{-\gamma/2}\right]^{-1} \tag{3.28}$$

The reuse number $N_r(i)$ of $i^{th}$ tier co-channel site has this relation derived in eqn.(3.33):

$$N_r(i) = \frac{1}{3} \cdot \left(\frac{d_i}{R}\right)^2$$

It is known that for every tier (say, $k^{th}$), there exists exactly six co-channel interfering sites, whose distance to the target site are all equal. For example, when $N_r = 1$, all neighboring sites are co-channel interference sites. When $N_r = 3$, all non-immediate but closest cell sites are co-channel interference sites. As depicted in Fig. 3.25, all six sites belonging to the $k^{th}$ tier group are denoted by **k**.

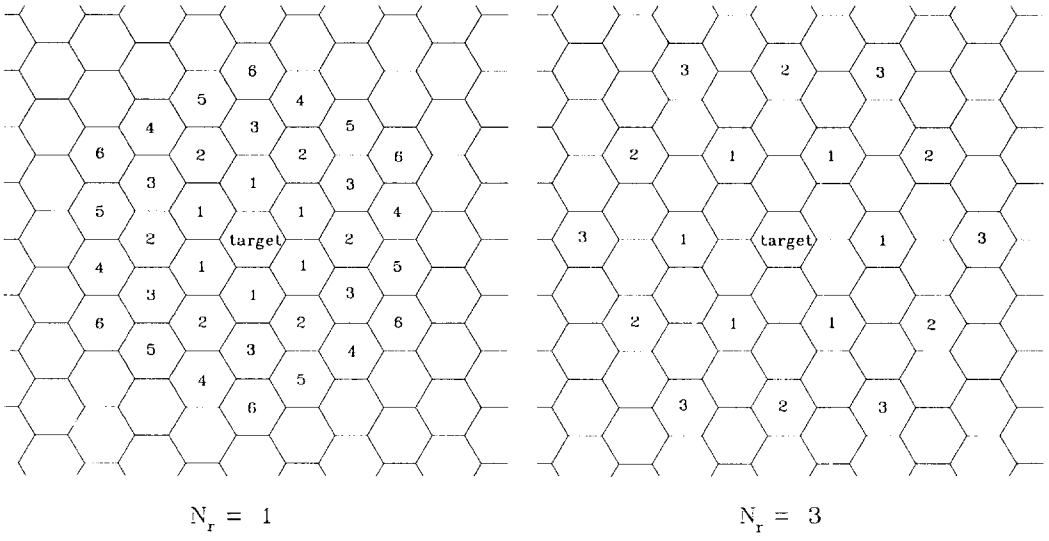



**Figure** 3.25: Six co-channel interfering sites belong to $k^{th}$ tier group

### 3.4.2 Radio Capacity: FDM/ODM vs. N-CDM/O-CDM

As seen from Fig. 3.25, inter-cell channel multiplexing with naive CDM affords reuse of the same bandwidth $W$ in every cell. The price paid for this convenience is increased interference due to a large number of co-channel interfering sites per unit area. To suppress co-channel interference, forward link data channels are spread-spectrum modulated with signature waveforms. If $W$ is the transmit bandwidth of a SSM signal, the actual data bandwidth is much smaller, say B, where $W = N_c B$. Given a data bandwidth of B the total sum rate $R_{sum}$ supported in each cell is a Shannon information theoretic quantity that depends on $\mathcal{E}_b/(I_o + N_o)$ as well as on the value of the spreading gain $N_c$ and the cross-correlation properties of assigned signature waveforms. In this model based on power ratio we see that the interference power is suppressed by a factor of $P_G$. This is the power gain after despreading. At the same time, the transmit bandwidth is also expanded by $N_c$. As we recall, for binary data modulation, $P_G \approx N_c$.

As we shall see in Chapter 6, under standard Gaussian approximation method and asynchronous reception, spread-spectrum modulation/demodulation offers a power gain of $N_c$ and a reduction in effective interference power by a factor of 3.

In FDM/ODM inter-cell interference suppression is offered only through propagation loss; transmit signals are not spread-spectrum modulated for multi-user interference mitigation. For this reason the reuse distance is larger in orthogonal inter-cell multiplexing policy than non-orthogonal N-CDM based multiplexing policy. If the reuse number is $N_r$ and the total allocated bandwidth is $W$, the cellular bandwidth is $W_c = W/N_r$. Given total bandwidth $W$, the maximum achievable rate $R_{sum}$ supported in each cell is an information-theoretic quantity that depends on $\mathcal{E}_b/(I_o + N_o)$, $N_r$ and intra-cell distance r between target mobile and base site. Unlike N-CDM/O-CDM, in orthogonal inter-cell multiplexing, it is possible to assign channels with different reuse numbers based on their intra-cell distances. This method of variable reuse exploits the concept of reuse partitioning. It is discussed in detail in Appendix 3A on page 152.

#### Effective Reuse Number

From eqn.(3.28) we can compute the Shannon rate $R_c(r)$ (in bits per second) of a target user per cell for given SIR:

$$R_c(r) = W_c \log\left[1 + \mathrm{SIR}(r)\right]$$

where $W_c = W/N_r$ is allocated bandwidth per cell. This achievable rate R(r) is a function of mobile distance r from the target base site. We can determine the average rate by averaging over mobile distances by assuming that mobile users are uniformly populated over the entire area of a hexagonal cell. For convenience, we set the cell radius R = 1. The bit energy $\mathcal{E}_b$ is adjusted such that at the cell edge (R = 1), the Shannon rate is normalized; i.e., SIR(R) = 1 and $R_c(R) = W_c$.

If the total bandwidth is $W$ Hz and average Shannon rate is $\overline{R_c}$, the effective reuse number is

$$N_{eff} = W/\overline{R_c} \tag{3.29}$$

For example, if $W$ Hz is allocated to each cell in a CDM-based network of reuse 1, and $\overline{R} = W$, then $N_{eff}$. In Table 3.1 we compare the effective reuse numbers of N-CDM/O-CDM and FDM/ODM policies for several different reuse numbers and channel models. We also list $N_{eff}$ with power control (for N-CDM/O-CDM), and with reuse partitioning (for FDM/ODM). From

**Table 3.1:** Effective Reuse Number $N_{eff}$ for Hybrid Channel Multiplexing Techniques (w/ PC = with power control, w/ RP = with reuse partitioning)

| $N_r$ | $\gamma$ | N-CDM/O-CDM | N-CDM/O-CDM w/ PC | FDM/ODM | FDM/ODM w/ RP |
|---|---|---|---|---|---|
| 1 | 4 | 2.58 | 1.78 | 6.38 | 6.38 |
| 3 | 4 | 4.55 | 3.85 | 4.23 | 3.48 |
| 4 | 4 | 5.63 | 4.82 | 3.76 | 2.77 |
| 7 | 4 | 6.45 | 6.32 | 5.55 | 4.13 |
| 1 | 2 | 2.78 | 1.91 | 6.44 | 6.44 |
| 3 | 2 | 5.03 | 4.11 | 4.78 | 3.66 |
| 4 | 2 | 6.10 | 5.05 | 3.98 | 3.01 |
| 7 | 2 | 6.93 | 6.77 | 5.78 | 4.52 |

the list we deduce that N-CDM/O-CDM with $N_r = 1$, along with power control, is the preferred choice. Since the total transmit power is constrained, for non-orthogonal channel multiplexing with N-CDM, a larger reuse number does not necessarily improve network capacity. This situation is also observed in orthogonal inter-cell channel multiplexing. However in FDM/ODM, a smaller reuse causes large interference that the cell capacity is reduced from its optimum. Reuse partitioning does offer some increase in capacity, but not enough to dominate N-CDM/O-CDM. Besides, implementation of reuse partitioning in FDM/ODM (channel swapping and mobile location tracking) is more complex than feedback power control of N-CDM/O-CDM.

## Notes and References

In this chapter, we first give concise definitions of various multi-user channel models. Each channel is characterized by the transition probability distributions between information source-sink pairs. We next show that the cellular channel is equivalent to our defined compound channel. In particular, we focus on the forward link of a cellular network—the broadcast plus interference compound channel. The last part of the chapter deals with several topics specifically pertinent to the cellular channel, namely the hexagonal geometry (since a hexagon is widely accepted as a theoretical shape of a cell) and the concept of cellular re-use. Most of the background material is borrowed from various texts in information theory and mobile radio communications. As references, we list the following texts by Cover [22], Csiszar and Korner [27], W. C. Y. Lee [71] and Rappaport [114]. Some of the information related to the broadcast channel was derived from P. Bergmans' Ph.D. thesis [8]. Several key concepts in hexagonal geometry overlap with those published by MacDonald [79]. The broadcast channel and its achievable rate region was first proposed by T. M. Cover [21]. Its capacity region was proved by P. P. Bergmans [7]. The role of cooperation was introduced by Bergmans. He also derived capacity regions for time-, frequency-, and code-division multiplexing schemes for a Gaussian noise channel. The interference channel was introduced by Shannon [127] through a two-way channel model. It was generalized by Carliel [12]. The capacity of a vector Gaussian multiple-access channel was derived by S. Verdu [149]. It is applies to the code-division multiplexing case, and is more general than the well-known Gaussian multiple-access capacity of Cover and Wyner [165]. The average broadcast capacity with multipath flat fading is studied by A. Goldsmith [44]. We extend P. P. Bergmans' results by including spread-spectrum multiplexing schemes. We also derive achievable rate regions of a broadcast plus interference (compound) channel under various statistical fading models. The cellular broadcast channel can be modelled as this kind of compound channel. We show that spread-spectrum multiplexing where signature codes belong to a special class of superimposed binary sequences has the largest rate region.

# Appendix 3A

## Hexagonal Cellular Geometry

At the heart of cellular radio engineering design is the concept of frequency reuse. A radio frequency bandwidth assigned to a base station is reused at a distant location by another base station. The recycling of bandwidth is made possible by the attenuation of high-frequency radio signal as it propagates away from a transmitter. In order to evaluate various performance parameters of a cellular radio communication system, the infinitely broad coverage area is partitioned into smaller regions that are called cells. The radiation pattern of an isotropic or omni-directional transmit antenna is accepted as a circle. That is, if a transmitter radiates from the center of a circle, mobiles located on the perimeter receive the same signal strength. The art of covering such a large region using equal-size regular polygons such that no gap or overlap exists among them is called "tessellation" [23]. Equivalently, the coverage layout is an imaginary grid of contiguous tiling of polygons of equal shape and size. The smallest unit in a very large grid is a *cell*. It can be shown that the only equal-size and *equally sided* polygons that tessellate a planar region are an equilateral triangle, a square and a hexagon. Among them the hexagon is the most suitable in approximating a circular coverage area since the difference in area between a hexagon and a circle is the smallest (see. Fig. 3.26). This fact alone does not

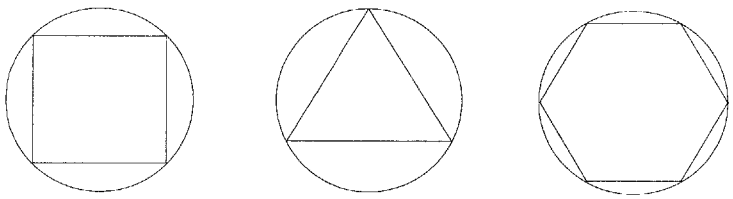


**Figure** 3.26: Valid polygons —square, triangle, hexagon— as a basic unit "cell" in a contiguous non-overlapping grid

justify the use of a hexagon as the shape of a cell. It is true that the overlap among adjacent circles is greater when a square grid —instead of a hexagon grid— is used. This is illustrated in Fig. 3.27. Thus, more circles (base stations) are required to blanket a coverage area when a square grid is used. This second fact does not necessarily illuminate the advantage of a hexagon over a square. We will have to wait until the end of this tutorial to give a succinct reason for the preference of hexagonally-shaped cells. First we introduce some simple conventions, notations and unique properties of hexagonal cellular geometry.

The radius R of a hexagon (see Fig. 3.28) is defined as the distance from the center to a vertex. The length of each side is also equal to R. The area A and the shortest distance from the center

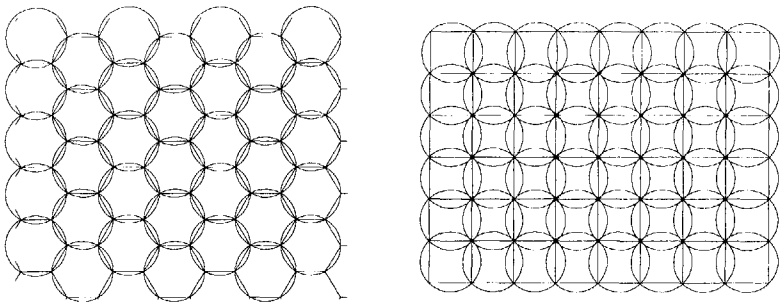


**Figure 3.27:** Coverage of an area using hexagonal and square grid. There is greater overlap of circles for a square grid.

to each side d are

$$A = \frac{3\sqrt{3}}{2} R^2, \qquad d = \frac{\sqrt{3}}{2} R$$

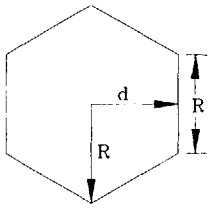



**Figure 3.28:** A hexagon and its dimension. R is the radius and its area $A = (3\sqrt{3}/2)\, R^2$. A hex unit is equal to 2d, where $d = (\sqrt{3}/2)R$.

## Coordinate System

Similar to the cartesian coordinates $(x, y)$, the hexagonal coordinate system uses two bases, except the angle between them is $\pi/3$ radians. As shown in Fig. 3.29, two types of hexagon grids, *flat-top* and *flat-side* grids, respectively can be constructed. The terminology is self-explanatory. For the flat-top grid, the coordinates are labelled $(u, v)$, while the primed notation $(u', v')$ is used for the flat-side grid. A *hex unit* along each basis is equal to $2d = \sqrt{3}\, R$. By convention, the *one-dimensional* or *linear* cellular grid model uses the flat-side hexagons since neighboring cells are contiguous. The one-dimensional model, which is used mainly as a "highway" footprint is shown in Fig. 3.30. For all other types of coverage such as built-up city area, suburban and rural locations, the most widely used model is the *two-dimensional* or *planar* cellular grid with either flat-side or flat-top hexagons. All hexagonal cells surrounding (and

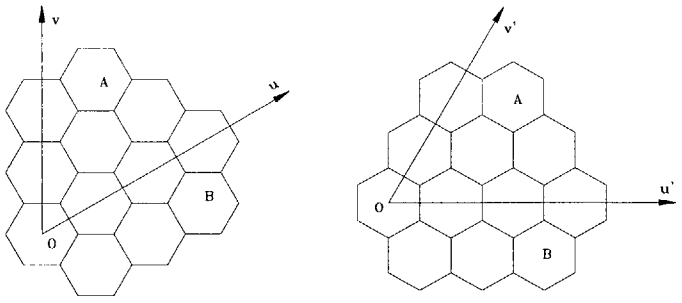



**Figure** 3.29: Flat-top and flat-side hexagonal grids.

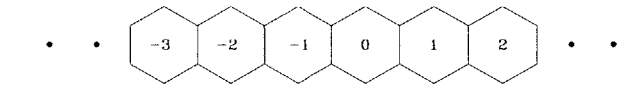



**Figure** 3.30: One-dimensional or linear hexagonal grid.

not necessarily touching) a target cell are called its *neighbors*. The distance between a cell and its *immediate* neighbor is $2d = \sqrt{3}R$, or 1 hex unit. Each hexagon in a grid is identified by indexing the *hexagonal coordinates* of its center. The *target* cell is always located at the origin with coordinates (0,0). All its neighbors are referenced with coordinates $(u, v)$ in hex units. For example, the coordinates for cells A and B in Fig. 3.31 are $(1, 2)$ and $(3, -1)$, respectively. Note that all cells in the first "quadrant" (see Fig. 3.32) have positive coordinate values. Likewise, both coordinates of the cells in the third quadrant are negative. In cellular system design we

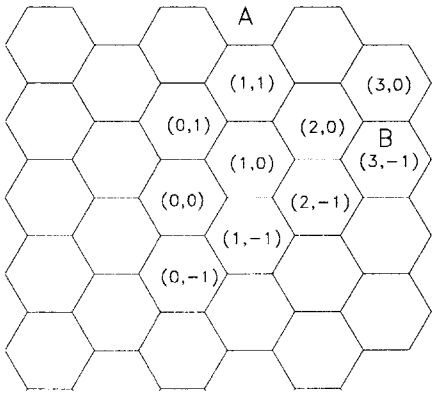



**Figure** 3.31: Identification of hex cells in a grid

are only interested in neighboring cells that are equidistant from the target cell; their exact coordinates are of secondary importance. In a *rotated* hexagonal coordinate system (see Fig.

3.33), each cell location is denoted by $(U, V, \theta)$, where $\theta$ is the relative *rotation angle* between u and U axes. Thus, a valid value of $\theta$ is $\frac{k\pi}{3}$, $k \in \mathbb{N}^5$. [A similar notation $(U', V', \theta)$ can be used for the flat-side grid.] The rotated coordinate system $(U, V, \theta)$ is preferred over the conventional $(u, v)$ system because the coordinates are always positive. In order to avoid any duplication in labelling, a coordinate such as $(0, V, \theta)$, $V \in \mathbb{Z}_+$ is not allowed. Unless noted, we always assume $U \geq V$ with the target cell at (0,0). All immediate neighboring cells of distance one hex unit

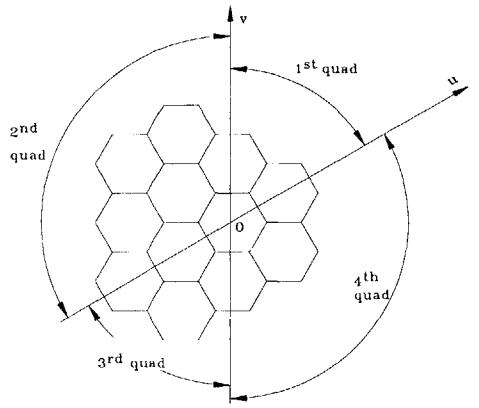



**Figure** 3.32: Four quadrants in a hexagonal coordinate system. Only cells in the first quadrant have coordinates $(u, v)$ with positive integers.

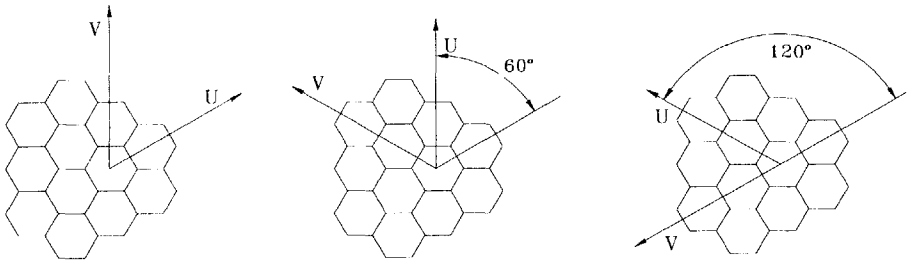



**Figure** 3.33: Rotated hexagonal coordinate system. Each cell is labelled using three parameters: $(U, V, \theta)$. The figure shows only the first three rotated angles, although there are a total of six.

are called *first-tier* cells. The next closest cells touching the first-tier cells are called *second-tier* cells. The next set of cells touching the second-tier cells are called *third-tier* cells and so on. Note that in general, not all $k^{th}$–tier cells are k hex units away from the target cell (see. Fig. 3.34).

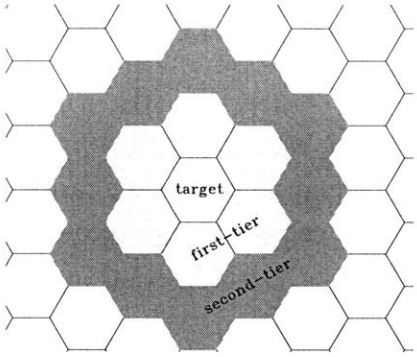
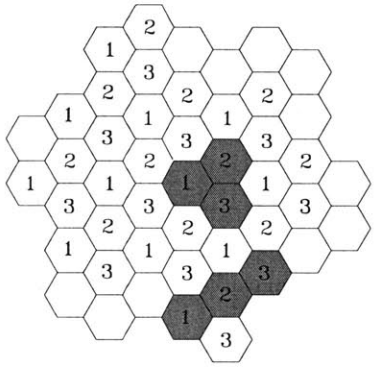



**Figure** 3.34: First and second tier cells. Every $k^{th}$–tier cell is not k hex units away from the target cell. Some have larger distances.

## Co-Channel Cells

For a set of positive integers U and V, all six neighboring cells with coordinates $(U, V, \theta)$ are equally distant. They are called *co-channel* cells. The terminology is derived from the fact that the target cell and its co-channel cells *co-share* the same frequency band. In other words, a frequency band or "channel" assigned to the target cell simultaneously *reused* in equidistant co-channel cells. Fig. 3.35 illustrates three sets of co-channel cells sharing channels 1, 2 and 3.



(1,1)

**Figure** 3.35: Cellular network with hexagonal cells. The reuse number is 3. The gray regions illustrate valid cluster shapes. The top cluster is 120° rotational invariant, while the bottom cluster has 180° rotational invariance.

**Proposition 3.1** *A target cell has only six neighbors of equal distance if the coordinates of each of*

*its neighbors is of the form* $(U, 0, \theta)$ *or* $(U, U, \theta)$. *For coordinates* $(U, V, \theta)$ *where* $U \neq V$, *the target cell has twelve equidistant neighbors.*

The proof is straightforward. For $(U, V, \theta)$ with $U \neq V$, two neighbors with respective coordinates $(U, V, \theta)$ and $(V, U, \theta)$ are equidistant from the target cell. Note that in the last case, although their are twelve equidistant cells, only six are co-channel cells.

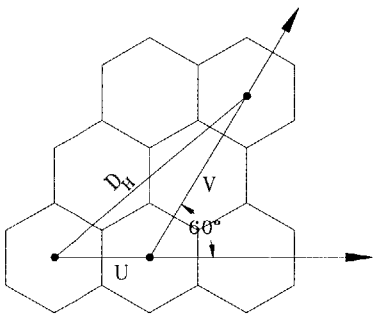



**Figure** 3.36: Reuse distance between two cells.

## Reuse Number

We have shown that six co-channel cells are equally distant. We now wish to compute this distance. Applying the *Law of Cosine* (see Fig. 3.36), it is easy to show that the *reuse distance* between the target and its co-channel cell (in hex units) is:

$$D_H = \sqrt{U^2 + V^2 + UV} \tag{3.30}$$

Since a hex unit equals $2d = \sqrt{3}R$, the distance in metric unit is:

$$D = \sqrt{3}\, R\, \sqrt{U^2 + V^2 + UV} \tag{3.31}$$

Equivalently,

$$\frac{D}{R} = \sqrt{3}\, \sqrt{U^2 + V^2 + UV} \tag{3.32}$$

The reuse number $N_r$ is defined as:

$$N_r = D_H^2 = \frac{1}{3}\left(\frac{D}{R}\right)^2 = U^2 + V^2 + UV \tag{3.33}$$

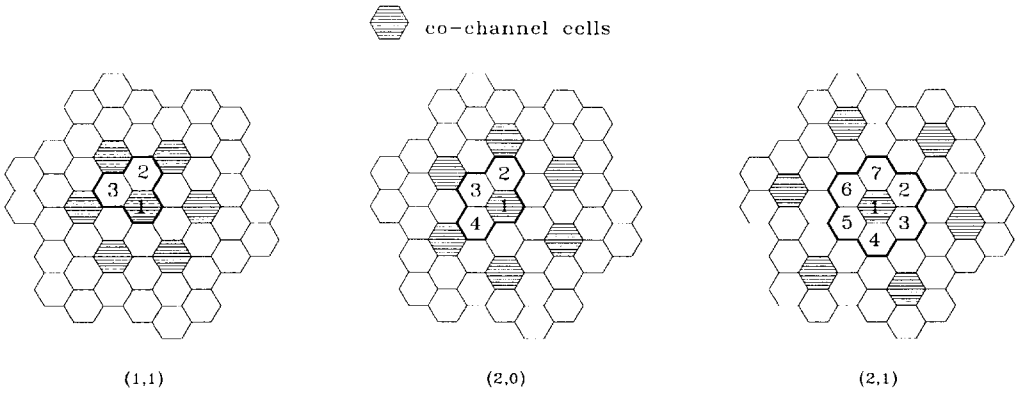



**Figure** 3.37: Various cluster patterns. The figure shows co-channel cells for reuse number N=3, 4 and 7.

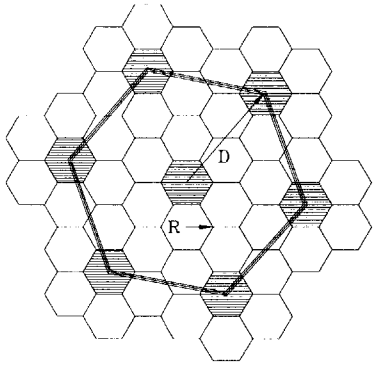



**Figure** 3.38: Enlarged hexagon with area B and radius D.

## Cell Cluster

In Fig. 3.37, a target cell along with its six co-channel cells are shown for three different cases: $(1,1,\theta)$, $(2,0,\theta)$ and $(2,1,\theta)$. By translation, we form another set (coset) of target and co-channel cells, and eventually exhaust all the cells in the (infinite) grid. A group of cells consisting of a target cell and its translated clones is called a *cluster*. It is defined as a group of cells that covers the entire grid without any gap or overlap when translated. Due to symmetry, it is seen immediately that by translating a target cell to its co-channel cell location, the entire grid can be constructed. The number of cells in a cluster is defined as the *cluster size* N.

**Proposition 3.2** *The cluster size* N *is equal to the reuse number* $N_T$.

When the centers of co-channel cells are connected as shown in Fig. 3.38, a larger hexagon is formed. Denote its area by B. By translation, the lines can also connect the center of masses of

the cluster. Since the angle between a pair of intersecting lines is $2\pi/3$, the area of the larger hexagon is:

$$B = NA + 6\left(\frac{1}{3}NA\right) = 3\,NA$$

On the RHS, the first term is the area of the cluster of the target cell, since this cluster is fully enclosed in the larger hexagon. The second term is the total area of partial clusters from its six co-channel cells, where each co-channel cell cluster contributes 1/3 of its area NA. Rearranging the above equation and using the fact that the ratio of the areas of the outer and inner hexagons is:

$$\frac{B}{A} = \frac{D^2}{R^2}$$

we obtain our desired result:

$$N = \frac{B}{3A} = \frac{1}{3}\frac{D^2}{R^2} = N_r$$

Since $U, V \in \mathbb{N}$ and the cluster size (reuse number) is:

$$N = N_r = U^2 + V^2 + UV \tag{3.34}$$

only certain integers are valid cluster sizes. Then, how can we determine these integer values without exhaustively substituting a pair of integers in (3.34), especially when N is large? Of course, it is most likely that we may never encounter large values of the reuse number N in cellular radio engineering design. However, in another communication setting such as vector quantization, it is often desirable to partition a plane into contiguous clusters of hexagons. In such a scenario, N can be very large. We state the following without proof.

**Proposition 3.3** N *is a valid cluster size if its factors are:*

(a) 1, 3, 4 *or* 7

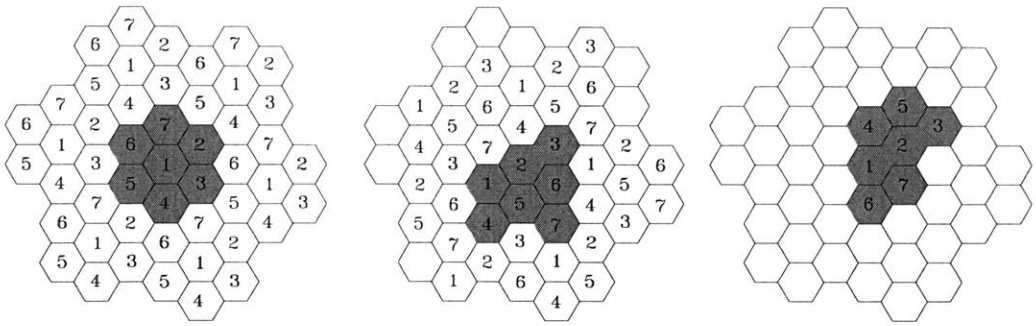
(b) *a prime number and its additive value is* 1, 3, 4 *or* 7

(c) *the square of an integer*

The additive value of an integer $abcd$ (in decimal form) is $(a + b + c + d)$.

## Rotational Invariance

It should be noted that knowledge of a valid cluster size is not sufficient in the construction of a valid cluster. As shown in Fig. 3.39, for a given cluster size, several shapes are possible and as illustrated, some are not valid clusters, i.e. gaps or overlaps exist when they are juxtaposed. A cell cluster is said to be $\rho^\circ$ *rotationally invariant* if its shape is invariant to a $\rho^\circ$ revolution about its center of mass. Due to symmetry, three possible locations of the center of mass are the center, the vertex and the midpoint that bisects each side. The corresponding clusters are 60°, 120° and 180° rotationally invariant. All three cases are illustrated in Fig. 3.37. Note that



(a) (b) (c)

**Figure** 3.39: Hexagonal cluster patterns for reuse 7. The shape in (a) in 60° rotational invariant, and its the preferred choice.

a 60° rotationally invariant cluster is both 120° and 180° rotationally invariant. However, 120° and 180° rotational invariance are mutually exclusive.

**Proposition 3.4** *A cluster in which co-channel cells are located at* $(U, U-1, \theta)$, $(U, U, \theta)$ *and* $(U, 0, \theta)$ *are* 60°, 120° *and* 180°*rotationally invariant, respectively.*

The proof is based on simple geometrical arguments. We now have a systematic approach of constructing a valid cluster shape for *cluster patterns* $(U, U-1)$, $(U, 0)$ and $(U, U)$. Then how do we construct valid cluster shapes for arbitrary cluster patterns $(U, V)$? We state the following without proof:

**Proposition 3.5** *A cluster pattern* $(U, V)$ *has a shape that is either* 60°, 120° *or* 180° *rotationally invariant.*

Since 60° rotational invariance is the most symmetric, we first check whether a cluster can be constructed with this symmetry. If not, it can be constructed with either 120° or 180° rotational invariance, but not both. Earlier in the section, we stated that cells in the $k^{th}$ tier are not equally distant. We now give a more precise statement:

**Proposition 3.6** *The number of cells* $N_t(k)$ *in the* $k^{th}$ *tier is* $6k$.

For a 60° rotationally invariant cluster, its cluster pattern is $(U, U-1)$. The corresponding cluster size $N$ is:

$$N_U = \sqrt{U^2 + (U-1)^2 + U(U-1)} = 3U^2 - 3U + 1 \tag{3.35}$$

Similarly, for $(U+1, U)$:

$$N_{U+1} = 3U^2 + 3U + 1 \tag{3.36}$$

It is easily deduced that the difference in the cluster size of $(U+1, U)$ and $(U, U-1)$ is $N_t(U)$. Subtracting (3.35) from (3.36),

$$N_t(U) = 6U$$

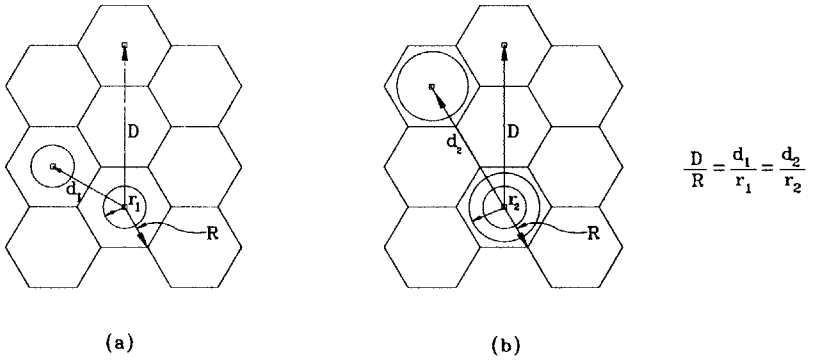
Based on the above result, the next statement follows naturally:

**Proposition 3.7** *Among the* $k^{th}$*–tier cells, there are* $k$ *groups of cells, each group consisting of six equally distant cells.*

For a given $(U, V)$ a target cell has six equidistant co-channel cells. These co-channel cells form a group. Since there are $6k$ cells in the $k^{th}$ tier, the number of such groups is $k$. Depending on the relationship between $U$ and $V$, members from two different groups may be equally distant from the target.

### Open Problem

This raises the question of whether we can state anything more about the distance properties of such groups. It is clear that the two groups with patterns $(U, V)$ and $(V, U)$ are equally distant from the target cell. Besides this, we are not able to give any general result. It is tempting to conjecture that a cluster size $N$ can be generated only by a unique pattern $(U, V)$ or $(V, U)$. This, however, is not true. For example, by an exhaustive search, we note that the patterns (7,0) and (5,3) have the same reuse number $N = 49$.




(a) (b)

**Figure** 3.40: Reuse partitioning of hexagonal cells.

## Reuse Partitioning

For a frequency reuse of $N_r$, the bandwidth per cell is $W/N_r$ Hz. Therefore larger cell bandwidth is obtained by tighter frequency reuse. For a mobile at distances $r$ from its base

($r < R$) is signal-to-interference power ratio (SIR) is:

$$\mathrm{SIR}(r) \approx \frac{1}{6}\frac{\frac{1}{r^{\gamma}}}{\frac{1}{d^{\gamma_1}} + \frac{1}{d^{\gamma_1}} + \ldots} \tag{3.37}$$

where $d_k$ is the (approximate) distance (see Fig. 3.41) between the mobile and a co-channel base station in group k. (A group is defined as a set of six equidistant co-channel cells. For first tier, there is only one group. Let's call it group 1. There are two groups in second-tier. Let's call them group 2 and 3, respectively, and so on.) The approximation becomes exact only when the mobile is located at the center of the target cell.

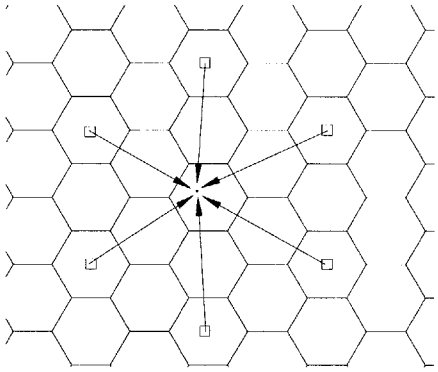


**Figure** 3.41: Approximate distances between target mobile and co-channel bases

Rewriting in a more compact form

$$\mathrm{SIR}(r) \approx \frac{1}{6}\frac{\frac{1}{r^{\gamma}}}{\frac{1}{d_{\mathrm{eff}}^{\gamma}}} \tag{3.38}$$

$$= \frac{1}{6}\left(\frac{d_{\mathrm{eff}}}{r}\right)^{\gamma} \tag{3.39}$$

where the *effective re-use distance* $d_{\mathrm{eff}}$ is expressed as

$$\frac{1}{d_{\mathrm{eff}}^{\gamma}} = \frac{1}{d_1^{\gamma_1}} + \frac{1}{d_2^{\gamma_2}} + \ldots$$

Within a cell, we define a *zone* as an annular region $z_i$ between two concentric circles of radii $r_i$ and $r_{i-1}$, $i \in \mathbb{Z}_+$ with $r_0 = 0$. In Fig. 3.40, there are three zones $z_1$, $z_2$ and $z_3$. Using the

following definitions:

$n_i$ : reuse number in zone i

$r_i$ : radius of zone i

$d_i$ : reuse distance of zone i

$c_i$ : number of channels (frequency bands, time slots etc.) allocated in zone i

Z : total number of zones

It is easy to see that $n_Z = N$, $r_Z = R$ and $d_Z = D$. Zone 1 is a circle and zone N is the difference in area between the hexagon of radius R and a circle of radius $r_{Z-1}$. As illustrated in Fig. 3.40, a channel assigned to a mobile in zone i of the target cell can be re-used by another mobile, also in zone i of its cell as long as

$$\frac{D}{R} \leq \frac{d_i}{r_i}$$

Equivalently,

$$SIR(R) \leq SIR(r_i)$$

Since the frequency reuse number is chosen such that all mobiles receive SIR above a specified threshold SIR*, we have

$$SIR^* \leq SIR(R) \leq SIR(r_i) \tag{3.40}$$

Therefore, a frequency band assigned to a user inside a *ring* of radius $r_i$ can be reused inside a similar ring (of radius $r_i$) located at a cell of distance $d_i$. In Fig. 3.40, the reuse number $N_r = 4$. Tighter reuse of 1 and 3 are possible inside rings of radii $r_1$ and $r_2$, respectively. If the total bandwidth W is partitioned into M frequency bands, the following equality must hold:

$$\sum_{i=1}^{Z} c_i n_i = M \tag{3.41}$$

Setting $SIR(R) = SIR(r_i)$, we can deduce the following:

$$Q \equiv \frac{r_i^2}{n_i} = \frac{R^2}{N} \tag{3.42}$$

If the channels per cell are uniformly distributed among the zones, i.e., the number of channels per unit area is constant, the following must be satisfied:

$$\frac{c_1}{r_1^2} = \frac{c_2}{(r_2^2 - r_1^2)} = \cdots = \frac{c_Z}{(r_Z^2 - r_{Z-1}^2)} \tag{3.43}$$

Substituting (3.42) in (3.43):

$$\frac{c_1}{Qn_1} = \frac{c_2}{Q(n_2 - n_1)} = \cdots = \frac{c_Z}{Q(n_Z - n_{Z-1})} \tag{3.44}$$

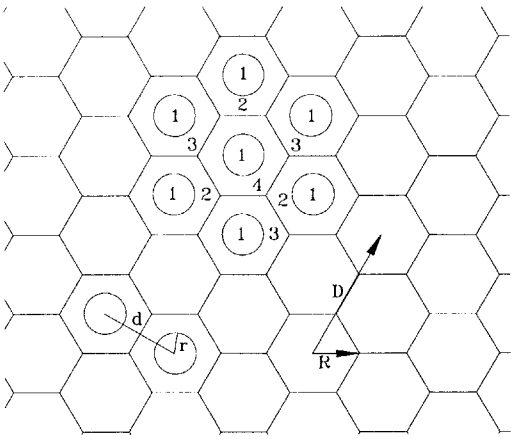



**Figure** 3.42: Channel assignment in reuse-partitioned cells for $N_r = 3$

By substituting the reuse numbers (1, 3, 4, 7, 9, 11, ...):

$$\frac{c_1}{1} = \frac{c_2}{2} = \frac{c_3}{1} = \frac{c_4}{3} = \cdots$$

With re-use partitioning, the total number of channels per cells (the cellular capacity) is

$$C_c = \sum_{i=1}^{Z} c_i$$

For $N_r$=7,

$$c_1 = \frac{M}{32}, \quad c_2 = \frac{M}{16}, \quad c_3 = \frac{M}{32}, \quad c_4 = \frac{3M}{32}$$

In the above example of re-use 7, $C_c = 7M/32$. Without reuse partitioning, the capacity is $M/7$. The increase in capacity due to reuse partitioning is 53%. If the cellular capacity is expressed in terms of bandwidth, then for re-use 7,

$$B_c = \left(\frac{7}{32}\right) W_{tot} = 0.2188\, W_{tot} \tag{3.45}$$

In another example shown in Fig. 3.42, channel 1 is assigned to ring 1 while channels 2, 3 and 4 are assigned to three rings of re-use 3. Since $c_2 = 2c_1$, we see immediately that

$$c_1 = \frac{M}{7}, \quad c_2 = \frac{2M}{7}$$

The cellular capacity $C_c = 3M/7$ and the effective re-use number is $7/3 \approx 2.3$.

### Area Adjustment

In eqn. (3.43) it is assume that all zones are annular. This however is not correct for the last zone since its area is the difference between a hexagon and a circle. We can rewrite eqns.

(3.43) and (3.44) with this area adjustment:

$$\frac{c_1}{r_1^2} = \frac{c_2}{(r_2^2 - r_1^2)} = \cdots = \frac{c_Z}{(k_h^* r_Z^2 - r_{Z-1}^2)}$$

$$\frac{c_1}{Qn_1} = \frac{c_2}{Q(n_2 - n_1)} = \cdots = \frac{c_Z}{Q(k_h^* n_Z - n_{Z-1})}$$

where

$$k_h^* = \frac{3\sqrt{3}}{2\pi}$$

With this correction, it is easily shown that

$$c_1 = \frac{M}{23.5}, \quad c_2 = \frac{2M}{23.5}, \quad c_3 = \frac{M}{23.5}, \quad c_4 = \frac{1.789M}{23.5}$$

and the bandwidth per hexagonal cell is

$$B_c^h = \left(\frac{5.789}{23.5}\right) W_{tot} = 0.2461\, W_{tot} \tag{3.46}$$

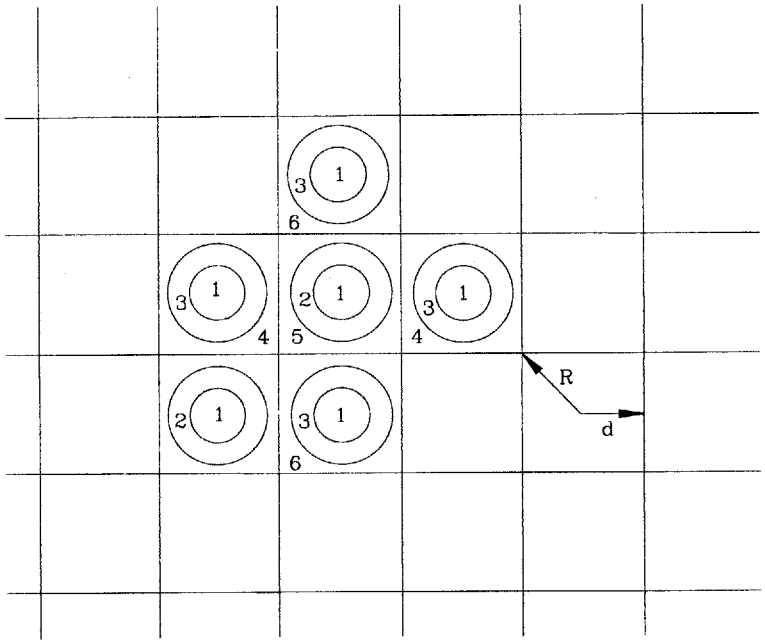



**Figure** 3.43: Cellular grid composed of equal-size squares

## Capacities of Hexagon- vs. Square-Shaped Cellular Networks

In a cellular grid (as shown in Fig. 3.43) composed of contiguous equal-size square cells of radius R, the following relations can be derived:

$$A = 2R^2, \qquad d = \frac{1}{\sqrt{2}} R$$

$$N_r^s = \frac{1}{2}\left(\frac{D}{R}\right)^2 = X^2 + Y^2$$

where X and Y are square units along x and y cartesian coordinates. Therefore, valid re-use numbers are

$$1, 2, 4, 5, 8, 9, 10, 13\ldots$$

and so on. With reuse partitioning we can calculate the cellular capacity of a square grid:

$$\frac{c_1}{r_1^2} = \frac{c_2}{(r_2^2 - r_1^2)} = \cdots = \frac{c_Z}{(k_s^* r_Z^2 - r_{Z-1}^2)}$$

where the area correction factor is

$$k_s^* = \frac{2}{\pi}$$

and

$$\frac{c_1}{1} = \frac{c_2}{1} = \frac{c_3}{2} = \frac{c_4}{1} = \frac{c_5}{3} = \frac{c_6}{1} = \frac{c_7}{1} = \cdots$$

In Fig. 3.43, channel 1 is reused in zone 1 of every cell. Channels 2 and 3 make up the group of channels with reuse number 2. For reuse 4, the group consists of 4 channels; 4, 5 and 6 are shown in the figure. For $D/R = \sqrt{21}$ (corresponding to hexagonal reuse $N_r^h = 7$), it is easy to see that for a square grid, $N_r^s = 10.5$. Valid square cluster sizes in this range are 10 and 13. Setting $n_Z = 10$, we can show that

$$c_1 = c_2 = c_4 = c_6 = \frac{M}{22.66}, \quad c_3 = \frac{2M}{22.66}, \quad c_5 = \frac{2M}{22.66}, \quad c_7 = \frac{0.89M}{22.66}$$

and the bandwidth per square cell is

$$B_c^s = \left(\frac{9.89}{22.66}\right) W_{tot} = 0.4365\, W_{tot} \tag{3.47}$$

# Appendix 3B

## Achievable Rate Regions: FDM >N-TDM

By setting $P = W = 1$, the achievable rate region of FDM with optimal power distribution is:

$$R_1 = \alpha \log\left(1 + \frac{1-(1-\alpha)(N_2-N_1)}{N_1}\right)$$
$$R_2 = (1-\alpha) \log\left(1 + \frac{1+\alpha(N_2-N_1)}{N_2}\right)$$

It is clear that the maximum achievable rates of both users are the single channel capacities: $R_1(\alpha = 1) = C_1^s$ and $R_2(\alpha = 0) = C_2^s$. The achievable rate region plot for the 2-user case is shown in Fig. 3.19(a). For N-TDM, the boundary is a straight line ($y = mx + b$):

$$R_1^N = -\left(\frac{C_1^s}{C_2^s}\right) R_2^N + C_1^s \qquad (*)$$

where the $y$ and $x$ intercept points are $C_1^s$ and $C_2^s$, the single-user channel capacities. Here, it is understood that $R_i^N$ implies $R_i^N(\alpha)$, a function of the time partition parameter $0 \leq \alpha \leq 1$. It appears all we need to show is the concavity of the FDM curve to prove the dominance of FDM over N-TDM. That is, we need to show:

$$\frac{d^2R_1}{dR_2^2} = \frac{d}{d\alpha}\frac{dR_1}{d\alpha}\frac{d\alpha}{dR_2}\frac{d\alpha}{dR_2} < 0$$

Unfortunately, the mathematics is messy and not as elegant as the alternative proof given by P. P. Bergmans. We will follow his work with a few minor changes. Rewriting $(*)$:

$$\frac{R_1^N}{C_1^s} + \frac{R_2^N}{C_2^s} - 1 = 0$$

For any rate pair $(R_1, R_2)$, define:

$$e = \frac{R_1}{C_1^s} + \frac{R_2}{C_2^s} - 1$$

For any point with coordinates $(R_1, R_2)$ to the right (i.e. outside of the triangle), $e > 0$. For points inside the triangle, $e < 0$.

The achievable rates for FDM with a prefixed bandwidth (constant $\beta$) partition are, with $P = W = 1$:

$$R_1 = \beta \log\left(1 + \frac{\alpha}{N_1\beta}\right)$$
$$R_2 = (1-\beta) \log\left(1 + \frac{1-\alpha}{N_2\beta}\right)$$

In general, it is understood that $R_i$ is a function of both power and bandwidth partition parameters $0 \leq \alpha, \beta \leq 1$. For an FDM curve with a fixed bandwidth partition, its y and x intercept points are the reduced bandwidth, single-user capacities, $C_1^s(\beta)$ and $C_2^s(\beta)$, with $W_1 = \beta$ and $W_2 = (1 - \beta)$ where:

$$C_i^s(\beta) = W_i \log\left(1 + \frac{1}{N_i W_i}\right)$$

This is shown in Fig. 3.19(b). In order to prove the dominance of FDM over N-TDM, we need only to show that for any bandwidth partition (arbitrary $\beta$), there exists a portion of the FDM curve outside the triangular region of N-TDM. For a fixed $\beta$, both intercept points of FDM are smaller than the intercept points, $C_1^s$ and $C_2^s$, of N-TDM. Next, it is straightforward to show that this FDM curve is concave:

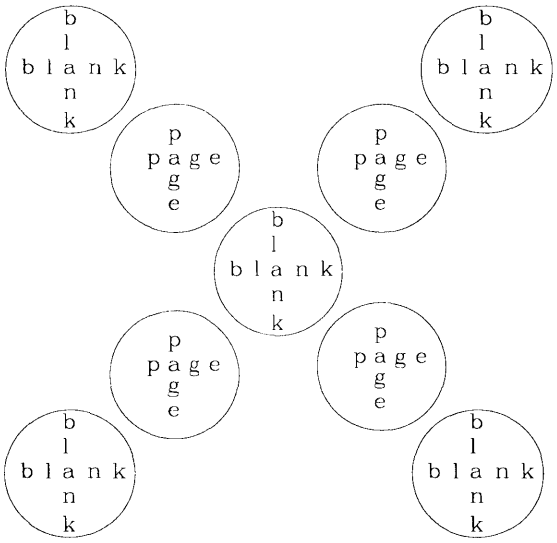
$$\frac{d^2R_2}{dR_1^2} = \frac{d}{d\alpha}\left(\frac{dR_2}{dR_1}\right) \cdot \frac{d\alpha}{dR_1} < 0$$

It appears that in order to show that FDM dominates over N-TDM, we only need to show that this FDM curve intersects the N-TDM straight line at two distinct points. In fact, we already know one such intersect point where the power and bandwidth partitions are equivalent ($\alpha = \beta$), $P_1 = \alpha P$ and $W_1 = \alpha W$. However, it is rather difficult to locate the other intersecting point since the FDM curve equation includes logarithmic terms. An alternative is to show that the rate-difference curve $e(\alpha, \beta)$ has a stationary point $\alpha^*$ that is not equal to $\beta$. (Remember that the two curves intersect at $\alpha = \beta$.) It is then straightforward to show that for $\alpha = \beta$:

$$\frac{d}{d\alpha} e(\alpha, \beta)\bigg|_{\alpha=\beta} = \frac{1}{C_1^s(1 + N_1)} - \frac{1}{C_2^s(1 + N_2)}$$

The slope of $e$ at $\alpha = \beta$ is equal to zero *only* if $N_1 = N_2$, which is expected since all multiplexing schemes are equivalent under equal-noise condition. For all other cases where $N_1 \neq N_2$, the slope of $e$ is non-zero; i.e., at the intersecting point, $\alpha$ is not a stationary point. Based on this result, and using the concavity of FDM curve, the dominance of FDM over N-TDM is proven.

The slope of $e$ for $\alpha = \beta$ is negative for $N_1 < N_2$.Therefore we know that the lower intersection point corresponds to the equal power-bandwidth partition case $\alpha = \beta$.


blank
page
blank
page
blank
page
page
blank
blank

# 4

# SIGNATURE WAVEFORMS AND SEQUENCES

## Summary

In Chapter 3, we propose code-division as the preferred channel multiplexing scheme for a transmitter in a cellular radio communication system, but little information is given about the design and construction of such codes. In an abstract sense a code is a set of instructions agreed upon between a transmitter and its receiver. In practice, information transfer in a CDM system is a two-step process: First, during the initial handshake phase of channel setup, a unique code sequence is assigned by the transmitter to a receiver. Second, messages from a source are encoded using this code sequence. Furthermore, this coded message is transmitted through free space in the form of electromagnetic waves. Hence, there is a bijective mapping between a coded message and its corresponding noise-like analog signal called *signature waveform*. The codes must be carefully chosen such that each receiver can readily decode its message from a composite broadcast plus interference signal. The cross-correlation function is a ruler that is used to measure the similarity between a pair of different code sequences or signature waveforms. Likewise the autocorrelation function measures the similarity between a code (or a waveform) and its phase-shifted version. Both cross- and auto-correlation functions can be further classified into three types: periodic, aperiodic and partial. A periodic correlation function is useful only in a special case where the time periods of signature waveforms are equal and the correlation window spans over this period. When the correlation window is less than the period, the correlation is measured in terms of two aperiodic correlation functions. In both periodic and aperiodic cases, the beginning (ending) epoch of the correlation window must be the

beginning (ending) epoch of the waveform. (We elaborate this remark graphically in the body.) If the above conditions (i.e., the same period between the waveforms, and correct alignment of time epochs) are not met, the only effective and meaningful tool is the partial correlation function. It is a function of both the time offset between the waveforms as well as the size of the correlation window. As we shall see, the partial correlation function is the most important, and also the most difficult to analyze and quantify deterministically. Most often, we must resort to its statistical or time-averaged properties. The main results presented in this chapter are:

- A special class of signature waveforms, known as binary orthogonal time functions, have zero periodic cross-correlation when the time offset is zero. Two types of orthogonal functions exist: *recursive* and *non-recursive*. The former is preferred since the periods of any two waveforms need not be equal; thus it offers seamless multi-rate transmission and reception capability. (See Chapter 6 for details.)

- In general, orthogonal waveforms and their corresponding code sequences have very poor aperiodic and partial correlation properties. This shortfall is easily compensated by superimposing an orthogonal code with a noise-like pseudo-random sequence. We coin such codes *superimposed orthogonal* sequences.

- We show that the choice of a particular pseudo-random sequence is not critical in the construction of superimposed orthogonal codes when the statistical properties of the partial auto-correlation function are used as the benchmark for system performance.

- There is little difference in correlation measure between recursive and non-recursive orthogonal codes when both are superimposed with the same pseudo-random sequence.

Most of our results presented in this chapter serve as background material for multi-rate CDM system design and analysis presented in Chapter 6.

## 4.1 Correlation Properties of Continuous-Time Signals

When characterizing the output waveform of a baseband filter (see Fig. 3.15) we must take into account its length or *period* (T sec.) as well as its occupied bandwidth (W Hz). Physically realizable (causal) finite-length waveforms are both time- and energy-limited. Using the Nyquist-Shannon Sampling Theorem, the dimension of a signal set

$$\mathbf{F}(t) = \left\{ f_1(t), f_2(t), \ldots, f_M(t) \right\}$$

where all waveforms in $\mathbf{F}(t)$ are of equal energy and equal period can be defined as

$$D \approx \lfloor 2TW \rfloor \tag{4.1}$$

where the product TW is commonly known as the passband *time-bandwidth product*. Using the Gram-Schimdt orthogonalization procedure it can be shown that $D \leq M$. Equality holds when the waveforms are mutually orthogonal, i.e.

$$\rho_{ij} = \int_{-\infty}^{\infty} f_i(t) f_j(t) dt = 0 \qquad f_i(t), f_j(t) \in F(t) \quad i \neq j \tag{4.2}$$

From eqn. (3.1) we know that a larger value of M implies a higher information content per waveform. However, for a fixed D, the waveforms become *correlated* ($\rho_{ij} \neq 0$) when $M > D$. For high-fidelity reception of a desired message from a code-division multiplexed signal, the waveforms must be carefully chosen such that they can be easily distinguished at the receiving end. The design of good signature waveforms is measured in terms of their auto- and cross-correlation functions. Each correlation function serves a different purpose:

- The cross-correlation function is a measure of the dissimilarity between a pair of signature waveforms. The mean-squared error of two *deterministic* signature waveforms is:

$$e_{ij}(\tau) = \int \left[ f_i(t) - f_j(t-\tau) \right]^2 dt = \mathcal{E}_i + \mathcal{E}_j - 2\,\rho_{ij}(\tau)$$

  where

$$\mathcal{E}_k = \int f_k^2(t)\, dt \qquad \rho_{ij}(\tau) = \int f_i(t) f_j(t-\tau)\, dt$$

  The parameter $\tau$ is the relative offset between the time epochs of signature waveforms. The waveforms are the most dissimilar when the mean-squared error $e_{ij}$ is maximized by minimizing their cross-correlation $\rho_{ij}$ —more precisely its absolute value $|\rho_{ij}|$ since correlation of $f_i(t)$ with both $f_j(t)$ and its inverse $-f_j(t)$ is of interest. Since the elemental waveforms that compose the broadcast composite signal can be time synchronized, we are particularly interested in the special case of $\rho_{ij}(0) = 0$ for all $i \neq j$ where all signature waveforms are mutually orthogonal. The *Dimensionality Theorem* states that for time-limited real pulses of period T and bandwidth W, the maximum number K of mutually orthogonal waveforms is $\approx TW$. A larger set ($> K$) of signature waveforms can be generated at the expense of $|\rho_{ij}| > 0$.

- The auto-correlation property of a signature waveform

$$\rho_j(\tau) = \int f_j(t) f_j(t-\tau)\, dt$$

  is critical for synchronization (tracking and locking) of carrier phase and time epoch of a signature waveform, especially when coherent detection is used at the receiver. A large spike at $\tau = 0$ such that $\rho_j(0) \gg \rho_j(\tau \neq 0)$ is desired. If the channel is time-dispersive, the received signal consists of delayed replicas of the transmitted signature

waveform. In order to reduce self-interference resulting from multiple delay components of the signature waveform, a low auto-correlation value is desired for any non-zero time offset within the channel delay spread $(\Delta\tau)_d$.

Ideally, we desire waveform pairs whose cross-correlation value $\rho_{ij}(\tau)$ is very small for any arbitrary time offset $\tau$, and whose auto-correlation values $\{\rho_i(\tau),\ \rho_j(\tau)\}$ are also very small for $\tau \neq 0$. They are often two conflicting requirements: The cross-correlation $\rho_{ij}(\tau)$ can be reduced by increasing the period T. However, if T is large, the initial acquisition and phase locking process can take too long, rendering the waveform impractical. Besides, it is well known from algebraic coding theory that it is impossible to design code sequences that possess ideal characteristics in both cross- and auto-correlation. See Sec. 4.7 for a brief discussion and a list of useful references on this topic .

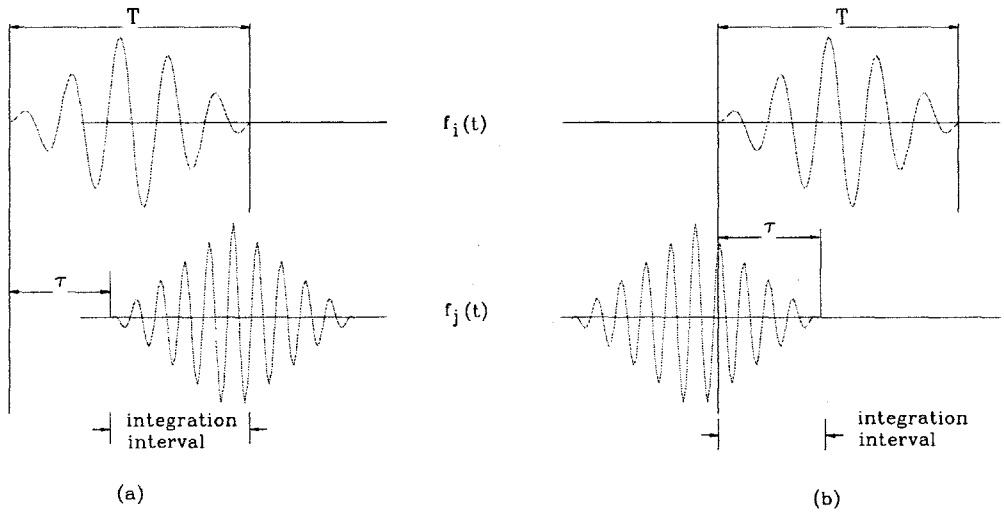



**Figure** 4.1: Aperiodic cross-correlation for symbol-time limited waveforms

### 4.1.1 Aperiodic Cross-Correlation

Physically realizable waveforms are strictly limited in time. Furthermore, if the waveforms are symbol-time-limited[1] —i.e., there is no overlapping of successive waveforms, $f_n(t) = 0$ for

[1]In communications literature the data rate is commonly measured in terms of symbols per second (sps). Other rates we will encounter are bits per second (bps) and chips per second (cps). Each symbol may carry several bits depending on the constellation set. Actual transmission, however, requires the generation of waveforms instead of symbols, and the number of waveforms transmitted per second is called the *symbol rate*, denoted by R. Its reciprocal T is the symbol time. If the period $T_p$ of the waveform is strictly limited to T, it is symbol-time-limited. In Chapter 5 we encounter time-overlapped waveforms where $T_p > T$; i.e., the beginning epoch of a waveform starts before the ending epoch of the previous waveform.

$t < 0$ and $t > T$— then demodulation and matched filtering are simplified. For such *aperiodic* waveforms, two kinds of cross-correlation functions exist:

$$\begin{aligned}\rho_{ij}(\tau) &= \int_0^T f_i(t)f_j(t-\tau)dt \qquad 0 \le \tau \le T \\ &= \int_\tau^T f_i(t)f_j(t-\tau)dt\end{aligned} \tag{4.3}$$

$$\begin{aligned}\hat{\rho}_{ij}(\tau) &= \int_0^T f_i(t)f_j(t+T-\tau)dt \qquad 0 \le \tau \le T \\ &= \int_0^\tau f_i(t)f_j(t+T-\tau)dt\end{aligned} \tag{4.4}$$

$f_i(t), f_j(t) \in F(t)$. They are called *continuous-time aperiodic* cross-correlation functions for an obvious reason: they both are aperiodic. They are illustrated in Fig. 4.1. In general, $\rho_{ij}(\tau) \neq \rho_{ji}(\tau)$, but $\rho_{ij}(\tau) = \hat{\rho}_{ji}(T-\tau)$.

### 4.1.2 Periodic Cross-Correlation

Since information is not conveyed on a one-shot basis but in a continuous stream of data modulated waveforms, it is more meaningful to deal with the *periodic* cross-correlation function. First we generate an infinite-time periodic waveform $p_i(t)$ by repeatedly concatenating symbol-time limited pulses $f_i(t)$, i.e. $p_i(t) = \sum_{n=-\infty}^{\infty} f_i(t-nT)$. Then, the *continuous-time periodic* cross-correlation function is:

$$C_{ij}(\tau) = \int_0^T p_i(t)p_j(t-\tau)dt \qquad -\infty < \tau < \infty \tag{4.5}$$

Similar to (4.3) and (4.4), we define the *continuous-time aperiodic* cross-correlation functions. For $0 \le \tau \le T$:

$$\hat{R}_{ij}(\tau) = \int_0^\tau p_i(t)p_j(t-\tau)dt = \hat{\rho}_{ij}(\tau) \tag{4.6}$$

$$R_{ij}(\tau) = \int_\tau^T p_i(t)p_j(t-\tau)dt = \rho_{ij}(\tau) \tag{4.7}$$

Note that both $R_{ij}(\tau)$ and $\hat{R}_{ij}(\tau)$ are periodic and well-defined for $-\infty < \tau < \infty$; but we still call them "aperiodic" cross-correlation functions[2] since they are respectively identical to $\rho_{ij}(\tau)$ and $\hat{\rho}_{ij}(\tau)$ for $0 \le \tau \le T$. These cases are illustrated in Fig. 4.2.
Of course,

$$C_{ij}(\tau) = R_{ij}(\tau) + \hat{R}_{ij}(\tau) \tag{4.8}$$

[2]Note that our definition of $R_{ij}(\tau)$ and $\hat{R}_{ij}(\tau)$ is not universal; they are called *partial* cross-correlation functions in Ref. [109] and in subsequent articles based on [109]. We reserve *partial* correlation for a different type of function (see Sec. 4.1.3).

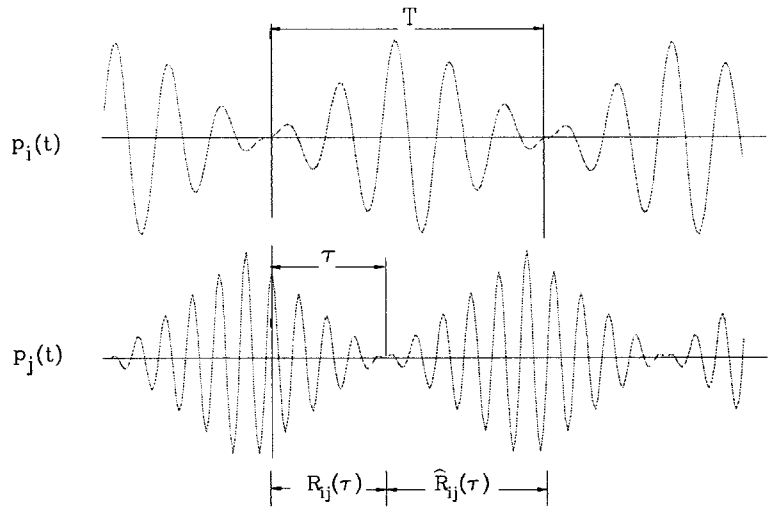



**Figure** 4.2: Periodic cross-correlation for symbol-time limited waveforms

Using the identities:

$$R_{ij}(\tau) = \hat{R}_{ji}(\tau) \tag{4.9}$$

$$\hat{R}_{ij}(\tau) = R_{ji}(\tau) \tag{4.10}$$

We immediately deduce that $C_{ij}(\tau) = C_{ji}(\tau)$. Hence $C_{ij}(\tau)$ is known as the "even" *continuous-time periodic* cross-correlation function. We can also define the "odd" *continuous-time periodic* cross-correlation function:

$$\hat{C}_{ij}(\tau) = \hat{R}_{ij}(\tau) - R_{ij}(\tau) \tag{4.11}$$

Note that $\hat{C}_{ij}(\tau) = -\hat{C}_{ji}(\tau)$.

### 4.1.3 Partial Cross-Correlation

Both periodic and aperiodic correlation are functions of a single variable: the delay parameter $\tau$. The correlation window for periodic correlation function spans over a period T. For aperiodic correlation functions, the correlation window is either $[0, \tau)$ or $[\tau, T)$, where $0 \leq \tau \leq T$. As depicted in Fig. 4.3, it is always from the beginning (ending) epoch of one waveform to the ending (beginning) of another waveform. Hence, knowledge of any two functions from the set $\{C_{ij}, R_{ij}, \hat{R}_{ij}\}$ is sufficient in determining the third function.

It is then apparent that a different type of correlation function must be defined when:

- Two waveforms have different periods. This is illustrated in Fig. 4.4. As seen from the figure, it is no longer meaningful to measure their cross-correlation in terms of periodic or aperiodic functions.

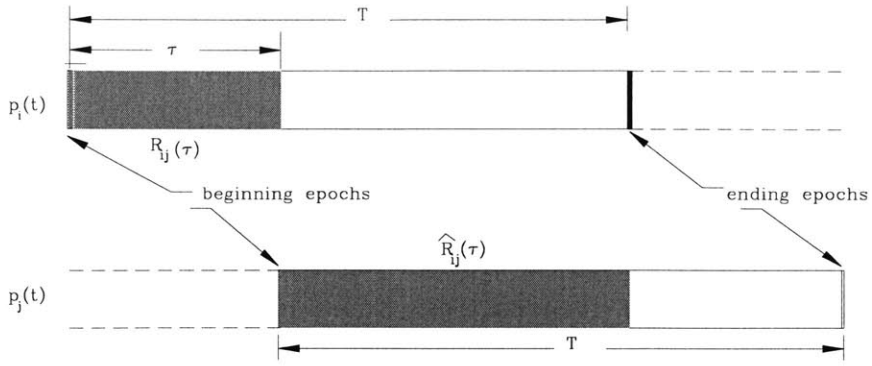



**Figure** 4.3: Beginning and ending epochs of aperiodic CC functions

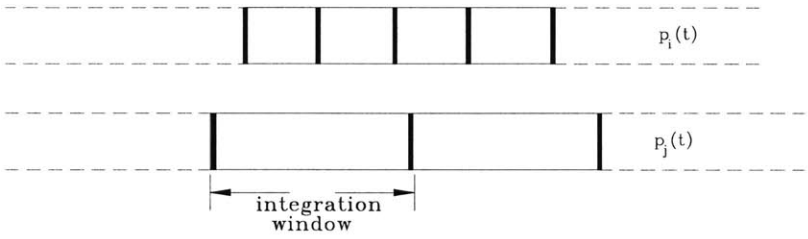



**Figure** 4.4: Correlation of two waveforms with different periods

- The beginning (ending) epoch of the correlation window is not the beginning (ending) epoch of a waveform, but rather shifted by an arbitrary non-zero offset. This condition exists even when both waveforms has the same period —as illustrated in Fig. 4.5.

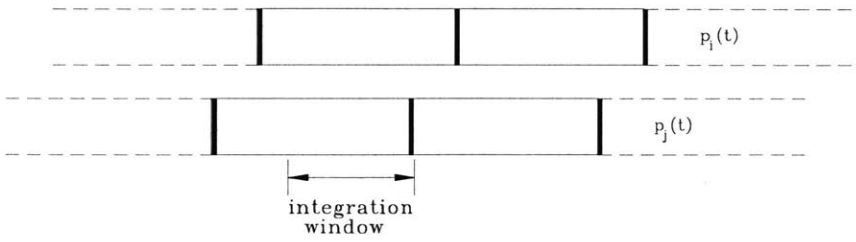



**Figure** 4.5: Correlation with a non-zero offset in time epoch

For an arbitrary offset $\tau$ and different periods between two waveforms, a cross-correlation function can still be defined. This is called the *partial cross-correlation* function. It is a three-parameter function: Besides the delay parameter $\tau$, it also depends on the start and stop points —or the length— of the integration window $(T_i, T_f)$. Please refer to Fig. 4.6 for a graphical

description. For the continuous case, the partial cross-correlation function is defined as:

$$\theta_{ij}(\tau; T_i, T_f) = \int_{T_i}^{T_f} p_i(t)p_j(t-\tau)\ dt \tag{4.12}$$

Of course,

$$\theta_{ij}(\tau;\tau, T) = R_{ij}(\tau)$$
$$\theta_{ij}(\tau;0,\tau) = \hat{R}_{ij}(\tau)$$
$$\theta_{ij}(\tau;0,T) = C_{ij}(\tau)$$

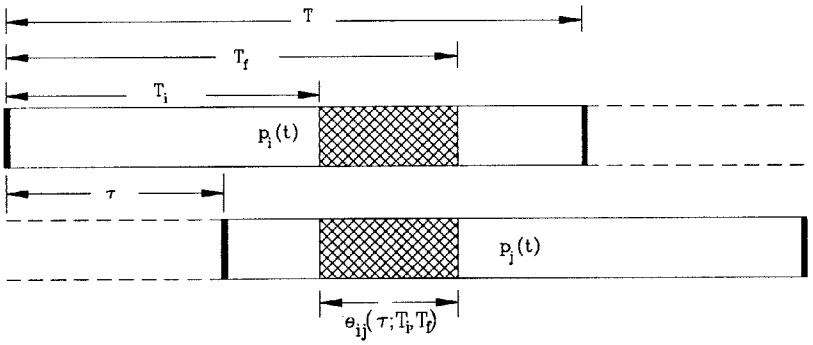



**Figure** 4.6: Partial cross-correlation function of two signature waveforms

### 4.1.4 Periodic, Aperiodic and Partial Auto-Correlation

For $i = j$, we obtain the periodic and aperiodic auto-correlation functions:

$$C_i(\tau) = \int_0^T p_i(t)p_i(t-\tau)dt \tag{4.13}$$

$$\hat{R}_i(\tau) = \int_0^\tau p_i(t)p_i(t-\tau)dt \tag{4.14}$$

$$R_i(\tau) = \int_\tau^T p_i(t)p_i(t-\tau)dt \tag{4.15}$$

Some care is required when dealing with the aperiodic auto-correlation functions. It is incorrect to denote $R_i(\tau) = R_{ii}(\tau)$; this implies $R_i(\tau) = R_{ii}(\tau) = \hat{R}_{ii}(\tau) = \hat{R}_i(\tau)$, which is a false statement. The partial auto-correlation function is:

$$\theta_i(\tau; T_i, T_f) = \int_{T_i}^{T_f} p_i(t)p_i(t-\tau)\ dt \tag{4.16}$$

For $\tau = 0$, the signal energy is:

$$\mathcal{E}_i = \int_0^T p_i^2(t)dt = \int_0^T f_i^2(t)dt \tag{4.17}$$

## 4.2 Correlation Properties of Discrete-Time Sequences

Consider a set of sequences $\mathbf{F}[n] = \{f_1[n], f_2[n], \ldots, f_M[n]\}$, where every sequence $f_j[n]$ has length (period) P. Using the same analogy as continuous-time waveforms, we can define the cross-correlation functions between two different sequences. Let $p_i[n]$ be the periodic repetition of $f_i[n]$. The aperiodic, periodic and partial cross-correlation functions are for $l \in \mathbb{N}^{P-1}$ are:

$$\hat{R}_{ij}[l] = \sum_{k=0}^{l-1} p_i[k]p_j[k-l] = R_{ji}[l] \tag{4.18}$$

$$R_{ij}[l] = \sum_{k=l}^{P-1} p_i[k]p_j[k-l] = \hat{R}_{ji}[l] \tag{4.19}$$

$$C_{ij}[l] = R_{ij}[l] + \hat{R}_{ij}[l] = C_{ji}[l] \tag{4.20}$$

$$\hat{C}_{ij}[l] = \hat{R}_{ij}[l] - R_{ij}[l] = -\hat{C}_{ji}[l] \tag{4.21}$$

$$\theta_{ij}[l; i, d] = \sum_{k=i}^{f} p_i[k]p_j[k-l] \tag{4.22}$$

where $d = f - i$. By setting $i = j$, the auto-correlation functions are $R_i[l]$, $\hat{R}_i[l]$, $C_i[l]$, $\hat{C}_i[l]$ and $\theta_i[l; i, d]$.

## 4.3 Signature Waveforms

In pulse-amplitude data modulation (PAM), $D = 1$ and $\mathbf{F}(t) = \{f(t)\}$. The shape and length of $f(t)$ is carefully designed to have special time-bandwidth characteristics. (See Chapter 5 for details on various pulse shapes and their time-frequency properties.) The two most common examples of $f(t)$ used in the literature are the brick-wall (rectangular) time-limited pulse $\Pi_T(t)$ and its frequency dual, the infinite-time sinc pulse $\Pi_W(t)$. T is used to denote the duration of the rectangular pulse, and $W$ is the bandwidth of the sinc pulse. A continuous stream of linearly data modulated PAM pulses is written as:

$$g(t) = \sum_{i=-\infty}^{\infty} x[i]f(t - iT) \tag{4.23}$$

where $x[i] \in \mathcal{X}$ is the information-bearing symbol from a PAM constellation set. In a multi-user setting, the pulse amplitude modulated signal for user k is:

$$g_k(t) = \sum_{i=-\infty}^{\infty} x_k[i]f_k(t - iT) \tag{4.24}$$

Note that $f_k(t)$ is now a special waveform chosen for user k. This is the *signature waveform* and it can also be expressed as a *spread-spectrum* modulated waveform:

$$f_k(t) = \sum_{n=0}^{N_c-1} a_k[n]\psi(t - nT_c) \qquad 0 \le t \le T \tag{4.25}$$

where $\{a_k[n]\}$ is the $k^{th}$ user's embedded, antipodal binary spread-spectrum (SS) sequence, i.e. $a_k[n] \in \mathbb{D}_2$ for $\forall\, k, n$. The common *chip* pulse $\psi(t)$ has chip duration $T_c$, and P is the period of the embedded binary SS sequence. The ratio of symbol-to-chip time intervals $N_c = T/T_c$ is the *Spreading Gain* (SG). Modulation by a chip sequence of rate $1/T_c$ effectively spreads or expands the bandwidth by $N_c$. In eqn. (4.25) we assume that the length (period) of the embedded sequence $\{a_k[n]\}$ is $N_c$. In general, a code sequence with a larger period P ($\gg N_c$) is commonly used to reduce self-induced interference. (See Sec. 4.6 for details.) Taking this into account, a spread-spectrum modulated, PAM signal can be rewritten as

$$g_k(t) = \sum_{i=-\infty}^{\infty} x_k[i] \sum_{n=iN_c}^{(i+1)N_c-1} a_k[n]\,\psi(t - nT_c) \tag{4.26}$$

An alternative expression is:

$$\begin{aligned} g_k(t) &= \sum_{i=-\infty}^{\infty} \sum_{n=0}^{N-1} x_k[i]a_k[n]\psi(t - iT - nT_c) \\ &= \sum_{i=-\infty}^{\infty} b_k[i]\psi(t - iT_c) \end{aligned} \tag{4.27}$$

where $b_k[i]$ can be interpreted as the *data modulated* signature sequence. For notational convenience, we drop the subscript and use N (instead of $N_c$) to denote the spreading gain. (In later sections where we describe orthogonal codes, N is used to denote the length of an orthogonal code and it may not equal $N_c$.) In effect signature waveforms are analog versions of the embedded discrete-time signature sequences. There exists four kinds of periods:

$T$ = Period of the signature waveform (in seconds)

$T_c$ = Period of the common chip waveform (in seconds)

P = Period of the embedded signature sequence (in samples)

$N_c$ = the number of chips contained in a period of a signature waveform

The various cross-correlation functions of any two signature waveforms can be expressed in terms of the cross-correlation function of the embedded sequence, plus the autocorrelation function of the common chip waveform. Depending on the relation between $N_c$ and P and the time epochs of the correlation window, the cross-correlation of two signature waveforms is either periodic, aperiodic or partial. We next study all three cases.

### 4.3.1 Periodic Correlation

Two different signature waveforms are orthogonal if:

$$\int_0^T f_k(t)\, f_j(t)\, dt = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a_k[n] a_j[m] \int_0^T \psi(t - nT_c)\psi(t - mT_c)\, dt = 0 \tag{4.28}$$

If $\psi(t) = \Pi_{T_c}(t)$, then orthogonality is satisfied when:

$$\sum_{n=0}^{N-1} a_k[n] a_j[n] = 0 \tag{4.29}$$

In general, it is impossible to design code sequences that remain orthogonal for any arbitrary time (chip) offset. For a special kind of code sequences called *superimposed orthogonal*, any two of them are mutually orthogonal if the sequences are *code synchronized*; that is, their beginning time epochs are aligned. In a code-division multiplexing, time synchronization can be classified as one of the following four:

1. *code synchronous*: The beginning time epochs of two spread-spectrum modulated (SSM) waveforms are aligned. It is implicitly assume that both waveforms have identical sequence period P and chip duration $T_c$.

2. *bit synchronous*: The beginning time epochs of two signature waveforms are aligned. It is implicitly assume that both waveforms have period T. In terms of SSM waveforms, there may exist a time offset equal to an integer multiple of T between two bit-synchronous signals.

3. *chip synchronous*: The beginning time epochs of two common chip waveforms are aligned. This is the weakest time alignment among them. Two SSM waveforms may be offsetted by a duration equal to an integer multiple of $T_c$.

4. *phase synchronous*: Phase synchronism arises only when CDM signals are RF/IF carrier modulated. Each carrier-modulated signal from a base site i has its own absolute carrier phase $\theta_i$ (measured according to some universal reference site) plus an additional phase shift induced by relative propagation delay $\tau_i$ such that the total phase offset is

   $$\phi_i = -2\pi f_o \tau_i + \theta_i$$

   For convenience, we assume the desired signal (from site 0) is demodulated phase coherently —i.e., $\phi_0 = 0$. A received signal is phase synchronous with the desired signal if $\phi_i = 2k\pi$ for any integer k.

It is obvious that code synchronism implies bit and chip synchronism, but not vice versa. Likewise two bit-synchronous waveforms are also chip-synchronous. Chip synchronism does not guarantees code or bit synchronism. There's no relation between phase and other time (code,bit or chip) synchronism; the former is a function of the carrier phase $\phi$ of analog waveforms while the latter three depend on the value of the relative propagation delay $\tau$ in relation to chip duration $T_c$ —i.e., whether $\tau$ is a multiple of $N_pT_c$, $NT_c$ or only $T_c$.

### 4.3.2 Aperiodic Correlation

When two spread-spectrum modulated signals $s_i(t)$ and $s_j(t)$ (both with period T) are also data modulated, then the expression of their periodic cross-correlation $C_{ij}(\tau)$ must take into account the numerical values of different data bits that are contained in the correlation window. This is illustrated in Fig. 4.7 where two data bits $\{x_j[-1], x_j[0]\}$ of user j overlap the correlation window of data bit $x_i[0]$ for user i.

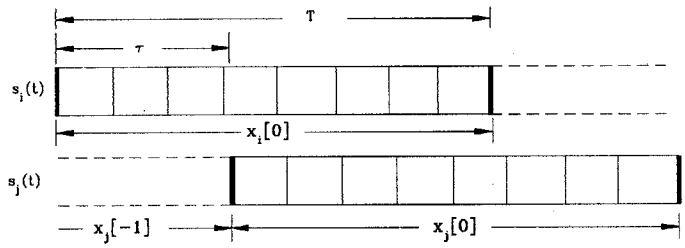



**Figure** 4.7: Overlapping of two different data bits of user j in the correlation window

When $N_c = P$, the continuous-time periodic cross-correlation $C_{ij}(\tau)$ is expressed in terms of the discrete-time aperiodic cross-correlation functions $\{R_{ij}[l], \hat{R}_{ij}[l]\}$ and the continuous-time auto-correlation functions $\{R_\psi(\tau), \hat{R}_\psi(\tau)\}$ of the common chip pulse. We consider two different cases where the time offset $\tau$ is and is not a multiple of chip duration $T_c$. In the following derivations, we consider cross-correlation between signature waveforms of two different users; however, the results are equally applicable to auto-correlation of a single signature waveform by replacing discrete cross-correlation with discrete auto-correlation.

**Case 1:** $\tau = lT_c$, $l \in \mathbb{N}^{N-1}$

When the delay $\tau = lT_c$ is a multiple of chip duration, the two signature waveforms are *chip-synchronous*, and the continuous-time aperiodic cross-correlation functions do not depend on

the shape of $\psi(t)$. This is illustrated in Fig. 4.8.

$$\begin{aligned}\hat{R}_{ij}(lT_c) &= \int_0^{lT_c} g_i(t)g_j(t-lT_c)\,dt \\ &= \sum_{k=0}^{l-1} b_i[k]\, b_j[k-l] \int_0^{T_c} \psi^2(t)\,dt \\ &= \hat{R}_{ij}[l]\, \mathcal{E}_\psi\end{aligned} \tag{4.30}$$

Similarly, it can be shown that

$$R_{ij}(lT_c) = R_{ij}[l]\, \mathcal{E}_\psi \tag{4.31}$$

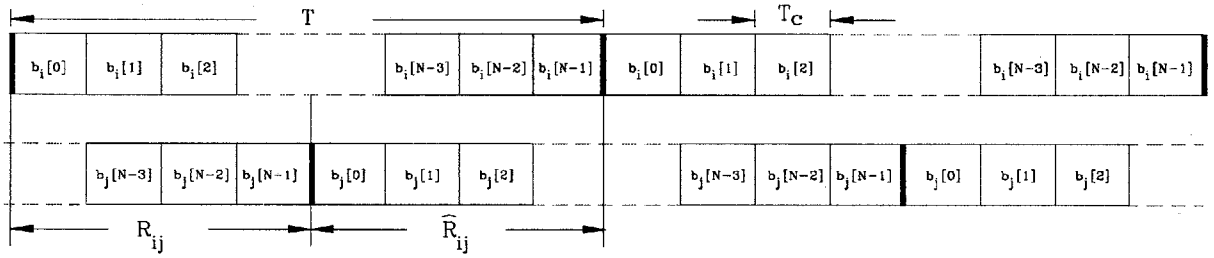



**Figure** 4.8: Aperiodic cross-correlation for chip-synchronous signature waveforms

**Case 2:** $\tau = lT_c + \Delta$, $l \in \mathbb{N}^{N-1}$, $0 \le \Delta < T_c$

If the two signature waveforms are *chip-asynchronous*, the continuous-time aperiodic cross-correlation functions depend on the shape of $\psi(t)$ through its aperiodic autocorrelation functions, $\rho_\psi(\tau)$ and $\hat{\rho}_\psi(\tau)$. (See eqns. (4.3) and (4.4)). As shown in Fig. 4.9, each chip of user j overlaps with two consecutive chips of user i. The correlation function is then broken up into two parts: discrete correlations with the first chip and then with the second chip of i, respectively. Each discrete correlation is weighed by the continuous auto-correlation of the chip pulse. This is illustrated by shaded regions in Fig. 4.10.

$$\begin{aligned}\hat{R}_{ij}(lT_c + \Delta) &= \int_0^{lT_c+\Delta} g_i(t)g_j(t-lT_c)\,dt \\ &= \sum_{k=0}^{l-1} b_i[k]\, b_j[k-l] \int_\Delta^{T_c} \psi(t)\psi(t-\Delta)\,dt \\ &+ \sum_{k=0}^{l} b_i[k]\, b_j[k-l] \int_0^{\Delta} \psi(t)\psi(t+T_c-\Delta)\,dt \\ &= \hat{R}_{ij}[l]\, \rho_\psi(\Delta) + \hat{R}_{ij}[l-1]\, \hat{\rho}_\psi(\Delta)\end{aligned} \tag{4.32}$$

$$
\begin{aligned}
R_{ij}(lT_c + \Delta) &= \int_0^{lT_c+\Delta} g_i(t) g_j(t - lT_c)\, dt \\
&= \sum_{k=l}^{N-1} b_i[k]\, b_j[k-l] \int_{\Delta}^{T_c} \psi(t)\psi(t-\Delta)\, dt \\
&+ \sum_{k=l+1}^{N-1} b_i[k]\, b_j[k-l] \int_0^{\Delta} \psi(t)\psi(t+T_c-\Delta)\, dt \\
&= R_{ij}[l]\, \rho_\psi(\Delta) + R_{ij}[l+1]\, \hat{\rho}_\psi(\Delta)
\end{aligned}
\tag{4.33}
$$

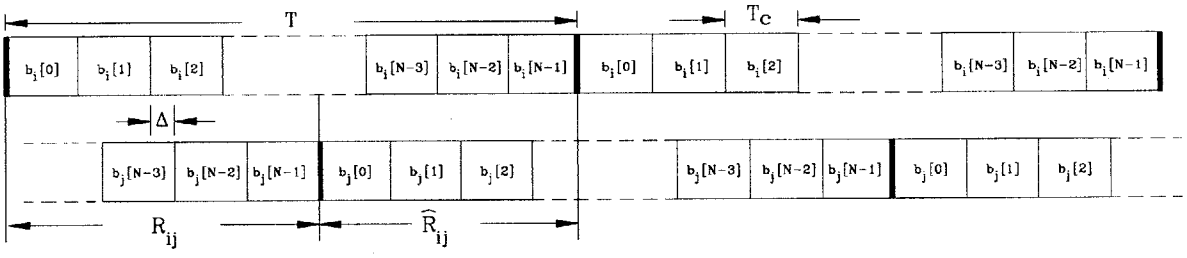



**Figure** 4.9: Aperiodic cross-correlation for chip-asynchronous signature waveforms

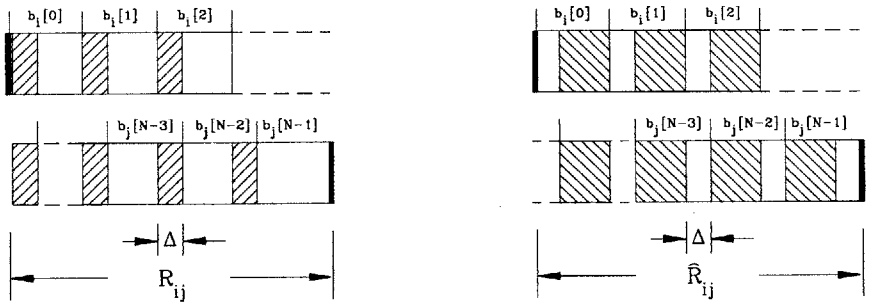



**Figure** 4.10: Two portions of periodic cross-correlation for chip-asynchronous waveforms

### 4.3.3 Partial Correlation

As demonstrated so far, when the period of the discrete signature sequence P equals the number of chips $N_c = T/T_c$ of the correlation window, the continuous-time aperiodic cross-correlation functions can be expressed in terms of their discrete-time counterparts and the autocorrelation of $\psi(t)$; i.e., there is an isomorphism

$$
\left\{ C_{ij}(t), R_{ij}(t), \hat{R}_{ij}(t) \right\} \longrightarrow \left\{ C_{ij}[n], R_{ij}[n], \hat{R}_{ij}[n], \rho_\psi(t), \hat{\rho}_\psi(t) \right\}
$$

This is a desired property for system analysis since there exists closed-form results on the periodic and aperiodic correlation properties of discrete sequences. However, in spread-spectrum modulated signals, the period of a code sequence is much longer than the spreading gain, $P \gg N$. Under such conditions the discrete-time *aperiodic* cross-correlation functions are replaced with their *partial* counterparts. If $T_i = iT_c$ and $T_f = fT_c$, $i, f \in \mathbb{Z}$, and the difference $d = f - i$, $f > i$,

$$\theta_{ij}(\tau; T_i, T_f) = \theta_{ij}(\tau; [i, d])$$

At the receiving end, the symbol matched filter typically integrates over the duration $T = T_f - T_i$, and we are interested in

$$\theta_{ij}(\tau; [i, N]) = \int_{iT_c}^{(i+N)T_c} g_i(t) g_j(t - \tau)\, dt \tag{4.34}$$

We next consider two different cases of time offset:

**Case 1:** $\tau = lT_c$, $l \in \mathbb{N}^{P-1}$

The two signature waveforms are *chip-synchronous*, and the partial cross-correlation function does not depend on the shape of $\psi(t)$.

$$\begin{aligned} \theta_{ij}(lT_c; [i, N]) &= \int_{iT_c}^{(i+N)T_c} g_i(t) g_j(t - lT_c)\, dt \\ &= \sum_{k=i}^{i+N-1} b_i[k]\, b_j[k-l] \int_0^{T_c} \psi^2(t)\, dt \\ &= \theta_{ij}[l; i, N]\, \mathcal{E}_\psi \end{aligned} \tag{4.35}$$

**Case 2:** $\tau = lT_c + \Delta$, $l \in \mathbb{N}^{N-1}$, $0 \le \Delta < T_c$

The two signature waveforms are *chip-asynchronous*, and the continuous-time partial cross-correlation function depend on the shape of $\psi(t)$ through its aperiodic autocorrelation functions, $\rho_\psi(\tau)$ and $\hat{\rho}_\psi(\tau)$.

$$\begin{aligned} \theta_{ij}(lT_c + \Delta; [i, N]) &= \int_{iT_c}^{(i+N)T_c} g_i(t) g_j(t - lT_c - \Delta)\, dt \\ &= \sum_{k=i}^{i+N-1} b_i[k]\, b_j[k-l] \int_\Delta^{T_c} \psi(t)\psi(t - \Delta)\, dt \\ &+ \sum_{k=i+1}^{i+N-1} b_i[k]\, b_j[k-l] \int_0^{\Delta} \psi(t)\psi(t + T_c - \Delta)\, dt \\ &= \theta_{ij}[l; i, N]\, \rho_\psi(\Delta) + \theta_{ij}[l-1; i, N]\, \hat{\rho}_\psi(\Delta) \end{aligned} \tag{4.36}$$

## 4.4 Pseudo-Random Binary Codes

In the previous sections, we show that a signature waveforms is a *spread-spectrum modulated* signal with its associated signature code sequence and a common chip pulse. The continuous time partial cross-correlation function of two signature waveforms can be expressed in terms of the discrete-time partial cross-correlation functions of their respective code sequences and the continuous-time partial auto-correlation functions of the chip pulse. In this section, we concentrate on the correlation properties of the embedded discrete code sequences —in particular, the binary-valued pseudo-random sequences. First, we present some definitions and preliminaries.

### 4.4.1 Orthogonality

In terms of notation two options exist when expressing binary sequences: one where the elements are from the Galois binary field $\mathbb{F}_2 = \{0, 1\}$, and another where they are from the integer ring $\mathbb{D}_2 = \{1, -1\}$. For example, two binary sequences or codes of length N can be expressed as:

$$\mathbf{b}_i = \Big(b_i[0], b_i[1], b_i[2], \ldots, b_i[N-2], b_i[N-1]\Big)$$

$$\mathbf{b}_j = \Big(b_j[0], b_j[1], b_j[2], \ldots, b_j[N-2], b_j[N-1]\Big)$$

where every element $(b_m[n]) \in \mathbb{F}_2$, $m = i, j$, $n \in \mathbb{N}^{N-1}$. Their inner-product over a binary field is:

$$\begin{aligned} \mathbf{b}_i \odot \mathbf{b}_j &= \sum_{n=0}^{N-1} b_i[n] \odot b_j[n] \\ &= b_i[0] \odot b_j[0] \oplus b_i[1] \odot b_j[1] \oplus \cdots \oplus b_i[N-1] \odot b_j[N-1] \end{aligned} \tag{4.37}$$

It is understood that for elements from $\mathbb{F}_2$, the summation and multiplication operations are *binary* modulo-2 addition '$\oplus$' and multiplication '$\odot$'. The truth tables for addition and multiplication in $\mathbb{F}_2$ are given in Table 4.1.

**Table** 4.1: Addition and Multiplication in $\mathbb{F}_2$ and $\mathbb{D}_2$

| A | B | A ⊕ B | A ⊙ B |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 |

| A | B | A + B | A · B |
|---|---|---|---|
| +1 | +1 | +2 | +1 |
| +1 | −1 | 0 | −1 |
| −1 | −1 | −2 | +1 |

Similarly, if we define two *antipodal* binary sequences of length N as:

$$\mathbf{c}_i = \Big(c_i[0], c_i[1], c_i[2], \ldots, c_i[N-2], c_i[N-1]\Big)$$

$$\mathbf{c}_j = \Big(c_j[0], c_j[1], c_j[2], \ldots, c_j[N-2], c_j[N-1]\Big)$$

where every element $(c_m[n]) \in \mathbb{D}_2$, $m = i, j$, $n \in \mathbb{N}^{N-1}$, then their inner-product over an integer ring is:

$$\begin{aligned} \mathbf{c}_i \cdot \mathbf{c}_j &= \sum_{n=0}^{N-1} c_i[n] \cdot c_j[n] \\ &= c_i[0] \cdot c_j[0] + c_i[1] \cdot c_j[1] + \cdots + c_i[N-1] \cdot c_j[N-1] \end{aligned} \tag{4.38}$$

Here, the arithmetic operations are real addition '+' and multiplication '$\cdot$'. (For convenience we often drop the "dot" in real multiplication.) In fact, the inner product also applies to any pair of sequences whose elements are from a Hilbert space.

### A Word on Notations

In terms of notations for arithmetic operators, any Galois field operation (addition '+', scalar (inner-product) multiplication '$\cdot$' or vector (outer-product) multiplication '$\times$' ) is closed in $\mathbb{F}_2$; thus, each corresponding symbol $\oplus$, $\odot$ or $\otimes$ is enclosed in a circle. For regular arithmetic operators, we use conventional notations: '+' and '$\cdot$'. The multiplication signs '$\cdot$' and '$\times$' are equivalent. For convenience, we often neglect to put neither; i.e.,

$$u \cdot v = u \times v = uv$$

The vector (outer-product) multiplication is defined as component-wise multiplication of two vectors of equal length. For Galois binary vectors,

$$\mathbf{b}_i \otimes \mathbf{b}_j = \Big[b_i[0] \odot b_j[0],\ b_i[1] \odot b_j[1],\ \ldots,\ b_i[N-1] \odot b_j[N-1]\Big]$$

For antipodal vectors,

$$\mathbf{c}_i \otimes \mathbf{c}_j = \Big[c_i[0] \cdot c_j[0],\ c_i[1] \cdot c_j[1],\ \ldots,\ c_i[N-1] \cdot c_j[N-1]\Big]$$

Depending on the type of vectors being studied, we can easily determine if the scalar multiplier is '$\cdot$' or '$\odot$'. Some authors also define $\mathbf{b}_i \otimes \mathbf{b}_j$ as the *Hadamard product*.

## 4.4.2 Level-Shifting

There is a bijective mapping between $b_i[n]$ and $c_i[n]$:

$$\begin{aligned} \mathbb{D}_2 &\Longleftrightarrow \mathbb{F}_2 \\ 1 &\longleftrightarrow 0 \\ -1 &\longleftrightarrow 1 \end{aligned} \tag{4.39}$$

From the truth table, it is immediate that modulo-2 addition $\oplus$ is equivalent to real multiplication, i.e.,

$$\begin{aligned} 0 \oplus 0 = 0 \quad &\longleftrightarrow \quad (+1)\cdot(+1) = +1 \\ 0 \oplus 1 = 1 \quad &\longleftrightarrow \quad (+1)\cdot(-1) = -1 \\ 1 \oplus 1 = 0 \quad &\longleftrightarrow \quad (-1)\cdot(-1) = +1 \end{aligned}$$

Another convention that will be used often is the "sign" description with the following map-

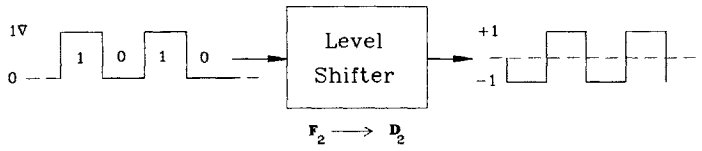



**Figure** 4.11: Conversion from finite binary to antipodal signals. This operation is known as level-shifting in circuit terminology. It can be implemented using a gain-2 amplifier with capacitor coupling at the output.

ping: $1 \rightarrow +$ and $-1 \rightarrow -$. It is the author's belief that the sign description of an antipodal sequence is more pleasing to the eye:

$$+++-+-- \quad \text{instead of} \quad 1,1,1,-1,1,-1,-1$$

The mapping between finite binary and antipodal sequences is called *level-shifting*. As shown in Fig. 4.11, a level-shifter is simply an amplifier with gain 2 and a dc bias of $-1$ volts.[3] In terms of notation, an upper-case letter is used to label a Galois binary sequence and its level-shifted binary sequence is denoted by a lower-case letter.

It is important to note that there is no equivalent mathematical representation of finite field multiplication $\odot$ in terms of real arithmetic operators. Similarly, the real product of two integers has no equivalent description in the finite binary field. The inner product of two binary antipodal sequences $\mathbf{c}_i \cdot \mathbf{c}_j \in \mathbb{Z}$ has an integer value, whereas the Galois inner-product is a closed operation: $\mathbf{b}_i \cdot \mathbf{b}_j \in \mathbb{F}_2$. These observations lead to algebraic results that cannot be transported from the finite field to the integer ring (or real field), or vice versa. For example, it is possible for a sequence $(1,1,1,1) \in (\mathbb{F}_2)^4$ to be self-orthogonal (i.e, orthogonal to itself) while no such vector exists over the real space. To avoid any confusion, our definition of orthogonality applies to the inner product of two antipodal signals over real space $\mathbb{R}$.

[3] A reader familiar with circuit theory knows that it can be implemented using a common-collector (or common-drain) amplifier with a gain of 2 and capacitor coupling at the output.

### 4.4.3 Properties of Signature Binary Sequences

For multi-user communications, we prefer (binary) signature sequences with the following statistical properties:

R-1: The sequence $\ldots, b_i[-2], b_i[-1], b_i[0], b_i[1], b_i[2], \ldots$ appears random, such that the sequence is memoryless and the outcome of each component $b_i[k] \in \mathbb{D}_2$ of the sequence is a fair coin flip: $P(b_i[k] = +1) = P(b_i[k] = -1) = 1/2$. Equivalently, the components are i.i.d. RVs with *bipolar symmetric Bernoulli* distribution (see Appendix 2B). The statistical mean and variance of the partial auto–correlation function of a *bipolar Binomial* distributed sequence $b_j$ are

$$\begin{aligned} E\Big[\theta_i[l; s, N]\Big] &= \sum_{k=s}^{s+N-1} E\Big[b_i[k] b_i[k-l]\Big] \\ &= \begin{cases} N & l = 0 \\ 0 & |l| \neq 0 \end{cases} \end{aligned} \tag{4.40}$$

$$\mathrm{Var}\Big[\theta_i[l; s, N]\Big] = \begin{cases} 0 & l = 0 \\ N & |l| \neq 0 \end{cases} \tag{4.41}$$

The derivation is straightforward. We know $\mathrm{Var}(\cdot) = E(\cdot)^2 - E^2(\cdot)$. By explicitly writing out the mean-squared value of the partial auto-correlation function for $l \neq 0$:

$$E\Big[\theta_i[l; s, N]\Big]^2 = E\left[\sum_{k=s}^{s+N-1} \sum_{m=s}^{s+N-1} b_i[k] b_i[k-l]\, b_i[m] b_i[m-l]\right]$$

Splitting into two summations of diagonal terms ($m = k$) and off-diagonal terms ($m \neq k$),

$$= E\left[\sum_{k=s}^{N} b_i^2[k] b_i^2[k-l] + 2 \sum_{k=s}^{s+N-2} \sum_{m=k+1}^{s+N-1} b_i[k] b_i[m] b_i[k-l] b_i[m-l]\right]$$

Since $E\Big[b_i[k] b_i[m]\Big] = 0$ for $k \neq m$, the second summation is zero,

$$E\Big[\theta_i[l; s, N]\Big]^2 = N$$

Combining with eqn. (4.40), the variance is as given in eqn. (4.41). The mean of the autocorrelation for zero offset is

$$\sum_{k=s}^{s+N-1} E\Big[\, b_i^2[k] \,\Big] = N$$

This is the "energy" $\mathcal{E}_i$ of a signature binary sequence of length N. Some authors prefer to normalize the chip value to $b_i[k] = \pm\frac{1}{\sqrt{N}}$ such that the resulting energy is unity. In this case the mean and variance of the *normalized* partial auto-correlation function can be expressed as

$$\begin{aligned} E\Big[\Theta_i[l;s,N]\Big] &= \sum_{k=s}^{s+N-1} E\Big[b_i[k]b_i[k-l]\Big] \\ &= \begin{cases} 1 & l=0 \\ 0 & |l|\neq 0 \end{cases} \end{aligned} \tag{4.42}$$

$$\mathrm{Var}\Big[\Theta_i[l;s,N]\Big] = \begin{cases} 0 & l=0 \\ \dfrac{1}{N} & |l|\neq 0 \end{cases} \tag{4.43}$$

where we have used the upper-case $\Theta$ to denote normalized partial correlation.

R-2: The binary sequence can be easily generated and synchronized at the designated receiver but nevertheless it appears random to other interfering receivers. The randomness condition is equivalent to

$$E\Big[\theta_{ij}[l;s,N]\Big] = 0 \tag{4.44}$$

$$\mathrm{Var}\Big[\theta_{ij}[l;s,N]\Big] = N \tag{4.45}$$

for all $|l| \neq 0$. In terms of normalized partial correlation:

$$E\Big[\Theta_{ij}[l;s,N]\Big] = 0 \tag{4.46}$$

$$\mathrm{Var}\Big[\Theta_{ij}[l;s,N]\Big] = 1/N \tag{4.47}$$

R-3: The number of such sequences must be large enough to assign a unique sequence to every transmitter-receiver pair in the network.

### 4.4.4 PN Sequence Postulates

The second requirement R-2 of a signature binary sequence stipulates that the sequence must be deterministic while possessing the "randomness" property. Such a sequence is pseudo-random. A binary sequence $b_i = \big(b_i[1],\ b_i[2], \ldots, b_{N_p}\big)$ of period $N_p$ where each element $b_i[k] \in \mathbb{D}_2$ is called *pseudo-random* if it satisfies the following postulates:

P-1: At most $|N_+ - N_-| = 1$ where $N_+$ and $N_-$ are the number of elements with '+1's and '−1's, respectively in $b_i$. Equivalently, $\Big|\sum_{n=1}^{N_p} b_i[n]\Big| \leq 1$.

P-2: A run of length r is defined as a string of consecutive r '+1's or '−1's. In a pseudo-random sequence $b_i$, if there are R runs,

- There are equal $(\frac{R}{2})$ number of +1 and −1 runs.
- There are $\frac{R}{n}$ runs of length $n/2$. For example, there are $\frac{R}{2}$ runs of length 1, $\frac{R}{4}$ runs of length 2, $\frac{R}{8}$ runs of length 4 and so on as long as $\frac{R}{n} \geq 1$.

P-3: The periodic auto-correlation function $C_i[l]$ is two-valued:

$$C_i[l] = \begin{cases} N_p & l = 0 \\ K & |l| \in \mathbb{Z}^{N_p - 1},\ N_p \gg K \end{cases} \tag{4.48}$$

### 4.4.5 Comments on Pseudo-Randomness

1. P-1 guarantees the randomness of the elements that make up the binary sequence over a period $N_p$. It, however, does not imply that the sequence itself is random. A non-random periodic sequence can also satisfy this property. For example, the alternating $\ldots + - + - + - \ldots$ and half-flip $\ldots + + + - - - \ldots$ sequences (and the orthogonal sequences introduced in Sec. 4.5) satisfy P-1 with $|N_+ - N_-| = 0$ (even $N_p$) despite the fact that they all highly "symmetric." Furthermore, as we shall see in the next section, Walsh orthogonal sequences have very large auto-correlation values for non-zero time-shifts. Thus, P-1 implies neither P-2 or P-3.

2. P-2 guarantees the randomness of $b_i$ over period $N_p$. As an example, consider the following pseudo-random sequence of length 7:

$$+ + + - + - -$$

There are 4 runs. They are $(+ + +)$, $(-)$, $(+)$ and $(--)$. (Since the number of +1 runs must equal the number of −1 runs, the total number of runs is necessarily an even integer.) There are two '+' runs and two '−' runs, satisfy the first condition of P-2. There are two runs of length 1 and a single run of length 2; hence, the second condition of P-2 is met. Note that the length of $(+ + +)$ run is irrelevant. P-2 is still satisfied if an additional '+' is added as a leading element to the sequence. However, the modified sequence:

$$+ + + + - + --$$

violates P-1. It is easily checked that it no longer satisfies P-3. Thus, P-2 implies neither P-1 or P-3.

3. As we saw in the above example, the length-7 sequence concatenated with a prefix of all "1"s, $\ldots + + + \ldots$ still satisfies P-2. Such a modified sequence, however, will have a large auto-correlation value for non-zero time-shift. P-3 guarantees that the periodic auto-correlation be very small for arbitrary non-zero phase-shifts for elements that are generated pseudo-randomly. It is highly unlikely that a string of heads will be observed if a fair coin is flipped repeatedly. However, it is fairly easy for non-random periodic sequences to satisfy P-3. The most obvious case is the impulse sequence $\delta[n] = [\ldots 00010000 \ldots]$. (It violates both P-1 and P-2.) As another less obvious example, consider the following sequence:

$$+ - + + - - + -$$

It can be checked that this sequence passes the test for both P-1 and P-2, but fails P-3. Thus, P-3 guarantees neither P-1 or P-2.

4. Based on above observations, we conclude that the three postulates are independent, and a sequence is "pseudo-random" only if all three conditions are satisfied.

### 4.4.6 Maximal-length Binary Sequences

There is a wealth of texts and research articles that analyze and characterize the unique properties of maximal-length (ML) binary sequences. The sources we have relied on are listed at the end of the chapter. We simply state that ML sequences satisfy all three pseudo-random postulates; i.e., *they qualify as pseudo-random sequences*. Its period is odd: $N_p = 2^n - 1$, $n \in \mathbb{Z}_+$. In one realization, the sequence is the output of a linear feedback shift-register (LFSR) with $n$ states (flip-flops). This is illustrated in Fig. 4.12.

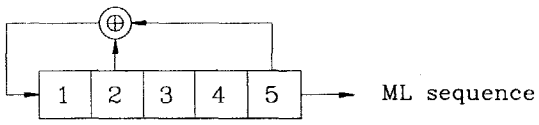



**Figure 4.12:** Linear feedback shift-register for ML sequence generation. The tap connections are [2,5].

The positions of tap connections identify a particular ML sequence. In the figure shown, the taps are [2, 5]. The sequence is *maximal-length* since the register *states* are all possible binary sequences of length $n$ except the all-zero sequence. Writing out the [2, 5] ML sequence explicitly:

$$\ldots 1111100110100100001010111011010 \ldots$$

Note that $(N_+ - N_-) = 1$; this is the *balance* property. The above sequence is generated with an initial state [11111]. If we start with a different initial state, we obtain the cyclically shifted

version of the above sequence. For example, with an initial state [01100] the resulting sequence is a right cyclic shift of 5 of the original sequence:

$$\ldots 0011010010000101011101101011111\ldots$$

For an ML sequence, the binary sum of an original sequence and its cyclically-shifted version is the sequence itself with another cyclic shift. For the above sequences, their sum is:

$$\ldots 0011010010000101011101101011111\ldots$$

By observation this is a right shift of 3 of the original sequence. This is the *shift-and-add* property. Using this property it can be shown that the *periodic* auto-correlation of an ML sequence is two-valued:

$$C_j[l] = \begin{cases} N_p & l = 0 \\ -1 & |l| \in \mathbb{Z}^{N_p-1} \end{cases} \tag{4.49}$$

The postulates do not reveal much about the partial auto-correlation properties of a PN sequence. For ML sequences the following hold:

- The average partial auto-correlation is equal to the periodic auto-correlation:

$$E_s\left[\theta_j[l; s, N]\right] = C_j[l] \tag{4.50}$$

- The variance of the partial auto-correlation function is:

$$\mathrm{Var}_s\left[\theta_j[l; s, N]\right] = \frac{1}{N}\left(1 + \frac{1}{N_p}\right)\left(1 - \frac{N}{N_p}\right) \tag{4.51}$$

The averaging is over all possible starting positions:

$$E_s\left[\theta_j[l; s, N]\right] = \frac{1}{N_p}\sum_{k=0}^{N_p-1}\left[\sum_{s=0}^{N-1} p_j[s+k]p_j[l+s+k]\right]$$

Swapping the summations:

$$= \frac{1}{N}\sum_{s=0}^{N-1}\left[\sum_{k=0}^{N_p-1} p_j[s+k]p_j[l+s+k]\right]$$
$$= \frac{1}{N}\sum_{s=0}^{N-1} C_j[l] = C_j[l]$$

In terms of the normalized partial auto-correlation function:

$$E_s\left[\Theta_j[l; s, N]\right] = \begin{cases} 1 & l = 0 \\ -1/N_p & |l| \in \mathbb{Z}^{N_p-1} \end{cases} \tag{4.52}$$

We know $\mathrm{Var}(\cdot) = \mathrm{E}(\cdot)^2 - \mathrm{E}^2(\cdot)$. By explicitly writing out the mean-squared value of the partial auto-correlation function:

$$\begin{aligned}\mathrm{E}_s\left[\theta_j^2[l;s,N]\right] &= \frac{1}{N_p}\sum_{k=0}^{N_p-1}\left[\sum_{s=0}^{N-1} p_j[s+k]p_j[l+s+k]\right]^2 \\ &= \frac{1}{N_p}\sum_{k=0}^{N_p-1}\left[\sum_{s=0}^{N-1}\sum_{q=0}^{N-1} p_j[s+k]p_j[l+s+k]p_j[q+k]p_j[l+q+k]\right]\end{aligned}$$

Using the shift-and-add property of an ML sequence, we know that

$$p_i[s+k]\cdot p_j[l+s+k] = p_j[l+s'+k]$$

is another cyclically shifted sequence. Similarly,

$$p_j[q+k]\cdot p_j[l+q+k] = p_j[l+q'+k]$$

Substituting:

$$\mathrm{E}_s\left[\theta_j^2[l;s,N]\right] = \frac{1}{N_p}\sum_{k=0}^{N_p-1}\left[\sum_{s'=0}^{N-1}\sum_{q'=0}^{N-1} p_j[l+s'+k]p_j[l+q'+k]\right]$$

Splitting the double summation term in the bracket into diagonal and off-diagonal terms:

$$= \frac{1}{N_p}\sum_{k=0}^{N_p-1}\sum_{s'=0}^{N-1} p_j^2[l+s'+k] + + \frac{1}{N_p}\sum_{k=0}^{N_p-1}\left[\sum_{s'=0}^{N-2}\sum_{q'=s}^{N-1} p_j[l+s'+k]p_j[l+q'+k]\right]$$

Using the shift-and-add property,

$$p_j[l+s'+k]p_j[l+q'+k] = p_j[l+r'+k]$$

for some integer $r' \neq s'$ and $r' \neq q'$. Using this result and simplifying

$$\mathrm{E}_s\left[\theta_j^2[l;s,N]\right] = N + \frac{1}{N_p}\sum_{s'=0}^{N-2}\sum_{q'=s}^{N-1}\sum_{k=0}^{N_p-1} p_j[l+r'+k]$$

Invoking the balance property,

$$\begin{aligned}&= N - \frac{N(N-1)}{N_p} \\ &= \frac{N}{N_p}\left(N_p - N + 1\right)\end{aligned}$$

which leads to

$$\begin{aligned} \mathrm{Var}_s\big[\theta_j[l;s,N]\big] &= \frac{N}{N_p}\,(N_p - N + 1) - (-1)^2 \qquad \text{for } l \neq 0 \\ &= (N-1)\left(1 - \frac{N}{N_p}\right) \end{aligned} \tag{4.53}$$

It is readily shown that for the normalized partial auto-correlation:

$$\begin{aligned} \mathrm{Var}_s\big[\Theta_j[l;s,N]\big] &= \frac{1}{N} - \frac{N-1}{N_p N} - \left(-\frac{1}{N_p}\right)^2 \qquad \text{for } l \neq 0 \\ &= \frac{1}{N}\left(1 + \frac{1}{N_p}\right)\left(1 - \frac{N}{N_p}\right) \end{aligned} \tag{4.54}$$

For $N_p \gg N$, $\mathrm{Var}_s\big[\Theta_j[l;s,N]\big] \simeq 1/N$. When we compare the values of the mean and variance [eqns. (4.52) and (4.54)] of the partial auto-correlation of an ML sequence with those of a Bernoulli random signature sequence [eqns. (4.42) and (4.43)], we notice that the approximations are quite good as long as $N_p \gg N$.

## 4.5 Binary Orthogonal Codes

Two *antipodal* binary sequences of length N are orthogonal if their inner-product is zero:

$$\mathbf{c}_i \cdot \mathbf{c}_j = \sum_{n=0}^{N-1} c_i[n]\, c_j[n] = 0$$

The orthogonality condition stipulates that N is even. In contrast, the period P of an ML pseudo-random binary sequence is odd. Strictly speaking, an orthogonal matrix $\mathbf{Q}$ of order $n$ is composed of orthonormal column vectors; i.e.,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 \mathbf{q}_2 \ldots \mathbf{q}_n \end{bmatrix}$$

where the column vectors are orthonormal:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

In what follows, we will relax this definition and allow consider a matrix as orthogonal if *all* its column vectors have equal norms, $\mathbf{q}_i^T \mathbf{q}_i = n$. Hence, an orthogonal matrix $\mathbf{Q}$ of order $n$ can be defined as:

$$\mathbf{Q}\mathbf{Q}^T = n\mathbf{I}_n$$

Since $\mathbf{Q}^{-1} = \mathbf{Q}^T$ implies $\mathbf{Q}^T\mathbf{Q} = n\mathbf{I}_n$, the distinct row vectors are also mutually orthogonal. A particular orthogonal matrix of interest to us is the *Hadamard* matrix $\mathcal{H}$, where the entries $(h_{ij}) \in \mathbb{F}_2$. By generating a Hadamard matrix of order $n$, we can designate its row vectors as orthogonal binary codes of length $n$. We next describe such a technique where binary orthogonal codes of length $2n$ are derived from those of length $n$.

### 4.5.1 Sylvester-type Orthogonal Codes

It is well-known that for $N = 2^k$, $k \in \mathbb{Z}_+$, a family of antipodal binary orthogonal sequences exist. They are called *Walsh-Hadamard* sequences since each binary orthogonal sequence of length N is a row vector of a $(N \times N)$ Hadamard matrix $\mathcal{H}_N$, where its entries $(h_{ij}) \in \mathbb{F}_2$. Starting with $\mathcal{H}_1 = [1]$ and

$$\mathcal{H}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.55}$$

Longer Hadamard sequences can be generated using the following recursive procedure:

$$\mathcal{H}_{2N} = \begin{bmatrix} \mathcal{H}_N & \mathcal{H}_N \\ \mathcal{H}_N & \overline{\mathcal{H}}_N \end{bmatrix} \qquad \overline{\mathcal{H}}_N = 1 \oplus \mathcal{H}_N \tag{4.56}$$

For $(h_{ij}) \in \mathbb{F}_2$, we denote the orthogonal matrix by $\mathcal{H}$. An equivalent *Walsh* orthogonal matrix $\mathcal{W}$ can be generated as follows:

$$\mathcal{W}_{2N} = \begin{bmatrix} \mathcal{W}_N & \mathcal{W}_N \\ \mathcal{W}_N & \overline{\mathcal{W}}_N \end{bmatrix} \qquad \overline{\mathcal{W}}_N = (-1)\mathcal{W}_N \tag{4.57}$$

where its entries $(w_{ij}) \in \mathbb{D}_2$ with

$$\mathcal{W}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{4.58}$$

Orthogonal sequences generated from Hadamard-Walsh matrices of (4.56) and (4.57) are called *Sylvester type*. Other types of orthogonal sequences such as Paley, Muller and Rademacher are described in the appendix at the end of this chapter. We will focus only on the Sylvester type since it has an equivalent binary tree representation, which is further discussed in Sec. 4.5.3. If we denote a vector of length $2^n$ ($n \in \mathbb{N}$) by $\mathbf{x}$, we can generate a pair of mutually orthogonal (sibling) vectors —$[\mathbf{x}\,\mathbf{x}]$ and $[\mathbf{x}\,\bar{\mathbf{x}}]$ — of length $2^{(n+1)}$ by concatenation and inversion. Such vectors are called *recursive*. Equivalently, any group of orthogonal codes that has a code tree description are called *recursive*. The Sylvester-type Hadamard-Walsh codes are conveniently recursive. Among the different types of Walsh codes listed in Appendix 4A, the Paley-type is the only example of non-recursiveness. It is in fact cyclic; i.e., a single cyclic shift of any Walsh code generates another valid orthogonal code. Next, we highlight several important properties of Walsh sequences.

#### A Word on Notation of Walsh codes

For a Sylvester-type $(N \times N)$ Walsh matrix $\mathcal{W}_N$, its orthogonal rows (code vectors) are denoted by $w_N[k]$ for the $k^{th}$-row vector of length N. If we wish to specify a particular position

in time of this vector, an additional index $m$ is used

$$w_N[m;k] = w_{km} \qquad w_{km} \in \mathcal{W}_N$$

where $w_{km}$ is the $(km)^{th}$ entry; i.e., it is the entry in the $k^{th}$ row, $m^{th}$ column of the matrix $\mathcal{W}_N$. If the length $N$ is known and its omission causes no ambiguity, we express the $k^{th}$ Walsh code in vector form:

$$\begin{aligned}\mathbf{w}_k[m] &= w_N[m;k] \\ &= \left(w_{k1}, w_{k2}, \ldots, w_{kN}\right)\end{aligned}$$

The above vector notation is consistently used whenever we are dealing with multi-channel or multi-user communication models. For such purposes, $\mathbf{w}_k$ denotes the Walsh code assigned to user $k$. It is implicitly assume that the orthogonal vectors are periodic; i.e,

$$w_k[m] = w_k[m + jN] \qquad \text{for any integer } j$$

When a Walsh code is modelled as a (continuous-time) waveform where each component $w_{km}$ is a rectangular pulse of duration $T_c$, it is denoted by $w_N(t;k)$.

### 4.5.2 Properties of Walsh sequences

(1) A Sylvester-type $(N \times N)$ Walsh matrix $\mathcal{W}_N$ is an orthogonal matrix:

$$\mathcal{W}_N \mathcal{W}_N^T = N\, I_N$$

Since $\mathcal{W}_N = \mathcal{W}_N^T$, it is also a *reflection* matrix (except for a scaling factor):

$$\mathcal{W}_N = N\, \mathcal{W}_N^{-1} = \mathcal{W}_N^T$$

It is then easy to show that, starting from $\mathcal{W}_2$ and by iteration, $\mathcal{W}_n \mathcal{W}_n^T = n\mathcal{W}_n \mathcal{W}_n^{-1} = nI_n$, proving that $\mathcal{W}_n$ is an orthogonal matrix. Note that there exists other types of $(N \times N)$ orthogonal matrices $\mathbf{Q}$, such as *rotation* and *permutation* matrices, that are not reflective, i.e. $\mathbf{Q} \neq N\, \mathbf{Q}^{-1}$ [135]. Due to its symmetry, both the columns and rows of the Walsh matrix can be used as orthogonal sequences. By definition an orthogonal matrix is composed of orthogonal column vectors. Thus changing their order —i.e., any column permutation— still preserves the mutual orthogonality property. For a Walsh matrix, this implies that a different set of orthogonal Walsh sequences can be generated by column permutation. These sequences are still "Walsh," but no longer recursive. Another important property of a reflection matrix is that any scaling of a column (or row) by a constant, or the addition of another column (or row) preserves the orthogonality condition. The

first condition implies that a Walsh code modulated by a data symbol (including negative integers) remains orthogonal to other Walsh codes of the same length. The second condition guarantees that the modulated sum of several Walsh codes (as in the case of multi-code transmission detailed in Chapter 6) is orthogonal to other Walsh codes of the same length.

(2) Using the Walsh matrix $\mathcal{W}_8$ as an example, we see that the first row of a Walsh matrix denoted by $w_N[1]$ is a vector of all 1's. All other Walsh sequences, $w_N[k]$, $k \in \mathbb{Z}_2^N$, are *balanced*, i.e., they all have equal number of 1's and $-1$'s.

$$\mathcal{W}_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \tag{4.59}$$

(3) A pair of higher-dimensional (longer length) Walsh codes can be expressed as a concatenation of two lower-dimensional Walsh codes:

$$\begin{aligned} w_{2N}[2l-1] &= \Big(w_N[l], +w_N[l]\Big) \\ w_{2N}[2l] &= \Big(w_N[l], -w_N[l]\Big) \end{aligned} \tag{4.60}$$

(4) An inverted Walsh sequence is orthogonal to other Walsh sequences (inverted or not) of the same length.i.e, $-w_N[k] \perp w_N[l]$ for $\forall\ l \neq k$. In contrast two Hadamard sequences (say, $H_N[k]$ and $H_N[l]$) of the same length are orthogonal, but an inverted version $\overline{H_N}[k]$ is not orthogonal to other Hadamard sequences of the same length. For example, $(0,0) \perp (0,1)$, but $(1,1)$ is not orthogonal to either $(0,1)$ or $(1,0)$.

(5) Since a Walsh sequences is a level-shifted version of a Hadamard sequence, it is commonly treated as a (continuous-time) waveform than as a discrete-time sequence. A plot of Walsh sequences as antipodal waveforms is shown in Fig. 4.13 where each sequence is considered as a train of rectangular chip pulses, with each "chip" pulse of duration $T_p$. The duration of a Walsh waveform of code length N is $N\,T_p$. It is immediate that the periodic cross-correlation is zero over this window:

$$w_N[k] \perp w_N[l] \longleftrightarrow \int_{t=0}^{NT_p} w_N(t;k) w_N(t;l)\, dt = 0$$

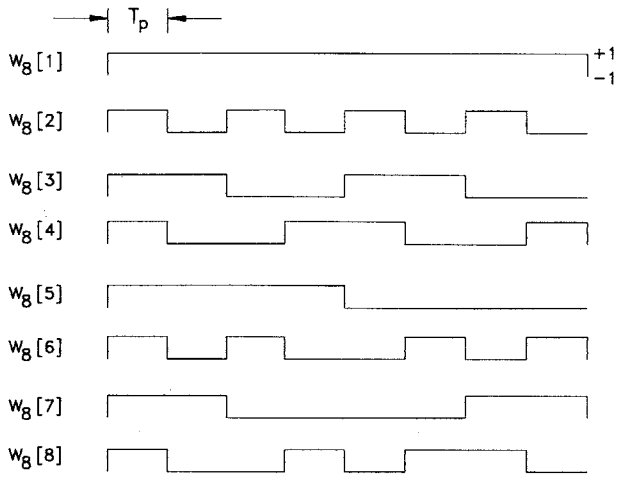



**Figure** 4.13: Walsh sequences as a train of rectangular chip pulses of length 8

Note that due to its recursive property, length $N = 8$ Walsh waveforms with odd indices are identical to Walsh waveforms of length $N/2 = 4$. Therefore it is only meaningful to include both the code length N and the chip pulse duration $T_c$ (or equivalently, the chip rate $R_c$) when declaring a set of orthogonal Walsh codes.

(6) For the same chip rate, two set of Walsh waveforms of lengths 4 and 8 are depicted in Fig. 4.14. Note that length-8 Walsh codes with odd indices are identical to length-4 Walsh codes. In particular, over the time window $[0, NT_p]$:

$$w_N[l] = w_{2N}[l] = w_{2N}[l + N] \tag{4.61}$$

A similar result can be generated over the time window $[NT_p/2, NT_p]$. In general, two Walsh codes of different lengths, $w_N[k]$ and $w_M[l]$ where $N \neq M$, and the same chip rate are orthogonal if they are not *relatives*. We give a precise definition of the term *relative* in the next section by first discussing the relationship among Walsh codes of various lengths and a binary code tree structure.

### 4.5.3 Orthogonal Binary Code Tree

An equivalent description and generation of Sylvester-type Walsh sequences is based on the structure of a *complete* binary code tree.

7 **Definition** *A tree is complete if it has no broken branches. A tree is binary if there is a maximum of two branches from each node. A branch is broken if a node does not contain two branches. In a code tree, a branch is also a sub-tree, and each node is a code.*

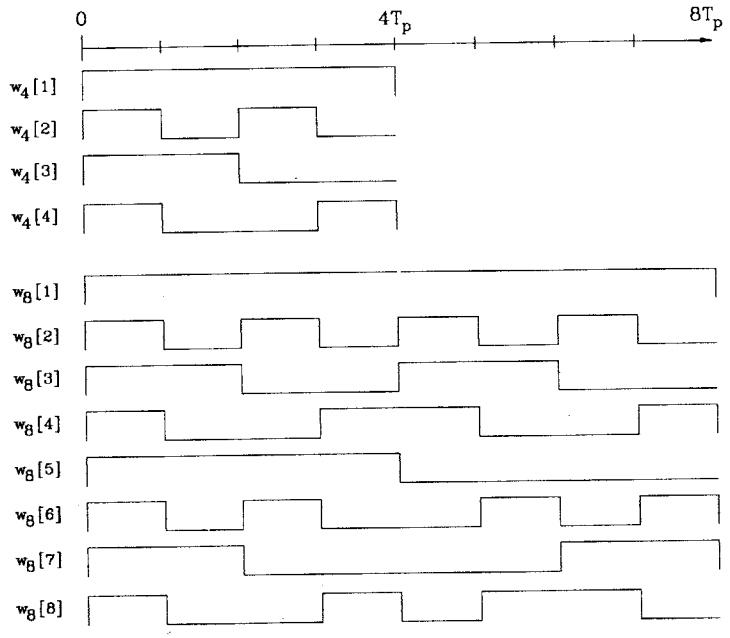
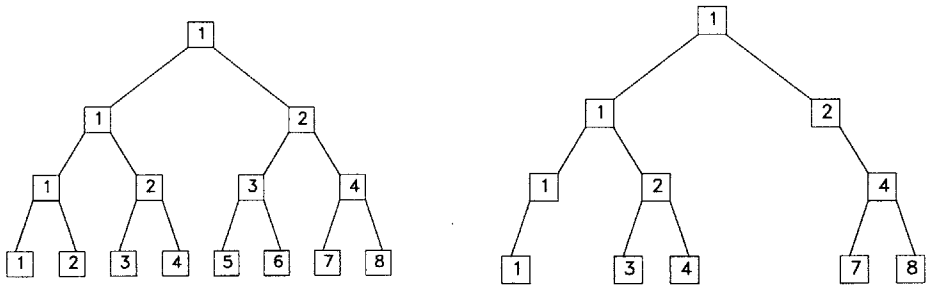



**Figure** 4.14: Walsh sequences as a train of rectangular chip pulses of lengths 4 and 8

An n example of a complete tree is shown in Fig. 4.15(a), whereas the tree in Fig. 4.15(b) is incomplete since it contains broken branches. Some authors also define a complete tree as a *balanced* tree since it is in equilibrium about its pivot—the central point of a root code. In order to identify the codes in the tree without ambiguity, each code is assigned a unique *layer number* and a *branch number*. Each code in a tree is identified using the following notation:

$$c_{ij} \quad \text{or} \quad (i,j) \qquad i = \text{layer number}, \quad j = \text{branch number}$$

The code layers are numbered sequentially from bottom to top, starting from 1. Thus, a higher-layer code has a shorter length (lower dimension) than a lower-layer code of higher dimension. Similarly, the branches in each code layer are numbered sequentially from left to right, starting from 1. The topmost code of a branch or tree is called its *root* code. For example, in Fig. 4.16, (5,1) is the root code of the tree while (3,1) is the root code of a branch (or a sub-tree) containing codes (2,1), (2,2), (1,1), (1,2), (1,3) and (1,4). All lower layer codes spanned from a higher layer code are defined as *descendant* codes. All high layer codes linking a particular code to the root code are called its *ancestor* codes. The immediate ancestor code is called the *mother* code. Two *sibling* codes are those generated from their mother code. They are also known as the *children* codes of the mother code. The lowest-layer codes are called *leaf* codes, or simply leaves. As shown in Fig. 4.16, the descendant codes of (2,4) are the siblings (1,7) and (1,8). The ancestor codes of (1,5) are (2,3), (3,2) and (4,1).



(a) Complete (b) Incomplete

**Figure 4.15:** Complete and incomplete binary code trees. In the incomplete tree, the branches from codes $c_{21}$ and $c_{32}$ are broken.

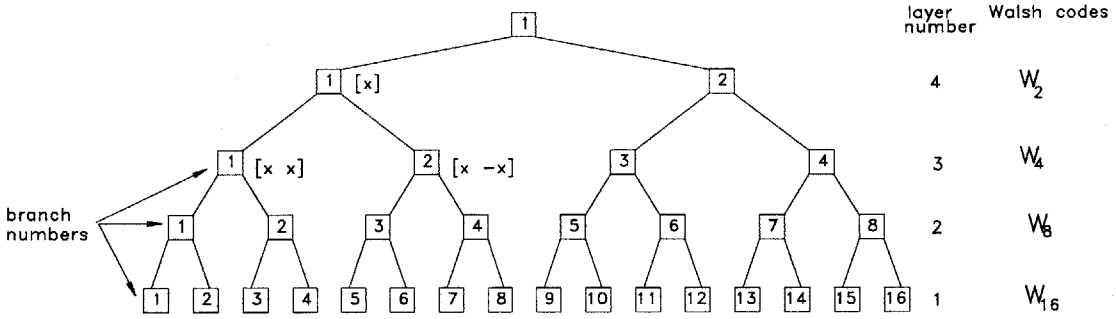



**Figure 4.16:** Complete binary code tree with 16 leaves. It is a graphical representation of the generation of length-16 Walsh codes.

### Prefix-free Condition

Each node in the complete binary tree is a Walsh code. All codes in the same layer have the same length. As a consequence of the property of a Walsh matrix, codes from the same layer are mutually orthogonal. The length of a mother code is half the length of its children. Hence, higher-dimensional children Walsh codes are generated from their mother code by concatenation. As illustrated in the top left half of the code tree in Fig. 4.16, two codes of layer 3 are generated recursively from their mother code of layer 4. Furthermore, any two codes of different layers are also orthogonal except when one is not an ancestor (or equivalently, a descendant code) of the other. This assertion can be verified in several ways. The first alternative is based on a Hadamard binary tree structure shown in Fig. 4.17 where each node is a *Hadamard* code, i.e., its elements $\in \mathbb{F}_2$. As before, longer-length children codes are recursively generated by concatenation of a lower-length mother code. Two *binary* Hadamard codes of different lengths are orthogonal if they are *prefix-free* binary codes. Prefix-free codes are treated in most information theory texts. (See Cover [22], pp. 81 for further explanation.) Two Hadamard (or

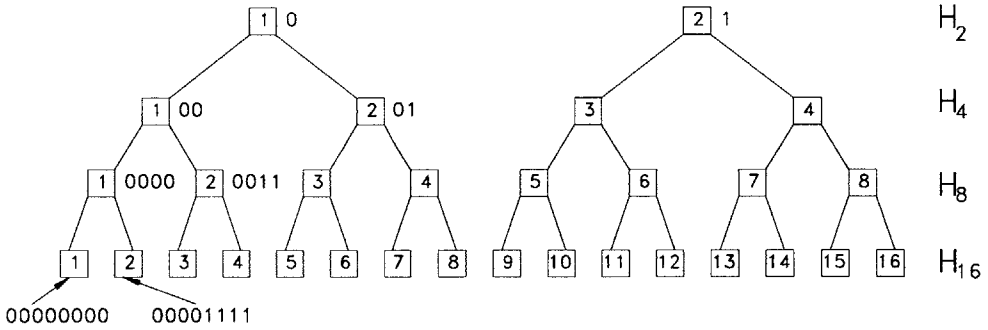



**Figure 4.17:** Complete binary tree with Hadamard codes. Two Hadamard codes are orthogonal if they are prefix-free binary codes.

Walsh) codes—of different lengths—are considered *relatives* if they have identical prefix. As far as we are concerned, two Hadamard codes of *different* lengths are relatives (non-orthogonal) if the prefix of the longer code is the shorter code itself. Thus, all codes linking a leaf and the root code inclusively are relatives. In Fig. 4.17, the codes (00) and (0011) are relatives (since prefix '00' is the same as the shorter code) while (01) is a non-relative of (0011).

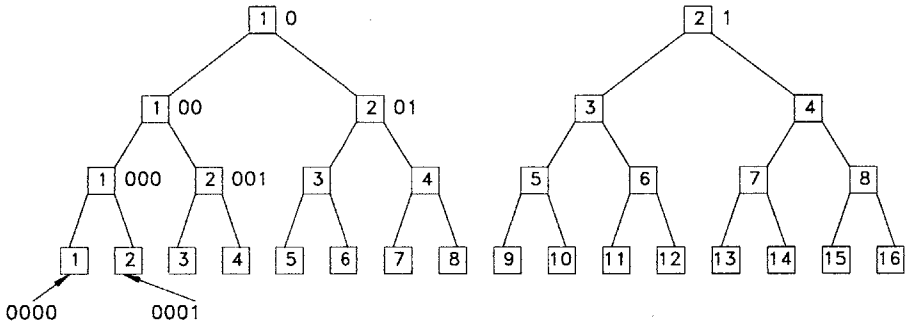



**Figure 4.18:** Description of binary prefix free codes in tree structure. Unlike binary Hadamard codes, a '1' or '0' is appended for lower layer codes.

The binary tree described in Fig. 4.17 is slightly different the tree structure of binary prefix free codes. For clarity we also illustrate in Fig. 4.18 the same tree structure, but with "conventional" labelling of binary prefix free codes, in which a '1' or '0' is appended to a lower layer descendant code —rather than concatenation as in Hadamard-Walsh codes. Regardless of slightly different methods of code generation, the mutual orthogonality property of binary prefix free codes still holds. At this point it seems the connection between Walsh-Hadamard and binary prefix free codes is trivial and redundant. In Chapter 7 we use this equivalence to prove an important theorem regarding the capacity of OVSF-CDM system.

### Walsh Codes as Continuous-time Waveforms

According to eqn. (4.61) and Fig. 4.14, the mother code is identical to its children in its defined time window. As evident from the figure, two relative waveforms—e.g. $w_4(t;1)$ and $w_8(t;1)$, or $w_4(t;1)$ and $w_8(t;5)$—are identical. Two prefix-free waveforms—e.g. $w_4(t;1)$ and $w_8(t;2)$—orthogonal:

$$\begin{aligned}
&\int_{t=0}^{4T_p} w_4(t;1)\, w_8(t;2)\, dt = 0 \\
&\int_{t=0}^{8T_p} w_4(t;1)\, w_8(t;2)\, dt = 0 \\
&\int_{t=0}^{8T_p} w'_4(t;1)\, w'_8(t;2)\, dt = 0
\end{aligned} \tag{4.62}$$

where $w'_N(t;k)$ is an infinite periodic train of Walsh waveforms:

$$w'_N(t;k) = \Big( \ldots, \pm w_N(t;k), \pm w_N(t;k), \ldots \Big)$$

The first two cross-correlation functions are aperiodic over duration of $4T_p$ and $8T_p$, respectively. The last is the periodic cross-correlation. As we shall see in Chapters 6 and 7, binary orthogonal codes of various lengths with the same chip duration are called *Orthogonal Variable-Spreading Factor* (OVSF) codes. The spreading factor is equal to the code length N.

## 4.6 Superimposed Orthogonal Binary Codes

It is obvious from Fig 4.14 that the periodic cross-correlation of any two Walsh codes is large and it is highly dependent on the time offset as well as the particular choice of Walsh codes. This is due to their highly regular structure from concatenation. For a special type of cyclic Paley-type Walsh codes (see Appendix 4A) derived from quadrature residuals, any cyclic offset of a particular Walsh code becomes another sibling Walsh code. Therefore, the application of orthogonal Walsh codes as user signature sequences is limited to broadcast transmission mode where code synchronism is possible. If, however, the channel generates multiple replicas of the transmit signal, then the received signal will suffer from a significant amount of self-noise that is induced by high cross-correlation values of Walsh codes.

As an alternative to orthogonal binary codes, we can assign pseudo-random (ML-type) codes as signature sequences. In Chapter 2, it is shown that a CDM system with non-orthogonal codes and naive decoding results in a rate region that is equivalent to naive CDM (N-CDM) —which is much smaller than various orthogonal multiplexing schemes such as O-CDM, FDM, TDM and N-TDM. Therefore, CDM with PN codes and single-user decoding would not even come close to being the most efficient resource allocation policy. Several researchers, including those at Qualcomm, have proposed a novel method where binary periodic orthogonal codes of period N

are superimposed with a PN sequence of length P, where $P \gg N$. In the literature, these codes are commonly described as "short" and "long" codes, respectively. The superimposed code can be expressed analytically as follows:

$$A_k[n] = \mathcal{H}_k[m] \oplus P[n] \tag{4.63}$$

where $\mathcal{H}_k[m]$ is the $k^{th}$ row of a Hadamard orthogonal matrix and $P[n]$ is a pseudo-random sequence. It is understood that all three sequences are periodic. The different indices $n$ and $m$ emphasize the fact that the clock rates of Hadamard and PN codes may be different. Let's denote the Walsh-Hadamard code rate by $R_w = mR$. The chip rate of the PN sequence is $R_c = nR$. We already know that the excess spreading factor $k_c$ is defined as $k_c = R_c/R_w$. Eqn. (4.63) can also be expressed in terms of their level-shifted binary sequences,

$$a_k[n] = \mathbf{w}_k[m] \cdot p[n] \qquad \text{where} \quad m = \left\lfloor \left\lfloor \frac{n}{k_c} \right\rfloor \right\rfloor$$

where $\mathbf{w}_k[m]$ is the $k^{th}$ row of a Walsh matrix and $p[n]$ is the antipodal equivalent of $P[n]$. To simplify the above expression, we can relabel the Walsh codes such that their new clock rate $\hat{R}_w$ equals the chip rate $R_c$. For $k_c = 2^m$ ($m \in \mathbb{N}$) we can interpret the new set of Walsh codes as a subset of Walsh codes of higher dimension. As an example, consider the set of Walsh codes of length 4: $\{w_4[j]\}$ and $k_c = 2$. By clocking each Walsh chip at twice the original rate, the resulting Walsh codes $\{\hat{w}_4[j]\}$ belong to the set of length 8. In particular,

$$\hat{w}_4[m; j] = w_8[2m - 1; j] \qquad 1 \le j \le 4$$

For convenience, we express the superimposed orthogonal sequence of user $k$ as

$$a_k[n] = \hat{\mathbf{w}}_k[n] \cdot p[n]$$

implying that both the PN sequence $p[n]$ and the embedded orthogonal code $\hat{\mathbf{w}}_k[n]$ are clocked at the same chip rate $R_c$.

### 4.6.1 Correlation Properties

We already know that for zero time offset, two superimposed codes are orthogonal over a Walsh code length:

$$\sum_{r=jN}^{(j+1)N-1} a_k[r]\, a_l[r] = N\, \delta[k - l]$$

When studying superimposed orthogonal waveforms, we are mostly interested in a case where the time offset is non-zero. This situation occurs when the channel generates multiple replicas of the transmit waveform at the receiver, or when another superimposed orthogonal waveform

from a neighboring cell interferes with the desired signal. In the former, we are dealing with the auto-correlation of the superimposed waveform for arbitrary non-zero time offsets. In the latter, it is the cross-correlation between two superimposed waveforms with different long PN codes. If the long code is a purely random Bernoulli sequence, then we are only interested in the statistical properties (i.e., the mean and variance) of auto- and cross-correlation functions. Fortunately, due to the i.i.d. property of a bipolar Bernoulli sequence, the mean and variance of the partial correlation of the resulting superimposed sequence is identical to those of a random sequence as listed in eqns. (4.40) and (4.41). In the previous sections it was shown that the ML binary sequence has correlation properties that approximates closely those of a random Bernoulli sequence when $N_p \gg N$. We must now determine if the superimposed orthogonal sequences possess correlation properties that approximate those of a random Bernoulli sequence. The mean and variance of the partial auto-correlation of a superimposed sequence can be determined as follows:

$$\begin{aligned} E\Big[\theta_i[l;s,N]\Big] &= \sum_{k=s}^{s+N-1} E\Big[a_i[k]\cdot a_i[k-l]\Big] \\ &= \sum_{k=s}^{s+N-1} E\Big[\hat{w}_i[k]p[k]\cdot \hat{w}_i[k-l]p[k-l]\Big] \end{aligned}$$

We ignore the trivial case of $E\Big[\theta_i[0;s,N]\Big] = N$. In the following we assume $l \neq 0$. Since the periods $N$ and $N_p$ are relatively prime, the expectation is computed over all possible relative time-offsets between the starting epochs of the PN code and the Walsh code. Hence,

$$E\Big[\theta_i[l;s,N]\Big] = \sum_{k=s}^{s+N-1} \frac{1}{N_p} \sum_{n=0}^{N_p-1} \hat{w}_i[k]p[k-n]\cdot \hat{w}_i[k-l]p[k-l-n]$$

Using the shift-and-add property of ML sequences,

$$p[k-n]p[k-l-n] = p[q-n]$$

for some integer $q \neq k$ and $q \neq (k-l)$, the expectation simplifies to

$$E\Big[\theta_i[l;s,N]\Big] = \sum_{k=s}^{s+N-1} \frac{1}{N_p}\left\{\sum_{n=0}^{N_p-1} p[q-n]\right\} \hat{w}_i[k]\hat{w}_i[k-l]$$

Due to the balanced property of ML sequences,

$$\frac{1}{N_p}\left\{\sum_{n=0}^{N_p-1} p[q-n]\right\} = -\frac{1}{N_p}$$

for any integer m. Substituting this result and simplifying

$$\mathrm{E}\Big[\theta_i[l;s,N]\Big] = -\frac{1}{N_p}\sum_{k=s}^{s+N-1}\hat{w}_i[k]\hat{w}_i[k-l]$$
$$= -\frac{1}{N_p}C_i[l]$$

Since the Walsh codes are periodic with period N, the starting time epoch s is irrelevant. We know that the periodic auto-correlation $C_i[l]$ is very sensitive to the values of time-offset l and code index i. However, we also know that its maximum absolute value is N, which leads to the following bound

$$\left|\mathrm{E}\Big[\theta_i[l;s,N]\Big]\right| \le \frac{N}{N_p} \tag{4.64}$$

We conclude that for $N_p \gg N$, the superimposed orthogonal sequence approximates the Bernoulli sequence in terms of the mean value of its partial auto-correlation. We next compute the variance by explicitly writing the mean-squared value:

$$\mathrm{E}\Big[\theta_i[l;s,N]\Big]^2 = \mathrm{E}\left[\sum_{k=s}^{s+N-1}\sum_{m=s}^{s+N-1} a_i[k]\,a_i[m]\,a_i[k-l]\,a_i[m-l]\right]$$

$$= \frac{1}{N_p}\left[\sum_{k=s}^{s+N-1}\sum_{m=s}^{s+N-1}\Big(\hat{w}_i[k]\,\hat{w}_i[m]\,\hat{w}_i[k-l]\,\hat{w}_i[m-l]\Big)\right.$$
$$\left.\sum_{n=0}^{N_p-1}\Big(p_i[k-n]\,p_i[m-n]\,p_i[k-l-n]\,p_i[m-l-n]\Big)\right]$$

Splitting the double summation into diagonal and off-diagonal terms,

$$= \frac{1}{N_p}\left[\sum_{k=s}^{s+N-1}\Big(\hat{w}_i^2[k]\,\hat{w}_i^2[k-l]\sum_{n=0}^{N_p-1}\Big(p_i^2[k-n]\,p_i^2[k-l-n]\Big)\right.$$
$$+\sum_{k=s}^{s+N-2}\sum_{m=k+1}^{s+N-1}\hat{w}_i[k]\,\hat{w}_i[m]\,\hat{w}_i[k-l]\,\hat{w}_i[m-l]$$
$$\left.\sum_{n=0}^{N_p-1}\Big(p_i[k-n]\,p_i[m-n]\,p_i[k-l-n]\,p_i[m-l-n]\Big)\right]$$

By applying the shift-and-add properties of ML sequences on the second term twice,

$$= N + \frac{1}{N_p}\left[\sum_{k=s}^{s+N-1}\sum_{m=s}^{s+N-1}\Big(\hat{w}_i[k]\,\hat{w}_i[m]\,\hat{w}_i[k-l]\,\hat{w}_i[m-l]\Big)\cdot\sum_{n=0}^{N_p-1}\Big(p_i[q-n]\,p_i[r-n]\Big)\right]$$
$$= N + \frac{1}{N_p}\sum_{k=s}^{s+N-1}\sum_{m=s}^{s+N-1}\hat{w}_i[k]\,\hat{w}_i[m]\,\hat{w}_i[k-l]\,\hat{w}_i[m-l]\sum_{n=0}^{N_p-1}p_i[s-n]$$

Using the balanced property, it simplifies to

$$= N - \frac{1}{N_p} \sum_{k=s}^{s+N-1} \sum_{m=s}^{s+N-1} \hat{w}_i[k]\, \hat{w}_i[m]\, \hat{w}_i[k-l]\, \hat{w}_i[m-l]$$

Splitting the double summation into diagonal and off-diagonal terms,

$$= N - \frac{1}{N_p} \left[ \sum_{k=s}^{s+N-1} \hat{w}_i^2[k]\, \hat{w}_i^2[k-l] + 2 \sum_{k=s}^{s+N-2} \sum_{m=k+1}^{s+N-1} \hat{w}_i[k]\, \hat{w}_i[m]\, \hat{w}_i[k-l]\, \hat{w}_i[m-l] \right]$$

The second summation can be bounded by its absolute maximum (the number of off-diagonal terms divided by 2):

$$\left| \sum_{k=s}^{s+N-2} \sum_{m=k+1}^{s+N-1} \hat{w}_i[k]\, \hat{w}_i[m]\, \hat{w}_i[k-l]\, \hat{w}_i[m-l] \right| \leq \frac{N(N-1)}{2}$$

which leads to

$$\begin{aligned} E\left[\theta_i[l;s,N]\right]^2 &\leq N - \frac{1}{N_p}\left[N \pm N(N-1)\right] \\ \mathrm{Var}\left[\theta_i[l;s,N]\right] &\lesssim N - \frac{1}{N_p}\left[N \pm N(N-1)\right] - \left(\frac{N}{N_p}\right)^2 \\ &\lesssim N\left[1 \pm \frac{(N-1)}{N_p}\right] \end{aligned} \tag{4.65}$$

For $N_p \gg N$, $\mathrm{Var}\left[\theta_i[l;s,N]\right] \lesssim N$. Note that above results hold for both recursive and non-recursive Walsh codes. Since the auto-correlation of Walsh codes is periodic, permutation of a pair of Walsh codes has no effect on its value.

## 4.7 Comments on Cross-Correlation of Signature Sequences

Compared to the wealth of information that is available on periodic discrete sequences, our treatment here is rather brief. We covered only the tip of the ice, limiting our scope to topics that are directly relevant to code-division multiplexing of spread-spectrum modulated signals. In all our analyses, we consider only binary sequences. There exists, however, non-binary sequences —also known as polyphase sequences— from larger Galois fields that possess unique properties that are not attainable in the binary space. Regarding PN sequences, we have dealt only with maximum-length sequences, mostly due to their unique balance and shift-and-add properties. Without the application of these two properties, most of our derivations cannot be simplified or approximated into closed-form expressions. For both ML sequences and ML-based superimposed sequences, we did not compute their cross-correlation values. It has been shown by Golomb [46] that the periodic cross-correlation of an ML binary sequence takes on at least

three different values. He further shows that for a given length $N_p$, such ML sequences with three-valued cross-correlation property is rather small ($\ll N_p$) . This may not be that critical of an issue for the downlink where the number of neighboring sites with significant interference level is not large. However, this is a serious matter in the uplink where the number of simultaneously active mobiles is quite large. For this reason the third-generation CDMA standard has adopted non-ML binary sequences such as Gold and Kasami codes as signature sequences. They still possess cross-correlation values in the same order as ML sequences; however, their auto-correlation is not two-valued. There exists considerable sidelobes. Regardless of the type of binary sequence, less is known about the properties of aperiodic and partial cross-correlation values. In terms of periodic auto-correlation, we already know that it is two valued; in fact, in the literature this type of periodic auto-correlation pattern is known as "impulsive" since its value is maximum when phase offsets are multiples of period $N_p$. For all other offsets, its value is $-1$.

For the non-binary case, there exists *perfect codes* whose periodic auto-correlation is zero for all non-multiple phase offsets of $N_p$. However, for both binary and non-binary cases, it is not possible to construct sequences that possess both ideal periodic auto- and cross-correlation values. The best we can achieve is non-binary sequences with impulsive auto-correlation and cross-correlation values no less than $\sqrt{N_p}$ — the Welch bound. As we discover in our analysis, it is rather the partial correlation properties of signature sequences that are of most value. Very little has been reported in this area mainly because it is difficult to construct combinatorial and algebraic frameworks for partial correlation. It is more common to measure and evaluate statistical or time-averaged properties partial correlation.

### 4.7.1 Correlation Zones

Recently, due to the popularity of synchronous spread-spectrum communication, several researchers have proposed binary and non-binary sequences with unique auto- and cross-correlation properties known as *the correlation zone*. Such codes, as elaborated in References [32], [33], [138], [139] have ideal zero auto- or cross-correlation values (i.e., periodic, aperiodic or partial) for a certain range of phase offsets called zones (See Fig. 4.19). For other phase offsets the correlation values may be large. These researchers are well aware of the fact that in a multipath channel, correlation values are significant only in a time window within the delay spread. If such codes become practical, it is likely that they will have a major impact on diversity Rake combining of multipath spread-spectrum signals. As we shall see in Chapter 6, Sec. 6.5, the non-ideal correlation values of signature sequences result in self-noise at the output of a Rake receiver. This self-noise either reduces the diversity gain or dramatically increases receiver complexity.

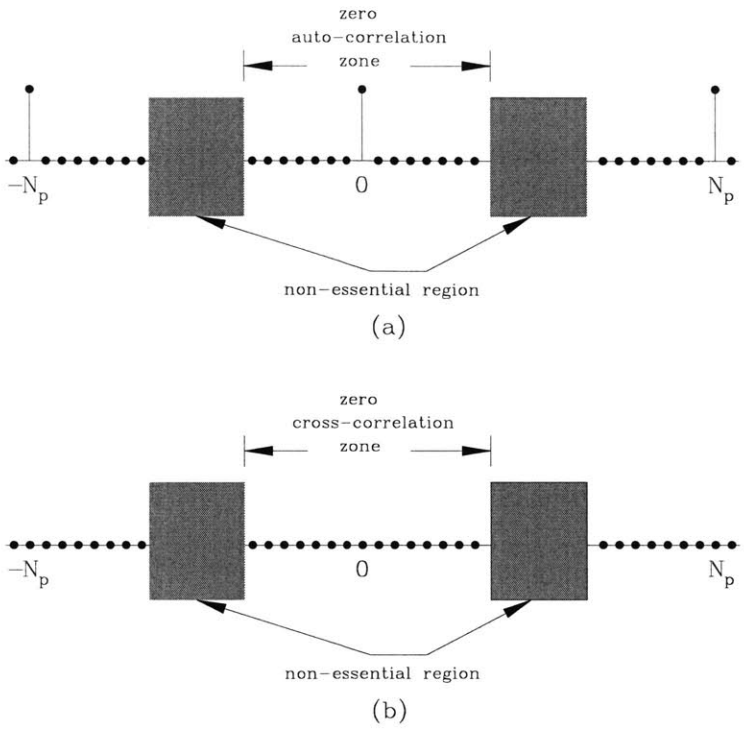



**Figure** 4.19: Zero correlation zones for periodic (a)auto-correlation and (b) cross-correlation. Correlation values outside the desired time window called "the zone" are non-essential.

## 4.8 Correlation Properties of Other Important Signals

Thus far our focus has been signature waveforms whose continuous-time cross-correlation is measured in terms of the discrete-time cross-correlation of embedded discrete binary sequences. For all purposes considered, they are spread-spectrum modulated. We are also interested in conditions where two signature waveforms, without spread-spectrum modulation and embedded discrete-time sequences, can still be orthogonal. As stated in (4.3), the cross-correlation function between two synchronous ($\tau = 0$), symbol-time limited signature waveforms is defined as

$$\rho_{ij} = \int_0^T s_i(t)s_j(t)dt$$

If the waveforms also carry information via PAM-type data modulation, then the data modulated signals are orthogonal as long as $\rho_{ij} = 0$ since

$$\int_0^T \{x_i s_i(t)\}\{x_j s_j(t)\}\, dt = x_i x_j \int_0^T s_i(t)s_j(t)dt = x_i x_j\, \rho_{ij}$$

where $x_k$ is PAM symbol for waveform $s_k(t)$. In particular, we are interested in waveform orthogonality when one signal is frequency-shifted, time-shifted or both time- and frequency-shifted version of the other.

### 4.8.1 Frequency-Shifted Signals

We are mainly interested in radio-frequency (RF) carrier (single-tone) waveforms where the RF carrier frequency is much higher than the symbol rate, $f_c \gg 1/T$. We study three unique cases:

#### Case 1: CW Signals with Zero Phase Offset

Denoting a CW signal with only an in-phase (cosine) component:

$$s_c^I(t) = \sqrt{\frac{2\mathcal{E}}{T}}\cos(2\pi f_c t) \qquad c = i, j \qquad 0 \le t \le T \tag{4.66}$$

$\mathcal{E}$ is the signal energy. We can similarly define the quadrature-phase (sine) component:

$$s_c^Q(t) = \sqrt{\frac{2\mathcal{E}}{T}}\sin(2\pi f_c t) \qquad c = i, j \qquad 0 \le t \le T \tag{4.67}$$

Assuming the RF carrier frequency is much higher than the symbol rate, $f_c \gg 1/T$, it is easily shown that:

$$\rho_{ij}^I = \int_0^T s_i^I(t)s_j^I(t)dt = \frac{\sin 2\pi(f_i - f_j)T}{2\pi(f_i - f_j)T} \tag{4.68}$$

$$\rho_{ij}^Q = \int_0^T s_i^Q(t)s_j^Q(t)dt = \frac{\sin 2\pi(f_i - f_j)T}{2\pi(f_i - f_j)T} \tag{4.69}$$

$$\rho_{ij}^{IQ} = \int_0^T s_i^I(t)s_j^Q(t)dt = \frac{\cos 2\pi(f_i - f_j)T - 1}{2\pi(f_i - f_j)T} \tag{4.70}$$

It can be deduced from the above equations that the minimum frequency separation $\Delta f = (f_i - f_j)$ required to maintain orthogonality is:

$$\Delta f = \begin{cases} 1/2T & s_i^I \perp s_j^I \quad \text{or} \quad s_i^Q \perp s_j^Q \\ 1/T & s_i^I \perp s_j^Q \end{cases} \tag{4.71}$$

In general, for two CW signals $s_c(t) = \sqrt{\mathcal{E}/T}\Big(\cos 2\pi f_c t + \sin 2\pi f_c t\Big)$, $c = i, j$ with both in-phase and quadrature components, they are orthogonal if $\Delta f = 1/2$.

**Case 2: CW Signals with Phase Offset θ**

Assuming a non-zero phase-offset θ between two CW signals:

$$\begin{aligned} s_i(t) &= \sqrt{\frac{2\mathcal{E}}{T}}\cos(2\pi f_i t) \\ s_j(t) &= \sqrt{\frac{2\mathcal{E}}{T}}\cos(2\pi f_j t + \theta) \end{aligned} \qquad 0 \le t \le T \tag{4.72}$$

Since $s_j(t)$ with a phase-offset θ can be considered as consisting of both in-phase and quadrature components, it can be readily shown that:

$$\Delta f(\theta) = 1/T \qquad s_i \perp s_j \tag{4.73}$$

In this case, it is immaterial whether $s_i(t)$ and $s_j(t)$ are cosine or sine functions. An important implication of eqns. (4.71) and (4.73) is studied further in Chapter 5.

### 4.8.2 Time-Shifted Signals

If a signal $s(t)$ is symbol-time limited, then it is trivial that $s(t)$ and its T-shifted versions are orthogonal:

$$\int_{\infty}^{\infty} s(t)s(t - kT)\,dt = 0 \qquad k \in \pm\mathbb{Z}_+$$

A more interesting case is when the pulse duration is larger than the symbol interval ($T_p > T$). For a family of $W$-Hz bandlimited sinc pulses $s(t - kT) = \text{sinc}\,2W(t - kT)$,

$$\begin{aligned} &\int_{\infty}^{\infty} s(t)s(t - kT)\,dt \\ &= \int_{\infty}^{\infty} \text{sinc}\,2W(t)\,\text{sinc}\,2W(t - kT)\,dt \\ &= \left(\frac{1}{2W}\right)^2 \int_{-W}^{W} e^{-j\pi kf/W}\,df \\ &= \frac{1}{2W}\ \text{sinc}(k) = 0 \quad \text{for} \quad k \in \pm\mathbb{Z}_+ \end{aligned} \qquad T = \frac{1}{2W} \tag{4.74}$$

### 4.8.3 Time- and Frequency-Shifted Signals

In a more general setting, we are interested in the requirements to maintain orthogonality between two carrier-modulated signals where the time shift is a non-zero multiple of T, the symbol duration:

$$\rho_{ij} = \int_{\infty}^{\infty} g_i(t)\,g_j(t - kT)\,dt \qquad k \in \pm\mathbb{Z}_+ \tag{4.75}$$

where

$$\begin{aligned} g_i(t) &= s(t)\ \cos(2\pi f_i t) \\ g_j(t-kT) &= s(t-kT)\ \cos\Big(2\pi f_j(t-kT)+\theta\Big) \\ &= s(t-kT)\ \cos\Big(2\pi f_j t+\theta_k\Big) \qquad \theta_k = \theta - 2\pi k f_j T \end{aligned}$$

Unless we give an explicit form of $s(t)$, we cannot say more about the cross-correlation property of $g_i(t)$ and $g_j(t)$. As we shall see in Chapter 5, time-frequency shifted signals are the most spectrally efficient pulses, especially when used in multi-carrier modulated systems.

## Notes and References

The correlation properties of linear shift-register sequences were studied by Golomb [46], Zierler [169], Gold [43] and others. Since both binary and non-binary sequences and their combinatorial and algebraic properties are of interest to coding theorists, they are also covered in error-control coding texts by Peterson and Weldon [101], MacWilliams and Sloane [80], Lin and Costello [77]. Texts on spread-spectrum communications by Dixon [30], Holmes [54] and Simon et al. [129] also include chapters dedicated to binary pseudo-random sequences with an emphasis on their application in spread-spectrum signalling. A series of journal articles by M. Pursely and his former students [109, 110, 119, 120] cover topics specific to spread-spectrum multiple-access communication systems. A more abstract treatment of algebraic codes can be found in the book by Lang [67].

The description of orthogonal binary codes using a tree structure follows the work by Adachi et al. [2]. Hadamard-Walsh matrices and their orthogonality properties can be found in any linear algebra textbook. We have used the texts by Strang [133], [134], [135] as references.

# Appendix 4A
# Generation of Binary Orthogonal Codes

## Muller-type Hadamard-Walsh Codes

A first-order Reed-Muller binary $(2^k, k)$ code has the following generator matrix:

$$\mathbf{G} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ 2^k - 1 \end{bmatrix}$$

where the value of each row is represented in decimal form. In order to generate all $2^k$ code vectors, the generator matrix must be expressed in binary form with elements from $\mathbb{F}_2$. As an example, for $k = 3$:

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Each code vector $\mathbf{v}$ of a linear block code is:

$$\mathbf{v} = \mathbf{u} \otimes \mathbf{G}^T$$

where $\mathbf{u}$ is a k– bit input (row) vector, and $\mathbf{v}$ is the resulting $n(= 2^k)$– bit output (row) vector. It is well known that Reed–Muller codes belong to a family of *Euclidean Geometry* (EG) codes. An $k^{th}$– order Reed-Muller code is equivalent to an $(k, 1)^{th}$– order EG code [77]. It is also known that Reed-Muller codes are orthogonalizable. In fact, if we use a non-cyclic generator matrix $\mathbf{G}$ as shown above, the resulting code vectors are mutually orthogonal. Non-cyclic Reed-Muller codes were first discovered by Muller [91]. For that reason, we define all orthogonal code vectors obtained using $\mathbf{G}$ as *Muller-type*. If we denote $\mathbf{U}$ as the matrix of all possible input

vectors:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{2^k} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

The matrix of all possible output code vectors is:

$$\mathbf{V} = \mathbf{U} \otimes \mathbf{G}^T = \mathbf{U} \otimes \mathbf{U}^T = \mathbf{G} \otimes \mathbf{G}^T$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

It is easily seen that the output code vector matrix $\mathbf{V}$ *is* the Hadamard matrix $\mathcal{H}$—$\mathcal{H}_8$ in the above case. Thus, the output matrix of a $(2^k, k)$ Reed-Muller code is the Hadamard matrix $\mathcal{H}_{2^k}$.

## Rademacher-type Hadamard-Walsh Codes

The following set of functions $\{r_i(t)\}$ on the interval $[0, 1)$ are known as Rademacher functions:

$$r_i(t) = (-1)^m \quad \text{if} \quad \frac{m}{2^i} \le t < \frac{m+1}{2^i} \qquad m \in \mathbb{N}$$

As examples, we list the first three of such functions:

$$r_1(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \end{cases}$$

$$r_2(t) = \begin{cases} 1 & 0 \le t < \frac{1}{4} \\ -1 & \frac{1}{4} \le t < \frac{1}{2} \\ 1 & \frac{1}{2} \le t < \frac{3}{4} \\ -1 & \frac{3}{4} \le t < 1 \end{cases}$$

$$r_3(t) = \begin{cases} 1 & 0 \le t < \frac{1}{8} \\ -1 & \frac{1}{8} \le t < \frac{1}{4} \\ 1 & \frac{1}{4} \le t < \frac{3}{8} \\ -1 & \frac{3}{8} \le t < \frac{1}{2} \\ 1 & \frac{1}{2} \le t < \frac{5}{8} \\ -1 & \frac{5}{8} \le t < \frac{3}{4} \\ 1 & \frac{3}{4} \le t < \frac{7}{8} \\ -1 & \frac{7}{8} \le t < 1 \end{cases}$$

A reader familiar with wavelets will recognize the above Rademacher functions are very similar to the scale-and-shift basis *scaling* functions. (Refer to the text by Strang and Nguyen [133] for a treatment on wavelets.) It is sometimes more convenient to express Rademacher functions graphically (see Fig. 4.20 (a)) or by "sign" notation:

$$\begin{aligned} r_1 &: \; ++++---- \\ r_2 &: \; ++--++-- \\ r_3 &: \; +-+-+-+- \end{aligned}$$

We can easily deduce the following (also see Fig. 4.20 (b)):

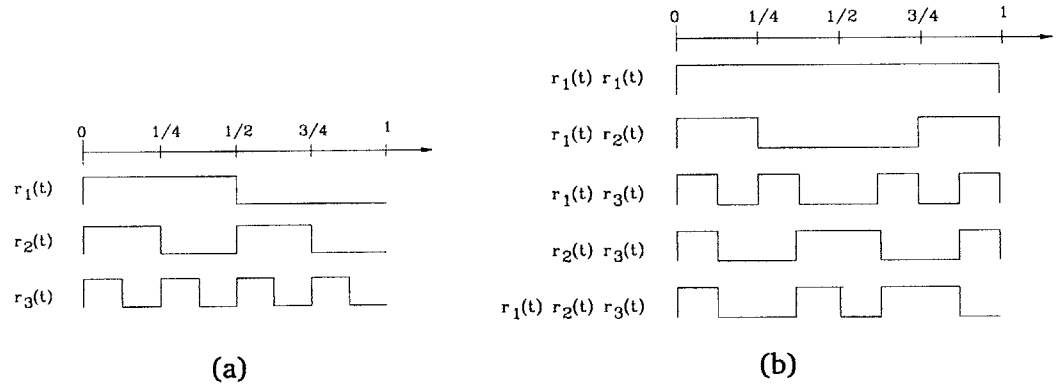



**Figure** 4.20: Rademacher functions and their products

$$\begin{aligned} r_i r_i &: \; +++++++++ \quad \forall i \in \mathbb{Z}_+ \\ r_1 r_2 &: \; ++----++ \\ r_1 r_3 &: \; +-+--+-+ \\ r_2 r_3 &: \; +--++--+ \\ r_1 r_2 r_3 &: \; +--+-++- \end{aligned}$$

It can be seen that the above three Rademacher functions and their distinct products make up the set of all Walsh waveforms of length 8 (see. Fig. 4.13). In general, it can be shown that a set of all Walsh waveforms of length $2^k$ can be generated from the set of Rademacher functions $\{r_i(t)\}_{i=1}^{k}$.

## Wavelet-type Hadamard-Walsh Codes

Wavelet functions of higher resolution are generated by taking the sums and differences of the original functions. Here, we give an example using four original functions $\{x_i(t)\}_{i=1}^{4}$. Generalizations for longer lengths are straightforward. Consider a set of eight functions as a column vector:

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix}$$

The sums and differences of the original functions are:

$$\mathbf{y} = \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \end{bmatrix} = \begin{bmatrix} x_1(t) + x_2(t) \\ x_1(t) - x_2(t) \\ x_3(t) + x_4(t) \\ x_3(t) - x_4(t) \end{bmatrix}$$

In matrix notation:

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix}$$

For the next higher resolution stage, the *alternate* sums and differences are:

$$\mathbf{z} = \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ z_4(t) \end{bmatrix} = \begin{bmatrix} y_1(t) + y_3(t) \\ y_2(t) + y_4(t) \\ y_1(t) - y_3(t) \\ y_2(t) - y_4(t) \end{bmatrix}$$

In matrix notation:

$$\mathbf{z} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \end{bmatrix}$$

After combining:

$$\mathbf{z} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \mathbf{x}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \mathbf{x}$$

We recognize the final matrix as the Walsh matrix $\mathcal{W}_4$. The generation of wavelet-type Walsh codes unique in the sense that the matrix entries are from a ternary set $\{-1, 0, 1\}$, whereas in the previous cases, the set is either bipolar $\{-1, 1\}$ or binary $\{0, 1\}$. In all types discussed so far, the lengths of Walsh codes are limited to powers of 2. We next describe the *Paley*-type where the length is $4k$, $k \in \mathbb{Z}_+$.

### Paley-type Hadamard-Walsh Codes

If $p$ is a positive odd prime number, then all non-zero squares modulo $p$ are defined as *quadratic residues*(QR) of $p$. Since $p$ is prime, we need to consider only the squared numbers $0^2$, $1^2$, $2^2$, ..., $(p-1)^2$. For $1 \leq a \leq p-1$,

$$(p-a)^2 \bmod p = a^2 \bmod p$$

Thus, we need to evaluate only half as many; i.e., $1^2$, $2^2$, ..., $\left(\frac{p-1}{2}\right)^2$. The remaining integers between 1 and $(p-1)$ that are not quadratic residues of $p$ are called *non-residues*. A Walsh-Hadamard matrix of order $n = (p+1) = 4k$, where $p$ is odd prime and $k$ is an integer, can be constructed as follows:

$$\mathcal{W}_n = \begin{bmatrix} 1 & 1 \\ 1^T & \mathbf{Q} - \mathbf{I} \end{bmatrix} \tag{4.76}$$

where $1$ is a row vector of all 1's of length $(n-1)$, and $\mathbf{Q} = (q_{ij})$ is the $(p \times p)$ *Jacobsthal* matrix with the following property:

$$q_{ij} = -q_{ji} = \chi(j-i)$$

Note that $\mathbf{Q}$ is *skewed-symmetric*, and $\chi$ is called the *Legendre symbol* with the following possible values:

$$\chi(i) = \begin{cases} 0 & i=\text{multiple of } p \\ 1 & i=\text{QR of } p \\ -1 & i=\text{non-QR of } p \end{cases}$$

We see immediately that **Q** is a diagonal matrix with the main diagonal terms of **Q** all equal to zero, $q_{ii} = 0$. In fact, it can be shown that **Q** is *circulant*; i.e., there is a wrap-around of the off-diagonal terms:

$$\mathbf{Q} = \begin{bmatrix} q_1 & q_p & \cdot & \cdot & q_2 \\ q_2 & q_1 & q_p & \cdot & q_3 \\ \cdot & q_2 & q_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & q_p \\ q_p & q_{p-1} & \cdot & q_2 & q_1 \end{bmatrix}$$

As an example, let $p = 7$.

$$1^2, 2^2, \ldots, \left(\frac{p-1}{2}\right)^2 = 1^2, 2^2, 3^2 = 1, 4, 2$$

The QR of p are 1,2 and 4. The non-residues are 3, 5 and 6. The components of the first column vector of **Q** become:

$$(q_1\ q_2 \cdots q_{p-1}\ q_p) = (0\ -\ -\ +\ -\ +\ +)$$

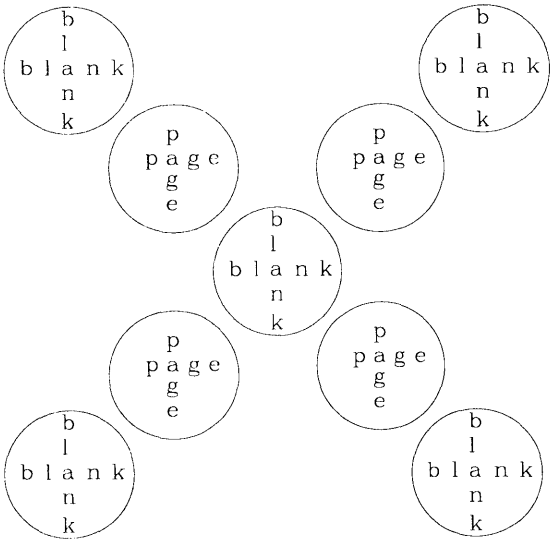
i.e., the components whose indices are QR of p take on value $-1$ while non-QR indexed components have value 1, leading to:

$$\mathbf{Q} = \begin{bmatrix} 0 & + & + & - & + & - & - \\ - & 0 & + & + & - & + & - \\ - & - & 0 & + & + & - & + \\ + & - & - & 0 & + & + & - \\ - & + & - & - & 0 & + & + \\ + & - & + & - & - & 0 & + \\ + & + & - & + & - & - & 0 \end{bmatrix}$$

Substituting in eqn.(4.76),

$$\mathbf{W}_n = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & + & - & + & - & - \\ + & - & - & + & + & - & + & - \\ + & - & - & - & + & + & - & + \\ + & + & - & - & - & + & + & - \\ + & - & + & - & - & - & + & + \\ + & + & - & + & - & - & - & + \\ + & + & + & - & + & - & - & - \end{bmatrix}$$

The above Walsh-Hadamard matrix is not a reflection matrix since it is not symmetric. It is proved in [80] that if a Walsh-Hadamard matrix of order n exists, then n is either 1, 2 or multiples of 4. It is, however, a conjecture that Walsh-Hadamard matrices exist whose orders are multiples of 4.


blank
page
blank
page
blank
page
page
blank
blank

# 5

# TIME-BANDLIMITED PULSE DESIGN

## Summary

Pulse shaping is an old technique in communication theory that is commonly applied to maximize the information rate for a given channel frequency response. It is true that reliable information exchange is guaranteed only for rates less than the channel capacity. This theoretical limit deduced from the Shannon capacity formula itself is achieved under the assumption that transmit pulses are shaped as infinite-length sinc pulses. If the notion of symbol time and bandwidth are relaxed, then a special class of "time-bandlimited" waveforms known as spheroidal prolate functions can be substituted in place of sinc pulses. In practice, physically realizable Nyquist pulse shapes that cause no intersymbol interference are used. Our main concern regarding the subject of pulse shaping is: "In a practical setting, does a particular type of carrier modulation scheme achieve a higher information rate than its alternatives?" In particular, we wish to compare the bandwidth efficiencies of single- and multi-carrier modulation formats. Finding a solution to this problem is the focal point of this chapter. We summarize our main results:

- Under ideal conditions of brick-wall channel response and infinite-length sinc transmit pulse shape, both single- and multi-carrier modulation formats are equivalent in bandwidth efficiency.
- Once the non-realizable sinc pulses are replaced with infinite-length Nyquist pulses, the multi-carrier format is more bandwidth efficient.
- Truncation of an infinite-length Nyquist pulse destroys the unique orthogonality property at the receiving end and invariably results in intersymbol interference. On the

other hand, symbol-time limited finite-length pulses cause inter-channel (or inter-carrier) interference—also commonly known as spectral leakage.

- For multi-carrier modulation, there exists a special class of finite-length pulse shapes that satisfies the Nyquist criterion for zero intersymbol interference with an added bonus of low spectral leakage.

The chapter is divided into two parts: the first deals with the measure of bandwidth, and the second is the design of finite-length Nyquist pulse shapes. When the term "bandwidth" is not clearly defined, it can lead to ambiguity in quantifying and interpreting the performance of a communication system. In the context of carrier modulation, the extent of the RF power spectrum of a train of transmit pulses, or the RF frequency response of the channel or the Fourier transform of the RF received signal all qualify as bandwidth. In this chapter, we limit our definition of bandwidth to the extent of the power spectral density of a train of transmit pulses. In general, the PSD of transmit pulses depends on both the property of a shaping filter as well as on the correlation properties of the modulating data sequence. We show that for linear modulation, the PSD is independent of the statistical property of the data sequence as long as the data symbols belong to an antipodal symmetrical (balanced) set. Hence the PSD becomes a function of the selected pulse shape only. We next describe various infinite and finite length pulse shapes that meet the Nyquist criterion for zero ISI. In particular, we highlight a class of finite-length time-overlapped pulses that satisfies the generalized Nyquist criterion and is thus suitable for multi-carrier modulation.

## 5.1 Time-Frequency Duality

When we compute channel capacity in Ch. 3, we implicitly assume that the transmit signal $s(t)$ is a stream of sinc pulses at a rate of $1/T_s$ sec. We know that sinc pulses have a brick-wall amplitude response and are the most spectrally efficient for transmission via an ideal channel of passband bandwidth $W$ Hz. For sinc pulses $T_s = 1/W$. On the other hand, when we study signature waveforms, we assume for simplicity that the spread-spectrum modulating chip pulses are non-overlapping rectangular pulses. By convention, a rectangular pulses is denoted by

$$\Pi_T(t) = \begin{cases} 1 & |t| < T/2 \\ 0 & |t| > T/2 \end{cases} \tag{5.1}$$

Its Fourier transform is

$$\mathcal{F}\left\{\Pi_T(t)\right\} = T\,\frac{\sin(\pi f T)}{\pi f T} = T\,\mathrm{sinc}(fT) \tag{5.2}$$

where $W = 1/T$. If $W$ is passband bandwidth, $W = 2/T$. The amplitude $|\Pi_W(f)|$ in decibel scale is plotted in Fig. 5.1. Notice the spectral leakage of a rectangular pulse outside the mainlobe

bandwidth of $W$ Hz. The first sidelobe is only about 15 dB below the mainlobe peak value. If the sinc function is used as the chip pulse, its amplitude response is $\Pi_W(f)$ — a brick wall— and there is no spectral leakage. However, sinc pulses are not practical:

- A finite-impulse response (FIR) baseband filter cannot be used for their generation due to their infinite time span
- At the receiving end, the integration window of a baseband filter must be infinite; i.e., an infinite-impulse response (IIR) baseband filter is required. Since information symbols must be decoded at a rate of $1/T$, long delays in baseband filtering are not acceptable.
- The succession of sinc pulses at a rate of $1/T$ satisfy the Nyquist criterion for zero intersymbol interference. However, any slight error in carrier frequency, symbol sampling rate and sampling instant leads to a large amount of ISI.
- Any channel distortion that is not compensated at the receiver results in a stream of distorted sinc pulses that are no longer Nyquist.[1] The accumulated ISI drives the AGC (automatic gain control) amplifier of the receiver into saturation, rendering the received signal useless.

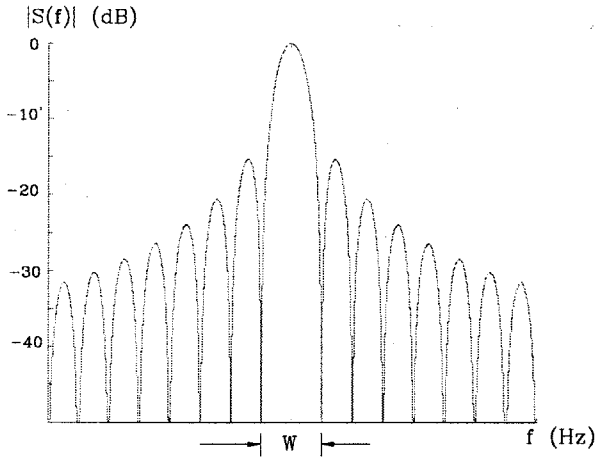



**Figure 5.1:** Multiple-access channel model with $m$ transmitters. Each transmitter $T_k$ communicates with its intended destination sink $S_k$ through the common receiver R.

Hence, we face a dilemma; a sinc Nyquist pulse train has an ideal flat frequency response, but it is impractical. A rectangular symbol-time-limited pulse train has infinite bandwidth, and if it is transmitted via a bandlimited channel, the output is severely distorted by ISI. The issue of

[1] A pulse shape is *Nyquist* if it satisfies the zero intersymbol interference criterion.

spectral leakage is of great concern for any RF communication service since strict regulatory guidelines on out-of-band spectral levels must be met.[2] There are several practical remedies. A sinc-like Nyquist pulse that is less sensitive to receiver and channel imperfections — and with good spectral characteristic— can be used. The most popular Nyquist pulse is the raised-cosine function. We should point out that raised cosine pulses are also infinite-length. Any truncation of sinc or raised cosine pulses results in spectral regrowth —due to Gibbs phenomenon— and non-zero ISI. If the Nyquist zero ISI condition is not imposed, then there are many more choices in pulse selection. The Gaussian pulse as applied in several European wireless systems is a good example. In this case, the main goal is spectral containment —Gaussian pulse has a bell-shaped smooth frequency response. Other non-Nyquist pulses can be constructed by shaping a rectangular pulse. In signal processing literature, various pulse shaping methods are collectively known as *windowing*. Some common window functions are Hamming, Bartlett and Kaiser. All are strictly symbol-time limited. Another option is to construct pulses that are still finite-length but with improved spectral containment by increasing the time span beyond the symbol duration. If such a pulse train is generated at rate $1/T$, the portion outside the range $[0, T]$ overlaps with an adjacent pulse. We define such pulses as *time-overlapped* functions.

## 5.2 Power Spectral Density

The channel of most communication systems —including the radio propagation channel— is band-limited. For a telephone line channel, bandwidth limitation may be due to the transmission medium (copper twisted pair) or a lowpass filter at the customer premise. In wireless communication, it is the regulatory body that defines a radio frequency range $[f_L, f_H]$ as the channel *bandwidth*. It may not be a strict end-to-end measure but a succession of frequency bands with acceptable power levels. An example of a *spectral mask* is shown in Fig. 5.2. It is the wireless system designer's duty to construct a carrier- and data-modulated pulse train with a power spectral density that is within the spectral mask. In our analysis, we treat the spectral mask as the "channel." The width and shape of the spectral content of a transmit signal is measured in terms of its *power spectral density* (PSD). If the power spectral density of the transmitted signal is non-zero outside the channel bandwidth, the received signal is distorted. This distortion results in intersymbol interference (ISI). Hence, a transmitted pulse train with a certain PSD that best matches the frequency response of the channel is selected such that distortion due to ISI is contained or completely eliminated. In general, the PSD of a train of modulated waveforms depends on the shape of the common pulse as well as on the auto-correlation of the embedded data sequence.

[2]Government and international bodies such as FCC and ITU set these mandatory guidelines such that two wireless systems with adjacently allocated RF bands do not interfere with each another.

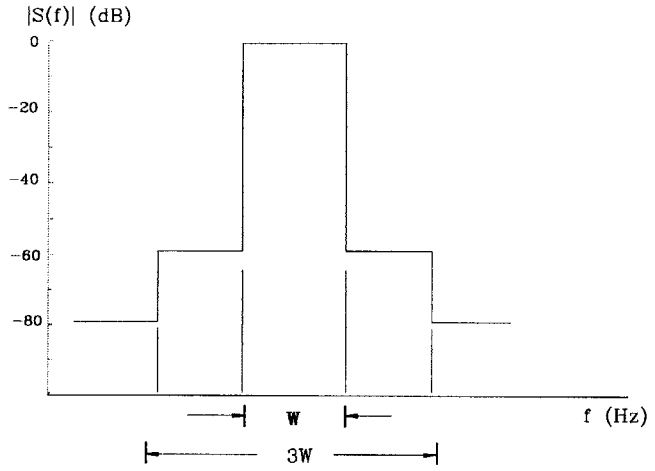



**Figure** 5.2: Multiple-access channel model with m transmitters

### 5.2.1 Deterministic Signals

Let g(t) be a *deterministic* baseband signal. If it is energy-limited, its spectral content is expressed by its Fourier transform G(f). If it is a power signal with infinite span, we can always truncate by taking a finite-time window of g(t)

$$g_w(t) = g(t)w(t) \tag{5.3}$$

where the window function

$$w(t) = \begin{cases} 1 & -\frac{T_w}{2} \le t \le \frac{T_w}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

$T_w$ is called the *window length*. The power $P_g$ is defined as the time-average or mean squared value of g(t):

$$P_g = \lim_{T_w \to \infty} \frac{1}{T_w} \int_{-T_w/2}^{T_w/2} g^2(t)\, dt \tag{5.5}$$

Substituting the time-limited energy signal of (5.3):

$$P_g = \lim_{T_w \to \infty} \frac{1}{T_w} \int_{-\infty}^{\infty} g_w^2(t)\, dt$$

Applying Parseval's Theorem:

$$\begin{aligned} P_g &= \lim_{T_w \to \infty} \frac{1}{T_w} \left[ \int_{-\infty}^{\infty} |G_w(f)|^2\, df \right] \\ &= \int_{-\infty}^{\infty} \lim_{T_w \to \infty} \frac{|G_w(f)|^2}{T_w}\, df \end{aligned}$$

The power spectral density $S_g(f)$ is:

$$S_g(f) = \lim_{T_w \to \infty} \frac{|G_w(f)|^2}{T_w} \tag{5.6}$$

For some window length $T_w$ we see that the shape and width of the PSD of $g(t)$ is approximately equal to the squared amplitude of its truncated version. If $g(t)$ is periodic —by infinite concatenation of $g_w(t)$— a window length $T_w = T$ is sufficient. The PSD expressed in decibel [dB] scale is

$$\begin{aligned} S_g(f)\ [\text{dB}] &= 20\ \log(|G_w(f)|) + K \qquad K = -10\log(T) \\ &= |G_w(f)|\ [\text{dB}] + K \end{aligned} \tag{5.7}$$

Thus, the PSD of $g(t)$ depends only on the magnitude of its Fourier transform $G_w(f)$.

### 5.2.2 Random Processes

If $g(t)$ is carrier-modulated, its PSD is translated and scaled but its shape and width remain unchanged. If $g(t)$ is used as a common pulse for data modulation, the PSD of the resulting modulated signal also depends on the auto-correlation of data sequence. Since data from an information source is random, the modulated signal is modelled as a sample function of a random process. From eqn. (4.23), a linear data modulated signal can be expressed as:

$$g(t) = \sum_{i=-\infty}^{\infty} x[i]\, f(t - iT) \tag{5.8}$$

where $x[i] \in \mathcal{X}$, a finite signal constellation set. Note that the common pulse $f(t)$ is not necessarily restricted to the symbol interval $[-T/2, T/2]$. In fact, it is possible for $g(t)$ to consists of a train of time-overlapping pulses. If each $x[i]$ is i.i.d., the modulation is *memoryless*. $\{x[i]\}$ is stationary if its joint distributions (equivalently, all its moments) are time-invariant. If $x[i]$ is *wide-sense stationary*, then its first and second moments are time-invariant:

$$E[x[i]] = \mu \qquad E(x[i]x[j]) = R_x[i-j] \tag{5.9}$$

The mean value of $g(t)$ is

$$\begin{aligned} E[g(t)] &= \sum_{i=-\infty}^{\infty} E[x[i]] f(t - iT) \\ &= \mu \sum_{i=-\infty}^{\infty} f(t - iT) \end{aligned}$$

It is periodic with period $T$. The auto-correlation function of $g(t)$ is:

$$\begin{aligned} R_g(t+\tau; t) &= E\Big[(g(t)g(t+\tau)\Big] \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E\Big[(x[i]x[j])\Big]\ f(t - iT)\, f(t + \tau - jT) \end{aligned} \tag{5.10}$$

Note that the second term on the RHS —the deterministic auto-correlation of $f(t)$— is a function of $\tau$ and T. The frequency-domain expression of $g(t)$ requires a two-dimensional Fourier transform of both $\tau$ and T. Alternatively, we can remove the dependence on T by averaging $R_g(t+\tau;t)$ over T. The time-averaged auto-correlation of $g(t)$ is:

$$\overline{R}_g(t+\tau;t) = \frac{1}{T}\int_{-T/2}^{T/2} \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} E\big(\big[x[i]x[j]\big]\big) f(t-iT)\, f(t+\tau-jT)\, dt \tag{5.11}$$

The above integral is further simplified by considering the following four cases:

**Case 1**: Non-Zero Mean, Wide-Sense-Stationary Sequence

$$\begin{aligned} E\big[x[i]\big] &= \mu \\ E\big[x[i]x[j]\big] &= R_x[i-j] \end{aligned} \tag{5.12}$$

Using the time-invariance property and a change in variables $l = i-j$, eqn.(5.11) simplifies to:

$$\begin{aligned} \overline{R}_g(\tau) &= \overline{R_g(t+\tau;t)} \\ &= \frac{1}{T}\int_{-T/2}^{T/2} \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} R_x[i-j]\, f(t-iT)\, f(t+\tau-jT)\, dt \\ &= \sum_{l=-\infty}^{\infty} R_x[l] \sum_{j=-\infty}^{\infty} \frac{1}{T}\int_{-T/2-jT}^{T/2-jT} f(t)\, f(t+\tau-lT)\, dt \\ &= \frac{1}{T}\sum_{l=-\infty}^{\infty} R_x[l] C_f(\tau-lT) \end{aligned} \tag{5.13}$$

where the time-average auto-correlation function of $f(t)$ is defined as:

$$C_f(\tau) = \int_{-\infty}^{\infty} f(t)\, f(t+\tau)dt \tag{5.14}$$

The random process $g(t)$ is *cyclo-stationary* due to periodicity of its mean and auto-correlation function $R_f(t)$, even though the embedded sequence is WSS. By taking the Fourier transform of both sides,

$$\overline{S}_g(f) = S_x(f)\, S_f(f) \tag{5.15}$$

Note that the PSD of the random sequence is periodic with period $1/T$:

$$S_x(f) = \sum_{l=-\infty}^{\infty} R_x[l] \exp(-j2\pi flT) \tag{5.16}$$

For a purely random sequence $\{x[i]\}$, it can be shown that:

$$R_x[l] = \begin{cases} \mathrm{Var}\big[x[i]\big] + \mu^2 & l = 0 \\ \mu^2 & l \neq 0 \end{cases} \tag{5.17}$$

Then,

$$S_x(f) = \mathrm{Var}[x[i]] + \mu^2 \sum_{l=-\infty}^{\infty} \exp(-j2\pi f l T) \tag{5.18}$$

Using Poisson's sum formula [97],

$$S_x(f) = \mathrm{Var}[x[i]] + \frac{\mu^2}{T} \sum_{l=-\infty}^{\infty} \delta\left(f - \frac{l}{T}\right) \tag{5.19}$$

Finally, the PSD of $g(t)$ consists of both continuous and periodic, discrete components:

$$S_g(f) = \mathrm{Var}\left[x[i]\right] S_f(f) + \frac{\mu^2}{T^2} \sum_{l=-\infty}^{\infty} \left|F\left(\frac{l}{T}\right)\right|^2 \delta\left(f - \frac{l}{T}\right) \tag{5.20}$$

where $F(f) = \mathcal{F}\{f(t)\}$.

**Case 2**: Zero Mean, Wide-Sense-Stationary Sequence
Set $\mu = 0$. Then eqn.(5.20) simplifies to:

$$S_g(f) = \mathrm{Var}[x[i]]\, S_f(f) \tag{5.21}$$

The zero-mean condition holds for symmetric PAM signal constellation sets. The PSD of $g(t)$ now depends only on the shape of a deterministic pulse $f(t)$.

**Case 3**: Non-Zero Mean, Cyclo-Stationary Sequence
In a spread-spectrum modulated signal, the embedded pseudo random code sequence has period P; hence, the sequence itself is cyclo-stationary:

$$\begin{aligned} E[x[i]] &= 0 \\ E[x[i]x[j]] &= R_x[i; i-j] \\ &= R_x[i + kP; i-j] \qquad k \in \mathbb{Z} \end{aligned} \tag{5.22}$$

Rewriting eqn.(5.13) with above modification,

$$\begin{aligned} \overline{R}_g(\tau) &= \frac{1}{T} \int_{-T/2}^{T/2} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} R_x[i; i-j]\, f(t - iT)\, f(t + \tau - jT)\, dt \\ &= \sum_{l=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} R_x[i; l] \frac{1}{T} \int_{-T/2-iT}^{T/2-iT} f(t)\, f(t + \tau - lT)\, dt \\ &= \frac{1}{T} \sum_{l=-\infty}^{\infty} \left[ \sum_{i=-\infty}^{\infty} R_x[i; l] \right] R_f(\tau - lT) \\ &= \frac{1}{T} \sum_{l=-\infty}^{\infty} \left[ \frac{1}{P} \sum_{i=1}^{P} R_x[i; l] \right] R_f(\tau - lT) \\ &= \frac{1}{T} \sum_{l=-\infty}^{\infty} \overline{R}_x[l]\, R_f(\tau - lT) \end{aligned} \tag{5.23}$$

where $\overline{R}_x[l]$ is the time-average of a periodic *statistical* auto-correlation function of a cyclo-stationary sequence $\{x[i]\}$. Hence, $\overline{R}_x[l]$ is analogous to the time-average correlation function $\overline{R}_f(t)$ of the deterministic pulse $f(t)$. By taking Fourier transforms on both sides of eqn.(5.23),

$$\overline{S}_g(f) = \overline{S}_x(f)\, S_f(f) \tag{5.24}$$

where $\overline{S}_x(f) = \mathcal{F}\{\overline{R}_x[l]$. Since the sequence is periodic, it is not a truly random sequence over the entire time span. However, within a period P, it is approximately random. We can deduce the expression for the PSD of $g(t)$ —similar to eqns. (5.17)–(5.20)— by substituting $\overline{R}_x[l]$ in place of $R_x[l]$.

**Case 4**: Zero Mean, Cyclo-Stationary Sequence
When the zero-mean condition is invoked, the PSD of $g(t)$ is a function of $f(t)$ only. For a zero-mean pseudo-random binary sequence, we have already shown that its time-average auto-correlation function is equal to its statistically averaged auto-correlation function.

## 5.3 Nyquist Criterion for Zero Interference

In the previous section, we show that the power spectral density $S_g(t)$ of a data modulated signal $g(t)$ is a product of the PSD's of the deterministic common pulse $f(t)$ and the embedded random sequence $\{x[i]\}$. If the mean value of the sequence is zero, we remove the dependence of $S_g(t)$ on $\{x_m[i]\}$. The simplest way of achieving a zero-mean discrete sequence is by selecting code symbols from a symmetric (balanced) constellation set such as

$$\mathbf{S} = \left\{ -s_M \; -s_{M-1} \; \ldots \; -s_2 \; -s_1 \; s_1 \; s_2 \; \ldots \; s_{M-1} \; s_M \right\}$$

where each bipolar pair $-s_n$ and $s_n$ are equally likely. The equally likely condition of a symmetric pair is guaranteed by scrambling[3] the sequence. In fact, this technique is almost always used in practice for spectral shaping. By imposing the zero-mean condition for the data sequence, we have effectively compartmentalize two optimization tasks in signal design: spectral shaping via the PSD of $g(t)$ for channel matching, and the design and construction of an embedded sequence $\{x_m[i]\}$ with low correlation properties. The latter is already covered in Chapter 4. As we stated earlier, if the PSD of a transmitted signal is larger than the channel bandwidth, the received signal is severely distorted. The main objective of spectral shaping is to transmit pulses at the highest possible rate with no intersymbol interference. Nyquist and others have studied this problem extensively. We summarize their key results:

[3]As an example, consider the 4-PAM set $\{-s_2 \; -s_1 \; s_1 \; s_2\} = \{00\ 01\ 10\ 11\}$. If a binary sequence has consecutive '1's such as ...11111... then the symbol $s_2$ is more likely than $-s_2$. This imbalance can be ameliorated by mod-2 addition of this binary sequence with a pseudo-random binary sequence.

### 5.3.1 Single-Carrier Nyquist Criterion

If a linearly modulated baseband signal

$$g(t) = \sum_{i=-\infty}^{\infty} x[i]\, f(t - iT)$$

is the input to a linear time-invariant channel with impulse response $c(t)$, the output is

$$\sum_{i=-\infty}^{\infty} x[i]\, r(t - iT)$$

where $r(t) = g(t) \star c(t)$. After baseband filtering at the receiver with impulse response $h(t)$, the output is

$$\sum_{i=-\infty}^{\infty} x[i]\, p(t - iT)$$

where

$$p(t) = r(t) \star h(t) = f(t) \star c(t) \star h(t)$$

Its samples at the symbol rate are

$$\sum_{i=-\infty}^{\infty} x[i]\, p(jT - iT)$$

The pulse $p(t)$ is *Nyquist* —i.e., it satisfies the Nyquist criterion— if

$$\sum_{i=-\infty}^{\infty} p\big[(j-i)T\big] = \delta_{ij} \tag{5.25}$$

At sampling instant $iT$, the only data symbol detected is $x[i]$. This is the zero intersymbol interference condition. Equivalently $p(t)$ is Nyquist if

$$P_+(f) \triangleq \sum_{j=-\infty}^{\infty} P(f - jf_o) = T \tag{5.26}$$

where $f_o = 1/T$. In words if the sum of the Fourier transform of $p(t)$ and its *aliased* copies has a flat spectrum, $p(t)$ is Nyquist. The above results are still valid if $g(t)$ is $f_c$ carrier-modulated and transmitted through a passband channel with impulse response

$$c_p(t) = \Re\big\{c(t)\exp(2\pi f_c t)\big\}$$

and carrier-demodulated at the receiver such that the output is $r(t)$. Hence, we interpret a pulse that satisfies eqn. (5.25) as *single-carrier Nyquist* or s-Nyquist for short.

It is obvious that a pulse with a brick-wall frequency response $\Pi_{f_o}(f)$ is Nyquist. The aliased

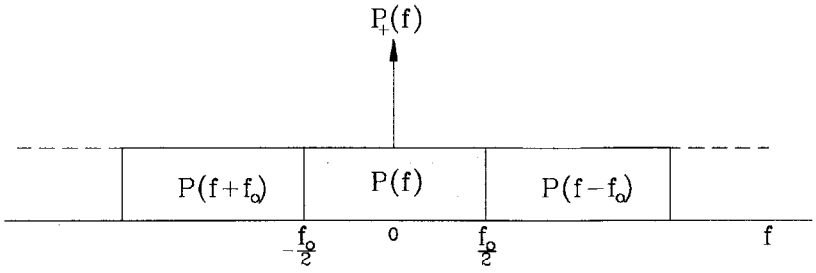



**Figure** 5.3: Fourier transform of sinc pulse and its aliased copies

Fourier transform $P_+(f)$ is plotted in Fig. 5.3. By observation, we deduce immediately that the Fourier transform must be an even function. It is clear that the sinc function is the only Nyquist pulse with the minimum baseband bandwidth of $f_o/2$ Hz, implying that the maximum symbol rate through a linear time-variant channel with baseband bandwidth $W$ Hz is $2W$ pulses per second. If we allow pulses with bandwidth larger than $f_o/2$ Hz, then there are infinitely many Nyquist pulses. The wider its bandwidth, the less spectrally efficient is the Nyquist pulse. For pulses whose bandwidth is strictly limited to the range $[f_o/2, f_o]$, we need consider spectral overlap from an adjacent aliased transform only. Eqn. (5.26) becomes:

$$P(f) + P(f - f_o) = T \qquad 0 < f < f_o \tag{5.27}$$

By defining the *excess bandwidth* $f_\tau = (f_o/2 + \Delta_e)$ and using the fact that $p(t)$ is real:

$$P\left(\frac{f_o}{2} + \Delta_e\right) + P\left(\frac{f_o}{2} - \Delta_e\right) = T \qquad |\Delta_e| < \frac{f_o}{2} \tag{5.28}$$

If $p(t)$ has linear phase $\theta_p(f) = -Kf$ ($K$ is a delay constant):

$$P(f) = |P(f)|e^{-j\theta_p}$$

The Nyquist criterion depends only on the shape of the amplitude response:

$$\left|P\left(\frac{f_o}{2} + \Delta_e\right)\right| + \left|P\left(\frac{f_o}{2} - \Delta_e\right)\right| = T \qquad |\Delta_e| < \frac{f_o}{2} \tag{5.29}$$

Since $|P(f)|$ is even we need consider the shape on one side only, say $0 < f < f_o$. Graphically, the amplitude condition is shown in Fig. 5.4. It can be broken into four parts: the gain T, even term $P_e(f)$, odd term $P_o(f)$ and dc bias $P_{dc}(f)$

$$|P(f)| = T\left(P_e(f) + P_o\left(f - \frac{f_o}{2}\right) + P_{dc}(f)\right) \qquad f \geq 0$$

Specifically,

$$P_e(f) = 0.5 \qquad |f| < \frac{f_o}{2} - f_r \tag{5.30}$$

$$P_o(f) = -P_o(-f) \qquad |f| < f_r \tag{5.31}$$

$$P_{dc}(f) = 0.5 \qquad |f| < \frac{f_o}{2} + f_r \tag{5.32}$$

The amplitude of the odd function is normalized such that

$$\left|P_o(\pm \frac{f_r}{2})\right| = 0.5$$

Outside their specified intervals, $P_e(f)$ and $P_o(f)$ are undefined.

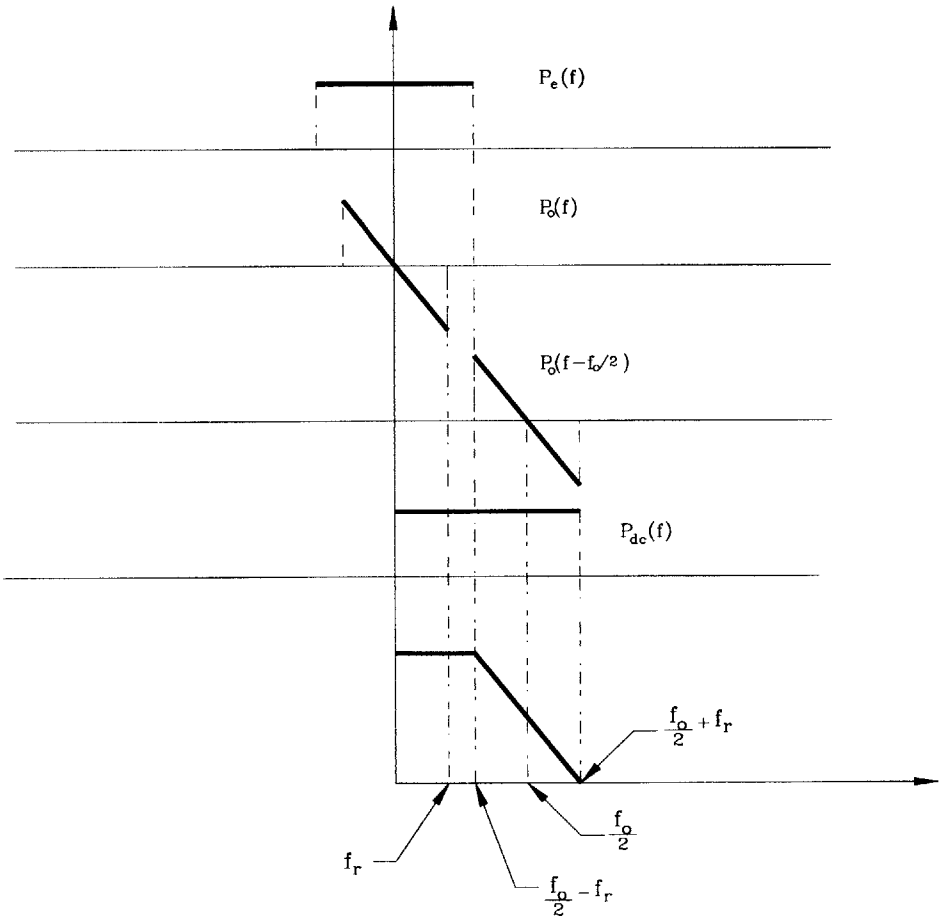



**Figure** 5.4: Amplitude response of Nyquist pulse in four parts

A good example is a family of pulses collectively known as raised-cosine pulses with Fourier

transform:

$$|P_{RC}(f)| = \begin{cases} T & |\omega| < \frac{f_o}{2} - f_r \\ \frac{T}{2}\left[1 - \sin\frac{\pi}{2}\left(\frac{f - f_o/2}{f_r}\right)\right] & \left|f - \frac{f_o}{2}\right| < f_r \\ 0 & |f| > \frac{f_o}{2} + f_r \end{cases} \tag{5.33}$$

The *roll-off factor* $r = 2f_r/f_o$ ranges from 0 (sinc pulse) to 1 (fully raised cosine). For a fully raised cosine function with $f_r = f_o/2$, $P_e(f)$ is non-existent. It is easily deduced from eqn. (5.33)

$$|P_{FRC}(f)| = \cos^2\left(\frac{\pi f}{2f_o}\right) \tag{5.34}$$

The plot of $P_+(f)$ for a fully raised cosine function is shown in Fig. 5.5. The symbol rate is now $W$ samples per second —half the rate of sinc pulse.

Note that both sinc and raised cosine pulses are strictly bandlimited; i.e., they have infinite

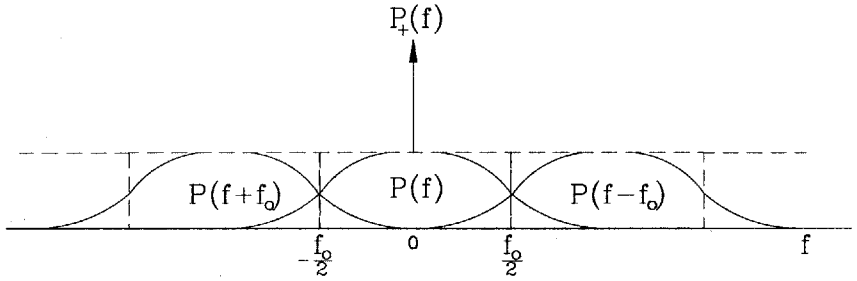



**Figure** 5.5: Fourier transform of fully raised cosine pulse and its aliased copies

length. It is true that raised cosine pulses are less susceptible to sampling rate and timing errors because of faster delay rate ($1/t^3$). However, any truncation by windowing leads to regrowth of sidelobes, and the Nyquist condition is no longer satisfied. A plot of $P_+(f)$ of a truncated sinc pulse is shown in Fig. 5.6.

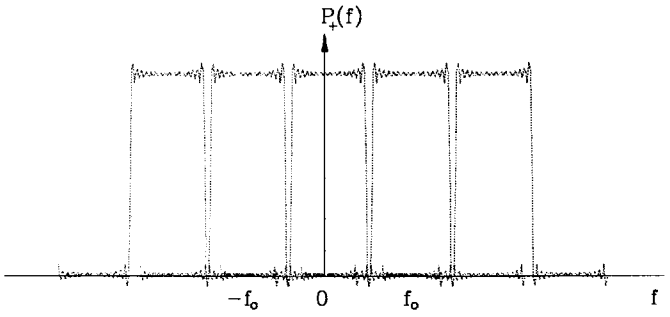



**Figure** 5.6: Fourier transform of truncated sinc pulse and its aliased copies

### 5.3.2 Multi-Carrier Nyquist Criterion

If the sum of K independent, linearly modulated baseband signal

$$\sum_{k=1}^{K} g_k(t) = \sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} x_k[i]\, f_k(t - iT)$$

is the input to a linear time-invariant channel with impulse response $c(t)$, the output is

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} x_k[i]\, r_k(t - iT)$$

where $r_k(t) = g_k(t) \star c(t)$. After baseband filtering at the receiver with impulse response $h_n(t)$, the output is

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} x_k[i]\, p_{kn}(t - iT)$$

where

$$p_{kn}(t) = r_k(t) \star h_n(t) = f_k(t) \star c(t) \star h_n(t)$$

Its samples at the symbol rate are

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} x_k[i]\, p_{kn}(jT - iT)$$

The pulse $p_{kn}(t)$ is *Generalized Nyquist* or g-Nyquist —i.e., it satisfies the Generalized Nyquist criterion— if

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} p_{kn}[(j - i)T] = \delta_{ij}\, \delta_{kn} \tag{5.35}$$

If only

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} p_{kn}[(j - i)T] = \delta_{ij}$$

then at a sampling instant $iT$ the detected data symbols are $\sum_{k=1}^{K} x_k[i]$. The desired symbol is $x_n[i]$. This is *inter-channel interference* (ICI) —i.e., intersymbol interference from *current* symbols $x_k[i]$, $i \neq n$ of other parallel channels. If only

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} p_{kn}[(j - i)T] = \delta_{kn}$$

At a sampling instant $iT$ the detected data symbols are $\sum_{i=-\infty}^{\infty} x_n[i]$. The desired symbol is $x_n[0]$. This is *inter-symbol interference* (ISI) —i.e., intersymbol interference from *past* and possibly *future* symbols $x_n[j]$, $j \neq i$ of the same channel. Finally if

$$\sum_{k=1}^{K} \sum_{i=-\infty}^{\infty} p_{kn}[(j - i)T] \neq \delta_{ij}\, \delta_{kn}$$

then the desired symbol $x_n[i]$ suffers from ISI, ICI and interference from the past and future symbols from other parallel channels

$$\sum_{\substack{k=1 \\ k\neq n}}^{K} \sum_{\substack{j=-\infty \\ j\neq i}}^{\infty} x_k[j]$$

This is *inter-block interference* (IBI). A Fourier transform condition analogous to the single-carrier case of eqn. (5.26) is

$$P_{kn+}(f) = \sum_{j=-\infty}^{\infty} P_{kn}(f - jf_o) = T\,\delta_{kn} \tag{5.36}$$

In words $p_{kn}(t)$ is g-Nyquist if it is single-carrier Nyquist and the sum of its Fourier transform and its aliased copies is zero for $k \neq n$. The second condition states that the Fourier transform of $p_{kn}(t)$ must be zero for $k \neq n$. Note that a single-carrier Nyquist condition is a necessary but not a sufficient prerequisite for multi-carrier Nyquist condition. Consider the following for $k \neq n$:

$$\begin{aligned} P_{kn}(f) &= \mathcal{F}\{p_{kn}(t)\} \\ &= \mathcal{F}\Big\{f_k(t) \star c(t) \star p_n(t)\Big\} \\ &= \mathcal{F}\Big\{r_k(t) \star p_n(t)\Big\} \\ &= R_k(f) \cdot P_n(f) \end{aligned}$$

Since the $k^{th}$ transmit and $n^{th}$ receive filters have non-zero frequency responses $P_{kn}(\omega) = 0$ if and only if $R_k(\omega) = 0$ or $R_k(\omega)$ and $P_n(\omega)$ do not overlap. The former condition $R_k(\omega) = 0$ is not acceptable since the $k^{th}$ receiver must be able to detect its transmitted symbols $\{x_k[i]\}$. The latter condition —the non-overlapping of $R_k(\omega)$ and $P_n(\omega)$— is possible only if $R_k(\omega)$ is carrier-modulated or $P_n(\omega)$ is a passband filter response, or both. The design of g-Nyquist pulses is postponed until we first describe the connection between zero ISI condition and the orthogonality principle.

### 5.3.3 Orthogonality Criterion

It is well-known that in AWGN channel, the optimal receive filter has an impulse response that matches the received pulse:

$$H_k(f) = \mathcal{F}\Big[h_k(t)\Big] = R_k^*(f)$$

Writing the Fourier transform of a linearly modulated signal as

$$\sum_{k-1}^{K} \sum_{i=-\infty}^{\infty} x_k[i]\, R_k(f) \exp(-j2\pi fiT)$$

The output of the $n^{th}$ matched filter is:

$$x_n[0]\left|R_n(f)\right|^2 + \sum_{i\neq 0} x_n[i]\left|R_n(f)\right|^2 \exp(-j2\pi fiT) + \sum_{\substack{k=1\\k\neq n}}^{K}\sum_{i} x_k[i]\, R_k(f)\, R_n^*(f) \exp(-j2\pi fiT)$$

The first term on the RHS is the desired symbol. For the g-Nyquist condition to hold,

$$\left|R_n(f)\right|^2 \exp(-j2\pi fiT) + \sum_{\substack{k=1\\k\neq n}}^{K}\sum_{i} R_k(f)\, R_n^*(f) \exp(-j2\pi fiT) = 0$$

Equivalently,

$$\sum_{i\neq 0} r_n(t-iT)\, r_n(t) + \sum_{\substack{k=1\\k\neq n}}^{K}\sum_{i} r_k(t-iT)\, r_n(t) = 0 \tag{5.37}$$

That is, the received pulse for the $n^{th}$ channel must be *orthogonal* to its past and future (time-shifted) pulses (i.e., the zero ISI condition)

$$\sum_{i\neq 0} r_n(t-iT)\, r_n(t) = 0 \tag{5.38}$$

and also be orthogonal to pulses from other channels (i.e., zero ICI and IBI conditions)

$$\sum_{\substack{k=1\\k\neq n}}^{K}\sum_{i} r_k(t-iT)\, r_n(t) = 0 \tag{5.39}$$

The implication of above results is that the design pulses that satisfy the g-Nyquist criterion is equivalent to the construction of a set of mutually orthogonal received pulses (time functions). We next describe two types of g-Nyquist orthogonal pulses.

### Sinc-type Pulses

Consider the following two pulses, both with a brick-wall amplitude response. For convenience, assume the phase of $R_1(f)$ is zero.

$$R_1(f) = \begin{cases} c & |f| \leq 1/T \\ 0 & |f| > 1/T \end{cases}$$

$$R_2(f) = \begin{cases} c\, e^{j\theta_2(f)} & |f| \leq 1/T \\ 0 & |f| > 1/T \end{cases}$$

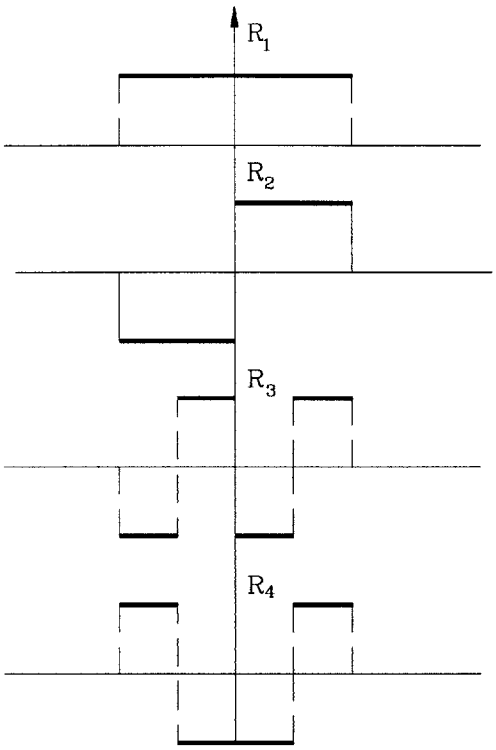



**Figure** 5.7: Frequency response of sinc-type g-Nyquist pulses

where

$$\theta_2(f) = \begin{cases} \pi/2 & 0 \le f \le 1/T \\ -\pi/2 & -1/T \le f \le 0 \end{cases}$$

Both pulses are illustrated in Fig. 5.7. We immediately deduce that

$$\sum_{j=-\infty}^{\infty} R_1(f - jf_o)\, R_2^*(f - jf_o) = 0$$

Similarly,

$$\sum_{j=-\infty}^{\infty} R_1^*(f - jf_o)\, R_2(f - jf_o) = 0$$

Since both pulses have brick-wall non-overlapping amplitude response, they satisfy the single-carrier Nyquist criterion:

$$\sum_{j=-\infty}^{\infty} \left| R_m(f - jf_o) \right| = 2c^2 \qquad m = 1, 2$$

It can be shown that the resulting g-Nyquist pulse is

$$p_{12}(t) = \mathcal{F}^{-}\Big\{P_{12}(f)\Big\} = c^2 \frac{\left(\cos\frac{2\pi t}{T} - 1\right)}{\pi t} \tag{5.40}$$

From the figure we see that $R_1(f)$ and $R_2(f)$ are identical to two-dimensional Walsh codes. It is then apparent that the design of sinc-type pulses (Walsh functions) that meet the g-Nyquist criterion is straightforward: For $n$ orthogonal pulses, define their phase responses such that their frequency responses are equivalent to the shape of $n$-dimensional Walsh codes. The only drawback is these pulses are not practical; they have steep (brick-wall) amplitude response and jumps in phase response.

### Chang Pulses

If excess bandwidth is allowed, there are numerous pulse shapes that satisfy the g-Nyquist criterion. We detail a class of $r = 1 (f_r = f_o/2)$, infinite-length, orthogonal and carrier-modulated pulses first proposed by R. Chang. It is assumed that the center frequencies of carriers are

$$f_i = (i + j - 0.5) f_o \qquad i \in \mathbb{Z}_+,\ j \in \mathbb{N} \tag{5.41}$$

The carrier frequency separation is $f_o$. For all carrier modulated pulses (any i)

$$R_i(f) = \begin{cases} |R_i(f)|\, e^{j\theta_i(f)} & |f - f_i| \le f_o \\ 0 & |f - f_i| > f_o \end{cases}$$

where

$$|R_i(f)|^2 = C_i + Q_i(f)$$

is composed of a DC bias term

$$C_i = 0.5 \qquad |f - f_i| \le f_o$$

and

$$Q_i(f) = \begin{cases} P_o(f - f_i - f_o/2) & f_i < f < f_i + f_o \\ -P_o(f - f_i + f_o/2) & f_i - f_o < f < f_i \end{cases}$$

The odd function $P_o(f)$ is defined in eqn. (5.31). It is seen that $Q_i(f)$ is even about $f_i$ and odd about $f_i \pm f_o/2$.

$$Q_i(f_i + f) = Q_i(f_i - f) \qquad 0 < f < f_i$$
$$Q_i(f_i' \pm f) = Q_i(f_i' \pm f) \qquad f_i' = f_i \pm f_o/2$$

The phase has a linear term plus a sinusoid

$$\theta_i(f) = k\pi\frac{f - f_i}{2f_o} + k_e \sin 2\pi\frac{f - f_i}{f_o} \qquad k \in \mathbb{N},\ k_e \in 2\mathbb{Z}_+$$

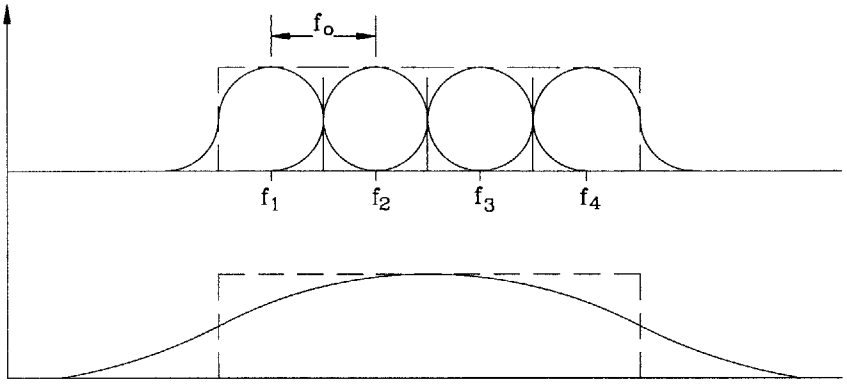



**Figure** 5.8: Bandwidth of single- and multi-carrier modulated raised cosine pulses

Compared to Eqns. (5.30)–(5.32), the above pulse is equivalent to a fully raised-cosine function except for a frequency translation by $f_i$. It is no surprise that an example of the squared amplitude of a carrier modulated g-Nyquist pulse is

$$|R_i(f)|^2 = C_i + Q_i(f) = \cos^2\left(\pi\frac{f - f_i}{2f_o}\right) \tag{5.42}$$

which is a frequency translated, fully raised-cosine pulse [cf. eqn. (5.34)]. If there are N carriers the total bandwidth is $(N + 1)f_o$. The symbol rate per channel is

$$R_s = \left(\frac{N}{N+1}\right) f_o \tag{5.43}$$

where $1/N$ is the roll-off factor of multi-carrier modulated signal. Compared to a single-carrier raised cosine pulse with $r > 1/N$, the multi-carrier modulation format is more bandwidth efficient. In summary, the overlapping concatenation of N frequency-shifted fully-raised cosine waveforms has a higher bandwidth efficiency than a single-carrier modulated raised cosine waveform with roll-off factor $r > 1/N$. Furthermore, because each sub-carrier in a multi-carrier modulated waveform has full roll-off factor ($r = 1$) its decay is smoother. The bandwidths of both single- and multi-carrier modulated waveforms are shown in Fig. 5.8.

## 5.4 Perfect Reconstruction Criterion

In the last section, it is shown that the bandwidth of a multi-carrier modulated signal is smaller that of a single-carrier modulated signal given that both achieve the same symbol rate

with zero ISI. However, in both cases of sinc-type and Chang pulses, the pulse length is infinite. If it is truncated, the zero ISI condition no longer holds. The important question is: "Can we design pulses that are g-Nyquist with finite length?" The answer is two-fold:

- If the pulse duration is symbol-time limited, there exists no g-Nyquist pulses regardless of pulse shaping —not counting the rectangular pulse.
- If the pulse duration is longer than $T_s$ but still finite, then there exists a class of time-overlapped functions that satisfy the g-Nyquist criterion.

Such time overlapped pulses have recently been discovered by H. Malvar. His research is concerned with filter bank analysis-synthesis techniques that are alias- and distortion-free. In signal processing applications such as speech coding and image compression, a signal is analyzed by a bank of bandpass filters, then sub-sampled, then over-sampled, and finally regenerated by a bank of synthesis filters. (See Fig. 5.22 for a block diagram illustration.) The condition where the original signal is recovered without any loss in information is called *Perfect Reconstruction* criterion, or simply PR condition. In this section we describe the relation between PR and g-Nyquist criteria. This bridge will aid us in the design of finite-length, time-overlapped g-Nyquist pulses. First, we review basic concepts in sampling.

### 5.4.1 Sampling Theorem Revisited

The sampling theorem for deterministic signals states that a baseband signal $g(t)$ that is strictly bandlimited to $W$ Hz —i.e., its Fourier transform $G(f)$ is zero for all frequencies $|f| > W$— is uniquely determined by its uniformly sampled values $\{g[nT_s]\}$ if $T_s \leq \frac{1}{2W}$. If $g(t)$ is a periodic power signal generated by concatenating symbol-time-limited pulses $g_w(t)$, the sampling theorem is still applicable by replacing the Fourier transform $G(f)$ with the power spectral density $|G_w(f)|^2/T$. For a random process its sample function is a power signal and its Fourier transform may not exist. In this case, the spectral content is measured in terms of its power spectral density. A random process is said to be *bandlimited* to $W$ Hz if its power spectral density (if it exists) $S_g(f) = 0$ for $|f| > W$. Similar to the Dirichlet's conditions for the deterministic signal, the PSD only exists for certain random processes. In fact, it is necessary that the process is wide-sense stationary. This is the well-known *Weiner-Khinchine* relation [68]. However, for a cyclo-stationary random process its average PSD can be computed. We can now restate the sampling theorem for random processes:

**Theorem 5.1 (Sampling Theorem)** *A baseband random process strictly bandlimited to W Hz if its (average) power spectral density is zero for all frequencies $|f| > W$. A sample function $g(t)$ is uniquely determined by its uniformly sampled values $\{g[nT_s]\}$ if $T_s \leq \frac{1}{2W}$.*

The minimum sampling rate $f_o = 2W$ is the *Nyquist rate*. By using time-shifted sinc functions as orthogonal basis functions, $g(t)$ can be synthesized as follows:

$$g(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\,\mathrm{sinc}\Big[2W(t-nT_s)\Big] \tag{5.44}$$

For a linearly modulated signal

$$g(t) = \sum_{i=-\infty}^{\infty} x[i]f(t-iT)$$

Its uniform samples are

$$g(nT_s) = \sum_{i=-\infty}^{\infty} x[i]f(nT_s-iT)$$

Substituting in eqn.(5.44):

$$g(t) = \sum_{i=-\infty}^{\infty} x[i] \sum_{n=-\infty}^{\infty} f(nT_s-iT)\,\mathrm{sinc}\Big[2W(t-nT_s)\Big]$$

At Nyquist rate $T_s = 1/2W$

$$= \sum_{i=-\infty}^{\infty} x[i] \sum_{n=-\infty}^{\infty} f\Big(\frac{n}{2W}-iT\Big)\,\mathrm{sinc}(2Wt-n)$$

We know from eqn. (4.74) that the time-shifted sinc functions $\mathrm{sinc}(2Wt-n)$ and $\mathrm{sinc}(2Wt-n')$ are orthogonal if $(n-n')$ is a non-zero integer. Furthermore, since the common pulse $\psi(t)$ has duration T, its $2WT$ samples — as $n$ varies from 0 to $2WT-1$ for $i=0$— are sufficient for synthesis. Recall that the *dimension* $D = 2WT$. This implies that instead of transmitting a pulse $x[i]f(t)$ every T sec., we can send a discrete sequence of length $2WT$:

$$x[i]\cdot\Big[f(0)\,f(\frac{1}{2W})\,f(\frac{2}{2W})\,\ldots\,f\Big(\frac{2WT-2}{2W}\Big)\,f\Big(\frac{2WT-1}{2W}\Big)\Big]$$

every T sec. To summarize, we review the sampling theorem to emphasize the fact that there is an equivalent representation of the continuous-time filter output in terms of its discrete-time uniform samples. Hence some insights can be gained in the construction of orthogonal g-Nyquist continuous-time waveforms by studying their corresponding discrete-time sequences. We do so in the next section.

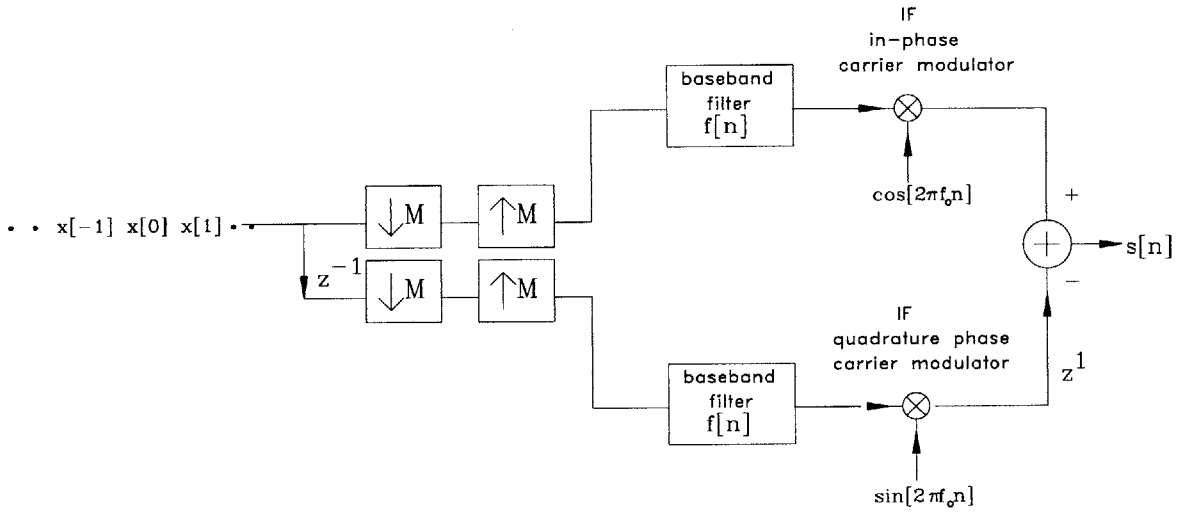



**Figure 5.9:** Discrete-time I/Q data modulator of complex-valued code symbols $x[j] = (x_I[j], x_Q[j])$ with pulse shaping FIR filter $f[n]$

### 5.4.2 Equivalent Discrete-time Representations

The continuous-time I/Q modulator of Fig. 3.15 can be replaced with its discrete-time version as shown in Fig. 5.9. All discrete-time functional blocks such as delay tap, decimator, interpolator, circular convolution are thoroughly discussed in digital signal processing texts; For continuity, we briefly review their application in multi-carrier modulated communication system framework.

- *Serial-to-Parallel Conversion:* In Fig. 5.9, the input stream of code symbols is split into odd and even symbols for in-phase and quadrature channel modulation. For multi-carrier modulation with M channels, a 1-to-M S/P converter must be devised. In discrete-time, a cascade of delay taps (denoted by a unit of Z-inverse $z^{-1}$) and M-*down-samplers* fulfill this function. This is illustrated in Fig. 5.10.

  If $x[n]$ is the input to a unit-delay tap, its output is

  $$x^{(1)}[n] = z^{-1}\{x[n]\} = x[n-1]$$

  In general, the $k^{th}$ output of a delay chain is

  $$x^{(k)}[n] = z^{-k}\{x[n]\} = x[n-k]$$

  If $x[n]$ is the input to an M-downsampler or decimator, the output is

  $$y^{(0)}[n] = \downarrow M\{x[n]\} = x[nM]$$

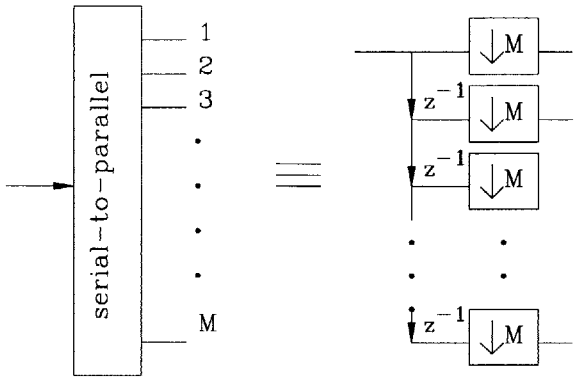



**Figure** 5.10: Discrete-time serial-parallel converter

When cascaded, the $k^{th}$ output of each delay-decimator tap is

$$y^{(k)}[n] = \downarrow M\{x[n-k]\} = x[nM-k]$$

Figure 5.11 illustrates serial-to-parallel conversion of a data stream for $M = 4$ parallel channels.

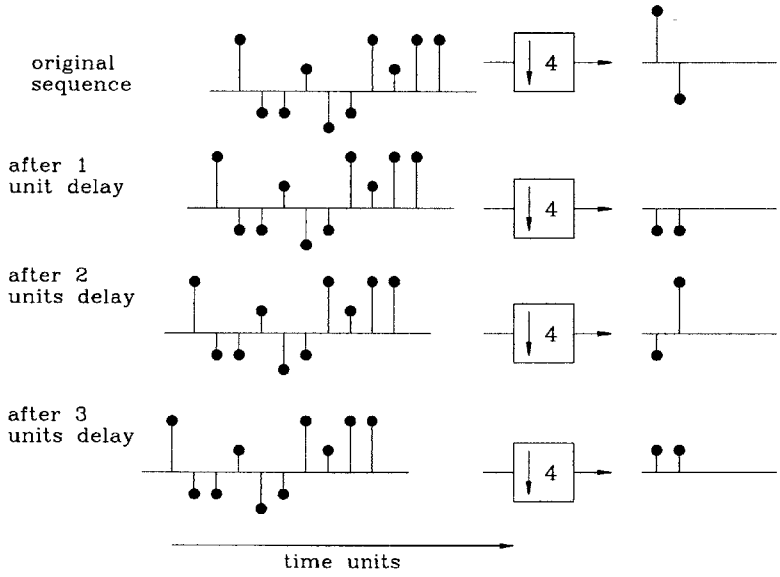



**Figure** 5.11: Cascade of delay and downsampling for serial-to-parallel conversion

- *Delta Modulation:* S/P conversion results in M parallel streams of phase-shifted code symbols. However, the time interval of a block of M code symbols must be the same

before and after S/P conversion. That is, the symbol rate of each output channel must be reduced by m; equivalently, zero symbols must be inserted between two consecutive code symbol outputs in each channel. This function is achieved by M-*up-sampling* or *interpolation*. This is illustrated in Figure 5.12.
If x[n] is the input to an interpolater, its output is

$$y[n] = \uparrow M\{x[n]\} \begin{cases} x[n/M] & n \bmod M = 0 \\ 0 & \text{otherwise} \end{cases}$$

A cascade of M down- and up-sampling (see Fig. 5.13) is equivalent to modulation by a periodic train of impulses of period M:

Cascaded output:

$$y[n] = (\uparrow M)(\downarrow M)\{x[n]\} = x[n] \cdot \delta_M[n]$$

where

$$\delta_M[n] = \sum_{r=-\infty}^{\infty} \delta[n - rM] = \begin{cases} 1 & n \bmod M = 0 \\ 0 & \text{otherwise} \end{cases}$$

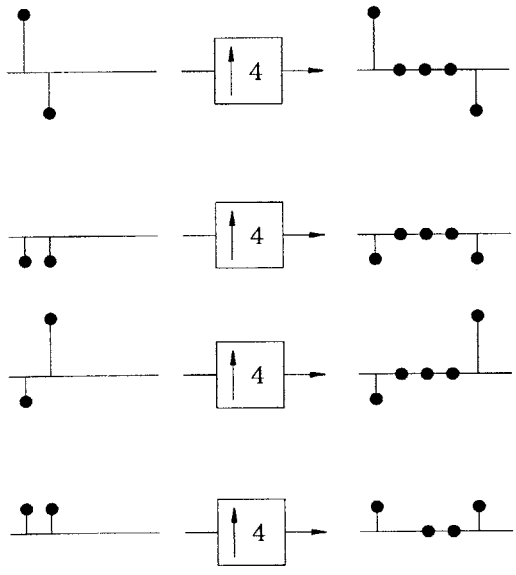



**Figure** 5.12: S/P conversion with upsampling for input-output rate matching

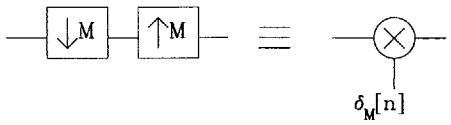



**Figure** 5.13: Discrete-time delta modulator

- *Transmit Pulse Shaping:* For convenience, we relabel the S/P converted output in $k^{th}$ channel as:

$$y_m[n] = (\uparrow M)(\downarrow M)z^{-m}\{x[n]\} = \begin{cases} x[n-m] & (n-m) \bmod M = 0 \\ 0 & \text{otherwise} \end{cases}$$

The transmit filter output is

$$g_m[n] = \sum_{k=0}^{N_w-1} f_m[k]\, y_m[n-k]$$

where $h_m[n]$ is the impulse response of a LTI causal filter of length (window) $N_w$. A block diagram of a generic multi-channel discrete-time modulator is shown in Fig. 5.14. For a special class of multi-carrier modulation format, each passband filter is equivalent to a cascade of low-pass filter and frequency translator (i.e., carrier modulator). This is illustrated in Fig. 5.15.

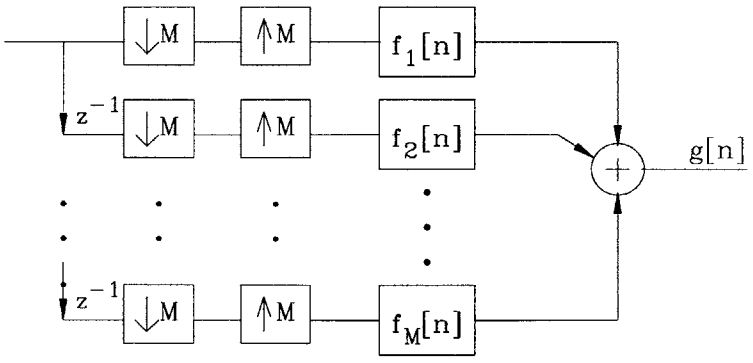



**Figure** 5.14: A block diagram of a generic discrete-time multi-channel modulator

- *Complex Carrier Modulation:* In the continuous case, the filter output $g_s(t)$ is frequency

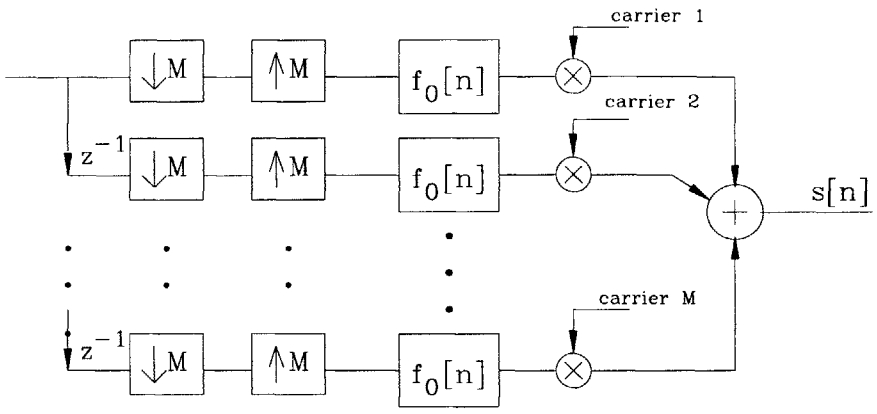



**Figure** 5.15: M-channel carrier-modulated communication system: transmitter

translated by an in-phase (cosine) or quadrature (sine) IF carrier time functions such that

$$s_I(t) = g_I(t)\cos(2\pi f_o t)$$
$$s_Q(t) = g_Q(t)\sin(2\pi f_o t)$$

From eqn. (3.5) it is understood that both in-phase and quadrature channels utilize the same baseband pulse-shaping filter $f(t)$:

$$g_s(t) = \sum_j f(t - jT_s)\, x_s[j] \qquad s = I, Q$$

An equivalent discrete-time IF carrier modulation in the $k^{th}$ channel can be expressed as:

$$s_{kI}[n] = g_{kI}[n]\cos\left[\frac{2\pi kn}{M}\right]$$
$$s_{kQ}[n] = g_{kQ}[n]\sin\left[\frac{2\pi kn}{M}\right]$$

where

$$g_{ks}[n] = \sum_j f(t - jT_s)\, x_{ks}[j] \qquad s = I, Q$$

An alternative description of discrete-time carrier modulation uses discrete-time Fourier carriers

$$W_M^k[n] = \exp\left[j\frac{2\pi kn}{M}\right]$$

The respective inverse carriers are

$$W_M^{-k}[n] = \exp\left[-j\frac{2\pi kn}{M}\right]$$

The resulting Fourier-carrier modulated signal is

$$s_k[n] = \left(x_k[n] * f[n]\right) W_M^k[n]$$

Since transmitted signals must be real, it is required in each block of M code symbols

$$x_k[n] = x^*_{M-k}[n] \tag{5.45}$$

Note that for M even $x_0[n] = x^*_M[n]$ and $x_{M/2}[n] = x^*_{M/2}[n]$, implying that the code symbols in the first and last channels must be real. (There are only M/2 distinct carriers.) The final transmit signal is

$$s[n] = \sum_{k=0}^{M-1} s_k[n]$$

Expanding,

$$= \sum_{k=0}^{M-1} \left(x_k[n] * f[n]\right) e^{j2\pi kn/M}$$

$$= \sum_{k=0}^{\frac{M}{2}-1} \left(x_k[n] * f[n]\right) e^{j2\pi kn/M} + \sum_{k=\frac{M}{2}}^{M-1} \left(x_k[n] * f[n]\right) e^{j2\pi kn/M}$$

$$= \sum_{k=0}^{\frac{M}{2}-1} \left(x_k[n] * f[n]\right) e^{j2\pi kn/M} + \sum_{k=1}^{\frac{M}{2}} \left(x_{M-k}[n] * f[n]\right) e^{j2\pi n\left(\frac{M-k}{M}\right)}$$

Using the relation of eqn. (5.45):

$$= \left\{ x_0[n] + x_{M/2}[n] e^{j\pi n} + \sum_{k=1}^{\frac{M}{2}-1} \left( x_k[n]\, e^{j2\pi kn/M} + x_k^*[n]\, e^{-j2\pi kn/M} \right) \right\} * f[n]$$

Further manipulation leads to

$$= \sum_{k=0}^{\frac{M}{2}-1} \left\{ x_{ke}[n] \cos\left(\frac{2\pi kn}{M}\right) - x_{ko}[n] \sin\left(\frac{2\pi kn}{M}\right) \right\} * 2\, f[n]$$

where the even and odd components of the *complex* code symbols

$$x_k[n] = x_{ke}[n] + j x_{ko}[n]$$

are given by

$$x_{ke}[n] = \frac{x_k[n] + x_k^*[n]}{2}$$

$$x_{ko}[n] = \frac{x_k[n] - x_k^*[n]}{2j}$$

In summary, the Fourier carrier modulated signal is equivalent to I/Q modulation with the constraint of eqn. (5.45)

$$s[n] = \sum_{k=0}^{M-1} \left(x_k[n] * f[n]\right) \exp\left(\frac{j2\pi kn}{M}\right)$$
$$= \sum_{k=0}^{\frac{M}{2}-1} \left\{ x_{ke}[n] \cos\left(\frac{2\pi kn}{M}\right) - x_{ko}[n] \sin\left(\frac{2\pi kn}{M}\right) \right\} * 2f[n] \qquad (5.46)$$

- *Real Carrier Modulation:* As described above, multiplication by inverse discrete Fourier transform is equivalent to complex I/Q data modulation. As an alternative, real modulation uses in-phase cosine carriers only. The transmit signal with cosine modulation is

$$s[n] = \sum_{k=0}^{M-1} \left(x_k[n] * f[n]\right) \cos\left(\frac{\pi}{M} kn\right) \qquad (5.47)$$

Note that two cosine carriers are orthogonal when their separation is only $\pi/M$, compared to Fourier carriers where the minimum separation for orthogonality is $2\pi/M$. Both modulation schemes are equivalent in the sense that the number of transmitted code symbols per interval $T_s$ is the same for a fixed bandwidth: the cosine carrier modulated signal has twice the number of carriers, each modulated by a real code symbol, whereas the Fourier modulated signal has half the number of carriers, each with in-phase and quadrature components carrying even and odd real code symbols, respectively. In the above cosine modulation format, the first channel has zero frequency. These functions

$$E_k^I[n] = \cos\left(\frac{\pi kn}{M}\right) \qquad (5.48)$$

are known as *Type-I discrete cosine transforms* (DCT-I). Other variations (types II, III and IV) are envisaged by offsetting the first carrier in time or frequency or both:

$$E_k^{II}[n] = \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)n\right] \qquad (5.49)$$

$$E_k^{III}[n] = \cos\left[\frac{\pi}{M}k\left(n+\frac{1}{2}\right)\right] \qquad (5.50)$$

$$E_k^{IV}[n] = \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}\right)\right] \qquad (5.51)$$

It is straightforward to show that for each type, the set of cosine functions are mutually orthogonal. The first four cosine basis functions of each type are plotted in Figs. 5.16 and 5.17.

Referring to eqns. (5.41) and (5.42), we comment that Chang pulses are continuous-time equivalent of Type II or IV cosine functions.

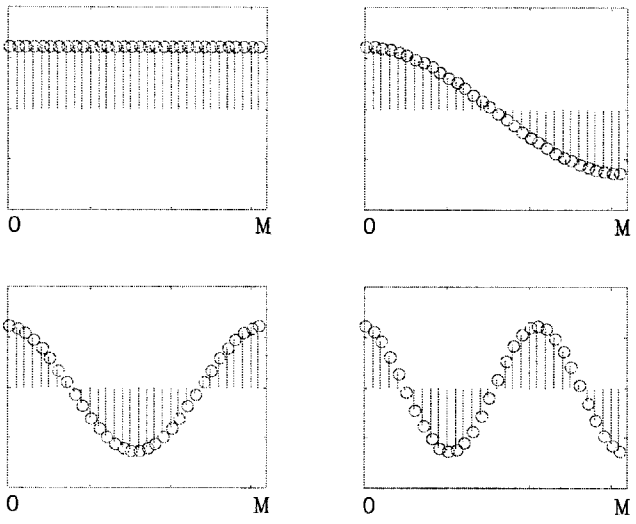



**Figure** 5.16: First four basis functions ($k = 0, 1, 2, 3$) of type I DCT. Due to zero offsets in time and frequency, it is the only DCT with odd or even symmetry about the mid-point.

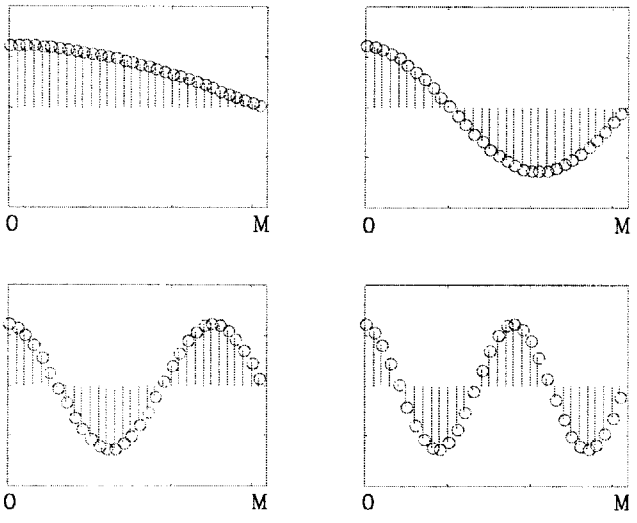



**Figure** 5.17: First four basis functions of type IV DCT. Unlike types I and III, there is no non-oscillating (DC) basis in type II and IV DCTs.

### 5.4.3 Discrete-time Generalized Nyquist Criterion

A block diagram of a transmitter of a general multi-channel communication system is depicted in Fig. 5.14. This is the discrete-time model where the input and output of every functional block are discrete-time sequences. Before transmission, it is understood that the transmit sequence g[n] must be D/A converted. For the special case of multi-carrier modulation as depicted in Fig. 5.15, a single prototype lowpass filter $f_0[n]$ suffices when combined with various IF carrier modulators. At the receiving end, the demodulation and detection blocks are mirror images of the transmitter. This is depicted in Fig. 5.18. (For convenience we omit the subscript "0" when denoting prototype filters.)

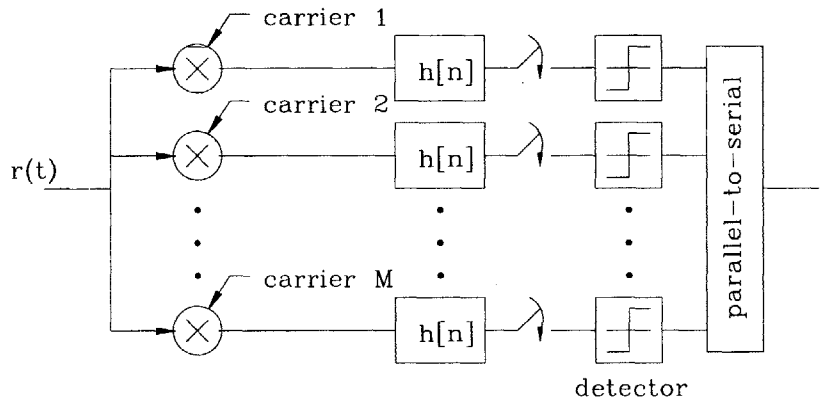



**Figure** 5.18: M-channel carrier-modulated communication system: receiver

Historically, communication systems are studied in the continuous-time domain. Therefore we are interested in designing transmit and receiver filter with impulse responses h(t) and f(t) such that the transmission rate is maximized and the receive filter outputs in each channel are ISI-free. Due to Shannon's capacity theorem, achievable information rate is a function of the received signal-to-noise ratio. In AWGN channel it is known that the SNR of the receive filter is maximized when the transmit and receiver filters f(t) and h(t) are a matched pair; i.e., F(f) is a complex-conjugate of H(f). This is the matched filter realization. The ISI-free condition is the generalized Nyquist criterion. In practice, it may not be possible to design filter responses that meet both criteria.

First, we discuss the Nyquist criterion in the discrete-time domain. Matched filter realization will be dealt with once we introduce time-overlapped pulses. We already know from Sec.5.3.3 that sinc-type pulses (whose Fourier transforms resemble Walsh codes) are g-Nyquist. If excess bandwidth is allowed, Chang pulses are g-Nyquist with each data channel containing a cosine carrier modulated signal. The drawback is both pulses are infinite in length, thus requiring IIR filters at both the transmitting and receiving sides. Any truncation or shortening of filter lengths

leads to non-zero ISI. Analogous to the conventional continuous-time Nyquist condition, we can derive the g-Nyquist criterion for discrete-time sequences. We detail step-by-step, starting with the input stream of code symbols $x[n]$:

- *Delayed Decimation:*

$$\begin{aligned} y_0[n] &= x[nM] \\ y_1[n] &= x[nM-1] \\ &\vdots \\ y_k[n] &= x[nM-k] \\ &\vdots \\ y_{M-1}[n] &= x[nM-M+1] \end{aligned}$$

- *After interpolation:*

$$\begin{aligned} g_0[n] &= \sum_{r=-\infty}^{\infty} x[n]\,\delta[n-rM] \\ g_1[n] &= \sum_{r=-\infty}^{\infty} x[n]\,\delta[n-1-rM] \\ &\vdots \\ g_k[n] &= \sum_{r=-\infty}^{\infty} x[n]\,\delta[n-k-rM] \\ &\vdots \\ g_{M-1}[n] &= \sum_{r=-\infty}^{\infty} x[n]\,\delta[n-M+1-rM] \end{aligned}$$

- *Transmit Filter Output:*

$$\begin{aligned} s_0[n] &= \sum_{l=-\infty}^{\infty} f_0[l]\,g_0[n-l] \\ &= \sum_{l=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} f_0[l]x[n-l]\,\delta[n-l-rM] \\ &= \sum_{r=-\infty}^{\infty} f_0[n-rM]x[rM] \end{aligned}$$

$$
\begin{aligned}
s_1[n] &= \sum_{r=-\infty}^{\infty} f_1[n-1-rM]x[rM+1] \\
&\vdots \\
s_k[n] &= \sum_{r=-\infty}^{\infty} f_k[n-k-rM]x[rM+k] \\
&\vdots \\
s_{M-1}[n] &= \sum_{r=-\infty}^{\infty} f_{M-1}[n-M+1-rM]x[rM+M-1]
\end{aligned}
$$

- *Transmitter Output:*

$$
\begin{aligned}
s[n] &= \sum_{k=0}^{M-1} s_k[n+k] \\
&= \sum_{k=0}^{M-1} \sum_{r=-\infty}^{\infty} f_k[n-rM]x[rM+k]
\end{aligned}
$$

For consistency, each elemental output $s_k[n]$ is time-advanced by $k$ units since it is delayed by the same amount during decimation. Assuming the channel is ideal with no delay: $c[n] = \delta[n]$, the received signal $r[n] = s[n]$.

- *Receive Filter Output:*

$$
\begin{aligned}
v_0[n] &= \sum_{l=-\infty}^{\infty} h_0[l]\, s[n-l] \\
&= \sum_{k=0}^{M-1} \sum_{r=-\infty}^{\infty} x[rM+k] \sum_{l=-\infty}^{\infty} h_0[l]\, f_k[n-l-rM] \\
v_1[n] &= \sum_{k=0}^{M-1} \sum_{r=-\infty}^{\infty} x[rM+k] \sum_{l=-\infty}^{\infty} h_1[l]\, f_k[n-l-rM] \\
&\vdots \\
v_{M-1}[n] &= \sum_{k=0}^{M-1} \sum_{r=-\infty}^{\infty} x[rM+k] \sum_{l=-\infty}^{\infty} h_{M-1}[l]\, f_k[n-l-rM]
\end{aligned}
$$

By separating code symbols in the zeroth block of the first channel:

$$x[0]\sum_{l=-\infty}^{\infty} h_0[n-l]f_0[l] + \sum_{r\neq 0} x[rM]\sum_{l=-\infty}^{\infty} h_0[n-l]\,f_0[l-rM]$$
$$+\sum_{k=1}^{M-1} x[k]\sum_{l=-\infty}^{\infty} h_0[n-l]\,f_k[l]$$
$$+\sum_{k=1}^{M-1}\sum_{r\neq 0} x[rM+k]\sum_{l=-\infty}^{\infty} h_0[n-l]\,f_k[l-rM] \tag{5.52}$$

The first term is the desired symbol. The second term is ISI from past and future code symbols in the same channel. The third term is ICI due to current symbols from other parallel channels. The fourth term is IBI due to past and future code symbols —i.e., symbols in another block— from other parallel channels. If the receive filter in each channel matches the corresponding transmit filter, and if $f_k[n]$ is an FIR filter of length $N_w$:

$$h_k[n] = f_k^*[N_w - 1 - n]$$

Since the filter tap weights are real,

$$h_k[n] = f_k[N_w - 1 - n] \tag{5.53}$$

Note that the receive matched filter $h(t)$ is also FIR. If the transmit pulses are orthogonal,

$$\sum_{l=0}^{N_w-1} f_k[l]\,f_n[l-rM] = \delta[k-n]\,\delta[r] \tag{5.54}$$

then the system meets the generalized Nyquist criterion. This is the discrete-time equivalent of eqn. (5.39).

### 5.4.4 Multi-Rate Filter Bank Design

We next describe the close relationship between the g-Nyquist criterion and Perfect Reconstruction Criterion. A block diagram of a multi-channel filter bank is depicted in Fig. 5.19. We briefly review the main properties of multi-channel filter bank. It is essentially a reverse process of a multi-channel communication system. As the name suggests, the system consists of a bank of M *analysis* and *synthesis* filters. An input signal is *analyzed* by passband filters $h_k(t)$ at center frequencies $f_k$ and bandwidth $B_k$. This is called *sub-band processing*. We are only interested in a class of uniform-bandwidth, critically sampled and carrier-modulated filter bank:

- *Uniform Bandwidth:* The bandwidths of all M bandpass filters are equal.
  For real modulation with cosine carriers,

$$B_k = \frac{\pi}{M} + \varepsilon$$

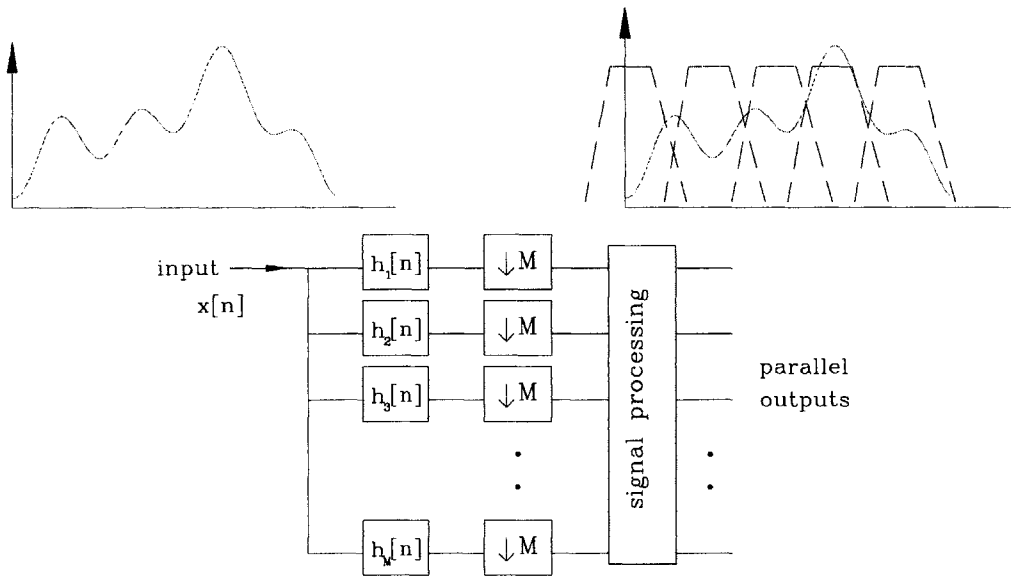



**Figure** 5.19: Sub-band processing with M-channel filter bank: analysis

where $\varepsilon$ is the excess bandwidth.
For complex modulation with Fourier carriers,

$$B_k = \frac{2\pi}{M} + \varepsilon$$

With real modulation the carrier spacing is approximately $\pi/M$. Over the entire frequency range $[0, 2\pi]$, $2M$ filters are required. With complex modulation, the carrier spacing is $2\pi/M$, resulting in $M$ bandpass filters.

- *Center Frequencies:* If the zeroth channel has a lowpass filter such that the center frequencies of the carriers are

$$\begin{aligned} f_k &= \frac{2\pi}{M}k \qquad k \in \mathbb{Z}^{M-1} \qquad \text{complex modulation} \\ f_k &= \frac{\pi}{M}k \qquad k \in \mathbb{Z}^{2M-1} \qquad \text{real modulation} \end{aligned} \tag{5.55}$$

This is known as *even stacking*. In *odd stacking*, every filter is bandpass with center frequencies

$$\begin{aligned} f_k &= \frac{2\pi}{M}\left(k+\frac{1}{2}\right) \qquad k \in \mathbb{Z}^{M-1} \qquad \text{complex modulation} \\ f_k &= \frac{\pi}{M}\left(k+\frac{1}{2}\right) \qquad k \in \mathbb{Z}^{2M-1} \qquad \text{real modulation} \end{aligned} \tag{5.56}$$

Both odd- and even-stacked frequency bands are illustrated in Fig. 5.20. It is important to note that for even stacking the center frequencies of all bands are non-zero; i.e., each

band is a frequency translation (carrier modulation) of a baseband signal. This difference between odd and even stacking becomes critical when the carriers themselves are complex-valued, as we shall see in Sec. 5.4.7.

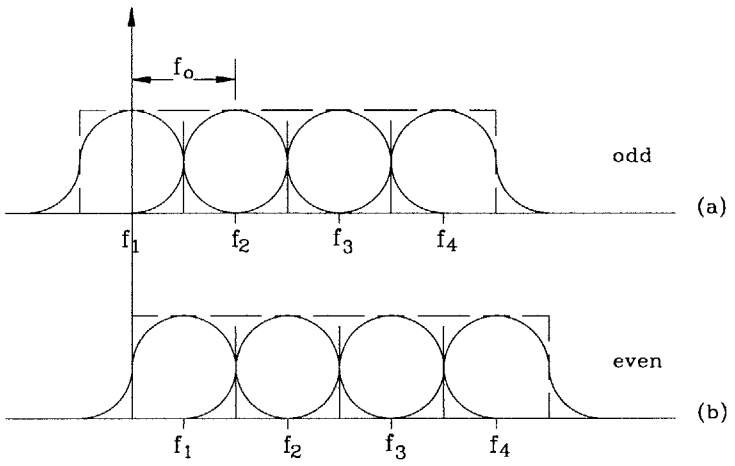



**Figure** 5.20: (a) odd and (b) even stacking of bandpass filters

- *Critical Sampling:* The filter output in each channel of an M-channel filter bank is critically sampled if an M-down sampler follows each filter. Decimation does not discard useful information; it only removes redundant information growth. It is well known that by proper design of a bank of analysis and synthesis filters, the original sequence $x[n]$ can be reconstructed from downsampled sequences. This is the Perfect Reconstruction condition.

- *Carrier Modulation:* All passband filters are derived from a single *prototype* filter $h[n]$. For $k \geq 1$

$$\begin{aligned} h_k[n] &= h[n] \cos\left[\frac{\pi n}{M}\left(k+l\right)\right] && \text{real modulation} \\ h_k[n] &= h[n] \exp\left[j\frac{2\pi n}{M}\left(k+l\right)\right] && \text{complex modulation} \end{aligned} \tag{5.57}$$

  where $l = 0$ for even stacking and 0.5 for odd stacking.

In general, the sub-sampled filter outputs are processed (e.g., compressed) for transmission or storage. At the other end, the original signal $x[n]$ is reconstructed or *synthesized* from stored or received samples in M sub-carriers. The synthesis operation consists of up-sampling followed by synthesis filtering with impulse responses $f_k[n]$, and finally the summation of sub-samples to reconstruct (a possibly) aliased and distorted version $\hat{x}[n]$ of the original signal. Comparing a multi-channel filter bank in Fig. 5.19 with a multi-channel communication system in Fig.

5.15, we notice that both systems are identical with the roles of analysis and synthesis filters reversed. We next follow the functional blocks of a carrier-modulated filter bank step-by-step:

- *Analysis Filter Output:*

$$y_k[n] = x[n] \star h_k[n] = \sum_r x[r]\, h_k[n-r] \tag{5.58}$$

- *After decimation:*

$$s_k[n] = y_k[nM] = \sum_r x[r]\, h_k[nM-r] \tag{5.59}$$

Assume there is no further processing such that $s_k[n] = r_k[n]$ for all k.

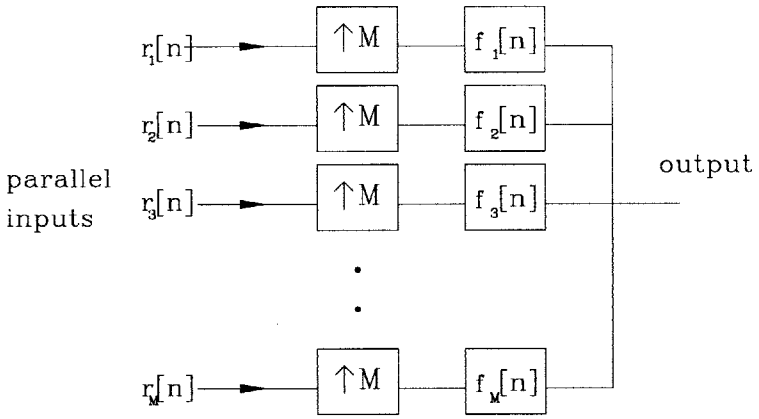



**Figure** 5.21: Sub-band processing with M-channel filter bank: synthesis

- *After interpolation:*

$$v_k[n] = \begin{cases} \sum_r x[r]\, h_k[n-r] & n \bmod M = 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.60}$$

- *Synthesis Filter Output:*

$$\begin{aligned} \hat{x}_k[n] &= \sum_l v_k[l]\, f_k[n-l] \\ &= \sum_l v_k[lM]\, f_k[n-lM] \\ &= \sum_l \sum_r x[r]\, h_k[lM-r]\, f_k[n-lM] \end{aligned} \tag{5.61}$$

■ *Reconstructed Signal:*

$$\hat{x}_n = \sum_{k=0}^{M-1} \sum_{l} \sum_{r} x[r]\, h_k[lM - r]\, f_k[n - lM] \tag{5.62}$$

The original signal $x[n]$ is recovered if there is no distortion or aliasing. This is known as the *Perfect Reconstruction Criterion*:

$$\sum_{k=0}^{M-1} \sum_{l} h_k[lM - r]\, f_k[n - lM] = \delta[n - r - D] \tag{5.63}$$

where D is an arbitrary time delay.

Substituting the PR condition of eqn. (5.63) in eqn. (5.62)

$$\hat{x}_n = x[n - D]$$

### 5.4.5 Key Observations

When the analysis and synthesis blocks of a critically decimated, uniformly modulated M channel filter bank are put side by side, we observe the complete system as shown in Fig. 5.22.

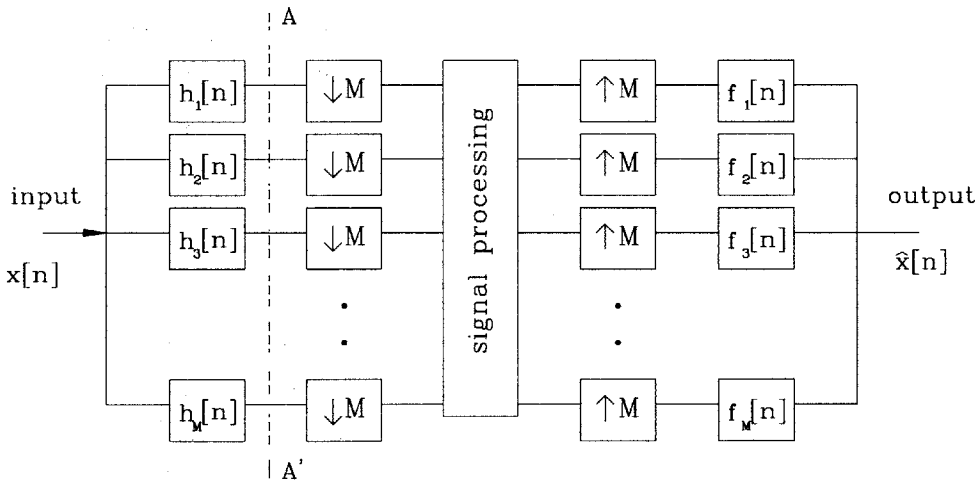



**Figure** 5.22: Standard form of an M-channel filter bank system

Without any signal processing of the channel sequences, it is possible to design a bank of analysis and synthesis filters such that the output $\hat{x}[n]$ is a delayed copy of the input $x[n]$; i.e., $\hat{x}[n] = x[n - D]$ for some arbitrary delay D. We know this condition as Perfect Reconstruction. Let's assume this is indeed the case; furthermore, we reconfigure the filter bank system along the dashed line $A - A'$ such that the analysis filter outputs are treated as parallel inputs to this

new system. This new configuration is illustrated in Fig. 5.23. It is assumed that the "channel" (the link between x[n] and $\hat{x}[n]$) is an ideal delay line. For the special case as depicted in Fig. 5.24 where bandpass filters are frequency translations of the lowpass prototype filter, the resulting reconfigured system is very similar to the multi-carrier communication system of Figs. 5.15 and 5.18, except for minor details which we discuss next.

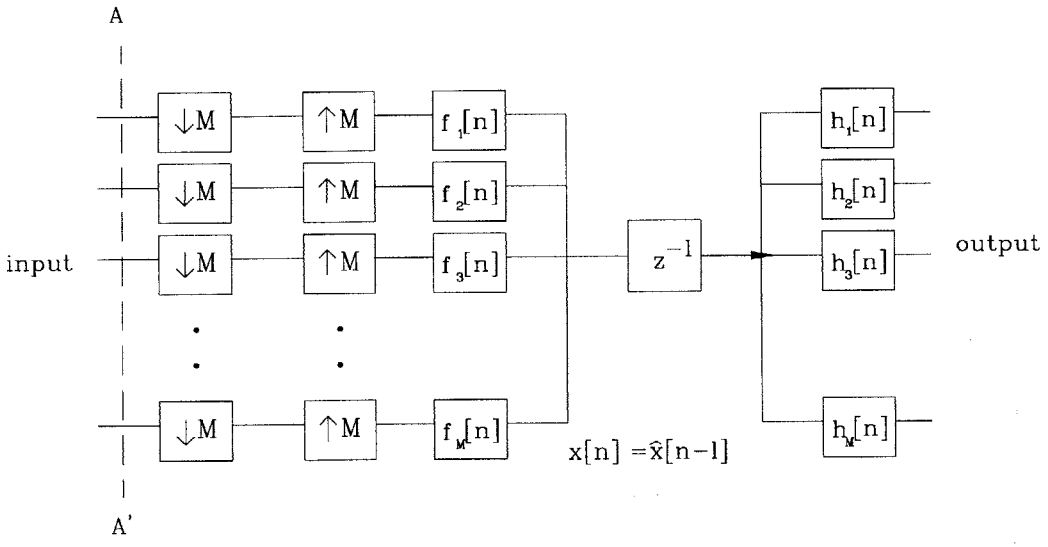



**Figure** 5.23: Reconfigured form of an M-channel filter bank system

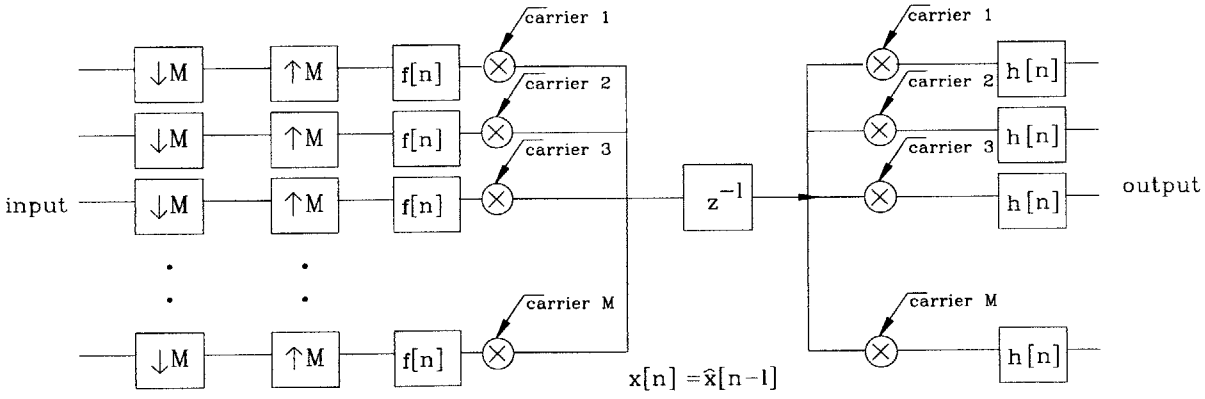



**Figure** 5.24: Reconfigured form of an M carrier modulated filter bank system

### Parallel-to-Serial Conversion

In a communication receiver the first stage is the carrier demodulator. Next, the baseband signal is matched filtered, sampled and its output is quantized. (A quantizer is also called a

hard limiter.) These steps are illustrated in Fig. 5.18. As a final step, the decoded symbols must be parallel-to-serial converted. The P/S conversion process is similar to S/P conversion with the interpolator replacing the decimator, and the order of delay and sampling reversed. This is depicted in Fig. 5.25.

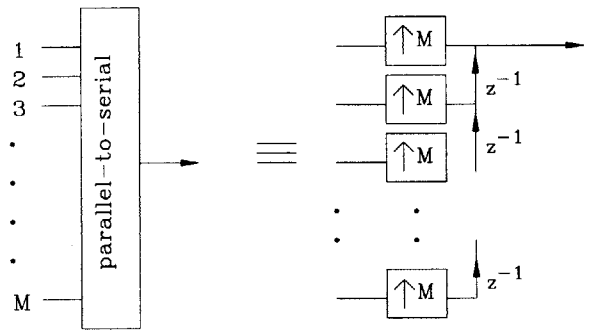



**Figure** 5.25: Parallel-to-Serial conversion in discrete time

By combining the parallel-to-serial converter with the M-channel filter bank receiver, and the inclusion of a delay chain at the input side for serial-to-parallel conversion, we can finally construct the complete multi-carrier communication system depicted in Fig. 5.26.

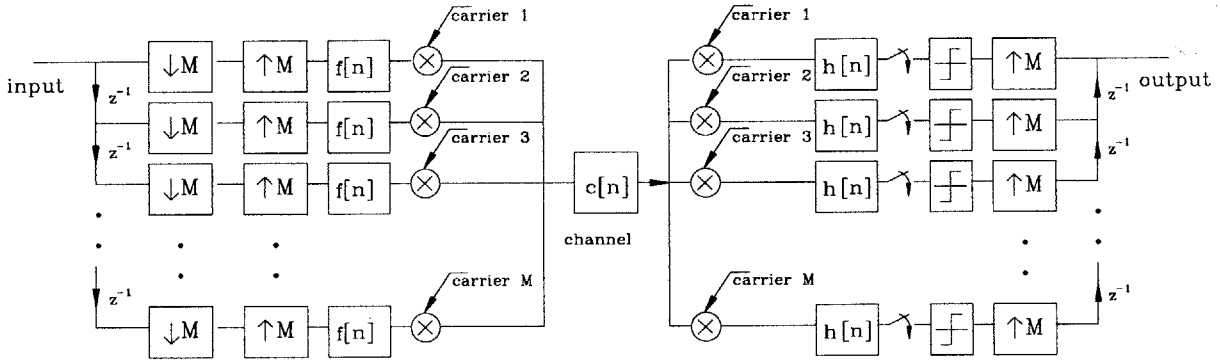



**Figure** 5.26: Detailed description of multi-carrier communication system

Note that the cascade of a bank of samplers, hard limiters, interpolators and two delay chains (as depicted in Fig. 5.27) is transparent to the overall operation of the complete system. The proof is straightforward:
Let the output samples of the first analysis filter $h_1[n]$ be

$$\ldots a_1 \ldots a_2 \ldots a_3 \ldots \tag{5.64}$$

such that there are $(M-1)$ irrelevant samples exist between $a_i$ and $a_{i+1}$ for some i. Similarly,

the output samples of the second analysis filter $h_2[n]$ can b expressed as

$$\ldots b_1 \ldots b_2 \ldots b_3 \ldots \tag{5.65}$$

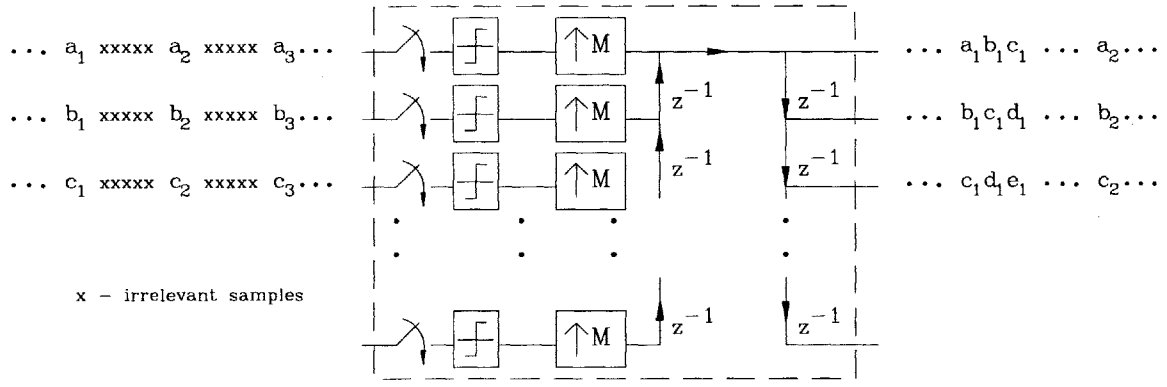



**Figure** 5.27: Transparency of the cascaded block of expander and delay chains

Here it is also understood that $(M-1)$ samples exist between $b_i$ and $b_{i+1}$ for some i. Other filter outputs are also labelled in a similar fashion. In the theory of matched filtering, only the output sample where the signal-to-noise ratio is maximum is significant. The sampler is synchronized such that its periodic samples occur at time instants where the SNR is maximum. Using this information, the outputs of parallel samplers are

$$\ldots a_1\, a_2\, a_3 \ldots$$

$$\ldots b_1\, b_2\, b_3 \ldots$$

and so on. Of course, the samplers perform the same function as decimators. Next these samples are quantized to one of the possible elements of the discrete signal constellation set. For the binary antipodal set, there are only two elements: $+1$ and $-1$. In this case, the quantizer is a zero threshold detector; i.e., any value greater than zero is quantized to $+1$ and other values less than or equal to zero are decoded as $-1$. As the final step, the decoded symbols must be parallel-to-serial converted before any other processing takes place. The output sequence of the first interpolator is

$$\ldots 00\, a_1\, 00 \ldots 00\, a_2\, 00 \ldots 00\, a_3 00 \ldots$$

Likewise, the output sequence of the second interpolator can be expressed as

$$\ldots 00\, b_1\, 00 \ldots 00\, b_2\, 0000 \ldots b_3 00 \ldots$$

Notice that there exists $(M-1)$ *zero-valued* samples between every pair of non-zero samples. Note also that the above sequences are not exactly identical to the analysis filter output sequences of eqns. (5.64) and (5.64). The output stream of the parallel-to-serial converter is

$$\ldots a_1 b_1 c_1 \ldots a_2 b_2 c_2 \ldots a_3 b_3 c_3 \ldots$$

This sequence can also be considered as the input information sequence of a general M carrier-modulated communication system. After passing through the delay chain, the parallel samples are

$$\ldots a_1 b_1 c_1 \ldots a_2 b_2 c_2 \ldots a_3 b_3 c_3 \ldots$$

$$\ldots b_1 c_1 d_1 \ldots b_2 c_2 c d_2 \ldots b_3 c_3 d_3 \ldots$$

$$\ldots c_1 d_1 e_1 \ldots c_2 d_2 e_2 \ldots c_3 d_3 e_3 \ldots$$

and so on. When the above parallel sequences are compared with the analysis filter output sequences of eqns. (5.64) and (5.64), they differ in the irrelevant positions. For example, the output of the first analysis filter has $(M-1)$ irrelevant samples between $a_1$ and $a_2$ that resulted from matched filter detection of channel 1. In contrast, the first output sequence of the delay chain has $(M-1)$ samples $b_1 c_1 d_1 \ldots$ (between $a_1$ and $a_2$) which are, in fact, relevant information samples of the remaining $(M-1)$ channels. However, due to decimation, the output sequences of the parallel downsamplers are

$$\ldots a_1\, a_2\, a_3 \ldots$$

$$\ldots b_1\, b_2\, b_3 \ldots$$

and so on. Thus the overall operation of the cascaded block (samplers, quantizers, P/S converter plus the delay chain) is transparent. In summary, we show that *a carrier-modulated M-channel filter bank analysis-synthesis processing system with perfect reconstruction is equivalent to an M-carrier modulated communication system with zero generalized ISI.* This implies that the design and construction of transmit and receive filter pairs that result in g-Nyquist pulses is equivalent to the design of analysis-synthesis filter bank that has perfect reconstruction property. This subject is discussed in the next section when we evaluate prototype lowpass filter designs. As a digression, we next give a brief treatment on actual implementation of multi-channel filter bank systems.

### 5.4.6 Polyphase Decomposition and Fast Computation

Throughout this chapter, our focus has been the description of continuous-time signal processing in terms of equivalent discrete-time sequences. Next, we introduce an important discrete-time domain concept known as the Perfect Reconstruction Construction; afterwards we describe

in detail its relationship to a continuous-time phenomenon called the zero-ISI Nyquist criterion. We have purposely omitted any discussion on the feasibility and computational complexity of a multi-carrier modulated communication system, mostly because this topic is treated thoroughly any standard text on multi-rate digital signal processing. We should, however, point out that the virtues of multi-carrier communication techniques (in terms of bandwidth efficiency) have been known for several decades. In the past the construction of multi-carrier modulators, frequency oscillators, and high-Q bandpass filters —for a large number of carriers/channels— was a formidable task. Furthermore, the orthogonality property that leads to zero ISI is only guaranteed when carrier frequencies and phases can be generated precisely without any error, jitter or time drift. All these issues were solved once the multi-carrier modulated signal is first generated in the discrete-time domain, followed by D/A conversion and analog RF carrier modulation. By applying certain novel mathematical results such as the noble identities and polyphase decomposition, it is possible to swap the order of filtering and up/down-sampling. This step further facilitates fast computation of carrier-modulated signals by using Fast-Fourier transform (FFT) and its inverse as carrier modulators and demodulators. For the special case of modulation and demodulation involving only in-phase cosine carriers, efficient and fast computation of channel samples is possible by using discrete-cosine transform (DCT) and its inverse. For an interested reader, we refer to the text by Malvar [82] which contains a set of computer codes for real-time FFT and DCT implementation.

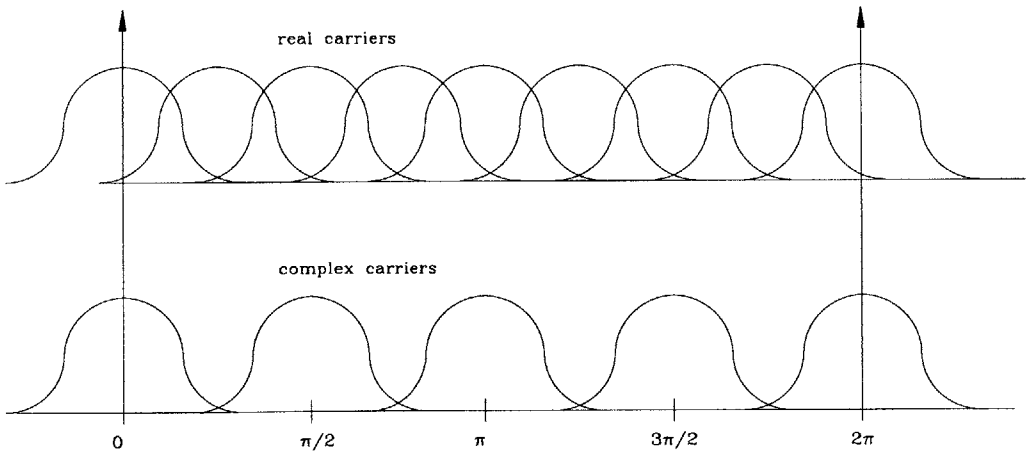



**Figure** 5.28: Bandwidth of real vs.complex carrier modulated signals

### 5.4.7 Bandwidth of Real vs. Complex Carrier Modulated Signals

One peculiarity about multi-carrier modulation in discrete-time domain is the fact that the resulting signal is still in baseband. Hence, when a complex carrier modulator such as a block

of inverse FFT is used for M-channel modulation, only half of M channels are modulated by distinct carriers. In fact there exists M/2 complementary-conjugate pairs of channels —one modulated by the in-phase (cosine) carrier and the other by the quadrature (sine) carrier. This fact has already been noted when we are studying Fourier (complex) and cosine (real) carriers.

At first glance this does not appear to be of any significance. In terms of bandwidth efficiency, both complex and real carrier modulated signals have the same PSD width. As illustrated in Fig. 5.28, the PSD width of a data modulated signal is the same whether data is modulated onto both carriers (in-phase and quadrature) or only onto the in-phase carrier. However, in order to maintain channel orthogonality in multi-carrier communication, the complex carriers must be spaced apart by twice the separation of real carriers. Since two the amount of independent information is carried on complex carriers, both schemes have the same information rate. Now, this is where the discrete modulation based on Fourier carriers becomes significant. In Fourier modulation via inverse DFT, the complex carriers are equal to complex (complementary-pair) roots on the unit circle for even M. Examples are shown in Fig. 5.29 for $M = 4$ and 8.

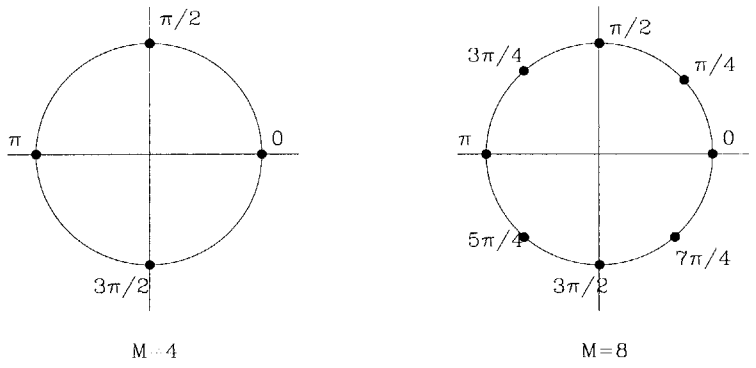



**Figure** 5.29: Complex roots of unity as complex carriers

Note that both roots 0 and $e^{j\pi}$ are real carrier pairs; thus this (real in-phase) carrier (whose center frequency is zero) can only carry real data. For non-spread spectrum modulated communications, this is a non-issue (in decreased spectral efficiency) when M is large. However, for channelization coding in spread-spectrum modulated communications, parallel channel orthogonality can no longer be maintained since two chips are missing in the quadrature branch. Therefore, code-division multiplexing with multi-carrier complex modulation cannot be implemented with DFT and IDFT blocks. Of course, this situation does not arise when center frequencies of all distinct carriers are non-zero (as in DCT II and IV based carriers) or when all carriers are real and carry real data only —such is the case for any DCT based multi-channel discrete modulation.

## 5.5 Prototype Window Design

The PR condition of eqn. (5.63) gives a clue on the design of analysis and synthesis filters for an M-channel filter bank. The task however still seems formidable due to the number of parameters involved. Fortunately, the PR condition is simplified considerably when every channel filter is a modulated version of a prototype filter $f[n]$ or $h[n]$.

$$\begin{aligned} \text{Analysis filter:} \quad & h_k[n] = h[n]\, m_k[n] \\ \text{Synthesis filter:} \quad & h_k[n] = p[n]\, m_k^*[n] \end{aligned}$$

where $m_k[n]$ denotes the modulating carrier function for the $k^{th}$ channel.
For Fourier-based complex modulation,

$$\begin{aligned} h_k[n] &= h[n]\, \exp\left(-j\frac{2\pi kn}{M}\right) \\ f_k[n] &= f[n]\, \exp\left(j\frac{2\pi kn}{M}\right) \end{aligned} \tag{5.66}$$

For cosine-based real modulation,

$$\begin{aligned} h_k[n] &= h[n]\, \cos\left(\frac{\pi kn}{M}\right) \\ f_k[n] &= f[n]\, \cos\left(\frac{\pi kn}{M}\right) \end{aligned} \tag{5.67}$$

Of course, it is understood that in both cases we can substitute carriers with frequency offset. By further stipulating that for each channel the analysis-synthesis filter be a matched-filter pair

$$f_k[n] = h_k^*[N_w - 1 - n] \tag{5.68}$$

we can rewrite the PR condition of eqn. (5.63) for complex modulation as

$$\sum_{k=0}^{M-1} \sum_{l} h_k[lM - r]\, f_k[n - lM] =$$

$$\sum_{l} f_k[n - lM]\, f_k[N_w - 1 - lM + r] \sum_{k=0}^{M-1} \exp\left(\frac{-j2\pi k}{M}[n - r]\right)$$

For a prototype synthesis filter whose length $N_w$ is equal to block length $M$

$$\sum_{l} f_k[n - lM]\, f_k[M - 1 - lM + r] \sum_{k=0}^{M-1} \exp\left(\frac{-j2\pi k}{M}[n - r]\right)$$

For $l \neq 0$, the prototype filter taps have weight zero. Thus we need to consider only the case $l = 0$.

$$f_k[n]\, f_k[M + r] \sum_{k=0}^{M-1} \exp\left(\frac{-j2\pi k}{M}[n - r]\right)$$

We know that for $n \neq r$ the summation of all complex roots of unity is zero. Therefore, by setting $n = r$ the PR condition implies

$$M f_k[n] f_k[M + n] = 1 \qquad \text{for } n \in \mathbb{Z}^{M-1} \tag{5.69}$$

The above result states that the square of every tap weight of the prototype synthesis filter must be unity. We immediately deduce that a train of impulses (a rectangular window)

$$f[n] = h[n] = \begin{cases} 1 & n \in \mathbb{Z}^{M-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.70}$$

is a valid prototype filter that satisfies the PR condition. This filter is, however, far from practical due to its sinc-type infinite bandwidth. If this prototype filter $f(t)$ is used for multi-carrier modulation and transmitted through a bandlimited channel, the channel output will be severely distorted, and the g-Nyquist condition is violated.

### 5.5.1 Time-Limited Pulses

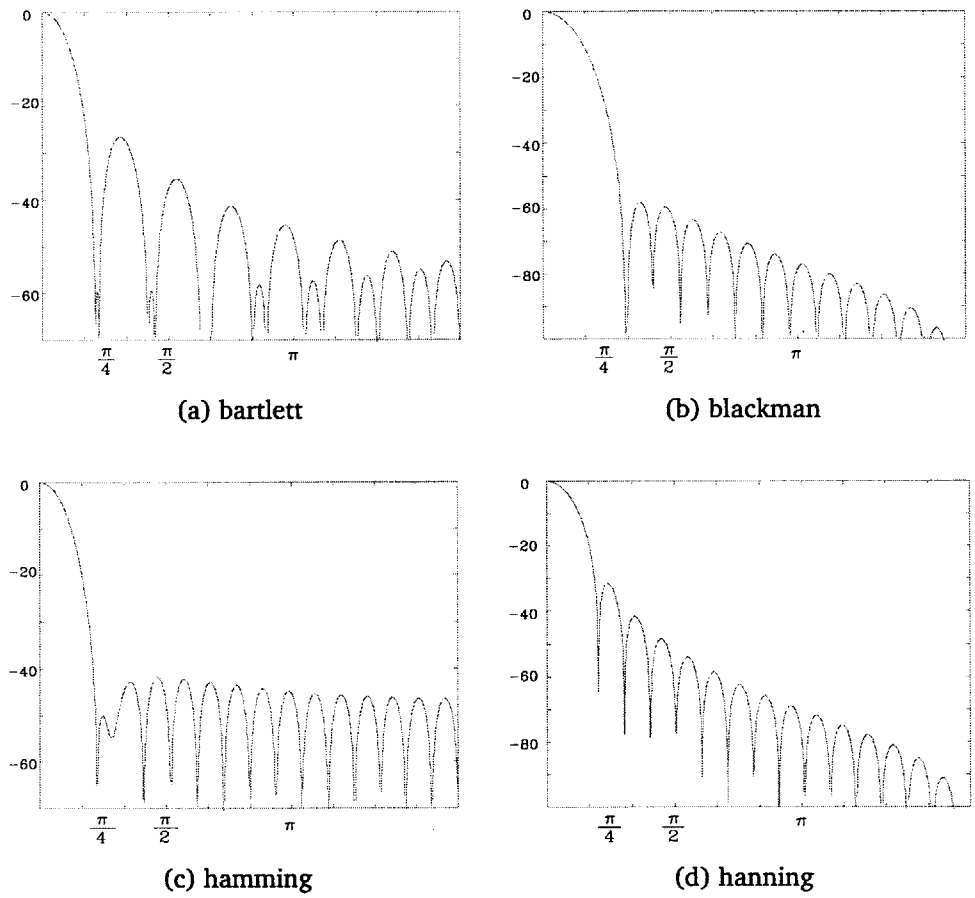
In discrete-time signal processing literature, the shaping of a generic pulse such as the impulse train of length M is known as *windowing*. The *windowed function* of length M

$$f_w(t) = f(t)\, w(t)$$

is a product of an impulse train of length M and a *window function*. Popular windows are Hamming (raised sinusoidal function), Bartlett (triangular function) and Kaiser (modified zeroth-order Bessel function). The frequency response is several windowed functions are plotted in Fig. 5.30.

Among is four windows the shaping amount is the greatest for the Blackman function. It is no surprise that its first sidelobe is much lower than those of other windows. It is evident from the plots that the sidelobe levels are lowered at the expense of a wider mainlobe. We provide another example where the free parameter $\alpha$ of the Kaiser function is varied. The resulting time windows and their corresponding frequency plots are shown in Fig. 5.31. The first sidelobe level drops from $-13$ dB for $\alpha = 0.5$ to $-38$ dB for $\alpha = 5$ at the expense of a wider mainlobe. This is a by-product of energy conservation. We mention this topic of windows in brief to remark that such pulse shaping techniques do not satisfy the generalized Nyquist or PR criteria. We can also deduce from eqn. (5.69) the following:

- In a complex (Fourier) carrier modulated, M-channel filter bank, any shaping of a rectangular prototype window of block length M does not meet the PR condition. Equivalently, in a multi-carrier modulated communication system, shaping of a symbol-time limited



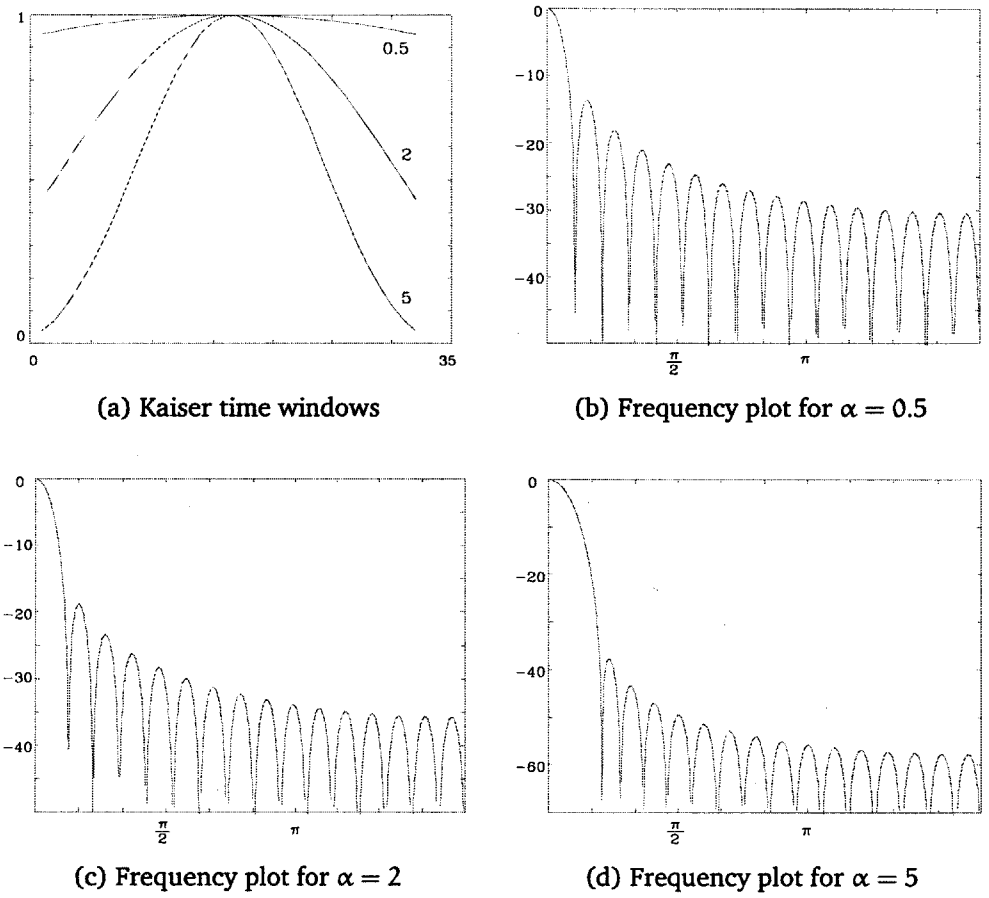

(a) bartlett (b) blackman

(c) hamming (d) hanning

**Figure** 5.30: Frequency response of various windows for length $M = 32$. Among the four windows, Bartlett (triangular) has the least smoothest transition. On the other hand, the Hamming window (raised sinusoidal) has very low sidelobes. Its mainlobe width is $\simeq \pi/4$. The Blackman window with the largest shaping has the lowest first sidelobe at $-60$ dB. However, its mainlobe width is the widest among the windows.

rectangular transmit pulse does not meet the g-Nyquist criterion. A vigorous proof of this statement is not given here. Its derivation is fairly straightforward. The reader can also refer to the text by Vaidyanathan [148] for a proof.

- In single-carrier modulation format, it is known that communication is most efficient when data is carried in both in-phase and quadrature sub-channels. There is no expansion in transmission bandwidth. As an example, QAM is always preferred over PAM since the former is twice the spectral efficiency as the latter. It is also known that in passband PAM, the spectrum of the modulated signal is symmetric about the carrier frequency $\pm f_c$ because the PAM signal set consists of real time functions only. That is, the same information is contained in either the upper (or equivalently, lower) sidelobe of the transmit




(a) Kaiser time windows

(b) Frequency plot for $\alpha = 0.5$

(c) Frequency plot for $\alpha = 2$

(d) Frequency plot for $\alpha = 5$

**Figure** 5.31: Kaiser windows of length $M = 32$ and corresponding frequency plots

modulated passband signal. In analog communication, this is called single-sideband modulation. It is shown in any undergraduate text on communication theory that the generation of a single-sideband modulated signal involves the use of a Hilbert transformer in baseband or a passband filtering with a steep cut-off —both techniques difficult to realize in practice. In most cases, its alternative double-sideband modulation format is preferred due to its simpler transmitter design. Interestingly, in multi-carrier modulation format, a sub-channel can be *real (cosine) carrier* modulated with *real data symbols*, or *complex (Fourier) carrier modulated* with *complex data symbols*. It is straightforward to see that both methods have the same spectral efficiency. As described in Chapter 4 Sec. 4.8, the frequency separation between Fourier carriers is double the spacing of cosine carriers if mutual orthogonality among carriers is maintained in both formats. Therefore, multi-carrier modulation with cosine functions is analogous to frequency-division multiplexing of single-sideband modulated signals —without the added complexity of sharp passband filters or baseband Hilbert transformers.

### 5.5.2 Time-Band-Limited Pulses

Up to this point, we have considered both symbol-time-limited finite-length pulses and infinite-length pulses as candidates for multi-carrier modulated waveforms. The goal is to design pulse shapes that meet the generalized Nyquist criterion when transmitted through a bandlimited channel. We show that infinite length pulses such as sinc-type pulses are impractical. A more practical raised-cosine pulses that were devised by Chang are g-Nyquist only when the receiver filter length is infinite. Any truncation of Change pulses would violate the g-Nyquist constraint. In practice, the class of truncated Change pulses may be a viable option if their induced ISI can be tolerated. For that matter, we showed that fully raised-cosine (Chang) pulses with multi-carrier modulation format is more spectrally efficient than raised cosine pulses with roll-off factor r and single carrier modulation if the number of sub-carriers (channels) $N > 1/r$. Next we state that in a complex-carrier modulated M-channel communication system with transmit pulses of length M, no pulse shaping can meet the g-Nyquist criterion. We obtain valuable insights into the design of finite-length pulses that meet g-Nyquist by studying M-channel analysis-synthesis filter bank system. Most important of all, we derive the PR criterion applicable to an analysis-synthesis filter pair that turns out to be equivalent to the g-Nyquist criterion for a transmit-receive filter pair. It is then shown that for a carrier-modulated filter bank, the PR condition depends only on the design of the impulse response (tap weights) of a prototype lowpass FIR filter. The impulse response of this prototype filter can be used as the prototype transmit pulse shape in multi-carrier modulation. As a side bonus, the transmit and receiver FIR filters are also a matched filter pair.

The remaining work in this chapter is the design of such prototype pulse shapes that meet the g-Nyquist and PR criteria. We define these pulses as time-overlapped (TO) functions due to their pulse length of 2KM. That is, these TO pulses are finite-length but much longer than the block length M. They were introduced by Henrique Malvar and their unique properties are thoroughly covered in his text and several other journal articles [81, 82]. When $K = 1$ (i.e., pulse length of 2M), the time-overlapped pulses are called *modulated lapped transform* (MLT) functions. For $K \in \mathbb{Z}_2$ they are collectively known as *extended lapped transform* (ELT) functions.

- *Modulated Lapped Transform Functions:*
  Prototype Window:

$$p[n] = \pm \sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right] \tag{5.71}$$

  Synthesis Filters:

$$f_k[n] = p[n]\sqrt{\frac{2}{M}}\cos\left[\left(k + \frac{1}{2}\right)\left(n + \frac{M+1}{2}\right)\frac{\pi}{M}\right] \tag{5.72}$$

- *Extended Lapped Transform Functions:*
  Prototype Window:

$$p[n] = -\frac{1}{2\sqrt{2}} + \frac{1}{2}\cos\left[\left(n+\frac{1}{2}\right)\frac{\pi}{M}\right] \tag{5.73}$$

  Synthesis Filters:

$$f_k[n] = p[n]\cos\left[\left(k+\frac{1}{2}\right)\left(n+\frac{M+1}{2}\right)\frac{\pi}{M}\right] \tag{5.74}$$

Note that both MLT and ELT functions use the same carrier function. In fact, by comparing the above equation with eqn. (5.51) we see that the carrier function is a translated type IV discrete cosine transform function: $E_K^{IV}\left[n+\frac{M}{2}\right]$. It is straightforward to show that MLT and ELT synthesis functions meet the Generalized Nyquist Criterion of eqn. (5.54) and the Perfect Reconstruction Criterion of eqn. (5.63). Prototype windows are plotted in Fig. 5.32.

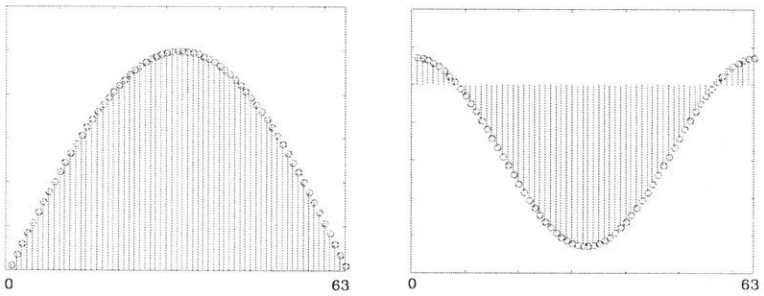



**Figure** 5.32: Prototype windows for (a) MLT (b) ELT functions

Both are symmetric (even) functions about its mid-point, implying that only half the tap weights of the FIR filter are involved in the prototype pulse design. The resulting MLT and ELT synthesis functions —precisely the first four— are plotted in Figs. 5.34 and 5.36. It is apparent that the passband synthesis functions are neither odd or even about its mid-point. This characteristic is due to modulation of a symmetric prototype function with a type IV discrete-time cosine function that has offsets in both time and frequency parameters. In Fig. 5.33 we plot the first two MLT synthesis functions $f_0(t)$ and $f_1(t)$ of length 64 and their symbol-time shifted versions. It is rather difficult to see graphically but all four waveforms are mutually orthogonal. This is the most fascinating property of time-overlapped pulses, applicable to both MLT and ELT, that all basis/synthesis functions satisfy all three types of orthogonality. We split both $f_0[n]$ and $f_1[n]$ at midpoint and relabel them as follows for $n \in [0, L-1]$:

$$f_{0+}[n] = f_0[n], \qquad f_{0-}[n] = f_0[n+L], \qquad f_{1+}[n] = f_1[n], \qquad f_{1-}[n] = f_1[n+L]$$

1. *Conventional Orthogonality*: Two synthesis functions $f_i[n]$ and $f_j[n]$ of length $M = 2L$ are orthogonal over a symbol-time correlation window $[0, L)$.

$$f_0[n] \perp f_1[n]$$

2. *Self Orthogonality*: A synthesis function $f_i[n]$ and its symbol-time shifted version $f_i[n + L]$ are also orthogonal over a correlation window $[L, 2L)$.

$$f_{0+}[n] \perp f_{0-}[n] \qquad f_{1+}[n] \perp f_{1-}[n]$$

3. *Overlap Orthogonality*: A synthesis function $f_i[n]$ and another synthesis function with a delay or advance equal to the symbol-time $f_j[n \pm L]$ are also orthogonal. Both leading and lagging time offsets must be considered since the synthesis functions are neither odd or even.

$$f_{0+}[n] \perp f_{1-}[n], \qquad f_{0-}[n] \perp f_{1+}[n], \qquad f_{0+}[n] \perp f_{1+}[n], \qquad f_{0-}[n] \perp f_{1-}[n]$$

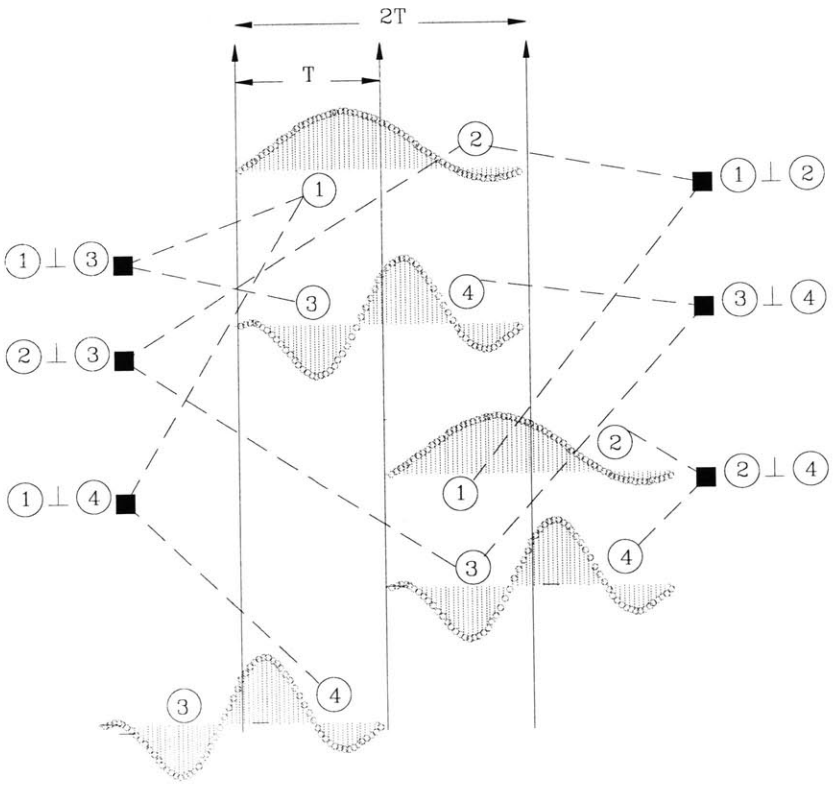



**Figure** 5.33: Orthogonality of MLT synthesis functions and their time offset versions

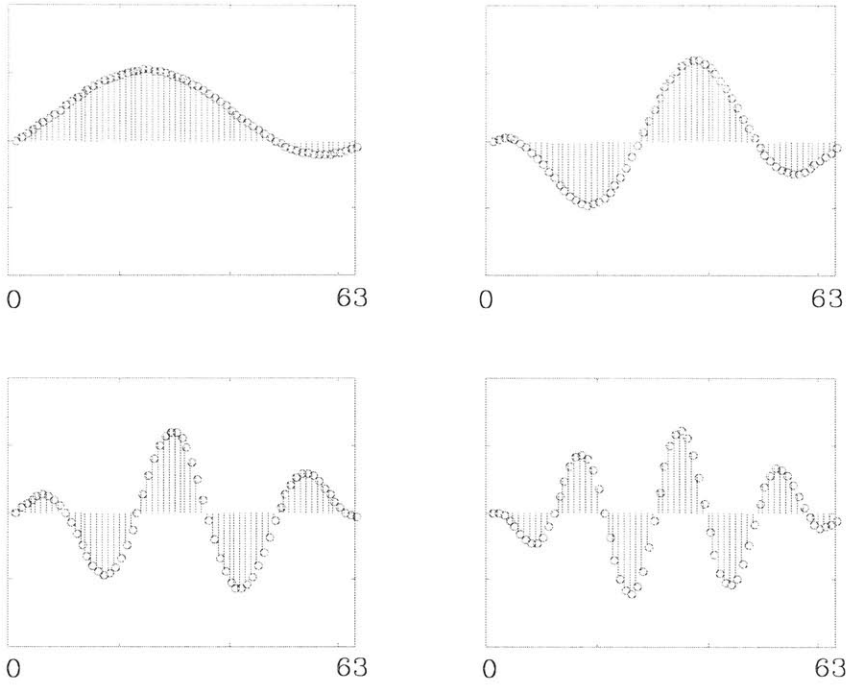



**Figure** 5.34: Synthesis basis functions of MLT

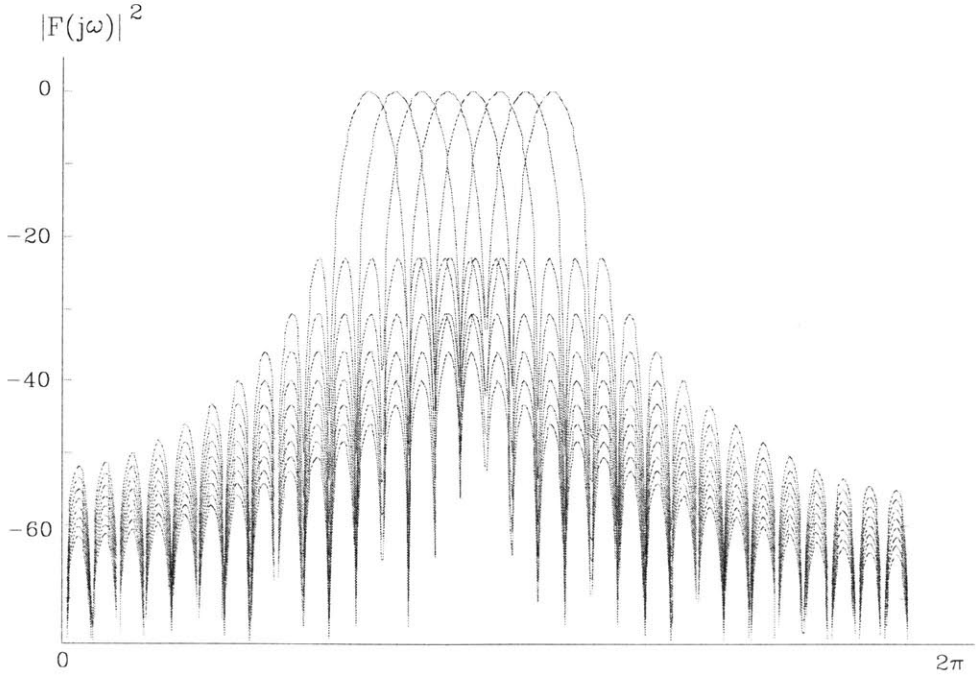



**Figure** 5.35: Amplitude response of MLT basis functions

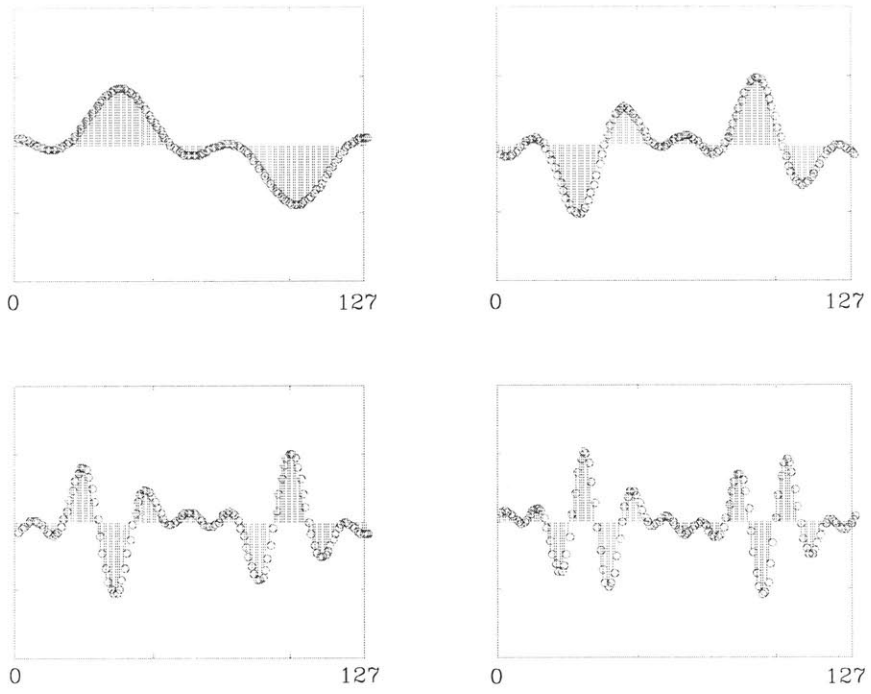



**Figure** 5.36: Synthesis basis functions of ELT

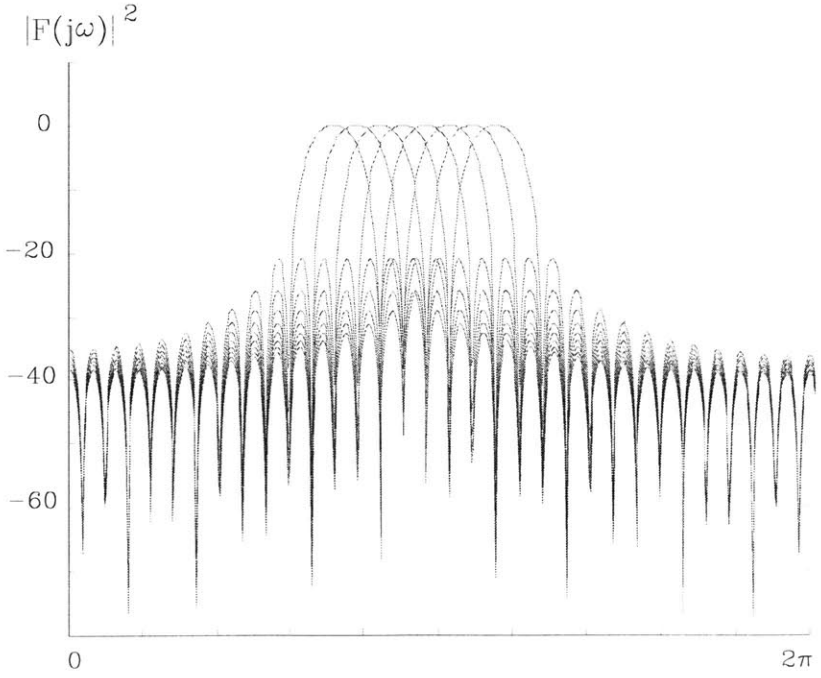
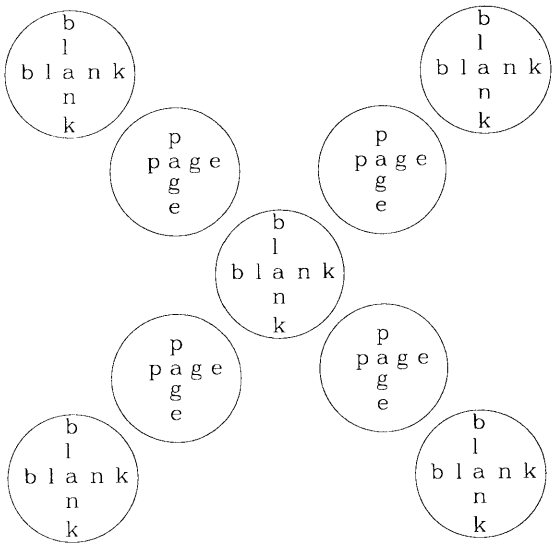



**Figure** 5.37: Amplitude response of ELT basis functions

## Notes and References

The Sampling Theorem is credited to C. E. Shannon, which appeared in [127], "Communications in the presence of noise." It was not discussed in his more well-known classic work [128]. Prior to Shannon, other researchers, H. Nyquist [93], E. T. Whittaker [159], J. M. Whittaker [160] and V. A. Kotel'nikov [65] have studied sampling of band-limited signals. They all considered sampling of *deterministic* signals. An extension to wide-sense stationary random processes was studied by Balakrishnan [5, 6]. For non-stationary processes, the works by Zakai [167], Piranashvilli [103] and Gardner [38] are cited. The sampling of linearly modulated waveforms for zero intersymbol interference was studied by Nyquist [93]. He proved that transmission at Nyquist rate is possible using sinc pulses with an ideal frequency response. R. Chang [17] showed that transmission speeds arbitrarily close to the Nyquist rate can be achieved using non-sinc like pulses with a gradual roll-off. These pulses satisfy the *generalized* Nyquist criterion of Schnidman [124]. However, both sinc and Chanc pulses are infinite-length in time. The notion of dimension for a band-limited signal over a window of T sec was proposed by Shannon [127]. The discrete-time analog of Nyquist criterion is the Perfect Reconstruction (PR) condition. Various forms of the PR condition were offered by A. Croisier, D. Esteban and C. Galand [26] and Smith and Barwell [132]. A special class of PR waveforms with good spectral properties is the family of time-overlapped (TO) pulses, first reported by Cassereau [14] where the pulse length is twice the symbol interval. H. Malvar [81, 82] extended this work by considering TO pulses longer than two symbol intervals. Since TO pulses have very-low out-of-band leakage, we also coin them time-bandlimited (TB) pulses.


blank
page
blank
page
blank
page
page
blank
blank

# 6

# MULTI-RATE TRANSMISSION SCHEMES

## Summary

For a band-limited channel, communication at a higher information rate invariably requires the use of multi-level signal constellation sets. This paradigm is the norm in digital subscriber loop (DSL) applications over a narrow-band twisted-pair telephone (TPT) channel. It also applies to long-haul over-the-horizon microwave (LOM) links. However, the majority of wireless systems including cellular networks adopt binary antipodal signal sets. The contributing factors may be power efficiency of linear amplification for constant-amplitude signals, receiver complexity constraint, limited energy source (of battery) for hand-held devices, higher spectral efficiency of binary modulation in low SNR operating region, and difficulty in tracking and locking of rapid fluctuation in received signal frequency, amplitude and phase. We observe several differences between cellular radio propagation (CRP) channel and the above mentioned twisted-pair telephone and microwave links.

- The frequency response of a TPT channel is highly frequency-selective (time dispersive) but relatively stationary. Therefore, channel state information can be accurately measured and relayed back through a feedback channel during channel set-up. Detection of a two-dimensional multi-level signal is feasible due to accurate estimation of channel gain and phase-coherent demodulation. Most of these attributes cannot be sustained in cellular radio since the channel is both time- and frequency-selective.

- Compared to cellular radio, a LOM channel is line-of-sight. In fact, the microwave towers are raised high enough that several Fresnel zones are free of any obstruction. Both transmitter and receiver are immobile. A large amount of signal processing complexity

and power consumption is tolerated at the receiver. These conditions may not be applicable to a CRP channel. In densely populated areas, no LOS path exists in cellular radio; reception from diffracted and scattered paths is the norm. Furthermore, power efficiency of receiver signal processors and RF transmit amplifier[1] of a mobile unit must be tightly controlled to improve battery life.

- Both TPT and LOM channels are single-user links; i.e., the physical medium itself —the copper wire or RF band— is not shared among many users. Hence, neither channel is interference- or transmit power-limited. In cellular radio an increase in transmit power at a base site may cause excessive interference at neighboring cells —reducing the signal-to-interference ratio (SIR) of receivers in neighboring cells at the expense of a higher SIR in the target cell. As a result the overall capacity of the entire network may be reduced.

If data modulation is limited to binary antipodal signal set, we know that there is a linear relationship between the bit rate $R_b = 1/T_b$ and signal dimension $D = WT_b$. In conventional FDM, the forward-link composite channel consists of a sum of M logical channels, each channel of bandwidth $W/M$. Similarly in conventional TDM, a logical channel (which is a time-slot) is a slice of duration $T_b/M$ of the entire frame of length $T_b$. A logical channel's information rate can be increased by transmitting a higher-dimensional signal via a larger bandwidth or time-slot. This is equivalent to the assignment of multiple frequency bands or time-slots, with each acting as a low-rate binary communication (logical) channel. In CDM where the actual transmission (chip) rate is independent of the information-bearing data rate, an interesting situation arises: For a fixed bandwidth (i.e, constant chip rate) it is possible to communicate at higher information rates by reducing the ratio between transmit bandwidth $W$ and information bit rate $R_b$. This ratio

$$P_G = \frac{W}{R_b} \tag{6.1}$$

which we defined earlier as "dimension" of binary signals is known as the *processing gain*. The relation between the optimal $P_G$ and cellular capacity in information-theoretic sense —in which unconstrained Gaussian signals are assumed— has already been addressed in Chapter 3. Here we are concerned with a comparative BER analysis between two binary, multi-rate CDM transmission techniques: one where all (low-rate) channels maintain the same $P_G$, and the other where each channel has a different $P_G$ to suit its respective information rate. The outcome of this analysis is consequential for the following reason: We stated in Chapter 3 that the transmission and reception of orthogonal multiple code channels per user can be replaced with a single code channel of reduced spreading —without having any effect on the achievable rate region of CDM. This conclusion guarantees that the scalability and complexity of a CDM receiver is on

[1]The power efficiency of a fixed base site is a non-issue. Indeed, the power efficiency of RF *transmit* amplifiers for mobile handsets is also not relevant to our thesis. We are only concerned with reception of CDM signals.

par with that of TDM; i.e., the receiver complexity is independent of the information rate. In this chapter, our goal is also prove that this conclusion is still applicable under practical settings with linear modulation-demodulation of binary signal sets through an AWGN channel with or without multipath fading.

As it turns out, methodologies used for BER computation of CDM signals themselves require further scrutiny for their accuracy. Historically, spread-spectrum modulated (SSM) communications systems —both single-user and multiple-access— are measured in terms of their ability to suppress unwanted interference or jamming signals. Moreover, for certain secure military applications, the processing gain may be in excess 1000 such that the PSD of a spread-spectrum modulated signal is hidden below the noise level. Under such conditions it is accurate to assume the received signal is composed of the desired SSM signal and undesired *disturbance* —which is a combination of unwanted interference and additive thermal noise. This disturbance is traditionally modelled as an AWGN process. We denote the one-sided PSD of this disturbance by

$$D_o = I_o + N_o$$

where $I_o$ and $N_o$ are one-sided PSDs of interference and noise processes, respectively. Since a Gaussian distribution is completely specified by its first and second moments (i.e., mean and variance), we only need to compute the mean of the desired signal and the variance of the disturbance to compute the BER. From signal detection theory we know that the BER of an antipodal binary spread-spectrum modulated signal is expressed in terms of a single parameter: the ratio of signal energy-to-disturbance PSD $\mathcal{E}_b/D_o$. Specifically (as detailed in the body),

$$\text{BER} \triangleq P_e = Q\left(\sqrt{\frac{2\mathcal{E}_b}{D_o}}\right) = Q\left(\sqrt{2P_G\,\text{SDR}}\right)$$

where $Q(\cdot)$ is the standard Gaussian complementary cumulative distributive function (CDF). The ratio $\mathcal{E}_b/D_o$ is related to the signal-to-disturbance power ratio (SDR) as follows:

$$\text{SDR} = \frac{P}{D_o W} = \frac{\mathcal{E}_b R_b}{D_o W} = \frac{\mathcal{E}_b}{D_o} \cdot \frac{1}{P_G}$$

If interference dominates over thermal noise, then the BER is a function of signal-to-interference power ratio (SIR) only:

$$P_e = Q\left(\sqrt{2P_G \frac{P}{I_o W}}\right)$$

Then a reduction in $P_G$ by $M$ can be compensated by increasing the transmit power of the SSM signal by the same factor —implying that the information rate of an SSM signal can be increased by reducing the processing gain without any degradation in BER. Related to this conjecture are three important questions:

- What is the role of $P_G$ in multi-user broadcast plus interference (composite) channel setting? What is the minimum $P_G$ that a traffic channel must maintain for reliable communication?

- In a composite channel model, how does a high-rate channel with reduced spreading affect (i.e., interfere) with other users in the system, even if its own BER is unaffected by a lower $P_G$?

- Under what conditions can we safely apply various Gaussian approximations in evaluating the BER? The Gaussian approximation, whenever it is valid, is desired because full knowledge of the statistical distribution of the disturbance is no longer required in BER analysis. Furthermore, it gives valuable insights into the effect of various system parameters on the overall system performance.

We attempt to answer the above questions in the remainder of this chapter. In brief, our findings are as follows:

- The resulting BER of CDM signalling is a function of many system parameters, including the signature sequences of both target and interfering users. We must therefore be aware that the exact BER itself is averaged over all possible combination of signature sequences. For this reason, it is assumed that all signature sequences are iid bipolar, symmetric Bernoulli random processes and the resulting BER applies to CDM signalling with random binary sequences.

- For CDM signalling via an ideal channel perturbed by AWGN alone, it is shown that BER results based on the standard Gaussian approximation (SGA) —the simplest in mathematical form— are optimistic when the interference population is low —for a fixed $\mathcal{E}_b/N_o$ (not $\mathcal{E}_b/I_o$ or SIR). For a fixed number of interferers, the SGA-based BER deviates considerably from the exact plot as $\mathcal{E}_b/N_o$ increases. On the other hand, the approximate BER curve based on improved Gaussian approximation —the most complex in mathematical form— is slightly more pessimistic than the exact curve as the interference population *increases*. For low interference population it gives very accurate results. Unlike SGA, its accuracy is immune to $\mathcal{E}_b/N_o$ variation. The BER curves based on conditional Gaussian approximation —with slight increase in complexity over SGA— closes the wide gap between exact and SGA based approximate curves. It still suffers from slight optimistic outcomes and susceptibility to $\mathcal{E}_b/N_o$ variation.

- This gap between the exact and approximate BER is narrowed when the channel undergoes flat fading. The choice of the statistical distribution of fading amplitude has little effect on this gap. Thus far we have implicitly assumed that the target receiver is a simple

linear, single-user matched filter based CDM decoder without any interference cancellation mechanism. For frequency selective fading channels, it is no longer meaningful to simply analyze the accuracy of various Gaussian approximations; we must also define the type of receiver structure since its decoding complexity influences BER curves.

- Based on exact BER curves, the performance of two target high-rate CDM users, parallel-channel, single-gain (PC-SG) and single-channel, reduced-gain (SC-RG) are almost identical; SC-RG user has a slightly lower BER for low interference population. This is likely due to the Gaussianity effect of larger interference population as observed by PC-SG user, even though the total interference power (variance) is the same in both cases. If the target receiver is low rate single-channel, single-gain, no measurable difference in observed in its exact BER curve when a PC-SG user is replaced by a SC-RG user. Therefore, for all practical purposes, we can accept that both high rate users affect other low rate CDM users in the system equally.

- For all three Gaussian approximations the value of spreading gain $N_c$ plays a minor role in their accuracies. Thus, the resulting approximate BER curves of SC-SG, PC-SG and SC-RG CDM users in AWGN channel —with or without flat fading— can be used for comparative system analysis with high level of confidence.

- The exact BER computation of CDM signalling and Rake combining in frequency-selective fading environment is rather difficult. On one hand we know how to compute the exact BER of multi-channel signalling and optimal maximum ratio combining —this is straightforward textbook material. On the other hand, we run into two thorny issues when we try to generalize the analytical results of multi-channel signalling to CDM domain: First, in CDM signal reception, the output of parallel channel receivers are correlated due to non-orthogonality of signature sequences assigned to parallel channels. The Rake receiver, compared to the optimal maximum-ratio combiner, is sub-optimal unless self-induced noise among parallel channel outputs is eliminated. Second, in CDM signalling, the received signal is perturbed by non-white, non-Gaussian disturbance —a combination of thermal noise and multi-user interference (MUI). Therefore, earlier analytical results based on AWGN can no longer be applied in a straightforward manner. Accurate statistical modelling of MUI become a critical issue.

- In wideband CDM signalling the resulting resolvable parallel channels are channel induced —not an elaborate system design as in other diversity systems based on space, time and frequency. The relative time delays between parallel channels is dependent on the channel impulse response. The maximum relative delay is the channel delay spread. in conventional Rake combining, any received signal whose arrival time is outside the

correlation window is treated as undesirable inter-symbol interference (ISI). Therefore, for a fixed delay spread, it is more likely that reception of SC-RG CDM signal generates a higher amount of ISI than PC-SG CDM user, resulting in higher BER for an SC-RG CDM channel.

- If the arrival epochs of resolvable paths are within the correlation window, the resulting self-interference between any two Rake finger outputs is *correlated*. In PC-SG CDM with $M$ parallel channels, correlated self-interference is composed of a single auto-correlation term and $(M-1)$ cross-correlation terms. In SC-RG, self-interference results from a single auto-correlation term. Since we assume signature sequences are random Bernoulli processes, their auto- and cross-correlation values for non-zero time offsets are the same. Based on this model, both PC-SG and SC-RG CDM channels have comparable BER curves.

## 6.1 Multi-Rate Transmission Techniques

It is known that the code sequence that achieves capacity of a single-user discrete-time Gaussian channel is Gaussian distributed with unconstrained amplitude levels. In practice, the signal amplitude set is constrained to a finite symmetric set such as the M-PAM set. For a continuous-time AWGN channel, capacity-achieving waveforms have brick-wall frequency response with unconstrained Gaussian amplitudes. Here it is understood that the waveforms are sample functions of a random process, and a brick-wall frequency response refers to a flat "white" transmit power spectral density within the channel bandwidth $W$. We know that sinc pulses have a flat frequency response; they also attain the highest symbol rate of $2W$ samples per second through an ideal bandlimited channel of $W$ Hz. If the single-user AWGN capacity is $C^s$ bits per second, we can determine the number of optimal bits per symbol through such channel as

$$C^s/2W$$

In practice, the achievable rate R (in bits per second) is less than $C^s$, mostly due to non-ideal coding and decoding. The actual symbol rate is also less than $2W$ samples per second due to pulse shaping (and consequential excess bandwidth). Hence, the target number of bits per symbol is

$$\frac{R}{(2-r)W} \tag{6.2}$$

where $r$ is the excess bandwidth roll-off factor. (The excess bandwidth is $rW$.) For a fully raised cosine pulse, the bit-to-symbol ratio is $R/W$. From Shannon's capacity formula, it is immediate that for SNR=1 (0 dB), $C^s/W = 1$. It is then deduced that for low SNR, $R/W \approx 1$. This is the *power-limited* channel. For high SNR, $R/W \gg 1$. By pumping a larger amount of power into a

signal of $W$ Hz, the bit rate can be increased considerably. For high SNR,

$$C^s/W \approx \log \text{SNR}$$

i.e., an exponential increase in power is required for a linear increase in capacity. This is the well-known rate penalty incurred by a *bandwidth-limited* channel for communication at a higher information rate. For a twisted-pair bandlimited telephone channel with a small usable bandwidth, pumping a high-power signal is the only option for a high bit rate. In contrast, a communication satellite with its constrained transmit power and large transmission bandwidth is most efficient when its operating point is in a very low SNR region.

The propagation channel of a terrestrial radio system is considered band-limited since the RF bandwidth is a scarce resource that is generally allocated by a regulatory body such as the Federal Communications Commission (FCC) in the United States. Due to frequency re-use, the CRP channel is also power-limited. In terrestrial broadcast services such as radio and television, power constraint is imposed by a regulatory body to limit the interference of a broadcast signal in a city into a neighboring city or town where another station may be using the same RF bandwidth. Likewise a high-density cellular radio network is effectively an interference-limited system, and regulation of transmit power at each base site is necessary to maximize the overall network capacity. We already know that in a low SNR region (i.e., a system with large interference level) the capacity is (approximately) linearly proportional to SNR. For this reason, many researchers have argued that it is counter-productive to transmit signals from a large amplitude set (e.g., M-PAM set with large M) through a cellular radio channel. Of course if a receiver demands a bite rate $R \gg W$, then the selection of a large amplitude set may be the only option.

In code-division multiplexed signal transmission, $R/W \ll 1$ due to spectral spreading. Interestingly it is possible to increase the information bit rate R —to a certain extent— without migrating to a higher signal constellation set. For example, with the binary antipodal set $\{1, -1\}$ the information (bit) rate $R = R_b$ is increased by reducing the processing gain[2] as long as $R/W < 1$. Note that if the bit rate is increased by N —i.e., $P_G$ is reduced by N— then the transmit power must be increased by the same factor N to maintain the same bit energy-to-noise density ($E_b/N_o$) regardless of the bit rate. This begs the question: *"Which scheme is the most spectrally efficient for higher transmission rates?"*

- a single-channel transmission mode with a fixed signal set (e.g., binary antipodal set) and reduced processing gain?, or
- parallel-channel transmission mode with a fixed signal set and no reduction in processing

[2]See our interpretation of (the difference between) processing gain $P_G$ and spreading gain SG (or spreading factor SF) on page 42.

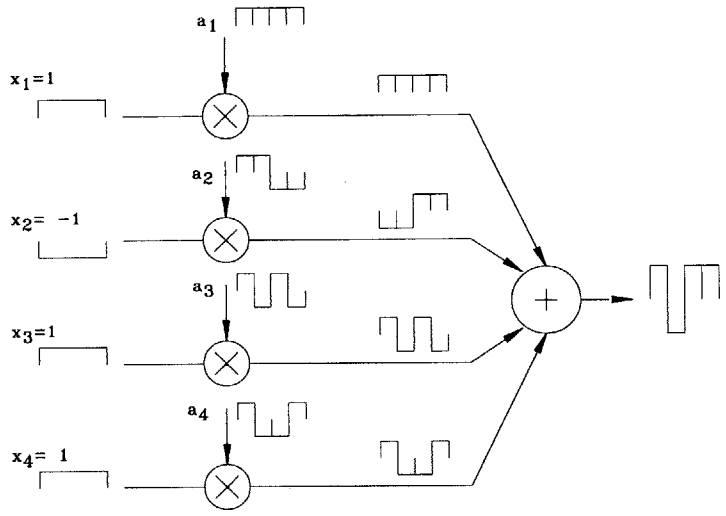



**Figure** 6.1: Code-division multiplexing of four standard-rate channels

gain? This is the power splitting option where an N-fold increase in the bit rate is achieved by transmitting N low-bit-rate channels in parallel; or

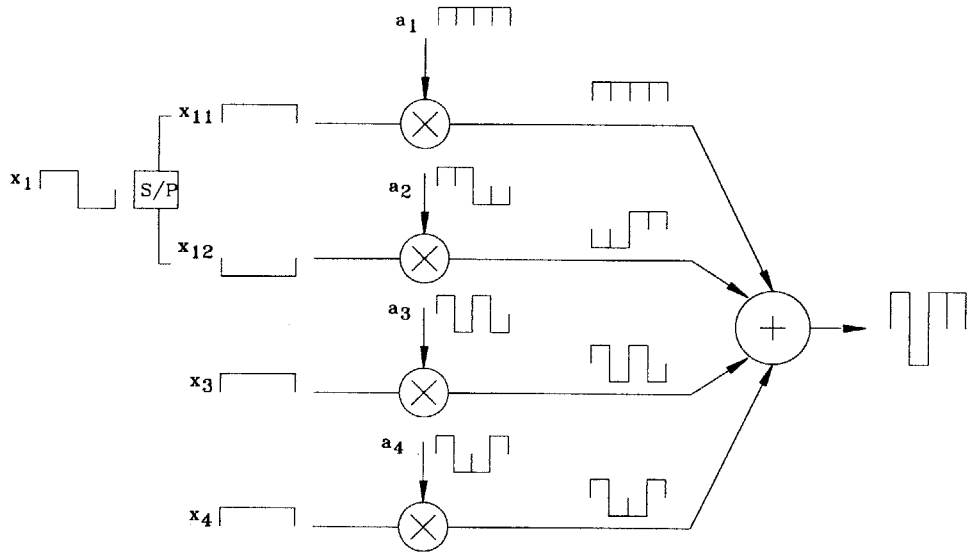
- a single channel transmission with a larger signal amplitude set and a fixed processing gain?

We do not address the third question regarding higher-order signal constellation sets. This topic is beyond the scope of our analysis. This subject has been studied for both non-spread spectrum and spread spectrum modulated signal sets [125][126], and it is widely accepted that for the same information rate, non-binary transmission mode is inferior to binary signalling in an AWGN channel with or without multi-user interference and multipath fading. We focus on three binary transmission schemes:

1. Standard single-rate transmission which we will call single-channel, single gain (SC-SG) mode. This is illustrated in Fig. 6.1 for four channels. To emphasize Walsh modulation, we omit spectrum spreading due to PN spreading.

2. A higher-rate, parallel-channel transmission mode with identical spreading gain in every channel, which we define as multi-channel, single-gain (PC-SG) mode. In the literature it is also known as Multi-Code (MC) transmission. It is illustrated in Fig. 6.2.

3. A higher-rate, single channel transmission mode with reduced spreading gain, which we define as single-channel, reduced-gain (SC-RG) mode. In the literature this mode of transmission is also known as Variable Spreading Gain (VSG) or Orthogonal Variable Spreading Factor (OVSF). It is illustrated in Fig. 6.3.




Code-division multiplexing of a high rate user using two standard-rate channels.

**Figure 6.2:** A higher-rate input sequence is split into M (=2 in the case) parallel low-rate streams such that the same SF ((=4 in the case) is maintained in all channels.

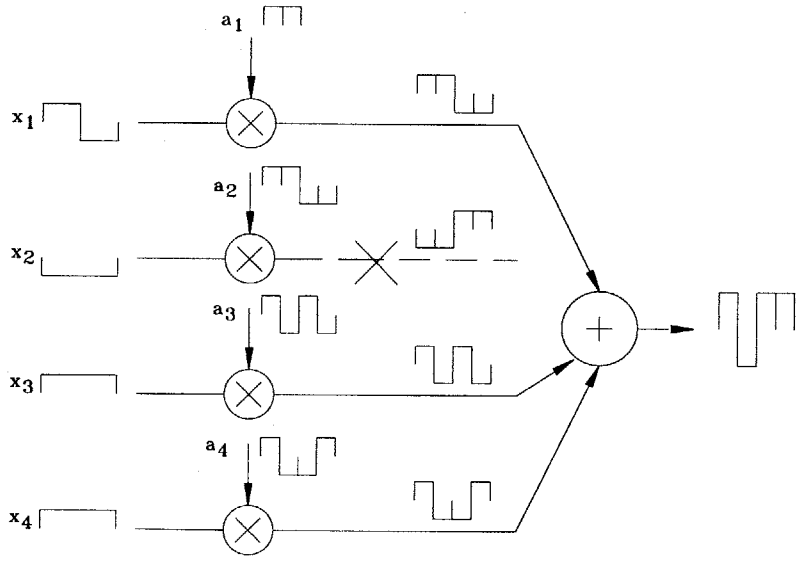



**Figure 6.3:** Code-division multiplexing of a high rate user using a reduced gain channel. A higher-rate input sequence is Walsh modulated using the same code sequence $\mathbf{a}_1=[\mathbf{a}_{11}\ \mathbf{a}_{11}]=[1111]$ as standard CDM. This operation is equivalent to Walsh modulation with a mother code $\mathbf{a}_{11}=[11]$ of reduced SF (=2 in this case). The other child code $\mathbf{a}_2$ is no longer permitted for Walsh modulation due to code blocking. See Chapter 7 for details.

The low SNR condition $R/W < 1$ stipulates that the processing gain $P_G > 1$. Its actual value can vary in the range $(1, N_{max}]$. When bit rate $R_b$ is increased by, say N, we can either pump more power into a single channel (by reducing $P_G$ by N) or split the power among N parallel channels with each maintaining the same $P_G$. A question of interest is whether an increase in transmit power by a factor N compensates for a reduction in $P_G$ by the same factor such that the bit-error probability at the receiving end remains unchanged. This will be addressed in subsequent sections. We first begin with the preliminaries: a CDM system model and spread spectrum modulated signal design.

### 6.1.1 System Model

The functional blocks of a code-division multiplexed transmitter are depicted in Fig. 6.4. The baseband section consists of a bank of channel encoders, data and spread-spectrum modulators. For our analysis, we ignore the channel encoder block since we are not interested in bandwidth expansion due to error-control coding. All BER analyses presented in this chapter are for (channel) uncoded bits.

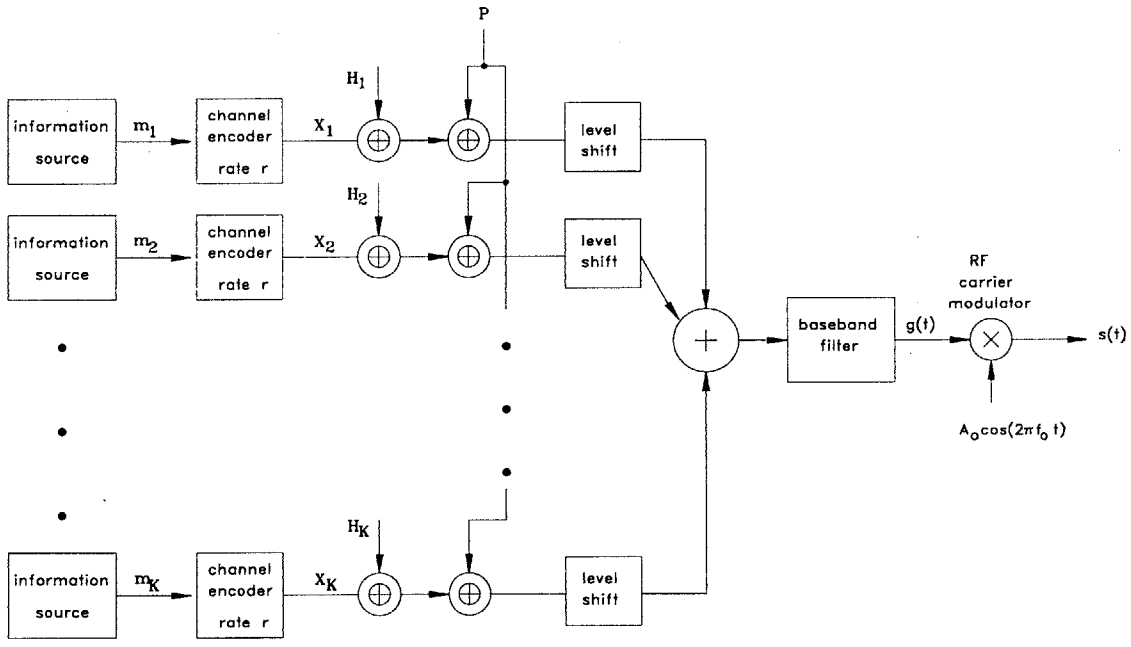



**Figure** 6.4: Channel multiplexing in CDM base transmitter

If the information sequence of user k

$$X_k[j] = \left( \ldots, X_k[-2], X_k[-2], X_k[-1], X_k[0], X_k[1], X_k[2], \ldots \right)$$

is binary-valued with elements from the finite set $\mathbb{F}_2$, then orthogonal modulation is equivalent to binary modulo-2 addition

$$V_k[m] = X_k[j] \oplus \mathcal{H}_k[m]$$

where $\mathcal{H}_k[m]$ is the $k^{th}$ row of an $(N \times N)$ Hadamard matrix and $m$ is the time index. Equivalently, $\mathcal{H}_k[m]$ is treated as a periodic[3] Hadamard code sequence of period N. We use different time indices (j for the data $X_k[j]$ and $m$ for the Hadamard sequences) to emphasize different clock rates such that

$$j = \left\lfloor \frac{m}{N} \right\rfloor$$

which implies that every data bit is modulo-2 added to a Hadamard sequence of length N. If the bit rate is denoted by $R_b$, the clock rate of a train of Hadamard sequences is $R_w = NR_b$. Similarly, spread spectrum modulation is written as

$$B_k[n] = V_k[m] \oplus P[n]$$

where $P[n]$ is a cell specific, periodic pseudo-random noise sequence with clock rate $R_c = nR_b$ and period P. The ratio $k_c = R_c/R_w = n/N$ is bandwidth expansion due to spectrum spreading. The period of the superimposed orthogonal sequence $B_k[n]$ is

$$N_c = k_c N \tag{6.3}$$

Alternatively we can combine orthogonal and PN modulation into a single binary operation:

$$B_k[m] = X_k[j] \oplus A_k[n]$$

where the superimposed orthogonal sequence is defined as

$$A_k[n] = P[n] \oplus \mathcal{H}_k[m]$$

Both interpretations will be useful in the following sections. Before linear combining of parallel SS-modulated sequences, every sequence must first be level-shifted. Following our convention, we denote the antipodal equivalents of all Galois binary sequences (labelled in upper-case letters) by their respective lower-case letters:

$$X_k[n] \longrightarrow x_k[n]$$
$$A_k[n] \longrightarrow a_k[n]$$
$$p[n] \longrightarrow p[n]$$
$$B_k[n] \longrightarrow b_k[n]$$

[3]It is understood that the spreading and modulating discrete-time sequences (PN and Hadamard) are periodic. For example, for a Hadamard sequence, $\mathcal{H}_k[m] = \mathcal{H}_k[lN + m]$ for $l \in \mathbb{Z}$.

Following our convention as stated on page 186, the antipodal equivalent of a Hadamard sequence $\mathcal{H}_k[m]$ is denoted by $\mathbf{w}_k[m]$:

$$\mathbf{w}_k[m] = 2\mathcal{H}_k[m] - \mathbf{1}_{1\times N}$$

If we wish to express a periodic Walsh sequence with the same chip rate, the modified sequence is labelled as

$$\hat{\mathbf{w}}_k[n] = \mathbf{w}_k\left[\left\lfloor \frac{n}{k_c} \right\rfloor\right]$$

such that the superimposed orthogonal sequence of user k is

$$a_k[n] = \hat{\mathbf{w}}_k[n] \cdot p[n]$$

A linearly data modulated signal for an input data sequence $c_k[i]$ can be expressed as

$$g_k(t) = \sum_i c_k[i]\, f(t - iT)$$

where $T = 1/R$. For a spread spectrum modulated signal,

$$f(t) = \sum_{l=0}^{N_c-1} \psi(t - lT_c) \tag{6.4}$$

where the *chip rate* $R_c = 1/T_c$ and

$$c_k[i] = x_k[i] \sum_{j=0}^{N_c} a_k[j + iN_c]$$

Equivalently,

$$c_k\left[\left\lfloor \frac{i}{N_c} \right\rfloor\right] = x_k\left[\left\lfloor \frac{i}{N_c} \right\rfloor\right] a_k[i] \tag{6.5}$$

The expression for a spread-spectrum modulated signal becomes

$$g_k(t) = \sum_i x_k[i] \sum_{r=iN_c}^{(i+1)N_c-1} a_k[r]\, \psi(t - iT - rT_c) \tag{6.6}$$

Using the relation $T = N_cT_c$,

$$= \sum_i x_k[i] \sum_{r=iN_c}^{(i+1)N_c-1} a_k[r]\, \psi\Big(t - \big[iN_c + r\big]\, T_c\Big) \tag{6.7}$$

An equivalent and simpler —nonetheless less revealing— expression is

$$= \sum_l b_k[l]\, \psi(t - lT_c) \tag{6.8}$$

This expression is useful and adequate when we model interfering signals from neighboring base sites. The sum of all $(K_c + 1)$ users is

$$g(t) = \sum_{k=0}^{K_c} g_k(t) = \sum_{l} \psi(t - lT_c) \sum_{k=0}^{K_c} b_k[l] \tag{6.9}$$

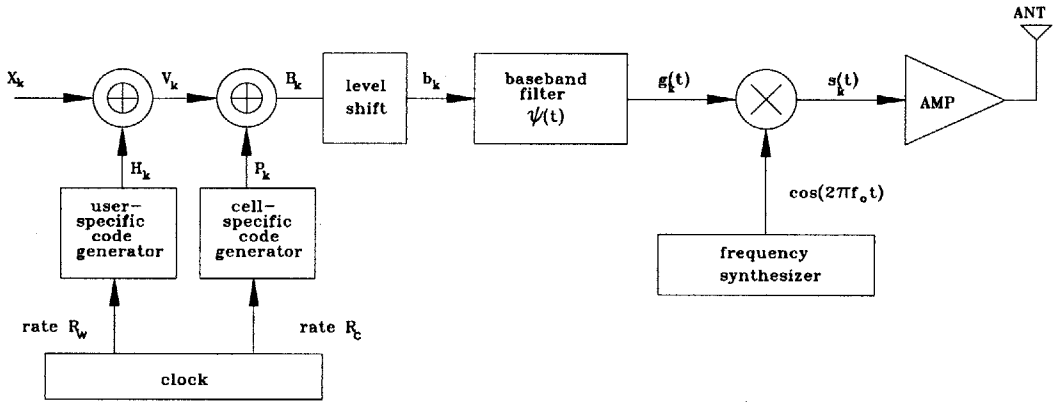



**Figure** 6.5: Data and carrier modulation in CDM transmitter

Each data and spread-spectrum modulated baseband signal belongs to one of the following three classes:

class 1: standard rate channel with processing gain $N_1$ and bit rate R

class 2: higher rate channel with processing gain $N_2 = N_1$ and bit rate MR corresponding to single-gain M parallel channel transmission mode

class 3: standard rate channel with processing gain $N_3 = N_1/M$ and bit rate R corresponding to variable-gain single channel transmission mode

We wish to evaluate the following:

1. The BER of a class 1 user when M class 1 signals plus other interfering signals are present in the system.

2. The BER of a class 1 user when class 2 plus other interfering signals are present in the system. This result will be compared with the BER from Step 1 to determine if a higher rate (PC-SG) signal worsens the reception quality of a standard rate (SC-SG) user.

3. The BER of a class 1 user when class 3 plus other interfering signals are present in the system. This result will be compared with the BER from Step 1 to determine if a higher rate (SC-RG) signal worsens the reception quality of a standard rate (SC-SG) user.

4. The BER of a higher rate class 2 user when other interfering signals are present in the system.

5. The BER of a higher rate class 3 user when other interfering signals are present in the system. This result will be compared to the BER from Step 4 to determine if PC-SG and SC-RG transmission modes are comparable in BER performance.

For class 2 multi-rate reception, the receiver decodes M parallel low-rate orthogonal channels. It is assumed the information in each of the M parallel channels is independently coded and modulated such that joint decoding of all M parallel channels offers no advantage over autonomous decoding of each channel separately. It is plausible that joint encoding of data in M parallel channels may be beneficial, especially in a multipath fading environment where self-interference among these channels exists. (See Sec. 6.5 for an interesting discussion on optimal combining techniques.) We, however, will not consider such option. This assumption is critical since it implies that we need not evaluate Steps 2 and 4; the reception of a class 1 signal or a channel of class 2 signal are indistinguishable. In short, we are concerned only with the interfering effects of the same- and reduced-spreading gain channels on a low- or high-rate user.

## 6.2 BER Analysis of Single-Carrier CDM Channels

If a single IF or RF carrier is used for frequency translation, the transmit signal is

$$\begin{aligned} s(t) &= \sqrt{2P_t}\, g(t) \cos\left(2\pi f_o t\right) \\ &= \sum_{k=0}^{K_c} s_k(t) \end{aligned}$$

where $P_t$ is the transmit power and the carrier modulated signal in each channel can be expressed as

$$s_k(t) = \sqrt{2P_{t*}} \cos\left(2\pi f_o t\right) \sum_j x_k[j] \sum_{r=jN_{c*}}^{(j+1)N_{c*}-1} a_i[r]\, \psi\Big(t - \left[jN_{c*} + r\right]T_c\Big)$$

the subscript $* = 1, 2, 3$ describes the class. Rewriting each class signal explicitly,

- Class 1: Single Channel, Single Gain (SC-SG)

$$s_k(t) = \sqrt{2P_{t1}} \cos\left(2\pi f_o t\right) \sum_j x_k[j] \sum_{r=jN_c}^{(j+1)N_c-1} a_k[r]\, \psi\Big(t - \left[jN_c + r\right]T_c\Big) \qquad (6.10)$$

- Class 2: Parallel Channels, single Gain (PC-SG)

$$s_k(t) = \sum_{m=1}^{M} s_{km}(t)$$

where

$$s_{km}(t) = \sqrt{2P_{t1}}\cos(2\pi f_o t)\sum_j x_{km}[j]\sum_{r=jN_c}^{(j+1)N_c-1} a_{km}[r]\,\psi\Big(t-[jN_c+r]T_c\Big) \tag{6.11}$$

It is understood that the parallel channels use mutually orthogonal code sequences

$$\sum_{r=jN_c}^{(j+1)N_c-1} a_{2k}[r]\,a_{2l}[r] = N_c\,\delta[k-l]$$

- Class 3: Single Channel, Reduced Gain (SC-RG)

$$s_k(t) = \sqrt{2MP_{t1}}\cos(2\pi f_o t)\sum_j x_k[j]\sum_{r=jN_c/M}^{(j+1)(N_c/M)-1} a_k[r]\,\psi\left(t-\left[\frac{jN_c}{M}+r\right]T_c\right) \tag{6.12}$$

In all cases, the bit energy $\mathcal{E}_b = PT_b$ is the same. ($T_b$ is the bit duration.) As a check, we can compute the power in each class. For class 1,

$$\begin{aligned}\text{Transmit Power} &= \frac{1}{T}\int_0^T s_1^2(t)\,dt\\ &= \frac{2P_{t1}}{T}\int\left\{\frac{1+\cos 4\pi f_o t}{2}\right\}\psi^2(t)\,dt\\ &= \frac{P_{t1}}{T}\int_0^T \psi^2(t)\,dt + \frac{P_{t1}}{T}\int_0^T \cos(4\pi f_o t)\,\psi^2(t)\,dt\end{aligned}$$

Due to symmetry the second term is approximately zero for $f_o \gg 1/T$, and the expression simplifies to

$$\text{Transmit Power} = P_{1t}$$

where we assume that the chip pulse is normalized for unit power:

$$\frac{1}{T_c}\int_0^{T_c}\psi^2(t)\,dt = 1$$

It is straightforward to check that for a single channel with reduced gain

$$P_{t3} = \frac{(\sqrt{2M})^2}{2}P_{t1} = M\,P_{t1}$$

For a class 2 channel,

$$P_{t2} = \int s_k^2(t)\,dt = \int \sum_{m=1}^{M}\sum_{n=1}^{M} s_{km}(t)s_{kn}(t)\,dt = \sum_{m=1}^{M}\int s_{km}^2(t)\,dt$$

The double summation is broken into diagonal and off-diagonal terms, and each of the latter terms equals zero since the parallel codes are mutually orthogonal. Therefore, $P_{t2} = M\,P_{t1}$.

Assuming the channel is LTI with unity gain and the receiver front-end thermal noise is modelled as zero-mean AWGN with one-sided PSD equal to $N_o$, the received signal can be expressed as

$$r(t) = s_0(t) + n(t) + I(t)$$

By convention the target signal is denoted as $s_0(t)$, irrespective of the class. The interfering signal $I(t)$ is

$$I(t) = \sum_{i=1}^{J_1} s_i^{(1)}\left(t-\tau_i^{(1)}\right) + \sum_{i=1}^{J_2} s_i^{(2)}\left(t-\tau_i^{(2)}\right) + \sum_{i=1}^{J_3} s_i^{(3)}\left(t-\tau_i^{(3)}\right)$$

where $J_c$ denotes the number of interfering signals belonging to class c. Each interfering signal is classified by a parenthesized superscript. Since the broadcast signal of each site is the sum of time-synchronized signals from parallel channels, they do not cause any interference at the receiver after SS demodulation. In the above equation we are referring to interference from broadcast signals of neighboring cells. Delayed replicas of the target signal (i.e., multipath components) are dealt with later in Sec. 6.5.

### 6.2.1 STD-CDM Interference Channels

We first consider a multi-cell CDM network in which every *interfering* channel has the same standard information rate. Therefore $J_2 = J_3 = 0$ and the target signal $s_0(t)$ belongs to one of three possible classes. For convenience we label the number of interferers as J by dropping the subscript '1'. We first consider the ideal AWGN channel model. A block diagram of a CDM receiver for user 0 detailing the most important functional units is depicted in Fig. 6.6.
At the target receiver the received IF signal is

$$r(t) = s_0(t) + \sum_{i=1}^{J} s_i\left(t-\tau_i\right) + n(t) \tag{6.13}$$

We note that the above expression of a received signal in a multi-cell forward link model is identical to that of uplink reception of a CDMA signal in chip asynchronous mode with J multi-access interfering users. The two significant differences are

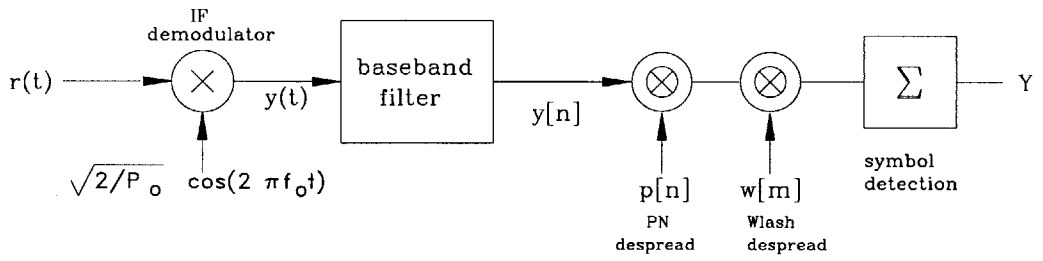



**Figure** 6.6: Carrier, spread-spectrum and data demodulation in a CDM receiver

1. The number of interfering signals J, which is equal to the number of neighboring base sites with strong interference power levels, may be much smaller than the number of interfering signals in a typical CDMA uplink model.

2. In uplink CDMA each interfering signal $s_i(t)$ is binary data and spread-spectrum modulated while in our model, $s_i(t)$ represents a broadcast compound signal —which itself is a linear sum of many elemental binary data and spread-spectrum modulated signals. As we shall see, this fact becomes significant when we consider the BER accuracy of various Gaussian approximation methods.

After IF demodulation,

$$r(t)\sqrt{\frac{2}{P_0}}\cos(2\pi f_o t)$$

$$= g_0(t)\left[1+\cos 4\pi f_o t\right] + \sum_{i=1}^{J}\sqrt{\frac{P_i}{P_0}}\, g_i(t-\tau_i)\left[\cos\phi_i + \cos(4\pi f_o t+\phi_i)\right] + z(t)$$

where

$$z(t) = \sqrt{\frac{2}{P_0}}\, n(t)\cos 2\pi f_o t$$

and the relative phase offset

$$\phi_i = -2\pi f_o \tau_i$$

In our derivation we did not include an absolute carrier phase term $\theta_k$ for each interfering signal, simply for the sake of clarity. Otherwise, the phase-offset term can be rewritten as

$$\phi_k = \theta_k - 2\pi f_o \tau_k$$

After removing double frequency terms by low-pass filtering

$$y(t) = g_0(t) + \sum_{i=1}^{J}\sqrt{\frac{P_i}{P_0}}\cos\phi_i\, g_i(t-\tau_i) + \sqrt{\frac{2}{P_0}}\, n(t)\cos 2\pi f_o t + z(t)$$

We assume the most general scenario of code and phase asynchronism. (The reader is referred to page 171 in Chapter 4 for classification of time synchronism.) Rewriting in terms of discrete-time sequences and common chip pulse

$$= \sum_{l} b_0[l]\,\psi(t - lT_c) \;+\; \sum_{i=1}^{J} \beta_i \sum_{l} b_i[l]\,\psi(t - \tau_i - lT_c) \;+\; z(t)$$

where

$$\beta_i = a_i \cos\phi_i \qquad a_i = \sqrt{\frac{P_i}{P_0}} = \frac{\alpha_i}{\alpha_0} \tag{6.14}$$

Each received power level $P_i$ is related to its respective channel gain $\alpha_i$ as follows:

$$P_i = \alpha_i^2 P_{tl}$$

If the channel is ideal, $\alpha_i = 1$ for all $i \in \mathbb{N}^J$. If we assume every channel induces attenuation only, then the gain $\alpha_i$ is treated as a constant in the range $(0, 1)$ with $\alpha_0 > \alpha_i$ for all $i \neq 0$. In the most general case of signal fading, each $\alpha_i$ is a RV assuming one of possible fading amplitude PDFs listed in Chapter 2.

By setting $\tau_i = q_i T_c + \Delta_i$, and after chip pulse baseband matched filtering

$$\begin{aligned} y[m] &= \int_{mT_c}^{(m+1)T_c} y(t)\,\psi(t - mT_c)dt \\ &= b_0[m]\,T_c \;+\; \sum_{i=1}^{J} \beta_i \sum_{l} \left\{ b_i[l - q_i]\,\rho_\psi + b_i[l - q_i - 1]\,\hat{\rho}_\psi \right\} \;+\; z[m] \end{aligned}$$

where the Gaussian noise sample is

$$z[m] = \sqrt{\frac{2}{P_0}} \int_{mT_c}^{(m+1)T_c} n(t) \cos 2\pi f_o t\,\psi(t - mT_c)dt$$

It is straightforward to show that

$$E\Big[z[m]\Big] = 0 \qquad \mathrm{Var}\Big[z[m]\Big] = \frac{N_o}{2}\frac{T_c}{P_0} \qquad E\Big[z[m]\,z[n]\Big] = 0 \quad m \neq n$$

After despreading for the target zeroth bit,

$$\begin{aligned} Y &= \sum_{m=0}^{N_c-1} y[m]\, a_0[m] \\ &= x_0[0]\, N_c\, T_c \\ &\quad + \sum_{i=1}^{J} \beta_i \left\{ \rho_\psi \sum_{l=0}^{N_c-1} b_i[l-q_i]\, a_0[l] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} b_i[l-q_i-1]\, a_0[l] \right\} \\ &\quad + \sum_{m=0}^{N_c-1} z_I[m]\, a_0[m] \end{aligned}$$

$$\boxed{Y = x_0[0]\, T_b + I + \eta} \tag{6.15}$$

where $\rho_\psi$ and $\hat{\rho}_\psi$ are aperiodic auto-correlation pair of common chip pulse $\psi(t)$. For convenience, we omit their dependence on $\Delta_i$. The interference vector is

$$\boxed{I = \sum_{i=1}^{J} I_i = \sum_{i=1}^{J} \beta_i \left\{ \rho_\psi \sum_{l=0}^{N_c-1} \zeta_i[l] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} \hat{\zeta}_i[l] \right\}} \tag{6.16}$$

The "cross-sequence" terms are

$$\zeta_i[l] = b_i[l-q_i]\, a_0[l] \qquad \hat{\zeta}_i[l] = b_i[l-q_i-1]\, a_0[l] \tag{6.17}$$

The noise term

$$\eta = \sum_{m=0}^{N_c-1} z[m]\, a_0[m] \tag{6.18}$$

is readily shown to be Gaussian distributed with zero mean and variance

$$\sigma_\eta^2 = N_c \frac{N_o}{2} \frac{T_c}{P_0} = \frac{N_o}{2} \frac{T_b}{P_0}$$

Note that the interference vector is a function of the following parameters:

- The total number of interferers J, which is a function of the cellular topology. For example, in a hexagonal configuration there are $6k$ $(6, 12, 18, 24, \ldots)$ $k^{th}$-tier cells surrounding the target cell.
- The propagation delay vector

  $$\Gamma = \left\{ \tau_1, \tau_2, \ldots, \tau_J \right\}$$

  in which each component $\tau_i$ is a function of the (possibly diffracted) distance between the interfering site and the target user modulo $T_b$. Since each signal path of an interferer traverses through a different path, the set $\Gamma$ consists of mutually independent RVs.

- The relative phase-offset vector

$$\Phi = \left\{\phi_1, \phi_2, \ldots, \phi_J\right\}$$

in which each component $\phi_i$ is relative phase difference between interferer $i$ and the target user's phase ($\phi_0 = 0$) modulo $2\pi$. It is a function of the propagation delay $\tau_i$ and the absolute carrier phase $\theta_i$. Since $\tau_i$ and $\theta_i$ are independent, we immediately deduce that the relative phase offset $\phi_i$ and $\tau_i$ are also independent. Since the absolute phases from each site are independent, the set $\Phi$ consists of mutually independent RVs. Their derived function set $\{\cos\phi_i\}$ appears in eqn. (6.16).

- Since $\tau_i = q_i T_c + \Delta_i$, the chip offset $q_i$ is a discrete RV with an integer support $[0, N_c - 1]$; the chip delay $\Delta_i$ is a real RV with support $[0, T_c)$. Each RV $\tau_i$ and its equivalent set $\{\Delta_i, q_i\}$ are interchangeable. For each path $i$, $\Delta_i$ and $q_i$ are independent.

- The spreading factor $N_c$, which is a function of the class of a target user.

- The continuous-time aperiodic auto-correlation function set $\left\{\rho_\psi, \hat{\rho}_\psi\right\}$ is a function of the shape of common chip pulse $\psi(t)$ and the chip delay $\Delta_i$. These two functions are uncorrelated since their correlation windows do not overlap. We already know that the set of chip delays $\{\Delta_i\}$ consists of independent RVs. Therefore, two function sets of independent RVs, $\left\{\rho_\psi(\Delta_i), \hat{\rho}_\psi(\Delta_i)\right\}$ and $\left\{\rho_\psi(\Delta_j), \hat{\rho}_\psi(\Delta_j)\right\}$ are mutually independent.

- Both cross-sequence terms $\zeta_i[l]$ and $\hat{\zeta}_i[l]$ are functions of the discrete chip offset $q_i$. For uncluttered notation, their dependency on $q_i$ has been omitted in above expressions. Both are functions of the target and interfering user's signature sequences $\{a_0[l], b_i[l]\}$. Since $\zeta_i[l]$ and $\hat{\zeta}_i[l]$ are functions of $\{a_0[l], b_i[l], q_i\}$ only, they are independent of $\mathbf{R}_\psi$ for each path $i$.

- The relative power gain $a_i = \sqrt{P_i/P_0}$ is a function of both the channel gain $\alpha_i$ and the transmit power level $P_{ti}$ of each interfering signal; i.e., it is a function of $M$. Since propagation paths are independent, the set $\{\alpha_i\}$ consists of mutually independent RVs. Likewise the set $\{a_i\}$ consists of mutually independent RVs. For each path $i$, the channel parameter set $C_i = \left\{\tau_i, \phi_i, \alpha_i\right\}$ consists of mutually independent RVs. Furthermore, the vector $\mathbf{C} = \{C_i\}$ consists of mutually independent RVs.

Lumping all above relevant terms we define the set of two mutually independent RVs for each interferer $i$

$$G_i = \left\{C_i, b_i[l]\right\} \tag{6.19}$$

such that

$$\mathbf{G} = \left\{G_1, G_2, \ldots, G_J\right\}$$

is the set of independent RVs that depends only on the channel state and interfering users' signature sequences. The target user's signature set $a_0[l]$ is common to all interferers. The power term $P_i$ is a function of a deterministic quantity M (class related) of the interferer while the number $N_c$ depends on the class of the target user.

### 6.2.2 Exact BER Computation

Before we derive the exact BER of received symbol Y in eqn. (6.15), we briefly review basic detection principles. If no interference is present, eqn. (6.13) reverts to the standard signal detection problem, perturbed in independent, zero-mean additive noise process

$$r(t) = s_0(t) + n(t)$$

Its equivalent vector form by geometrical projections (along orthonormal basis functions) is eqn. (6.15) without the interference vector

$$Y = \pm A + \eta$$

where $s_0(t) = \pm Af(t)$ and $f(t)$ is the only orthonormal basis function for 2–PAM binary signals. Assuming data bit $+1$ is sent, it is straightforward to show that for white noise with PSD $\sigma^2$:

$$E\Big[Y|+\Big] = A \qquad Var\Big[Y|+\Big] = \sigma^2 \tag{6.20}$$

where $\sigma^2$ is the variance of $\eta$. The noise samples at bit sampling times are uncorrelated. If the noise process $n(t)$ is Gaussian, then the noise samples mutually independent. If the received bits are ISI-free, the BER for a train of bit sequence is the same as that of one-bit (one-shot) transmission. Then the conditional probability of error is

$$P_{e|+} = Q\left(\sqrt{\frac{A^2}{\sigma^2}}\right)$$

where the *complementary* cumulative distribution function (CDF) of a standard Gaussian RV is

$$Q(x) = 1 - F_X(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2}\, du \tag{6.21}$$

where the RV $X \sim \mathcal{N}(0,1)$. Since we know that the energy of data symbol $s_0(t)$ —carrying one bit of information— is $\mathcal{E}_b = A^2$, and using the expression $\sigma^2 = N_o/2$ for the noise variance, the bit-error probability of a stream of equally likely memoryless input bits is

$$P_{e|+} = P_{e|-} = P_e = Q\left(\sqrt{\frac{2\mathcal{E}_b}{N_o}}\right)$$

Equivalently, using the following relation between the baseband signal power P and the bit rate $R_b$ and the maximum Nyquist symbol rate $R_s$ through an ideal channel of $W$ Hz,

$$\mathcal{E}_b = P R_b \qquad R_s = R_b = 2W$$

Here it is assumed that its symbol rate $R_s$ equals the bit rate $R_b$ since the symbol $s_0(t)$ carries one bit of information every T sec. After substitutions, the bit-error probability can be expressed as

$$P_e = Q\left(\sqrt{SNR}\right) \qquad SNR = P/\mathcal{N} \tag{6.22}$$

where $\mathcal{N} = (N_o W)$ is the noise power. If the noise is not Gaussian distributed, all above expressions still hold, except for the Q-function which is replaced by another suitable complementary CDF function. With the interference inclusive, the received symbol is

$$Y = \pm A + \mathbf{I} + \eta \qquad A = x_0[0]\, T_b$$

Conditioned on a particular set of values $\mathbf{W} = \{a_0[l], \mathbf{G}\}$ for the interference vector $\mathbf{I}$, the conditional error probability can still be expressed as

$$P_{e|\mathbf{W}} = Q\left(\frac{A^*}{\sigma^*}\right)$$

where

$$E\left[Y\middle|\mathbf{W};+\right] = A^* = A + \sum_{i=1}^{J} I_i \qquad Var\left[Y\middle|\mathbf{W};+\right] = (\sigma^*)^2 = N_o/2 \tag{6.23}$$

which leads to an average bit-error probability by conditioning on all possible statistical outcomes of the interference vector:

$$\overline{P_e} = \int Q\left(\sqrt{SNR^*}\right) f_{\mathbf{W}}(\mathbf{x})\, d\mathbf{x} \qquad SNR^* = \left(\frac{A^*}{\sigma^*}\right)^2 \tag{6.24}$$

where $f_{\mathbf{W}}(\mathbf{x})$ is the joint pdf of all J interfering signal parameters. As an alternative, we can compute the exact BER by first determining the joint pdf of the disturbance RV $D = (\mathbf{I} + \eta)$ and then deriving its complementary CDF function. Since the noise and interference terms are statistically independent, we can write the pdf of D as

$$f_D = f_\eta \star f_{\mathbf{I}}$$

Whichever option we choose to compute the BER, it is tantamount that we derive the joint pdf of the interference vector $\mathbf{I}$. First, we note that each interference term $I_i$ is a function of $G_i$. Using our knowledge of the mutually independent set $\mathbf{G}$, the joint pdf of the sum of interferences $\mathbf{I}$ is the linear convolution of the marginal pdfs all components:

$$f_{\mathbf{I}} = f_{I_1} \star f_{I_1} \cdots \star f_{I_J}$$

Equivalently, in terms of the moment-generating functions (MGF)

$$\mathcal{M}_{\mathrm{I}} = \mathcal{M}_{\mathrm{I}_1} \cdot \mathcal{M}_{\mathrm{I}_2} \cdots \mathcal{M}_{\mathrm{I}_J}$$

If all the components are iid, the MGF simplifies to $\left(\mathcal{M}_{\mathrm{I}}\right)^J$. All that remains is to derive an expression of the marginal pdf of each interference term —which we do so next. First, we expand an interference term in eqn. (6.16):

$$\mathrm{I}_i = a_i \cos\phi_i \left\{ \rho_\psi \sum_{l=0}^{N_c-1} \zeta_i[l] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} \hat{\zeta}_i[l] \right\} \tag{6.25}$$

In order to proceed, we assume that every signature sequence is an independent bipolar, symmetric Bernoulli process such that

$$\mathrm{E}\left[ b_i[l]\, b_j[m] \right] = \delta_{ij}\, \delta_{lm} \tag{*}$$

$$a_0[l]\, a_0[m] = \delta_{lm} \tag{**}$$

Since the discrete-time sequences are binary antipodal, the value of their element-wise products —$\zeta_i[l]$ and $\hat{\zeta}_i[l]$— are either 1 or $-1$. Therefore, the product terms are bipolar Bernoulli RVs (see Appendix 2B). Since both outcomes of each RV are equally likely, both RVs are symmetric. Their respective summations

$$\mathrm{R}_{\zeta_i} = \sum_{l=0}^{N_c-1} \zeta_i[l] \tag{6.26}$$

$$\hat{\mathrm{R}}_{\zeta_i} = \sum_{l=0}^{N_c-1} \hat{\zeta}_i[l] \tag{6.27}$$

are symmetric bipolar Binomial RVs, and it can be shown that they are statistically independent. The proof is as follows. The correlation of the sums is

$$\mathrm{E}\left[ \mathrm{R}_{\zeta_i} \hat{\mathrm{R}}_{\zeta_i} \right] = \sum_{l=0}^{N_c-1} \sum_{m=0}^{N_c-1} \mathrm{E}\left[ \hat{\zeta}_i[l] \zeta_i[l] \right] \tag{6.28}$$

The diagonal terms are zero by (*). Similarly, the off-diagonal terms are also zero by (**). We see that each interference term

$$\mathrm{I}_i = a_i \cos\phi_i \left\{ \rho_\psi\, \mathrm{R}_{\zeta_i} + \hat{\rho}_\psi\, \hat{\mathrm{R}}_{\zeta_i} \right\}$$

can be expressed in terms of the power gain RV $a_i$, a derived function $\cos\phi_i$ of phase offset RV, two bipolar Binomial RVs, and two derived functions $\rho_\psi$ and $\hat{\rho}_\psi$ of the chip delay $\Delta_i$. All terms are mutually independent. Note that other RVs $\tau_i$ and $q_i$ are no longer relevant

because we have assumed the sequences are Bernoulli processes. If instead we are dealing with deterministic (known) sequences, they exact values of $\hat{R}_{\zeta_i}$ and $R_{\zeta_i}$ can be determined for a given $q_i$.

In order to derive explicit closed-form expression for the pdf of $f_{I_i}$ we assume the following:

1. The phase offset $\phi_i$ is uniformly distributed in the range $[0, 2\pi)$. We know from Papoulis [97] that the pdf of $R_\phi \triangleq \cos\phi$ is an inverse sinusoidal function:

$$f_{R_\phi}(u) = \begin{cases} \dfrac{1}{\pi\sqrt{1-u^2}} & |u| < 1 \\ 0 & |u| \geq 1 \end{cases}$$

2. The power gain $a_i = \sqrt{P_i/P_0}$ is treated as a deterministic quantity. This assumption is valid if we interpret the channel gain $\alpha_i$ as path loss due to distance propagation. (See Chapter 2, Section 2.1 for details.) If we choose to include signal fading in the model, then $\alpha_i$ is a RV with one of the possible fading distributions outlined in Appendix 2A. We defer the BER analysis of fading models to subsequent sections.

3. Chip delay $\Delta_i$ is uniformly distributed in the range $[0, T_c)$.

4. The chip pulse is rectangular and chip-time limited; thus the expressions for the auto-correlation functions simplify to

$$\lambda \triangleq \hat{\rho}_\psi = \Delta_i$$
$$\lambda' \triangleq \rho_\psi = T_c - \hat{\rho}_\psi = T_c - \Delta_i$$

For the sake of completeness we next derive an alternative expression as proposed by Morrow [89]. First, we manipulate the second summation term as follows

$$\sum_{l=0}^{N_c-1} b_i[l - q_i - 1]\, a_0[l]$$
$$= \hat{\zeta}_i[0] + \sum_{l=1}^{N_c-1} b_i[l - q_i - 1]\, a_0[l]\, a_0[l-1]\, a_0[l-1]$$

Next, define the partial autocorrelation function of length 1

$$\theta_0[1; l, 0] = a_0[l]\, a_0[l-1]$$

For convenience, let's abbreviate it to $\theta_0[l]$. The summation further simplifies to

$$= \hat{\zeta}_i[0] + \sum_{l=1}^{N_c-1} b_i[l-q_i-1]\,\theta_0[l]\,a_0[l-1]$$
$$= \hat{\zeta}_i[0] + \sum_{l=1}^{N_c-1} \zeta_i[l-1]\,\theta_0[l]$$

By a change of variables $l' = l-1$

$$= \hat{\zeta}_i[0] + \sum_{l'=0}^{N_c-2} \zeta_i[l']\,\theta_0[l'+1]$$

We can also rewrite the first summation term as

$$\zeta_i[N_c-1] + \sum_{l=0}^{N_c-2} \zeta_i[l]$$

By combining the two summation terms

$$\frac{I_i}{\beta_i} = \rho_\psi\,\zeta_i[N_c-1] + \hat{\rho}_\psi\,\hat{\zeta}_i[0] + \sum_{l=0}^{N_c-2} \zeta_i[l]\left\{\theta_0[l+1]\,\hat{\rho}_\psi + \rho_\psi\right\}$$

For a set of integers in the range $\mathbb{N}^{N-2}$, let's define

$$\mathbf{U} = \left\{ l \,\middle|\, \theta_0[l+1] = a_0[l]\,a_0[l+1] = 1 \right\}$$

and its complement

$$\mathbf{U}^c = \left\{ l \,\middle|\, \theta_0[l+1] = a_0[l]\,a_0[l+1] = -1 \right\}$$

such that

$$\mathbf{U} \cup \mathbf{U}^c = \mathbb{N}^{N_c-2} = \{0, 1, 2, \ldots, N_c-2\}$$

and

$$\mathcal{U} + \mathcal{U}' = N_c - 1$$

where the cardinalities $\mathcal{U} = |\mathbf{U}|$ and $\mathcal{U}' = |\mathbf{U}^c|$. Rewriting the interference term of eqn. (6.25),

$$I_i = \beta_i\left\{\rho_\psi\,\zeta_i[N_c-1] + \hat{\rho}_\psi\,\hat{\zeta}_i[0] + (\rho_\psi + \hat{\rho}_\psi)\sum_{l\in\mathbf{U}} \zeta_i[l] + (\rho_\psi - \hat{\rho}_\psi)\sum_{l\in\mathbf{U}^c} \zeta_i[l]\right\}$$
$$= \beta_i\left\{\rho_\psi\,\zeta_i[N_c-1] + \hat{\rho}_\psi\,\hat{\zeta}_i[0] + C_\psi\,\Omega + \hat{C}_\psi\,\Omega'\right\}$$

where the even and odd periodic auto-correlation functions are defined as

$$C_\psi = \rho_\psi + \hat{\rho}_\psi \qquad \hat{C}_\psi = \rho_\psi - \hat{\rho}_\psi$$

For rectangular chip-time-limited pulses

$$C_\psi = T_c \qquad \hat{C}_\psi = T_c - 2\Delta_i$$

Note that the RVs

$$\Omega = \sum_{l \in U} \zeta_i[l] \qquad \Omega' = \sum_{l \in U^c} \zeta_i[l]$$

are independent, symmetric bipolar Binomial RVs. Similarly, $\zeta_i[N_c - 1]$ and $\hat{\zeta}_i[0]$ are a pair of mutually independent symmetric, bipolar Bernoulli RVs. Furthermore, it is easily seen that all four terms $\Omega$, $\Omega'$, $\zeta_i[N_c - 1]$ and $\hat{\zeta}_i[0]$ are mutually independent. For rectangular chip-time-limited pulses, the original expression for each interference term is

$$I_i = a_i \cos\phi_i \left\{ \left(T_c - \Delta_i\right) R_{\zeta_i} + \Delta_i \hat{R}_{\zeta_i} \right\}$$

After above transformations, it can be rewritten as

$$a_i \cos\phi_i \left\{ \left(T_c - \Delta_i\right) \zeta_i[N_c - 1] + \Delta_i \hat{\zeta}_i[0] + T_c \Omega + \left(T_c - 2\Delta_i\right) \Omega' \right\}$$

Note that in the original form both Binomial RVs $R_{\zeta_i}$ and $\hat{R}_{\zeta_i}$ are of length $(N_c - 1)$. When their PMFs are conditioned, we must consider all $N_c^2$ distinct outcomes. In the transformed form, the Binomial RVs $\Omega$ and $\Omega'$ have lengths that are dependent according to the relation $\mathcal{U} + \mathcal{U}' = N_c - 1$. For an arbitrary case where the length of $\Omega = x$, then the length of $\Omega' = N_c - 2 - x$. The total number of possible outcomes is $(x + 1) + (N_c - 1) - x = N_c$. Since the value of $x$ spans from 0 to $N_c - 2$, the total number of PMF conditioning equals $N_c(N_c - 1)$. By including the number of conditioning for Bernoulli RVs $\zeta_i[N_c - 1]$ and $\hat{\zeta}_i[0]$, we conclude that there is no difference in the amount of computational complexity between the original and transformed expressions.

## 6.3 Various Gaussian Approximations

As the name implies, every Gaussian approximation technique is a convenient tool (with reduced computational complexity) that computes an approximate BER of a desired CDM received signal perturbed by multi-user interference and zero-mean additive white Gaussian noise. In each variation the interference vector I is modelled as a Gaussian RV to various degrees of conditioning. The conditioning parameters are

1. channel variables: relative propagation delay $\tau_i$, carrier phase offset $\phi_i$, channel gain $\alpha_i$ (mostly due to path loss)

2. system design parameter: the spreading factor $N_c$

3. system design parameters: statistical or deterministic correlation of signature sequences $\{b_i[l]\}$ of interference signals plus the target user's signature sequence $a_0[l]$

Since the impulse response of a cellular radio channel is time-varying random process, its variable set $C_i = \{\tau_i, \phi_i, \alpha_i\}$ is a random vector. Therefore, we can only meaningfully measure the *average* BER of a CDM signal in which the expectation is over the joint pdf of **C**. Fortunately, the random RVs in the set **C** are mutually independent and the averaging of BER is achieved by taking successive expectations over each marginal distribution of $\tau_i$, $\phi_i$ and $\alpha_i$, respectively:

$$\overline{P_e} = \int\int\int P_e(\Gamma, \Phi, \underline{\alpha})\, f_\Gamma(\mathbf{u})\, f_\Phi(\mathbf{v})\, f_{\underline{\alpha}}(\mathbf{w})\, d\mathbf{u}\, d\mathbf{v}\, d\mathbf{w}$$

The remaining problem is the analysis of signature sequences and their correlation properties. In a multiple-access channel model where a single receiver jointly decodes CDMA signals from many scattered sources, it is prudent to assume that the signature sequences are deterministic —since the same signature sequences must be regenerated for despreading. In this model a deterministic value of bracketed terms in eqn. (6.16) can be computed; or as an alternative, we can derive some tight bounds. This has been the paradigm of early research activity in CDMA BER analysis. For the broadcast model where the target receiver has no knowledge of the signature sequences of interfering users, it is more meaningful to assume their sequences are random with certain statistical distributions. The most popular model assumes an interfering signature sequence as a stream of independent, symmetric and bipolar Bernoulli RVs. This may be due to its simplicity and a good statistical fit between a Bernoulli random sequence and a practical binary PN sequence. (See Chapter 4, Sec. 4.6 for details.) Therefore, in every Gaussian approximation technique that we cover next, it is assumed that every signature sequence is a symmetric, bipolar Bernoulli random process.

From the perspective of system analysis, the Gaussian approximation of the interference **I** is very attractive because we have a ton of knowledge in communications theory, specifically to signal detection in additive Gaussian noise. We know that the BER is a non-linear function of a single parameter: the signal-to-noise power ratio SNR, or equivalently, the ratio of bit energy to noise spectral density $\mathcal{E}_b/N_o$. [ In fact, we know that the BER $P_e = Q(\sqrt{SNR})$. ] Complexity of numerical computation is low. However, in the past averaging of the Q-function (say, over the RV $\theta$) is cumbersome because the conditioning RV —the signal-to-noise ratio SNR— appears in the limit on the integral

$$\overline{P_e} = \int Q\Big(\sqrt{SNR(\theta)}\Big)\, f_\theta(u)\, du = \frac{1}{\sqrt{2\pi}} \int\int_{SNR(\theta)}^{\infty} e^{-v^2}\, f_\theta(u)\, dv\, du$$

Here the bar over $P_e$ denotes statistical average. In order to reduce the computational complexity, it is desirable to perform statistical averaging over $SNR(\theta)$ (resulting in $\overline{SNR}$) rather than over $P_e$. As we shall see in this section:

- The *Standard Gaussian Approximation* (SGA) measures the average bit-error rate $\overline{P_e}$ by first computing the mean of SDR over all random parameters $\mathbf{W} = \{\mathbf{C}, b_i[l], a_0[l]\}$. Then

$$\overline{P_e} = Q\left(\sqrt{\overline{SDR}}\right)$$

- The *Conditional Gaussian Approximation* (CGA) measures the average bit-error rate $\overline{P_e}$ in two steps: First, it determines conditional $\overline{P_e}\,\big|\,\{\phi_i\}$ by computing the conditional mean of SNDR over all random parameters $\{\tau_i, \alpha_i, b_i[l], a_0[l]\}$, except for the random phase $\phi_i$. Second, the BER is averaged over the PDF of phase $\phi_i$.

$$\overline{P_e} = \int Q\left(\sqrt{\overline{SDR}}\right) f_{\Phi}(\mathbf{u})\, d\mathbf{u}$$

  Note that the above is a vector integral; i.e., the averaging is performed successively over the PDF of each independent RV $\phi_i$. The merit of singling out the phase term in CGA is explained later in this section.

- The *Improved Gaussian Approximation* (IGA) measures the average bit-error rate $\overline{P_e}$ in two steps: First, it computes the SDR averaged over signature sequences $\{b_i[l]\}$ and $a_0[l]$. Second, the average BER is computed by averaging over the channel random vector $\mathbf{C}$. Hence,

$$\overline{P_e} = \int Q\left(\sqrt{\overline{SDR}}\right) f_{\mathbf{C}}(\mathbf{u})\, d\mathbf{u}$$

  Note that IGA measures an *approximate* BER since averaged SDR (averaged over the statistics of signature sequences) is not conditionally Gaussian. The *exact* average BER is computed by averaging over all random parameters $\mathbf{W}$ of the interference vector $\mathbf{I}$:

$$\overline{P_e} = \int Q\left(\sqrt{SNR}\right) f_{\mathbf{W}}(\mathbf{u})\, d\mathbf{u}$$

Before we give a thorough discussion on each Gaussian approximation method, we wish to point out that nowadays computational complexity that arises from statistical averaging of the Q-function is no longer relevant. Due to faster and more efficient mathematical tools (see Sec. 6.6.5) and powerful computing machines, the exact BER can be computed with moderate effort. However, we are still interested in various Gaussian approximation schemes because we gain some insights into the effect of different random parameters and their intertwined relationships on the overall performance of a CDM system.

### 6.3.1 Standard Gaussian Approximation (SGA)

Referring to eqn.(6.15),

$$Y = x_0[0]\, T_b + (I + \eta)$$

we can combine the interference plus noise term as the disturbance D that disrupts the reception of the desired signal. Then it is easy to show that conditioned of the set of RVs $\mathbf{C} = \{C_i\}$ where $C_i = \{\phi_i, \tau_i, \alpha_i\}$:

$$E\Big[Y\Big|\mathbf{C};+\Big] = T_b = E\Big[Y\Big|+\Big] \tag{6.29}$$

Note the conditioning difference in the above with that of eqn. (6.23). Here the spread-spectrum sequences are treated as a stream of independent (symmetric and bipolar) Bernoulli RVs such that

$$E\Big[I_i\Big|C_i;+\Big] = \beta_i\Big\{\rho_\psi \sum_{l=0}^{N_c-1} E\big[b_i[l-q_i]\big]\, a_0[l] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} E\big[b_i[l-q_i-1]\big]\, a_0[l]\Big\} = 0$$

Using compact notations of eqns.(6.26) and (6.27), we can compute the conditional variance of the disturbance

$$\mathrm{Var}\Big[D\Big|\mathbf{C};+\Big] =$$

$$E\Bigg[\sum_{i=1}^{J}\sum_{j=1}^{J} \beta_i\beta_j\Big\{\rho_\psi R_{\zeta_i} + \hat{\rho}_\psi \hat{R}_{\zeta_i}\Big\}\Big\{\rho_\psi R_{\zeta_j} + \hat{\rho}_\psi \hat{R}_{\zeta_j}\Big\}\Big|\mathbf{C}\Bigg] + \sigma_\eta^2$$

The off-diagonal terms ($i \neq j$) equal zero according to (*), which simplifies the expression to

$$\sum_{i=1}^{J} E\Bigg[a_i^2\,(\cos\phi_i)^2\Big\{\hat{\rho}_\psi^2\,\hat{R}_{\zeta_i}^2 + \rho_\psi^2\,R_{\zeta_i}^2 + 2\,\rho_\psi\hat{\rho}_\psi R_{\zeta_i}\hat{R}_{\zeta_i}\Big\}\Big|\,C_i\Bigg] + \sigma_\eta^2 \tag{6.30}$$

By virtue of the relation in (*) and the symmetry and bipolarity of the desired user's sequence (**), the third term in the bracket (after averaging) is zero. If we treat the power gains $a_i$ as constants among all interferers (with the reference $\alpha_0 = 1$), then $a_i = \alpha_i$ for all interfering signals. For the sake of simplicity we further assume $\alpha_i = \alpha$. Averaging over the set of iid uniformly distributed phase $\phi_i \in [0, \pi)$

$$\mathrm{Var}\Big[D\Big|\tau_i;+\Big] = \frac{\alpha^2}{2}\sum_{i=1}^{J} E\Big[\hat{\rho}_\psi^2\hat{R}_{\zeta_i}^2 + \rho_\psi^2 R_{\zeta_i}^2\Big|\tau_i\Big] + \sigma_\eta^2$$

Note that

$$E\Big[R_{\zeta_i}^2\Big] = \sum_{l=0}^{N_c-1}\sum_{m=0}^{N_c-1} E\Big[[b_i[l-q_i]\,a_0[l][b_i[m-q_i]\,a_0[m]\Big] = N_c$$

by virtue of (*) and (**). Similarly, $E\Big[\hat{R}_{\zeta_i}^2\Big] = N_c$. Averaging over the uniformly distributed RV $\tau_i \in [0, T_b)$, and after some grungy steps

$$\begin{aligned}
\mathrm{Var}\Big[D\Big|+\Big] &= \frac{\alpha^2 N_c}{2}\Bigg[\sum_{i=1}^{J}\sum_{i=0}^{N_c-1}\Big\{\frac{1}{T_b}\int_{iT_c}^{(i+1)T_c}\Big(\Delta_i^2 + (T_c-\Delta_i)^2\Big)\,d(\Delta_i)\Big\}\Bigg] + \sigma_\eta^2 \\
&= \frac{\alpha^2 N_c}{2}\,J\,\frac{1}{T_b}\Big\{\frac{2}{3}N_c T_c^3\Big\} + \frac{N_o}{2}\frac{T_b}{P_0} \\
&= \frac{\alpha^2 J}{3N_c}T_b^2 + \frac{N_o}{2}\frac{T_b}{P_0}
\end{aligned} \tag{6.31}$$

The average *received* signal-to-disturbance power ratio can be expressed as

$$\overline{\mathrm{SDR}}(\alpha) = \left( \frac{T_b^2}{\frac{\alpha^2 J}{3N_c} T_b^2 + \frac{N_o}{2} \frac{T_b}{P_0}} \right)$$

$$= \left( \frac{\alpha^2 J}{3N_c} + \frac{N_o}{2\mathcal{E}_b} \right)^{-1} \tag{6.32}$$

where the target user's *received bit energy* $\mathcal{E}_b = P_0 T_b$. Without any interference, the above expression reverts to the conventional Gaussian model with $\mathrm{SNR} = 2\mathcal{E}_b/N_o$. A more insightful expression of $\overline{\mathrm{SDR}}$ is in terms of the *received chip energy* $\mathcal{E}_c = \mathcal{E}_b/N_c$:

$$\overline{\mathrm{SDR}}(\alpha) = \frac{N_c \mathcal{E}_c}{\alpha^2 J \mathcal{E}_c/3 + N_o/2}$$

For each chip detection, spread-spectrum modulation with a spreading factor $\mathrm{SF} = N_c$ offers a power gain of $N_c$ over other interfering signals, but has no effect on additive Gaussian noise. Note that the *effective chip energy* of each interferer is $\mathcal{E}_c/3$. This reduction in the total number of interference by a factor of 3 is a direct result of chip and phase asynchronism. It is straightforward to show that

- if all interfering chip pulses are aligned, this factor is reduced to 2. This scenario applies to both bit and code synchronism.
- If only the phase offsets are all equal to zero (assuming there is chip asynchronism), this factor reduces to 2/3.
- if all interfering chip pulses are aligned and the phase offset is zero for all signals, this factor is reduced to 1. This condition of chip and phase synchronism is highly unlikely in a multi-cell interference setting. However, it perfectly mimics the broadcast model where code and phase synchronism occurs. Therefore, it is important that codes with very low cross-correlation properties be chosen in a broadcast setting. For that matter orthogonal codes are optimal.

The *Standard Gaussian Approximation* (SGA) uses the signal-to-disturbance ratio $\mathrm{SDR}(\alpha)$, *averaged* over all the statistics of the channel $\{\tau_i, \phi_i\}$ plus the interference signals $\{b_i[l]\}$ and the desired user's sequence $a_0[l]$ to compute the bit-error rate:

$$P_e = Q\left(\sqrt{\overline{\mathrm{SDR}}(\alpha)}\right) \tag{6.33}$$

If we wish to treat the channel gains $\{\alpha_i\}$ as RVs, then the above error probability can be averaged over their statistics. As obvious from our derivation, the accuracy of SGA depends on

the validity of the Gaussianity of the interference vector **I**. That is, the Gaussian approximation of **I** rests on the applicability of the *Central Limit Theorem* (CLT) which states (loosely according to Feller [34]) that the cumulative distribution function (CDF) of a sum of J independent, well-behaved RVs approaches that of a Gaussian RV for large J. All elemental RVs in the sum are well-behaved if none has a variance much larger than the rest. In a general setting it is not necessary that the RVs have the same distribution, but it is sufficient to guarantee the well-behaved stipulation. Rewriting the interference vector explicitly

$$\mathbf{I} = \sum_{i=1}^{J} I_i = \sum_{i=1}^{J} a_i \cos\phi_i \left\{ \rho_\psi \sum_{l=0}^{N_c-1} b_i[l-q_i]\, a_0[l] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} b_i[l-q_i-1]\, a_0[l] \right\}$$

we can deduce the following:

- For large J, the CLT guarantees that the interference vector **I** is approximately Gaussian distributed as long as every interfering term $I_i$ is well-behaved; i.e., there is no term with a dominantly large $\alpha_i$. The numerical value of the spreading factor $N_c$ or the statistical distributions of the embedded RVs $\alpha_i$, $\phi_i$, $\tau_i$ and $\{b_i[l]\}$ are less critical.

- For the special case of chip and phase synchronism (see page 171 in Chapter 4 for classification of time synchronism), $\hat{\rho}_\psi = 0$, $\rho_\psi = T_c$ and $\cos\phi_i = 1$, and the interference vector simplifies to

$$\mathbf{I} = \sum_{i=1}^{J} I_i = \sum_{i=1}^{J} \alpha_i T_c \sum_{l=0}^{N_c-1} b_i[l-q_i-1]\, a_0[l]$$

  If all interfering signals are well-behaved, i.e., if power control guarantees that $\alpha_i^2 \simeq \alpha^2$ for all i, then **I** is a nested summation of $(N_c \cdot J)$ symmetric, bipolar Bernoulli RVs. Note that the symmetry condition implies that the resulting Binomial RV **I** has zero mean and its pdf is symmetric about the origin. It is well known that a Binomial pdf approximates a Gaussian bell curve. However, since the Binomial RV has finite support (in contrast to a Gaussian RV with infinite support), the approximation in CDF is more accurate for large lengths. If, however, there exists a dominant interferer with a large $\alpha_i$, the Gaussian approximation fails because the well-behaved stipulation of the CLT is violated —irrespective of the length $(N_c \cdot J)$ of the Binomial RV[4] It has been shown by other researchers by computer simulation plots that the SGA fails when interfering power levels are non-uniform. Note that for chip and phase synchronism with power control, it is a large value of the product $(N_c \cdot J)$ that attains the accuracy of the standard Gaussian approximation model, and not the value of each term $N_c$ or J.

[4]The *near-far effect* in which a high-power interfering signal overpowers the desired signal is subtle in a multi-user environment. As we have seen in the BER expression based on SGA, it is the *total* power from *all* interfering signals that is critical, not the power level of a particular interferer. However, the BER based on SGA is less accurate (i.e., more optimistic than the actual value) when a dominant interfering signal exists.

- If there is chip synchronism only, the interference vector is

$$I = \sum_{i=1}^{J} I_i = T_c \sum_{i=1}^{J} a_i \cos\phi_i \sum_{l=0}^{N_c-1} b_i[l - q_i - 1]\, a_0[l]$$

Each Binomial RV in the inner summation has a shorter length ($N_c$) and it is weighted by two other RVs $\alpha_i$ and $\cos\phi_i$. Since $(\cos\phi_i)^2 \leq 1$, it has minimal effect on the variance of each interference term $I_i$, under the condition that the channel power gains $\alpha_i$ are well controlled. However, we know that if $\phi_i$ is uniformly distributed in $[0, \pi)$, its function $\cos\phi_i$ has a arc-sine density function. Therefore, if X is a symmetric, bipolar Binomial RV, the resulting RV $X\cos\phi_i$ is no longer Binomial. In fact, its density function is no longer symmetric or bell-shaped. Furthermore, it is well-known that the sum of independent RVs whose PDFs are smooth and symmetric approximate the Gaussian CDF and its bell-shaped PDF faster as the number of terms in the sum grows. We conclude that with phase asynchronism, a large value of J (along with power control) is necessary for the accuracy of the standard Gaussian approximation model. Note that in this situation, the value of $N_c$ is of less significance.

- Finally, we note that based on the observation from the previous case, the role of the spreading factor $N_c$ is quite limited in affecting the accuracy of the standard Gaussian approximation model. For the typical case of chip asynchronism and power control, the high value of $N_c$ cannot compensate for a low number of interferers (small J). This situation, where the SGA fails and the resulting BER becomes optimistic— has been noted by many researchers who have plotted the SGA-based BER for various values of $N_c$. As for as we are concerned, it implies that the standard Gaussian model will not detect any difference in BER between a class 2, high-rate (PC–SG) target user (with a fixed $N_c$) and a class 3, high-rate (SC–RG) target user (with a reduced $N_c$).

### 6.3.2 Conditional Gaussian Approximation (CGA)

In the previous section we hinted that the product

$$V_i = X_i \cos\phi_i$$

in which $X_i$ is a bipolar symmetric Binomial RV and $\phi_i$ is a uniformly distributed RV, has an asymmetric distribution. Therefore, the joint pdf of the sum of several $V_i$'s:

$$V_1 + V_2 + V_3 + \ldots$$

which is the convolution of the marginal pdfs, does not resemble a "bell-shaped" curve that typifies the Gaussian density function —unless the number of terms in the sum is very large.

Based on this knowledge, we can afford to compute average SNR without averaging over the distribution of the phase vector $\Phi$. From the previous section,

$$\mathrm{Var}\Big[\mathrm{D}\Big|\mathbf{C};+\Big] = \sum_{i=1}^{J} \mathrm{E}\left[a_i^2(\cos\phi_i)^2\left\{\hat{\rho}_\psi^2\,\hat{R}_{\zeta_i}^2 + \rho_\psi^2\,R_{\zeta_i}^2\right\}\Big|\,C_i\right] + \sigma_\eta^2$$

Again, assuming the channel gain $\alpha_i$ is a fixed parameter $\alpha$, and using the randomness property of the signature sequences according to (*) and (**),

$$= N_c\,\alpha^2 \sum_{i=1}^{J} \mathrm{E}\left[(\cos\phi_i)^2\left\{\hat{\rho}_\psi^2 + \rho_\psi^2\right\}\Big|\,C_i\right] + \sigma_\eta^2$$

Averaging over iid propagation delays $\{\tau_i\}$ and after some massaging

$$\mathrm{Var}\Big[\mathrm{D}\Big|\Phi;+\Big] = \frac{2}{3}\frac{\alpha^2 T_b^2}{N_c}\sum_{i=1}^{J}(\cos\phi_i)^2 + \frac{N_o}{2}\frac{T_b}{P_0}$$

which leads to the following expression for average $\mathrm{SDR}(\alpha,\Phi)$:

$$\overline{\mathrm{SDR}}(\alpha,\Phi) = \frac{N_c\,\mathcal{E}_c}{\left(\frac{2\mathcal{E}_c}{3}\right)\alpha^2\cos^2\Phi + N_o/2} \qquad \cos^2\Phi = \sum_{i=1}^{J}(\cos\phi_i)^2$$

Assuming the phase offsets are iid with a uniform PDF,

$$\overline{P_e} = (2\pi)^{-J}\int_{\phi_1=0}^{2\pi}\int_{\phi_2=0}^{2\pi}\cdots\int_{\phi_J=0}^{2\pi} Q\left(\sqrt{\overline{\mathrm{SDR}}(\alpha,\Phi)}\right)\mathrm{d}\phi_1\,\mathrm{d}\phi_2\ldots\mathrm{d}\phi_J \tag{6.34}$$

### 6.3.3 Improved Gaussian Approximation (IGA)

In the previous section we showed that the average BER based on conditional Gaussian approximation as

$$\overline{P_e} = \int Q\left(\sqrt{\overline{\mathrm{SDR}}(\alpha,\Phi)}\right) f_\Phi(\mathbf{u})\,\mathrm{d}\mathbf{u}$$

If we condition the interference vector $\mathbf{I}$ over the statistics of $\{b_i[l]\}$ and $a_0[l]$ only, the average BER based on the improved Gaussian approximation is

$$\overline{P_e} = \int Q\left(\sqrt{\overline{\mathrm{SDR}}(\mathbf{C})}\right) f_C(\mathbf{u})\,\mathrm{d}\mathbf{u}$$

the conditional signal-to-disturbance ratio is

$$\overline{\mathrm{SDR}}(\mathbf{C}) = \frac{T_b^2}{N_c\,Z + \sigma_\eta^2}$$

where

$$\mathbf{Z} \triangleq \mathrm{Var}\Big[\mathbf{I}\Big|\mathbf{C}\Big] = \sum_{i=1}^{J} \mathrm{E}\left[a_i^2\,(\cos\phi_i)^2\Big\{\hat{\rho}_\psi^2 + \rho_\psi^2\Big\}\,\Big|\,C_i\right]$$

From probability theory we know that if a RV g is a function of another RV X, the mean of g can be computed in two ways:

$$\mathrm{E}\Big[g(X)\Big] = \int u\, f_g(u)\, du = \int g(x)\, f_X(x)\, dx$$

Some authors coin this equivalence the *Law of Expectation*. Using this analogy, we can rewrite the average BER as

$$\overline{P_e} = \int Q\left(\frac{T_b}{\sqrt{N_c\,\mathbf{Z} + \sigma_\eta^2}}\right) f_{\mathbf{Z}}(\mathbf{u})\, d\mathbf{u} \tag{6.35}$$

Explicit derivation of the above integral can be found in [89],[74]. In a more general setting, we can compute the variance of **I** by conditioning over the statistics of $a_0[l]$ only. Note that

$$\mathrm{E}\Big[I_i\Big|C_i;+\Big] = \beta_i\left\{\rho_\psi \sum_{l=0}^{N_c-1} \mathrm{E}\big[b_i[l-q_i]\,a_0[l]\big] + \hat{\rho}_\psi \sum_{l=0}^{N_c-1} \mathrm{E}\big[b_i[l-q_i-1]\,a_0[l]\big]\right\} = 0$$

as long as either $b_i[l]$ or $a_0[l]$ is treated as a symmetric, bipolar Bernoulli process. In this case the conditional variance is

$$\begin{aligned}\mathbf{Z}' &\triangleq \mathrm{Var}\Big[\mathbf{I}\Big|\mathbf{C}, b_i[l]\Big] \\ &= \sum_{i=1}^{J} \mathrm{E}\left[a_i^2\,(\cos\phi_i)^2\Big\{\hat{\rho}_\psi^2\,\hat{R}_{\zeta_i}^2 + \rho_\psi^2\,R_{\zeta_i}^2 + 2\rho_\psi\hat{\rho}_\psi\,\theta_i[1; l-q_i, N_c-1]\Big\}\,\Big|\,C_i, b_i[l]\right]\end{aligned}$$

The above equation is *almost* identical to eqn.(6.30) except for the third (cross-product) term. It can be shown that

$$\begin{aligned}&\mathrm{E}\Big[\rho_\psi\hat{\rho}_\psi R_{\zeta_i}\hat{R}_{\zeta_i}\,\Big|\,C_i, b_i[l]\Big] \\ &= \mathrm{E}\Big[\rho_\psi\hat{\rho}_\psi\,\Big|\,C_i\Big] \sum_{l=0}^{N_c-1}\sum_{m=0}^{N_c-1} \mathrm{E}\Big[a_0[m]a_0[l]\,b_i[l-q_i]b_i[l-q_i-1]\,\Big|\,C_i, b_i[l]\Big] \\ &= \mathrm{E}\Big[\rho_\psi\hat{\rho}_\psi\,\Big|\,C_i\Big] \sum_{l=0}^{N_c-1} \mathrm{E}\Big[b_i[l-q_i]b_i[l-q_i-1]\,\Big|\,C_i, b_i[l]\Big] \\ &= \mathrm{E}\Big[\rho_\psi\hat{\rho}_\psi\,\theta_i[1; l-q_i, N_c-1]\,\Big|\,C_i, b_i[l]\Big]\end{aligned}$$

The off-diagonal terms vanish due to (**). Unlike eqn.(6.28), the diagonal terms do not vanish since the condition (*) does not hold. Unless the signature sequence $b_i[l]$ is memoryless, the partial auto-correlation term $\mathrm{E}\Big[\theta_i[1; l-q_i, N_c-1]\,\Big|\,b_i[l]\Big] \neq 0$. We know from Chapter 4 that

all periodic PN sequences have memory (initial states). Hence, for known PN sequences, the conditioning of the variance of I over $\{C_i, b_i[l]\}$ is a formidable task. If the PN sequences are unknown at the target receiver, we must then supply statistical properties (such as auto- and cross-correlation) in order to evaluate the conditional variance of I. We therefore stress an important fact that the randomness (Bernoulli) property of the signature sequences simplifies the derivation of all Gaussian approximation techniques considerably. Otherwise, without such closed-form expressions for the average BER, it is rather difficult to deduce any meaningful insights into the role of channel parameters $\{\tau_i, \phi_i, \alpha_i\}$ on the BER performance in AWGN channel. The numerical results of above Gaussian approximation schemes are presented in Sec. 6.6.5. Next, we analyze the BER performance when the channel induces multipath and fading.

## 6.4 BER Analysis in Flat Fading Channel

If the channel induces signal fading, the received signal as expressed in eqn.(6.13) can be rewritten as

$$r(t) = \alpha_0 s_0(t) + \sum_{i=1}^{J} \alpha_i s_i(t - \tau_i) + n(t)$$

After demodulation and despreading the target zeroth bit is

$$Y = \alpha_0 x_0[0] T_b + \mathbf{I} + \eta' \tag{6.36}$$

The explicit expression of the interference vector $\mathbf{I}$ is still the same as eqn.(6.16). The Gaussian noise term $\eta'$ is zero mean with variance

$$\sigma^2_{\eta'} = \frac{N_o}{2} \frac{T_b}{P_{t1}}$$

Note that this noise term $\eta'$ (without normalization by $\alpha_0$) is not the same as $\eta$. However, they are related:

$$\sigma^2_{\eta'} = \alpha_0^2 \frac{N_o}{2} \frac{T_b}{P_0} = \alpha_0^2 \sigma^2_{\eta}$$

It is straightforward to show that applying SGA

$$E\left[Y \middle| \{\alpha_i\}; +\right] = \alpha_0 T_b \tag{6.37}$$

$$\mathrm{Var}\left[D \middle| \{\alpha_i\}; +\right] = \frac{T_b^2}{3N_c} \sum_{i=1}^{J} \alpha_i + \sigma^2_{\eta'} \tag{6.38}$$

such that the signal-to-disturbance ratio conditioned on fading amplitudes is

$$\overline{\mathrm{SDR}}(\{\alpha_i\}) = \frac{N_c \alpha_0^2 \mathcal{E}_c'}{\frac{\mathcal{E}_c'}{3} \cdot \sum_{i=1}^{J} \alpha_i^2 + \frac{N_o}{2}}$$

Note that $\mathcal{E}_c'$ denotes the *transmit* chip energy, which is identical for all users in the network. At the target receiver 0, the received chip energy of an interferer i is $\mathcal{E}_{ci} = \alpha_i^2 \mathcal{E}_c'$. After conditioning on fading statistics, the average BER based on SGA is

$$\overline{P_e} = \int_{\alpha_0=0}^{\infty} \int_{\alpha_1=0}^{\infty} \cdots \int_{\alpha_J=0}^{\infty} Q\left(\sqrt{\overline{SDR}(\{\alpha_i\})}\right) d\alpha_0\, d\alpha_1 \ldots d\alpha_J \tag{6.39}$$

In compact form

$$\overline{P_e} = \int \overline{P_e}(\underline{\alpha}) f_{\underline{\alpha}}(\mathbf{u})\, d\mathbf{u}$$

where the channel gain vector

$$\underline{\alpha} = \{\alpha_0, \alpha_1, \ldots, \alpha_J\}$$

We can readily derive the average BER based on CGA

$$\overline{P_e} = \int \overline{P_e'}(\underline{\alpha}) f_{\underline{\alpha}}(\mathbf{u})\, d\mathbf{u}$$

where the fade-conditioned BER $\overline{P_e'}(\underline{\alpha})$ is similar to the expression in eqn.(6.34) except the conditional SDR is

$$\overline{SDR}(\underline{\alpha}, \Phi) = \frac{N_c \alpha_0^2 \mathcal{E}_c'}{\frac{2\mathcal{E}_c'}{3} \sum_{i=1}^{J} \alpha_i^2 \cos^2 \phi_i + \frac{N_o}{2}}$$

Similarly, we can derive the BER based on IGA

$$\overline{P_e} = \int \overline{P_e''}(\underline{\alpha}) f_{\underline{\alpha}}(\mathbf{u})\, d\mathbf{u}$$

where the fade-conditioned BER $\overline{P_e''}(\underline{\alpha})$ is similar to the expression in eqn.(6.35) except for a slight modification

$$\overline{P_e''}(\underline{\alpha}) = \int Q\left(\frac{\alpha_0 T_b}{\sqrt{N_c \mathbf{Z} + \sigma_{\eta'}^2}}\right) f_Z(\mathbf{u})\, d\mathbf{u}$$

In Chapter 2 we noted that the PDFs of fading power (amplitude squared) tend to have skewed shapes. We also noted that scaling of a Binomial RV by $\cos^2 \phi$ destroys the symmetry of Binomial PDF. Here we have an interesting situation where every interference term $I_i$ is scaled by both $\cos^2 \phi_i$ and $\alpha_i^2$. Without explicit derivation of the resulting PDF, it is difficult to speculate on the Gaussianity of each interference term's PDF. On the other hand, the signal mean $E[Y|\alpha_0; +]$ itself (scaled by $\alpha_0$) is a random variable. Averaging the conditional BER over the PDF of $\alpha_0$ tends to "smoothen" the plot; i.e., several authors have shown that the inaccuracy gap between the exact conditional BER and those based on Gaussian approximations is narrowed by conditioning over $\alpha_0$. Regardless of the accuracy of Gaussian methods, we note that CDM BER analysis with flat fading is interesting only in academic sense since most RF channels for wideband CDM communications are frequency-selective.

## 6.5 BER Analysis in Frequency Selective Fading Channel

Before we analyze BER performance of CDM signal in frequency selective fading, we briefly review the requirement and merit of diversity combining. If the same signal $s(t)$ is transmitted over L logical channels, the output of logical channel $i$ is

$$s_i(t) = s(t) \star f_i(t)$$

where $f_i(t)$ is the transmit filter response of channel $i$. This type of communication is called *multi-channel*. The received complex[5] baseband signal can be expressed as

$$r(t) = \sum_{l=1}^{L} \alpha_l \, e^{-j\phi_l} \, s_l(t - \tau_l) + n(t) \tag{6.40}$$

The baseband receiver, as shown in Fig. 6.7 consists of a bank of L matched filters. We will coin each component receiver a "finger." The total number of channels L is called the *diversity order*. Initially assume the set of channel gains and carrier phase offsets $\{\alpha_i, \phi_i\}$ as fixed parameters known or can be estimated accurately by the receiver.

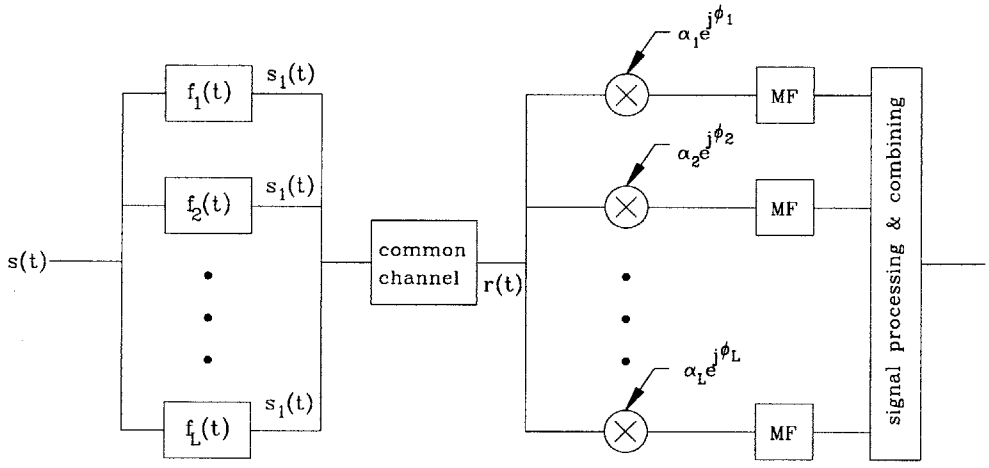



**Figure** 6.7: Diversity combining of multi-channel signal

The sampled output of finger $i$ can be expressed as

$$\begin{aligned} Q_i[n] &= \Re\left[\int_{t=nT+\tau_i}^{(n+1)T+\tau_i} \alpha_i \, e^{j\phi_i} \, r(t) s_i(t - \tau_i) \, dt\right] \\ &= \alpha_i^2 \mathcal{E}_s + \sum_{l \neq i} \alpha_i \alpha_l \cos(\phi_i - \phi_l) \, C_{il}(\tau) \; + \; z_i[n] \end{aligned} \tag{6.41}$$

[5]For our analysis, we assume $s(t)$ carries binary information, implying it is real.

where $\mathcal{E}_s$ is the symbol energy. The periodic cross-correlation[6] function is

$$C_{li}(\tau) = \int s_l(t-\tau_l)s_i(t-\tau_i)\,dt \qquad \tau = \tau_{il} = \tau_i - \tau_l$$

With a slight abuse in notation, we omit the subscripts by shortening the index to $\tau$ instead of $\tau_{il}$. The additive noise sample is

$$z_i[n] = \Re\left[\alpha_i e^{j\phi_i} \int_{t=nT+\tau_i}^{(n+1)T+\tau_i} n(t)s_i(t-\tau_i)\,dt\right] \tag{6.42}$$

For convenience we drop the brackets by considering the zeroth sample. Rewriting:

$$Q_i = \alpha_i^2 \mathcal{E}_s + \mathcal{S}_i + z_i$$

The second term is the *self-induced* (SI) noise generated at the output of finger i:

$$\mathcal{S}_i = \sum_{l\neq i} \alpha_i\alpha_l \cos(\phi_i - \phi_l)\, C_{il}(\tau) \tag{6.43}$$

The logical channels can be separated in frequency (FDM mode) or time (TDM mode). By careful design (as we detailed in Ch. 4, Sec. 4.8) the cross-correlation function can be set to equal zero. If the additive noise process is Gaussian, then the finger outputs $\{Q_i\}$ are Gaussian RVs. If the noise process is white, then the outputs are uncorrelated. Finally, if n(t) is AWGN, the outputs are iid Gaussian RVs. The combined output of the fingers (assuming SI noise is zero) is

$$Q = \sum_{l=1}^{L} Q_l = \mathcal{E}_s \sum_{l=1}^{L} \alpha_l^2 + z \qquad z = \sum_{l=1}^{L} z_l \tag{6.44}$$

from which

$$E[Q] = \mathcal{E}_s \sum_{l=1}^{L} \alpha_l^2 \qquad \mathrm{Var}[Q] = \mathrm{Var}[z]$$

It is well known that for multi-channel communication, this type of diversity combining receiver —where each finger "matches" not only the waveform $s_i(t)$ but also the gain $\alpha_i$ and phase shift $\phi_i$— maximizes the output signal-to-noise ratio if the outputs (i.e., noise samples) are independent. Furthermore, if the noise samples are Gaussian, it has been shown to be optimal —in minimum BER sense. Assume n(t) is zero-mean AWGN with one-sided PSD of $N_o$. The noise samples $z_1, z_2, \ldots$ at finger outputs are mutually independent with zero mean and variance

$$\mathrm{Var}[z_i] = \alpha_i^2 \mathcal{E}_s \frac{N_o}{2}$$

[6]It is true that both $s_l(t)$ and $s_i(t)$ are channelized versions of the same signal s(t). For current interest, when we use the term "cross" in defining the correlation function, we are referring to two output signals from different channels.

The signal-to-noise ratio of the combiner output is

$$\mathrm{SNR}_o = \frac{2\mathcal{E}_s}{N_o} \sum_{i=1}^{L} \alpha_i^2$$

Since Q is a Gaussian RV, the BER is

$$P_e = Q\left(\sqrt{\mathrm{SNR}_o}\right)$$

If the channels subject no attenuation and the transmit power is split among channels such that $\alpha_i^2 \mathcal{E}_s$ is the signal energy in channel i, then $\sum_{i=1}^{L} \alpha_i^2 = 1$. The BER under this scenario is the same as single-channel communication with transmit energy $\mathcal{E}_s$. Thus, multi-channel communication offers no advantage over its single-channel counterpart. Of course, it is well known that multi-channel communication is used to improve overall link quality in a case where any or all of the component channels are not reliable at all times —i.e., each is susceptible to high bit error rates due to noise, interference or signal fading. In single-channel communication with (flat) fading, the conditional BER (conditioned on fading amplitude $\alpha$) is

$$P_e^{(s)}(\alpha) = Q\left(\sqrt{\alpha^2 \frac{2\mathcal{E}_s}{N_o}}\right)$$

Averaging over the fade statistic, the average BER is

$$P_e^{(s)} = \int Q\left(\sqrt{u^2 \frac{2\mathcal{E}_s}{N_o}}\right) f_\alpha(u)\, du$$

For example, if $\alpha$ is a Rayleigh RV with mean-squared value $E[\alpha^2] = \overline{\alpha^2}$,

$$P_e^{(s)} = \frac{1}{2}\left(1 - \Pi\right) \qquad \Pi = \sqrt{\frac{\mu}{1+\mu}}, \quad \mu = \frac{\mathcal{E}_s}{N_o}\overline{\alpha^2}$$

On the other hand, in multi-channel communication with iid fade amplitudes, it can be shown that

$$P_e^{(m)} = \left[\frac{1}{2}\left(1 - \Pi\right)\right]^L C[L;\Pi] \qquad C[L;\Pi] = \sum_{k=0}^{L-1} \binom{L-1+k}{k} \left[\frac{1}{2}\left(1+\Pi\right)\right]^L$$

For the purpose of comparison, in high SNR region where $\mu \gg 1$,

$$P_e^{(s)} \approx \frac{1}{4\mu} \qquad P_e^{(m)} \approx \left(\frac{1}{4\mu}\right)^L \binom{2L-1}{L} \tag{6.45}$$

It is apparent that when compared to single-channel mode, the BER of a multi-channel signal falls faster at a steeper slope that is directly proportional to the number of channels L. In the limit[7]

$$P_e^{(m)} \to Q\left(\sqrt{\frac{2\mathcal{E}_s}{N_o}}\right) \quad \text{as} \quad L \to \infty$$

[7]The BER does not approach 0 as $L \to \infty$ since energy is conserved; i.e., the energy per channel is $\mathcal{E}_s/L$.

### 6.5.1 Rake Receiver

In spread-spectrum modulated communication via a multipath fading channel, each path can be modelled as a distinct logical channel with independent fade statistic $\{\alpha_i, \phi_i, \tau_i\}$. Historically, a multipath diversity combiner for spread-spectrum modulated signal is called a Rake receiver. We briefly outline several important properties of the Rake combiner. The background material is referenced mostly from tutorial papers by Turin [143], [146].

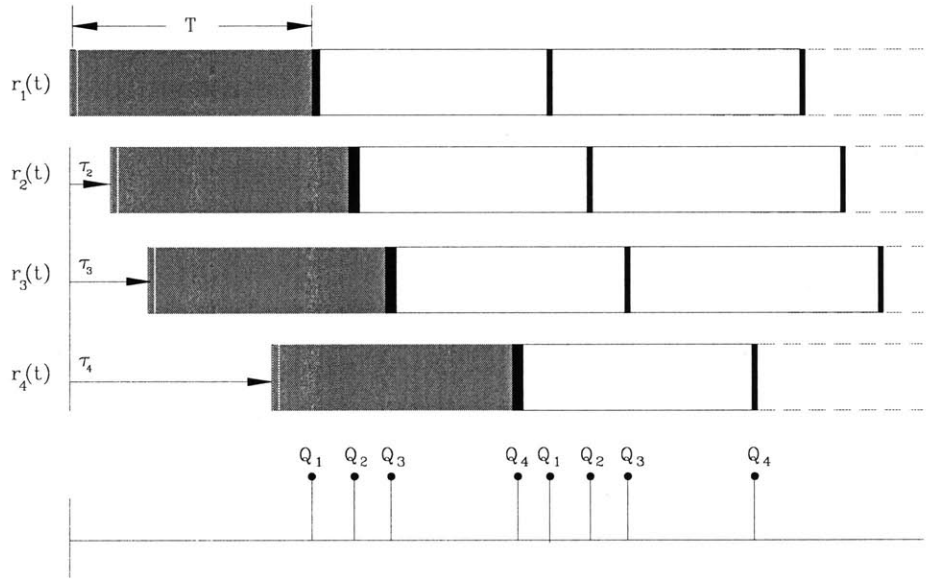



**Figure** 6.8: Resolvable replicas of a SS modulated signal and their sampling instants

Fig.6.8 illustrates the arrival times of $L_r = 4$ replicas of the transmit signal. Set $\tau_1 = 0$. The relative delays of other paths are $\tau_2 < \tau_3 < \tau_4$ and resolvable ($\tau_{i+1} - \tau_i \geq T_c$). The first Rake finger is locked to the signal in path 1 and so on. At this point we ignore interference from neighboring sites. Using eqn. (2.16) we can express the received signal as

$$r(t;\tau) = \sum_{i=1}^{L_r(t)} \alpha_i(t)\, e^{-j\,\phi_i(t)}\, s\Big(\tau - \tau_i(t)\Big) + n(t) \tag{6.46}$$

After matched filtering the sampled output of finger 1, denoted by $Q_1$ is:

$$\begin{aligned} Q_1 &= \Re\left[\alpha_1(t)\, e^{j\,\phi_1(t)} \int_0^T r(t;\tau)\, s\Big(\tau - \tau_1(t)\Big)\, d\tau\right] \\ &= \alpha_1^2\, \mathcal{E}_s + \sum_{i=2}^{L_r} \mathcal{S}_{1i} + \eta_1 \end{aligned} \tag{6.47}$$

where the noise sample at the output of finger i is

$$\eta_i = \Re\Big[\alpha_i e^{j\phi_i} \int_{t=\tau_i}^{T+\tau_i} n(t)s(t-\tau_i)\,dt\Big] \tag{6.48}$$

The self-induced noise from other paths is

$$\begin{aligned}\mathcal{S}_{1i} &= \Re\left[\alpha_1\ \alpha_i\ e^{j\phi_1}\ e^{-j\phi_i} \int_0^T s(\tau-\tau_1)\,s(\tau-\tau_i)\,d\tau\right] \\ &= \alpha_1\ \alpha_i\ \cos(\phi_1-\phi_i)\ C_s(\tau_{i1})\end{aligned} \tag{6.49}$$

Before proceeding, we point out several striking differences between the conventional multi-channel communication system versus SSM communication in multipath fading by comparing eqns. (6.40)–(6.43) with (6.46)–(6.49):

- *Time-Varying Diversity Order*: In multi-channel communication, the number of channels L is a fixed design parameter. In multipath fading, the number of resolvable paths $L_r(t)$ is a time-varying random process that depends on the characteristics of propagation medium.

- *Auto- vs. Cross-Correlation*: In multi-channel communication, the transmit signal $s_i(t)$ in each logical channel i can be carefully constructed to reduce or even eliminate self-induced noise. In SSM communication, $s_i(\tau) = s(\tau-\tau_i)$ is a delayed replica of the transmit $s(t)$ in which the delay parameter $\tau_i(t)$ itself is time-varying. It is rather difficult to design signature waveforms that induce little or no self-noise for any arbitrary $\tau_i$. If finger outputs $\{Q_i\}$ contain SI noise, the maximum-ratio combining (MRC) method is no longer optimal.

- *Independence of Gaussian Noise Samples*: In multi-channel communication, SI noise is measured in terms of periodic *cross*-correlation function $C_{ij}(\tau)$. In contrast, SI noise at the output of a Rake finger is a function of the periodic *auto*-correlation function $C_s(\tau_{il})$. As we have elaborated in Chapter 4 during the discussion on superimposed sequences, $C_s(\tau_{il})$ is ultimately a function of the partial auto-correlation function of the embedded discrete superimposed sequence of period $N_p$ when $N_pT_c \gg T$. As stated earlier, it is rather difficult to design signature waveforms with small $C_s(\tau_{il})$ for an arbitrary $\tau_{il}$. The consequences are as follows. From eqn.(6.42) we can determine the cross-correlation of two Gaussian noise samples

$$\begin{aligned}E\Big[z_i[n]z_j[n]\Big] &= \alpha_i\alpha_j\ \cos(\phi_i-\phi_j) \\ &\quad \int_{u=nT+\tau_i}^{(n+1)T+\tau_i} \int_{v=nT+\tau_j}^{(n+1)T+\tau_j} E\Big[n(u)n(v)\Big]\,s_i(u-\tau_i)s_j(v-\tau_j)\,du\,dv \\ &= \alpha_i\alpha_j\ \cos(\phi_i-\phi_j)\frac{N_o}{2} \int_{t=nT+\tau_i}^{(n+1)T+\tau_i} s_i(t-\tau_i)s_j(t-\tau_j)\,dt\end{aligned}$$

Since the noise samples have zero mean, any two of them are uncorrelated (i.e., independent for Gaussian samples) if $E[z_i[n]z_j[m]] = 0$, a condition guaranteed by $C_s(\tau_{il}) = 0$. In FDM or TDM mode, it is possible to construct channel signals $s)i(t)$ and $s_j(t)$ that are disjoint in frequency (in terms of their Fourier transforms) or time overlap for an offset $\tau_i - \tau_j$. For a Rake receiver, we can derive an analogous expression for the cross-correlation of two Gaussian noise samples. We see that

$$E[\eta_i[n]\eta_j[n]] = 0 \quad \text{if and only if} \quad \int_{t=nT+\tau_i}^{(n+1)T+\tau_i} s(t-\tau_i)s(t-\tau_j)\,dt = 0$$

which is unattainable for an arbitrary $n$ and random $\tau_i$ and $\tau_j$. This implies that the Rake finger noise samples are correlated and the MRC scheme is no longer optimal.

Ignoring SI noise, the output of Rake finger 1 is

$$Q_1 = \alpha_1^2 \mathcal{E}_s + \eta_1$$

After combining outputs of Rake fingers, the combiner output is

$$\begin{aligned} Q &= \sum_{i=1}^{L_r} Q_i \\ &= \mathcal{E}_s \sum_{i=1}^{L_r} \alpha_i^2 + \eta \qquad \text{where } \eta = \sum_{i=1}^{L_r} \eta_i \end{aligned}$$

Conditioned on the fading amplitudes $\underline{\alpha}$, it can be shown that the signal-to-noise ratio of at the output of a Rake receiver is

$$SNR_o(\underline{\alpha}) = \frac{2\mathcal{E}_s}{N_o} \sum_{i=1}^{L_r} \alpha_i^2$$

This is equivalent to the standard form of the output SNR of a maximal-ratio combiner with $L_r$-branch diversity. However, in a practical setting, we must deal with non-ideal conditions that are encountered in CDM signalling over wideband multipath fading channel:

- *Longer Delay Spread*: In Fig. 6.8 the relative delays are purposely chosen so that all are resolvable and their sampling instants are contained in the symbol interval T; i.e., the delay spread $(\Delta\tau)_d < T$. This defines the *under-spread* condition. Ideally we wish to construct signature waveforms with zero periodic autocorrelation $C_s(\tau_i) = 0$ for a relative delay $\tau_i$ as small as possible[8]. If the minimum $\tau_i$ that affords zero auto-correlation is the chip duration $T_c$, then $T_c$ is known as the Rake combiner's *time resolution*. Thus a Rake

[8]The reasoning is that the finer the time resolution, the higher number of scattered transmit energy components the Rake receiver is able to gather. From eqn.(6.45) we know that the larger $L_r$ moves the BER curve away from the single-channel bound towards the AWGN channel bound.

receiver is said to resolve multiple signal paths with a time resolution of $T_c$. If the delay spread is larger than the symbol duration — the *over-spread* condition— such that the relative delay of a multipath signal $\tau_i > T$, the channel induces intersymbol interference (ISI). Theoretically, the replica with delay $\tau_i > T$ can still be used in Rake combining as long as the sampling instants do not overlap. We illustrate such a scenario in Fig. 6.9 for path 5. However, in practice a Rake finger is a tap-delay line filter where each delay unit equals $T_c$. If $T = N_c T_c$ the filter consists of $N_c$ delay taps. Thus, the weight of a multipath signal outside a window of T cannot be tapped for signal combining. It is known from Chapter 2 that the delay spread $(\Delta\tau)_d$ is a time-varying positive variable that depends only on the characteristics of the propagation medium, whereas the symbol duration T (or equivalently, the symbol rate $R_s$) is a design parameter which can be chosen freely by the network designer. The relation between $(\Delta\tau)_d$ and T becomes critical when we compare two multi-rate schemes: PC-SG and SC-RG.

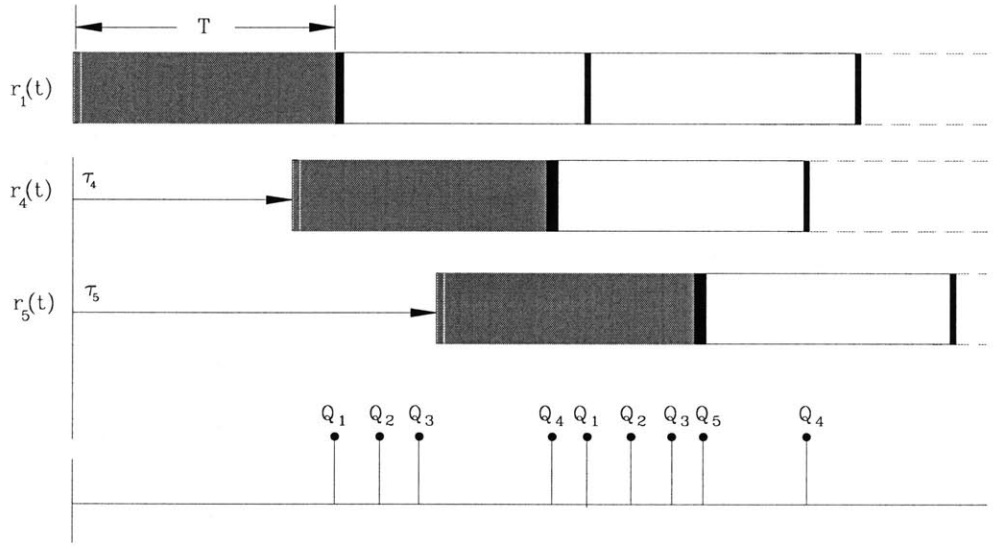



**Figure** 6.9: Intersymbol interference caused by a multipath signal with delay $\tau_5 > T$

- *Non-zero partial correlation*: In spread-spectrum modulated communication the transmit signal

$$s(t) = \sum_j x[j]\, f(t - jT)$$

is data modulated and in the special case where bit- and code-synchronism are equivalent ($T = N_p T_c$), its periodic autocorrelation in eqn.(6.49) must be expressed in terms of the

two aperiodic autocorrelation functions:

$$\begin{aligned}
&\int_0^T s(\tau)\, s(\tau - \tau_i)\, d\tau \\
&\quad = x[-1]\, x[0] \int_0^{\tau} f(\tau)\, f(\tau - \tau_i)\, d\tau + x[0] \int_{\tau}^{T} f(\tau)\, f(\tau - \tau_i)\, d\tau \\
&\quad = \pm R_f(\tau_i) + \hat{R}_f(\tau_i)
\end{aligned}$$

Depending on the sign (the product of $x[-1]$ and $x[0]$) the overall autocorrelation value may no longer be zero regardless of the type of signature waveform used. Furthermore, if $s(t)$ is modulated by a long PN sequence such that $N_p \gg N_c$, the periodic autocorrelation must be expressed in terms of the discrete-time partial autocorrelation function. By applying eqn.(4.36),

$$\begin{aligned}
C_s(\tau_i) &= \int_0^T s(\tau)\, s(\tau - \tau_i)\, d\tau \\
&= \theta_b[l; 0, N_c]\, \rho_\psi(\Delta) + \theta_b[l-1; 0, N_c]\, \hat{\rho}_\psi(\Delta)
\end{aligned}$$

where $\tau_i = (lT_c + \Delta)$ and

$$s(t) = \sum_{j=-\infty}^{\infty} b[j]\psi(t - jT_c)$$

Recalling our discussion in Chapter 4 on spread-spectrum modulated waveforms, each is generated by modulating an antipodal discrete, periodic PN sequence with a periodic train of common chip pulses. We know that maximum-length (ML) PN codes have a normalized *periodic* autocorrelation $C(\tau)$ equal to $-1/N_p$ for $\tau \geq T_c$, where $N_p$ is the code period. However, we are now dealing with the partial auto-correlation of the embedded sequence. Regardless of the type PN sequence used, it is very unlikely that $\theta_b[l; 0, N_c] \approx 0$ for any arbitrary chip delay $l$. Therefore, we should be aware that each Rake finger output suffers from self-induced noise from other paths. In time and frequency diversity systems, the independence constraint for zero cross-correlation is guaranteed by assigning frequency bands that are separated by at least the coherence bandwidth, and time slots that are separated by more than the coherence time of the channel. The zero self-interference condition is satisfied by assigning non-overlapping frequency bands or time-slots. In both cases the ISI issue is non-existent since the diversity branches are purposely generated by the system designer to combat signal degradation due to fading.

### 6.5.2 Correlated and Uncorrelated Multipath Interference

In the literature, several authors have distinguished SI noise into two categories: *correlated* and *uncorrelated* self-interference. When path delay $\tau_{ij}$ between two received replicas is less

than symbol duration T, the SI noise generated at the output of finger i or j is said to be correlated. This scenario is illustrated in Fig. 6.11. In contrast, when $\tau_{ij} > T$ as depicted in Fig. 6.10, the self-interference is uncorrelated. Note the difference between correlatedness of SI noise from the ISI-causing self-noise as illustrated in Fig. 6.9:

$$\begin{aligned} \text{Correlated SI noise} &: \quad \tau_{ij} = \tau_j - \tau_i < T \quad \text{assuming } \tau_j > \tau_i \\ \text{Uncorrelated SI noise} &: \quad \tau_{ij} > T \\ \text{ISI-causing SI noise} &: \quad \tau_i > T \quad \text{for any path } i \end{aligned}$$

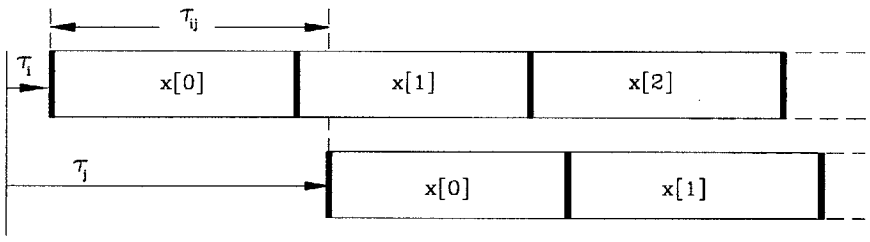



**Figure** 6.10: Uncorrelated self-interference when path delay $\tau_{ij} > T$

We immediately deduce that if SI noise samples between paths i and j are uncorrelated, then at least path i causes ISI. On the other hand, if the signal in path i causes ISI, its self-interference may or may not be correlated with replicas from other paths. Referring to Fig. 6.9, we see that paths 4 and 5 have correlated self-interference while paths 3 and 5 are uncorrelated. We now explain the correlatedness.

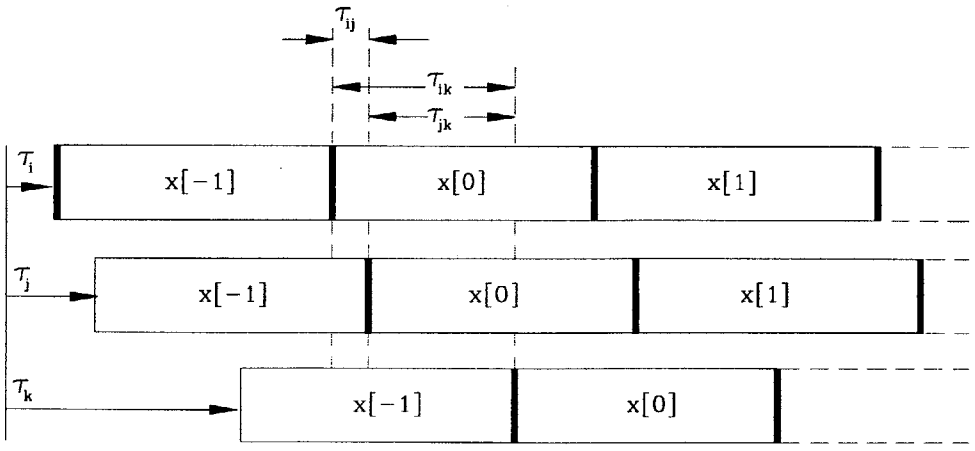



**Figure** 6.11: Correlated self-interference when path delay $\tau_{ij} < T$

### Correlated Self-Interference (C-SI)

The SI noise at the output of finger i can be written as

$$\mathcal{S}_i = \beta_{ij}\Big\{x[-]\,\hat{R}_i(\tau_{ij}) + x[0]\,R_i(\tau_{ij})\Big\} + \beta_{ik}\Big\{x[-]\,\hat{R}_i(\tau_{ik}) + x[0]\,R_i(\tau_{ik})\Big\}$$

where the aperiodic auto-correlation functions are

$$\hat{R}_i(\tau_{ij}) = \int_{t=0}^{\tau_{ij}} s(t)s(t-\tau_{ij})dt \;= R_j(\tau_{ij})$$

$$R_i(\tau_{ij}) = \int_{t=\tau_{ij}}^{T} s(t)s(t-\tau_{ij})dt = \hat{R}_j(\tau_{ij})$$

For short-hand we will use

$$\beta_{ij} = \alpha_i\,\alpha_j\,\cos(\phi_i - \phi_i), \qquad x[-] = x[-1], \qquad x[+] = x[1]$$

Factoring out the data bits,

$$\mathcal{S}_i = x[-]\left\{\beta_{ij}\,\hat{R}_i(\tau_{ij}) + \beta_{ik}\,\hat{R}_i(\tau_{ik})\right\} + x[0]\left\{\beta_{ij}\,R_i(\tau_{ij}) + \beta_{ik}\,R_i(\tau_{ik})\right\} + x[+]\left\{-\right\}$$

Similarly for fingers j and k:

$$\mathcal{S}_j = x[-]\left\{\beta_{jk}\,\hat{R}_j(\tau_{jk})\right\} + x[0]\left\{\beta_{ij}\,R_i(\tau_{ij}) + \beta_{jk}\,R_j(\tau_{jk})\right\} + x[+]\left\{\beta_{ij}\,\hat{R}_i(\tau_{ij})\right\}$$

$$\mathcal{S}_k = x[-]\left\{-\right\} + x[0]\left\{\beta_{ik}\,R_i(\tau_{ik}) + \beta_{jk}\,R_j(\tau_{jk})\right\} + x[+]\left\{\beta_{ik}\,\hat{R}_i(\tau_{ik}) + \beta_{jk}\,\hat{R}_j(\tau_{jk})\right\}$$

Note that in each finger l, the terms associated with data bit x[0] are "overlapping" or "correlated" since the same auto-correlation functions also appear in other finger outputs. Similarly, terms associated with x[−] are correlation between signals in path l with other time *delayed* paths (paths whose delays $\tau_k > \tau_l$). In contrast, terms associated with x[+] are correlation between signals in path l with other time *advanced* paths (paths whose delays $\tau_i < \tau_l$). Hence, in general,we can express the SI noise at the output of an arbitrary path l as

$$\mathcal{S}_l = \sum_{i<l} \beta_{il}\left\{x[0]\,R_i(\tau_{il}) + x[+]\,\hat{R}_i(\tau_{il})\right\} + \sum_{k>l} \beta_{lk}\left\{x[0]\,R_l(\tau_{lk}) + x[-]\,\hat{R}_l(\tau_{lk})\right\}$$

Alternatively, with an abuse in notation:

$$\mathcal{S}_l = x[0]\sum_{j\neq l}\beta_{jl}\,R_j(\tau) + x[+]\sum_{i<l}\beta_{il}\,\hat{R}_i(\tau) + x[-]\sum_{k>l}\beta_{lk}\,\hat{R}_l(\tau) \tag{6.50}$$

where it is understood that

$$\tau = \tau_{jl} \quad \text{for} \quad l > j \qquad \text{and} \qquad \tau = \tau_{lj} \quad \text{for} \quad l < j$$

It is interesting to note that the first summation term is identical to the input of a synchronous decorrelator receiver (see the text by Verdu [149] for details) for multi-user detection, except in our case

1. We are dealing with auto- (instead of cross-) correlation functions. To a lesser importance, the correlation functions in the Rake receiver are aperiodic, whereas in a decorrelator they are periodic.

2. In our case each correlation function is scaled by a fade statistic $\beta_{ij}$.

3. The transmit data vector of length L

   $$\mathbf{x}[0] = \Big[ x[0], x[0], \ldots, x[0] \Big]$$

   is an L-fold repetition of the transmit data bit x[0], whereas in a decorrelator the data vector consists of independent data bits from L different users:

   $$\mathbf{u}[0] = \Big[ u_1[0], u_2[0], \ldots, u_L[0] \Big]$$

   such that the input (column) vector of a decorrelator can be expressed as

   $$\mathbf{Y} = \mathbf{u}[0]\, \mathbf{C}_s + \mathbf{z}$$

   where $\mathbf{C}_s$ is the (normalized) periodic cross-correlation vector and $\mathbf{z}$ is the Gaussian noise vector. Ignoring the second and third summation terms in eqn.(6.50) the output vector of a Rake receiver can be written as

   $$\mathbf{Q} = \mathbf{x}[0]\, \mathbf{R}_s + \underline{\eta}$$

   where

   $$\mathbf{Q} = \Big[ Q_1, Q_2, \ldots, Q_L \Big]$$

   $$\underline{\eta} = \Big[ \eta_1, \eta_2, \ldots, \eta_L \Big]$$

For maximum ratio diversity combining with coherent detection, the Rake receiver must accurately estimate channel statistics $\{\alpha_i, \phi_i, \tau_i\}$ for each finger. Thus, we can assume that the set of scaling factors $\{\beta_{ij}\}$ is known to the receiver. The receiver also has full knowledge of the auto-correlation values of its signature sequence. In essence the correlation matrix $\mathbf{R}_s$ is known to the Rake receiver. Referring to linear multi-user CDMA detection schemes we can multiply the

output vector **Q** with a suitable matrix to either *decorrelate* or *minimize the mean-squared error*. Before we execute this step we must first deal with the two summation terms in eqn.(6.50). It is apparent that both terms can be treated as ISI from past and future data bits and thus can be removed by decision-feedback equalization.

To summarize, in this section on correlated self-interference (C-SI) we emphasize the fact that the Rake receiver with coherent MRC is far from optimum, mainly due to non-zero auto-correlation of replicas from different paths. All references that we come across that address C-SI as undesirable noise and no attempt is made to remove it by de-correlation or decision-feedback equalization.

#### Uncorrelated Self-Interference (U-SI)

When $\tau_{ij} > T$, signal in path j causes ISI to output signal in finger i. If we assume the Rake receiver has tap-delay line implementation and thus cannot correlated with signal paths outside the symbol time correlation window, then U-SI must be treated as unwanted noise.

#### Correlated and Uncorrelated Channel Interference (CI)

In multipath fading model, the output of each Rake finger suffers not only SI from delayed replicas of its own signal, but also from delayed replicas in other parallel channels. We define this type of interference as *channel interference*. Depending on the value of each propagation delay (relative to T), CI is either correlated or uncorrelated with the desired symbol time window. For correlated CI, since each receiver has no knowledge of the signature sequences of other channels, no interference cancellation —analogous to the de-correlator proposed for correlated SI— is possible. Therefore, CI is treated as unwanted noise at Rake finger outputs.

### 6.5.3 Gaussian Approximation of Multi-User Interference

Thus far, we have classified four types of *intra-cell* interference: correlated and uncorrelated self-interference (C-SI and U-SI), and correlated and uncorrelated channel interference. The latter two have the same effect as unwanted noise and we collectively call them as CI. The only parameter common to all types of intra-cell interference is the propagation delay set $\{\tau_i\}$ —which is known or can be estimated with good accuracy by each Rake receiver. Therefore, the contribution of each unwanted noise —in terms of energy— can be computed deterministically from periodic auto- or cross-correlation of desired signal and unwanted noise signal. For this step, the receiver must have complete knowledge of actively assigned Walsh code channels. Alternatively, such knowledge of assigned codes is not necessary if it plans to treat CI as unwanted noise.

When we introduce inter-cell interference, their arrival times are not known and random to the target receiver. If an inter-cell arrival time falls within the correlation window, it contributes one unit of unwanted noise. If its arrival time falls outside the window, it contributes one unit of unwanted noise to the next target data bit. That is, for inter-cell interference, it makes no difference what its absolute arrival time epoch is relative to the correlation window. Thus, we assume the unwanted noise contributed by each inter-cell interferer is indistinguishable from an intra-cell channel interference contributed by another parallel active code channel. If there exists —on average— $K_c$ actively assigned forward link code channels per cell, the total number of interfering channels $J_c$ in each Rake finger input is $(K_c - 1 + K_c J)$. If $J_c$ is large, the accuracy of Gaussian approximation of total CI is improved. Since we know that BER averaged over the statistic of fading amplitude improves the accuracy of GA, we will measure total interference power (variance) using SGA. (We did not measure any improvement in BER accuracy by substituting CGA or IGA in place of SGA.)

## 6.6 Comparative Analysis of BER Curves of Target CDM Users

In eqn. (6.17) we treat each interfering signal as an arbitrary pseudo-random sequence not known to the target receiver. When each interferer's information bit $x_i[m]$ is modulated with a superimposed orthogonal sequence $a_i[n]$, the resulting sequence is another superimposed orthogonal sequence $b_i[n]$ of chip rate $R_c$. The spreading gain —and the corresponding information rate— affects the receive power level $P_i$ only. Therefore, our derived formulae and other analytical results from previous sections still apply to multi-rate PC-SG and SC-RG CDM channels. The only difference is that for a higher-rate SC-RG CDM target user, its spreading gain $N_c$ is reduced —which subsequently affects the values of cross-correlation terms ($\zeta_i[l]$ and $\hat{\zeta}_i[l]$) and the noise term $\eta$. To maintain fairness in comparative study, its power gain $\beta_i$ [see eqn. (6.14)] must be re-scaled such that its bit energy $\mathcal{E}_b$ (and thus, $\mathcal{E}_b/N_o$) is the same as the bit energy of a standard SC-SG CDM channel or a component low rate channel in a PC-SG CDM high rate channel.

### 6.6.1 AWGN Only Channel

Our goal in this section is to determine the accuracy of BER curves that are generated using various Gaussian approximation methods. As a benchmark we also include the exact (average) BER curve, computed according to eqn.(6.24). In particular, we are interested in the accuracy of Gaussian approximations as system parameters such as channel bandwidth (or equivalently, the spreading gain $N_c$), the interference population J and link quality (measured in $\mathcal{E}_b/N_o$) are varied. Eight BER curves —as shown in Figs. 6.12–6.15— are presented. The four plots

measure exact and approximate BERs based on Gaussian approximations as a function of interference population J for

1. narrow bandwidth, high signal-to-noise ratio: $N_c$=64, $\mathcal{E}_b/N_o$=12 dB.
2. wide bandwidth, high signal-to-noise ratio: $N_c$=256, $\mathcal{E}_b/N_o$=12 dB.
3. narrow bandwidth, low signal-to-noise ratio: $N_c$=64, $\mathcal{E}_b/N_o$=5 dB.
4. wide bandwidth, low signal-to-noise ratio: $N_c$=64, $\mathcal{E}_b/N_o$=5 dB.

### 6.6.2 Flat Fading Channel

Similar to the AWGN channel model, we analyze the effects of system parameters —the spreading gain $N_c$, the interference population J and link quality $\mathcal{E}_b/N_o$ on BER in flat fading Gaussian channel. Two BER curves —as shown in Figs. 6.16–6.17— are presented where the pdf of fading amplitude is Rayleigh. The two plots measure exact and approximate BERs based on Gaussian approximations as a function of interference population J for

1. narrow bandwidth, high signal-to-noise ratio: $N_c$=64, $\mathcal{E}_b/N_o$=12 dB.
2. wide bandwidth, high signal-to-noise ratio: $N_c$=256, $\mathcal{E}_b/N_o$=12 dB.

### 6.6.3 Frequency Selective Fading Channel

Two BER plots in frequency selective fading, both with three resolvable paths, are shown in Figs. 6.18–6.19. In the first plot, the relative delays are selected such that in one pair of plots, only correlated self-noise is observed while in the other pair, both correlated and uncorrelated SI are observed. The second plots are similar except that time delays that generate ISI are used.

### 6.6.4 Comments on BER plots

From BER plots in AWGN, we deduce that SGA is optimistic while the reverse is true for IGA. Of course, the gap between IGA and exact BER curve is much smaller. This gap shrinks in flat fading model. There is little gap between the exact curve and all other Gaussian approximation curves. As we have seen from these BER plots in AWGN, flat fading and frequency selective fading, the difference between PC-SG CDM and SC-RG CDM is barely noticeable in almost all cases. However, when taking into account seamless multi-rate provisioning and reduced receiver complexity, single-channel CDM with reduced gain is the preferred option.

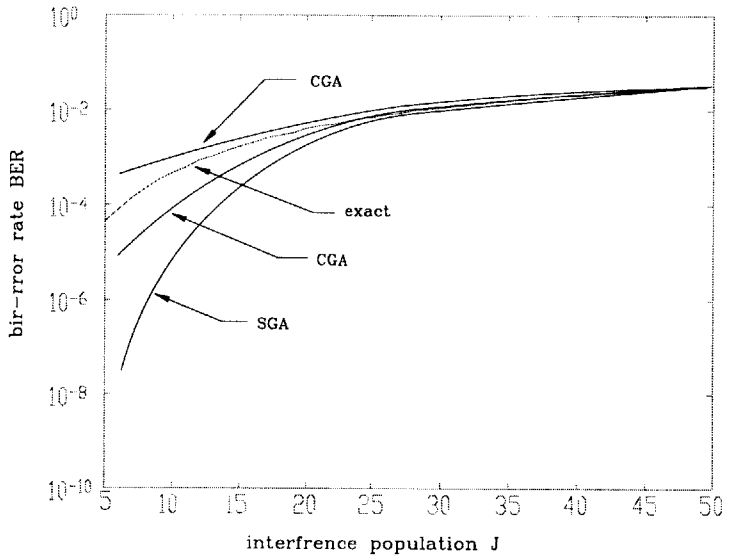



**Figure 6.12:** AWGN-BER of narrowband, high SNR CDM signalling as a function of interference population J. $N_c = 64$, $\mathcal{E}_b/N_o = 12$ dB.

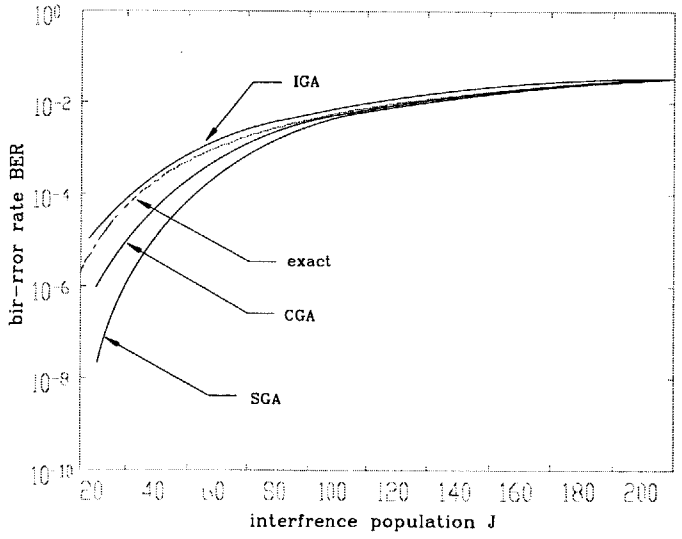



**Figure 6.13:** AWGN-BER of wideband, high SNR CDM signalling as a function of interference population J. $N_c = 256$, $\mathcal{E}_b/N_o = 12$ dB.

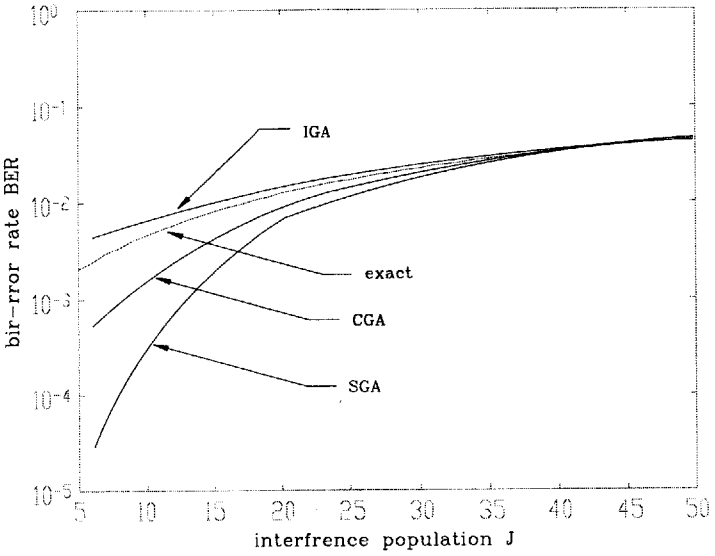



**Figure 6.14:** AWGN-BER of narrowband, low SNR CDM signalling as a function of interference population J. $N_c = 64$, $\mathcal{E}_b/N_o = 5$ dB.

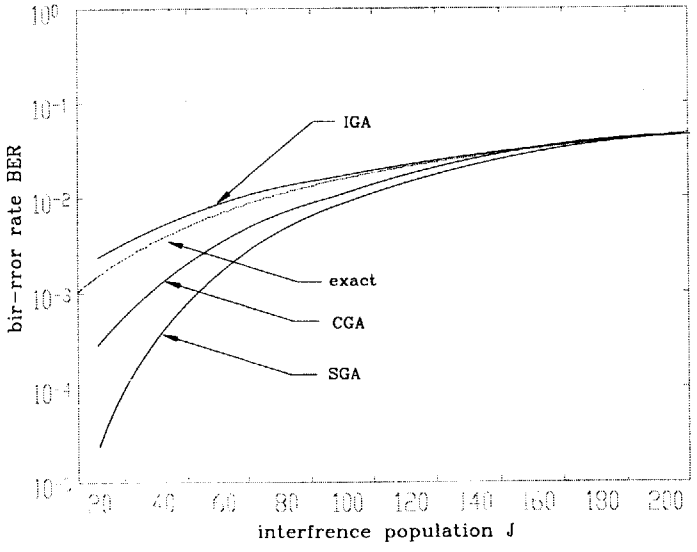



**Figure 6.15:** AWGN-BER of wideband, low SNR CDM signalling as a function of interference population J. $N_c = 256$, $\mathcal{E}_b/N_o = 5$ dB.

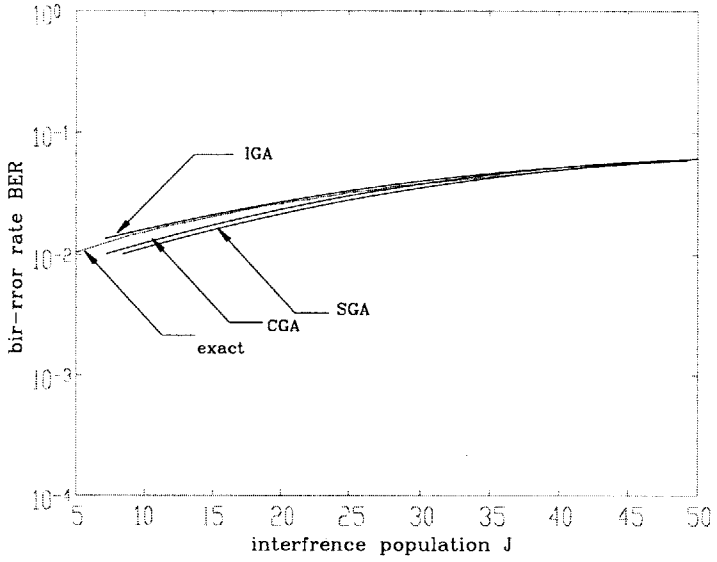



**Figure 6.16:** Flat Rayleigh fading-BER of narrowband, high SNR CDM signalling as a function of interference population J. $N_c = 64$, $\mathcal{E}_b/N_o = 12$ dB.

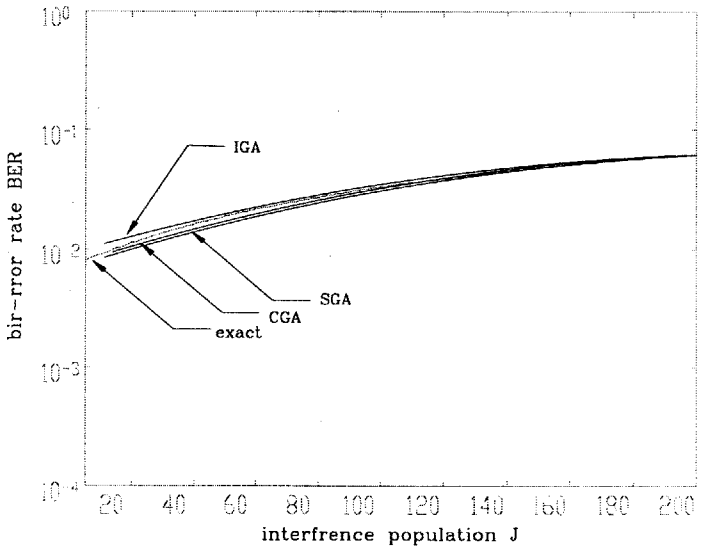
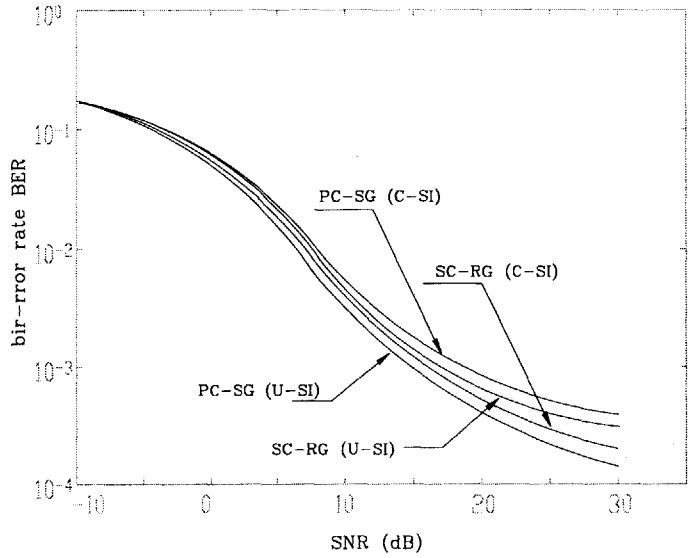
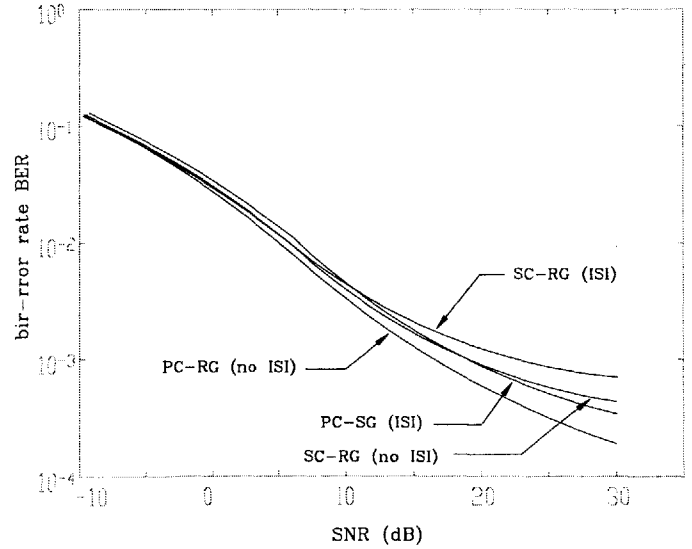



**Figure 6.17:** Flat Rayleigh fading-BER of wideband, high SNR CDM signalling as a function of interference population J. $N_c = 256$, $\mathcal{E}_b/N_o = 12$ dB.




Frequency selective Rayleigh fading-BER of wideband, high SNR

**Figure** 6.18: CDM signalling as a function of signal-to-noise power ratio SNR. $N_c = 256$, $\mathcal{E}_b/N_o = 12$ dB. Non ISI case.




Frequency selective Rayleigh fading-BER of wideband, high SNR

**Figure** 6.19: CDM signalling as a function of signal-to-noise power ratio SNR. $N_c = 256$, $\mathcal{E}_b/N_o = 12$ dB. ISI case.

### 6.6.5 Comments on Numerical Methods for CDM BER Analysis

When dealing with and manipulating the standard Gaussian cumulative distribution function (CDF) or its complementary Q function, the most difficult part is its lower limit x:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{u^2}{2}\right) du$$

which is also the indexing parameter. When we average the Q function over x, the resulting equation in closed form is often difficult to derive except for certain probability density functions. For numerical computation, it tends to be unsteady since the extent of x is unbounded, at least toward infinity. However, we are fortunate that a few years ago, J. W. Craig derived an alternative integral form of the Q function

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{u^2}{2\sin^2\theta}\right) d\theta$$

It is clear that this is definite integral where the indexing parameter x is no longer a limit. This new form only works for $x \geq 0$. Since the standard Gaussian pdf is symmetric, this is a simple constraint that can be dealt with easily. In all our numerical computations, including computer-generated BER plots, this simpler integral of the Q function was used. Details on this form and its variations are detailed in the text by Simon and Alouini [131].

## Notes and References

In the following, we highlight only certain authors/researchers and their work(s) that can be considered as pioneering or had a profound impact on the advancement of CDM/CDMA technologies.

Among texts that treat topics in spread-spectrum communications, the collaborative work of M. K. Simon et al. [129] is considered the "Bible." It thoroughly covers all aspects of both direct-sequence and frequency-hopped SS communication systems. This includes such applications as anti-jamming and low-intercept military applications, ranging and location (GPS type) commercial applications as well as historical account of SS systems. Its only lack is in the area of multi-user communication with SS modulating codes. Other excellent books that cover similar topics but with a different flavor are those by Dixon [30] and Holmes [54]. The latter emphasizes on coherent reception of direct-sequence SS modulated signals. The earliest analytical work on phase-coded (direct-sequence) SS modulation/demodulation for multi-user communication can be attributed to M. Pursley [109]. It measures (the approximate) bit-error rate of a spread-spectrum multi-access (SSMA) signal through an AWGN channel in a multi-user environment using the received signal-to-ratio (SNR) as the benchmark. These two articles may been have the first to propose the *standard* or simplified Gaussian approximation method. The sequel [110] computes various bounds for the BER based on the correlation properties on signature sequences. However, it appears the idead of SS modulated signals for multi-user communication was popularized by a tutorial by R. Pickholtz et al. [102] —based on the number of times it's been cited in future published articles. Subsequent articles by M. Pursley and his students ([112],[89], [74], [39]) extend his earlier results by modifying the standard Gaussian approximation method to better approximate the exact BER. In particular, Morrow and Lehnert proposed the *improved* Gaussian approximation in [89].

Spread-spectrum multiple access with superimposed orthogonal sequences was first proposed by K. S. Gilhousen of Qualcomm Inc. The best source to study Qualcomm's CDMA architecture (besides the official U.S. Interim Standard-95 documents [61]) is the text by one of its founders A. J. Viterbi [154]. Before the introduction of Qualcomm's Walsh code based CDM, SSMA systems assume the assignment of non-orthogonal pseudo-random SS sequences. This is the main reason earlier works have put so much emphasis on the design and correlation properties of PN sequences and the resulting BER. In Qualcomm's CDM where the Walsh code length $N(= 64)$ is much less than the period of the PN sequence $N_p(= 32767)$, the spread-spectrum modulating sequence can be conveniently modelled as a random Bernoulli sequence. This model opens the door for CDMA BER analysis with random spreading sequences. The auto- and cross-correlation of random sequences can now be expressed in terms of their statistical properties rather than in terms of deterministic values or bounds based on a pair of known,

deterministic PN sequences. The number of articles in this area is far too many to list here. We only cite a journal article by Fong et al. [35] since it is one of a limited few that compares the BER performance of superimposed orthogonal CDM versus conventional PN sequence based CDM.

The inclusion of signal fading in a CDM/CDMA system model results in a more realistic BER analysis since almost all CDM/CDMA systems communicate over the wireless link. The analysis of a flat fading model is straightforward; furthermore, it is more of academic value than practical significance since a typical wideband spread-spectrum signal undergoes frequency selective fading. Optimal diversity combining of multipath signals was first proposed by D. G. Brennan [11]. It was made accurate by R. Price [105]. He and P. Green [106] successfully coined the term "Rake" receiver for such multipath combiner. The multipath combining and SNR enhancement of a Rake receiver was popularized by G. Turin [143]–[146]. His insightful amalgamation of theory and practice for wireless multipath fading channels highlights the diversity combining effect of SSMA signals, similar to that attained by time, frequency or antenna diversity scheme.

The universal standardization of third-generation wireless standards based on wideband CDMA fostered the study of BER analysis for multi-rate schemes. Among the many published works, we attribute Chih-Lin I et al. [58],[59], [42] of AT&T as the inventors of multi-code (parallel-channel, single-gain) transmission mode. Its counterpart the single-code, reduced-gain mode is patented by K. S. Gilhousen of Qualcomm [41]. Based on standard Gaussian approximation, Ottoson and Svensson [95] showed that both schemes are comparable in BER analysis. Ramakrishna and Holtzman confirmed that the equality of both multi-rate schemes holds under the improved Gaussian approximation model. Both studies focus on AWGN channels. By contrast M. Fan et al. [31] used an exact model to analyze the bit-error rate of a chip-synchronous CDMA system and showed slight differences between multi-code and OVSF schemes —and their dependence on received SNR and user population. They then showed that both schemes are equal in a flat-fading model due to the averaging effect of the fading amplitude. The significance of inter-path interference on the BER in CDMA was first pointed out by Chan [16]. As a follow-up, the performance of a CDMA system in a multipath fading environment with the inclusion of self-interference was studied by Cheun [19]. In his analysis the standard Gaussian approximation is used. Subsequent analysis by Hwang and Lee [57] showed that the Gaussian approximation breaks down as the spreading factor is reduced. Based on similar arguments, Zhang et al. [168] claim that multi-code outperforms OVSF in a multipath fading channel.

As far as we know, combining methods that reduce inter-path interference or inter-path interference plus Gaussian noise —as proposed in this chapter— have not appeared elsewhere. In fact, these procedures are very similar to linear de-correlator and MMSE detection methods [149] that have been proposed for joint detection of multi-user CDMA signals.

# Appendix 6A

## Linear Transformation of Multi-Code Signals

By observing Fig. 6.4 it is obvious that the level-shifted baseband sequence (before pulse shaping) can be expressed as

$$p[n] \cdot v[m]$$

where $p[n]$ is the common (cell-specific) periodic, level-shifted PN sequence, and the Walsh modulated data sequence can be expressed as

$$v[m] = \sum_{k=1}^{N} x_k \left[ \left\lfloor \frac{m}{N} \right\rfloor \right] \cdot w_k[m] \tag{6.51}$$

Here $w_k[m]$ denotes the $m^{th}$ element of the $k^{th}$ Walsh code. In terms of the Walsh matrix $\mathcal{W}_N$,

$$w_k[m] = w_{km} \in \mathcal{W}_N$$

is the $km^{th}$ entry ($k^{th}$ row and $m^{th}$ column) of the Walsh matrix. For a block of input data sequence

$$\left([x_1[i], x_2[i], \ldots, x_N[i]\right) \tag{6.52}$$

of length N, $v[m]$ is the sum of all Walsh codes (of length N), each code $w_k[m]$ weighed by its corresponding data symbol $x_k[i]$. In the above equations, we purposely use different indices $m$ and $n$ for the Walsh and PN sequences to highlight the fact that their clock are not necessarily the same. In matrix notation, the Walsh modulated sequence of eqn. (6.51) can be written as

$$\mathbf{v} = \mathbf{x} \cdot \mathcal{W}_N \tag{6.53}$$

where the vector x is an input data block of length N as given in eqn. (6.52). The Walsh matrix

$$\mathcal{W}_N = \left[\mathbf{w}_1^T \; \mathbf{w}_2^T \; \ldots \mathbf{w}_N^T\right]^T$$

is expressed in terms of its orthogonal row vectors $\{\mathbf{w}_k\}$. The superscript T denotes vector transposition. Note that in multiple-channel, single-gain (PC-SG) transmission, the modulating matrix $\mathbf{M}$ consists of a subset of orthogonal vectors of the Walsh matrix:

$$\mathbf{M} = \left[\mathbf{w}_i^T \; \mathbf{w}_j^T \; \ldots \mathbf{w}_k^T\right]^T$$

where $\mathbf{M}$ is an M-byN matrix consisting of any M orthogonal row vectors from $\mathcal{W}_N$. We assume there are $M < N$ parallel CDM channels per user.

## Constant-Amplitude Signals

It is apparent that the Walsh modulated sequence, which is also the linear combination of N antipodal binary sequences, is no longer binary. In fact it is a multi-level sequence where its elements

$$v_j[m] \in \mathbf{S} = \{s_0, s_1, \ldots, s_M\} \tag{6.54}$$

belongs to a constellation set **S** such that its elements

$$s_i = 2i - M \qquad i \in \mathbb{N}^M$$

Note that the subscript j in $v_j[m]$ denotes the index of the component of the vector **v** and not a user's index. For example, if $M = 4$ the constellation set is

$$\mathbf{S}_4 = \{-4, -2, 0, 2, 4\}$$

It is easy to see that when M is odd, the multi-level signal set is equivalent to $(M+1)$-ary pulse amplitude modulation set $\mathbf{S}_{PAM}$.

**Table** 6.1: Possible outputs of a Walsh modulated sequence

| X | v[m] | Index | $T_1$ | $T_2$ | Index |
|---|---|---|---|---|---|
| 0000 | (4, 0, 0, 0) | | | | |
| 0001 | (2, 2, 2, -2) | ✓ | 1 | (2, 2, 2, 2) | 0 |
| 0010 | (2, -2, 2, 2) | ✓ | 4 | (2, -2, 2, -2) | 5 |
| 0011 | (0, 0, 4, 0) | | | | |
| 0100 | (2, 2, -2, 2) | ✓ | 2 | (2, 2, -2, -2) | 3 |
| 0101 | (0, 4, 0, 0) | | | | |
| 0110 | (0, 0, 0, 4) | | | | |
| 0111 | (-2, 2, 2, 2) | ✓ | 8 | (-2, 2, 2, -2) | 9 |
| 1000 | (2, -2, -2, -2) | ✓ | 7 | (2, -2, -2, 2) | 6 |
| 1001 | (0, 0, 0, -4) | | | | |
| 1010 | (0, -4, 0, 0) | | | | |
| 1011 | (-2, -2, 2, -2) | ✓ | 13 | (-2, -2, 2, 2) | 12 |
| 1100 | (0, 0, -4, 0) | | | | |
| 1101 | (-2, 2, -2, -2) | ✓ | 11 | (-2, 2, -2, 2) | 10 |
| 1110 | (-2, -2, -2, 2) | ✓ | 14 | (-2, -2, -2, -2) | 15 |
| 1111 | (-4, 0, 0, 0) | | | | |

To further elaborate the multi-level properties of $v[m]$, we consider the $(4\times 4)$ Walsh matrix:

$$\mathcal{W}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix} \tag{6.55}$$

By assigning alternating $(+/-)$ weights in the summation of Walsh codes, all 16 possible output sequences of $v[m]$ can be generated. They are shown in Table 6.1 for $M = N = 4$. The first column lists all 16 possible input vectors. The input sequence is $\mathbf{X} = \big(X[1],\ X[2],\ X[3],\ X[4]\big)$ and the mapping for level-shifting is

$$X[i] = 0 \rightarrow 1, \qquad X[i] = 1 \rightarrow -1$$

The second column lists the resulting Walsh modulated vectors. The third column checks the resulting vector (from the same row) whose elements $v_j \in 2\mathbb{D}_2 = \{-2, 2\}$. Such a vector is called *constant-amplitude*. Let T denote a set of constant amplitude, length-4 vectors whose elements are from the set $2\mathbb{D}_2$. It is immediate that T consists of 16 entries, some of which belong to the subset $T_1 \subset T$ of valid Walsh output sequences. The remaining complementary vectors in the set $T_2 \subset T$ are listed in the fifth column. Note that $T_1$ and $T_2$ are mutually exclusive and exhaustive subsets of T; i.e., $T_1 = (T_2)^c$ where the superscript "c" denotes the complement. For convenience, we also list the equivalent decimal value of each binary vector of T in the columns 4 and 6. For example, the decimal value of $(-2,-2,-2,2) \rightarrow (1,1,1,0)$ is 14. Using this notation, we can state that

$$T_1 = \{1, 2, 4, 7, 8, 11, 13, 14\}$$
$$T_2 = \{0, 3, 5, 6, 9, 10, 15\}$$

We also define the domain set $S_1$ consisting on input vectors whose corresponding range is $T_1$. When expressed in decimal format,

$$S_1 = \{1, 2, 4, 7, 8, 11, 13, 14\}$$

It is seen that $T_1 = S_1$ and $(S_1)^c = T_2$. It is no coincidence that exactly half of all possible Walsh modulated vectors (those in $T_1$) are constant-amplitude. Due to equal energy constraint, a vector with an element $\pm 4$ must contain three zeroes. By permutation, there exists eight such vectors and the remaining eight are constant-amplitude. This issue of constant-amplitude signal design to not directly related to our study of CDM transmission in the downlink. This is due to the fact that the linearly combined signal in a CDM transmitter is always multi-level — independent of the type of multi-rate transmission scheme used, PC-SG or SC-RG. Furthermore,

the issue of power amplifier linearity and its efficiency is not critical at the base site. However, if multi-code (PC-SG) transmission were to be used in the uplink mobile handset, the amplitude fluctuation of the transmit signal can cause unwanted signal distortion and spectral leakage. For that matter, we are interested in particular cases where a Walsh modulated sequence retains the constant-amplitude property.

## Precoding

In the above fairly simple example, $N = M = 4$. We would like to be able to understand the partitioning of the sets $T_1$ and $T_2$ for larger values of N and M. This information is useful when measuring the peak-to-average amplitude ratio of the Walsh modulated signal. (A higher peak-to-average amplitude ratio may require less power-efficient linear amplifiers [24].) It has been shown in [157] that a constant-amplitude Walsh modulated sequence of length 4 can be constructed for a block of 3-bit input sequence by appending a fourth data bit that satisfies the "parity-check" equation. This coding technique is known as *precoding*. The technique is straightforward: For every 3-bit input sequence

$$\big(X[i],\, X[i+1],\, X[i+2]\big)$$

a parity-check $X[i+3]$ that satisfies the following equation

$$X[i+3] = \overline{X[i] \oplus X[i+1] \oplus X[i+2]} \tag{6.56}$$

is appended such that their corresponding Walsh modulated sequence $v[n]$ is constant-amplitude. The over-bar denotes binary (modulo-2) complement. In terms of antipodal sequences, the parity-check equation is

$$x[i+3] = -\big(x[i] \cdot x[i+1] \cdot x[i+2]\big) \tag{6.57}$$

In order to generate constant-amplitude vectors, we know that the valid input vectors must belong to $S_1$. By observing Table 6.1. we note that the first three bits $\big(X[i],\, X[i+1],\, X[i+2]\big)$ of vectors in $S_1$ span all possible 3-bit patterns from [000] to [111]. The only requirement now is to devise a boolean relationship that results in the fourth bit being equal to $X[i+3]$. It can be deduced that the fourth bit is an inverted sum of the first three. By iteration, it has been shown by Wada et al. [157] that for any complete set of Walsh codes of length $N = 4^m$ for $m \in \mathbb{Z}$, a block of input sequence of length $3^m$ can be appended with parity-check bits such that the resulting Walsh modulated sequence is constant-amplitude. The drawback of precoding is that the effective bit rate is reduced by a factor of $(3/4)^m$. Furthermore, if a single CDM channel carries a bit rate of R bps, then possible multiple bit rates of PC-SG transmission, combined with precoding, are limited to $3^m$ R bps.

### Walsh Code Symmetry

The construction of the parity-check equation in eqn. (6.56) is straightforward since it is based on the observation of all possible input vectors in Table 6.1. In the section we present an alternative observation that takes into account the "symmetric" structure of a Walsh matrix. For the $(2 \times 2)$ matrix

$$W_2 = \begin{bmatrix} + & + \\ + & - \end{bmatrix} \tag{6.58}$$

All possible outputs are

$$[+\ 0] \qquad [-\ 0] \qquad [0\ +] \qquad [0\ -]$$

It is apparent that the construction of constant-amplitude vectors is not possible. Next, consider the $(4 \times 4)$ matrix

$$W_4 = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix} \tag{6.59}$$

Note that $w_4[1]$ and $w_4[3]$ (as well as $w_4[2]$ and $w_4[4]$) siblings[9]. All possible sums of $w_4[1]$ and $w_4[3]$ are

$$[+\,+\ 00] \qquad [00\ +\,+] \qquad [-\,-\ 00] \qquad [00\ -\,-]$$

It is obvious that the above sums are identical to those in eqn. (6.6.5) when Walsh chips are treated as square waveforms; i.e., with $(++)$" in place of $(+)$. Similarly, for $w_4[2]$ and $w_4[4]$ all possible sums are

$$[+\,-\ 00] \qquad [00\ +\,-] \qquad [-\,+\ 00] \qquad [00\ -\,+]$$

They are also identical to the sums in eqn. (6.6.5) when the basic unit $(+-)$ is substituted for $(+)$. In every summation we notice that the "zero" output occurs in either the first or the second half of the vector, depending on the values of the input sequence. This result is an innate property of *recursive* Walsh codes since a pair of children codes (**xx** and $\mathbf{x}\bar{\mathbf{x}}$) are generated by a mother code (**x**) by concatenation and inversion. We now have enough knowledge to construct the parity-check equation of a precoder:

1. We first freely choose any pair of sibling Walsh codes. Their data modulated sum introduces a zero vector in the beginning or latter half of the output vector.

2. We then choose another pair of sibling codes. One of them is data modulated while the other is parity-checked such that it inserts the zero vector in the opposite half. Thus, the sum of all data modulated vectors is constant-amplitude.

[9] In general, $w_N[k]$ and $w_N[k + N/2]$ are siblings. See Ch. 4, Sec. 4.5.3 for a discussion on the tree structure of Walsh codes.

It is easy to see that the above construction is valid for the sum of any four Walsh codes of arbitrary length N. The above procedure for $M = 4$ parallel channels can be generalized for $M = 16$ and higher $M = 4^m$, $m > 2$. For the $(16 \times 16)$ matrix, it can be shown that the sums of siblings are (assuming the weights are all 1's)

$$\begin{aligned}
v_1 &= w_{16}[1] + w_{16}[9] = [++, ++, ++, ++, 00000000] \\
v_2 &= w_{16}[2] + w_{16}[10] = [+-, +-, +-, +-, 00000000] \\
v_3 &= w_{16}[3] + w_{16}[11] = [++, --, ++, --, 00000000] \\
v_4 &= w_{16}[4] + w_{16}[12] = [+-, -+, +-, -+, 00000000] \\
v_5 &= w_{16}[5] + w_{16}[13] = [++, ++, --, --, 00000000] \\
v_6 &= w_{16}[6] + w_{16}[14] = [+-, +-, -+, -+, 00000000] \\
v_7 &= w_{16}[7] + w_{16}[15] = [++, --, --, ++, 00000000] \\
v_8 &= w_{16}[8] + w_{16}[16] = [+-, -+, -+, +-, 00000000]
\end{aligned}$$

We did not show the outcome of the sum for other input combinations. It is, however, straightforward to see that the effect of different input sequences (e.g. $+-$, $--$, $-+$) is t change the signs or swap the zero vector (from second half to first half) or a combination of both. Note that on the first half, we have created a Walsh matrix of length 8. Note also that $s_k$ and $s_{k+4}$ are sibling pairs. Their sums are

$$\begin{aligned}
s_1 &= v_1 + v_5 = [++, ++, 00, 00, 00000000] \\
s_2 &= v_2 + v_6 = [+-, +-, 00, 00, 00000000] \\
s_3 &= v_3 + v_7 = [++, --, 00, 00, 00000000] \\
s_4 &= v_4 + v_8 = [+-, -+, 00, 00, 00000000]
\end{aligned}$$

We now have enough knowledge to construct the parity-check equation of a precoder:

1. We first freely choose two input bits for any pair of sibling Walsh codes. For convenience, we choose the first sibling pair $w_{16}[1]$ and $w_{16}[9]$. Depending on the value of the 2-bit input sequence (say, $[x_1\, x_2]$), the zero vector (say $Z_1$) will lie in the first or second half of the modulated vector. Let the $k^{th}$ output vector of a pair of siblings be $v_k$. (For easy tractability, we will give an example by letting $v_1 = [x\ x\ 0\ 0]$, i.e., the zero vector lies in the second half. The symbol "x" denotes any non-zero vector.)

2. The input bits $[x_3\, x_4]$ for the next pair of siblings $w_{16}[5]$ and $w_{16}[13]$ are freely chosen. Here we have two different outcomes: its zero vector (say, $Z_2$) lies in the same half as $Z_1$, or it lies in the other half. We consider only the first case where both zero vectors $Z_1$ and $Z_2$ lie in the same half. The algorithm for the other case follows along the same reasoning.

(In our example $v_5 = [x\ x\ 0\ 0]$.) Their sum $s_1 = (v_1 + v_5)$ now has a zero vector $Z_3$ (of length 4) either in the first or the second "quarter." (In our example, set $s_1 = [x\ 0\ 0\ 0]$.) In order to maintain constant-amplitude output, the next sum $s_2$ must have a zero vector in the same half as $Z_1$, and another zero vector (of length 4) in the other quarter of $Z_3$. For the next set of siblings, say $w_{16}[2]$ and $w_{16}[10]$, two input bits $[x_5\, x_6]$ are freely chosen. If the output $v_2$ has a zero vector $Z_4$ in the same half as $Z_1$, then its complementary sibling pair must be modulated such that their sum $s_2$ has a zero vector in the other quarter of $Z_3$. Let's assume this is the case. Then the sibling pair $w_{16}[3]$ and $w_{16}[11]$ are modulated by two parity-check bits. Hence, we have completed constant-amplitude output in one-half of the final output.

3. The next pair of siblings, $w_{16}[4]$ and $w_{16}[12]$ (and similarly for its complementary pair $w_{16}[5]$ and $w_{16}[13]$) must have corresponding zero vectors $Z_5$ and $Z_6$ in the other half of $Z_1$. To satisfy this constraint, only one of two input bits is free. Let's denote these free input bits by $x_7$ and $x_8$.

4. For the next set of siblings $w_{16}[6]$ and $w_{16}[14]$, only one of two input bits is free since its zero vector lies in the same half as $Z_5$. Denote this bit as $x_9$. Their sum $s_3 = v_3 + v_7$ has a zero vector $Z_7$ in one of the two possible quarters.

5. For the last remaining pair of siblings $w_{16}[7]$ and $w_{16}[15]$ and its complementary pair $w_{16}[8]$ and $w_{16}[16]$, all four input bits must be parity-check such that their sum $s_4$ has a zero vector in the same half as $Z_5$ and another zero vector in the other quarter of $Z_7$. This step completes the constant-amplitude property on the other half.

Thus far we have showed precoding procedures that result in constant-amplitude output vectors. In particular we give modulation algorithms for 4- and 16-parallel PC-SG transmission modes. Procedures for other transmission rates where the number of parallel channels equals $M = 4^m$ can be derived analogously. We now address the question of whether other rates besides $M = 4^m$ for feasible for PC-SG constant-amplitude transmission. We use $M = 8$ as an example. For the $(8 \times 8)$ matrix, it can be shown that the sums of siblings are (assuming the weights are all 1's)

$$v_1 = w_8[1] + w_8[5] = [+\,+\,+\,+,\ 0000]$$
$$v_2 = w_8[2] + w_8[6] = [+\,-\,+\,-,\ 0000]$$
$$v_3 = w_8[3] + w_8[7] = [+\,+\,-\,-,\ 0000]$$
$$v_4 = w_8[4] + w_8[8] = [+\,-\,-\,+,\ 0000]$$

It is obvious that the sum of two output vectors listed above does not result in a constant-amplitude vector. Thus the precoding procedure is not possible for $M = 8$.

### Linear Parity-Check Coding

Based on above discussions on precoding, we witness some other interesting properties of Walsh modulated sequences as tabulated in Table 6.1. Note that each of the remaining eight input vectors of $(S_1)^c$ in column 1 has the same first three bits as a vector $\mathbf{v}$ in $T_1$. For example, [0000] and [0001] have the same first three bits. We also notice that for every vector $\mathbf{v}$ in $T_1$, there exists a complementary (constant-amplitude) vector $\mathbf{u}$ that belongs to $T_2$. (The complement of $\mathbf{v}$ (in column 3) is listed in the same row in column 5.) For example, $[2,2,2,2] \in T_2$ is the complement of $[2,2,2,-2]$. It may then be possible to linearly map every input vector $\mathbf{x}_i \in (S_1)^c$ to a vector $\mathbf{u}_i \in T_2$. It can be seen from columns 3 and 5 that all vectors in the sets $T_1$ and $T_2$ have the same weight distribution. We also observe that any vector $\mathbf{U}$ of $T_2$ can be generated from a three-bit input vector $\mathbf{Y}$ using the following matrix equation:

$$\mathbf{U} = \mathbf{Y} \otimes \mathcal{G}_p \tag{6.60}$$

where

$$\mathcal{G}_p = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{6.61}$$

is the generator matrix of a linear, systematic parity-check code. The symbol $\otimes$ denotes vector multiplication in binary Galois field GF(2). (The truth tables of GF(2) arithmetic is given in Table 4.1.) We know immediately that due to the linearity of block coding, the output vector $\mathbf{U} \in T_2$. Furthermore, any precoded Walsh modulated sequence can be generated using the same parity-check generator matrix by modifying the modulating Walsh matrix:

$$\mathcal{W}_N^* = \left[\mathbf{w}_1^T \; \mathbf{w}_2^T \; \mathbf{w}_3^T \; -\mathbf{w}_4^T\right]^T \tag{6.62}$$

such that the fourth row vector is sign-inverted. For any three-bit input vector $\mathbf{Y}$, we denote the output of a linear block code with a generator matrix $\mathcal{G}_p$ by

$$\mathbf{Q} = \mathbf{Y} \otimes \mathcal{G}_p$$

Its level-shifted antipodal sequence is denoted as $\mathbf{q}$. It is straightforward to show that

$$\mathbf{v} = q \cdot \mathcal{W}_N^*$$

is a precoded, Walsh modulated sequence of $T_1$. The orthogonal modulating matrix is the modified Walsh matrix of eqn. (6.62). If we, however, repeat the above steps with another orthogonal matrix (the identity matrix $\mathbf{I}$), it is readily shown that

$$\mathbf{v} = q \cdot 2\mathbf{I}_N$$

the output vector $\mathbf{v} \in T_2$. Since precoding involves inversion of the sum of three input bits, its coding operation is non-linear; that is, the output vector cannot be expressed as in terms of an input vector and a generator matrix. However, as shown above, by modifying the conventional Walsh matrix, we show that precoding is in fact a parity-check encoding followed by orthogonal modulation. We then show that parity-check encoding followed by another orthogonal (identity) matrix results in a dual code. Since the (code) vectors from both sets $T_1$ and $T_2$ have the same weight distribution, their (Euclidean or Hamming) distance properties are also the same.

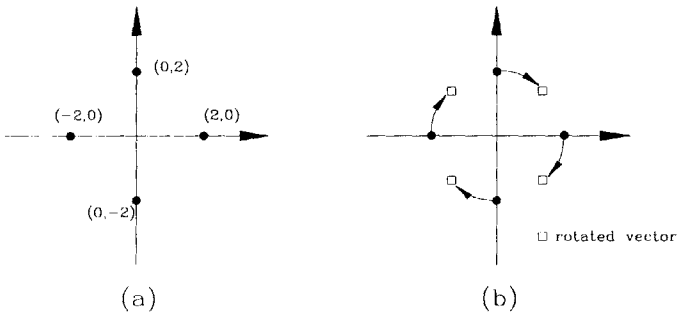



**Figure** 6.20: Constellation of Walsh modulated output (a) original (b) $\pi/4$ rotated

## Relation between OVSF and Multi-Code Transmission Schemes

For a special case where the number of parallel channels $M = 2^m$ the multi-level output vector can be rotated such that the resulting sequence is antipodal and constant-amplitude. We first describe this technique for the simplest case of $M = 2$ and later generalize it for larger $M$. For the purpose of illustration we assume $M = N$. For a 2-bit input sequence, all four possible Walsh modulated output vectors are shown in Fig. 6.20(a) as points in a two-dimensional constellation. The $(x, y)$ basis vectors are $(1, 0)$ and $(0, 1)$. This constellation resembles that of 4-QAM or QPSK. It is well known that by $\pi/4$ rotation along the origin, the new signal points become constant-amplitude. This is illustrated in Fig. 6.20(b). It is equivalent to $\pi/4$-QPSK set. The $\pi/4$ (clockwise) rotation matrix $\mathbf{R}_2$ is

$$\mathbf{R}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \mathbf{X}_2 \tag{6.63}$$

A rotation matrix belongs to a special class of square matrices known as orthonormal matrices. A higher-order rotation matrix can be generated recursively according to the following rule:

$$\mathbf{R}_{2N} = \frac{1}{\sqrt{2N}} \begin{bmatrix} \mathbf{X}_N & \mathbf{X}_N \\ \mathbf{X}_N & \overline{\mathbf{X}_N} \end{bmatrix}$$

All orthonormal matrices $\mathbf{Q}$ satisfy the following identity:

$$\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{-1}$$

That is, an inverse exists and it is equal to its transpose. For reflection matrices $\mathbf{T}$ such as the recursive Walsh matrix, we already know that it satisfies another equality:

$$\mathbf{T}^{\mathrm{T}} = \mathbf{T}^{-1} = \mathrm{k}\,\mathbf{T}$$

where k is a normalization constant. We will set $\mathrm{k} = 1$ since the matrix can always normalize the original reflection matrix such that $\mathbf{T}^{-1} = \mathbf{T}$. Note that for a rotation matrix, $\mathbf{R} \neq \mathrm{k}\mathbf{R}^{-1}$. We see immediately that the Walsh matrix (due to its reflective property) can be used for both parallel channel modulation and demodulation. Another reflection matrix is the identity matrix.

We now show that the reflection property also applies to the product matrix of Walsh and rotation matrices

$$\begin{aligned}\mathbf{P}_2 &= \mathcal{W}_2 \cdot \mathbf{R}_2 \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = 2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\end{aligned}$$

For an arbitrary $\mathrm{N} = 2^{\mathrm{m}}$ it can be shown by induction using eqn. (6.63) that the product matrix $\mathbf{P}_{\mathrm{N}}$ is an alternating identity matrix $\mathbf{J}_{\mathrm{N}}$ such that

$$\mathbf{D}_{2\mathrm{N}} = \sqrt{2\mathrm{N}} \begin{bmatrix} \mathbf{J}_{\mathrm{N}} & \mathbf{0}_{\mathrm{N}} \\ \mathbf{0}_{\mathrm{N}} & \mathbf{J}_{\mathrm{N}} \end{bmatrix}$$

where $\mathbf{0}_{\mathrm{N}}$ is an $(\mathrm{N} \times \mathrm{N})$ zero matrix and

$$\mathbf{J}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

At the demodulator the inverse of the product matrix is

$$\begin{aligned}\mathbf{J}^{-1} &= \left(\mathcal{W} \cdot \mathbf{R}\right)^{-1} = \mathbf{R}^{-1} \cdot \mathcal{W}^{-1} \\ &= \mathbf{R}^{\mathrm{T}} \cdot \mathcal{W}^{\mathrm{T}} = \left(\mathcal{W} \cdot \mathbf{R}\right)^{\mathrm{T}} = \mathbf{J}^{\mathrm{T}}\end{aligned}$$

Since $\mathbf{J}$ is diagonal, it is also a reflection matrix ($\mathbf{J}^{-1} = \mathbf{J}$). Note that in general, for two arbitrary orthonormal matrices,

$$\mathbf{Q}_1 \cdot \mathbf{Q}_2 \neq \left(\mathbf{W} \cdot \mathbf{R}\right)^{-1}$$

Based on above results, a rotated, Walsh modulated output vector can be expressed as

$$\mathbf{s}_{\mathrm{R}} = \mathbf{x} \cdot \mathcal{W} \cdot \mathbf{R}$$

where $\mathbf{x}$ is a level-shifted antipodal binary input sequence. If we replace the rotation matrix by a normalized Walsh matrix (a reflection matrix), the reflected output is

$$\begin{aligned} \mathbf{s}_T &= \mathbf{x} \cdot \mathcal{W} \cdot \mathbf{T} \\ &= \frac{1}{\sqrt{N}} \mathbf{x} \cdot \mathcal{W} \cdot \mathcal{W} = \frac{1}{\sqrt{N}} \mathbf{x} \end{aligned}$$

This, in fact, is the output of OVSF modulation. The reflected output along the dashed line m—n is depicted in Fig. 6.21.

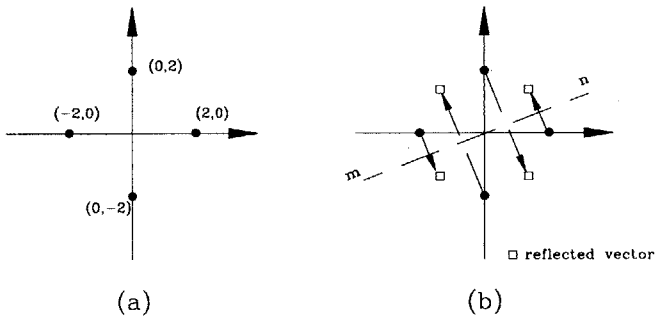



**Figure** 6.21: Constellation of Walsh modulated output (a) original (b) reflected

If the channel and its induced additive noise are transparent to orthogonal transformations such as rotation and reflection, both OVSF and multi-code transmission modes have the same system performance. Note that an orthogonal transformation does not change the distance property of a signal set. An example of a transparent channel is the ideal AWGN channel since the Gaussian noise process is spherically symmetric (or rotationally invariant). We illustrate this model in Fig. 6.22.

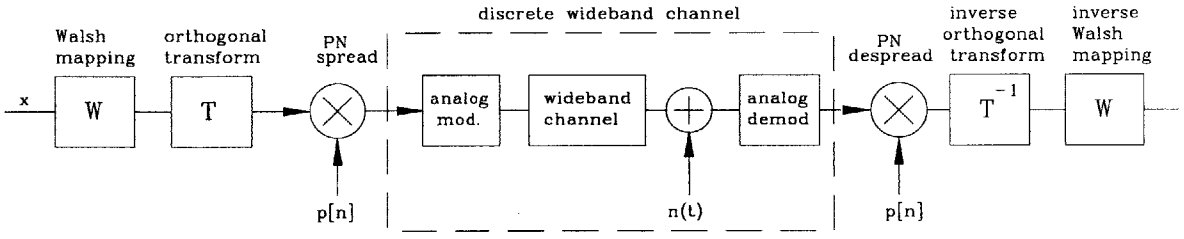



**Figure** 6.22: Modified model of a SSM communication system with the inclusion of linear orthogonal transformations

# 7

# RESOURCE ALLOCATION ALGORITHMS

## Summary

In Chapter 3 we show that a resource allocation policy based on code-division multiplexing provided several advantages over time and frequency sharing. The equivalence of orthogonal CDM to TDM and FDM schemes in terms of achievable rate region is predicated upon the assignment of many parallel orthogonal code channels to each receiver in a broadcast setting. In Chapter 6 it is proven that it may not be necessary after all to use parallel binary orthogonal code channels; instead, a single binary orthogonal code with variable spreading and duty cycle suffices for an arbitrary rate transmission. Bit-error-rate analysis of uncoded binary data in Chapter 6 shows that both single and multiple code channel schemes are equivalent in the sense that one can be substituted in place of the other without affecting the BER of the designated receiver or other interfering receivers in the system. Therefore, it is concluded that CDM with single binary code channels is the preferred option due to its simpler, and readily scalable receiver design structure.

In this chapter it will be shown that there is a catch to this equivalence. The BER analysis is a measure of the link quality per channel. It does not measure the overall throughput of the system. The throughput of single-channel, reduced gain CDM scheme (SC-RG CDM) can still be smaller than its alternative parallel-channel, single gain CDM (PC-SG CDM) if a condition called *orthogonal code blocking* is not removed. Code blocking is a code assignment constraint that occurs only when recursive orthogonal codes are used as signature sequences. It is irrelevant when the set of orthogonal codes are non-recursive. However, it is also important to note that recursive orthogonal codes facilitate seamless multi-rate data transmission. Details on the nature of code blocking and procedures for its removal are presented in this chapter.

There are two alternative methods of removing code blocking: horizontal and vertical code reassignments. Both involve the release of an assigned binary code and reassignment of another binary code. Hence, this is a (code) scheduling task that must be handled by a broadcaster in coordination with its target receivers. If the newly assigned code is a relative of the released code, we call such procedure *vertical reassignment*. If the reassigned code is a sibling, the procedure is called *horizontal reassignment*. We show that the former is suited for bursty data connections where a large ratio of peak-to-average transmission rate is tolerated. The latter scheme supports zero latency; thus, it is the preferred option for delay-sensitive constant-rate applications such as voice and live video. In practice, the central controller (the broadcaster) in a CDM based communication system must adopt a single resource allocation policy that efficiently handles both bursty and constant-bit-pipe connections. The implication is that the chosen reassignment policy may need to support both horizontal and vertical reassignments at the same time. Such a procedure will be called *combined reassignment* policy. It becomes apparent that SC-RG CDM with combined reassignment is not only equivalent to PC-SG CDM in terms of system throughput, but is also the preferred multi-rate transmission scheme due to its natural ability to support statistical multiplexing of mixed (bursty and constant rate) data channels.

We summarize the main results below:

- OVSF code blocking is unique only to recursive orthogonal codes. We wish to eliminate code blocking for two reasons: First, in a single-channel multi-rate CDM system with constant rate connections, the network capacity is reduced by as much as 25% (due to code blocking) compared to an alternative CDM multi-rate system that employs parallel orthogonal code channels. We, however, do not promote parallel channel CDM (multi-code CDM) due to its increased receiver complexity for multi-rate reception. Second, with a large capacity penalty due to code blocking in SC-CDM, a competing multiplexing scheme such as TDM may seem more appropriate for multi-rate communication.

- Horizontal code reassignment is ideal for constant rate, delay-intolerant connections. We devise a code reassignment algorithm that is optimal in minimizing the number of reassignments. The algorithm prefers over-population of the code tree.

- Vertical code reassignment is ideal for bursty connections that are delay tolerant. We propose an efficient algorithm that maximizes bandwidth utilization and also guarantees finite bounded delay. The algorithm prefers under-population of the code tree.

- In order for a SC-CDM based system to support both constant rate and bursty connections, OVSF codes must be reassigned both horizontally and vertically. At first glance it appears these assignment policies are in conflict since one prefers under-population while the

other desires over-population. We prove otherwise by describing a combined assignment strategy that affords statistical multiplexing of mixed-rate channels —a provision that is not readily adaptable in a parallel channel CDM based system.

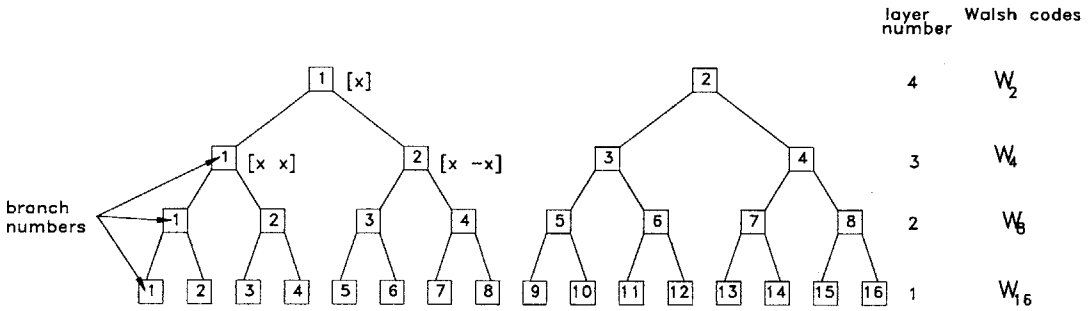



**Figure** 7.1: Binary tree structure of OVSF codes

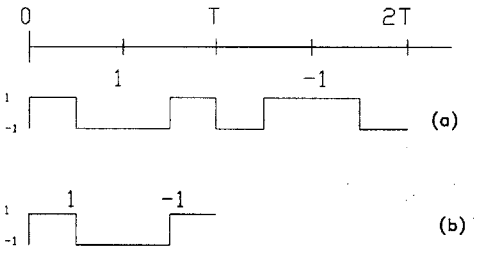



**Figure** 7.2: Similarity between two Walsh coded sequences. Each sequence cannot be decoded uniquely since one Walsh code is an ancestor of the other.

## 7.1 Orthogonal Variable Spreading Factor Codes

The properties of recursive binary orthogonal codes of variable length are already discussed in Chapter 4. A quick review is given here. Fig. 7.1 shows an orthogonal variable spreading factor (OVSF) code tree with 16 leaves. All (16+8+4+2)=30 nodes in the tree are valid OVSF codes. The topmost code is the *root* code. From the root code [X], its two *children* codes — [XX] and [X$\overline{\text{X}}$]— can be generated by concatenation. If [X] is a vector of length $n$, we assume [X] $\in (\mathbb{D}_2)^n$; i.e., an OVSF code is antipodal binary-valued. All OVSF codes in the same row are called *cousin* codes. All codes linking code A to the root code (inclusive) through the shortest path are called *ancestor* codes of A. All codes branching from an OVSF code are its *descendant* codes. The *relatives* of a code comprise of both its descendants and ancestors. (A warning

about our convention: A cousin code is not a relative!) *Two codes are orthogonal if and only if one is not a descendant of the other.* Thus, a code and its relatives are not orthogonal. Once a particular code is assigned (to a receiver), simultaneous use of its descendant or ancestor code(s) is not allowed because the encoded sequences may be indistinguishable. An example is shown in Fig. 7.2 where a binary data sequence $(+1,-1)$ is encoded (or modulated) using two different Walsh codes. The top waveform (labelled (a)):

$$(+1,-1,-1,+1,-1,+1,+1,-1)$$

is the resulting encoded sequence modulated by a Walsh code $W_4(4) = (+1,-1,-1,+1)$. The encoded sequence spans over an interval $[0, 2T]$. The bottom waveform, labelled (b):

$$(+1,-1,-1,+1)$$

uses a shorter length Walsh code $W_2(2)$. A shorter length code supports a higher data rate, and this is the reason why the encoded sequence spans only over the interval $[0, T]$. We see immediately that within the interval $[0, T]$, both waveforms (a) and (b) are identical. In essence, the most important property of OVSF codes is the constraint that only mutually orthogonal codes can used for simultaneous data transmission and channel multiplexing.

Another insightful interpretation of OVSF codes is its equivalence to binary prefix-free codes when $[X] \in \mathbb{F}_2$. By definition a binary code is prefix-free if its prefix is not the same as another valid binary code. Because OVSF codes are generated recursively, a code and its relatives have the same prefix. In our example, $W_2(1) \Leftrightarrow [X] = 0$ and $W_2(12) \Leftrightarrow [\overline{X}] = 1$,

$$W_2(1) \Leftrightarrow (0) \qquad W_4(2) \Leftrightarrow (01) \qquad W_8(4) \Leftrightarrow (011)$$

Note that in binary prefix-free code construction, a child code is tagged either '0' or '1'. (Cautionary note: Binary prefix-free codes described here are "conventional" type. See page 191 for a different description based on binary Hadamard sequences.) Since they all have the same prefix, they are not mutually orthogonal. However, Walsh codes that are not relatives, for example, $W_4(4) \Leftrightarrow (11)$ and $W_2(1) \Leftrightarrow (0)$ are orthogonal. If we label each node in the binary tree by an binary vector over $\mathbb{F}_2$, we see that the assignment of OVSF codes for channel multiplexing is equivalent to the construction of variable-length binary prefix-free codes. (Prefix-free codes are discussed in [22]).

If the maximum spreading gain is $N_{max}$ and the information rate of a leaf code is R bps, the chip rate is $N_{max}$ R cps (chips per second). Equivalently, each information bit $\{+1,-1\}$ is mapped to a leaf code A or its inverse $\overline{A}$ of length $N_{max}$. Since two binary orthogonal codes of length 2L are generated from their *parent* code of length L, OVSF codes are called "recursive." It is this recursive property of OVSF codes that allows a code channel to support variable transmission rates. Since the underlying communication system uses a single chip rate for

transmission, we deduce immediately that the parent code of a leaf supports information rate of 2R bps. By going one step further, we know that the grandparent code of a leaf code supports an information rate of 4R bps and so on. In general, an OVSF code channel can vary its data rate by powers of 2 by sliding *vertically* to its relative code. There is no change in information rate if the sliding is *horizontal.*

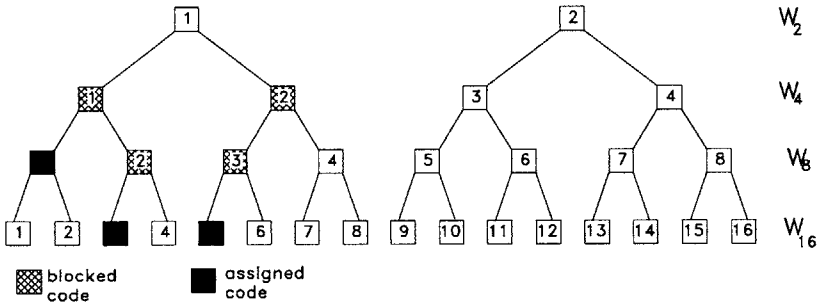



**Figure** 7.3: OVSF code blocking

### 7.1.1 Code Blocking in OVSF-CDM

In parallel channel-single gain CDM (PC- SG CDM), k mutually orthogonal leaf codes are allocated to a receiver requesting an information rate of kR bps. In OVSF-code-based single channel-CDM (SC-CDM), the system may not be able to support the same rate even though k leaf codes are vacant. Consider the example shown in Fig. 7.3. Assume the system is code-limited; i.e., it can support a maximum capacity of $N_{max}$ R bps with $N_{max}$ leaves, and each leaf supports R bps. If codes $W_8(1)$, $W_{16}(3)$ and $W_{16}(5)$ are already assigned, the "used" capacity in the left sub-tree is 4R bps, and the unused capacity is $(8-4) = 4R$ bps. However, codes $W_4(1)$, $W_4(2)$, $W_8(2)$ and $W_8(3)$ are *blocked* by their respective descendant codes. Only $W_8(4)$ of rate 2R is available for assignment, even though the free capacity in the sub-tree is 4R. As a result, a new connection requesting 4R bps is blocked in SC-CDM. In PC-CDM the call is supported by assigning leaves $W_{16}(4)$, $W_{16}(6)$, $W_{16}(7)$ and $W_{16}(8)$.

**8 Definition** *We define OVSF code blocking as the condition that a new connection cannot be supported although the system has excess capacity to support the requested information rate.*

## 7.2 Dynamic Assignment of OVSF Codes

It is important to note that code blocking only occurs in higher layer codes. It is evident from the code tree (Fig. 7.3) that the higher the requested data rate, the larger the blocking

probability. Overall, the system throughput of SC-RG CDM will be much less than that of PC-SG CDM due to code blocking. For practical reasons a resource allocation policy that requires a single-channel receiver design —regardless of the information rate— is desired. Our immediate concern is to develop a procedure where SC-RG CDM can maintain the same system throughput as PC-SG CDM with some modifications to its code assignment policy. In the following sections, we propose two very different alternatives that remove code blocking completely. They both require the release and reassignment of OVSF codes while the connection is active; i.e., they are make-before-break options where information transfer is uninterrupted.

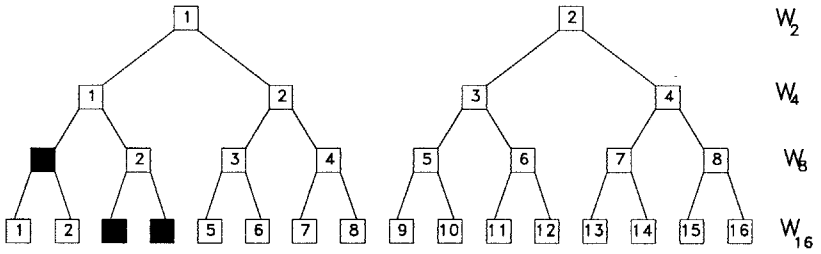



**Figure** 7.4: Complete overpopulation of OVSF sub-tree

### 7.2.1 Horizontal Reassignment

The code blocking case in the left sub-tree of Fig. 7.3 can be removed if the active channel with $W_{16}(5)$ releases its assigned code, and it is reassigned the vacant $W_{16}(4)$. Now, $W_4(2)$ of rate 4R is available for a new connection. The new topology of the sub-tree, as shown in Fig. 7.4, is called *completely overpopulated.* The assigned codes are packed as densely as possible. Coding blocking does not occur when a (sub) tree is completely overpopulated. Because the code sliding is horizontal, all active channels retain their information rates after code reassignments. It is easy to see that this policy is suitable for applications that require constant bit-rate connections with zero latency. The optimality criterion is the minimization of the number of disturbances (i.e., the number of code reassignments) to active channels such that a new call with its requested rate can be supported. We present such an optimal horizontal code assignment algorithm in Sec. 7.3.

### 7.2.2 Vertical Reassignment

Let us revisit the code blocking case in the left sub-tree of Fig. 7.3. Without horizontal reassignments, a user requesting rate 4R has two options: it must be content with an assigned

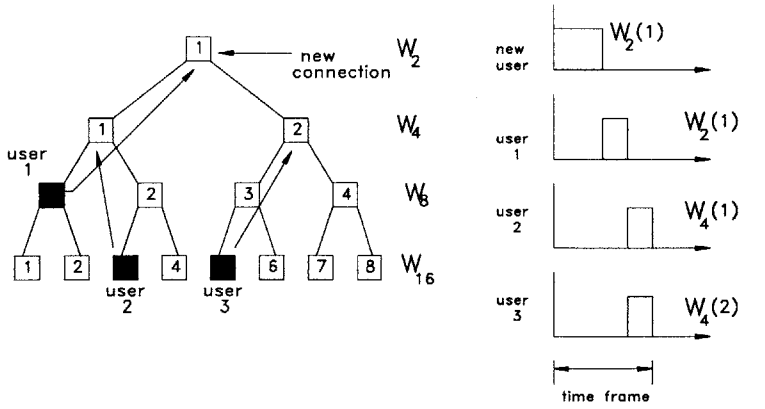



**Figure** 7.5: Vertical reassignment of OVSF codes

code $W_8(4)$ of 2R, or it can be assigned a parent code $W_4(1)$ of 8R with 50% duty cycle for an average rate of 4R. The latter approach has immediate consequences on the rest of the system with active channels. All remaining active channels must also slide their assigned codes to their respective parent codes. This scenario is shown in Fig. 7.5. Obviously, this is a combination of vertical code sliding and time sharing. In the extreme case (Fig. 7.6) where each active channel slides to the root code, the resulting policy is clearly time-division multiplexing. A major drawback of TDM is its large peak-to-average information rate ratio, resulting in a large delay gap between information bursts. It is clear that CDM with vertical code sliding and time-sharing is a compromise between CDM with horizontal reassignment (no latency) and TDM (maximum latency). Its attractive features are the complete removal of code blocking and reduced complexity in code scheduling since no code swapping is involved; instead, a receiver needs to adjust its correlation window in accordance with the information rate. In Sec. 7.4, we describe an efficient vertical reassignment algorithm that minimizes latency and backlog of information packets.

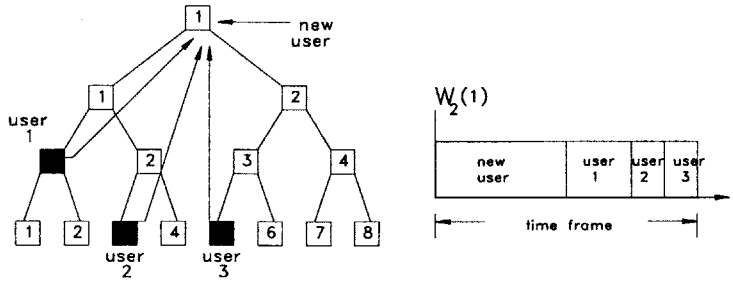



**Figure** 7.6: Extreme case of vertical reassignment: time-division multiplexing

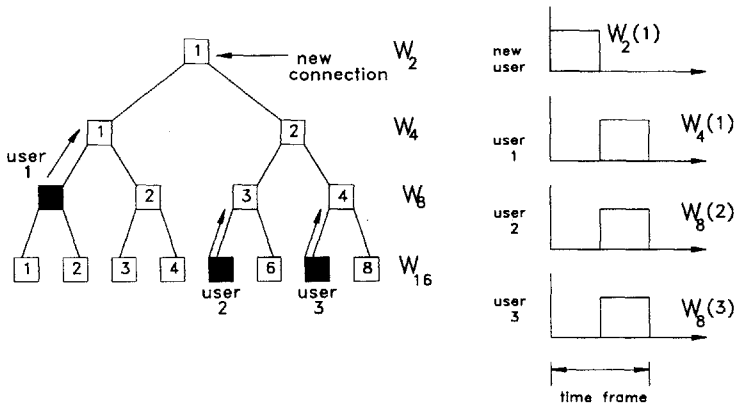



**Figure** 7.7: complete Under-population of OVSF sub-tree

### 7.2.3 Combined Reassignment

In the horizontal reassignment policy, code reassignments take place only once when a new call arrives. This is the call-by-call basis reassignment. In vertical reassignment, the time sharing operation takes place in every time frame. Because of the burstiness of information packets from each information source, the reassignment strategy must be preformed dynamically on a packet-by-packet basis. It is also possible to combine both horizontal and vertical reassignment policies. The benefits of this synergic approach is in the reduction of horizontal reassignments (reduction in control signal overhead), statistical multiplexing capability (maximizing total capacity) and increased smoothness (less choppiness) of information transfer (bounded latency). The downside is an increase in complexity for code scheduling —mainly due to conflicting objectives in desirable topology configurations for vertical and horizontal reassignment policies. Let us elaborate further. In horizontal reassignment policy, the optimality criterion is to support a new call with the minimum number of reassignments for active channels. The ideal topology of a code tree is complete overpopulation where assigned codes are all concentrated in a portion of the tree such that the rate of a vacant code is maximized. (This is illustrated in Fig. 7.4.) For vertical reassignments, the opposite is desired; i.e., a completely underpopulated code tree (see Fig. 7.7) has the smallest peak-to-average information rate ratio. The ultimate goal of a combined assignment policy is to strike a balance between horizontal and vertical reassignments.

It is foreseen that future mobile communications systems must be able to support all kinds of mobile data applications that are bursty, steady, delay-sensitive etc. A CDM broadcaster that supports only one –but not both– reassignment policy is not viable in practice. A simple, but not very efficient, option is to partition the code tree into two sub-trees. One sub-tree is dedicated to constant-bit rate connections (with horizontal reassignment policy) while the other sub-tree

supports vertical sliding with time-sharing. We notice immediately that the peak information rate of a connection is reduced by tree partitioning. Moreover, an active channel must relinquish its active channel and queue for a new channel once its mode of information transfer is changed in nature, say from constant to bursty rate connection. We can do better than a simple tree partitioning. This topic is further discussed in Sec. 7.5.

## 7.3 DCA Algorithm for Constant-Rate Channels

For constant-rate connections that demand zero latency, the only dynamic code assignment (DCA) policy is horizontal reassignment. The algorithm involves three steps:

1. A check on the system capacity to see if a new connection with its requested rate can be supported by reassigning OVSF codes for ongoing connections. If not, the new connection is blocked.

2. If there is enough excess capacity, the target OVSF code must be evaluated. Based on the requested information rate, we are required to only search for sibling codes from a single layer of the tree. The code that minimizes the number of reassignments is our target code.

3. Once the target code is determined, we must reassign all of its actively assigned descendant codes. This step may trigger the reassignment of OVSF codes in other parts of the tree.

### 7.3.1 Optimal Dynamic Code Assignment Algorithm

Our goal is to design an algorithm that minimizes the number of necessary reassignments of occupied codes to support a new connection. This is our chosen optimality criterion. We have selected this criterion because it causes the least amount of interruption to ongoing connections. If another optimality criterion is desired —such as overpopulation of a code tree, then the resulting reassignment algorithm will be very different from what we propose here. The key idea underlying our optimal algorithm is to associate a *cost function* with each candidate *branch*, and to assign the root code of a minimum-cost branch to the new call. (A xR branch has an OVSF code of rate xR as its root code.) The cost function is only defined when there is excess system capacity, i.e., when the capacity-test equation (7.2) is a strict inequality —before the new call is accepted.

9 **Definition** *Under the assumption that the system has excess capacity, the cost of reassigning an occupied code* C *is defined as the minimum number of code reassignments (including* C *itself) necessary to reassign* C *to some other branch* D *and subsequent reassignments of all active descendant codes in branch* D.

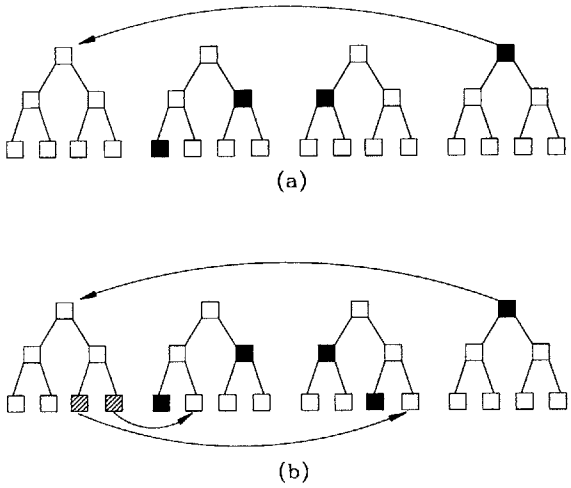



**Figure 7.8:** Cost of reassigning a 4R code. In (a), the cost is 1 since there is an empty branch. In (b), the total cost is 3 since two leaves (each of cost 1) must also be reassigned.

Since the reassignment of a leaf code of rate R results in no additional code reassignments but itself, by definition its cost is 1. Fig. 7.8 illustrates the cost of reassigning a 4R code. As can be seen from the figure, the cost depends on the topology of other branches. When there is an immediate vacancy in another branch, the cost is only 1 (Fig. 7.8(a)). The cost is 3 for the topology in Fig. 7.8(b). It is equally useful to measure the cost function of a branch; this is the sum of costs of reassigning all active codes contained in the branch. We see immediately that the maximum cost of any 2R branch is 2, that of any 4R branch is 4 and so on. The challenge of implementing an optimal algorithm lies in searching for a minimum-cost branch efficiently. This is because the cost of reassigning an occupied code in one branch is a function of the cost of reassigning codes in other branches. A straightforward implementation of the algorithm using the cost function would compute the cost of all branches using a recursive procedure. This would require an exponential amount of computations because the procedure must be iterated for descendant codes as well. Fortunately, several key observations (Theorems 7.2 and 7.3) regarding the optimality of code reassignment can help us significantly reduce the complexity of searching for a minimum-cost branch. Before we present an efficient search algorithm for a minimum-cost branch, we first outline the steps in the execution of our optimal DCA algorithm:

1) Check if the new call with rate kR bps can be supported, i.e. its requested rate is within the system capacity (Theorem 7.1). If so, go to Step 2. If not, block the call.

2) Find a minimum-cost branch where the root code supports rate kR bps. (Theorem 7.2 ensures that there is no need to search for branches with higher root code rates.) This step

can be achieved by using, for example, an optimal topology search algorithm described later in this section.

3) Once a minimum-cost branch is found, the root code of the branch is assigned to the new call. If the branch is empty, the root code is assigned to the call and the process is complete. Otherwise, it is necessary to reassign the (occupied) descendant codes of the branch as described in step 4.

4) Reassign to another branch the code with the highest data rate among the descendant codes first. If there are more than one descendant code with the highest data rate, it can be chosen arbitrarily among them. (This procedure is sufficient to maintain optimality of the algorithm. See Theorem 7.3.)

5) To reassign a code (with the highest data rate among the descendant codes), go back to Step 2 by treating it as a new call requesting its rate. The above algorithm guarantees that a code is assigned to a minimum-cost branch in each iteration. It can be proven that the resulting total number of code reassignments is always minimum.

**Capacity Test: Code-limited Case**

Let R bps be the data rate supported by a leaf code. The code tree consists of $N_{max}$ leaves. Then, in an idealized code-limited (single-cell) case, the system capacity is equal to $N_{max}R$ bps, where $N_{max} = 2^m$, $m \in \mathbb{Z}_+$. This is based on the assumption that all assigned codes are mutually orthogonal, and there is no multiple access interference among them. Thus, the system has $N_{max}$ parallel single-user channels, each channel supporting a data rate of R bps. If L is the total number of users in the system and $k_i = 2^{n_i}$, $n_i \in \mathbb{Z}_+$, is the *rate factor* of user $M_i$ whose assigned data rate is $k_iR$ bps, then we must have:

$$\sum_{i=1}^{L} k_i \leq N_{max} \tag{7.1}$$

Rewriting (7.1):

$$\sum_{i=1}^{L} 2^{-r_i} \leq 1 \tag{7.2}$$

where the integer $r_i = m - n_i$. Since each assigned OVSF code in a CDM system is equivalent to a *prefix-free* binary code, the capacity check (7.2) is simply an alternate form of Kraft's inequality. We now state the following:

**Theorem 7.1** *If the requested data rate of an incoming call is within the system capacity, i.e. it satisfies Kraft's inequality, (7.1) or (7.2), the call can be supported by code reassignments.*

Proof of Theorem 7.1: It is well known [22] that given a set of integers $r_i$, $i \in \mathbb{Z}_1^L$, satisfying Kraft's inequality, prefix-free binary codes of code lengths $r_i$ can be constructed. From (7.2), OVSF codes with data rates $k_iR$ bps can then be constructed correspondingly. □

### Capacity Test: Interference-limited Case

An ideal code-limited CDM system is meaningful in a single or isolated cell wireless environment. However, in a multiple-cell wireless environment, multiple-access interference from neighboring cells must also be taken into account when computing the system (or cell) capacity. As a result, the capacity is less than $N_{max}R$ bps [40]. In this case, the capacity test using Kraft's inequality is modified as follows:

$$\sum_{i=1}^{L} 2^{-r_i} \leq \frac{1}{D} \tag{7.3}$$

where $D > 1$ is the *effective reuse number* of a CDM system in a multiple-cell environment. For example, if the total number of codes that can be supported in the forward link of a CDM system is 44 (out of 64) [40], D equals 64/44. In a code-limited case, we assume that the rate factor of user $M_i$ (of rate $k_iR$ bps) is $k_i$. In a multiple-cell environment, the rate factor $k_i$ must also take into account the additional interference from users in neighboring cells.

**Theorem 7.2** *In step 2 of the above DCA algorithm, to assign a code of rate* $kR$ *bps, it is sufficient to consider only branches of root code rate* $kR$ *bps to maintain optimality of the algorithm.*

The proof is given in Appendix 7.A.
Consider the example shown in Fig. 7.8. Assume that a second-layer code needs assignment. There are 8 second-layer branches (indexed 1 through 8 from left to right). Theorem 7.2 states that branches 7 and 8 need not be considered as possible candidates because their root (parent) code in third-layer branch 4 is already occupied. In order to assign a code with branch number 7 or 8, we must first reassign the parent code. It is more costly to reassign a higher layer code than other lower layer descendent codes (leaf codes in this case). Among the second-layer branches, 4 and 5 can also be eliminated from the candidate pool because both are already assigned.

**Theorem 7.3** *In step 4 of the DCA algorithm, reassigning a code with the highest data rate among the descendant codes first is sufficient to maintain optimality of the algorithm.*

The proof is given in Appendix 7.B.

When a minimum-cost branch is located, all its occupied descendant codes (if any) must be reassigned. It is of course possible to reassign leaf codes first. However, all leaf code reassignments have a cost of 1, and the new branch numbers can be chosen arbitrarily. Erroneous

selection of new branch numbers for leaf codes (or lower layer codes in general) can lead to higher costs for higher layer codes that need reassignments. Theorem 7.3 states that the minimum cost criterion is maintained if a code with the highest data rate is reassigned first.

### 7.3.2 Minimum-Cost Branch Search Procedures

To complete the description of our code reassignment algorithm, we present several algorithms that search for a minimum-cost branch. It is important to stress that any of the following searches can be used to locate a target branch. Some searches will indeed identify the minimum-cost branch while others will not. Regardless of the selected search procedure, the reassignment algorithm completely removes code blocking. As we shall see next, searches that locate a branch with the minimum cost are generally more complex.

#### Exhaustive Search

In an exhaustive search algorithm, the cost of every candidate branch is computed. This would require computing the cost of reassigning every (occupied) descendant code contained in a branch. Since reassigning a code to another branch would involve reassigning codes of that branch and other branches, a chain reaction type of reassignments occurs. The cost function must be computed recursively. The complexity of such recursive computations can be prohibitively large if the target code is of high information rate. The exhaustive search method, by its nature, does locate a minimum-cost branch at the cost of complexity. It also does not exploit some topological properties of binary sub-trees. We next describe two search methods that are far more efficient by taking into account the locations and patterns of assigned codes in a branch.

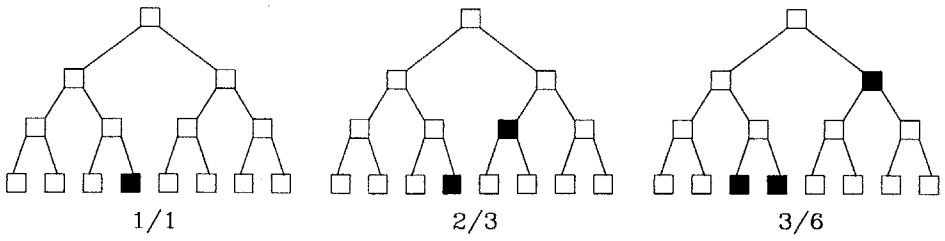



**Figure** 7.9: Code pattern description for 8R branches

#### Code Pattern Search

In a majority of cases, branches with more descendant codes and/or occupied capacity will have a higher cost. This property is exploited in the code-pattern search algorithm. First, we

define the *code pattern* A/B of a branch, which is defined in terms of A, the total number of occupied descendant codes and B, their occupied capacity. For example, three different code patterns of 8R branches in Fig. 7.9 are denoted as 1/1, 2/3 and 3/6, respectively. In most cases, a branch with code pattern A/B has a smaller cost than a branch with code pattern C/D if

a) $A < C$ or

b) $A = C$ and $B < D$

Since the code pattern of each branch is very easy to compute, an efficient heuristic search algorithm is simply to pick the branch with the smallest code pattern (based on lexicographical order). All possible code patterns for candidate 4R and 8R branches are shown in Figs. 7.22—7.29. The code pattern description for branches with higher code rate factors (16R and beyond) can be constructed in a similar fashion. The code pattern search method does not always yield a minimum-cost branch. This is mainly because the code pattern does not uniquely characterize the locations and the composition of the codes within a branch. It is possible for two branches with the same code pattern to have different costs. This occurs when one is underpopulated while the other is overpopulated. In the following, we describe an optimal search algorithm that makes use of the code pattern of a branch as well as the code structure within a branch.

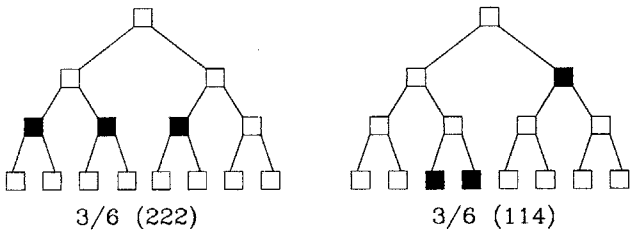



**Figure** 7.10: Extended code pattern description for 8R branches

### Topology Search

In order to describe the code structure, we would extend the code pattern description of a branch. Let us use an example to illustrate this. When two branches (say, 3/6) have the same code pattern but different descendant codes, an extension such as 3/6(114) or 3/6(222), describing the descendant code composition, is used. This is illustrated in Fig. 7.10. An extension to a code pattern is crucial in the search of minimum-cost branch because two branches with the same code pattern but different code composition do not necessarily have the same cost. (See Appendix 7.C for tables listing cost comparison between any two branches.) A branch can be characterized completely using a *grid pattern*. A grid pattern is a sequence of numbers and X's, where the numbers represent the code rate factor of assigned codes in a branch and X's show the locations of vacant R codes. Examples of grid patterns are shown in Fig. 7.11. Note

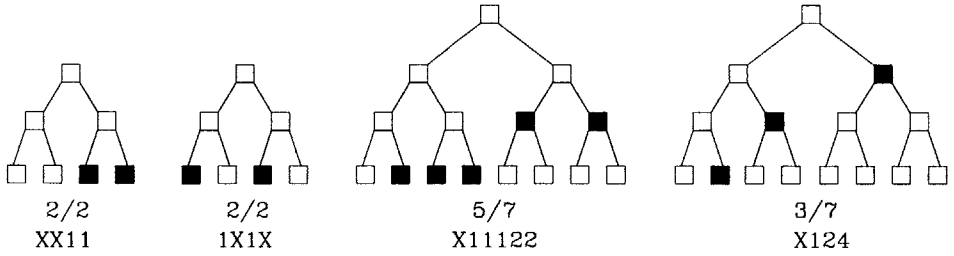



**Figure** 7.11: Grid pattern description for 8R branches

the significance of grid patterns. It shows two branches with the same code pattern (and code composition), but the cost of reassigning a 2R code into the 1X1X branch is twice that of the XX11 branch.

The topology search is considered as an extension to the code pattern search because it still uses the extended code pattern A/B of a branch to first locate a possible minimum-cost branch. Afterwards, a cost comparison table, such as Table 7.3 or 7.4, is used to check if further testing of the branch is required. Tables 7.3–7.7 in Appendix 7.C outline a list comparison tests of cost for all possible 4R and 8R branches. The key idea is that given two branches, one can determine which one has a smaller cost without actually measuring the exact cost of each one. Branches that require further tests are marked by # in the tables. The required tests for each branch are also marked by *. The tables are computed by searching for all possible branches with unique extended code patterns, and comparing the costs of these branches in each case. Note that the topology search method is very efficient since the tables can be computed off-line. Although we have only described the method for the cases of 4R and 8R codes, it can be easily generalized to include branches of higher code rate factors. Thus, our topology search method is not only optimal, but also computationally efficient for moderately high code rate factors. The complexity of a DCA scheme depends mainly on the type of search algorithm used to locate a minimum-cost branch. Once a branch is selected, the algorithm must also specify a list of codes that must be reassigned along with their respective branch numbers. Several options are available to a system designer.

(1) If fast computational time is a premium, the code pattern search algorithm can be used to locate a minimum-cost branch. The search involves off-line table look-up only. Since this algorithm is generally not optimal in locating a minimum-cost branch, the number of code reassignments may be higher.

(2) If computational time is not critical, optimal topology search algorithm can be used to minimize the number of reassignments. Compared to code pattern search, this algorithm uses extra comparison tests to identify a minimum-cost branch.

(3) Regardless of the search algorithm used for a minimum-cost branch, it may take several time slots to reassign descendant codes error-free (by using receiver acknowledgments). However, fast connection time for a new call is still possible. This is achieved by assigning a lower rate vacant code in the minimal-cost branch to the new call. Once all occupied codes in the branches have been reassigned, the rate indicator (layer number) is used to adjust the data rate.

### 7.3.3 Simulation Model and Results

In the preceding two sections, we outlined various schemes of implementing a spectral-efficient and cost-effective OVSF-CDM system. The complexity and the task of reassigning OVSF codes is located on the central controller's (base station's) side. By simulation, we first analyze the loss in spectral efficiency of an OVSF-CDM system that does not apply DCA. Second, we measure the total number of codes that must be reassigned if DCA is activated. The flow chart of the simulation program is shown in Fig. 7.12. The simulation uses the following parameters:

- Call arrival process is Poisson with mean arrival rate $\lambda = 1 \sim 16$ calls/unit time (Fig. 7.13), $4 \sim 64$ calls/unit time (Fig. 7.15).
- Call duration is exponentially distributed with a mean value of $1/\mu = 0.25$ units of time.
- Maximum spreading factor $N_{max} = 64$ (Fig. 7.13), 256 (Fig. 7.15).
- Possible OVSF code rates: R, 2R, 4R and 8R.
- Capacity test: code-limited.
- DCA scheme: optimal and sub-optimal search algorithms.

An outline of the simulation program is as follows:

1) Input parameters such as call arrival rate, duration, code rate distributions are entered.
2) For each new call, the capacity check (7.2) is performed. If the call cannot be supported, the call is dropped. Otherwise, one of the reassignment subroutines is executed.
3) If the call rate is 8R, the 8R subroutine assigns a code to the new call. If descendant codes need to be reassigned, it is directed to other reassignment subroutines.
4) If the call rate is 4R, 2R or R, the execution is similar to Step 3.
5) Once the reassignment is complete, the last step is to record the number of blocked calls and the number of reassigned codes.
6) The process is repeated by returning to Step 2.

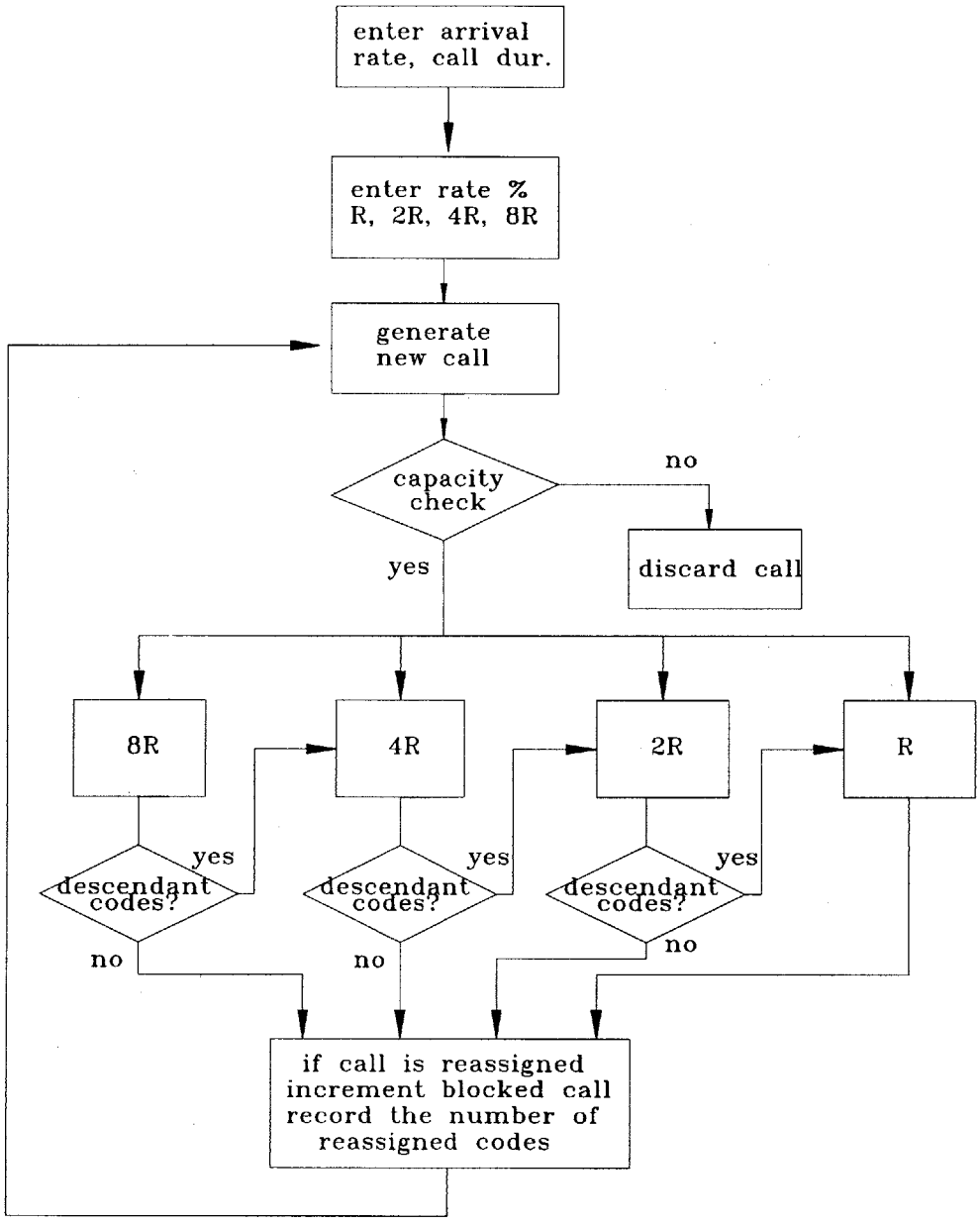



**Figure 7.12:** Flow chart of Proposed DCA algorithm
The algorithm can be extended to include reassignments of 16R codes and beyond. Note that reassignments occur only after a new call is initiated. As an alternative, reassignments can occur after a call terminates.

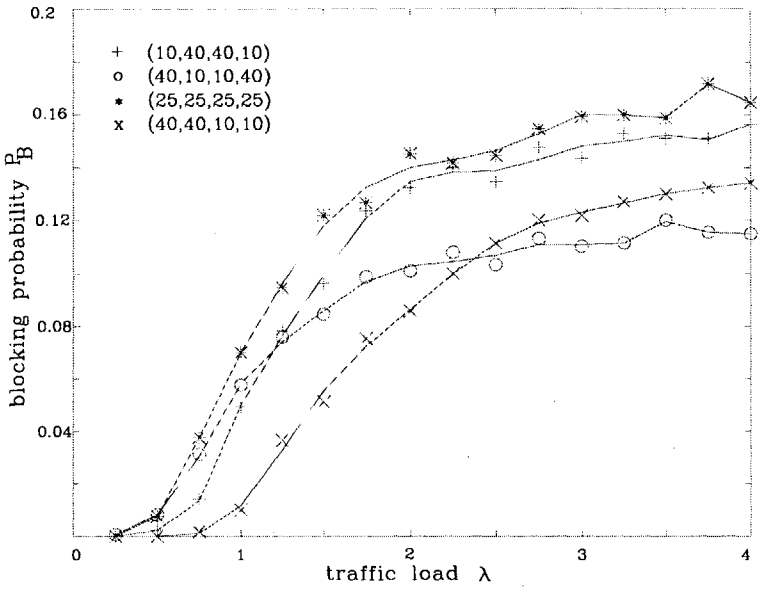



**Figure** 7.13: Plot of blocking probability vs. traffic load for $N_{max} = 64$.

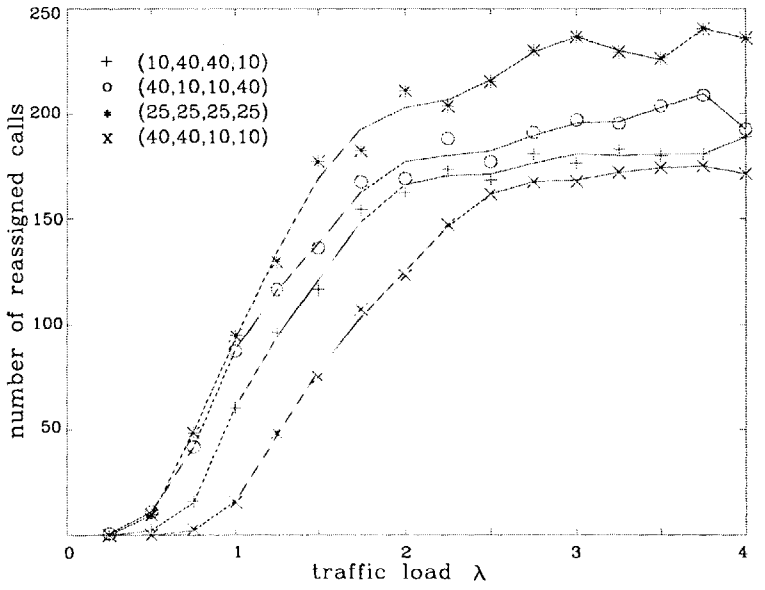



**Figure** 7.14: Plot of number of reassigned codes vs. traffic load for $N_{max} = 64$

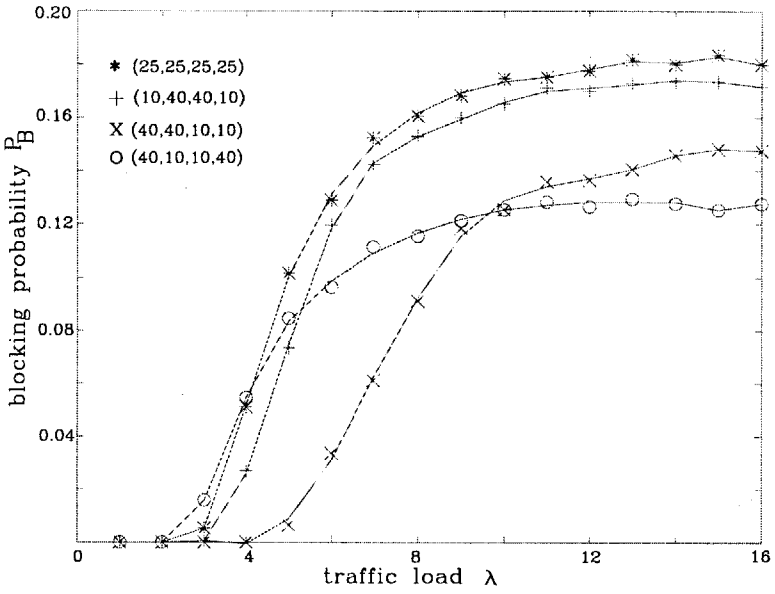



**Figure** 7.15: Plot of blocking probability vs. traffic load for $N_{max} = 256$.

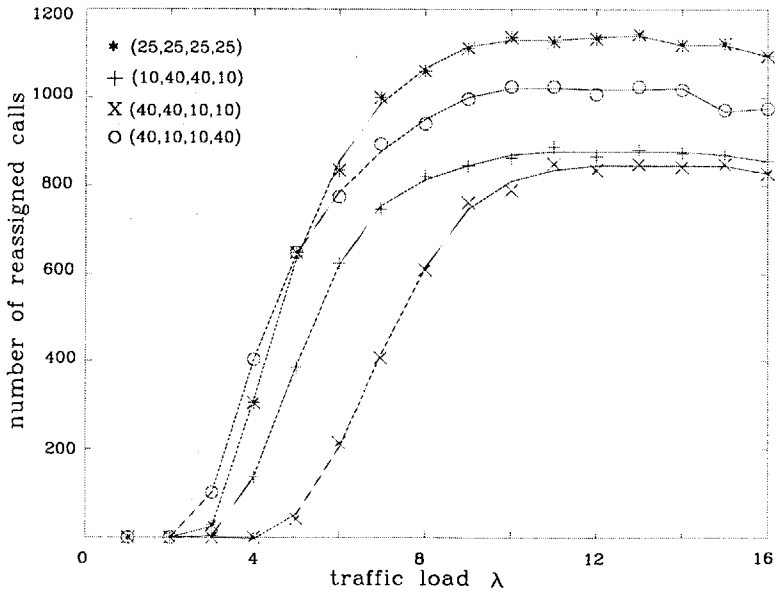



**Figure** 7.16: Plot of number of reassigned codes vs. traffic load for $N_{max} = 256$

Figures 7.13 and 7.15 show plots of code blocking probability, i.e. probability of calls that would be blocked if DCA is not implemented, versus traffic load $\lambda/\mu$ (average service rate average call duration). We have chosen $N_{max}$ = 64 and 256 to reflect the values chosen for IS-95 and W-CDMA standards [114, 28]. The total number of accepted calls (calls that satisfy the capacity constraint) is 1000 ($N_{max}$ = 64) and 4000 ($N_{max}$ = 256) for each traffic load simulated. The number of codes that must be reassigned to support the accepted calls are shown in Figs. 7.14 and 7.16. Each point on a curve in the figures represents the average value over 10 simulations. The simulation shows that both the blocking probability and the number of code reassignments are the largest when the code rates are uniformly distributed. The (R, 2R, 4R, 8R) = (10, 40, 40, 10) distribution has the second largest blocking rate. Its number of code reassignments is, however, less than that of (40, 10, 10, 40) distribution. These results are expected because in (10, 40, 40, 10) distribution, calls of rate 2R and 4R have a higher probability of satisfying the capacity check of (7.2), and also getting blocked without reassignments. The number of code reassignments in (10, 40, 40, 10) is less than (40, 10, 10, 40) because, in general, it costs more to reassign a higher-layer code (in this case, 8R).

By using the blocking probabilities and the code rate distribution, we can calculate the loss in spectral efficiency of an OVSF-CDM system if codes are not reassigned. For example, if the blocking probability is $P_B$ (the percentage of calls blocked due to code blocking) for a certain traffic load, the loss in spectral efficiency $\gamma$ is:

$$\begin{aligned}\gamma &= P_B \frac{\text{loss in capacity due to code blocking}}{\text{capacity without code blocking}} \\ &= P_B \frac{2S_2 + 4S_4 + 8S_8}{P_1 + 2P_2 + 4P_4 + 8P_8}\end{aligned} \tag{7.4}$$

where $P_i$ is the percentage of calls of rate iR over the duration of simulation. Thus,

$$P_1 + P_2 + P_4 + P_8 = 1$$

$S_i$ is the percentage of rate iR among calls that are subject to code blocking:

$$S_i = \frac{P_i}{P_2 + P_4 + P_8}$$

As an example, for a rate distribution of (25, 25, 25, 25):

$$\gamma = P_B \frac{2 \times \frac{1}{3} + 4 \times \frac{1}{3} + 8 \times \frac{1}{3}}{\frac{1}{4}(1 + 2 + 4 + 8)}$$

The plots in Figs. 7.13–7.16 were generated using the sub-optimal code pattern search (CPS) algorithm. Table 7.1 lists the loss in spectral efficiency without DCA for several different code rate distributions. The fourth column of the table shows the average number of code reassignment per accepted call if DCA is implemented using CPS algorithm.

Table 7.1: Loss in spectral efficiency due to OVSF code blocking (Code Pattern Search Algorithm)

| $N_{max}$ | traffic load | (R, 2R, 4R, 8R) | spectral efficiency loss | reassigned codes per call |
|---|---|---|---|---|
| 64 | 4 | (10,40,40,10) | 0.17 | 0.19 |
| 64 | 4 | (40,10,10,40) | 0.18 | 0.21 |
| 64 | 4 | (25,25,25,25) | 0.21 | 0.24 |
| 64 | 4 | (40,40,10,10) | 0.19 | 0.17 |
| 256 | 16 | (10,40,40,10) | 0.19 | 0.22 |
| 256 | 16 | (40,10,10,40) | 0.19 | 0.26 |
| 256 | 16 | (25,25,25,25) | 0.22 | 0.39 |
| 256 | 16 | (40,40,10,10) | 0.21 | 0.21 |

Similar results generated by using the optimal code pattern plus topology search algorithm are tabulated in Table 7.2. As expected, the number of reassignments is less with the optimal scheme —however, little difference, if any, is observed in spectral efficiency loss.

Table 7.2: Loss in spectral efficiency due to OVSF code blocking (Code Pattern plus Topology Search Algorithm)

| $N_{max}$ | traffic load | (R, 2R, 4R, 8R) | spectral efficiency loss | reassigned codes per call |
|---|---|---|---|---|
| 64 | 4 | (10,40,40,10) | 0.18 | 0.10 |
| 64 | 4 | (40,10,10,40) | 0.18 | 0.14 |
| 64 | 4 | (25,25,25,25) | 0.20 | 0.16 |
| 64 | 4 | (40,40,10,10) | 0.20 | 0.12 |
| 256 | 16 | (10,40,40,10) | 0.19 | 0.15 |
| 256 | 16 | (40,10,10,40) | 0.19 | 0.18 |
| 256 | 16 | (25,25,25,25) | 0.21 | 0.27 |
| 256 | 16 | (40,40,10,10) | 0.21 | 0.13 |

## 7.4 DCA Algorithm for Bursty-Rate Channels

By reassigning OVSF codes horizontally, the data rate of each code channel is unchanged. This method is useful in applications where constant bit-rate connections are required. It also applies to delay-sensitive connections. It, however, has several drawbacks.

- First, the number of simultaneous connections that horizontal DCA supports is hard-

limited —also known as *dimension-limited* since the system can support a maximum of $N_{max}$ connections, which is equal to the number of leaves in an OVSF binary code tree.

- Second, for each horizontal reassignment, the base transmitter must send control information to a designated receiver to request a switch to its newly assigned code, which is a cousin Walsh code. This process requires transmission of control signalling as overhead and coordination between transmitter-receiver pairs. Furthermore, reassignments are required whenever the system experiences code blocking.

- Third, it cannot support statistical multiplexing since a connection does not release its assigned code even if it is idle.

These three issues can be resolved as follows:

- *Virtual Codes*: Hard limitation of parallel orthogonal codes is addressed by assigning *virtual* orthogonal codes. By assigning virtual codes, the total number of forward link channels is no longer dependent on maximum spreading gain $N_{max}$. The specifics of this topic are covered in Sec. 7.6.2.

- *Vertical Reassignment*: In vertical reassignment, the length (dimension) of a Walsh code —as opposed to a newly assigned cousin Walsh code— must be specified in accordance with information-bearing data rate. The frequency of reassignment does not depend on incoming/outgoing call traffic statistic, but is rather a design parameter. As a result, rate adaptation is possible at a finer time frame level.

- *Statistical Multiplexing*: Statistical multiplexing is afforded by combining vertical reassignments with time sharing. Information packets from each source are queued in parallel buffers (see Fig. 7.17). In each time frame, packets at the head of the queue from selected channels are multiplexed and transmitted according to some quality-of-service (QoS) control mechanism.

It is obvious that this procedure is suitable only for information content that is delay-tolerant. Since our goal is information transfer via parallel OVSF-CDM code channels, the chosen QoS criterion is bounded latency (i.e., finite buffer size) in all parallel channels and maximization of total network capacity. We propose a heuristic algorithm based on the concept of credit-reward system. An outline of this algorithm is as follows:

(1) Each user $i$ is assigned a leaf code. Let's denote its requested rate by $k_iR$.

(2) Each user $i$ is tracked by its "credit" and "usage" denoted by $C_i$ and $U_i$, respectively. Initially, $C_i = 0$ for all users. In the beginning of every time frame, $C_i$ is increment by $k_i$. In one time frame, if transmission of information is possible at its requested *average* rate,

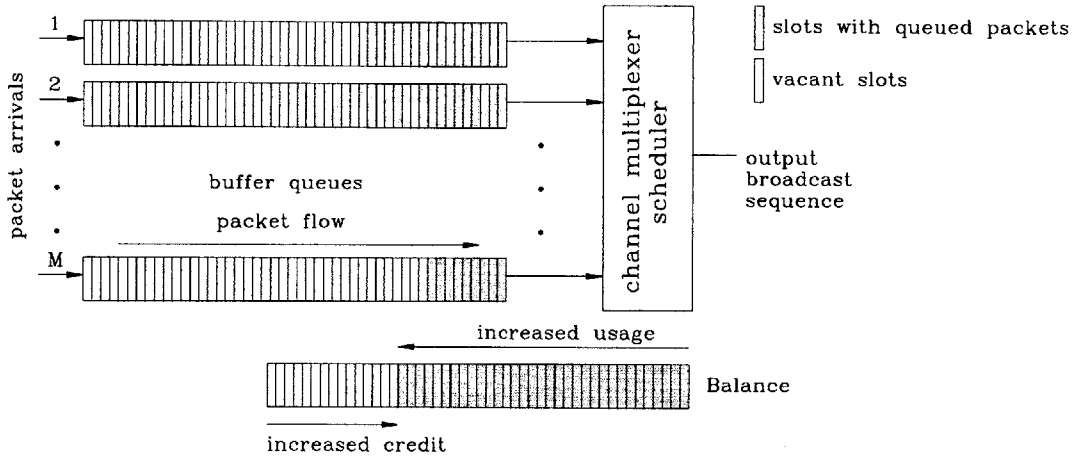



**Figure** 7.17: Buffer queues in parallel OVSF-CDM channels

its usage $\mathcal{U}_i$ is said to be k units. At the end of every time frame, $\mathcal{C}_i$ is reduced by $\mathcal{U}_i$—the amount of information transferred per time frame.

(3) After one time frame, the balance $\mathcal{B}_i$ of user i is:

$$\mathcal{B}_i = \mathcal{C}_i - \mathcal{U}_i$$

If the balance $\mathcal{B}_i$ reaches the "overdraft" limit $B_{min}$ at the beginning of each time frame, no information is transferred for user i. On the other hand, in order not to favor a user that has been idle for a long period, we also put a "credit limit" $C_{max}$ on the maximum amount of credit that each user can accumulate. Thus, the maximum value of $\mathcal{B}_i$ is $C_{max}$.

The value of overdraft limit $B_{min}$ must be chosen carefully to maximize the system capacity. The credit limit $C_{max}$ protects users against unbounded delays. In most cases, these values are chosen on a trial-and-error basis by simulating various traffic models with different rate distributions. In this short review we have not addressed the role of call admission for new connections. Note that the balance $\mathcal{B}_i$ is a measure of buffer size. We wish to control the call admission policy for new connections —thus increasing the network capacity— while choosing efficient values for $C_{max}$ and $C_{max}$ such that the balance $\mathcal{B}_i$ is bounded by a some pre-set finite value. A detailed discussion of DCA using vertical reassignment is given in Reference [64]. It is shown therein that all requested rates are met *on average* by vertical DCA. The transfer of information is, however, very choppy—with the leaf codes suffering the worst. Since choppiness is equivalent to increased latency, it is not a desirable outcome for certain delay-sensitive applications. The reference also covers the merits and efficiency of vertical DCA.

## 7.5 DCA Algorithm for Mixed-Rate Channels

One disadvantage of SC-RG CDM compared to PC-SG CDM is its limitation of information rates to certain quantized levels, $kR$ bps where $k = 2^n$, $n \in \mathbb{N}$. It is shown in [28] that information rates between the quantized levels can be achieved by varying the rate from one time slot to another. Of course, we already know that any arbitrary rate can be supported by vertical reassignments with time sharing. Our interest here is the support of arbitrary rates to receivers in combination with dynamic horizontal reassignments. An important distinction between the combined DCA and vertical DCA schemes is that in the latter, transmission of information at a non-zero minimum rate is supported. In combined DCA, each channel has a maximum and a minimum ($\geq 0$) information rate provisioning, with the average rate located somewhere in between. Equivalently the channel is less choppy since the peak-to-average rate ratio is reduced considerably. It appears that there is much to gain by combining horizontal and vertical reassignment strategies in any resource allocation policy that removes code blocking. However, we have hinted that the design of an efficient algorithm that incorporates combined OVSF code reassignments is not an easy matter. The difficulty arises from conflicting optimization strategies in the algorithms that use either the horizontal or vertical code reassignments, but not both. In horizontal DCA, the goal is to pack assigned codes tightly as possible by reassignments. This is the overpopulation strategy. In contrast, vertical DCA prefers codes that are spread out such that large duty cycles are maintained for all active channels. This is the underpopulated case. We next describe reassignment strategies of our combined DCA:

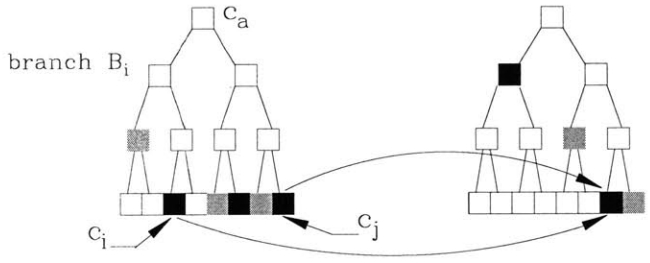



**Figure** 7.18: Reassignment of leaf codes to another branch

- *Horizontal Reassignments for Reduced Choppiness*: In vertical DCA, each user $i$ is classified by its maximum $R_i(\max)$ and minimum $R_i(\min)$ rate requirements. Initially, user $i$ is assigned a leaf code, say $c_i$. The spreading gain of the assigned OVSF code $c_i$ may be reduced such that its ancestor code, say $c_a$ supports a rate $R_{c_a} \geq R_i(\max)$. Let $c_a$ be the root code of a branch, say $B_i$. Note that any of the leaf codes not equal to $c_i$ in branch $B_i$ can be used in the initial assignment phase of user $i$, without affecting the maximum supported rate. If there exists another user, say $j$ that is assigned another leaf code $c_j$ in

branch $B_i$, its impact on user i —and vice versa— is the likelihood of increased choppiness in their data rates. This choppiness can only be removed by reassigning horizontally a leaf code to either user i or j, with the requirement that the new leaf code belong to another branch. This situation is illustrated in Fig. 7.18.

- *Horizontal Reassignments for Minimum-rate Guarantee*: In order to meet the minimum rate requirement of user i, an OVSF code $c_m$ whose rate $R_{c_m} \geq R_i(\text{min})$ must be assigned. Initially user i is assigned a leaf code $c_i$, and later reassigned to an ancestor code, preferably $c_m$. Let us assume $B_m$ is a branch whose root code is $c_m$. For this step we must reassign horizontally all users whose assigned codes are the leaf codes of branch $B_m$. Since we know a priori the minimum rate $R_i(\text{min})$ of user i before establishing a connection, we can perform a minimum-cost search procedure (i.e., the optimal horizontal DCA) for a code whose rate is greater or equal to $R_i(\text{min})$.

- *Modification of Horizontal DCA Algorithm*: For a user with a bursty connection, its assigned root code (whose rate $\geq R_i(\text{max})$) in a branch can also be shared with other users. For constant rate connections with 100% duty cycle, no code sharing is allowed. Therefore, when searching for a minimum-cost branch, all branches whose root code is assigned to bursty users no longer qualify as candidates. Similarly, when a new bursty user demands a channel with its requested $\{R_i(\text{max}), R_i(\text{min})\}$, the best candidate branches are those whose root codes are shared by other bursty users. To meet its demand, it may be necessary to carry out horizontal reassignments as detailed in the previous two steps.

To summarize, four different types of horizontal reassignments exist in combined DCA:

1. Leaf codes of bursty connections in a branch are reassigned to reduce rate choppiness.
2. Codes for any type of connection —either bursty or fixed rate— are reassigned to meet the minimum rate request of a bursty connection.
3. Codes of constant-rate connections are reassigned to meet the maximum rate request of a bursty connection.
4. Codes of constant-rate connections are reassigned to support a new connection.

We emphasize that our proposed scheme is equivalent to *dynamic partitioning* of branches such that bursty connections share codes within a *branch*. For a fixed rate connection, it is assigned the root code of a branch for the entirety of its connection. It is clear that our proposed combined DCA scheme is more efficient that *static partitioning* of the entire OVSF tree into two sub-trees, where each sub-tree is for bursty or fixed rate connections only. We, however, do not include a viable admission policy for both bursty and fixed-rate connections since this is a separate issue from the code scheduling and reassignment policies. In practice, it is likely that the

cellular network operator will offer several classes of QoS for bursty and fixed rate connections with different price ranges.

## 7.6 System Requirements for Code Reassignment

In terms of hardware complexity single channel-VG-CDM is preferred over parallel channel SG-CDM since the former requires a single decoder at the mobile terminal regardless of the information rate. In the ITU's 3G wideband CDMA standard, each traffic channel in the forward (base-to-mobile) link is time-slotted. Both control information and user data bits are time-multiplexed in each time slot. See Fig. 7.19 for an illustration. In SC-RG CDM, a user's data rate can be varied every time slot by adjusting the rate field (indicating the transmitted data rate) that is embedded in the control information. However, the standard does not specifically address the procedure of resource allocation, i.e., OVSF code assignment for rate adaptation and packet switching on a time-slot basis.

In this section, we propose the necessary protocols for the implementation of a code scheduling scheme that dynamically assigns OVSF codes to mobile users on a time-slot basis such that the total throughput of the system is maximized while an average data rate guarantee to provided to each mobile user. By using a dynamic traffic scheduler, our protocol supports per-connection guaranteed rates, while allowing high degrees of burstiness with small backlogs. This is achieved without the need for a mobile user to overbook its required rate, thereby maximizing the system throughput. Since the required control signalling to implement our scheme is based on the 3G W-CDMA standard, our results are directly applicable to systems using this standard. Even when OVSF codes are optimally assigned, the system capacity may be under-utilized because the maximum number of OVSF codes is hard-limited. This is true when at least one channel has a low duty-cycle; i.e., it remains inactive most of the time. For such a case, we propose a new protocol —which is not included in the 3G W-CDMA standard— that soft-limits the number of active users by permitting many users to share a common OVSF code.

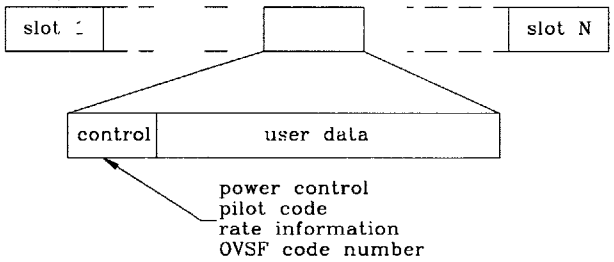



**Figure 7.19:** Time-division multiplexing of control and data channels within a time-slot. This is the chosen format in ITU's 3G W-CDMA standard.

### 7.6.1 Control Signalling for Horizontal Reassignments

In order to inform the user of a newly assigned OVSF code, some type of control channel signaling is necessary. Two possible options are in-band and out-of-band control channel signaling. An in-band control channel signaling mode (see Fig. 7.19) has already been proposed for third-generation CDMA systems [28], where the pilot signal, transmit power control, rate information and other control signals are time-multiplexed with the user data. (From Fig. 7.1 we know that the layer number conveys the rate information. The branch and layer numbers are sufficient parameters to identify a unique OVSF code.) We can modify the W-CDMA header such that not only the rate information but also the branch number can be included in the header of a time slot. By time slotting, the new protocol supports horizontal reassignment of OVSF codes on a dynamic basis.

In order for the Walsh generator at the receiving end to switch from one code to another, the encoding of data with the newly assigned code can be delayed for a number of time slots. Fig. 7.20 illustrates an example where the branch number of a reassigned OVSF code is embedded in slot 1. The data mapping using this code, however, does not begin until slot k. Other demodulation and detection functions such as carrier recovery, code acquisition, symbol time synchronization and pseudo-noise (PN) code despreading need not be altered. Power control signaling also remains the same since another code from the same layer is reassigned.

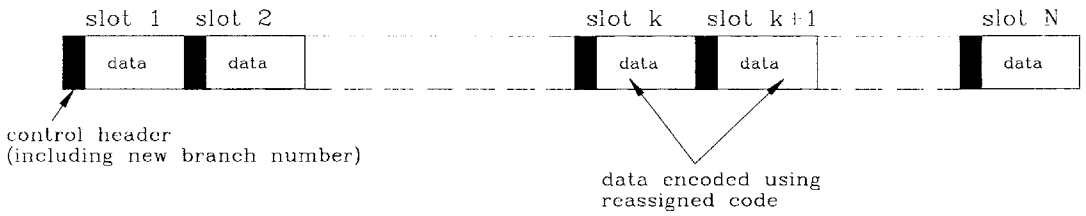



**Figure** 7.20: Delay of encoded data for Walsh code switching

Ultimately, our goal is to modify the control channel protocol without increasing the number of bits in the control header for relaying code reassignment information to the receiver. Expanding the portion of the control header –while keeping the length of a time slot fixed– reduces data throughput. One option is to use the rate information field to relay both the layer number (rate) and branch number (code), but not at the same time. In IMT-2000 W-CDMA standard [28], seven layers —corresponding to spreading factors 256, 128, 64, 32, 16, 8 and 4— are specified . Therefore, three bits (before error control coding and spreading) —say from 1 to 7— are sufficient to indicate the data rate or layer number. The unused three-bit pattern "000" can be used as a flag to signal the mobile that the control header in future time slots contains (reassigned) code information. Since there are at most 256 branch numbers in IMT-2000

systems, 8 bits (which can be transmitted over three time slots) are necessary to relay the code assignment information. An alternative to in-band control signaling is the use of cell-specific common control signalling channels such as paging and broadcast channels. It is foreseen in the future that mobile handsets will be capable of parallel channel (multi-code) reception. If the handset is equipped with at least two baseband receiving units, one can be used to decode information data while the other monitors control signals such as code reassignment. Regardless of the scheme selected for control channel signaling, in-band or out-of-band, an acknowledgment mechanism from mobiles can be included to insure that every receiver is tuned to its reassigned orthogonal code.

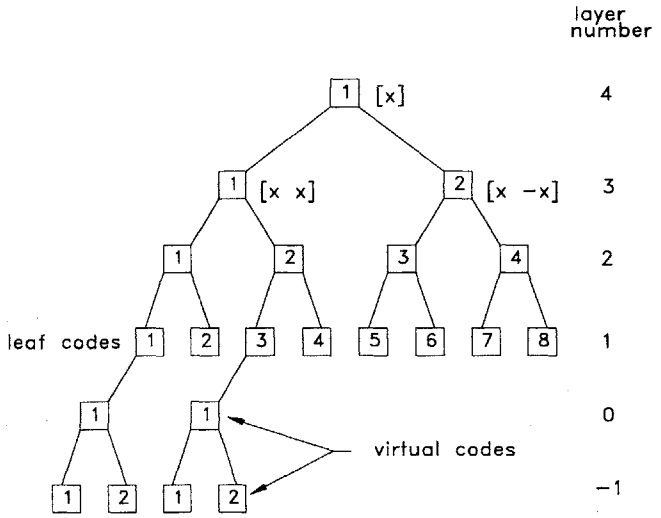



**Figure** 7.21: Generation of virtual codes for bursty connections

### 7.6.2 Control Signalling for Vertical Reassignments

We stated earlier that the W-CDMA standard already supports vertical reassignments by inclusion of a rate field on the control header. This protocol implicitly assumes that each user is assigned at least a leaf code, guaranteeing $R_i(\min) = R$. We however can foresee applications where $0 \leq R_i(\min) \leq R$. In such situations, the system capacity will be under-utilized since there is no statistical multiplexing of code rates. In essence, we wish to assign *virtual codes* to each bursty user such that the number of *active* —but not necessarily receiving— users is much larger than the number of leaf codes. This is illustrated in Fig. 7.21. Conceptually, this protocol is equivalent to TDM where the number of receivers is potentially much larger than the number of time-slots in a single frame. To implement such a protocol, the control header of the W-CDMA time-slot must be modified to include the *identification* of a virtual code.

## Notes and References

Most of the material come from two sources [86],[64]. Although both algorithms can be used in single- and multi-carrier based CDM systems, these references focus only on the former. This is because no such multi-carrier CDM system is operational to date. In CDM two important issues in networking are the support of variable data rates and bounded latency. In [86] a horizontal code reassignment scheme is proposed for constant-rate, delay sensitive code channels. Since horizontal reassignment of an OVSF code does not change its supported data rate, both rate and delay requirements are satisfied. The only drawback of this scheme is that a code channel is not released until a call is completed. For most non-voice applications, a user is "on" only for a short burst, and remains idle for a long duration. For a user to be "on" at all times, it must be assigned a *virtual* code channel. For such a connection, better use of system capacity is possible by statistical multiplexing. This is the underlying concept of the algorithm proposed in [64]. Here, the code reassignments are vertical. The algorithm is well-suited for bursty and delay-tolerant connections. Since a code channel can be time-shared by many users (though not simultaneously), the system no longer suffers from a "code or dimension limit." Its main drawback is data choppiness, i.e. a large value in the ratio of its peak to average data rate. Both algorithms can be used in ITU W-CDMA standards [28] with minor or little modification in control channel signalling. Finally, by combining horizontal and vertical reassignment schemes, we can eliminate several drawbacks that exist in each scheme. The combined code reassignment scheme presented in this chapter is new; it covered in [86, 64] and has yet to appear in any public domain archives.

An important topic not addressed in this chapter is variable rate support by combining OVSF-CDM with multi-code CDM. It is foreseen in the future that mobile handsets will have the capability of simultaneously decoding two or more parallel CDM channels. Since each parallel channel can carry increased data rate by reducing its spreading gain, there is greater flexibility —from a transmitter's point of view— in offering variable rates when OVSF- and multi-code CDM schemes are combined. Furthermore, the reassignment algorithms —for the removal of code blocking— are somewhat simplified because reassignments of higher layer codes are less probable; i.e., a high rate channel request can be split into two or more low rate channel requests via multi-code parallel encoding and decoding. It is likely that with combined transmission schemes, the instantaneous transmission rates will be less choppy when using vertical code reassignment algorithm.

# Appendix 7A

## Proof of Theorem 7.2

It suffices to show that given any procedure **A** of reassigning a code C to a branch $\mathcal{K}$ with an occupied root code D of a higher rate than C, we can derive another procedure **B** of reassigning C to another branch with less cost. In procedure **A**, D needs to be reassigned to the root code of another branch $\mathcal{Q}$. There is an empty sub-branch $\mathcal{K}'$ of $\mathcal{K}$ with its unoccupied root code being an immediate descendant of D. In $\mathcal{Q}$, consider its left sub-branch $\mathcal{Q}_1$ and right sub-branch $\mathcal{Q}_2$. Since D is reassigned to the root code of $\mathcal{Q}$, the occupied codes of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ need to be reassigned to other branches. Let $\mathcal{Q}_1'$ be the subset of codes in $\mathcal{Q}_1$ that are reassigned to $\mathcal{K}'$ and $\mathcal{Q}_1''$ be the subset of codes in $\mathcal{Q}_1$ that are reassigned to other branches. Similarly, let $\mathcal{Q}_2'$ be the subset of codes in $\mathcal{Q}_2$ that are reassigned to $\mathcal{K}'$ and $\mathcal{Q}_2''$ be the subset of codes in $\mathcal{Q}_2$ that are reassigned to other branches. Let $\mathcal{Q}_3'$ be the set of codes other than $\mathcal{Q}_1'$ and $\mathcal{Q}_2'$ that are reassigned to $\mathcal{K}'$. Let $\mathcal{Q}_3''$ be all other reassigned codes in procedure **A**. The cost of procedure **A** is therefore

$$2 + |\mathcal{Q}_1'| + |\mathcal{Q}_1''| + |\mathcal{Q}_2'| + |\mathcal{Q}_2''| + |\mathcal{Q}_3'| + |\mathcal{Q}_3''|$$

where the cardinality $|\mathcal{Q}_x'|$ denotes the number of codes in $\mathcal{Q}_x'$. (The first term above is due to the reassignment of C and D alone.) We now derive a procedure **B** that reassigns code C to branch $\mathcal{Q}_2$ with cost 1 less than the above, but without reassigning D. In procedure **B**, the codes in $\mathcal{Q}_1'$, $\mathcal{Q}_2'$, and $\mathcal{Q}_3'$ are all reassigned to branch $\mathcal{Q}_1$. This is possible because $\mathcal{Q}_1$ has the same capacity as $\mathcal{K}'$. The codes in $\mathcal{Q}_1''$, $\mathcal{Q}_2''$, and $\mathcal{Q}_3''$ are reassigned to other branches as in procedure **A**. The code C is reassigned to a code in $\mathcal{Q}_2$. The cost of procedure **B** is therefore at most

$$1 + |\mathcal{Q}_1'| + |\mathcal{Q}_1''| + |\mathcal{Q}_2'| + |\mathcal{Q}_2''| + |\mathcal{Q}_3'| + |\mathcal{Q}_3''| \qquad \square$$

# Appendix 7B
# Proof of Theorem 7.3

Before we prove theorem 3, we state the following simple fact:

**Lemma 7.1** *Let C(T) be the minimum cost of a branch in set T. Then*

$$C(T') \leq C(T) \qquad \textit{for any} \quad T' \supseteq T.$$

The proof is trivial since by adding more branches, the reassignment of codes into these branches cannot increase the cost of reassigning codes into existing branches.
Let the two occupied codes (of the same branch) that must be reassigned be $c_1$ and $c_2$, and $R_2 < R_1$, where $R_i$ is the data rate of code $c_i$.

The theorem uses the following idea: Given a procedure of reassigning $c_2$ before $c_1$, we can always derive another procedure of reassigning $c_1$ before $c_2$ with equal or less cost.

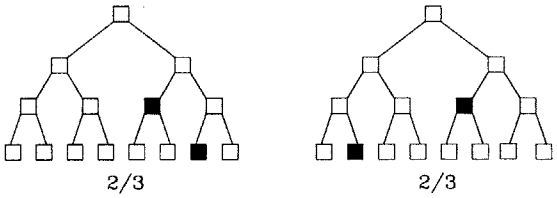
- *Case 1*: We first consider the case where
  (a) $c_2$ is reassigned first, say to code $c_k$, and
  (b) during the reassignment of $c_1$, $c_2$ is not reassigned.
  In this case, $c_1$ cannot be reassigned to the mother code of $c_k$, say $c_m$, which has the same rate as $c_1$ because of code blocking by $c_2$. So $c_1$ needs to be reassigned to other branches excluding $c_m$; let $\mathcal{C}$ denote such cost of reassigning $c_1$. Now let us reassign $c_1$ first by reserving code $c_k$ for $c_2$ and do not consider code $c_k$ in the reassignment of $c_1$; let $\mathcal{C}'$ denote such cost of reassigning $c_1$. Observe that the branch with $c_k$ as the root code is a subset of the branch with $c_m$ as the root code. Thus by Lemma 7.1, $\mathcal{C}' \leq \mathcal{C}$. In other words, we show that in this case reassigning $c_1$ first will yield a lower or equal cost.

- *Case 2*: We next consider the case where
  (a) $c_2$ is reassigned first, say to code $c_k$, and
  (b) during the reassignment of $c_1$, $c_2$ is reassigned to another code $c'_k$.
  Let $\mathbf{S}_0$ denote the state (i.e. locations of all occupied codes) of the code tree initially before $c_2$ is reassigned the first time. Let $\mathbf{S}_1$ denote the state of the code tree after $c_2$ is assigned to code $c_k$. During the process of reassigning $c_1$ from state $\mathbf{S}_1$, $c_2$ is reassigned from code $c_k$ to code $c'_k$. Now let us repeat the above code reassignment procedures from state $\mathbf{S}_0$ to state $\mathbf{S}_1$ except that $c_2$ is not reassigned to $c_k$. By keeping $c_k$ unoccupied, we repeat the same code reassignment procedures above to reassign $c_1$ from state $\mathbf{S}_1$. Afterwards, we assign $c_2$ to $c'_k$. Note that the code reassignment is identical for every code as in the above procedures except for $c_2$. It is reassigned from its original location in state $\mathbf{S}_0$ to $c'_k$ directly without being reassigned to $c_k$, reducing the cost by 1. Therefore, in this case, we also show that by reassigning $c_1$ first will yield a lower cost. □

# Appendix 7C

## Code Patterns

In selecting the optimal reassignment policy, we need to determine the branch with the minimum cost. As a first step, we denote every branch B by its *code pattern* $x/y$, where $x$ is the number of assigned codes and $y$ is the rate-sum of assigned codes in branch B. By definition, a leaf code has rate 1. Branch B with code pattern $x/y$ belongs to group $x$. In Appendix 7D we tabulate a comparative listing of the relative costs of branches. It is true that a code pattern is not a complete description of assigned codes in a branch. As shown below, two branches with




the same code pattern have different topologies. The left branch is over-populated, while the right is an example of an under-populated branch. The left branch can support 4R code without any reassignment. In contrast, the cost is 1 in the right branch to support a new connection of rate 4R.

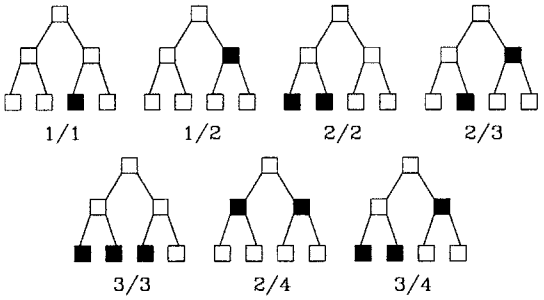



**Figure** 7.22: Code patterns for 4R branches

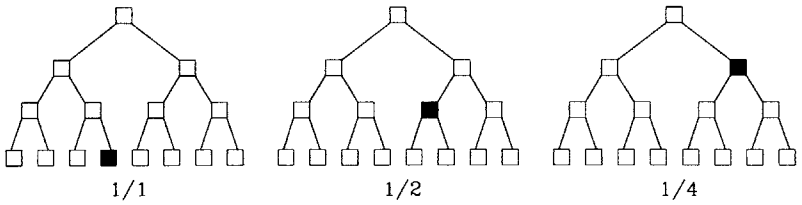



**Figure** 7.23: Code patterns for 8R branches in group 1

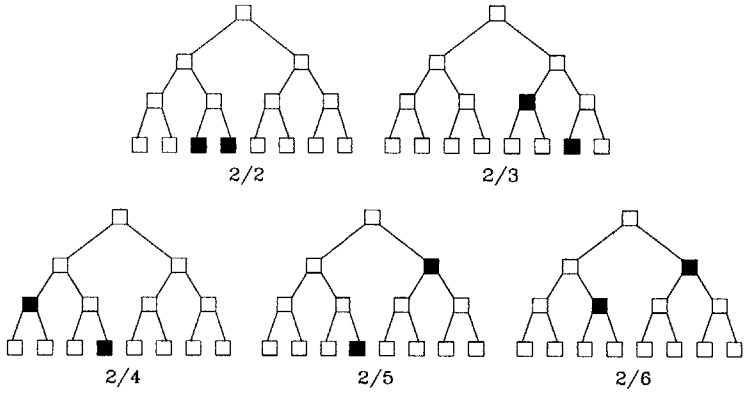



**Figure** 7.24: Code patterns for 8R branches in group 2

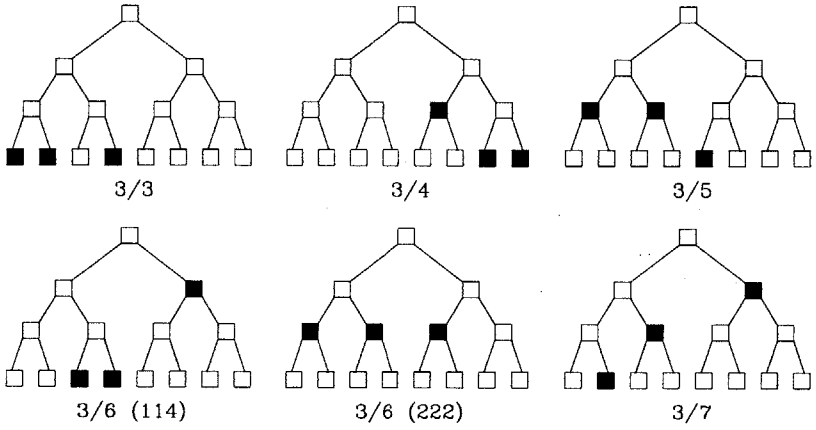



**Figure** 7.25: Code patterns for 8R branches in group 3

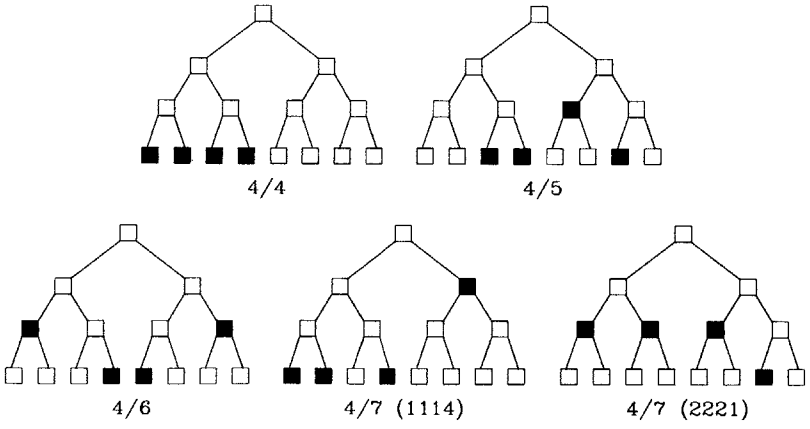



**Figure** 7.26: Code patterns for 8R branches in group 4

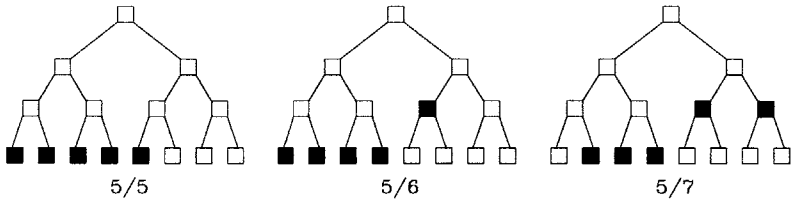



**Figure** 7.27: Code patterns for 8R branches in group 5

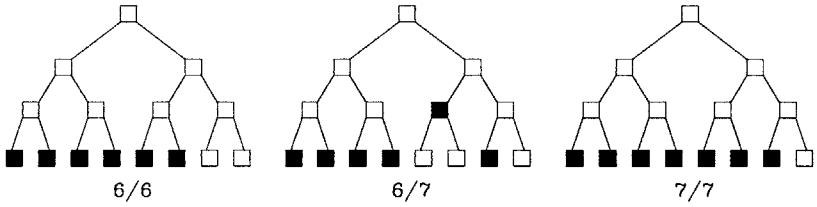



**Figure** 7.28: Code patterns for 8R branches in groups 6 and 7

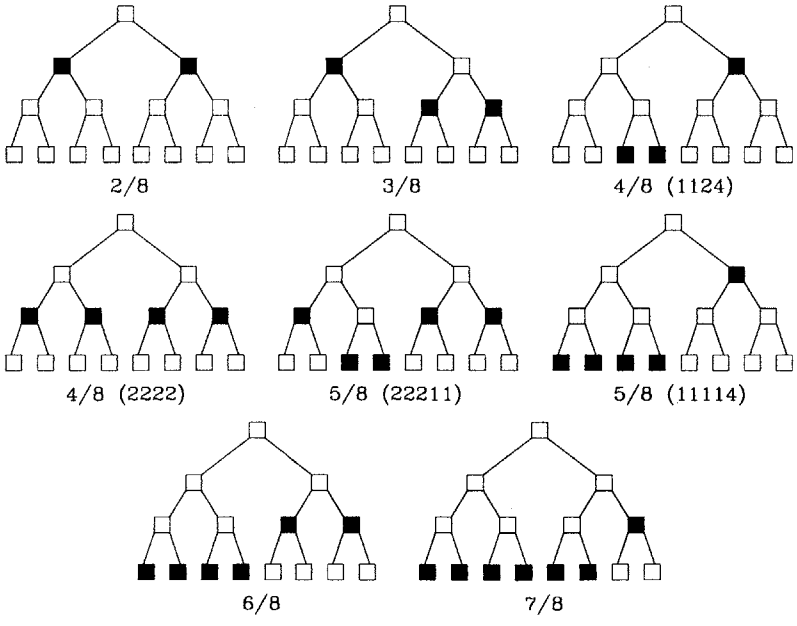



**Figure** 7.29: Code patterns for 8R branches in group 8

# Appendix 7D
# Cost Comparison Tables

In the following it is assumed that there is extra capacity to support a channel of designated rate. Since it takes one reassignment to move a code of rate R, its cost is set to 1. For a 2R branch, the cost is either 1 or 2, depending on the vacancy of another 2R branch in the remaining portion of the tree. For 4R branches, seven possible patterns exist as shown in Fig. 7.22. It is clear that 1/1 has the lowest cost (of 1) among them. Next is 1/2. The reasoning is as follows: If 1/2 is a candidate for reassignment, it implies 1/1 branches do not exist in the remaining part of the tree. If there exists 2/2 or another 1/2, then 1/2 has a cost of 1. (2/2 has a vacant 2R.) If no 1/2 or 2/2 exists, then there must be *at least* two 4R branches of 2/3 or 3/3, or one of each. (Otherwise, the system capacity is exceeded and a channel of 4R cannot be supported.) The cost of 2/2 is now 2 since the 2R code in its branch and an R code from 2/3 or 3/3 need reassignment. Note that $1/2 < 2/2$ because the cost of 2/2 is always 2. If the candidate branch is 2/3, then there must exist at least three 3/3 branches. (Both 2/4 and 3/4 have full loads; all fully loaded are excluded in cost comparison.) In this case, both 2/3 and 3/3 have a cost of 3. Since a tie is not allowed, we arbitrarily set $2/3 < 3/3$. Using this methodology, the cost comparison tables of higher-level branches such as 8R and 16R can be generated. The process is tedious but the reasoning is straightforward.

**Table** 7.3: Cost comparison of 4R branches

| | | |
|---|---|---|
| 1/1 | < | 1/2, 2/2, 2/3, 3/3, 2/4, 3/4 |
| 1/2 | < | 2/2, 2/3, 3/3, 2/4, 3/4 |
| 2/2 | < | 2/3, 3/3, 2/4, 3/4 |
| 2/3 | < | 3/3, 2/4, 3/4 |

**Table** 7.4: Cost comparison of 8R branches in groups 1 & 2

| | | |
|---|---|---|
| 1/1 | < | 1/2 |
| 1/2 | < | 1/4 |
| 1/4 | < | 2/2 |
| 2/2 | < | 2/3 |
| 2/3 | < | 2/4 |
| 2/4 | < | 2/5, 2/6, 3/3 |

**Table** 7.4 (contd.): Cost comparison of 8R branches in groups 1 & 2

| | | | |
|---|---|---|---|
| 2/5 | > | 3/4 | unless 3/4 is overpopulated with grid 211XXXX |
| 2/5 | < | 3/5 | |
| 2/5 | < | both 3/6's, 3/7 | |
| 2/6 | < | 3/3 | if there are vacant 4R and 2R<br>i.e., there exists branch 3/4 or 4/4 with vacant 4R |
| 2/6 | < | 3/4 | if there are vacant 4R and 2R<br>i.e., 4/4 with vacant 4R or 3/4 with 211XXXX |
| 2/6 | < | 3/5 | if 4R is vacant in another branch |
| 2/6 | < | 3/6(411) | |
| 2/6 | = | 3/6(222) | |
| 2/6 | < | 3/7 | |
| 2/6 | < | 4/4 | unless 4/4 is 1X1X1X1X,<br>and no vacant 2R in other branches |
| 2/6 | < | 4/5, 4/6, 4/7 | |
| 2/8 | < | all 3's | if there are 2 vacant 4R |

**Table** 7.5: Cost comparison of 8R branches in group 3

| | | | |
|---|---|---|---|
| 3/3 | < | all 3, 4, 5, 6, 7 | |
| 3/4 | < | 3/5 | |
| 3/4 | < | 3/6(411) | unless 3/6 grid is 41X1X and<br>there is no vacant 2R in other branches<br>AND 3/4 grid is 211XXXX |
| 3/4 | < | 3/6(222), 3/7, 4, 5, 6, 7 | |
| 3/5 | < | 3/6(411), 3/6(222), 3/7 | |
| 3/5 | < | 4/4 | unless 4/4 grid is 1X1X1X1X<br>and there are no other vacant 2R |
| 3/5 | < | 4/5 | unless 4/5 grid is 21X1X1X<br>and there are no other vacant 2R |
| 3/5 | < | 4/6, 4/7, 5, 6, 7 | |

**Table** 7.5 (contd.): Cost comparison of 8R branches in groups 3

| | | | |
|---|---|---|---|
| 3/6(411) | < | 3/6(222) | |
| 3/6(411) | < | 3/7 | |
| 3/6(411) | < | 4/4 | if 4/4 has a vacant 4R |
| 3/6(411) | < | 4/5 | |
| 3/6(411) | < | 4/6 | |
| 3/6(411) | < | 4/7(4111), 4/7(2221), 5/5 | |
| 3/6(411) | < | 5/6 | unless 3/6 grid is 411XX<br>and there are no other vacant 2R |
| 3/6(411) | < | 5/7, 6, 7 | |
| 3/6(222) | < | 3/7 | |
| 3/6(222) | < | 4/4 | if 4/4 has vacant 4R and 2R |
| 3/6(222) | < | 4/5 | |
| 3/6(222) | < | 4/6 | if there are 3 vacant 2R in other branches |
| 3/6(222) | < | 4/7(4111), 4/7(2221), 5/5 | |
| 3/6(222) | < | 5/6 | unless there are no vacant 2R |
| 3/6(222) | < | 5/7, 6, 7 | |
| 3/7 | > | 4/5, 4/6 | |
| 3/7 | < | 4/7(4111) | |
| 3/7 | = | 4/7(2221) | |
| 3/7 | < | 5/5 | unless 5/5 grid is 1X1X111X, and there<br>are less than 2 vacant 2R in other branches |
| 3/7 | < | 5/7, 6, 7 | |
| 3/8 | < | 4/4 | if there are 2 vacant 2R |
| 3/8 | < | 5/5 | if 5/5 grid is 11111XXX, and |
| | | | there are 2 vacant 2R in other branches |
| 3/8 | < | 5/6, 5/7, 6, 7 | if there are 2 vacant 2R |

**Table** 7.6: Cost comparison of 8R branches in group 4

| | | | |
|---|---|---|---|
| 4/4 | < | all 4, 5, 6, 7 | |
| 4/5 | < | 4/6, both 4/7's, 5, 6, 7 | |
| 4/6 | < | 4/7(4111), 4/7(2221) | |
| 4/6 | < | 5/5 | unless there is no vacant 2R |
| 4/6 | < | 5/6 | unless 5/6 grid is 2111X1X and AND 4/6 grid is 2211XX |
| 4/6 | < | 5/7, 6, 7 | |
| 4/7(2221) | < | 4/7(4111) | |
| 4/7(2221) | < | 5/5 | if there are 3 vacant 2R |
| 4/7(2221) | < | 5/6 | if there are 3 vacant 2R |
| 4/7(2221) | < | 5/7 | |
| 4/7(2221) | < | 6/6 | if there is at least one vacant 2R |
| 4/7(2221) | < | 6/7, 7/7 | |
| 4/7(4111) | > | 5/5 | |
| 4/7(4111) | > | 5/6 | |
| 4/7(4111) | = | 5/7 | |
| 4/7(4111) | < | 6/6 | if there are 2 vacant 2R in other branches |
| 4/8(4211) | < | all other patterns | |
| 4/8(2222) | < | 4/7(4111) | if there are 4 vacant 2R |

**Table** 7.7: Cost comparison of 8R branches in groups 5, 6 and 7

| | | | |
|---|---|---|---|
| 5/5 | < | all 5, 6, 7 | |
| 5/6 | < | 5/7, 6, 7 | |
| 5/7 | < | 6/6 | if there are 2 vacant 2R |
| 5/7 | < | 6/7, 7/7 | |
| 5/8(41111) | > | all other patterns | |
| 5/8(22211) | > | all other patterns | |
| 6/6 | < | 6/7, 7/7 | |
| 6/7 | < | 7/7 | |
| 6/8 | > | all other patterns | |
| 7/7 | < | 7/8 | |

# Appendix 7E

# Example of Horizontal Reassignments

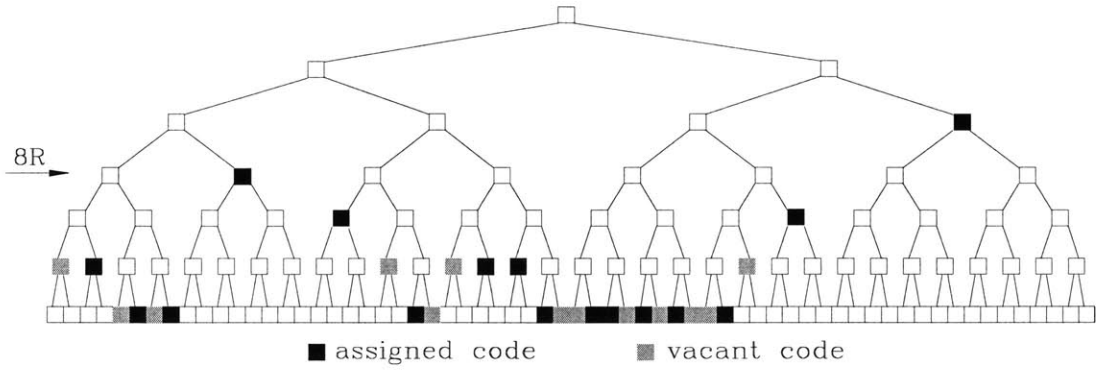



**Figure** 7.30: CDM code tree with 64b leaves

The code tree consists of 64 leaves. Each leaf code has rate R. The target rate is 8R. As shown in Fig. 7.30, the assigned and vacant codes are shaded black and gray, respectively. The total available capacity is found to be 17R by summing the rates of vacant (gray) codes. The following notation is used to label branches and codes: $B_i(j)$ and $C_i(j)$ are the branch and code of rate $i$ with index $j$ (from left to right, with leftmost set to 1) respectively. Among the target branches of 8R, $B_8(2)$, $B_8(7)$ and $B_8(8)$ are not candidates since they are fully loaded. $B_8(2)$ is an assigned code while $B_8(7)$ and $B_8(8)$ are children of an assigned code $B_16(4)$. After removing branches that are not in contention, the new topology of the tree is shown in Fig. 7.31. From Figs. 7.24 and 7.25, the code patterns of remaining 8R branches are:

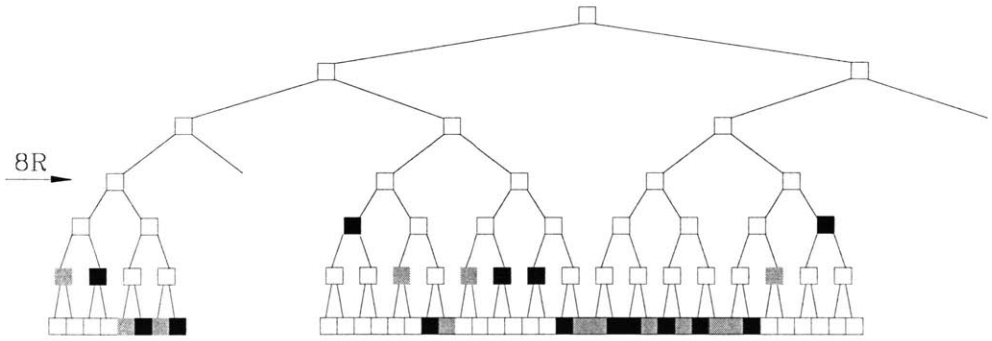



**Figure** 7.31: New topology of tree with 8R candidate branches

$$B_8(1) = 3/4 \text{ (underpopulated)}$$
$$B_8(3) = 2/5$$
$$B_8(4) = 3/5 \text{ (underpopulated)}$$
$$B_8(5) = 4/4 \text{ (severely underpopulated)}$$
$$B_8(6) = 2/5$$

From Tables 7.4—7.7, the cost comparisons are:

$$B_8(1) < B_8(3)$$
$$B_8(3) = B_8(6)$$
$$B_8(6) < B_8(5)$$
$$B_8(5) < B_8(4)$$

Note that because $B_8(5)$ is severely underpopulated and there are no other vacant 2R branches, its cost is lower than $B_8(4)$. The minimum-cost branch is $B_8(1)$. Its descendants $C_2(2)$, $C_1(6)$ and $C_1(8)$ must be reassigned. After removing $B_8(1)$, the resulting topology with 2R target branches is shown in Fig. 7.32. Next, fully loaded 2R branches are removed, resulting in the final topology of Fig. 7.33. At this point, the reassignment is straightforward. A vacant 2R branch with the lowest index is chosen. The complete reassignments to vacate 8R branch $C_8(1)$ are:

$$C_2(2) \longrightarrow C_2(11)$$
$$C_1(6) \longrightarrow C_1(24)$$
$$C_1(8) \longrightarrow C_1(25)$$

For the sake of simplicity, we choose an example where the reassignment is not *nested*; that is, the reassignment of descendant codes — $C_2(2)$, $C_1(6)$ and $C_1(8)$ — does not lead to reassignment of codes in other branches. If our example is modified such that the target 8R branch has a descendant code of rate 4R, then it is necessary in the next step to find a minimum cost 4R branch. This step may result in reassignments of 2R and R codes from other branches. This is an example of *nested* reassignment.

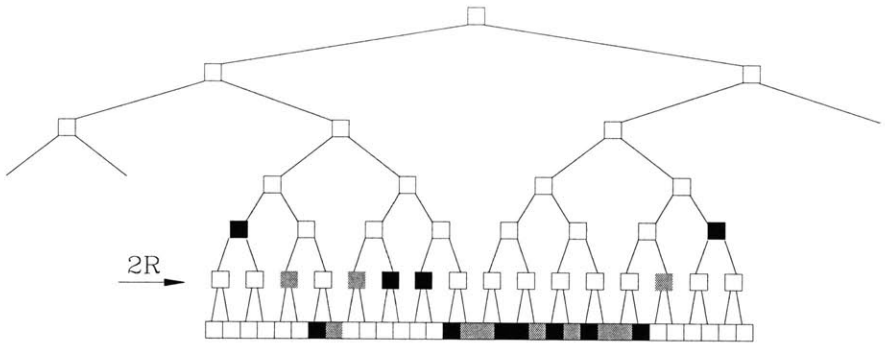



**Figure** 7.32: New topology of tree with 2R candidate branches

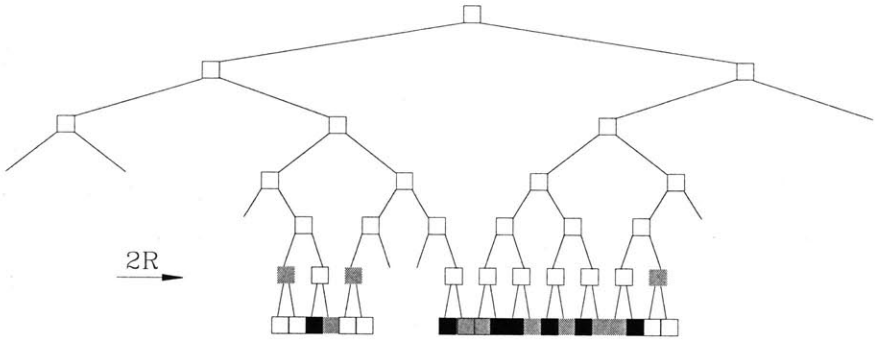



**Figure** 7.33: Final topology of tree with 2R candidate branches

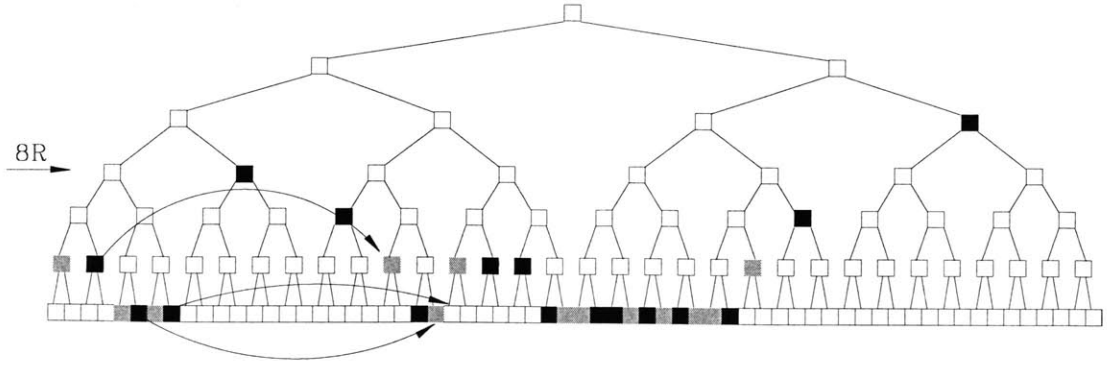
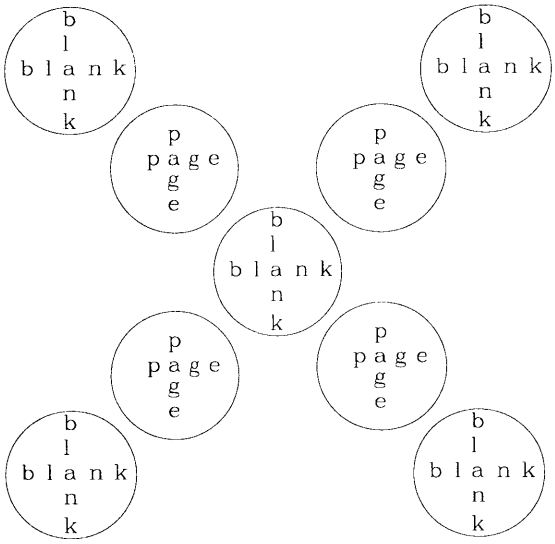



**Figure** 7.34: Reassignment of descendant codes of 8R branch


blank
page
blank
page
blank
page
blank
page
blank
blank
blank

# CONCLUSION

There is no denying that future wireless systems and applications will dominate every aspect of our daily lives at home and in business environment. It is likely that most users —nomadic or otherwise— will embrace the ubiquity concept —any time, anywhere— of mobile computing. In the near future we will also demand wireless infrastructure to provide "any type" of application, including non-voice services such as streaming/live video, high-resolution imaging and other bandwidth hungry services, and at the same time, maintaining a certain guaranteed level of QoS as user population grows. Unfortunately, current wireless systems are not designed to handle heavy traffic loads. We have witnessed voice quality degradation (due to low SIR) in IS-95 CDMA network during busy peak usage hours. Similarly, we have suffered from dropped or busy/blocked calls using a GSM handset due to capacity saturation at hotspots. This matter is similarly evident in public Wi-Fi IEEE 802.11x wireless Ethernet coverage areas where the throughput dramatically falls when many users are accessing the network.

In this monograph we set out to achieve two major objectives:

- Learn current and future trends in wireless applications and services, and understand how current 1G and 2G networks fail or fall short in addressing such needs.
- Based on lessons learned, design a robust and efficient cellular wireless network that can meet the ever-growing appetite for bandwidth-hungry applications by using the most efficient signal transmission and detection methods

### 7.6.3 Topics Covered

In the following we give a summary of topics that were addressed, and in most cases, resolved:

- When we tried to compare the efficacy of 2G networks, particularly the dominant networks of GSM and IS-95, we encounter difficulty in completeness and fairness. For exam-

ple, wireless systems are duplex communication infrastructures. The uplink and downlink performance must be measured separately for frequency-division duplexed systems such as these. Therefore, we quickly limit our scope to cover only downlink communication where there is greater potential for high rate transmission needs (per channel) and likelihood of traffic congestion. Now, we ask ourselves, do we compare downlinks of GSM versus IS-95? Rather, should we study the pros and cons of TDM- and CDM-based networks? We also quickly realize channel multiplexing in the downlink of a cellular wireless network is a two-fold process: primary sharing of total resources among cells within the network, and secondary sharing of allocated bandwidth among forward link channels per autonomous cell. It is complete only after we specify every network configuration in terms of inter- and intra-cell multiplexing schemes. Thus, GSM has FDM/TDM structure while IS-95 is FDM/CDM.

- First, we measure the capacity and simultaneously achievable rate regions of various cellular multiplexing schemes: FDM/FDM, FDM/TDM, FDM/CDM, etc. We observe that all intra-cell orthogonal channel multiplexing techniques, TDM, FDM and orthogonal CDM (O-CDM) have the same ARR. The optimal scheme —superposition coding and successive interference cancellation— which dominates over the orthogonal schemes is not pursued as a viable candidate due to its receiver complexity, double feedback requirement and partial understanding in the design of superposition codes. The more pressing issue is the ARR among various inter-cell channel multiplexing schemes. Here, there is no clear winner. Depending on background noise levels and amount of interference from neighboring cells and propagation loss, FDM may have a larger ARR than N-CDM with SIC, or vice versa. it became apparent that in measuring the efficiency of a cellular broadcast network, we must take into account the joint ARR when an inter-cell channel multiplexing technique is paired with an intra-cell channel multiplexing technique. We notice immediately that the only two viable candidates for inter-cell channel multiplexing are FDM and N-CDM. The former is preferred if we wish to maintain inter-cell channel orthogonality for channel separation (zero cross-correlation), whereas the latter applies non-orthogonal coding with non-zero but controlled amount of cross-correlation. In inter-cell FDM, interference arises from distant (co-channel) cells that reuse the same frequency band. The only interference suppression mechanism is propagation path loss. In inter-cell N-CDM, every neighboring cell is an interferer. It uses information-independent coding (and to a lesser extent, propagation loss) to suppress interference. We then measure the ARR of two cellular broadcast networks based on these paradigms —FDM/ODM versus N-CDM/O-CDM. The winner is N-CDM/O-CDM.

- Next, we design and construct two type of signature code sequences suitable for inter- and

intra-cell channel multiplexing, respectively. For N-CDM, conventional pseudo-random codes were used. For O-CDM, we studied both recursive and non-recursive orthogonal codes. When combined, the resulting signature sequences are known as superimposed. By computing statistical correlation properties, we prove that superimposed signature sequences maintain the same randomness characteristics of random PN codes when the code length is long. Furthermore, the choice of orthogonal codes —recursive or non-recursive— has little effect on overall correlation property of the superimposed sequence. Over a shorter window of $N_c$, superimposed codes still remain orthogonal when used as broadcast codes. For seamless multi-rate transmission, recursive orthogonal binary codes are preferred. Hence, by taking into account all the factors listed thus far, our chosen resource allocation policy is N-CDM/O-CDM with PN spreading sequences for Inter-cell channelization and binary, superimposed, recursive orthogonal codes for intra-cell channelization.

- All the operations listed so far can be implemented in discrete-time domain. Both data- and spread-spectrum modulation are considered discrete-time signal processing schemes. After level-shifting and discrete-to-analog conversion, the impulse samples must be shaped via a baseband lowpass prototype filter for waveform generation, after which the pulse stream is frequency translated to RF by carrier modulation. As is well known, pulse shaping and carrier modulation are analog signal processing techniques that must be jointly optimized. For carrier modulation, we have two choices: single versus multiple carriers. In single-carrier case, a pulse shape that either minimizes out-of-band power leakage or results in zero ISI at the output of receive filter or both is preferred. Zero ISI for single-carrier modulated signal is known as conventional or single-carrier Nyquist criterion. For multi-carrier case, the role of pulse shaping is more important since the amplitude and phase response of chosen pulse determines the amount of inter-carrier and inter-block interference —besides ISI. A pulse that causes no distortion at the receiver output (zero ISI, zero ICI and zero IBI) is said to meet the generalized Nyquist criterion. In particular, we are interested in g-Nyquist pulses that are finite in length; i.e., they can be synthesized and analyzed using FIR filters. Such pulses do exist; we coin them time-overlapped pulses. By using such pulses along with a bank of properly spaced cosine carriers, we showed that the resulting multi-carrier modulated signal is g-Nyquist and more bandwidth efficient than a similarly designed single-carrier modulated signal with Nyquist pulse shaping. Therefore, the baseband output sequence —which is the sum of binary superimposed, recursive orthogonal, data modulated sequences— must be serial-to-parallel converted and shaped via a bank of identical baseband filters, each with time-overlapped impulse response. Each parallel channel is modulated by a cosine carrier of identical phase. The sum of parallel output channels is the final composite broadcast signal.

- After describing various signal processing stages of a transmitter, from data modulation of information symbols to the generation of output RF analog waveform, we must carry out the detection operations at the receiving end. In particular, we studied bit-error rate analysis of binary data and spread-spectrum modulated CDM signalling in a multi-user interference environment. We considered BERs for both low- and high-rate CDM data channels with a common chip rate. For high rate CDM data transmission, we compared the performance of parallel-channel single gain scheme against single-channel reduced gain scheme. For all practical purposes, their BER performance are comparable. From the receiver's perspective, SC-RG CDM is preferred for lower receiver complexity.

- We then point out that for variable rate transmission, SC-RG CDM and PC-SG CDM are comparable on a per-channel basis. When overall network capacity is measured, a CDM network that supports SC-RG CDM channels may have much lower spectral efficiency due to a scheduling constraint known as code blocking. To eliminate code blocking completely, we propose two different strategies based on reassignments of signature codes in active CDM component channels. We also devise an algorithm that supports multiplexing of both constant bit rate and bursty logical channels.

### 7.6.4 Topics not Covered

- When studying the composite channel for cellular broadcasting, we did not address variation in ARR of various channel multiplexing schemes when the transmission medium induces multipath fading. Furthermore, we did not cover the severity of different fading distributions on the ARR.

- When we studied superimposed orthogonal codes, less emphasis was placed on the construction and selection of PN sequences for inter-cell channel multiplexing. It is likely that when superimposed, certain sequences offer better auto- and cross-correlation values than other alternatives. This is a subject outside of our scope and expertise.

- Our summary on multipath fading applies mainly to single-carrier modulated signals. For multi-carrier modulation, we must take into account the correlation of channel impulse responses among sub-bands that make up the entire bandwidth.

- Related to correlated multipath fading is performance analysis of multi-carrier modulated signals. We propose multi-carrier modulation with time-overlapped pulses since this combination offers higher spectral efficiency in power spectrum width (compared to single-carrier modulated signal) plus zero distortion (i.e., satisfying the g-Nyquist criterion.) We, however, did not measure the BER performance of such signals. Therefore, we

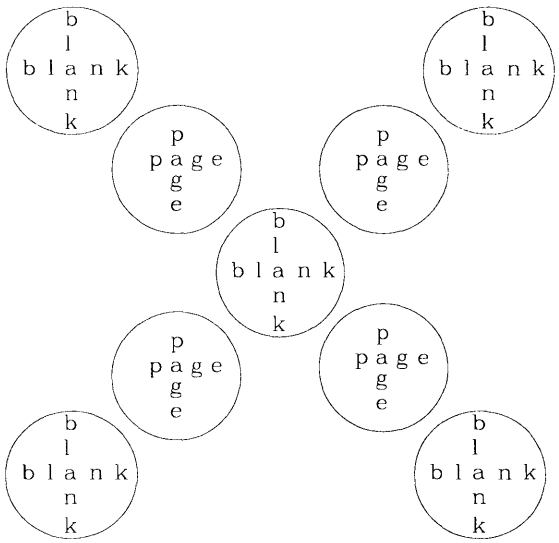
were not able to report an unbiased BER comparison between single- and multi-carrier modulated CDM signalling.

### 7.6.5 Topics of Interest

- As we pointed out in Chapters 3 and 6, CDM discrete-time sequences can be spread in time of frequency or in both domains. When spread in time, the resulting signal is the conventional single-carrier direct-sequence CDM signal. When spread in frequency, it is commonly known as multi-carrier CDM signal. other equivalent terms are OFDM-CDM and multi-tone CDM. By spreading in both time and frequency domains, we can generate a hybrid CDM signal whose sub-carrier bandwidth is larger than that of OFDM-CDM, but much smaller than the total bandwidth of single-carrier CDM signal. Since Walsh code lengths exist for powers of 2, the total number of sub-carriers can vary as 2, 4, 8 16, 32, 64 and so on. it would be interesting to measure the BER performance of hybrid CDM signalling for a given multipath fading channel response as the number of sub-carriers to varied.

- In discussing zero correlation zones, we briefly mentioned that such codes do exist for inter-cell channel multiplexing that results in zero auto or cross-correlation over a finite-length window known as correlation zone. Since assigned signature codes are super-imposed, it would be interesting to study this topic further, and analyze if such zero correlation zones exist for superimposed orthogonal binary sequences.

- In describing OVSF code reassignment strategies, we hinted that in the future, handsets may have increased processing power to decode several parallel CDM channels simultaneously. If this were the case, the total requested rate can be split among parallel CDM channels, with each channel supporting variable rate through OVSF-CDM. Under such circumstances, there is more flexibility in assigning and reassigning codes to remove code blocking and statistical multiplexing. The design and synthesis of an efficient resource allocation policy that can assign, say k, parallel channels at once to every receiver is of great interest, for both academic and practical purposes.


blank
page
blank
page
blank
page
page
blank
blank

# SYMBOLS & NOTATIONS

■ *Non-subscripted Letters*

| | |
|---|---|
| c ..... | speed of light |
| d ..... | generic distance between a transmitter and a receiver |
| D ..... | dimension of signal set |
| I ..... | number of information sources in the network |
| J ..... | number of interfering single-rate CDM users |
| K ..... | 1) constraint length of a convolutional code<br>2) shaping parameter of Ricean probability density function |
| L ..... | number of independent channels in multi-channel diversity combining |
| M ..... | 1) generic number of users (sinks or sources) in the network<br>2) number of parallel channels in a high-rate PC-SG CDM channel<br>3) number of bits per symbol |
| N ..... | length of an orthogonal (Walsh) code |
| P ..... | period of a generic discrete-time signature sequence |
| R ..... | 1) number of receivers in the network<br>2) generic number of symbols or waveforms per second (symbol rate) |
| r ..... | roll-off factor of a practical Nyquist pulse |
| S ..... | number of information sinks in the network |
| T ..... | 1) number of transmitters in the network<br>2) generic time duration of a symbol or waveform (in sec.) |
| W ..... | generic (passband) bandwidth |

## ■ *Non-subscripted Symbols*

| | |
|---|---|
| $\lambda$ ..... | wavelength (typically in meters) of continuous-time RF signal, $\lambda = c/f_c$ where c is the speed of light |
| $\gamma$ ..... | exponent of propagation loss, typically as power of distance: $d^{\gamma}$ |
| $\tau$ ..... | propagation delay |
| $\alpha$ ..... | 1) fixed attenuation 2) free parameter in a fading pdf |
| $\beta$ ..... | free parameter in a fading pdf |
| $\nu$ ..... | Doppler frequency shift |

## ■ *Calligraphic Symbols*

| | |
|---|---|
| $\mathcal{B}_i$ ..... | balance of channel i in OVSF-CDM multiplexing |
| $\mathcal{C}_i$ ..... | accrued credit of channel i in OVSF-CDM multiplexing |
| $\mathcal{C}^o$ ..... | capacity region of optimal channel multiplexing scheme |
| $\mathcal{C}^s$ ..... | capacity region where every channel has single-user capacities |
| $\mathcal{E}$ ..... | generic symbol for energy of pulse, symbol or sequence |
| $\mathcal{E}_b$ ..... | bit energy |
| $\mathcal{E}_c$ ..... | energy of a common chip pulse |
| $\mathcal{E}_D$ ..... | energy per dimension |
| $\mathcal{E}_s$ ..... | symbol energy, same as $\mathcal{E}$ |
| $\mathcal{M}$ ..... | message book |
| $\mathcal{R}^F$ ..... | ARR of any ODM (FDM, TDM or O-CDM) scheme |
| $\mathcal{R}^{NC}$ ..... | ARR of naive CDM scheme |
| $\mathcal{R}^{NT}$ ..... | ARR of naive TDM (simple time sharing) scheme |
| $\mathcal{U}_i$ ..... | usage of channel i in OVSF-CDM multiplexing |
| $\mathcal{V}$ ..... | symbol code book of broadcaster |
| $\mathcal{X}$ ..... | symbol code book of generic source |

## ■ *Subscripted Letters and Symbols*

| | |
|---|---|
| $A_i$ ..... | generic amplitude of transmit signal of user $i$ |
| $B_c$ ..... | coherence bandwidth |
| $B_d$ ..... | Doppler spread |
| $c_{ij}$ ..... | OVSF code with layer no. $i$ and branch no. $j$ |
| $f_c$ ..... | generic passband RF carrier frequency |
| $f_i$ ..... | passband carrier frequency of $i^{th}$ channel |
| $f_o$ ..... | generic passband IF carrier frequency |
| $\mathcal{H}_N$ ..... | $(N \times N)$ Hadamard orthogonal matrix with entries from Galois field $\mathbb{F}_2$ |
| $I_i$ ..... | 1) $i^{th}$ source in the network<br>2) interference sample from neighboring site |
| $K_c$ ..... | number of mobiles in a cell |
| $k_c$ ..... | excess spreading factor; equal to $N_c/N$ |
| $m_i$ ..... | message from source $i$ |
| $L_p$ ..... | propagation loss as the ratio of receive and transmit powers |
| $N_c$ ..... | number of chips in a data symbol, equal to $T/T_c$ |
| $N_{eff}$ ..... | effective reuse number |
| $N_{max}$ ..... | maximum spreading factor |
| $N_o$ ..... | one-sided Gaussian noise PSD |
| $N_r$ ..... | frequency reuse number |
| $N_p$ ..... | period of a discrete-time pseudo-random sequence |

■ *Subscripted Letters and Symbols (contd.)*

$P_e$ ..... bit error probability

$P_G$ ..... processing gain, same as PG

$P_r$ ..... receive power

$P_t$ ..... transmit power

$r_c$ ..... rate of an information-dependent channel code

$R_i$ ..... $i^{th}$ receiver in the network

$R_{max}$ ..... maximum bit rate per channel supported by system

$R_p$ ..... pulse rate; if pulse width is $T_c$, $R_p = R_c$

$R_s$ ..... symbol rate, interchangeable with R

$S_i$ ..... $i^{th}$ sink in the network

$T_i$ ..... $i^{th}$ transmitter in the network

$T_p$ ..... pulse width; inverse of the pulse rate $R_p$

$T_s$ ..... symbol duration, interchangeable with T

$W_c$ ..... cell bandwidth

$W_m$ ..... mobile bandwidth

$\mathcal{W}_N$ ..... $(N \times N)$ Walsh orthogonal matrix with entries from integer ring $\mathbb{D}_2$

$W_{ss}$ ..... bandwidth of spread-spectrum modulated signal

$W_{tot}$ ..... total bandwidth allocated to entire network

$\beta_i$ ..... relative channel power gain of interfering user i

$(\Delta\tau)_d$ ..... delay spread

$(\Delta t)_c$ ..... coherence time

$\delta_{ij}$ ..... Kronecker delta function, same as $\delta[i-j]$
$\delta_{ij} = 1$ for $i = j$ and zero elsewhere

$\phi_i$ ..... relative carrier phase offset of interfering user i

$\theta_i$ ..... absolute carrier phase of interfering user i

$\tau_i$ ..... relative propagation delay offset of interfering user i

## ■ *Discrete-time Sequences*

$a_j[n]$ ..... superimposed orthogonal modulating sequence of user/channel j

$b_j[n]$ ..... data and spread-spectrum modulated sequence of user/channel j

$f_i[n]$ ..... 1) impulse response of transmit/synthesis filter in channel i of multi-carrier/filter bank system

$f_o[n]$ ..... 1) impulse response of prototype transmit/synthesis lowpass filter of multi-carrier/filter bank system

$h_i[n]$ ..... 1) impulse response of receive/analysis filter in channel i of multi-carrier/filter bank system

$h_o[n]$ ..... 1) impulse response of prototype receive/analysis lowpass filter of multi-carrier/filter bank system

$s[n]$ ..... discrete-time transmit sequence equivalent to $s(t)$ in continuous-time

$X_j[n]$ ..... Galois-binary data sequence of user j

$x_j[n]$ ..... antipodal-binary data sequence of user j

$\delta[n]$ ..... discrete-time impulse sequence $\delta[0] = 1$ and zero elsewhere

$\delta[i-j]$ ..... Kronecker delta function, same as $\delta_{ij}$ $\delta[i-j] = 1$ when $i = j$ and zero elsewhere

$\delta_M[n]$ ..... discrete-time periodic impulse sequence of period M $\delta[kM] = 1$ for all integer k and zero elsewhere

$\theta_{ij}[l; i, d]$ ..... (un-normalized) partial cross-correlation of two discrete sequences with chip offset l, starting position i and length d

$\theta_{ij}[l; i, d]$ ..... (un-normalized) partial auto-correlation of a discrete sequence with chip offset l, starting position i and length d

$\Theta_{ij}[l; i, d]$ ..... (normalized) partial cross-correlation of two discrete sequences with chip offset l, starting position i and length d

$\theta_{ij}[l; i, d]$ ..... (normalized) partial auto-correlation of a discrete sequence with chip offset l, starting position i and length d

## ■ *Continuous-time Functions*

$c(t)$ ..... linear time-invariant baseband channel impulse response

$c(t;\tau)$ ..... linear time-variant baseband channel impulse response

$C_{ij}(\tau)$ ..... continuous-time periodic cross-correlation function

$C_j(\tau)$ ..... continuous-time periodic auto-correlation function

$c_p(t)$ ..... linear time-invariant passband channel impulse response

$c_p(t;\tau)$ ..... linear time-variant passband channel impulse response

$f(t)$ ..... 1) baseband transmit filter impulse response
2) synthesis prototype filter in a multi-channel filter bank

$g(t)$ ..... output of baseband transmit filter $f(t)$
usually it is already data (and spread-spectrum) modulated

$g_k(t)$ ..... output of baseband transmit filter $f_k(t)$ of $k^{th}$ user/channel

$h(t)$ ..... 1) baseband receive filter impulse response
2) analysis prototype filter in a multi-channel filter bank

$L(t)$ ..... number of resolvable multipaths, same as $L_r(t)$

$L_u(t)$ ..... number of unresolvable multipaths

$n_I(t)$ ..... complex Gaussian noise in-phase component

$n_Q(t)$ ..... complex Gaussian noise quadrature component

$R_c(t;\tau)$ ..... LTV channel auto-correlation function

$R_c(\tau)$ ..... delay spread function

$R_{ij}(\tau)$ ..... continuous-time aperiodic cross-correlation function
with $[0,\tau]$ window of periodic waveforms

$\hat{R}_{ij}(\tau)$ ..... continuous-time aperiodic cross-correlation function
with $[\tau,T]$ window of periodic waveforms

$R_j(\tau)$ ..... continuous-time aperiodic auto-correlation function
with $[0,\tau]$ window of a periodic waveform

$\hat{R}_j(\tau)$ ..... continuous-time aperiodic auto-correlation function
with $[\tau,T]$ window of a periodic waveform

$r(t)$ ..... baseband received signal

■ *Continuous-time Functions (contd.)*

| | |
|---|---|
| $r_p(t)$ ..... | passband received signal |
| $s(t)$ ..... | sum of carrier and data-modulated signals $\sum_k s_k(t)$ |
| $s_k(t)$ ..... | carrier and data-modulated signal of $k^{th}$ user/channel |
| $\mathcal{S}_c(f;\tau)$ ..... | LTV scattering function, Fourier transform of $R_c(t;\tau)$ |
| $U(x)$ ..... | unity function; equal to 1 for $x \geq 0$ |
| $v(t)$ ..... | baseband receive signal after carrier demodulation |
| $x(t)$ ..... | output of discrete-to-continuous time (D/C) delta modulator |
| $y(t)$ ..... | output of baseband receive filter $h(t)$ |
| $\tau_n(t)$ ..... | arrival time delay in $n^{th}$ resolvable path relative to zeroth path |
| $\alpha_n(t)$ ..... | channel (amplitude) gain in $n^{th}$ resolvable path |
| $\theta_{ij}(\tau; T_i, T_f)$ ... | continuous-time (un-normalized) partial cross-correlation function with offset $\tau$ and $[T_i, T_f]$ window |
| $\theta_j(\tau; T_i, T_f)$ ... | continuous-time (un-normalized) partial auto-correlation function with offset $\tau$ and $[T_i, T_f]$ window |
| $\Theta_{ij}(\tau; T_i, T_f)$ .. | continuous-time (normalized) partial cross-correlation function with offset $\tau$ and $[T_i, T_f]$ window |
| $\Theta_j(\tau; T_i, T_f)$ ... | continuous-time (normalized) partial auto-correlation function with offset $\tau$ and $[T_i, T_f]$ window |
| $\rho_{ij}(\tau)$ ..... | continuous-time aperiodic cross-correlation function of finite-length waveforms with $[0, \tau]$ window |
| $\hat{\rho}_{ij}(\tau)$ ..... | continuous aperiodic CCF of finite-length waveforms with $[\tau, T]$ window |
| $\rho_j(\tau)$ ..... | continuous-time aperiodic auto-correlation function of a finite-length waveform with $[0, \tau]$ window |
| $\hat{\rho}_j(\tau)$ ..... | continuous aperiodic ACF of finite-length waveform with $[\tau, T]$ window |
| $\delta(t)$ ..... | continuous-time impulse function |
| $\psi(f)$ ..... | Doppler spread function |
| $\psi(t)$ ..... | impulse response of transmit chip filter |

## ■ *Mathematical Operators*

| | |
|---|---|
| '$\star$' ..... | continuous-time convolution |
| '$*$' ..... | discrete-time linear convolution |
| '$\oplus$' ..... | modulo-2 addition (exclusive-OR operation) |
| '$+$' ..... | real addition |
| '$\odot$' ..... | modulo-2 multiplication |
| '$\cdot$' ..... | real multiplication, same as '$\times$' |
| '$\otimes$' ..... | element-wise vector multiplication |
| $\mathcal{F}[\,\cdot\,]$ ..... | Fourier transform |
| $\mathcal{F}^{-}[\,\cdot\,]$ ..... | inverse Fourier transform |
| $\mathrm{E}[\,\cdot\,]$ ..... | statistical mean |
| $\mathrm{Var}[\,\cdot\,]$ ..... | statistical variance |
| $\Re\{\,\cdot\,\}$ ..... | real part of expression in brackets |

## ■ *Limits and Inequalities*

| | |
|---|---|
| $[m, n]$ ..... | the set of all real numbers between $m$ and $n$ (both inclusive and $n > m$), same as $m \leq x \leq n$ |
| $(m, n]$ ..... | the set of all real numbers between $m$ and $n$ (inclusive of $n$ and $n > m$), same as $m < x \leq n$ |
| $[m, n)$ ..... | the set of all real numbers between $m$ and $n$ (inclusive of $m$ and $n > m$), same as $m \leq x < n$ |
| $[m, n]$ ..... | the set of all real numbers between $m$ and $n$ (both exclusive and $n > m$), same as $m < x < n$ |
| $\left\lfloor \frac{x}{M} \right\rfloor$ | the set of all all integer multiples of $M$, same as $\pm M\,\mathbb{N}$ $\{\ldots, -2M, -M, 0, M, 2M, \ldots\}$ |

## ■ *Poly-lined Symbols for Number Sets*

$\mathbb{Z}$ ..... the set of all integers

$\mathbb{Z}_i$ ..... $\{i, i+1, i+2, \ldots, \infty\}$ for integer $i > 0$

$\mathbb{Z}_+$ ..... the set of all positive integers, same as $\mathbb{Z}_1$

$-\mathbb{Z}_+$ ..... the set of all negative integers, same as $\mathbb{Z}_-$

$2\mathbb{Z}_+$ ..... the set of all positive even integers

$\mathbb{Z}_i^n$ ..... $\{i, i+1, i+2, \ldots, n\}$, both $i$ and $n$ are positive integers

$\mathbb{Z}^i$ ..... $\{1, 2, \ldots, i-1, i\}$, same as $\mathbb{Z}_1^i$

$-\mathbb{Z}^i$ ..... $\{-1, -2, \ldots, -i+1, i\}$

$\pm\mathbb{Z}^i$ ..... $\{-i, -i+1, \ldots, -1, 1, \ldots, i-1, i\}$

$\pm\mathbb{Z}^i$ ..... the set of all positive and negative integers, excluding zero

$\mathbb{N}$ ..... the set of all natural numbers
same as $\mathbb{Z}_0$ or $\{0, \mathbb{Z}_+\}$

$\mathbb{N}^i$ ..... $\{0, 1, 2, \ldots, i-1, i\}$
same as $\mathbb{Z}_0^i$ or $\{0, \mathbb{Z}^i\}$

$-\mathbb{N}^i$ ..... $\{0, -1, -2, \ldots, -i+1, -i\}$
same as $-\mathbb{Z}_0^i$ or $\{0, -\mathbb{Z}^i\}$

$\pm\mathbb{N}^i$ ..... $\{-i, -i+1, \ldots, -1, 0, 1, \ldots, i-1, i\}$
same as $\pm\mathbb{Z}_0^i$

$\mathbb{F}_2$ ..... binary Galois field $\{0, 1\}$
same as $\mathbb{Z}_0^1$

$\mathbb{D}_2$ ..... antipodal binary set $\{-1, 1\}$
same as $\pm\mathbb{Z}^1$

$\mathbb{D}_M$ ..... antipodal M-ary set (M is even and every element is odd)
$\{-(M-1), -(M-3), \ldots, -1, 1, \ldots, (M-3), (M-1)\}$

$\mathbb{D}$ ..... the set of all odd integers

$|\mathbb{D}|$ ..... the set of all positive odd integers

## ■ *Parameterized Functions*

| | |
|---|---|
| $\Gamma(\alpha)$ ..... | gamma function |
| $f(r)$ ..... | generic expression for a probability density function, generally the time variable t is not used for indexing |
| $F(r)$ ..... | generic expression for a cumulative distribution function |

## ■ *Notations for Probability Density Functions*

| | |
|---|---|
| $X \sim \Gamma(\alpha, \beta)$ ..... | gamma pdf with free parameters $\alpha$ and $\beta$ |
| $X \sim \mathcal{N}(\mu, \sigma^2)$ ..... | Gaussian pdf with mean $\mu$ and variance $\sigma^2$ |
| $X \sim \chi^2(r)$ ..... | standard chi-squared pdf with r degrees of freedom |
| $X \sim \chi^2(r, \sigma^2)$ ..... | central chi-squared pdf with r degrees of freedom and $\sigma^2$ |
| $X \sim \chi^2(r, \sigma^2, \mu^2)$ ..... | non-central chi-squared pdf with r degrees of freedom |
| $X \sim \chi(n)$ ..... | standard chi pdf with n degrees of freedom |
| $X \sim \chi(n, \sigma^2)$ ..... | central chi pdf with n degrees of freedom and $\sigma^2$ |
| $X \sim \chi(n, \sigma^2, \mu^2)$ ..... | non-central chi pdf with n degrees of freedom |
| $R \sim \chi(1, \sigma^2)$ ..... | one-sided Gaussian pdf with $\sigma^2$ |
| $R \sim \chi(2, \sigma^2)$ ..... | Rayleigh pdf with $\sigma^2$ |
| $R \sim \chi^2(2, \sigma^2, \mu^2)$ ..... | Rice or Nakagami-n pdf |
| $R \sim q(q, \Omega)$ ..... | Nakagami-q pdf with $\Omega$ |
| $Y \sim bB(p)$ ..... | bipolar Bernoulli pdf with $\Pr(Y = 1) = p$ |
| $V \sim bBin(n, p)$ ..... | bipolar Binomial pdf with $\Pr(V = 1) = p$ |
| $X \sim \Gamma(\alpha, \beta)$ ..... | gamma pdf with free parameters $\alpha$ and $\beta$ |

■ *Acronyms*

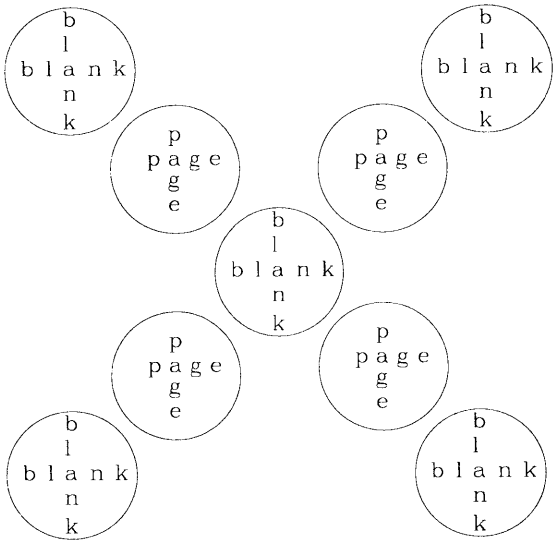
| | |
|---|---|
| AC ..... | auto-correlation |
| ACF ..... | auto-correlation function |
| AF ..... | amount of fading (a measure of fading severity) |
| AFD ..... | average fade duration |
| ARR ..... | (simultaneously) achievable rate region |
| AWGN ..... | additive white Gaussian noise |
| BER ..... | bit-error rate; same as $P_e$ |
| BT ..... | bandwidth-time period product as a measure of excess bandwidth |
| CC ..... | cross-correlation |
| CCF ..... | cross-correlation function |
| CDMA ..... | code-division multiple access |
| CI ..... | channel interference (which occurs at Rake finger output due to non-ideal cross-correlation between parallel channels) |
| CW ..... | continuous-wave (unmodulated) signal |
| DCT ..... | discrete cosine transform |
| DFT ..... | discrete Fourier transform |
| DS-CDM ..... | direct-sequence CDM, same as phase-coded SSM |
| ELT ..... | extended lapped transform |
| FDM ..... | frequency-division multiplexing |
| FDMA ..... | orthogonal-division multiple access |
| FF ..... | flat fading |
| FFT ..... | fast computation of discrete samples of DFT |
| FH-CDM ..... | frequency-hopped CDM |
| FSF ..... | frequency selective fading |
| IBI ..... | interblock interference (i.e., ISI from past symbols in other sub-carriers/channels) |

■ *Acronyms (contd.)*

| | |
|---|---|
| ICI | inter-channel or inter-carrier interference<br>(i.e., ISI from current symbols in other sub-carriers/channels) |
| iid | independent, identically distributed (random variables) |
| I/O | input-output device (e.g., communication channel) |
| I/Q | in-phase and quadrature modulation format |
| I-SI | ISI causing self interference |
| ISI | intersymbol interference |
| LCR | level-crossing rate |
| LOS | line-of-sight |
| LTI | linear time-invariant |
| LTV | linear time-varying |
| ML | maximal-length (pseudo-random sequence) |
| MLT | modulated lapped transform |
| MSK | minimum shift keying modulation |
| N-TDM | naive time-division multiplexing |
| N-CDM | naive code-division multiplexing |
| N-CDMA | naive code-division multiple access |
| O-CDM | orthogonal code-division multiplexing<br>its use as a collective representation of TDM, FDM and O-CDM |
| ODM | orthogonal-division multiplexing |
| OFDM | orthogonal frequency-division multiplex<br>a multi-carrier modulation format in digital domain |
| O-QPSK | offset quadrature shift keying modulation |
| OVSF | orthogonal variable spreading factor |
| PC-SG | parallel-channel, single-gain CDM transmission<br>also known as multicode-CDM/CDMA |
| PDF | probability density function |

■ *Acronyms (contd.)*

| | |
|---|---|
| PG | processing gain, same as $P_G$ |
| PN | pseudo-noise sequence |
| PSD | power spectral density |
| QPSK | quadrature shift keying modulation |
| RF | radio frequency |
| SC-RG | single-channel, reduced-gain CDM transmission<br>also known as OVSF-CDM/CDMA or VSG-CDMA |
| SC-SG | single-channel, single-gain CDM transmission<br>also known as conventional CDM/CDMA |
| SDR | signal-to-disturbance power ratio<br>disturbance is the sum of interference and Gaussian noise |
| SF | spreading factor, same as spreaing gain, denoted by $N_c$ |
| SNR | signal-to-noise power ratio |
| SI | self-induced noise or self interference (that exists at<br>Rake finger output due to non-ideal auto-correlation) |
| SIC | successive interference cancellation |
| SIR | signal-to-interference ratio |
| U-SI | uncorrelated self interference |
| SHF | super-high frequency range (3 – 30 GHz) |
| SSM | spread-spectrum modulation/multiplexing |
| TDM | time-division multiplexing |
| TDMA | time-division multiple access |
| TH-CDM | time-hopped CDM |
| UHF | ultra-high frequency range (300 – 3,000 MHz) |
| VHF | very-high frequency range (30 – 300 MHz) |
| WSS | wide-sense stationary (random process) |
| WSSUS | wide-sense stationary, uncorrelated scattering radio channel |


blank
page
blank
page
blank
page
blank
page
blank
blank

# BIBLIOGRAPHY


[1] F. Adachi, M. Sawahashi and H. Suda, "Wideband CDMA for Next-Generation Mobile Communications Systems," *IEEE Comm. Mag.*, vol. 36, no. 9, Sept. 1998, pp. 56-69.

[2] F. Adachi, M. Sawahashi and K. Okawa, "Tree-structured generation of orthogonal spreading codes with different lengths for forward link of DS-CDMA mobile radio," *IEEE Elect. Lett.*, vol. 33, Jan. 1997, pp. 27-8.

[3] P. Agin and F. Gourgue, "Comparison Between Multicode with Fixed Spreading and Single Code with Variable Spreading Options in UTRA/TDD," $2^{nd}$ *IEEE Workshop on Sig. Proc. Adv. Wireless Comm.*, 1999, pp 325-28.

[4] D. Ayyagari and A. Ephremides, "Cellular Multicode CDMA Capacity for Integrated (Voice and Data) Services," *IEEE J. Sel. Areas Comm.*, Vol. 17, No. 5, May 1999, pp. 928-38.

[5] A. V. Balakrishnan, "A note on the sampling principle for continuous signals," *IRE Trans. Info. Theory*, vol. 3, June 1957, pp. 143-6.

[6] ____, "Essentially band-limited stochastic processes," *IEEE Trans. Info. Thy.*, vol. 11, 1965, pp. 145-56.

[7] P. P. Bergmans and T. M. Cover, "Cooperative Broadcasting," *IEEE Trans. Info. Thy.*, vol. 20, no. 3, May 1974, pp. 317-24.

[8] P. P. Bergmans, *Degraded broadcast channels*, Ph.D. thesis, Elect. Eng. Dept., Stanford U., June 1972.

[9] R. E. Blahut, *Digital Transmission of Information*, Addison-Wesley, 1990.

[10] ____, *Digital Transmission of Information*, Addison-Wesley, 1990.

[11] D. G. Bernnan, "On the maximum signal-to-noise ratio realizable from several noisy signals," *Proc. IRE*, vol. 43, no. 1, Oct. 1955, pp. 1530.

[12] Aydano B. Carliel, "Interference Channels," *IEEE Trans. Info. Thy.*, vol. 24, no. 1, Jan. 1978, pp. 60-70.

[13] ____, "A case where interference does not reduce capacity," *IEEE Trans. Info. Thy.*, Sept. 1975, pp. 569-70.

[14] P. Cassereau, *A New Class of Optimal Unitary Transforms for Image Processing*, Master's Thesis, Mass. Inst. of Tech., Cambridge, MA, May 1985.

[15] P. Cassereau, D. H. Staelin and D. de Jager, "Encoding of images based on a lapped orthogonal transform," *IEEE Trans. Comm.*, vol. 37, Feb. 1989, pp. 189-93.

[16] N. Chan, "Multipath propagation effects on a CDMA cellular system," *IEEE Trans. Veh. Tech.*, vol. 43, Nov. 1994, pp. 848-55.

[17] R. W. Chang, "Synthesis of band-limited orthogonal signals for multi-channel data transmissions," *Bell Sys. Tech. J.*, Aug. 1966, pp. 1052-64.

[18] U. Charash, "Reception through Nakagami fading multipath channels with random delays," *IEEE Trans. Comm.*, vol. 27, no. 4, Apr. 1979, pp. 657-70.

[19] K. Cheun, "Performance of Direct-sequence Spread- spectrum RAKE Receivers with Random Spreading Sequences," *IEEE Trans. Comm.*, vol. 45, Sept. 1997, pp. 1130-43.

[20] G. Cooper and C. McGillem, *Modern Communications and Spread Spectrum*, McGraw-Hill, 1986.

[21] T. M. Cover, "Broadcast Channels," *IEEE Trans. Info. Thy.*, vol. 18, no. 1, Jan. 1972, pp. 2-14.

[22] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[23] H. S. M. Coxeter, *Introduction to Geometry*, 2nd ed., Wiley, New York, 1969.

[24] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, Artech House, 1999.

[25] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall, 1983.

[26] A. Croisier, D Esteban and C. Galand, "Perfect Channel Splitting by use of Interpolaton/Decimation/Tree Decomposition Techniques," *Intl. Conf. Inf. Sci. Sys.*, Patras, Greece. August, 1976.

[27] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akademiai Kiado, Budapest, Hungary, 1986. (ISBN 963 05 7440 3)

[28] E. Dahlman and K. Jamal, "Wide-band Services in a DS-CDMA based FPLMTS System," *Proc. IEEE Veh. Tech. Conf.* Atlanta, GA, May 1996, pp. 1656-60.

[29] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 3rd ed., Brooks/Cole Publishing, 1991.

[30] R. C. Dixon, *Spread-Spectrum Systems*, Wiley-Interscience, New York, 1976.

[31] M. Fan, C. Hoffmann and K. Siu, "Error-rate analysis for multirate DS-CDMA transmission schemes," *IEEE Trans. on Comm.*,vol. 51 , Nov. 2003, pp. 1897-1909.

[32] P. Z. Fan, N. Suehiro, N. Kuroyanagi and X. M. Deng, " Class of binary sequences with zero correlation zone", *IEEE Elect. Lett.*, vol. 35, no. 10, May 1999, pp. 777-9.

[33] P. Z. Fan and M. Darnell, " Construction and comparison of periodic digital sequence sets", *IEEE Proc. Comm.*, vol. 144, no. 6, Dec. 1997, pp. 361-6.

[34] W. Feller, *An Introduction to Probability Theory and Its Applications,* vols. 1 & 2, Wiley, New York, 1966.

[35] M. Fong, V. K. Bhargava and Q. Wang, "Concatenated Orthogonal/PN Spreading Sequences and Their Application to Cellular DS-CDMA Systems with Integrated Traffic," *IEEE J. Sel. Areas Comm.*, vol. 14, no. 3, Apr. 1996, pp. 547-58.

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

[37] Abbas E. Gamal and T. M. Cover, "Multi-user information theory," *Proc. IEEE*, vol. 68, no. 12, Dec. 1980, pp. 1466-83.

[38] W. A. Gardner, "A sampling theorem for non-stationary random processes," *IEEE Trans. Info. Thy.*, vol. 18, no. 3, May 1972, pp. 808-9.

[39] E. A. Geraniotis and M. B.Pursley, "Error Probability for Direct-Sequence Spread-Spectrum Multiple Access Communications—Part II: Approximations," *IEEE Trans. Commun.*, vol. 30, May 1982, pp. 985-95.

[40] K. S. Gilhousen et. al., "On the Capacity of a Cellular CDMA System," *Proc. IEEE Trans. Veh. Tech.*, vol. 40, May 1991, pp. 303-11.

[41] K. S. Gilhousen, U.S. Patent No. 5,751,761.

[42] R. D. Gitlin et al., U.S. Patent No. 5,442,625.

[43] R. Gold, "Optimal Binary Sequences for Spread Spectrum Multiplexing," *IEEE Trans. Info. Thy.*, Oct. 1967.

[44] A. Goldsmith, "Multiuser capacity of cellular time-varying channels," *Twenty-Eighth Asilomar Conf. Sig., Sys. & Comp.*, vol. 1, 1994, pp. 83-8.

[45] ____, "The capacity of downlink fading channels with variable rate and power," *IEEE Trans. Veh. Tech.*, vol. 46, no. 3, Aug. 1997, pp. 569-80.

[46] S. W. Golomb, *Shift Register Sequences*, Holden-Day, 1967.

[47] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products,* $6^{th}$ ed., Academic Press, San Diego, CA, 2000.

[48] L. J. Greenstein et al., "Channel and propagation models for wireless system design I," *IEEE J. Sel. Areas Comm.*, vol. 20, no. 3 , Apr. 2002, pp. 493-5.

[49] F. Hansen and F. I. Meno, "Mobile fading-Rayleigh and log-normal superimposed," *IEEE Trans. Veh. Tech.*, vol. 26, Nov. 1977, pp. 332-5.

[50] S. Hara and R. Prasad, "Overview of Multicarrier CDMA," *IEEE Comm. Mag.*, Dec. 1997, pp. 126-33.

[51] H. Harashima and H. Miyakawa, "A Method of Code Conversion for A Digital Communication Channel with Intersymbol Interference," *IEEE Trans. Comm.*, vol. 20, Aug. 1972, pp. 774-80.

[52] M. J. Ho and G. L. Stuber, "Co-channel interference of microcellular systems on shadowed Nakagami fading channels," *Proc. IEEE Veh. Tech. Conf.*, Secaucus, NJ, May 1993, pp. 568-71.

[53] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed., Prentice Hall, 1995.

[54] J. K. Holmes, *Coherent Spread Spectrum Systems*, John Wiley & Sons, 1982.

[55] J. Holtzman, "A simple accurate method to calculate spread spectrum multiple-access error probabilities," *IEEE Trans. Comm.*, vol. 40, no. 3, March 1992, pp. 461-4.

[56] R. S. Hoyt, " Probability functions for the modulus and angle of the normal complex variate," *Bell Syst. Tech. J.*, vol. 26, Apr. 1947, pp. 318-59.

[57] K. C. Hwang and K. B. Lee, "Performance analysis of low processing gain DS/CDMA systems with random spreading sequences," *IEEE Comm. Lett.*, vol. 2, no. 12, Dec. 1998, pp. 315-7.

[58] Chih-Lin I et al., "IS-95 Enhancements for Multimedia Services," *Bell Labs Tech. J.*, Autumn 1996, pp. 60-87.

[59] C-L. I and R. D. Gitlin, "Multi-code CDMA wireless personal communication networks," *Proc. IEEE Intl. Conf. Comm.* Seattle, WA, vol. 2, June 1995, pp. 1060-4.

[60] ____, "Variable Spreading Gain CDMA," *AT&T Lab Notebook*, August 1993, pp. 9-11.

[61] TIA/EIA Interim Standard-95, *Mobile station - base station compatibility standard for dual-mode wideband spread spectrum cellular system*, July 1993.

[62] W. C. Jakes, ed., *Microwave Mobile Communications*, Wiley & Sons, New York, 1974.

[63] W. Jang, B. Vojcic and R. Pickholtz, "Joint Transmitter-Receiver Optimization in Synchronous Multiuser Communications over Multipath Channels," *IEEE Trans. Comm.*, vol. 46, no. 2, Feb. 1998, pp. 269-78.

[64] A. C. Kam, T. Minn and K. Y. Siu, "Providing Rate Guarantees and Fair Access for Bursty Data Traffic in W-CDMA," *IEEE J. Sel. Areas Comm.*, Vol. 19, Nov. 2001, pp. 2121-30.

[65] V. A. Kotel'nikov, "On the transmission capacity of "ether" and wire in eletrocommunications," *Izd. Red. Upr. Svyazi RKKA*, Moscow, 1933.

[66] K. C. Hwang and K. B. Lee, "Performance analysis of low processing gain DS/CDMA systems with random spreading sequences," *IEEE Comm. Lett.*, vol. 2, no. 12, Dec. 1998, pp. 315-7.

[67] S. Lang, *Algebra*, 3rd ed., Addison-Wesley, 1993.

[68] B. P. Lathi, *Modern Digital and Analog Communication Systems*, HRW Saunders, 2nd ed., 1989.

[69] E. Lee and D.G. Messerschmitt, *Digital Communication*, 2nd ed., Kluwer Academic Publishers, 1994.

[70] S. J. Lee, H. W. Lee, and D. K. Sung, "Capacities of Single-code and Multicode DS-CDMA Systems Accommodating Multiclass Services," *IEEE Trans. Veh. Tech.*, vol. 48, no. 2, March 1999, pp. 376-84.

[71] W. C. Y. Lee, *Mobile Communications Engineering*, 2nd ed., McGraw-Hill, New York, 1998.

[72] ____, "The most spectrum-efficient duplexing system: CDD," *IEEE Comm. Mag.*, vol. 40, no. 3, Mar. 2000, pp. 163-6.

[73] ____, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Trans. Veh. Tech.*, vol. 39, no. 3, Aug. 2000, pp. 187-9.

[74] J. S. Lehnert and M. B. Pursley, "Error probabilities for binary direct-sequence spread-spectrum communications with random signature sequences," *IEEE Trans. Comm.*, vol. 35 no. 1, Jan. 1987, pp. 87-97.

[75] Lifang Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels I: Ergodic Capacity," *IEEE Trans. Info. Thy.*, vol. 47 no. 3, March 2001, pp. 1083-1102.

[76] ____, "Capacity and optimal resource allocation for fading broadcast channels II: Outage Capacity," *IEEE Trans. Info. Thy.*, vol. 47, no. 3, Mar. 2001, pp. 1103-27.

[77] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, 1983.

[78] Fuyung Ling, "Matched filter bound for time-discrete multipath Rayleigh fading channels," *IEEE Trans. Comm.*, vol. 43, no. 2/3/4, Feb./Mar./Apr. 1995, pp. 710-13.

[79] V. H. MacDonald, "The Cellular Concept," *Bell Sys. Tech. J.*, vol. 58, no. 1, 15-43, Jan. 1979.

[80] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, 1977.

[81] H. S. Malvar, "Modulated QMF filter banks with perfect reconstruction," *IEEE Elect. Lett.*, vol. 26, no. 13, Jan. 1990, pp. 906-7.

[82] ____, *Signal Processing with Lapped Transform*, Artech House, 1992.

[83] S. Matsufuji and N. Suehiro, "Complex Hadamard matrices related to bent sequences," *IEEE Trans. Info. Thy.*, vol. 42, no. 2, Mar. 1996, pp. 637.

[84] J. E. Mazo, "Exact matched filter bound for two-beam Rayleigh fading," *IEEE Trans. Comm.*, vol. 39, no. 7, Jul. 1991, pp. 1027-30.

[85] L. B. Milstein and M. S. Lim, "On the performance of a higher order alphabet size in CDMA," *IEEE Comm. Lett.*, vol. 1, no. 1, Jan. 1997, pp. 9-11.

[86] T. Minn and K. Y. Siu, "Dynamic Assignment of Variable-Spreading Factor Codes in W-CDMA," *IEEE J. Sel. Areas Comm.*, vol. 18, no. 8, Aug. 2000, pp. 1429-40.

[87] ____, "Linear transformation of multi-level signal set in multi-code CDMA," *Proc. IEEE Veh. Tech. Conf.*, Rhodes, Greece, May, 2001, pp. 1239-43.

[88] ____, "Variable Chip Rate CDMA," *Proc. IEEE Veh. Tech. Conf.*, Tokyo, Japan, May 2000, pp. 2267-71.

[89] R. K. Morrow and J. S. Lehnert, "Bit-to-bit Error Dependence in Slotted DS/SSMA PAcket Systems with Random Signature Sequences," *IEEE Trans. Comm.*, vol. 37, Oct. 1989, pp. 1052-61.

[90] M. Mouly and M. Pautet, *The GSM System for Mobile Communication*, 1992, ISBN: 2-9507190-0-7.

[91] D. E. Muller, "Applications of Boolean Algebra to Switching Circuit Design and to Error Detection," *IRE Trans.*, EC-2, Sept. 1954, pp. 38-49.

[92] M. Nakagami, " The m-distribution: a general formula of intensity distribution of rapid fading," in W. C. Hoffman ed., *Statistical Methods in Radio Wave Propagation*, Oxford Pergamon Press, 1960, pp. 3-36.

[93] G. Nyquist, "Certain topics in telegraph transmission theory," *AIEE Trans.* vol. 47, 1928, pp. 617-44.

[94] T. Ojanpera, R. Prasad, *Wideband CDMA for Third Generation Mobile Communications*, Artech House, London, 1998.

[95] T. Ottosson and A. Svensson, "Multi-Rate Performance in DS/CDMA Systems," Technical Report No. 14, Dept. of Information Theory, Chalmers University of Technology, Goteborg, Sweden, 1995.

[96] R. Padovani, "Reverse link performance of IS-95 based cellular systems," *IEEE Pers. Comm. Mag.*, vol. 1, no. 3, 1994, pp. 28-34.

[97] A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, 1962.

[98] ____, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 3rd ed., 1991.

[99] D. Parsons, *The Mobile Radio Propagation Channel*, 1st ed., Wiley & Sons, New York, 1992.

[100] "Performance of Multirate Techniques in DS-CDMA Systems," http://perth.mit.edu/.

[101] W. W. Peterson and E. J. Weldon Jr., *Error Correcting Codes*, 2nd ed., MIT Press, 1972.

[102] R. Pickholtz, D. Schilling and L. Milstein, "Theory of Spread spectrum Communications: A Tutorial," *IEEE Trans. Comm.*, vol. 30, no. 5, May 1982, pp. 855-84.

[103] Z. A. Piranashvilli, "On the problem of interpolation of random processes," *Theory Prob. Appl.*, vol. 12, 1967, pp. 647-57.

[104] H. V. Poor and S. Verdu, "Probability of Error in MMSE Multiuser Detection," *IEEE Trans. Info. Thy.*, vol. 43, no. 3, May 1997, pp. 858-71.

[105] R. Price, "Optimal detection of random signals in noise with applications to scatter-multipath communication," *IRE Trans. Info. Theory*, vol. 2, Dec. 1956, pp. 125-35.

[106] R. Price and P. Green, "A communication technique for multipath channels," *Proc. IRE*, vol. 46, Dec. 1958, pp. 555-70.

[107] J. G.Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, New York, 1995.

[108] J. G. Proakis and D. G. Manolakis, *Signal Processing: Principles, Algorithms and Applications*, Prentice Hall, 1996.

[109] M.B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication—Part I: system analysis," *IEEE Trans. Comm.*, vol. 25, Aug. 1977, pp. 795-9.

[110] M. B. Pursley and D. V. Sarwate, "Performance evaluation for phase-coded spread-spectrum multiple-access communication—Part II: Code Sequence Analysis," *IEEE Trans. Comm.*, vol. 25, Aug. 1977, pp. 800-3.

[111] ____, "Bounds on aperiodic corss-correlation for binary sequences," *IEEE Elect. Lett.*, vol. 12, 1976, pp. 304-5.

[112] ____,"Error Probability for Direct-Sequence Spread-Spectrum Multiple Access Communications—Part I: Upper and Lower Bounds," *IEEE Trans. Commun.*, vol. 30, May 1982, pp. 975-83.

[113] S. Ramakrishna and J. Holtzman, "A Comparison between Single Code and Multiple Code Transmission Schemes in a CDMA System," *Proc. IEEE veh. Tech. Conf.*, vol. 2, 1998, pp. 791-5.

[114] T. Rappaport, *Wireless Communications: Principles and Practices*, Prentice Hall, 1996.

[115] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, 1944, pp. 292-332.

[116] ____, "Statistical properties of a sine wave plus random noise,"*Bell Syst. Tech. J.*, vol. 27, Jan. 1948, pp. 109-57.

[117] S. M. Ross, *Stochastic Processes*, 2nd ed., John Wiley & Sons, 1996.

[118] ____, *Introduction to Probability Models*, 7th ed., Academic Press, 2000.

[119] D. V. Sarwate and M. B. Pursley, "Partial Correlation Effects in Direct-Sequence Spread-Spectrum Multiple-Access Communication Systems," *IEEE Trans. Comm.*, vol. 32, May, 1984, pp. 567-73.

[120] ____, "Cross-correlation properties of pseudorandom and related sequences," *Proc. IEEE*, vol. 68, May 1980, pp. 593-618.

[121] D. V. Sarwate, "Bounds on crosscorrelation snd sutocorrelation of sequences," *IEEE Trans. Info. Thy.*, vol. 25, 1979, pp. 720-4.

[122] Hiroshi Sato, "On degraded Gaussian two-user channels," *IEEE Trans. Info. Thy.*, vol. 24, vol. 5, Sept. 1978, pp. 637-40.

[123] ____, "An outer bound on the capacity region of broadcst channels," *IEEE Trans. Info. Thy.*, vol. 24, vol. 3, May 1978, pp. 374-81.

[124] D. A. Schindman, "A generalized Nyquist criterion and optimum linear receiver for a pulse modulation system," *Bell Sys. Tech. J.*, July 1967, pp. 2163-77.

[125] P. Shamain, and L. B. Milstein, "Using higher order constellations with minimum mean square error receiver for severe multipath CDMA channel," *Proc. IEEE Intl. Conf. Comm.*, 1998, pp. 1035-9.

[126] ____, "Using Higher Order Constellations with Minimum Mean Square Error (MMSE) Receiver for Severe Multipath CDMA Channel," *Proc. IEEE Intl. Symp. Pers., Indoor & Mobile Radio Comm.*, vol. 3, 1998, pp. 1035-9.

[127] C. E. Shannon, "Communications in the presence of noise," *Proc. IRE*, vol. 37, Jan. 1949, pp. 10-21.

[128] ____, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, 1948.

[129] M. K. Simon et al., *Spread Spectrum Communications Handbook*, revised ed., McGraw-Hill, 1994.

[130] M. K. Simon, S. M. Hanedi and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, PTR Prentice Hall, 1995.

[131] M. K. Simon and M. Alouini, *Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis*, John Wiley & Sons, 2000.

[132] M. J. T. Smith and T. P. Barwell, "A procedure for designing exact reconstruction filter banks for tree-structured subband coders," *Proc. IEEE Intl. Conf. Acoust., Speech and Sig. Proc.* March 1984, pp. 27.1.1-4.

[133] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.

[134] G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, 1986.

[135] ____, *Linear Algebra*, Wellesley-Cambridge Press, 1993.

[136] N. Suehiro, "A signal design without co-channel interference for approximately synchronized CDMA systems," *IEEE J. Sel. Area. Comm.*, vol. 12, no. 5, 1994, pp. 837-41.

[137] H. Suzuki, "A statistical model for urban radio propagation," *IEEE Trans. Comm.*, vol. 25, Jul. 1977, pp. 673-80.

[138] X. H. Tang and P. Z. Fan, "A class of pseudonoise sequences over GF(P) with low correlation zone," *IEEE Trans. Info. Thy.*, vol. 47, no. 4, May 2001, pp. 1644-9.

[139] ____, "Bounds on aperiodic and odd correlations of spreading sequences with low and zero correlation zone ," *IEEE Elect. Lett.*, vol. 37, no. 19, Sept. 2001, pp. 1201-3.

[140] CATT, "TD-SCDMA Radio Transmission Technology for IMT-2000 candidate submission", *ITU Proposal for G3 RTT*, September, 1998.

[141] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Info. Theory*, vol. 45, Mar. 1999, pp. 641-57.

[142] ____, "Multiaccess fading channels– Part 1:polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Info. Thy.*, vol. 44, no. 7, Nov. 1998, pp. 2796-2815.

[143] G. Turin, "Introduction to Spread-Spectrum Antimultipath Techniques and Their Applications to Urban Digital Radio," *Proc. IEEE*, vol. 68, no. 3, March 1980, pp. 328-53.

[144] ____,"On optimal diversity reception," *IRE Trans. Info. Thy.*, vol. 7, Jul. 1961, pp. 311-329.

[145] ____,"On optimal diversity reception II," *IRE Trans. Comm. Syst.*, vol. 7, Mar. 1962, pp. 22-31.

[146] G. L. Turin et al.,"A statistical model of urban multipath propagation," *IEEE Trans. Veh. Tech.*, vol. 21, Feb. 1972, pp. 1-9.

[147] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proc. IEEE*, vol. 78, Jan.1990, pp. 56-93.

[148] ____, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[149] S. Verdu, *Multiuser Detection*, Cambridge U. Press, New York, 1998.

[150] ____, "Minimum Probability of Error For Asynchronous Gaussian Multiple Access Channels," *IEEE Trans. Info. Thy.*, vol. 32, no.1, Jan. 1986, pp. 85-96.

[151] ____, "Capacity regin of Gaussian CDMA channels: The symbol-synchronous case," *Proc. of the Twenty-Fourth Annual Allerton Conf. Comm., Control and Comp.*, vol. 32, Oct. 1986, pp. 1025-39.

[152] S. Verdu and S. Shamai, "Spectral Efficiency of CDMA with Random Spreading," *IEEE Trans. Info. Thy.*, vol. 45, no. 2, Mar. 1999, pp. 622-39.

[153] P. Viswanath, V. Anantharam, D. Tse, "Optimal Sequences, Power Control, and User Capacity of Synchronous CDMA Systems with Linear MMSE Multiuser Receivers," *IEEE Trans. Info. Thy.*, vol. 45, no. 6, Sept, 1999, pp. 1968-83.

[154] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communications*, Addison-Wesley, 1995.

[155] A. J. Viterbi and J. K. Omura, *Principles of Digital communication and Coding*, McGraw-Hill, 1979.

[156] B. Vojcic and W. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," *IEEE Trans. Comm.*, vol. 46, no. 10, Oct. 1998, pp. 1346-1355.

[157] T. Wada, T. Yamazato, M. Katayama, and A. Ogawa, "A Constant Amplitude Coding for Orthogonal Multi-code CDMA Systems," *IEICE Trans. Fund. Comm.*, vol. E80-A, no. 12, December 1997, pp. 2477-83.

[158] L. R. Welch, "Lower bounds on the maximum cross-correlation of signals," *IEEE Trans. Info. Thy.*, vol. 20, 1974, pp. 397-9.

[159] E. T. Whittaker, "On the functions which are represented by the expansion of interpolating theory," *Proc. Roy. Soc. Edinburgh,* vol. 35, 1915, pp. 181-94.

[160] J. M. Whittaker, "The Fourier theory of cardinal functions," *Proc. Math. Soc. Edinburgh*, vol. 1, 1929, pp. 169-76.

[161] S. G. Wilson, *Digital Modulation and Coding*, Prentice Hall, 1996.

[162] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965.

[163] G. Wu, A. Jalali and P. Mermelstein, "On Channel Model Parameters for Microcellular CDMA Systems," *IEEE Trans. Veh. Tech.*, vol. 44, no. 3, Aug. 1995, pp. 706-11.

[164] R. Yarlagadda and J. E. Hershey, "Analysis and synthesis of bent sequences," *IEE Proc.*, vol. 136, no. 2, Mar. 1989, pp. 112-23.

[165] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Info. Thy.*, Vol. 40 , Nov. 1994, pp. 1713-27.

[166] N. Yee, J-P. Linnartz and G. Fettweis, "Multicarrier CDMA in indoor wireless radio networks," *Proc. IEEE Intl. Symp. Pers., Indoor & Mobile Radio Comm.*, Yokohama, Japan, Sept. 1993, pp. 109-13.

[167] M. Zakai, "Band-limited functions and the sampling theorem," *Inform. Contr.*, vol. 8, 1965, pp. 143-58.

[168] R. Zhang, T. Tihung, H. Zhang and P. He, "BER Performance Comparison of Single Code and Multicode DS/CDMA Channelization Schemes for High Rate Data Transmission," *IEEE Comm. Lett.*, vol. 5, no. 2, Feb. 2001, pp. 67-9.

[169] N. Zierler, "Linear Recurring Sequences," *J. Soc. Indust. and Appl. Math.*, vol. 7, 1959, pp. 31-48.