Adaptive Turbo-Bit-Interleaved Coded Modulation for Wireless Channels

by

Francis Syms

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

Adaptive Turbo-Bit-Interleaved Coded Modulation for Wireless Channels

Francis Syms

A thesis submitted in conformity with the requirements for the Degree of Master of Applied Science, Graduate Department of Electrical and Computer Engineering, in the University of Toronto, 2001

Abstract

The use of bit-interleaved coded modulation (BICM) over Doppler fading channels has been analyzed. We show that bit-interleaving outperforms symbol interleaving in both slow and fast fading channels with one propagation path. In multipath, slow fading channels, which characterize third generation wireless environments, we show that bit-interleaving is again better than symbol interleaving when a RAKE receiver is employed and the number of resolved paths is small. We generate a simulation model for 3G propagation environments and show that inter-symbol interference (ISI) induces an error floor when a large number of propagation paths are used in RAKE reception. Furthermore, an adaptive channel scheme has been designed for use in third generation (3G) wireless systems. Our system, based on punctured turbo bit-interleaved coded modulation, provides spectral efficiencies between 2/3 and 16/5 bits/symbol/2D and operates at symbol signal-to-noise ratios as low as -1.4dB. Using small data block sizes required for 3G (5000 bits was chosen), our system tracks capacity to within 2dB at BER= 10^{-3} and 2.85dB at FER=10^{-2}.

Acknowledgements

I am deeply grateful to my supervisor, Prof. F. Kschischang, for his guidance and support throughout the course of my graduate studies. His numerous comments and careful reading of the manuscript are greatly appreciated.

Thank you, Patricia, for your never ending encouragement, love, and patience.

I would like also to thank my friends in the Communications Group, namely Sujit, Ivo, and Steve, for many interesting discussions and comments about my manuscript. I gratefully acknowledge the research assistantship provided by Prof. F. Kschischang.

Contents

1 Introduction **1**

1.1 Third Generation (3G) Wireless Communications 1

1.1.1 First and Second Generation Cellular Networks 1

1.1.2 CDMA and Third Generation Cellular Networks 4

1.1.3 3G Channel Considerations and Power Control 4

1.2 Bit-Interleaved Coded Modulation 6

1.3 This Work 8

2 Channel Models **10**

2.1 Wireless Channels 10

2.1.1 Clarke's Model for Flat Fading 15

3 Bit-Interleaved Coded Modulation **24**

3.1 BICM System Overview 24

3.1.1 Encoder 25

3.1.2 Decoder 28

3.1.3 Iterative Decoder (BICM-ID) 29

3.2 Performance 30

3.3 Capacity 33

3.4 BICM Complexity 34

4 3G Wireless Channels **39**

4.1 Single-Path (Frequency Non-Selective) Performance 39

4.2 Channel Model Definition 40

4.3 Spreading 42

4.4 Rake Receiver 43

4.5 Multipath (Frequency Selective) Performance 45

5 Adaptive 3G Channel Coding **49**

5.1 Service Class Constraints 50

5.2 LDD and Speech Class Operating Modes 52

5.2.1 AWGN Propagation Selection 52

5.2.2 Selection of Modes 52

5.3 Operating Mode Performance 54

5.4 Frame Size Constraints and Complexity 56

6 Conclusions **59**

A BICM Metric Derivation **62**

B Turbo Codes **64**

B.1 Iterative-Decoding 64

B.2 Interleaving 65

B.3 Puncturing 66

C Bit-Error-Rate (BER) Curves **69**

D Decoding Complexity **73**

D.1 Viterbi Decoding 74

D.2 Turbo Coding 75

D.3 Branch Metric Computation 76

Acronyms

1G	First Generation Wireless Systems	Page 1
2G	Second Generation Wireless Systems	Page 1
AMPS	Advanced Mobile Phone System	Page 1
FDMA	Frequency-Division Multiple Access	Page 2
TDMA	Time-Division Multiple Access	Page 2
CDMA	Code-Division Multiple Access	Page 2
WWII	World War Two	Page 2
DS-SS	Direct Sequence Spread Spectrum	Page 3
FH-SS	Frequency Hopping Sequence Spread Spectrum	Page 3
PN	Pseudo-Random Noise	Page 3
W-CDMA	Wideband Code-Division Multiple Access	Page 4
TDD	Time Division Duplex	Page 5
FDD	Frequency Division Duplex	Page 4
ISI	Inter-Symbol Interference	Page 5
SNR	Signal-to-Noise Ratio	Page 5
TCM	Trellis Coded Modulation	Page 6
AWGN	Additive White Gaussian Noise	Page 7
BICM	Bit-Interleaved Coded Modulation	Page 7
IID	Independent Identically Distributed	Page 8
BICM-ID	Bit-Interleaved Coded Modulation with Iterative Decoding	Page 8
RMS	Root Mean Squared	Page 14
PSD	Power Spectral Density	Page 17

FFT	Fast Fourier Transform	Page 19
IFFT	Inverse Fast Fourier Transform	Page 20
IDFT	Inverse Discrete Fourier Transform	Page 21
PSK	Phase Shift Keying	Page 24
SP	Set Partitioning	Page 27
MSB	Most Significant Bit	Page 27
LSB	Least Significant Bit	Page 27
ML	Maximum Likelihood	Page 28
CSI	Channel State Information	Page 28
QPSK	Quadrature Phase Shift Keying	Page 34
QAM	Quadrature Amplitude Modulation	Page 37
BER	Bit Error Rate	Page 39
TDL	Tapped Delay Line Model	Page 42
EGC	Equal Gain Combining	Page 45
MRC	Maximum Ratio Combining	Page 45
LDD	Low Data Delay	Page 49
QOS	Quality of Service	Page 49
CRC	Cyclic Redundancy Check	Page 51
FER	Frame Error Rate	Page 51
SICM	Symbol Interleaved Coded Modulation	Page 59
RSCC	Recursive Systematic Convolutional Code	Page 65
FBA	Forward Backward Algorithm	Page 76

Chapter 1

Introduction

1.1 Third Generation (3G) Wireless Communications

In the year 2000, worldwide mobile phone sales totalled 412.7 million, a 45.5 percent increase from 1999 sales. By 2005, it is estimated that there will be more than half a million basestations in Europe alone [1]. No longer do these networks service only voice traffic. Users want to be able to use their phone to surf the internet - checking things such as email, stock quotes, and the weather. Third generation (3G) wireless systems are being developed to meet this challenge.

1.1.1 First and Second Generation Cellular Networks

The first cellular networks were rooted in the analog domain. These first generation (1G) networks were based upon the Advanced Mobile Phone System (AMPS) [2] . Each user was assigned a bandwidth of 30kHz in which their voice signal was modulated and transmitted using analog frequency modulation. The second generation (2G) networks, were digital in nature. These networks, which are primarily still in use today, provided a ten-fold increase in capacity over the first generation analog

systems [3]. The range of 2G systems can be classified according to three main standards: IS-54, GSM, and IS-95. These networks are differentiated by the way they handle multiple users. The first two standards, IS-54 and GSM, are descendants of AMPS and are used in North America and Europe respectively. These systems use a combination of frequency-division multiple access (FDMA) and time-division multiple access (TDMA). FDMA separates users by assigning each a disjoint frequency band while TDMA assigns each a disjoint transmission time interval. The third standard, IS-95, uses a multiple access technique known code-division multiple access (CDMA). It is this standard which has stood the test of time and has formed the basis for the global third generation (3G) networks. In CDMA, all users share the same time and frequency intervals. Each user is separated by use of unique code. The time and frequency allocations of the three multiple access schemes are illustrated in Figure 1.1.

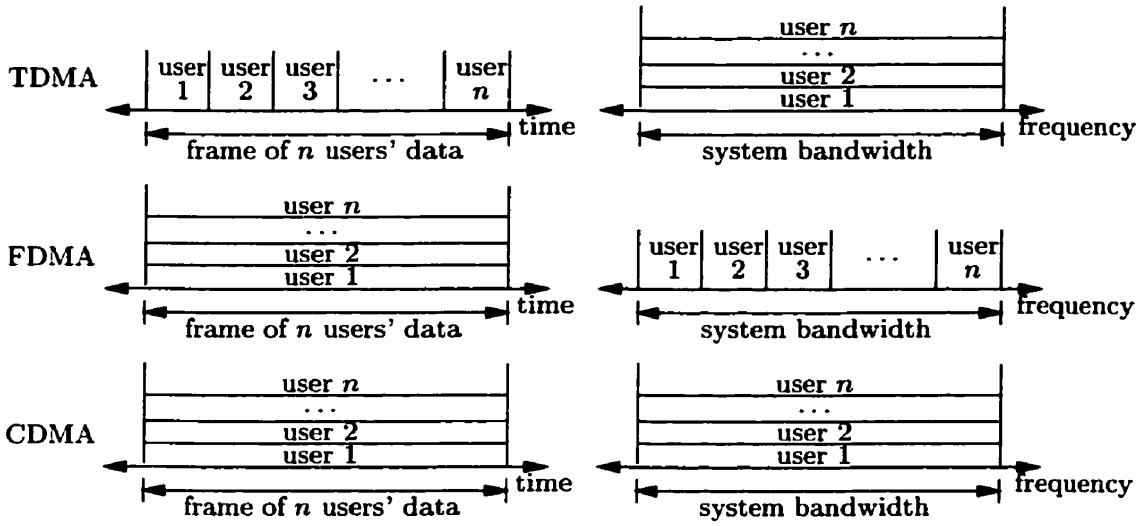


Figure 1.1: Frequency and Time Allocation for Multiple Access Schemes

Spread Spectrum

First used by the military in WWII , spread spectrum modulation [4, 5] is a technique whereby the data signal is spread out in frequency (before transmission) to a

bandwidth which is much larger than the signal data rate. This technique has an inherent advantage over other communication strategies - it provides excellent interference rejection. This ability is due to the way the signal is spread out in frequency. Of the many different type of spread spectrum, the two important ones are direct sequence (DS) and frequency hopping (FH) . In frequency hopping, the data signal is narrowband modulated but its carrier frequency is changed frequently (within a fixed band). In direct-sequence (DS) spread spectrum, the data signal is multiplied by a high rate pseudo-random noise (PN) sequence as shown in Figure 1.2. On receipt of the message the receiver de-spreads it to its original form by multiplying the incoming signal by the same PN sequence. Formed by shift registers with a large number of states, these sequences have the property that they appear random to observers without a-priori knowledge of the sequence. Any narrowband interference in the channel is effectively spread out in frequency by the de-spreader in the receiver thereby minimizing its effect on the data signal. Direct sequence spread spectrum is the technique most often used in CDMA systems.

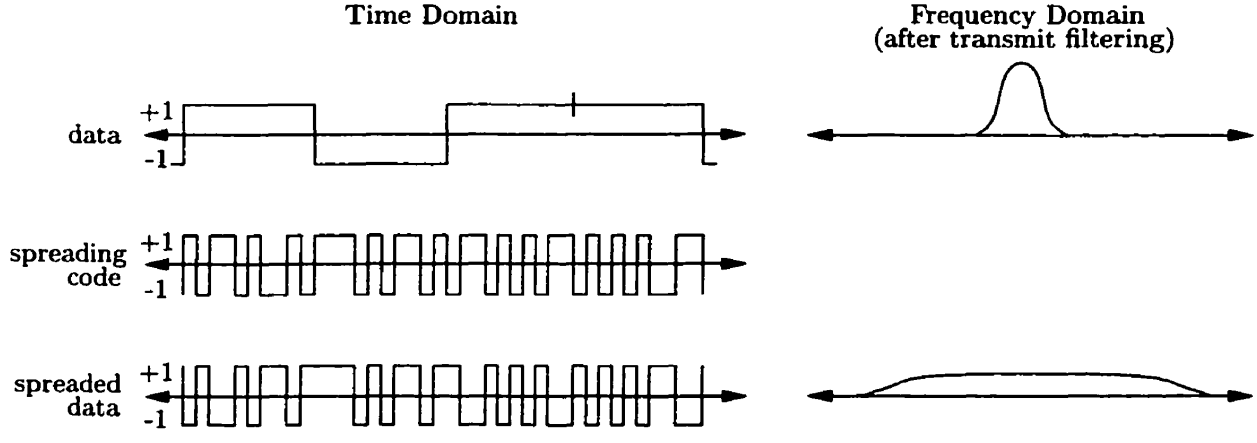


Figure 1.2: Direct Sequence Spread Spectrum

1.1.2 CDMA and Third Generation Cellular Networks

CDMA was first proposed for cellular systems in 1978 [6] and was adopted for use in IS-95. The telecommunications industry has recognized the merits of spread spectrum and all proposed third generation standards specify the use of CDMA. The need for a new standard is clear. The 2G networks with data rates of less than 100kbps cannot support the data rates required for internet traffic [7]. Users want the flexibility of global travel without having to worry about their terminal not being supported by the local network; hence, a global standard is required. The first goal is realistic, the second overly ambitious. In the end, two main global 3G CDMA standards were developed: CDMA2000 and W-CDMA [8]. The main reason for the development of two unique standards was backwards-compatibility. In North America, much IS-95 CDMA equipment is already in place. CDMA2000 was developed such that some of this legacy equipment could be used. European and Japanese systems are not restricted to backwards-compatibility since their 2G system is TDMA/FDMA based. Their version of the 3G standard, known as wideband CDMA (W-CDMA) has many similarities to the North American standard. The main difference lies in the choice of chip rate. W-CDMA uses a rate of 3.098Mchips/s whereas CDMA2000 uses rates which are multiples of 1.2288Mchips/s. Both however, aim to provide the same rates. As a minimum, two different data rates are to be supported: 144kbps (with terminal speeds of up to 120km/h) and 2Mbps (with fixed terminal location). Also frame sizes for data transmission are on the order of 10-20ms.

1.1.3 3G Channel Considerations and Power Control

Third generation cellular networks are comprised of both downlink (forward) and uplink (reverse) channels. On the downlink channels, the basestation is the transmitter and the mobile users are the receivers. On uplink channels, the roles are reversed. These channels are either located in disjoint frequency bands, known as frequency division duplex (FDD) communication, or in disjoint time intervals, known as time

division duplex (TDD) communication. In the uplink, each mobile is assigned a different PN spreading (scrambling) sequence and communicates with the basestation asynchronously (with respect to other mobiles). In the downlink of the 3G system, transmission to all users is synchronous. Furthermore, each user's data signal is made orthogonal to the others by assigning each an orthogonal spreading code. Upon reception, however, the data signals are no longer orthogonal. In communication over wireless propagation environments, receivers tend to see multiple, delayed copies of the transmitted signal. As a result, the orthogonality property of the signals is destroyed. This causes severe problems (interference) for a user who is trying to decode their intended received signal. There are several techniques which are used to mitigate this effect. One way is through equalization. Each of these delayed copies (transmit paths) contain valuable information about the data being transmitted. If received properly, the channel can be considered to exhibit time diversity. The Rake receiver, introduced by Price and Green [9] in 1958 exploits this effect. It is a technique used at the receiver to linearly equalize ISI in a multipath propagation environment. The basic operation of the Rake receiver is as follows. A receiver will contain a set number of 'fingers'. The purpose of these fingers is to coherently receive (via a matched filter) the paths with the best signal power. Each path is weighted and linearly combined to produce the output signal, as shown in Figure 1.3. The second way this interference problem is mitigated is through the use of power control. Each terminal or base station adjusts its power, while maintaining the required performance level, such that its interference-effect on other receivers is minimized. This is done by attempting to maintain a constant signal-to-noise (SNR) at every receiver. In wireless channels, keeping SNR constant is not easy. Fading causes the constant transmitted signal amplitude to vary over time, thereby varying the SNR. The frequency of the fading in a 3G propagation environment is on the order of 200Hz when moving at a speed of 120km/h and transmitting at a carrier frequency of 2GHz. While power control information is sent to the receiver at approximately 2kbps in 3G, there is usually a

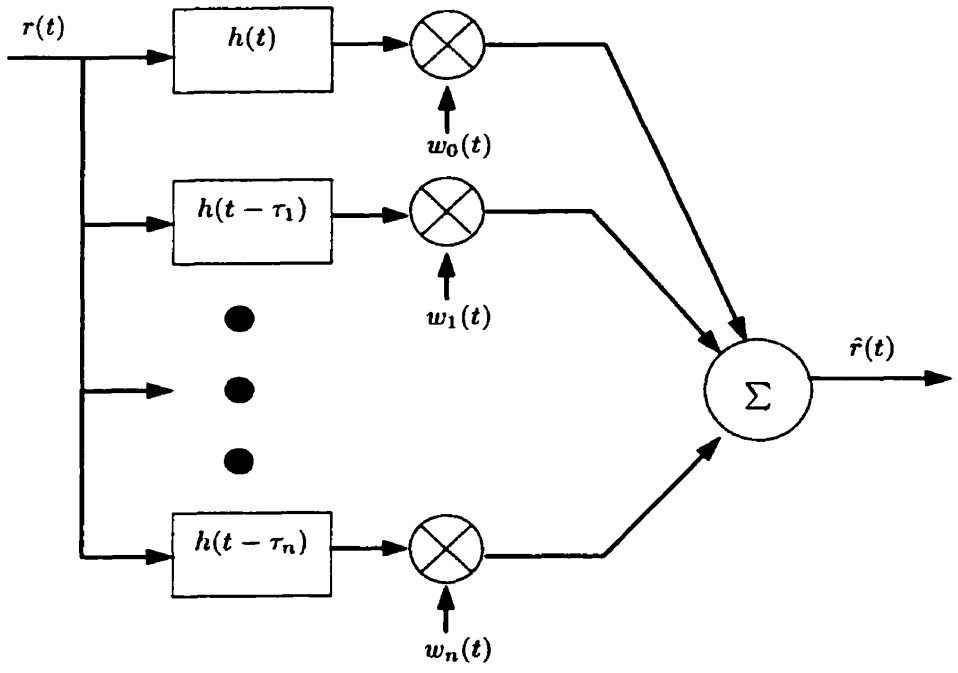


Figure 1.3: RAKE Receiver with n-fingers where $h(t-\tau_i)$ is the matched filter tuned to each path and $w_i(t)$ is scaling factor for each path

1-2 frame lag before the actual power is changed. As a result there is a range about the average which the SNR will vary over. Known as the fading margin, it is an important consideration when determining the target (average) SNR. System performance must not be compromised by these variations. As a result the channel coding used must be both robust in this type of environment and, at the same time, provide high spectral efficiency given a small frame size. Our work focuses on channel coding techniques for the downlink which work well under these system and environmental constraints. Coded modulation, and in particular bit-interleaved coded modulation, is able to satisfies these constraints.

1.2 Bit-Interleaved Coded Modulation

The invention of trellis coded modulation(TCM) [10] and multilevel coding [11] in the 1970's merged the principles of modulation and coding together into a single entity

which was appropriately called "coded modulation". These techniques provided significant performance improvements in the bandwidth-limited regime of the additive white Gaussian noise (AWGN) channel. Unfortunately, direct application of these ideas to wireless fading channels does not result in the same performance improvements [12]. Fading effectively causes errors which are bursty in time. Attempts were made to design new codes which employed a high degree of time diversity [12]. The objective of these codes is to interleave the symbols at a depth exceeding the coherence time of the fading process. Zehavi [13] embraced this idea and proceeded to take it one step further. In his approach, he recognized that performance could be further improved by making the diversity equal to the number of bits rather than the number of symbols along any error event. This diversity was achieved by bit-wise interleaving at the encoder output. In a later publication, Caire *et al.* [14] named this technique *Bit-Interleaved Coded Modulation* (BICM). A block diagram illustrating BICM is presented in Figure 1.4. In [14], a detailed theoretical analysis of the BICM system was presented and it was proved that in terms of capacity, BICM is sub-optimal to TCM.

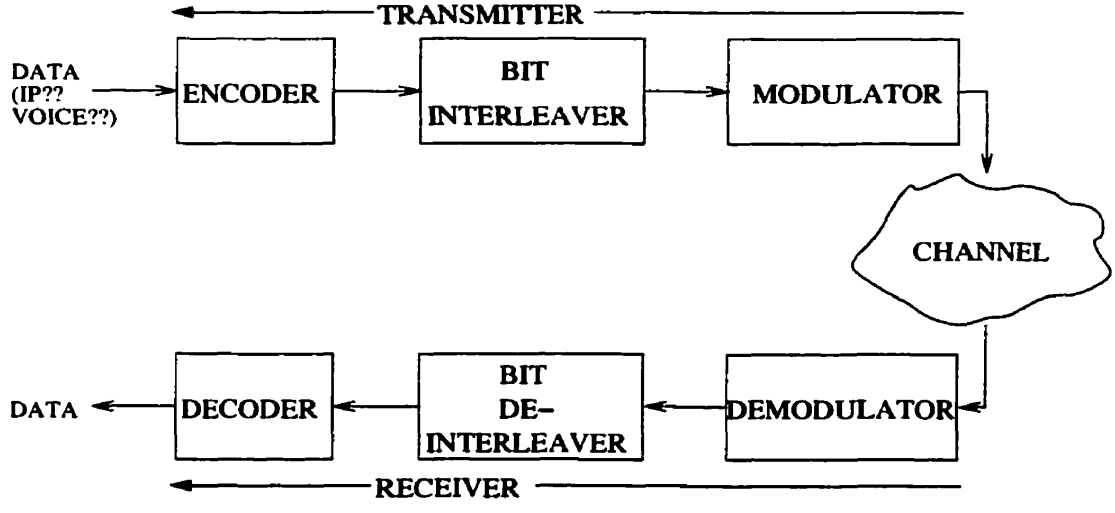


Figure 1.4: Bit-Interleaved Coded Modulation (BICM)

Despite its sub-optimality, BICM performs well in a fading channel. At a bit error

rate of $P_b = 10^{-4}$, BICM has been shown to provide a SNR improvement of approximately 0.5dB when compared with symbol interleaved TCM [13] and approximately 1.2dB when compared to conventional multi-level coded modulation [15] over IID Rayleigh channels. On AWGN channels, however, the performance using the same codes is at least 0.5 to 1dB worse than TCM and multi-level coding. This may be explained by the fact that the bit-interleaving causes an inherent "random modulation" [13] in the system which significantly reduces the free Euclidean distance. Li and Ritcey [16] proposed a scheme which combats this effect. Their scheme, called Bit-Interleaved Coded Modulation with Iterative Decoding (BICM-ID), utilizes iterative decoding in combination with hard-decision [16] or soft decision [17] feedback. They have shown that these techniques can provide up to 1dB improvement over BICM in both AWGN and IID-Rayleigh environments.

1.3 This Work

The robustness of BICM to IID fading suggests that it may work well in a 3G propagation environment. Furthermore, turbo codes in combination with BICM may provide the physical layer groundwork for a system which is not only robust to fading and multipath but performs well when compared to Shannon's capacity. This idea is the underlying thread which binds together the work presented in this thesis. Specifically, this thesis has several objectives:

1. To generate and analyze results for BICM over AWGN and IID Rayleigh Channels
    - Generate capacity curves for BICM over AWGN and IID Rayleigh channels for various modulation schemes using different constellation mapping techniques.
    - Analyze the decoding complexity of BICM.

2. To generate and analyze results for BICM over a propagation environment which exhibits Doppler fading and multipath propagation.

   - Analyze the 3G specs and develop a propagation model which can easily be implemented.
   - Build an environment for simulation of Doppler and multipath fading.
   - Generate and analyze results for BICM over one path slow and fast fading propagation environments.
   - Generate and analyze results for BICM over one path 3G slow fading propagation environment.
   - Generate and analyze results for BICM over a realistic 3G slow fading, frequency selective propagation environment.

3. To design an adaptive channel coding system for 3G wireless communications using turbo codes.

   - Design the adaptive rate strategy.
   - Simulate a chosen turbo code in combination with BICM for various puncture masks and modulation schemes.
   - Select optimal configurations for use in the system and analyze performance.
   - Analyze the decoding complexity of Turbo Codes in combination with BICM.

This thesis begins, in Chapter 2, with an outline of the channels used. Chapter 3 introduces and analyzes BICM. The performance of BICM over wireless channels is presented in Chapter 4 followed by the proposed adaptive BICM system in Chapter 5. The thesis is summarized and future work is discussed in Chapter 6.

Chapter 2

Channel Models

When evaluating the performance of any proposed system, it is important to model the system's environment as closely as possible. In this work, we examined performance in a number of channel environments. This chapter will describe these channel models. They can be categorized into four types:

1. Additive White Gaussian Noise (AWGN) Channel
2. Independent Identically Distributed (IID) Rayleigh Fading Channel
3. Correlated Rayleigh Fading Channel (Flat Fading Channel)
4. Multipath, Correlated Rayleigh Channel (Frequency Selective Fading Channel)

The AWGN channel is well known [18, 19, 20] and will not be explained any further here. The last three items are an intrinsic part of the wireless channel. They are ordered by increasing complexity and will be described in Section 2.1.

2.1 Wireless Channels

When transmitting a signal over a wireless channel, the receiver will almost always see multiple copies of the signal. These copies, varying in amplitude, phase, and possibly

frequency, will combine in a constructive or destructive manner at the receiver. They are due to reflections from the ground or man-made structures such as buildings. It is the combination of these received waves at the receiver which results in small-scale fading. Small-scale fading, often known simply as fading, will be the focus of this section. There is another type of fading, known as large-scale fading, which deals with the transmission of signals over large distances. Large-scale fading is typically characterized by two effects: (i) free space propagation loss which is modelled as the inverse m^{th} power of distance and (ii) shadowing due to large structures such as mountains which is modelled as the log normal distribution. For more information on this type of fading see Rappaport [18].

Small-scale fading variations can be modelled by assuming that the transmitted signal passes through a linear filter with a time varying impulse response. The impulse response, $h(t,\tau)$, completely characterizes the channel and is a function of both time variations due to motion, t, and channel multipath delay, τ, for a fixed instance in time. By convolving the transmitted signal, $x(t)$, with the impulse response, $h(t,\tau)$, the signal which is present at the receiver, $y(t)$, can be determined. This received signal, $y(t)$, can be written as

$$y(t) = x(t) \star h(t,\tau) + \eta(t) = \int_{-\infty}^{\infty} x(\tau)h(t,\tau)d\tau + \eta(t) \tag{2.1}$$

where $\star$ is the convolution symbol and $\eta(t)$ is AWGN with two sided power spectral density $N_0/2$.

By assuming that the channel is band-limited and band-pass (as suggested by Rappaport [18]), the impulse response, $h(t,\tau)$, can be represented in terms of its complex baseband (or low-pass) impulse response, $h_b(t,\tau)$.

$$h(t,\tau) = Re\left\{h_b(t,\tau)e^{j2\pi f_c t}\right\}, \tag{2.2}$$

where f_c is the carrier frequency in Hertz (Hz).

Along any one propagation path from the transmitter to the receiver, the signal will be scaled by a complex fading envelope which varies in time and frequency. Modelled as a single impulse, $\delta(t - \tau(t))$, the low-pass response can be written as

$$h_b^{1-path}(t, \tau) = \alpha(t, \tau) e^{-j2\pi f_c \tau(t)}, \tag{2.3}$$

where $\alpha(t, \tau)$ and $\tau(t)$ are the amplitude and delay respectively.

In reality, there is a continuous spectrum of multipath components. Receivers, however, have limited resolution. For this reason, the channel is usually modelled as having a discrete number of multipath components. Each component is considered to be located in one of N delay bins along the impulse response delay axis. Figure 2.1 provides an illustration of this type of channel.

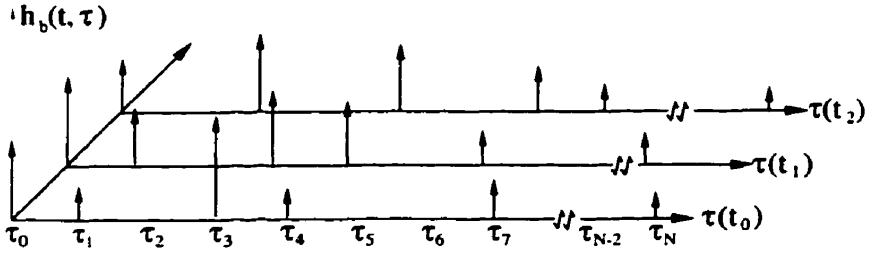


Figure 2.1: An example of the time varying discrete-time impulse response for the multipath radio channel

The channel response based on all resolvable multipath components is given by

$$h_b(t, \tau) = \sum_{i=0}^{N-1} \alpha_i(t, \tau_i) e^{-j2\pi f_c \tau_i(t)} \delta(\tau - \tau_i(t)) \tag{2.4}$$

and the resulting low-pass channel output, $r(t)$ is

$$r(t) = \sum_{i=0}^{N-1} \alpha_i(t, \tau_i) e^{-j2\pi f_c \tau_i(t)} c(t - \tau_i(t)) + \eta(t) \tag{2.5}$$

Equation 2.5 analytically characterizes the channel; however, for clarity in dis-

cussion, it is best to split the model into several categories. The channel impulse response causes the transmitted signal to be spread out in both time and frequency. The frequency spreading, known as the Doppler spread, is generally attributed to motion by either the transmitter, receiver, or objects in the environment. The time spreading, known as the delay spread, is a result of multipath generated by reflections of the transmitted signal with the environment. Depending on how the received signal varies with respect to these channel variations, the channel can be classified as one of four types:

- Frequency-Selective Fast Fading
- Frequency-Selective Slow Fading
- Flat Fast Fading
- Flat Slow Fading

A brief description of these four types is presented next. An illustration of their relationships is presented in Figure 2.2. For a more detailed description, see [18, 19, 20].

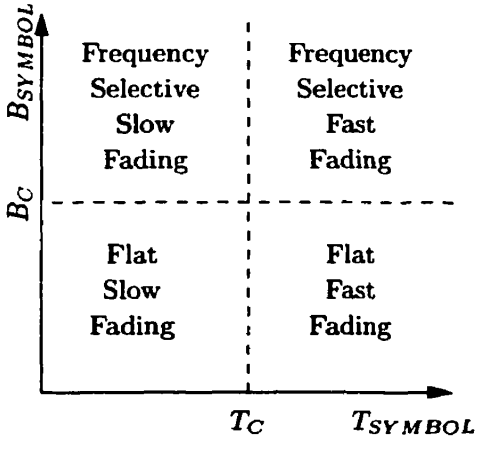


Figure 2.2: Fading types as a function of signal bandwidth and symbol period where T_C and B_C represent the channel coherence time and coherence bandwidth respectively

Frequency-Selective Versus Flat Fading

If the channel induces no observable multipath components, the channel is considered to be flat. This happens if all spectral components for a given signal see the same channel characteristics (equal gain and linear phase). The rms delay spread is defined as the standard deviation of the distribution of delays for a given received multipath signal. It is a good indication of how much ISI will be induced by the channel. The range of frequencies over which the channel is considered to be constant is known as the coherence bandwidth (B_C). If the bandwidth of the signal is larger than the coherence bandwidth, frequency-selective fading will occur. Different portions of the signal will be affected by the channel in different ways - inducing multipath effects.

Fast versus Slow Fading

Fast or slow fading is a measure of the rate at which the channel impulse response varies with respect to the signal. The frequency of this variation is known as the Doppler spread. If the Doppler spread is much less than the signal bandwidth, we have slow fading. Otherwise, it is considered to be fast fading. Coherence bandwidth was used to describe the range of frequencies over which the channel response was invariant. A similar parameter is used here. The coherence time is defined as the time interval over the channel response can considered constant.

Our work will examine the behavior of BICM over these four types of channels. The flat fading channel (fast and slow) was implemented using Clarke's model [21] and is discussed in Section 2.1.1 and illustrated in Figure 2.3. The frequency-selective channel was generated using the tapped-delay line model [20]. This model, illustrated in Figure 2.4, generates the multipath effect by delaying, attenuating and combining the transmitted signal. In the figure, L refers to the total number of resolvable paths. Each branch is attenuated by a flat fading waveform, $\alpha(t)$. The fading waveform, $\alpha(t)$, is generally assumed to have a Rayleigh distribution [18]. The reasoning behind this is as follows. The received signal is composed of a transmitted signal which

has spread out in time. As stated above, the receiver resolves the received energy into one or more paths. Each resolved path may be composed of multiple copies of the transmitted signal, each slightly different in amplitude or phase. Using the Central Limit Theorem [20], as the number of copies increases, the received signal will assume a Gaussian distribution. The fading waveform is the envelope of the sum of two quadrature Gaussian noise sources which is, by definition, a random variable with a Rayleigh distribution. If there is motion by the receiver or transmitter, $\alpha(t)$ is generated using the Clarke model. If there is no motion whatsoever, $\alpha(t)$ is generated using independent, identically, distributed(IID) Rayleigh random variables.

Both flat and frequency-selective channels can experience either slow or fast fading. A convenient way to indicate the "speed" at which a channel is fading is by using the normalized ratio $2f_m/f_s = 2f_mT_s$. The maximum Doppler shift, given in Hz, is represented by f_m and is equivalent to the velocity of the receiver relative to the transmitter divided by the wavelength of the carrier frequency (v/λ). The bandwidth and symbol period of the signal is represented by f_s and T_s respectively. Figure 2.5 illustrate both slow and fast fading envelopes. Slow fading is generated using a normalized ratio of 1/128 and fast fading was generated using 1/8. Both envelopes were generated using the Clarke model.

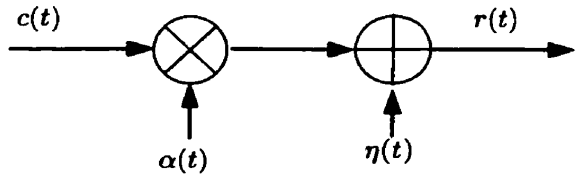


Figure 2.3: Flat fading channel model

2.1.1 Clarke's Model for Flat Fading

Clarke [21] stated that the electric field which is incident on the mobile is made up of N randomly distributed plane waves (polarized in the same direction as the antenna). Each wave has both a random arrival angle and a random phase. The electric field

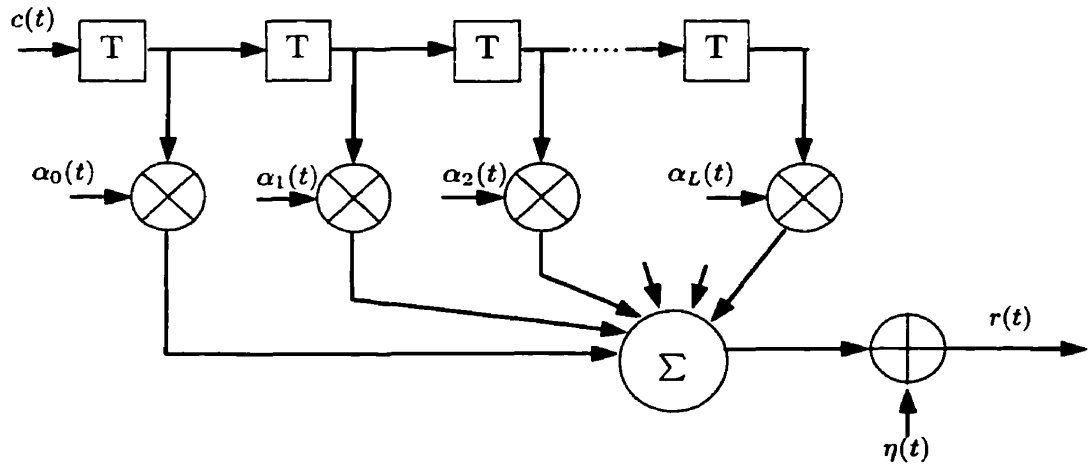


Figure 2.4: Tapped-delay line model for the frequency-selective fading channel

of this polarized plane wave is

$$E_Z(t) = T_C(t)\cos(2\pi f_C t) - T_S(t)\sin(2\pi f_C t) \tag{2.6}$$

The terms $T_C(t)$ and $T_S(t)$ are given by

$$T_C(t) = \sum_{n=0}^{\infty} \alpha_n \cos\left(2\pi f_m t \cos(\theta_n) + \phi_n\right) \tag{2.7}$$

$$T_S(t) = \sum_{n=0}^{\infty} \alpha_n \sin\left(2\pi f_m t \cos(\theta_n) + \phi_n\right) \tag{2.8}$$

where α_n, θ_n, and ϕ_n are the amplitude, arrival angle, and initial phase of the n^{th} arriving wave respectively. Under the assumption of an isotropic receiving antenna, the arrival angle, θ_n, is a random variable which is uniformly distributed over $[-\pi, \pi)$. Both $T_C(t)$ and $T_S(t)$ are considered to be uncorrelated, wide-sense stationary random variables with a zero-mean Gaussian distribution. This can be said given that there are a large number of scattered waves present. An illustration of this model is presented in Figure 2.6.

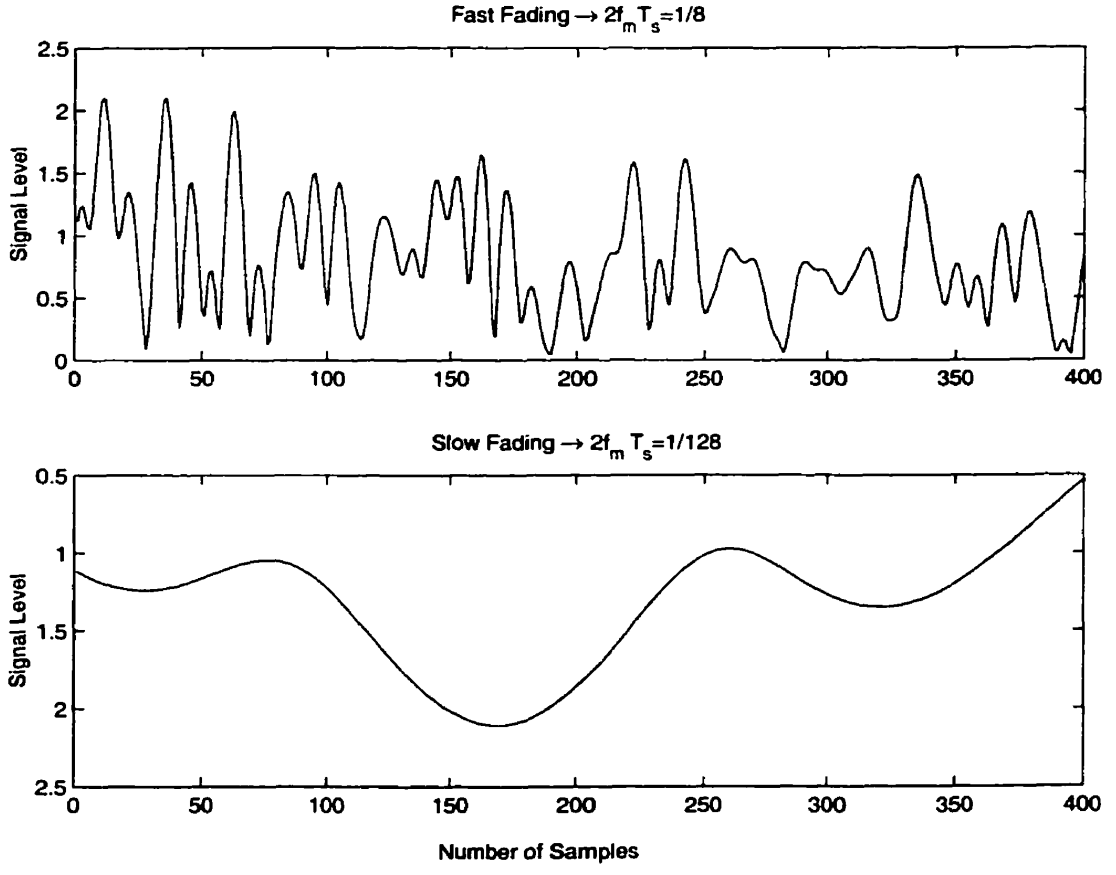


Figure 2.5: Fading envelopes for slow and fast fading generated via the Clarke model

The instantaneous signal envelope of the E-field is given by

$$|E_Z(t)| = \sqrt{T_C^2(t) + T_S^2(t)} = \alpha(t) \tag{2.9}$$

The envelope, $\alpha(t)$, is Rayleigh distributed and is the fading waveform which combines with the transmitted signal in the channel. Gans [22] showed that the received E-field wave has the power spectral density

$$S_{E_Z}(f) = \frac{1.5}{\pi f_m \sqrt{1 - \left(\frac{f - f_c}{f_m}\right)^2}}. \tag{2.10}$$

The power spectral density(PSD) is plotted in Figure 2.7. This PSD can be seen as

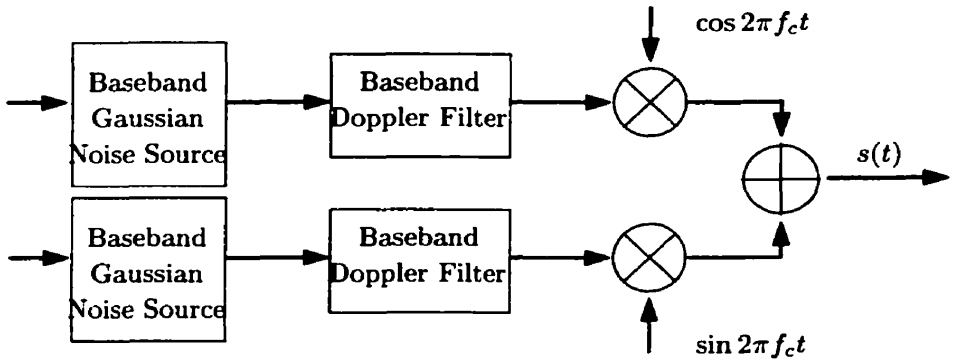


Figure 2.6: Clarke's model

the frequency response of a linear time invariant filter $H_c(f) = \sqrt{S(f)}$. Smith [23] used this filter to come up with a procedure for implementing the Clarke model. This is the approach taken in this work and will be discussed now.

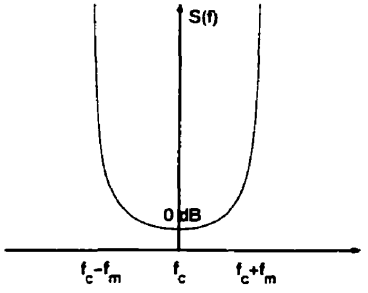


Figure 2.7: Doppler power spectrum of a received signal in a fading channel

In order to generate the Clarke model, one needs to create the complex Gaussian random vector $T_C(t) + jT_S(t)$. By generating Gaussian random variables and passing them through the filter $h_c(t)$(the inverse discrete Fourier transform of $H_c(f)$), discrete values of this vector can be created. One problem, however, is that the filter is difficult to implement because its is time-unlimited and has infinite amplitude in the frequency domain. For this reason, approximations must be made. In Smith's implementation, the discrete representations of $T_C(t)$ and $T_S(t)$ are first created in the frequency domain using an approximated, filter frequency response, $H_c(f)$. The representations are then converted back to the time domain where they are combined to form the appropriate fading envelope, $\alpha(t)$. This process is illustrated in Figure 2.8. The exact

procedure for generating N samples of the envelope is $\alpha(t)$ is as follows:

1 Determine the values for the following parameters:

- Carrier frequency (f_c) - e.g. 2GHz
- Doppler frequency (f_m) - e.g. for velocity= 120km/h, $f_m = \frac{vf_c}{c} = 222.22Hz$ where c is the speed of light
- Signal bandwidth (BW_s) - e.g. 4.096MHz

2 Generate N samples of complex white Gaussian noise in the frequency domain for both the in-phase ($\beta_I[f]$) and quadrature ($\beta_Q[f]$) components of the model under the following constraints. The variables $T_C(t)$ and $T_S(t)$ are real valued. This means that their frequency counterparts must exhibit Hermitian symmetry. Also, given that the IDFT is

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j\frac{2\pi kn}{N}}$$

we can see $X[k]$ must be real valued when $k = 0$ and $k = N/2$ in order for $x[n]$ to be real. The generated noise samples are effectively scaled representations of these frequency counterparts and therefore must have the same properties. Also, it is recommended that N be a power of 2 such that FFT algorithm can be used.

3 Generate S samples of the Doppler filter, $H_c[f]$, where S is given by

$$S = \frac{f_m}{f_c}N$$

According to Mahvidi [24], for accurate representation, the number of samples, S, should be greater than 128. The samples are generated using (2.10) with f_c set to zero. They should be evenly distributed between $[-f_m, +f_m)$. It is not necessary to generate a sample for $+f_m$ since the

IFFT will assume periodicity and infer a value for this point [25]. At $-f_m$, the amplitude of the frequency response is infinite. Since this is unrealizable, an approximation for this sample must be made. A value may be chosen by interpolating a value based on the two samples adjacent to the sample at point $-f_m$.

4 Compute the component-wise multiplication of the noise samples, $\beta_I[f]$ and $\beta_Q[f]$ generated in Step 2 with the filter samples, $H_c[f]$, generated in Step 3.

5 Perform an IFFT on each of the two results in the previous step. The N-point outputs of the IFFT are the sampled in-phase and quadrature components of the vector $T_C(t) + jT_S(t)$.

6 Determine the sampled fading envelope, $\alpha[t]$, by computing the magnitude of the sampled complex vector $T_C(t) + jT_S(t)$. This is accomplished by calculating the magnitude squared of $T_C[t]$ and $T_S[t]$ generated in Step 5, adding them together and taking the square root of the result.

7 Normalize the envelope such that the root mean squared (rms) value is unity. This is the final result which may be used to represent the fading envelope, α, in a simulation.

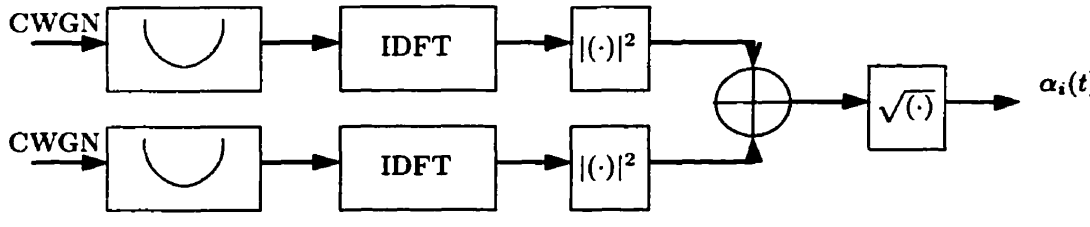


Figure 2.8: Implementation of Clarke's model

IDFT Optimization

One problem with this implementation is that it is computationally expensive. The effort required primarily has to do with performing the N-point IDFT. An FFT algorithm is typically used to calculate the IDFT which requires roughly $N \log_2 N$ operations. When generating fading envelopes for 3G simulations, signal bandwidths on the order of 4MHz (3.84Mcps) are used with Doppler filter bandwidths of approximately 2Hz (3kM/h) 450Hz (120km/h). In order to achieve good resolution on the filter, an IDFT would have to be executed on sample sizes on the order of 10^6 points or 20 bits (translating to 10^6 I and Q filtered noise samples each). The FFT would require approximately 20 million operations. This can prove to be very difficult to do in real time. Furthermore, the number of noise samples which must be pre-generated is large (2 million). While these may not be a large problem in software, a more efficient way would be needed to implement such a model in hardware (20 bit FFTs are unrealistic). By modifying the implementation, the size of both the IDFT which must be performed and the data which must be pre-generated can be reduced significantly. The general formula for the IDFT is as follows:

$$\tilde{x}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}[k] \exp\left(j \frac{2\pi}{N} kn\right) \tag{2.11}$$

There are two distinct properties exhibited by our frequency spectrum which can be used to optimize the IDFT for our application. They are as follows:

1. The spectrum appears as if it has been zero padded
2. The spectrum exhibits Hermitian symmetry $X[f] = X^*[-f]$

The first property lets us redefine the IFFT such that the we sum over the non-zero elements only

$$\tilde{x}[n] = \frac{1}{N} \sum_{k=i}^{L+i} \tilde{X}[k] \exp\left(j \frac{2\pi}{N} kn\right) \tag{2.12}$$

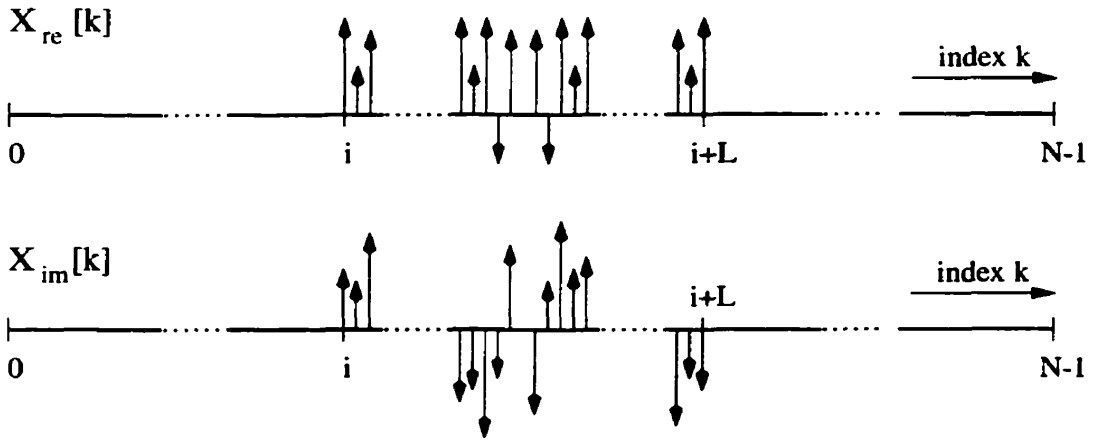


Figure 2.9: Example Frequency Spectrum

The second property tells us that the time domain signal will be real valued. Using Euler's equation($e^{j\omega} = \cos\omega + j\sin\omega$), (2.12) can be rewritten as:

$$\begin{aligned}
\tilde{x}[n] &= \frac{1}{N}\sum_{k=i}^{L+i}\left\{\left(\tilde{X}_{re}[k] + j\tilde{X}_{im}[k]\right)\left(\cos\left[j\frac{2\pi}{N}kn\right] + j\sin\left[j\frac{2\pi}{N}kn\right]\right)\right\} \\
\tilde{x}[n] &= \frac{1}{N}\sum_{k=i}^{L+i}\left\{\left(\tilde{X}_{re}[k]\cos\left[j\frac{2\pi}{N}kn\right] - \tilde{X}_{im}[k]\sin\left[j\frac{2\pi}{N}kn\right]\right)\right. \\
&\quad \left. + j\left(\tilde{X}_{re}[k]\sin\left[j\frac{2\pi}{N}kn\right] + \tilde{X}_{im}[k]\cos\left[j\frac{2\pi}{N}kn\right]\right)\right\}
\end{aligned}$$

Since the result must be real-valued, the imaginary terms will cancel out and the following equation will remain:

$$\tilde{x}[n] = \frac{1}{N}\sum_{k=i}^{L+i}\left\{\tilde{X}_{re}[k]\cos\left[j\frac{2\pi}{N}kn\right] - \tilde{X}_{im}[k]\sin\left[j\frac{2\pi}{N}kn\right]\right\} \qquad (2.13)$$

To summarize, the number of operations required to compute an DFT/IDFT is roughly $N\log_2 N$ for the FFT implementation. Using the same measure, the new modified technique will be roughly ML where M is the number of points required to represent the fading envelope. Obviously, the overall number of computations required to generate a large fading envelope is still less for the FFT (20×2^{20} operations for

20 bit IDFT using FFT as opposed to 2^{40} for the modified method). The advantage of the method is that the a fading envelope of any desired sample size can be easily produced. This leads to less data storage at any one time, smaller IDFT, and faster computation.

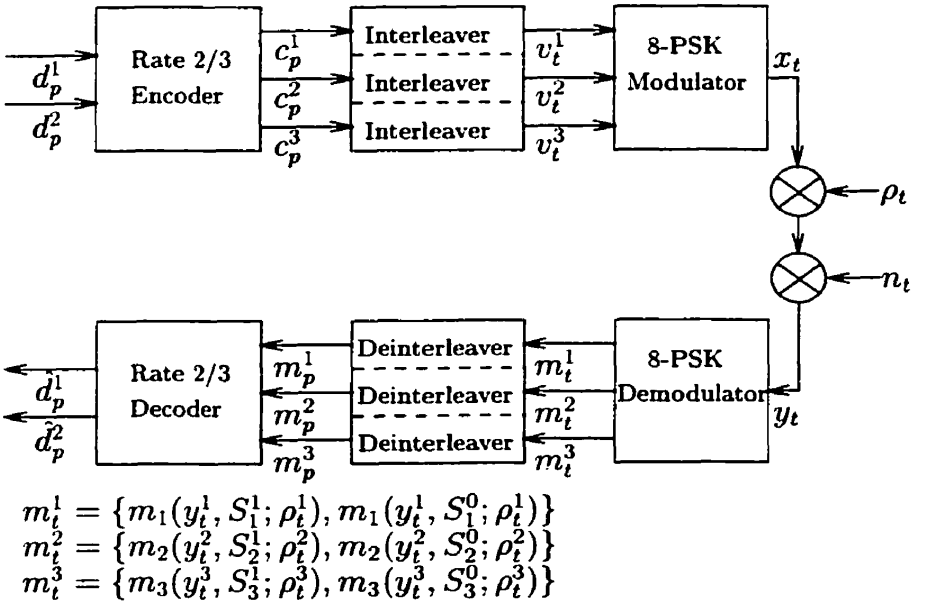
Chapter 3

Bit-Interleaved Coded Modulation

Bit-interleaved coded modulation is a technique that combines coding, modulation, and interleaving in a clever way such that the diversity of a communications system is increased. It has been shown that while Euclidean distance is an important factor in a AWGN channel, it is not so important in a fading channel [19]. What is more important in these environments is diversity. Diversity is defined as the spreading out of information across time or space. In the TCM type systems, the only way to increase diversity is either increase the constraint length or to avoid parallel transitions within the code. BICM, invented by Zehavi [13], successfully increases diversity while allowing for use of standard codes designed for the AWGN channel. This section will describe the BICM system as well as the standard TCM system. The performance of BICM will be examined over the AWGN and IID Rayleigh channel propagation models.

3.1 BICM System Overview

In total the BICM system is comprised of three distinct elements: FEC coder/decoder, interleaver/de-interleaver, and modulator/demodulator, as shown in Figure 3.1, which depicts a BICM system using a rate-2/3 code with 8-PSK modulation. Although a



$$m_t^1 = \{m_1(y_t^1, S_1^1; \rho_t^1), m_1(y_t^1, S_1^0; \rho_t^1)\}$$
$$m_t^2 = \{m_2(y_t^2, S_2^1; \rho_t^2), m_2(y_t^2, S_2^0; \rho_t^2)\}$$
$$m_t^3 = \{m_3(y_t^3, S_3^1; \rho_t^3), m_3(y_t^3, S_3^0; \rho_t^3)\}$$

Figure 3.1: Zehavi's [13] BICM system using a rate-2/3 code with 8-PSK modulation

specific case is shown, BICM is flexible in allowing any type of code (punctured or not) to be used in combination with virtually any modulation type. The best way to understand BICM is by examining the encoder and decoder separately.

3.1.1 Encoder

1. The data bits are encoded using an (n, k) convolutional code such that

$$D = [d_0^1, d_0^2, \cdots, d_0^k, d_1^1, d_1^2, \cdots, d_1^k, \cdots, d_p^1, d_p^2, \cdots, d_p^k] = [D_0, D_1, \cdots, D_p]$$

and

$$C = [c_0^1, c_0^2, c_0^3, \cdots, c_1^n, c_1^1, c_1^2, c_1^3, \cdots, c_1^n, \cdots, c_p^n, c_p^1, c_p^2, c_p^3, \cdots, c_p^n] = [C_0, C_1, \cdots, C_p]$$

where D_p and C_p are the k-tuple of input bits and the n-tuple of encoded bits of the p^{th} input signal respectively. The labels d_p^j and c_p^j represent the jth bit of D_p and C_p respectively.

2. Each of the n outputs of the encoder is fed into an independent, random interleaver.

3. Once interleaved, the binary n-tuple is mapped to one of $M = 2^n$ channel signals in the signal set S where $x = \mu(v^1, v^2, \cdots, v^n)$ and v^i are the points in the constellation. The constellation mapping used is Gray labelling for BICM without iterative decoding [14] and mixed labelling for BICM with iterative decoding [26]. Three constellation mappings for 8-PSK are shown in Figure 3.2 and are described as follows:

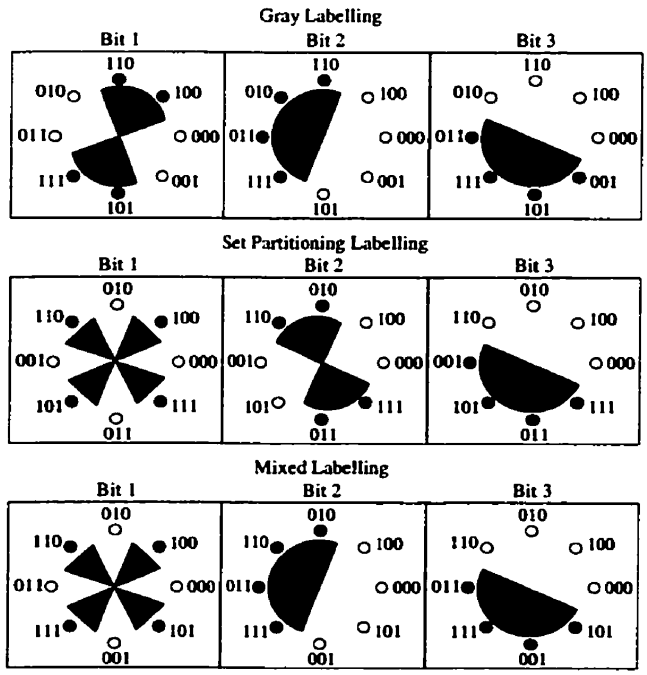


Figure 3.2: BICM Gray, Set-partitioning, and Mixed labelling for 8PSK - the furthest bit to the right is the MSB

- Gray labelling - The objective is to minimize the number of bit errors which occur when a specific constellation point is transmitted but the receiver thinks that the adjacent point has been sent. It can be thought of as maximizing diversity. This is done by assigning constellation points with a binary representation which differs by only one bit between adjacent

points. For example in 8PSK, binary labels 101 and 111 are mapped to adjacent points.

- Set-partitioning (SP) labelling - The idea is to provide greatest protection to the most significant bits and the least protection to the least significant bits thereby maximizing effective Euclidean distance. This is done by recursively splitting the constellation points in half, assigning each half to the remaining most significant bits. For example in 8PSK, all points which have a label with a one as the MSB are adjacent to one another. No points in which the LSB is one are adjacent to one another. In this scheme, if the receiver again incorrectly chooses the adjacent point, there is a high probability that the MSB will be correct.
- Mixed labelling is a combination of both set-partitioning and Gray labelling. It is easiest to think of mixed labelling in reference to set partitioning labelling. If b_2, b_1, b_0 represents a binary label for natural labelling, mixed labelling can be represented as $b_2, (b_2 + b_1), b_0$ and Gray labelling is $b_2, (b_2 + b_1), (b_1 + b_0)$.

For example, the partitioning of the 8-PSK constellation into subsets for Gray labelling is done as follows:

$S = \{x : x(k) = \sqrt{2E_s} exp\left(j\frac{k\pi}{4}\right), k = 0, 1, \ldots, 6, 7\}$

$S_1^0 = \{x(k), k = 0, 3, 4, 7\}; \qquad S_1^1 = \{x(k), k = 1, 2, 5, 6\}$

$S_2^0 = \{x(k), k = 0, 1, 6, 7\}; \qquad S_2^1 = \{x(k), k = 2, 3, 4, 5\}$

$S_3^0 = \{x(k), k = 0, 1, 2, 3\}; \qquad S_3^1 = \{x(k), k = 4, 5, 6, 7\}$

where S_i^b is the constellation subset such that i^{th} bit within the label has the value b

4. The signal is then transmitted over the channel.

3.1.2 Decoder

1. The received signal is used to calculate two bit-metrics for each bit in the n-tuple. For example, with 8-PSK, six bit-metrics are calculated at each time instant t. In conventional BICM, a sub-optimal bit-metric,$m_i^A(y_t^i, S_i^b; \rho_t^i)$, is used instead of the optimal bit-metric, $m_i(y_t^i, S_i^b; \rho_t^i)$. This sub-optimal approximation of the optimal maximum-likelihood(ML) bit-metric, derived in Appendix A, reduces the number of computational operations required for decoding. For AWGN and Rayleigh channels, the bit-metrics are

$$\begin{aligned} m_i(y_t^i, S_i^b; \rho_t^i) &= \log\left\{\max_{\mathbf{x}\in S_i^b} p_{\theta_t}(\mathbf{y}_t|\mathbf{x})\right\} \\ m_i^A(y_t^i, S_i^b; \rho_t^i) &= -\min_{x\in S_i^b} \| y_t - x \|^2 \qquad \text{(AWGN)} \\ &= -\min_{x\in S_i^b} \| y_t - \rho_t x \|^2 \qquad \text{(Rayleigh} - \text{Perfect CSI)}. \end{aligned}$$

   where θ_t and ρ_t are the amplitude and phase of the fading envelope at time t respectively. The CSI is the channel-state information and collectively refers to the channel affects on signal amplitude, phase, and delay. In AWGN, ρ_t^i=1.

2. The trellis branch-metrics are then formed by de-interleaving and summing the bit-metrics corresponding to all possible values of $\hat{C}_p$.

$$m^A(Y_p, S; \rho_p) = \sum_{i=1}^{n} \left\{(1-\hat{c}_t^i)\ m_i^A(y_p^i, S_i^0; \rho_p^i) + \hat{c}_p^i\ m_i^A(y_p^i, S_i^1; \rho_p^i)\right\}$$

   where $Y_p = (y_p^1, y_p^2, \cdots, y_p^n)$, $\rho_p = (\rho_p^1, \rho_p^1, \cdots, \rho_p^n)$, and the de-interleaved metrics are written using the time sequence p instead of t. The symbol $\hat{}$ denotes an estimated value.

3. The trellis branch-metrics are fed into the Viterbi decoder which determines the

binary sequence $\hat{C}$ with the highest cumulative sum of metrics.

$$m^A(Y, S; \rho) = \sum_{p=1}^{N} m^A(Y_p, S; \rho_p)$$

where N represents the length of the codeword.

3.1.3 Iterative Decoder (BICM-ID)

In iterative decoding, the output of the decoder is fed back to the demodulator. The demodulator uses the decoding information to help decide which point in the constellation subset it should select. The exact procedure is as follows.

1. The first round of decoding using BICM-ID is identical to that of BICM.

2. The subsequent rounds of decoding all begin with this step. The iterative-decoding bit-metric, $m_i^{ID}(y_p^i, S_i^1; \rho_p^i)$, is re-calculated as follows for Rayleigh channels with perfect CSI:

$$\begin{aligned}
m_1^{ID}(y_p^1, S_1^b; \rho_p^1) &= \| y_p - \rho_p \mu([b, \hat{v}_p^2, \hat{v}_p^3, \cdots, \hat{v}_p^n]) \|^2 \\
m_2^{ID}(y_p^2, S_2^b; \rho_p^2) &= \| y_p - \rho_p \mu([\hat{v}_p^1, b, \hat{v}_p^3, \cdots, \hat{v}_p^n]) \|^2 \\
&\vdots \quad \vdots \quad \vdots \\
m_n^{ID}(y_p^n, S_3^b; \rho_p^n) &= \| y_p - \rho_p \mu([\hat{v}_p^1, \hat{v}_p^2, \hat{v}_p^3, \cdots, b]) \|^2 .
\end{aligned}$$

   where $\hat{v}_p^i$ are the decoded bits out of the Viterbi decoder. A similar calculation is performed for AWGN channels.

3. Branch-metric calculation - identical to the Step 2 for BICM (Section 3.1.2).

4. Viterbi decoding - identical to the Step 3 for BICM (Section 3.1.2).

Steps 2 and 3 above apply to hard-decision feedback only. Soft-decision feedback employs the use of the BCJR algorithm over the trellis. More details on this decoding

strategy can be found in [17, 26, 27].

3.2 Performance

The performance of BICM is discussed in this section. We examine and contrast the performance of BICM, BICM-ID and symbol-interleaved coded modulation (CM) over various channels. The results in this section were simulated using a Rate 2/3, 8-state, non-recursive convolutional code. This code is the same as the one used by Zehavi (octal generators{(4,2,6,),(1,4,7)}) and is optimized for diversity [28, page 331]. The simulations were run with a frame size of 2502 bits and gray constellation labelling was used for both CM and BICM. Mixed labelling was used for BICM-ID. A random spread interleaver [29] was used which had a minimum spread of 26. All performance curves are presented in terms of bit-error rate(BER) versus normalized information bit signal-to-noise ratio (SNR). The normalized bit SNR, in dB is determined by the following relation

$$\text{BitSNR} = 10\log\left\{\frac{E_b}{N_0}\right\} = 10\log\left\{\frac{E_s}{N_0 \cdot R}\right\} \tag{3.1}$$

$$= 10\log\left\{\frac{E_s}{(2\sigma^2)\cdot(R_{code}/m)}\right\} \tag{3.2}$$

where E_b, E_s, and m are the average bit energy, average symbol energy, and number of bits per symbol respectively. The code rate is represented by R_{code} and the overall rate is given by R. The AWGN noise power density, given by N_0 is equal to twice the variance of the AWGN noise($N_0 = 2\sigma^2$).

Figures 3.3 and 3.4 show the performance of the three schemes over the AWGN and Rayleigh channels. These results confirm the results presented by Zehavi [13] and by Li and Ritcey [26] over the same channels with thirteen decoding iterations.

Notice that for the AWGN channel, CM always outperforms BICM. At a BER

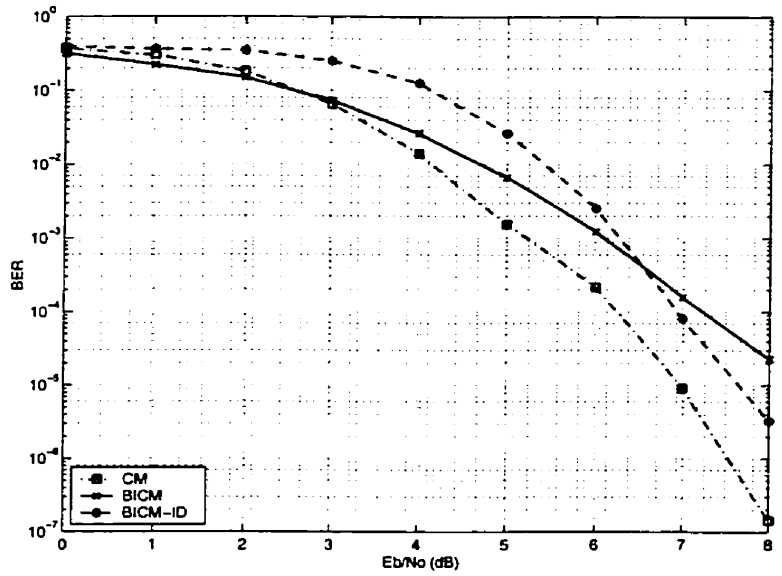


Figure 3.3: BICM over the AWGN channel using 8PSK and an 8-state NR 2/3 Convolutional Code

of 10^{-5}, TCM is approximately 1.5dB better than CM. As indicated earlier, on the AWGN channel with coherent detection, a reliable measure of performance is the minimum Euclidean distance [30, 31]

$$d^2_{free} = \min_{(s_n)\neq(s'_n)} \left[\sum_n d^2(s_n, s'_n) \right]^{1/2}, (s_n), (s'_n) \in C \tag{3.3}$$

the asymptotic error performance of the TCM, BICM, and BICM-ID is dominated by the error event at minimum Euclidean Distance.

Li and Ritcey [16] calculated d^2_{free} for the three systems and showed that indeed the Euclidean Distance for BICM was less than TCM. In terms of BER versus SNR, CM also outperforms BICM-ID with ten iterations over the AWGN channel. Although at a BER of 10^{-5}, CM is only 0.5dB better than BICM-ID. Li and Ritcey [26] have shown that at high SNR, BICM-ID will, in fact, perform better than CM.

On the Rayleigh fading channel, the asymptotic error performance is dominated by the *time diversity* of the coded modulation scheme. In general, the error probability

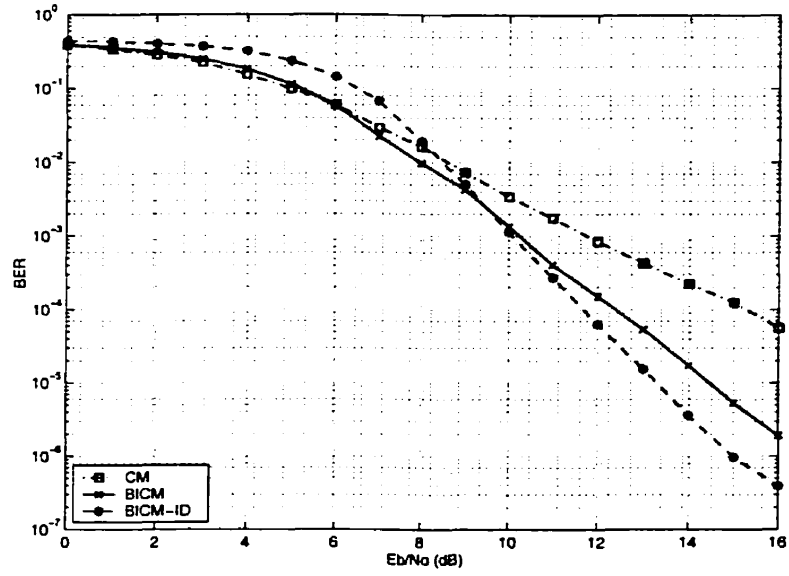


Figure 3.4: BICM over the IID Rayleigh channel using 8PSK and an 8-state NR 2/3 Convolutional Code

on a fading channel is directly proportional to the probability of a deep fade [19, page 43]. This error probability can be reduced by using the *time diversity technique*. This technique transmits the same symbol multiple times over a channel in the hope only some of the transmissions will be affected by deep fades. In a fading channel, the reduction in error probability that results from using the time diversity technique is equivalent to the reduction achieved by increasing the effective length of the code. In other words, time diversity on a fading channel can be defined as the minimum number of distinct symbols between any two codewords(hamming distance) [19, pp 161-165].

Again, in [16] it was shown that the Hamming distance was indeed the larger in BICM than in TCM. This is confirmed by observing the simulation results presented in Figure 3.4. At a BER of $P_b = 10^{-}5$, BICM-ID(10) outperforms TCM by approximately 2dB whereas BICM outperforms TCM by approximately 1dB.

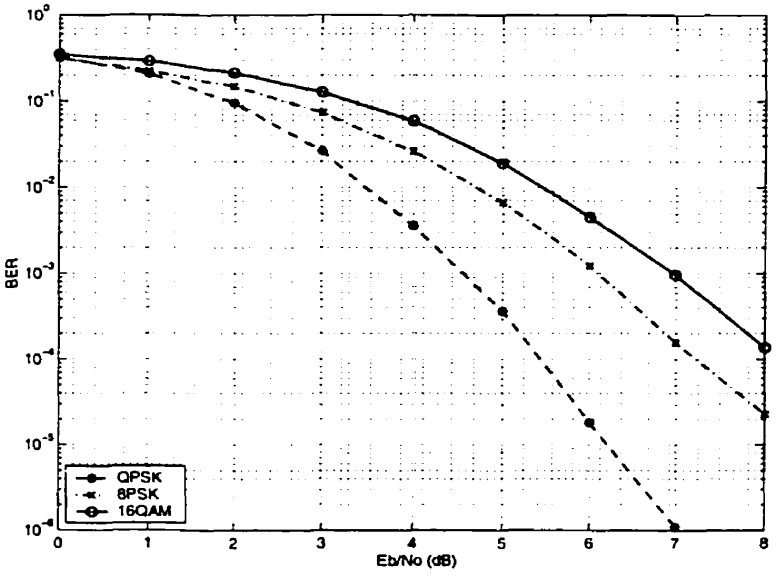


Figure 3.5: BICM over the AWGN channel using an 8-state NR 2/3 Convolutional Code

3.3 Capacity

Capacity is a measure of the amount of information which can be reliably sent over a given channel. Shannon showed that the capacity of a discrete, memoryless channel is given by:

$$C = \max_{p(x)} I(\mathbf{x};\mathbf{y}),$$

where $p(x)$ is the input distribution and $I(\mathbf{x};\mathbf{y})$ is the mutual information between two random variables, $\mathbf{x}$ and $\mathbf{y}$ [32]. The capacity of coded modulation was derived by Ungerboeck [10]. Under the constraint of uniform input distribution, it can be written (in terms of information bits per N complex dimensions) as

$$AMI_{TCM} = I(\mathbf{x};\mathbf{y}|\theta) = m - E_{\mathbf{x},\mathbf{y},\theta}\left[\log_2 \frac{\sum_{z\in S} p_\theta(\mathbf{y}|\mathbf{z})}{p_\theta(\mathbf{y}|\mathbf{x})}\right] \tag{3.4}$$

where $\mathbf{x}$, $\mathbf{y}$ and θ are the input, output and the channel state parameters respectively. The transition pdf is represented by $p_\theta(\ )$. S is the channel signal set and $m = \log_2 |S|$.

In BICM, the interleaver has the effect of creating m parallel independent channels

over which the data is transmitted [14]. The capacity becomes the average of the AMI's for each of the m independent channels. The resulting capacity for BICM with perfect CSI and uniform inputs is

$$C_{BICM} = I(\mathbf{x};\mathbf{y}|\theta) = m - \sum_{i=1}^{m} E_{\mathbf{x},\mathbf{y},\theta}\left[\log_2 \frac{\sum_{\mathbf{z}\in S} p_\theta(\mathbf{y}|\mathbf{z})}{\sum_{\mathbf{z}\in S_b^i} p_\theta(\mathbf{y}|\mathbf{x})}\right]. \tag{3.5}$$

By means of the data-processing theorem [32, page 32], they proved that

$$C_{BICM} \leq C_{CM}. \tag{3.6}$$

Furthermore, Caire *et al.* [14, 33] used (3.5) to develop capacity curves for QPSK, 8-PSK and 16QAM BICM with SP and Gray labelling. We have regenerated the curves and have included capacity with mixed labelling. Note that mixed labelling is equivalent to SP labelling for QPSK and 16QAM These curves are shown in Figures 3.6 through to and 3.11. The curves have been generated by collecting 2048 noise samples at each point. It can easily be observed from the figures that performance of BICM using Gray labelling is the best and is very close to that of CM. In fact, it is equal to CM for QPSK. Mixed labelling performs better that SP for 8PSK. Mixed labelling is the same as SP for QPSK and 16QAM.

3.4 BICM Complexity

In Appendix D, we derive the complexity involved in decoding BICM, TCM, and BICM-ID using convolutional codes. If we define the total number of operations required for decoding TCM as H, then by using a rough approximation of the equations in Tables D.7-D.4, it can be said that the total number of operations required for Rayleigh decoding of BICM is $(2M)H$ for the optimal metric and $(1.5M)H$ for the sub-optimal metric when using convolutional codes where M represents the number of signals in the constellation. For Rayleigh decoding of BICM-ID, the total number

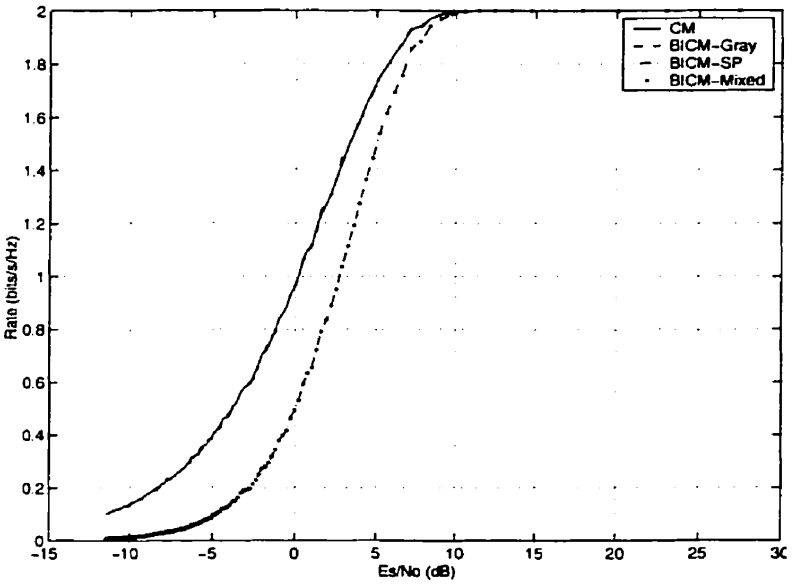


Figure 3.6: BICM and CM capacity versus SNR for QPSK over AWGN with coherent detection (SP denotes set-partitioning labelling)

of operations is roughly $(2M + 3\ell)H$ for the optimal metric and $(1.5M + 3\ell)H$ for the sub-optimal metric where ℓ represents the number of decoding iterations.

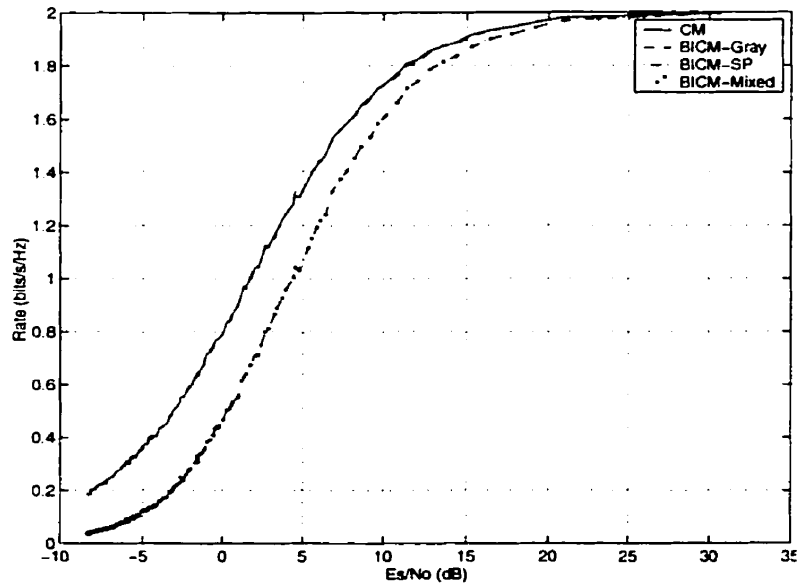


Figure 3.7: BICM and CM capacity versus SNR for QPSK over Rayleigh fading with coherent detection (SP denotes set-partitioning labelling)

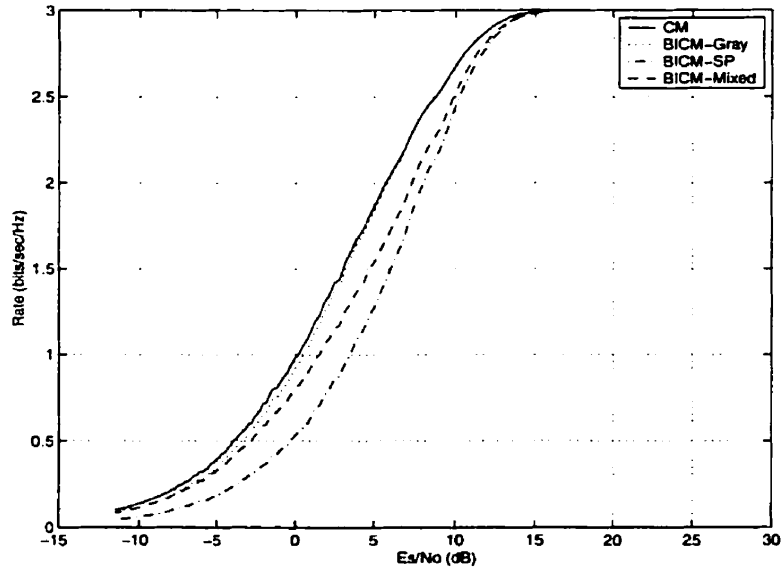


Figure 3.8: BICM and CM capacity versus SNR for 8-PSK over AWGN with coherent detection (SP denotes set-partitioning labelling)

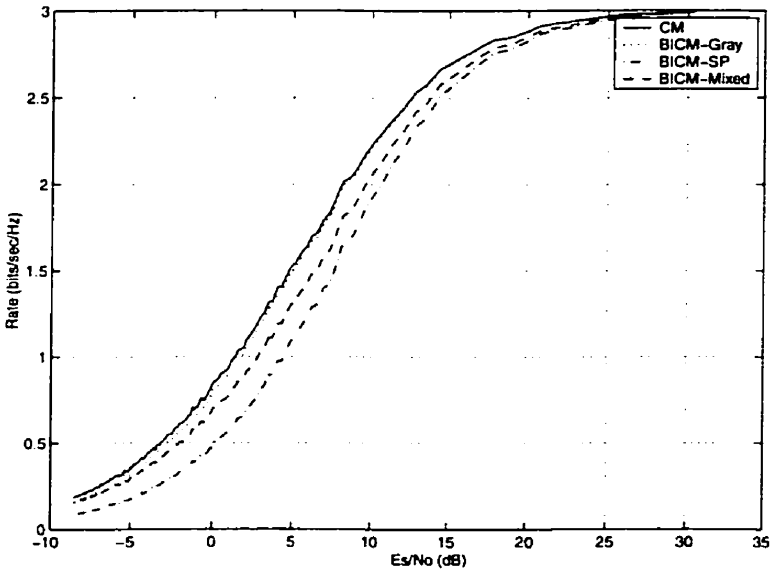


Figure 3.9: BICM and CM capacity versus SNR for 8-PSK over Rayleigh fading with coherent detection (SP denotes set-partitioning labelling)

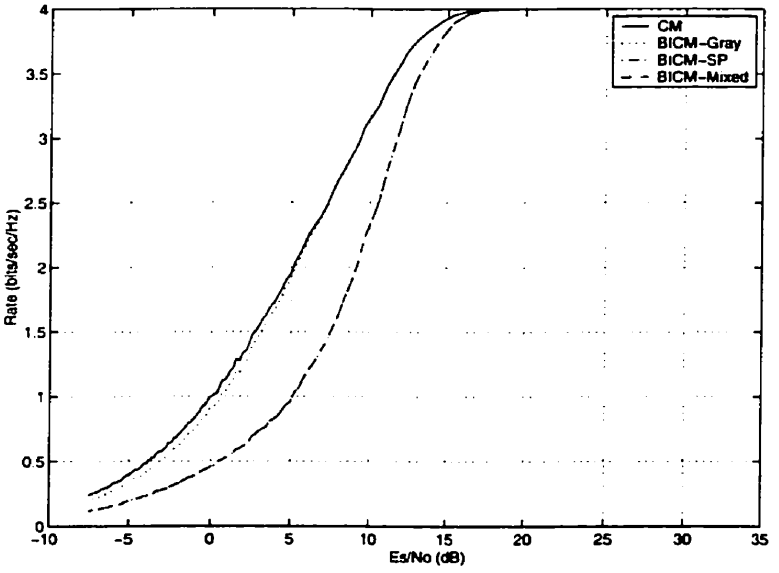


Figure 3.10: BICM and CM capacity versus SNR for 16-QAM over AWGN fading with perfect CSI (SP denotes set-partitioning labelling)

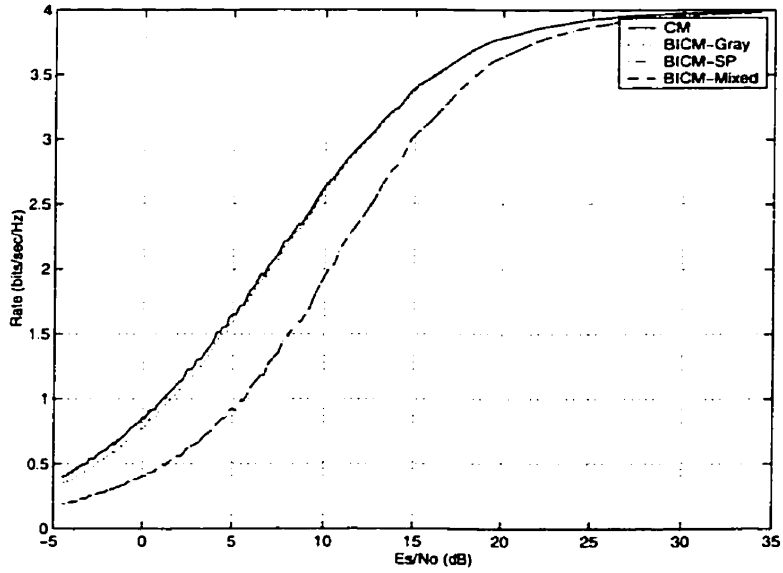


Figure 3.11: BICM and CM capacity versus SNR for 16-QAM over Rayleigh fading with perfect CSI (SP denotes set-partitioning labelling)

Chapter 4

3G Wireless Channels

This chapter will examine the performance of BICM in a 3G wireless type channel. The sections in this chapter will detail the selected 3G channel model, briefly discuss spreading and the Rake receiver, and present the results of the multipath simulations. First, however, BICM will be examined in a single-propagation environment.

4.1 Single-Path (Frequency Non-Selective) Performance

Before performance of BICM in multipath is analyzed, it is useful to first examine the single-path case. Single-path propagation in a wireless channel is also referred to as frequency non-selective propagation. Figure 4.1 shows the performance of BICM in a single-path environment for both slow and fast fading conditions. The first observation which can be made is that BICM outperforms CM with symbol-interleaving by almost an order of magnitude at an SNR of 14dB. The BER for BICM is roughly 2×10^{-5} when CM with symbol interleaving has a BER of 10^{-4}. Furthermore, it can be readily seen that BICM exhibits the same performance for both fast and slow fading. This may be attributed to the fact that the block length is relatively small (2502 bits), which limits the distance which the bits can be separated. Small block

lengths are required for 3G as will be discussed in Chapter 5.

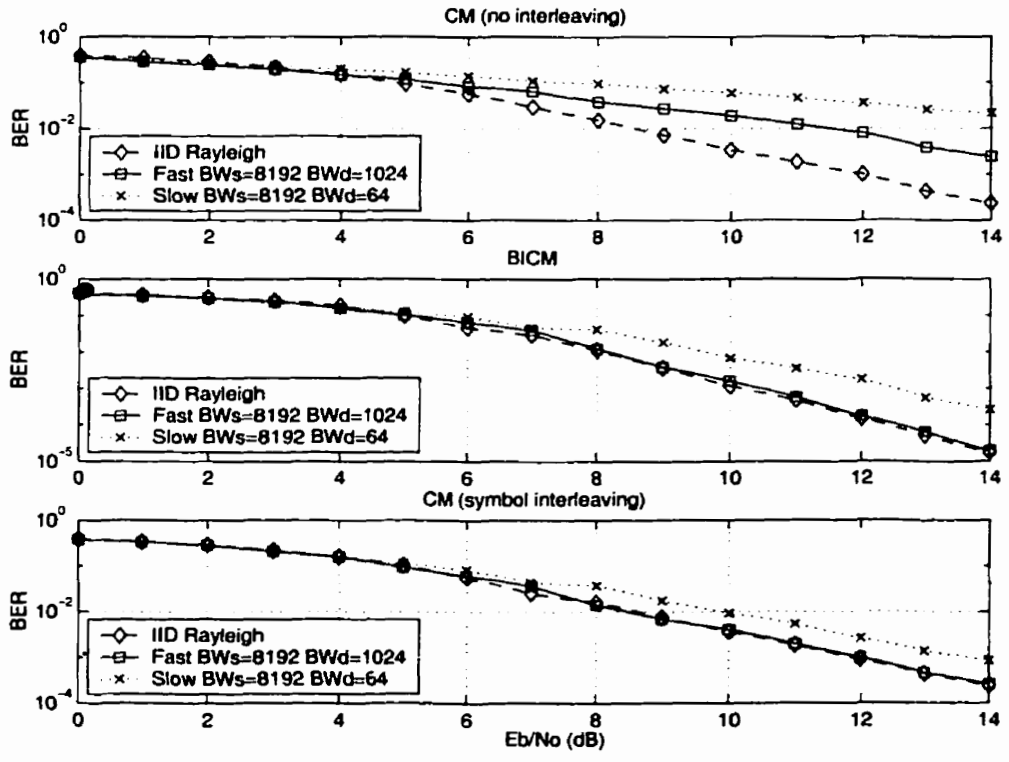


Figure 4.1: BICM and CM performance for frequency selective (slow and fast) propagation using 8PSK and an 8-state NR 2/3 Convolutional Code - octal(65,57) (coded block length=2502 bits)

4.2 Channel Model Definition

The multipath channel chosen is similar to the one used by Melis and Giovanni [34] and represents the 3GPP vehicular A propagation environment as defined by the European Telecommunications Standard Institute [35]. The tapped delay line model of our channel is illustrated in Figure 4.2. Melis' channel parameters are presented in Table 4.1 along with the values used in these simulations.

When generating the fading envelope for each path, the mobile velocity was set to 120km/h. Spreading, using Walsh codes (see Section 4.3), was applied to the modulated signal with the chip rate chosen to be 4.096Mchips/s (3GPP). This combination

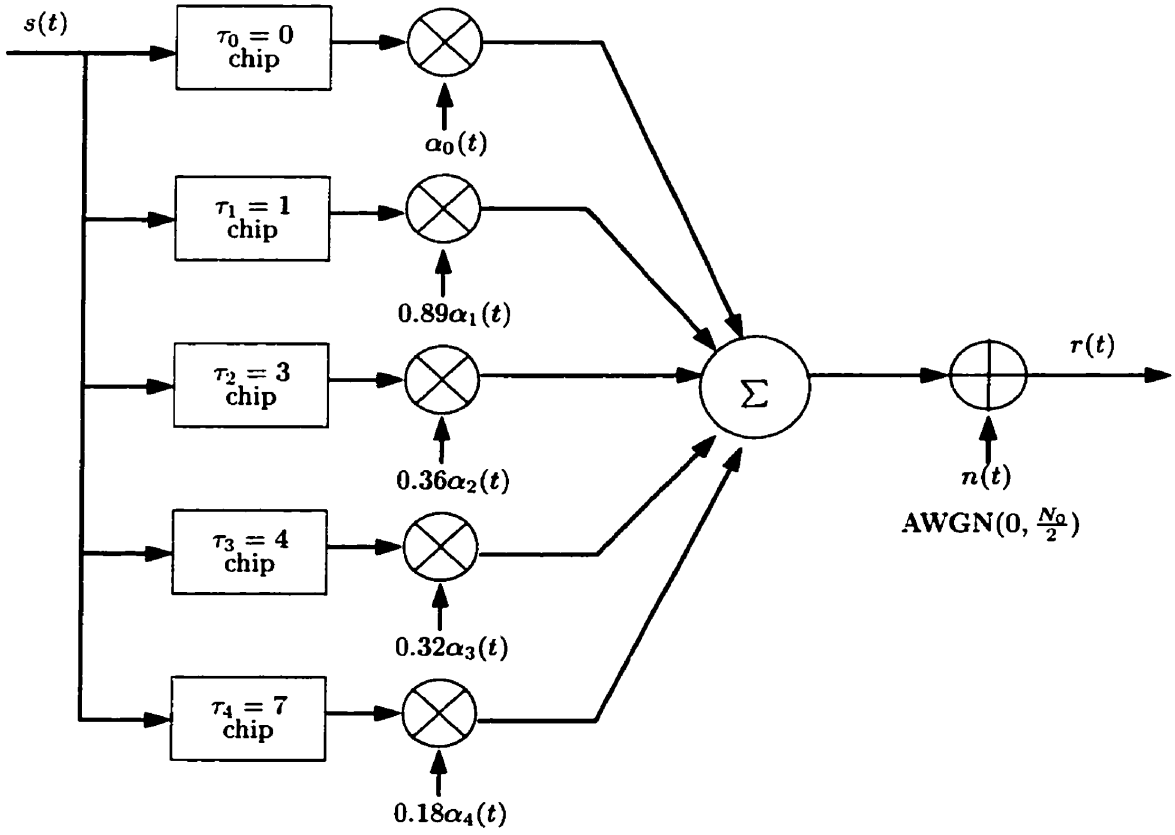


Figure 4.2: Tapped Delay Line Model Used in Simulations

Tap	Parameters	
	Delay (ns)	Avg. power (dB)
1	0	0
2	310	-1.0
3	710	-9.0
4	1090	-10.0
5	1730	-15.0
6	2510	-20.0

Tap	Normalized Parameters	
	Chip Delay	Atten Weight α
1	0	1
2	1	0.89
3	3	0.36
4	4	0.32
5	7	0.18
6	10	0.1

Table 4.1: The Tapped-Delay Line(TDL) model parameters of the Vehicular A channel

of chip rate and velocity resulted in a very slow fading envelope being applied to each path (normalized ratio $2f_m/f_s \approx 0.0001$). To simply the simulations, it is assumed that the received signal was sampled once at the center of each chip interval. This eliminates the need for pulse shaping, but will only work provided that the multipath delays are integer multiples of the chip interval. For this reason, the propagation delays were normalized and rounded such that they are multiples of the chip interval.

4.3 Spreading

Spreading is a technique used in CDMA systems and is well described elsewhere [3, 18, 20]. Walsh codes of equal length are orthogonal to each other and are often used for channel separation in the downlink. These codes are the rows of a Hadamard matrix. In brief, the matrix of size 2^n is constructed as follows:

$$H_1 = [+] \quad \text{, and } H_{2^n} = \begin{bmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{bmatrix}$$

where n is a positive integer and $+$ and $-$ represent positively and negatively signed elements respectively . The Hadamard matrix of size eight is

$$H_8 = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix}$$

It is the last row of the matrix which is the Walsh code (W8) that is used in our simulations. The Walsh codes of length sixteen and thirty-two which are also used are:

$$\begin{aligned} W16 &= \begin{bmatrix} W8 & -W8 \end{bmatrix} \\ &= \begin{bmatrix} + & - & - & + & - & + & + & - & - & + & + & - & + & - & - & + \end{bmatrix} \\ W32 &= \begin{bmatrix} W16 & -W16 \end{bmatrix} \end{aligned}$$

The last row of the Walsh matrix of size sixteen is the one chosen for this work.

4.4 Rake Receiver

In section 4.2, is was shown that the 3G channel model is frequency selective - it has multiple propagation paths. To mitigate the ISI, we implemented a RAKE receiver as presented in Figure 4.3. In order to use such a receiver, you need to have good knowledge of the specific arrival times of each path. This information is usually obtained from the pilot channel in a CDMA system. Perfect knowledge of the specific

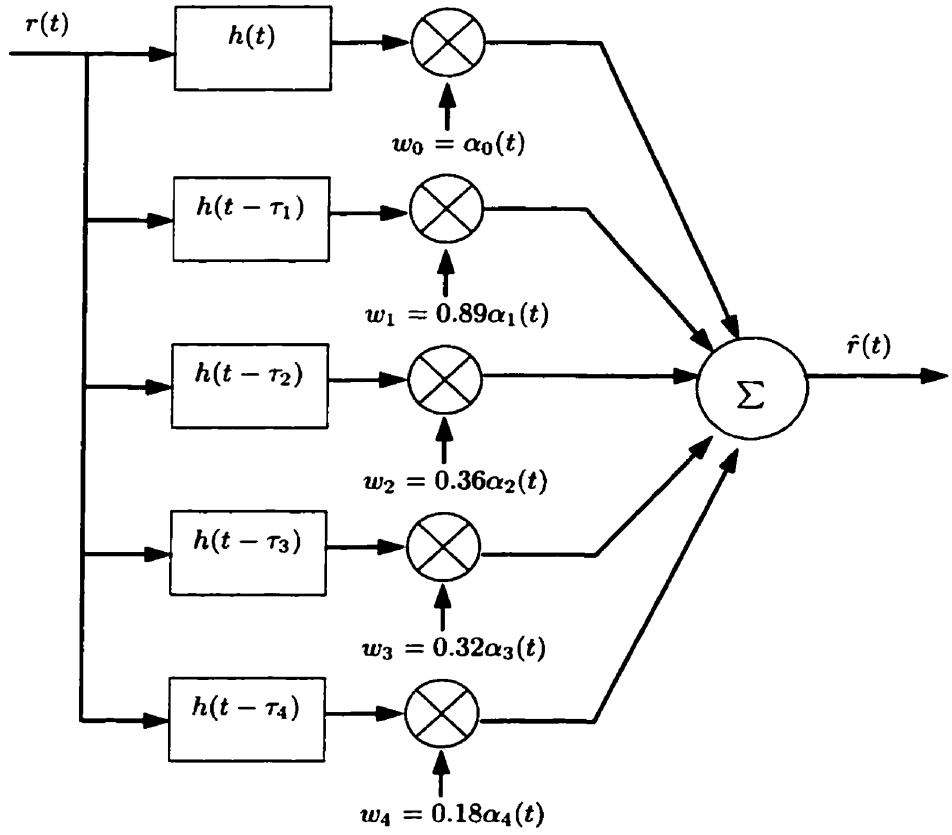


Figure 4.3: Maximum-Ratio Combining RAKE Receiver Used in Simulations where $h(t)$ is the Matched Filter

path delays will be assumed in our implementation. Once all the paths are coherently tuned, there is still the question of how to combine them. There are several types of combining strategies - the two major ones being: maximum-ratio combining (MRC) and equal-gain combining (EGC) [18]. Equal-gain combining weights all paths equally and coherently adds them together. Maximum-ratio combining, shown to be the optimal linear combining strategy [36, 37], weighs each of the fingers by their channel attenuation. The obvious requirement for this method is that good amplitude channel state information is required. Like delay, this measurement is often performed on the pilot signal. In a 3G propagation environment, the fading envelope changes slowly; therefore, it is relatively easy to make this measurement. For this reason, we will assume perfect amplitude channel state information and will use the maximum-ratio combining technique.

4.5 Multipath (Frequency Selective) Performance

In this section, the performance of BICM is examined in a multipath environment. Figures 4.4, 4.5, 4.6, and 4.7 illustrate the performance in two multipath channel configurations. The channels and Rake receivers used are 3-path/3-finger and 5-path/5-finger models based on the UMTS model shown in Figure 4.2. The 3-path/3-finger model only used the first three taps of the UMTS model. The data is encoded, modulated and then spread with the length 16 Walsh code as described in Section 4.3.

There are several observations which can be made from these figures. The most important of which is that bit-interleaving provides the same performance when compared to symbol-interleaving over the 3-Path channel and, in fact, provides worse performance over the 5-Path channel. This result supports the theory that as the number of resolvable paths increases, the channel begins to appear more and more Gaussian-like [19]. As shown in Section 3.2, bit interleaving performs better than symbol interleaving over a Rayleigh channel but worse over an AWGN channel. The

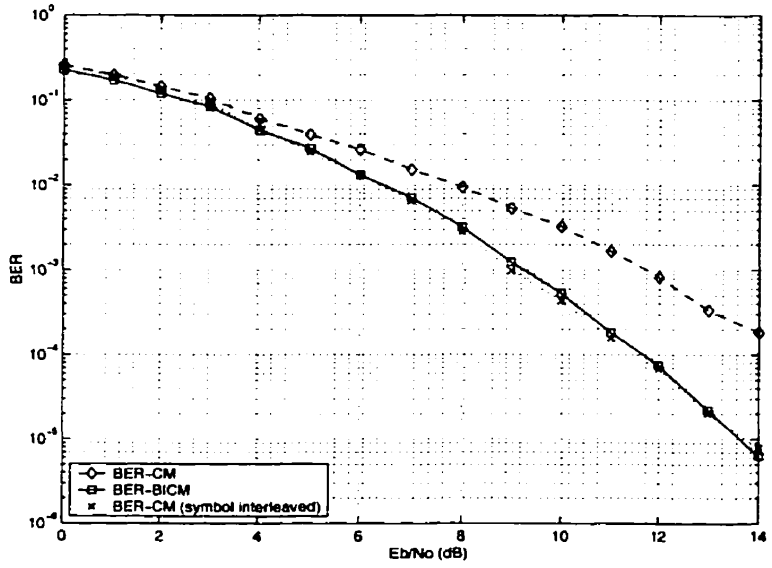


Figure 4.4: BICM over the 3-path, multipath fading channel using a 3-finger RAKE receiver with perfect CSI for 8PSK using an 8-state NR 2/3 Convolutional Code - octal(65,57)

second interesting effect is the 16-QAM error floor in Figure 4.7. This error floor can be attributed to the autocorrelation properties of the Walsh spreading code (shown in Figure 4.8). We see from the figure that the correlation of the sequence with shifted versions of itself is relatively high. It appears that this may induce an error floor at high SNRs. We also observed that as the spreading factor becomes larger relative to the delay spread of the channel, the error floor begins to appear at higher SNRs. To mitigate this effect, one would need to either increase the spreading factor or use a better code and/or interleaver. In the next section we will explore the use of turbo codes in such a system.

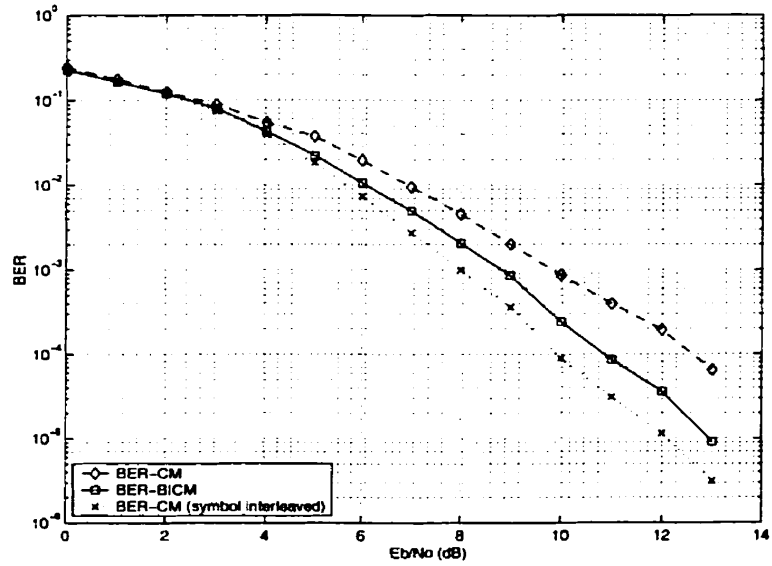


Figure 4.5: BICM over the 5-path, multipath fading channel using a 5-finger RAKE receiver with perfect CSI for QPSK, 8PSK, and 16QAM using an 8-state NR 2/3 Convolutional Code

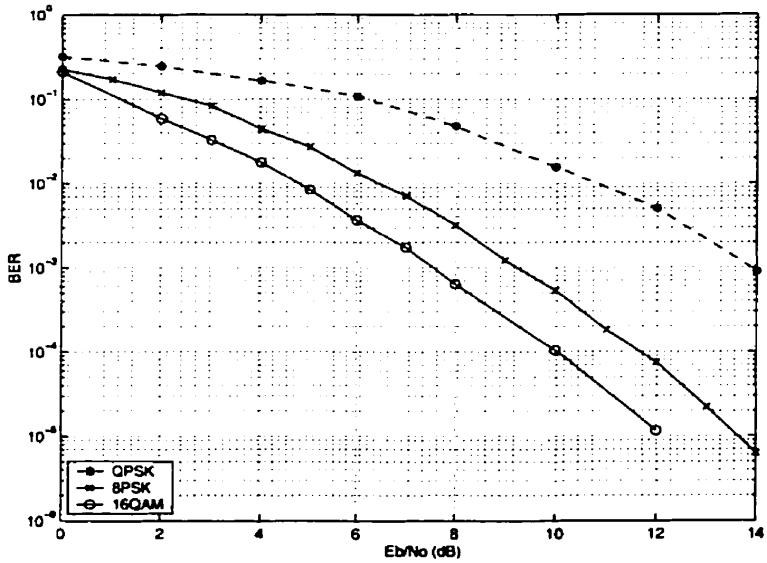


Figure 4.6: BICM over the 3-path, multipath fading channel using a 3-finger RAKE receiver with perfect CSI for 8PSK using an 8-state NR 2/3 Convolutional Code

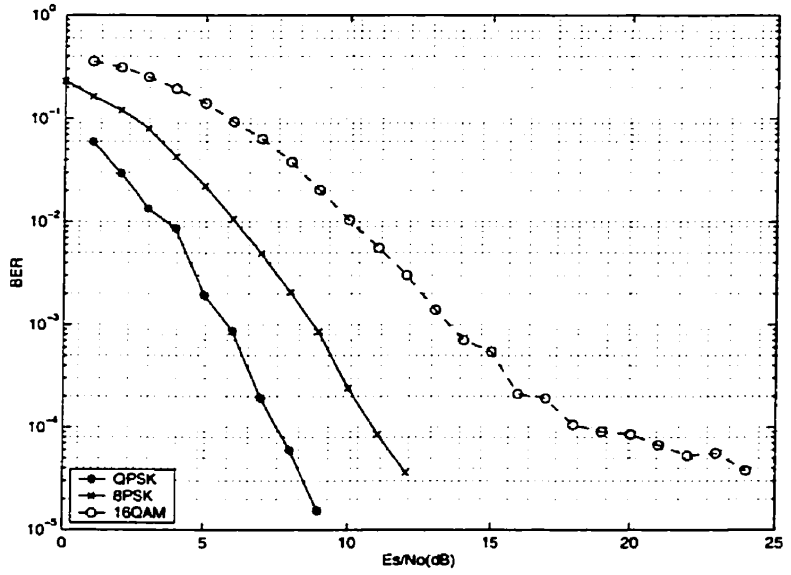


Figure 4.7: BICM over the 5-path, multipath fading channel using a 5-finger RAKE receiver with perfect CSI for QQPSK, 8PSK, and QPSK using an 8-state NR 2/3 Convolutional Code

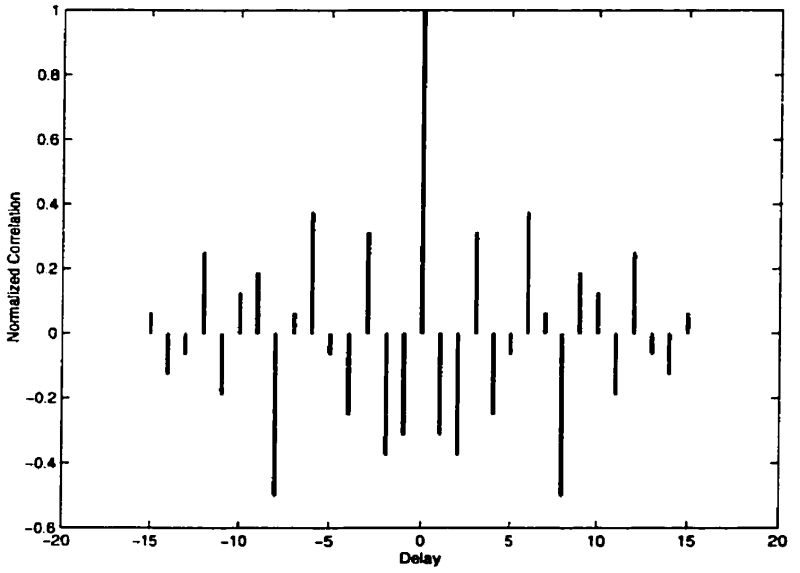


Figure 4.8: Autocorrelation of Walsh 16 spreading code of length 16

Chapter 5

Adaptive 3G Channel Coding

The UMTS standard is divided up into various classes of service. The main two classes of service which are used for transmission of real-time data are the speech and low data-delay (LDD) services [34, 35]. The parameters for each of these two service classes is presented in 5.1.

Service Class	Bit Rate	Minimum BER	Maximum Delay	Frame Length	Chip Rate	Spreading Factor
Speech	8kb/s	10^{-3}	20ms	10ms	4.096Mchips/s	8-256
LDD	144kb/s	10^{-6}	50ms	10ms	4.096Mchips/s	8-256

Table 5.1: Parameters for the Real-Time UMTS classes of service

Generally speaking, in a mobile environment, users who are close to the base station will see a better downlink SNR than users who are near the cell boundary. It is important that all of these users achieve a minimum quality of service. One important performance metric which determines the quality of service (QOS) is the data error rate. Shannon pointed out that, as the SNR drops, the maximum achievable data rate must decrease as well in order to maintain a specified error rate. In a non-adaptive rate type system, the base station must choose a channel coding scheme which meets the QOS for the user with the poorest SNR. For example, a user who is close to the base station may have an $E_s/N_o = 9$dB while a user near the cell boundary may have

an $E_s/N_o = 3$dB. The maximum achievable rates for these SNRs are 3.16 bits/s/Hz and 1.58 bits/s/Hz respectively. If the base station is forced to select a channel coding scheme which accommodates the worse of the two SNRs, the user close to the base station will have their maximum available data rate cut in half. One solution to mitigate this problem is to have an adaptive channel coding scheme which is based on the user's SNR.

In this chapter, an adaptive channel coding scheme based on BICM using a 16-state turbo code for the two UMTS classes of service is proposed. For each class of service, a set of configurations (consisting of various code rates and modulation types) will be selected which will guarantee system rate performs within a fixed distance from capacity over a range of SNRs. The turbo code used is explained in detail in Appendix B.

5.1 Service Class Constraints

In an UMTS system, the two classes of service mentioned above would exist on two different types of data (traffic) channels. Typically, in the downlink, a user would be assigned one or more traffic channels which would carry voice and/or IP type traffic. The speech channel would carry voice (or latency intolerant) type traffic whereas the LDD channel would carry IP (or latency tolerant) type traffic. These channels would be made orthogonal to each by assigning each channel a different orthogonal spreading code (usually a different row of the Walsh matrix which as described in Section 4.3).

In the LDD class of service, a delay of up to 50ms can be tolerated. The BER, however, must be less than 10^{-6}. One way to achieve this is to use selective-repeat ARQ. This topic is well discussed in the literature (see [38, 39, 29]) and will be briefly described here. Simply stated, the transmitter(base station) continuously sends frames of coded data to the receiver(user terminal). Before turbo encoding

in the transmitter, each frame will have been CRC encoded. The CRC bytes will have been appended to the data entering the turbo code. As the receiver decodes the frames, a CRC check will be executed. If the frame fails the check, the receiver will prompt the transmitter to send the corrupt frame again. As discussed, with turbo codes, an error floor will tend to appear at lower BERs. This problem can be mitigated through the choice of a better interleaver [40]. With the spread interleavers used in this work, the error floor appears at approximately 10^{-5}. For this reason, the selective-repeat ARQ is a viable option for the LDD channel. An ARQ scheme will attempt to keep the data error rate at zero. In order to satisfy this strategy, there can be no bound on the number of retries allowed. In reality, however, there is a bound and in the LDD class the bound is chosen such that the overall latency is less than 50ms. When using an ARQ scheme, there are two questions which must be answered. First, what should the FER operating point be? By choosing the minimum raw FER to be 10^{-2}, a BER of roughly 10^{-4} can be achieved. Second, what is the maximum allowable number of retries? With a delay tolerance of 50ms and assuming there is some processing time associated with an ARQ request, a choice of a maximum three retries of 10ms frames would be appropriate. Given these settings, the corresponding post-ARQ FER would be roughly 10^{-6} with a BER of 10^{-12}. This scheme fulfills the requirements of the LDD class of service and will be basis for choosing a set of LDD operating configurations.

Unlike the LDD class, the speech service class cannot tolerate latency. It can, however, tolerate a reasonably high BER of 10^{-3}. This means that, with a frame size of 10ms and a maximum delay of 20ms, there is only enough time to send one block of data. Based on this constraint, a similar set of operating configurations will be chosen for the speech class.

5.2 LDD and Speech Class Operating Modes

The simulations which are used as the basis for selection of operating modes as performed over an AWGN channel. The motivation for this is as follows.

5.2.1 AWGN Propagation Selection

There are two major differences between traffic channels which experience AWGN only and ones which experience both AWGN and frequency-selective fading (see Chapter 4). The first is that the data transmission in 3G must contend with varying SNR due to the fading. In a slow fading channel such as one experienced in 3G systems, the Doppler bandwidth is approximately 3Hz and 200Hz for motion of 3km/h and 120km/h. In 3G, the power control updates happen at 1500Hz - much faster than the fading rate. This means that assuming that we are able to obtain good channel state estimates, power control will mitigate the fading and the SNR will be kept constant at the receiver. The other significant difference is the multipath propagation. By employing a RAKE structure in the receiver, the self-interference caused by multipath and, for that matter, multiple access interference can be modelled as AWGN noise. . Furthermore, we demonstrated that the performance of BICM in multipath fading is more like the AWGN performance than the performance in single path, Rayleigh propagation. Based on these assumptions we can say that the combination of the RAKE receiver with fast power control makes the multipath, fading channel conditionally Gaussian. As a result, it makes more sense to design the system around simulations performed in an AWGN environment.

5.2.2 Selection of Modes

The operating modes were drawn from a set of eighteen configurations which included a 16-state turbo code punctured to six different rates (1/3, 2/5, 4/9, 1/2, 2/3, and 4/5) and three modulation schemes with three different spectral efficiencies (2, 3,

and 4 bits/symbol/2D). From these eighteen combinations of modulation and code rates, ten different operating modes were selected - providing channel coding rates ranging from 2/3 up to 16/5 bits/symbol/2D. The turbo code and puncturing masks used are described in Appendix B. The modulation consisted of three Gray mapped constellations - namely QPSK, 8PSK, and 16QAM with spectral efficiencies of 2, 3, and 4 bits/symbol/2D respectively. Detailed performance curves for all the puncture masks and modulations are presented in Appendix C

An illustration of the proposed system transmitter is presented in Figure 5.1. In 3G systems, channel state information, such as SNR, is typically measured from the reference channel referred to as the pilot channel. The SNR must be estimated and many such algorithms can be found in the literature. In this proposed system, the SNR would be measured at the receiver antenna via the pilot channel and used to determine the operating configuration which the transmitter should use to send data. The receiver would have a target SNR which the transmitter would attempt to satisfy. The receiver would evaluate its measured SNR against the target SNR. It would then proceed to tell the transmitter to increase or decrease its SNR such that the target SNR is met. The target SIR would be chosen such that the required data rate is achieved and the transmitted power is within the allotted range. The performance of all eighteen simulated configurations is presented in Table 5.2. The winning configurations are displayed in Table 5.3.

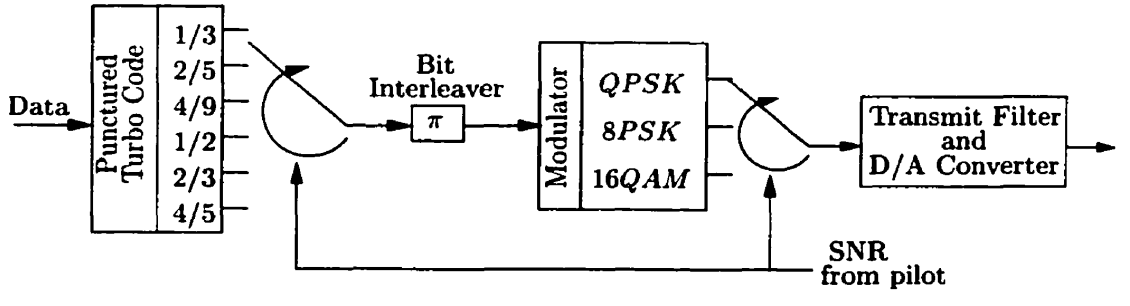


Figure 5.1: Adaptive 3G Channel Coding Transmitter using BICM

Rate (bits per symbol)	Puncture Mask Index	Modulation	E_s/N_o for BER$\leq 10^{-3}$	E_s/N_o for FER$\leq 10^{-2}$	E_s/N_o at Shannon's Capacity
2/3	0	QPSK	-1.14	-0.95	-2.31
4/5	1	QPSK	0.21	-0.43	-1.30
8/9	2	QPSK	0.74	1.22	-0.70
1	3	QPSK	1.46	1.85	0
4/3	4	QPSK	3.50	3.80	1.82
8/5	5	QPSK	5.02	5.79	3.08
1	0	8PSK	1.74	1.83	0
6/5	1	8PSK	3.22	3.54	1.13
4/3	2	8PSK	3.92	4.25	1.81
3/2	3	8PSK	5.09	5.35	2.62
2	4	8PSK	7.09	7.54	4.77
12/5	5	8PSK	8.85	9.84	6.31
4/3	0	16QAM	3.57	3.70	1.81
8/5	1	16QAM	4.98	5.15	3.08
16/9	2	16QAM	5.89	6.05	3.85
2	3	16QAM	6.73	7.02	4.77
8/3	4	16QAM	9.33	9.75	7.28
16/5	5	16QAM	11.15+	11.98+	9.13

Table 5.2: Simulation results for BICM using the 16-state punctured turbo code (octal(37,21)) and a spread interleaver of size 5000.

5.3 Operating Mode Performance

As previously stated, simulations were performed for eighteen configurations - six puncture masks and three modulation schemes. There are several observations worth noting. The distance from capacity is proportional to the level of puncturing. A plot, at a BER=10^{-3}, of the performance of all the modes simulated against capacity is shown in Figure 5.2. As the level of puncturing increases, so does the distance that the E_S/N_0 for a given error rate is apart from the E_S/N_0 at capacity. These ranges are presented in Table 5.4 for the three modulation schemes. Also as puncturing increases, the error floor prominent in turbo codes, starts to emerge at higher error rates. The SNR difference between BER= 10^{-3} and FER= 10^{-2} is at least three times larger at rate-4/5 (high puncturing) than it is at rate-1/3 (no puncturing).

Modes	Rate (bits per symbol)	Puncture Mask Index	Modulation	E_s/N_o for BER$\leq 10^{-3}$	E_s/N_o for FER$\leq 10^{-2}$	E_s/N_o at Shannon's Capacity
0	2/3	0	QPSK	-1.14	-0.95	-2.31
1	4/5	1	QPSK	0.21	-0.43	-1.30
2	8/9	2	QPSK	0.74	1.22	-0.70
3	1	3	QPSK	1.46	1.85	0
4	4/3	4	QPSK	3.50	3.80	1.82
5	8/5	5	QPSK	5.02	5.79	3.08
6	16/9	2	16QAM	5.89	6.05	3.85
7	2	3	16QAM	6.73	7.02	4.77
8	8/3	4	16QAM	9.33	9.75	7.28
9	16/5	5	16QAM	11.15+	11.98+	9.13

Table 5.3: Modes of the system using the punctured turbo code with octal generators (37,21) and an interleaver of size 5000.

	BER= 10^{-3}	FER= 10^{-2}
QPSK	1.17 – 1.94	1.36 – 2.51
8PSK	1.74 – 2.54	1.83 – 3.03
16PSK	1.76 – 2.02	1.91 – 2.85

Table 5.4: Distance from Capacity for System Configurations

Furthermore, it is evident that 8PSK performs significantly worse relative to capacity than does QPSK or 16QAM. As a result, when the operating modes were selected, no 8PSK configurations were selected. The primary selection criteria for choosing operating modes was that the highest rate was providing given for the smallest E_S/N_0. Another consideration was that the system should switch between different modulation schemes at only one SNR point - possible making implementation easier. For these two reasons, ten operating modes were selected supporting symbol SNRs as low as -1.14 dB.

This SNR value is further lowered by the processing gain provided by using CDMA. It is the chip SNR, not the symbol SNR , which is observed at the receiver

antenna. The chip SNR, E_C/N_0, supported by this systems is calculated as follows:

$$\text{minimum supported } E_C/N_0 = -1.14\text{dB} - 10 \times \log_{10}(\text{SF})$$

where SF is the spreading factor. For example, for SF= 8 and SF= 256, the minimum E_C/N_0 is -10.17 and -25.22 respectively.

5.4 Frame Size Constraints and Complexity

The 3G standard constrains the frame size to fit within a 10ms interval. Hence, the number of bits which can be send in each frame is limited. Also, in 3G, the coded data is spread as part of the CDMA processing. This further limits the number of bits available per frame. Table 5.5 contains the maximum number of bits permitted for different modulation schemes and spreading factors (assuming a chip rate of 4.096Mcps) Furthermore, this table also lists the minimum code rate required to satisfy the LDD class data rate of 144kb/s.

A spreading factor of 32 would correspond to 31 orthogonal users in the forward link of a CDMA system. The first Walsh row is typically assigned to the pilot channel.

	Frame Length in Bits			Minimum Code Rate		
Modulation Type	SF=8	SF=16	SF=32	SF=8	SF=16	SF=32
BPSK	5120	2560	1280	9/32	9/16	N/A
QPSK	10240	5120	2560	9/64	9/32	9/16
8PSK	15360	7680	3840	3/32	3/16	3/8
16QAM	20480	10240	5120	9/128	9/64	9/32

Table 5.5: Coded Frame Size and Minimum Code Rate Required to Satisfy the LDD Frame Size of 10ms

In Appendix D, the complexity of Turbo-BICM decoding is roughly derived. This derivation is useful for comparing the complexity of different modulation schemes and data block lengths. The number of operations required to decode BICM-Turbo is a

follows:

$$\text{multiplications} \quad (M)|E| + 2\ell(6|E| + 2) \tag{5.1}$$

$$\text{additions} \quad (M-1)(n-1))\,|E| + 2\ell(2|E| - 2) \tag{5.2}$$

for a code of rate k/n where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations.

	2500 bits	5000 bits	10000 bits
QPSK	$1.075E6$	$2.15E6$	$4.3E6$
8PSK	$1.115E6$	$2.23E6$	$4.46E6$
16QAM	$1.195E6$	$2.39E6$	$4.78E6$

Table 5.6: Number of Operations Required to Decode Turbo-BICM

For the simulations presented here, the number of multiplications required is shown in Table 5.6. In general, the number of multiplications and additions required is proportional to $300K$ and $100K$ respectively where K is the number of information bits per block. Every extra data bit per block translates to an additional 400 operations which must be done when decoding. For this reason, data block sizes of 5000 and less were chosen for use in this system.

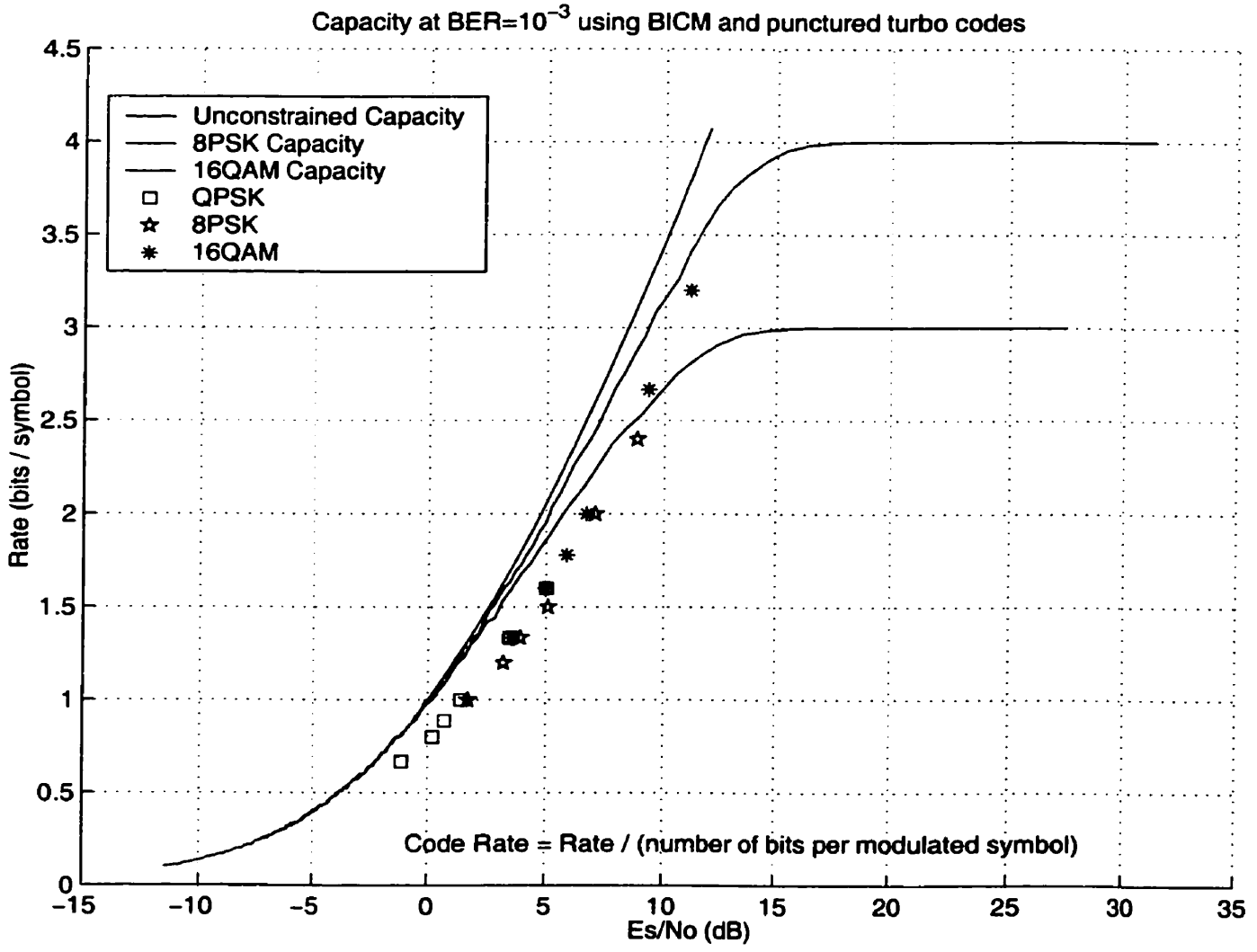


Figure 5.2: Performance of All Configurations at BER= 10^{-3} compared against Capacity

Chapter 6

Conclusions

The performance of bit-interleaving versus symbol-interleaving in conjunction with coded modulation has been evaluated over a variety of channels, as shown in Table 6.1. While the performance of BICM and SICM had been previously examined over AWGN and IID Rayleigh environments, we evaluated the two using more realistic channels. We showed that over a one-path Doppler fading environment, BICM outperforms SICM. We also showed that for short block lengths (2500 bits), IID Rayleigh propagation is a valid assumption for estimating performance in a fast fading environment when either bit or symbol-interleaving is used. In 3G environments the winner is not as clear for short block lengths. Our results suggest that as the number of resolvable paths increases, the channel will look more and more Gaussian-like provided that RAKE reception is employed and that perfect channel state information is available. In our comparison, 3-finger diversity reception is the threshold - less that 3-fingers, BICM outperforms SICM; greater that 3-fingers, SICM is better. While the multipath provides diversity and mitigates fading, we have also shown that the ISI induces an error floor at high SNR - particularly noticeable for 16QAM. Since BICM was shown to be $2M$ times more complex than SICM (where M is the constellation size) it is recommended that it be used in 3G environments only when the number of RAKE fingers (and resolvable paths) is low.

Channel Characteristics			RAKE		
Noise	Fading	Paths	Fingers	Winner	Notes
AWGN	None	1	1	BICM	
None	IID Rayleigh	1	1	SICM	Perfect CSI
AWGN	Doppler Fast	1	1	BICM	Perfect CSI
AWGN	Doppler Slow	1	1	BICM	Perfect CSI
AWGN	Doppler 3G	3	3	BOTH	Perfect CSI, SF=16
AWGN	Doppler 3G	5	5	SICM	Perfect CSI, SF=16

Table 6.1: Best Interleaving Strategy (Bit Versus Symbol) over Various Propagation environments

Turbo coding in combination with BICM performs well when compared to capacity. We showed that by using a standard punctured turbo code with bit-interleaving and higher-order modulation, a set of operating modes can be determined for use in 3G systems which tracks capacity at a fixed distance. Our implementation provided spectral efficiencies between 2/3 and 16/5 bits/s/Hz and tracked AWGN capacity with a distance of 2.1dB (+/- 0.75dB) at a FER= 10^{-2} and a distance of 1.6dB (+/- 0.43dB) at a BER= 10^{-3}. This capacity distance is applicable over a dynamic range of 12.64dB. While the performance is good, turbo codes are expected to be outstanding. This holds true for large block sizes. In our implementation, we examined smaller block sizes, a 3G constraint. Furthermore, a spread interleaver was chosen for use in this work. Recently, other types of interleavers has shown to give better performance [40].

We showed that BICM performs well in Doppler and multipath fading. The advantage to BICM is that it is robust over a variety of channel conditions. We have presented a methodical and adaptive approach to designing the physical layer components of a third generation wireless system. To really appreciate just how resistant BICM is to fading in 3G, one would have to simulate power control. This is also a worth exploring. Furthermore, the objective of our system was to maximize the rate for any particular user. Another perspective is to view the system as providing adaptive capacity instead of adaptive rate. If more capacity is required for the system,

lower the data rate of all users. This allows the base station to reduce the transmit power required for each user and thereby increase system capacity. Although we have covered many aspects of the system design, there are more areas yet to be covered. We have, however, provided a good starting point for future work.

Appendix A

BICM Metric Derivation

In conventional BICM [14, 13, 41, 16], the demodulator produces a set of maximum-likelihood bit-metrics. A derivation of these metrics is presented next and is a summarized version of one presented in [14].

The transition probability density function of the channel is defined as

$$p_\theta(\mathbf{y}|\mathbf{x}) : \theta \in \mathbb{C}^{N'}; \mathbf{x}, \mathbf{y} \in \mathbb{C}^{N}$$

where $\mathbb{C}^N$, $\mathbb{C}^{N'}$, and θ represent N-dimensional Euclidean space, N'-dimensional Euclidean space, and the channel state respectively. θ is assumed to be independent of the $\mathbf{x}$ and the channel is assumed to be memoryless, so that

$$p_{\{\theta\}}(\{\mathbf{y}\}|\{\mathbf{x}\}) = \prod_{j=1}^{k} p_{\theta_j}(\mathbf{y_j}|\mathbf{x_j})$$

where the sequences $\{\theta\} = (\theta_1, \ldots, \theta_k)$, $\{\mathbf{x}\} = (\mathbf{x_1}, \ldots, \mathbf{x_k})$, and $\{\mathbf{y}\} = (\mathbf{y_1}, \ldots, \mathbf{y_k})$ and k is the sequence length. The likelihood function for $\mathbf{y}$ is

$$p_\theta\left(\mathbf{y}|\ \ell^i(\mathbf{x}) = b\right) = \sum_{\mathbf{x} \in S_i^b} p_\theta\left(\mathbf{y}|\mathbf{x}\right) P(\mathbf{x}|\ \ell^i(\mathbf{y}) = b)$$

where $\ell^i(\mathbf{x})$ denotes the ith bit of the label of $\mathbf{x}$ and S_i^b is the subset of the signals $\mathbf{x} \in S$ whose label has the value $b \in \{0, 1\}$ in position i.
Next, a uniform(equiprobable) input distribution is assumed.

$$p_\theta\left(\mathbf{y}|\ \ell^i(\mathbf{x}) = b\right) = 2^{-(M-1)} \sum_{\mathbf{x} \in S_i^b} p_\theta\left(\mathbf{y}|\mathbf{x}\right)$$

where M represents the number of points in the signal constellation. The log likelihood function is then obtained

$$\log\left\{p_\theta\left(\mathbf{y}|\ \ell^i(\mathbf{x}) = b\right)\right\} = \log\left\{2^{-(M-1)}\right\} + \log\left\{\sum_{\mathbf{x} \in S_i^b} p_\theta\left(\mathbf{y}|\mathbf{x}\right)\right\}$$

Finally, the constant term is removed from the equation and the BICM ML *bit* metric at time t is defined.

$$m_i(y_t^i, S_i^b; \rho_t^i) \quad = \quad \log\left\{\sum_{\mathbf{x} \in S_i^b} p_{\theta_t}\left(\mathbf{y_t}|\mathbf{x}\right)\right\} \tag{A.1}$$

where ρ represents the fading amplitude. It takes on a value of one for AWGN channels.

Because this metric may be too computationally complex for implementation, a log-sum approximation $\left(\log \sum_j z_j \simeq \max_j \log z_j\right)$ is often used. This sub-optimal approximation is appropriate as long as the sum in the left-hand side of the equation is dominated by a single term. The approximation is a good estimate at high signal-to-noise ratios (SNR) [13]. The sub-optimal BICM bit-metric is

$$m_i(y_t^i, S_i^b; \rho_t^i) \quad = \quad \log\left\{\max_{\mathbf{x} \in S_i^b} p_{\theta_t}\left(\mathbf{y_t}|\mathbf{x}\right)\right\}. \tag{A.2}$$

Appendix B

Turbo Codes

First discovered by Berrou *et al.* [42], turbo coding is a strategy by which two constituent codes are concatenated together in a parallel fashion - separated by an interleaver. An example of turbo encoder is presented in Figure B.1. The 16-state turbo encoder has two constituent recursive, systematic, convolutional codes - each with octal generators $37, 21$. This standard code has been shown to perform well in an AWGN environment [43] and is the one chosen for use in this work.

Turbo codes have been shown to perform within a dB of capacity [42] for large block sizes. The key ingredient for this remarkable performance lies in the iterative decoding strategy used.

B.1 Iterative-Decoding

In decoding of turbo codes, each constituent is soft decoded using input information which is the decoded information from the other code, as shown in Figure.

The decoding of each code is performed using the BCJR (or Forward/Backward Algorithm) [44]. A set of calculated messages are passed in both the forward and reversed directions of the trellis. At each depth of the trellis, a message (A_i for forward and B_j for reverse) is calculated based on the message at the previous depth(A_{i-1} and

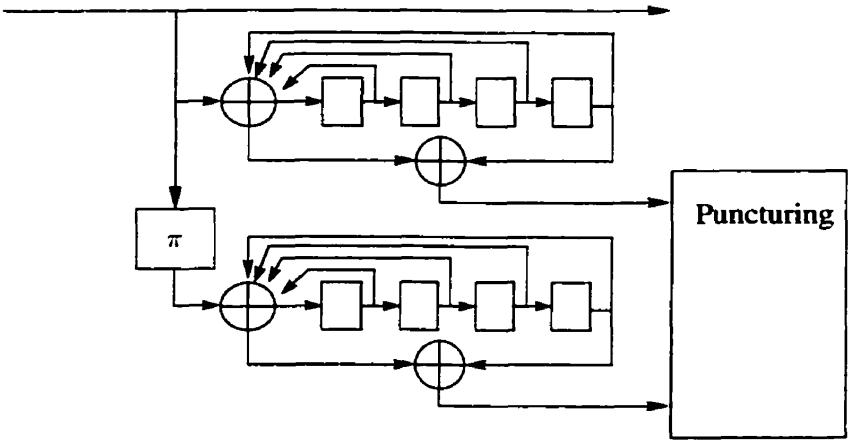


Figure B.1: Turbo Code: 16-state with identical rate-1/2 constituent RSC codes with octal generators 37, 21

B_{j-1}), and the branch symbol and bit probabilities. We will not go into details here but an example of the algorithm is shown in Figure B.3. The final output probabilities of each decoder are derived by multiplying the branch symbol probabilities at each depth by the forward/backward messages from adjacent vertices.

B.2 Interleaving

The objective of interleaving is take points which are adjacent to one (and correlated) and spread them out within time or space. This is accomplished by reordering the information within a block before transmission. The strategy used when reordering the information is important. Three examples of interleavers which are often used in Turbo Codes are as follows:

1. "Turbo"Interleavers - Bits are written row by row into a square matrix. Then they are read out column by column.

2. Random Interleavers - Bits are reordered in a uniform random fashion. Each bit has same probability of being placed in any of the new positions.

3. Spread Interleavers - Bits are reordered in uniform random fashion but are

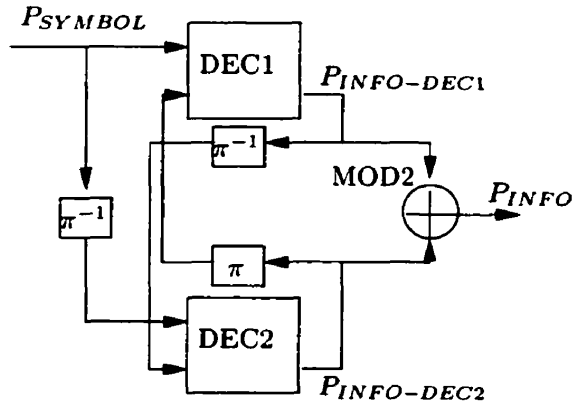


Figure B.2: Turbo Decoding where P_{SYMBOL}, $P_{INFO-DEC1}$ and $P_{INFO-DEC2}$ are vectors containing the probability values for the coded symbols, decoded bits for decoder 1 and decoded bits for decoder 2 respectively

constrained such that adjacent bits must be mapped to locations which are a minimum distance from one another. A spread interleaver is often discussed in terms of its minimum spread distance as discussed above.

In turbo codes, interleavers are central to the performance of turbo codes. They are key to ensuring good distance properties. It has been shown that interleavers which permutates the data in some random way are better the matrix based interleavers. For this reason, spread interleavers have been chosen for use in this work.

B.3 Puncturing

Often, systems will be required to provide multiple data rates. Puncturing is a strategy by which the same base code is used to provide multiple data rates. The advantage to doing this is that only one encoder/decoder pair needs to be implemented in hardware - reducing the cost and space required. In puncturing, the data is encoded as usual but only a percentage of the redundant bits generated by the encoder are actually sent. For example, if the base encoder is a systematic and rate-1/3, eight redundant bits will be generated for every four data bits which have been input. If four redundant bits are stripped off, a rate-1/2 code is produced. When decoding, the

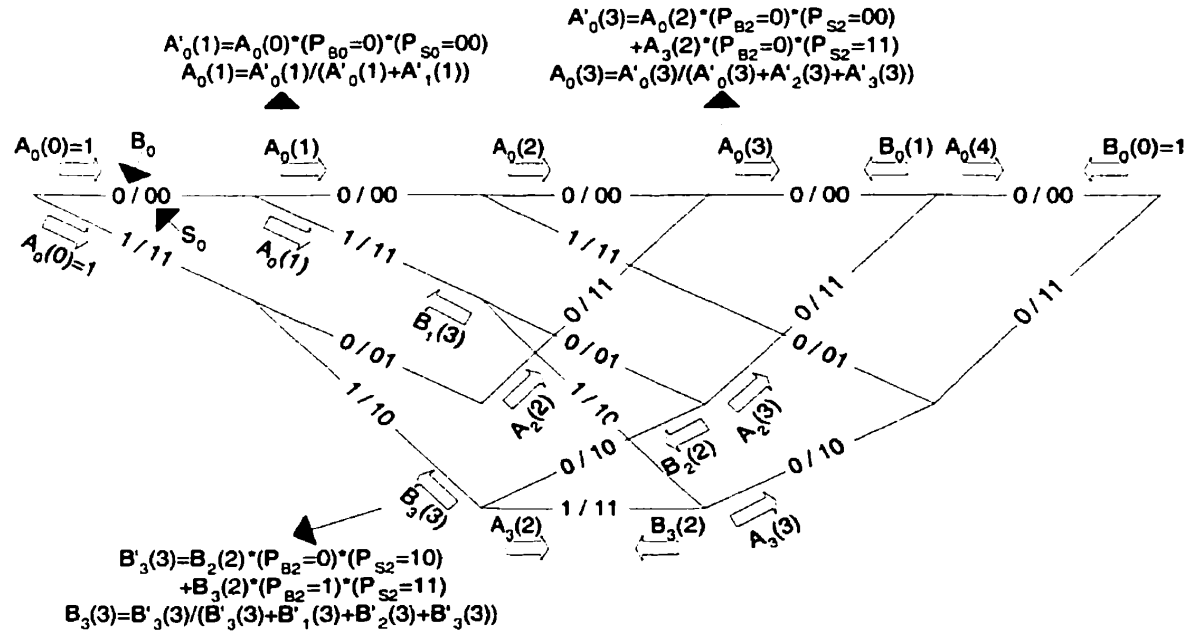


Figure B.3: Example: Forward Backward Algorithm on a 4-State Trellis

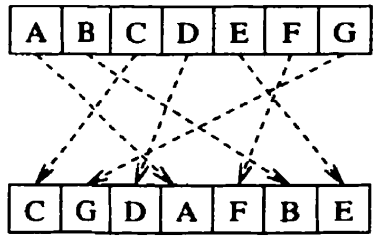


Figure B.4: Interleaving Example

same base decoder is used. Erasures are placed in the locations where the stripped bits should be. The puncturing masks used are presented in Table B.1 and have been determined to give good performance by Rowitch and Milstein [45].

Puncture Index	0	1	2	3	4	5
Punctured Code Rate	1/3	2/5	4/9	1/2	2/3	4/5
Systematic Bits	11111111	11111111	11111111	11111111	11111111	11111111
Parity1	11111111	11111110	00111110	00101110	00000110	00000010
Parity2	11111111	10111111	00011111	00011110	00000110	00000010

Table B.1: Turbo code puncturing masks based on a puncturing period of eight

Appendix C

Bit-Error-Rate (BER) Curves

The simulation results for the 16-state turbo code are illustrated in the figures presented in this section.

This code is punctured to achieve code rates of 1/3, 2/5, 4/9, 1/2, 2/3, and 4/5. The coded outputs are modulated using QPSK, 8PSK, or 16QAM. Organized by rate, each figure contains a plot of BER versus E_b/N_o for each of QPSK, 8PSK and 16QAM. In most cases, the results are presented for information block lengths of 1000 and 5000. Also the bit SNR required to achieve capacity for each of the modulation types is highlighted via a line on the plot (the leftmost line is QPSK, followed by 8PSK and then 16QAM). The equation used to calculate the SNR which achieves capacity (see [32]) is as follows:

$$\frac{E_b}{N_0} = \frac{E_s}{N_0} \cdot \frac{1}{R} = \frac{2^R - 1}{R} \tag{C.1}$$

where E_b, E_s, and N_0 are the average bit energy, average symbol energy and the noise power spectral density respectively. The channel coding rate, R, is defined as the punctured turbo code rate times the spectral efficiency of the modulation used. For example, if the code rate is 2/3 and the modulation is 8PSK, the channel coding rate is 2 bits/symbol/2D.

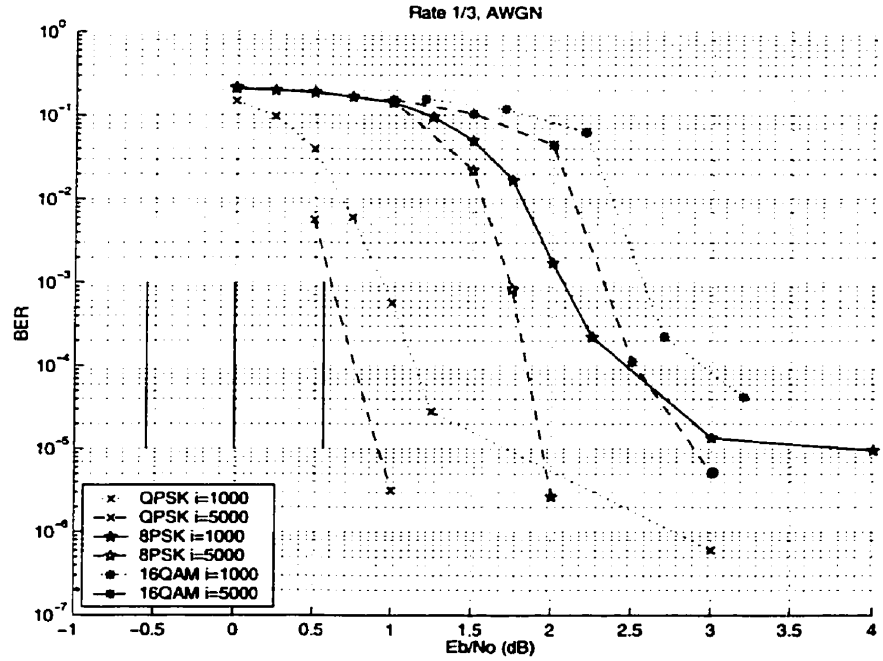


Figure C.1: BER for Rate-1/3 punctured turbo code

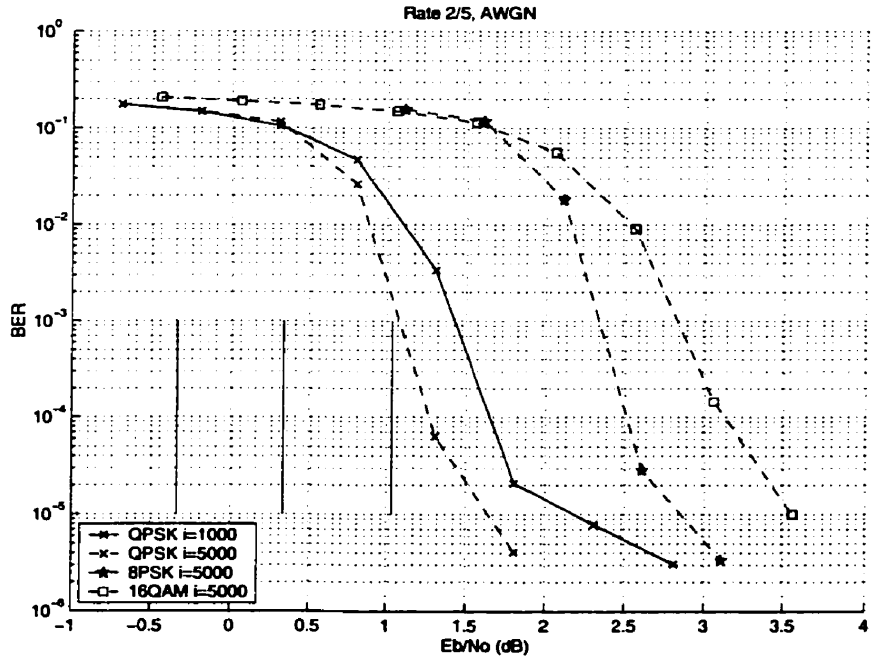


Figure C.2: BER for Rate-2/5 punctured turbo code

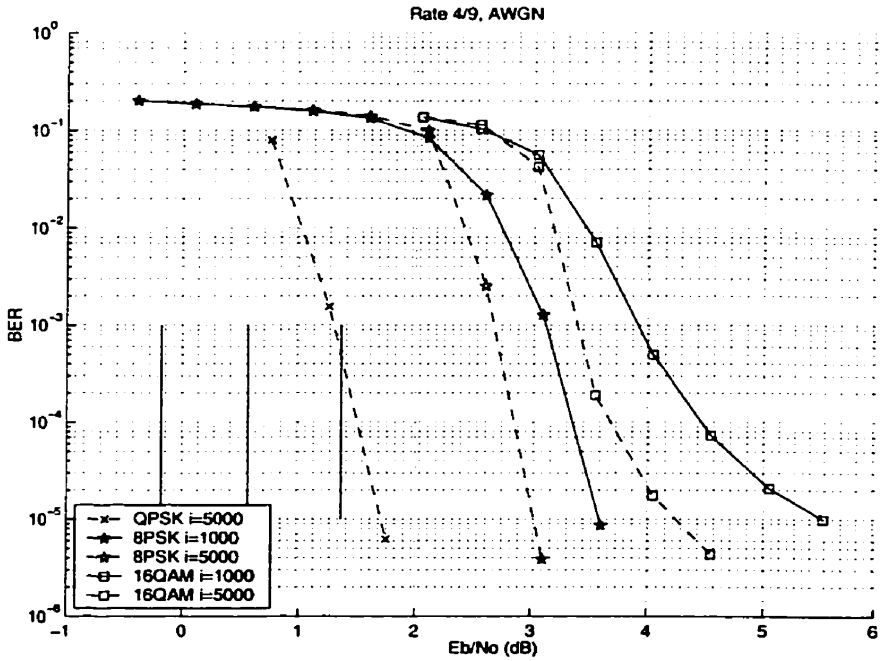


Figure C.3: BER for Rate-4/9 punctured turbo code

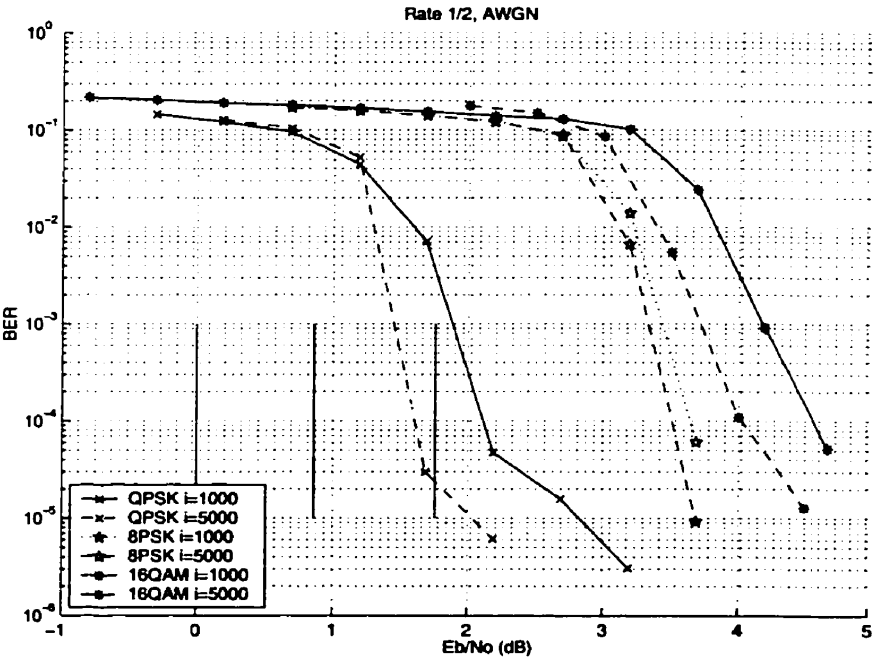


Figure C.4: BER for Rate-1/2 punctured turbo code

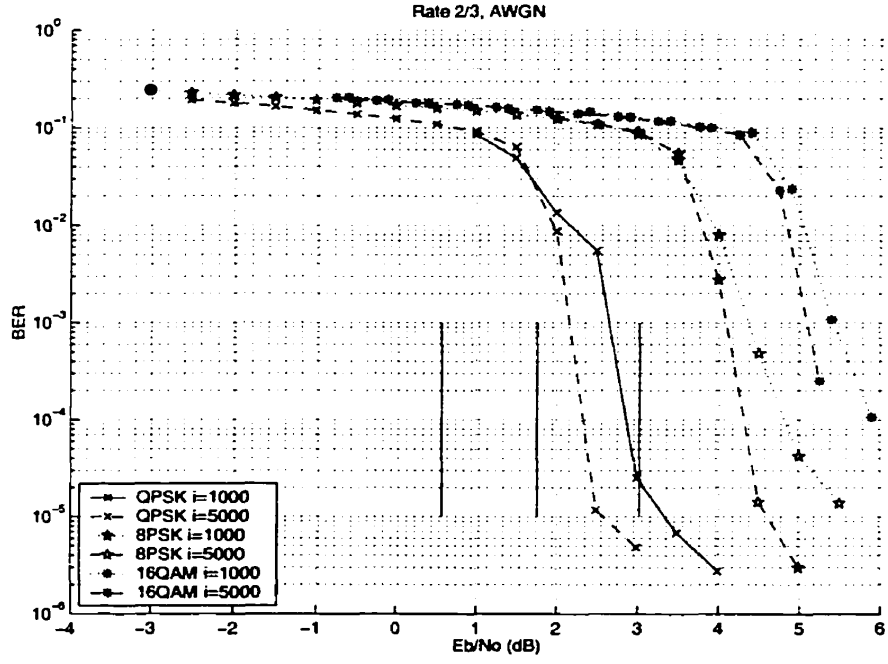


Figure C.5: BER for Rate-2/3 punctured turbo code

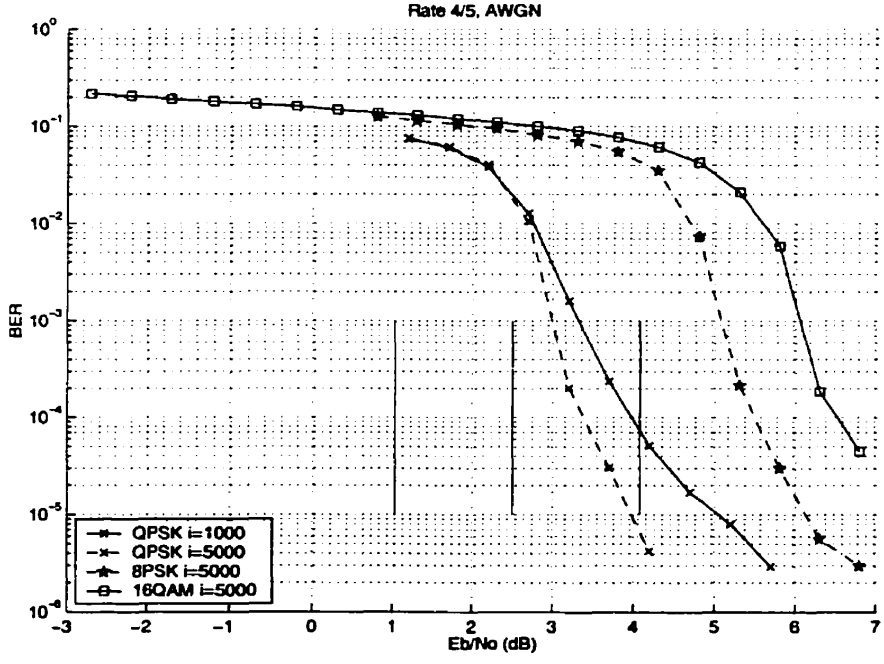


Figure C.6: BER for Rate-4/5 punctured turbo code

Appendix D

Decoding Complexity

The computational effort required to decode TCM, BICM, and BICM-ID(hard-decision feedback) using Viterbi and Turbo decoding is evaluated in this section. Specifically, the number of addition and multiplication operations required to decode an (n,k) code are calculated in the next two sub-sections. The results of the derivations are presented in Tables D.1-D.8. It can be readily seen that TCM is by far, the simplest with respect to the number of additions and multiplications required for decoding. The same ranking holds true for Turbo codes.

		Number of Additions - Viterbi - AWGN
TCM		$2\|E\| - \|V\| + 1$
BICM	ML	$((M-1)(n-1)+1)\|E\| - \|V\| + 1$
	sub-ML	$((M/2-1)(n-1)+1)\|E\| - \|V\| + 1$
BICM-ID	ML	$((M+\ell-2)(n-1)+1)\|E\| - \|V\| + 1$
	sub-ML	$((M/2+\ell-2)(n-1)+1)\|E\| - \|V\| + 1$

Table D.1: Number of additions required to decode an (n, k) convolutional code (Viterbi decoding) on an AWGN channel(with coherent detection) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

		Number of Multiplications - Viterbi - AWGN
TCM		$2\|E\|$
BICM	ML	$(M/2+1)\|E\|$
	sub-ML	$(M/2+1)\|E\|$
BICM-ID	ML	$(M/2+\ell)\|E\|$
	sub-ML	$(M/2+\ell)\|E\|$

Table D.2: Number of multiplications required to decode an (n,k) convolutional code (Viterbi decoding) on an AWGN channel(with coherent detection) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

		Number of Additions - Viterbi - Rayleigh
TCM		$2\|E\|-\|V\|+1$
BICM	ML	$((M-1)(n-1)+1)\|E\|-\|V\|+1$
	sub-ML	$((M/2-1)(n-1)+1)\|E\|-\|V\|+1$
BICM-ID	ML	$((M+\ell-2)(n-1)+1)\|E\|-\|V\|+1$
	sub-ML	$((M/2+\ell-2)(n-1)+1)\|E\|-\|V\|+1$

Table D.3: Number of additions required to decode an (n,k) convolutional code (Viterbi decoding) on a Rayleigh channel(with perfect CSI) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

D.1 Viterbi Decoding

Decoding of the BICM system using convolutional codes is generally done using the Viterbi [46] algorithm. The number of operations required to implement the Viterbi algorithm is well documented and a derivation is presented in McEliece [44]. Specifically, the number of addition and multiplication operations for a given Viterbi decoded code C are:

$$\begin{aligned} \textit{multiplications} &= |E| \\ \textit{additions} &= |E|-|V|+1 \end{aligned}$$

		Number of Multiplications - Viterbi - Rayleigh
TCM		$2\|E\|$
BICM	ML	$(M+1)\|E\|$
	sub-ML	$(M+1)\|E\|$
BICM-ID	ML	$(M+2\ell-1)\|E\|$
	sub-ML	$(M+2\ell-1)\|E\|$

Table D.4: Number of multiplications required to decode an (n,k) convolutional code (Viterbi decoding) on a Rayleigh channel(with perfect CSI) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

		Number of Additions - Turbo - AWGN
TCM		$\|E\|$
BICM	ML	$((M-1)(n-1))\|E\| + 2\ell(2\|E\|-2)$
	sub-ML	$((M/2-1)(n-1))\|E\| + 2\ell(2\|E\|-2)$
BICM-ID	ML	$((M+\ell-2)(n-1))\|E\| + 2\ell(2\|E\|-2)$
	sub-ML	$((M/2+\ell-2)(n-1))\|E\| + 2\ell(2\|E\|-2)$

Table D.5: Number of additions required to decode an (n,k) Turbo code (FBA decoding) on an AWGN channel(with coherent detection) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

where $|E|$ and $|V|$ are the number of edges and vertices in the trellis respectively.

D.2 Turbo Coding

A similar derivation of the complexity of turbo codes has been performed by Sauve [47]. The number of addition and multiplication operations for a given code which is Turbo decoded via the forward-backward algorithm for a rate-1/2 turbo code are:

$$\begin{aligned} \textit{multiplications} &= 2\ell(6|E|+2) \\ \textit{additions} &= 2\ell(2|E|-2) \end{aligned}$$

		Number of Multiplications - Turbo - AWGN
TCM		$\|E\| + 2\ell(6\|E\| + 2)$
BICM	ML	$(M/2)\|E\| + 2\ell(6\|E\| + 2)$
	sub-ML	$(M/2)\|E\| + 2\ell(6\|E\| + 2)$
BICM-ID	ML	$(M/2 + \ell - 1)\|E\| + 2\ell(6\|E\| + 2)$
	sub-ML	$(M/2 + \ell - 1)\|E\| + 2\ell(6\|E\| + 2)$

Table D.6: Number of multiplications required to decode an (n, k) Turbo code (FBA decoding) on an AWGN channel(with coherent detection) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

		Number of Additions - Turbo - Rayleigh
TCM		$\|E\|$
BICM	ML	$((M-1)(n-1))\|E\| + 2\ell(2\|E\| - 2)$
	sub-ML	$((M/2-1)(n-1))\|E\| + 2\ell(2\|E\| - 2)$
BICM-ID	ML	$((M+\ell-2)(n-1))\|E\| + 2\ell(2\|E\| - 2)$
	sub-ML	$((M/2+\ell-2)(n-1))\|E\| + 2\ell(2\|E\| - 2)$

Table D.7: Number of additions required to decode an (n, k) Turbo code (FBA decoding) on an AWGN channel(with coherent detection) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

where $|E|$ and $|V|$ are the number of edges and vertices in the trellis respectively. The variable ℓ represents the number of decoding iterations required.

D.3 Branch Metric Computation

Next, the computational effort required to calculate the branch-metrics will be estimated. In TCM, the ML symbol metrics are the same as the branch-metrics. In a Gaussian channel, the received symbol is subtracted from the estimated symbol and the result is squared. In trellis terms, this translates to $|E|$ multiplications and $|E|$ additions. In a Rayleigh channel with perfect CSI, the received signal is multiplied by

		Number of Multiplications - Turbo - Rayleigh
TCM		$\|E\| + 2\ell(6\|E\| + 2)$
BICM	ML	$(M)\|E\| + 2\ell(6\|E\| + 2)$
	sub-ML	$(M)\|E\| + 2\ell(6\|E\| + 2)$
BICM-ID	ML	$(M + 2\ell - 1)\|E\| + 2\ell(6\|E\| + 2)$
	sub-ML	$(M + 2\ell - 1)\|E\| + 2\ell(6\|E\| + 2)$

Table D.8: Number of multiplications required to decode an (n, k) Turbo code (FBA decoding) on a Rayleigh channel(with perfect CSI) where $|E|$ and $|V|$ represent the number of edges and vertices in the trellis respectively, M represents the constellation size, and ℓ represents the number of decoding iterations

the fade amplitude before it is subtracted from the estimated symbol. This translates $2|E|$ multiplications and $|E|$ additions.

In BICM, the symbol metrics are used to form bit-metrics which, are used in turn, to form the branch-metrics. The calculation of the optimal(ML) bit metric (A.1) is complicated. As in TCM, the received signal must be subtracted from the estimated constellation signal point. This same calculation is done for different signal points for a total of $M/2$ times (where M represents the constellation size). The results of these calculations are summed together($M/2 - 1$ additions) to form the bit-metric. In the next step, n bit-metrics are added together($n-1$ additions) to form the branch-metric. In trellis terms, this translates to $(M/2)|E|$ multiplications and $(M - 1)(n - 1)|E|$ additions for a Gaussian channel and $M|E|$ multiplications and $(M - 1)(n - 1)|E|$ additions for a Rayleigh channel with perfect CSI.

The sub-optimal BICM bit-metric (A.2) calculation is simpler. It requires the same $M/2$ distance calculations but, in this case, the optimal metric the results are not summed together. The calculation requires $(M/2)|E|$ multiplications and $(M/2 - 1)(n - 1)|E|$ additions for a Gaussian channel and $M|E|$ multiplications and $(M/2 - 1)(n - 1)|E|$ additions for a Rayleigh channel with perfect CSI. In general, it can be seen that sub-optimal metric calculation requires half as many addition operations as the optimal(ML) metric calculation.

BICM-ID requires the same number of operations as BICM for first pass decoding. On subsequent decoding cycles, calculation of the bit-metric is less complex. The decoder feeds back an estimated signal constellation point. For this reason the distance calculation needs is performed only once on the $\ell - 1$ subsequent decoding cycles(ℓ represents the total number of iterations). In total, the optimal bit-metric calculation requires $((M/2) + \ell - 1)|E|$ multiplications and $(M + \ell - 2)(n - 1)|E|$ additions for a Gaussian channel and $(M + 2\ell - 2)|E|$ multiplications and $(M + \ell - 2)(n - 1)|E|$ additions for a Rayleigh channel with perfect CSI.

Sub-optimal BICM-ID metric calculation is similar. The sub-optimal bit-metric calculation requires $((M/2) + \ell - 1)|E|$ multiplications and $(M/2 + \ell - 2)(n - 1)|E|$ additions for a Gaussian channel and $(M + 2\ell - 2)|E|$ multiplications and $(M/2 + \ell - 2)(n - 1)|E|$ additions for a Rayleigh channel with perfect CSI.

Bibliography

[1] Dataquest Inc., "Gartner dataquest predicts half a billion mobile phones should be sold during 2001," Tech. Rep., March 2001, http://www3.gartner.com.

[2] V. H. MacDonald, "The cellular concept," *Bell System Technical Journal*, vol. 58 (1), pp. 15–41, 1979.

[3] E. Sousa, "CDMA mobile radio communications, - ece1543," Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Mar. 2001.

[4] R. C. Dixon, *Spread Spectrum Systems*, Wiley, New York NY, 1st edition, 1984.

[5] D. L. Schilling, R. L. Pickholtz, and L. M. Milstein, "Spread spectrum goes commercial," *IEEE Spectrum*, pp. 40–45, Aug. 1990.

[6] G. R. Cooper and R. W. Nettleton, "A spread-spectrum technique for high capacity mobile communications," *IEEE Trans. on Veh. Tech.*, vol. VT-7(4), pp. 39–40, Nov. 1978.

[7] E. Sousa, "Class notes for mobile communications - ECE1543," Edward S. Rogers Sr. Department of Electrical and Computer Engineering,University of Toronto, Mar. 2001.

[8] IMT2000, "The road to IMT2000," Tech. Rep., 2001, http://www.itu.int/imt.

[9] R. Price and P. E. Green, "A communication technique for multipath channels," *Proc. IRE*, vol. 46, pp. 555–570, March 1958.

[10] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, May 1982.

[11] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.

[12] C.-E. Sundberg and N. Seshadri, "Coded modulations for fading channels: An overview," *Europ. Trans. Telecomm. (ETT)*, pp. 309–324, May-June 1993.

[13] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, 1992.

[14] G. Caire, G. Taricco, , and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[15] U. Wachsmann, J. Huber, and P. Schramm, "Serial concatenation of interleaved codes: Performance and analysis, design, and iterative decoding," *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, p. 5, Aug. 1998.

[16] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Select. Areas Commun.*, vol. 17(4), pp. 715–724, Apr. 1999.

[17] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding using soft feedback," *Electronic Letters*, vol. 34(10), pp. 942–943, May 1998.

[18] T. S. Rappaport, *Wireless Communications: Principles and Practices*, Prentice Hall, Englewood Cliffs NJ, 1996.

[19] S. Jamali and T. Le-Ngoc, *Coded-modulation techniques for fading channels*, Kluwer, New York NY, 1994.

[20] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York NY, 3rd edition, 1996.

[21] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell System Technical Journal*, vol. 47, pp. 957–1000, 1968.

[22] M. J. Gans, "A power spectral theory of propagation in the mobile radio environment," *IEEE Trans. on Veh. Tech.*, vol. VT-21, pp. 27–38, Feb. 1972.

[23] J. I. Smith, "A computed generated multipath fading simulation for mobile radio," *IEEE Trans. on Veh. Tech.*, vol. VT-24(3), pp. 39–40, Aug. 1975.

[24] H. Mahvidi, "Performance of latency constrained turbo-like codes over correlated fading channels," M.A.Sc. thesis, University of Toronto, Toronto ON, 2000.

[25] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, Prentice-Hall, Upper Saddle River NJ, 3rd edition, 1996.

[26] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Int. Conf. Communication*, vol. 2, pp. 858–863, May 1999.

[27] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan. 1997.

[28] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Englewood Cliffs NJ, 1983.

[29] R. J. Mantha, "Hybrid automatic repeat request schemes using turbo codes and low density parity check codes," M.A.Sc. thesis, University of Toronto, Toronto ON, 1999.

[30] S. Benedetto, M. Mondin, and G. Montorsi, "Performance evaluation of trellis-coded modulation schemes," *Proc. IEEE*, vol. 82(6), June 1994.

[31] F. Kschischang, "Class notes for error control codes - ECE1501," Esdward S. Rogers Sr. Department of Electical and Computer Engineering, University of Toronto, Oct. 1999.

[32] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York NY, 1991.

[33] G. Taricco G. Caire and E. Biglieri, "Capacity of bit-interleaved channels," *Electronic Letters*, vol. 32, pp. 1060–1061, June 1996.

[34] B. Melis and G. Romano, "UMTS W-CDMA: Evaluation of radio performance by means of link level simulations," *IEEE Pers. Commun.*, pp. 42–49, June 2000.

[35] 3GPP, "Selection procedures for the choice of radio transmission technologies of the UMTS," *TS UMTS 30.03 version 3.1.0*, Nov. 1997.

[36] D. Brennan, "Linear diversity combining techniques," *Proc. IRE*, vol. 47, pp. 1075–1102, June 1959.

[37] M. S Alouini, S. Kim, and A. Goldsmith, "Rake reception with maximum-ratio and equal-gain combining for DS-CDMA systems in Nakagami fading," *IEEE Int. Conf. Univ. and Pers. Comm.*, vol. 2, pp. 708–712, 1997.

[38] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice Hall, Englewood Cliffs NJ, 1995.

[39] S. Kallel, "Efficient hybrid ARQ protocols with adaptive forward error correction," *IEEE Trans. Commun.*, pp. 291–289, Feb. 1994.

[40] S. N. Crozier, "New high-spread high-distance interleavers for turbo codes," *Proc. Bienn. Symp. Info. Theory*, pp. 3–7, May 2000.

[41] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, pp. 169–171, Nov. 1997.

[42] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," *IEEE Int. Conf. Communication*, pp. 1064–1070, May 1993.

[43] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance and analysis, design, and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 909–926, May 1988.

[44] R. J. McEliece, "On the BCJR trellis for linear block codes," *IEEE Trans. Inform. Theory*, vol. 42(4), pp. 1072–1092, July 1996.

[45] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Commun.*, vol. 48(6), pp. 948–959, June 2000.

[46] A. Viterbi, "Error bounds for convolutional codes and as asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

[47] P.-P. Sauve, "Multibit turbo decoding of turbo codes," M.A.Sc. thesis, University of Toronto, Toronto ON, 1998.

[48] R. J. McEliece, "The trellis complexity of convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42(4), pp. 1855–1864, Nov. 1996.

[49] J. Wozencraft and B. Reiffen, *Sequential Decoding*, MIT Press, Cambridge MA, 3rd edition, 1961.

[50] G. Forney, *Concatenated Codes*, MIT Press, Cambridge MA, 1966.