UNIVERSITY OF CALIFORNIA, SAN DIEGO

Time Varying Channels: Characterization, Estimation, and Detection

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Nathan D. Ricklin

Committee in charge:

Professor James R. Zeidler, Chair
Professor Robert R. Bitmead
Professor William S. Hodgkiss
Professor Lawrence B. Milstein
Professor Bhaskar D. Rao

2010

The dissertation of Nathan D. Ricklin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2010

DEDICATION

To choosing the more adventurous of two options.

EPIGRAPH

Hey man of science with your perfect rules of measure,
Can you improve this place with the data that you gather?
—Bad Religion

Summit or death, either way I win.
—Rob Slater

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My academic experience at UCSD has been closely intertwined with my climbing career. When I first arrived here in San Diego, as when I began climbing, I asked myself questions like "am I good enough for this?"

My closest climbing partners have been my closest academic peers, and we all pushed each other in both arenas. We bounced ideas off of each other, had late-night whiteboard sessions. And we climbed bigger and badder things, pushing our limits on unknown ground at altitude in the Sierra Nevada.

As years passed by, my youthful self-doubt was soon replaced with confidence. Confidence that my work and ideas were academically sound, that I wasn't a complete idiot. And confidence that we could execute our craziest ideas in the mountains, turn thought into action. We could climb for 29 hours non-stop, we could pick a line in a photograph and show up and climb it, we could fly half-way round the world to climb in the mountains we always dreamed of. We could even go to Mongolia to legitimitely search for the Tomb of Genghis Khan.

I have to thank my friends for this, because this growth didn't come solely from within.

I'd like to thank my advisor, Professor Zeidler, for picking me out of a list of applicants believing that I could do a good job, and then for always nudging me in the right direction and giving me the freedom to work on my own ideas, on my own time, at my own pace; Professor Milstein and Professor Rao for their excellent classes in communications and signal processing which inspired my research; Professor Bitmead and Professor Hodgkiss for agreeing to serve on my committee.

The material in this thesis has either appeared in or been submitted to journal publications and conference proceedings. The dissertation author was the primary investigator and author of this material.

Chapter 3, in part, is a reprint of the paper "Data-Aided Joint Estimation of Carrier Frequency Offset and Frequency-Selective Time-Varying Channel", Nathan D. Ricklin and James R. Zeidler, as it appeared in the proceedings of IEEE ICC, Beijing, 2008.

Chapter 4, in part, is a reprint of the papers "Block Detection of Multiple Symbol DPSK in a Statistically Unknown Time-Varying Channel" Nathan D. Ricklin and James R. Zeidler, as it appeared in the proceedings of IEEE ICC, Dresden, 2009; "Multi-

ple Symbol Differential Detection of MPSK over Statistically-Unknown, Frequency-Flat, Time-Varying Channels" Nathan D. Ricklin and James R. Zeidler, accepted for publication in the IEEE Transactions on Communications.

Chapter 5, in part, has been submitted for publication to IEEE Globecom 2010, under the title "Constant Magnitude Correlative Coding for ICI Reduction and Blind Detection in OFDM" Nathan D. Ricklin and James R. Zeidler, March 2010. Chapter 5, in part, has also been submitted for publication to the IEEE Transactions on Communications, under the title "Constant Magnitude Correlative Coding for OFDM: ICI Reduction and Low-Complexity Blind Detection" Nathan D. Ricklin and James R. Zeidler, May 2010.

VITA

2003	B. S. in Electrical Engineering, University of Maryland, College Park
2007	M. S. in Electrical Engineering, University of California, San Diego
2010	Ph. D. in Electrical Engineering, University of California, San Diego

PUBLICATIONS

N. Ricklin and J.R. Zeidler, "Data-Aided Joint Estimation of Carrier Frequency Offset and Frequency-Selective Time-Varying Channel", *IEEE International Conference on Communications*, Beijing, 2008.

N. Ricklin and J.R. Zeidler, "Block Detection of Multiple Symbol DPSK in a Statistically Unknown Time-Varying Channel", *IEEE International Conference on Communications*, Dresden, 2009.

N. Ricklin, "Cleaver: Cleavage Dreamer and Southeast Arete" *American Alpine Journal*, New York, NY, American Alpine Club, pp 114-115, 2009.

N. Ricklin and J.R. Zeidler, "Multiple Symbol Differential Detection of MPSK over Statistically-Unknown, Frequency-Flat, Time-Varying Channels", *IEEE Transactions on Communications*, in submission, 2010.

N. Ricklin and J.R. Zeidler, "Constant Magnitude Correlative Coding for ICI Reduction and Blind Detection in OFDM", *IEEE Global Communications Conference*, in submission, 2010.

N. Ricklin and J.R. Zeidler, "Constant Magnitude Correlative Coding for OFDM: ICI Reduction and Low-Complexity Blind Detection", *IEEE Transactions on Communications*, in submission, 2010.

N. Ricklin, "Mt. Langley: Horizontal Thought Movement" *American Alpine Journal*, New York, NY, American Alpine Club, 2010.

ABSTRACT OF THE DISSERTATION

Time Varying Channels: Characterization, Estimation, and Detection

by

Nathan D. Ricklin

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California, San Diego, 2010

Professor James R. Zeidler, Chair

This dissertation is concerned with digital communications systems operating over channels that vary in time due to mobility of the transmitter or receiver. Velocity of the transmitter or receiver causes a Doppler shift or smearing in frequency which hinders parameter estimation and degrades data detection performance.

To aid in our goal of improved estimation and detection performance, we discuss several characterization methods, both probabilistic and deterministic, that can accurately model the time-varying channel and that enable efficient parameter estimation and data detection.

The first topic we address is that of carrier frequency offset estimation in the presence of a time-varying frequency-selective channel. The scheme is data-aided, meaning that a block of pilot data known to the receiver is first transmitted through the channel. Joint estimators of both the CFO and the time-varying frequency-selective channel are developed using both probabilistic and deterministic channel models. The probabilistic CFO estimator is shown to achieve the Cramer-Rao Lower Bound, and the deterministic estimator can perform well, but depends heavily on the choice of channel parameterization in the deterministic model.

The second topic we focus on is concerned with blind data detection over time-varying channels. The problem we consider is recovery of a short block of time-domain data symbols transmitted over a channel with time-varying multiplicative distortion. The receiver does not know the data, the channel, or any statistics that could parameterize a probabilistic description of the channel. We design data detection algorithms based on a deterministic parameterization of the channel that can be implemented efficiently with the sphere decoding algorithm. Additionally we provide techniques for a priori selection of the channel parameterization that result in near-optimal performance over a wide range of channel conditions.

In our final topic we consider blind detection over a frequency-selective time-varying channel. A short block of symbols is transmitted in the frequency domain, as either an OFDMA frequency allocation or as a subset of a larger OFDM symbol. The multiplicative distortion across the symbols is varying, due to frequency-selectivity. In addition, the transmitted symbols interfere with one another due to the channel time-variation, manifested as intercarrier interference. We introduce a novel method of correlatively coding the symbols across frequency before transmission that reduces the power of the ICI at the receiver. The scheme results in transmitted symbols that are constant in magnitude, which enables the application of efficient detection algorithms also based either probabilistic or deterministic parameterization of the channel. The resulting blind detector is efficient and can outperform an OFDM system which knows the channel perfectly.

Chapter 1

Introduction

1.1 Complexity of Communications Channels

As communications channels become more and more complex, so do the models that characterize them, and so must the techniques used to recover data transmitted through them. Here we introduce increasingly more complicated types of communications channels and briefly discuss the implications of that increased complexity on the requirements for estimation and detection.

1.1.1 The AWGN Channel

Figure 1.1: The AWGN channel.

Conceptually, the simplest communication channel is the additive white Gaussian noise (AWGN) channel. As illustrated in Fig. 1.1, the channel consists of just a single path from the transmitter to the receiver, with some additional thermal noise at the receiver. It's the simplest because it's easy to analyze, and estimation and detection methods are also easy to derive and efficient to implement.

Another property of the AWGN channel is separability; signals transmitted on orthogonal frequencies, on orthogonal time slots, or on orthogonal codes are also orthogonal at the receiver. It is this property that allows efficient implementation of detection

algorithms. The optimal detection method of a transmitted sequence of symbols is to optimally detect each symbol separately.

1.1.2 The Multipath Fading Channel

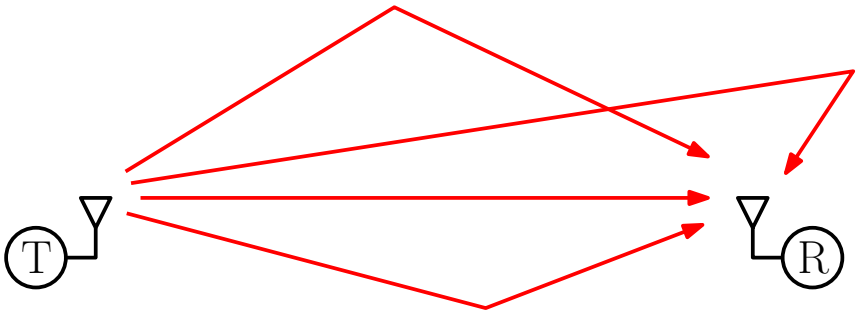

Figure 1.2: The multipath fading channel.

In more complex environments where there are scatterers surrounding the receiver reflecting multiple copies of the transmitted signal, a simple line-of-sight model such as the AWGN channel does not suffice to characterize the channel. Each path that the transmitted signal follows on its way to the receiver is a different length, and so the signal arriving from each path arrives with a different phase. The receiver then sees a sum of many copies of the same signal, each with different phases. The paths all sum constructively or destructively, and the receiver effectively sees a single copy of the transmitted signal with a certain phase and a certain gain, depending on the summation of the signal across all the received paths.

The gain caused my multiple paths degrades the performance of communications systems, since the times that it is very small lower the effective SNR and hurt performance more than the times that it is very large helps performance.

Additionally, the unknown gain introduces the need for channel estimation, since the gain must be known for optimal signal detection in noise.

1.1.3 The ISI Channel

If a portion of the path lengths of the multipath fading channel are very long, then copies of the transmitted signal can arrive after one symbol duration, effectively interfering with detection of the next symbol. This is known as intersymbol interference (ISI), and can degrade system performance if not dealt with. When ISI is present,

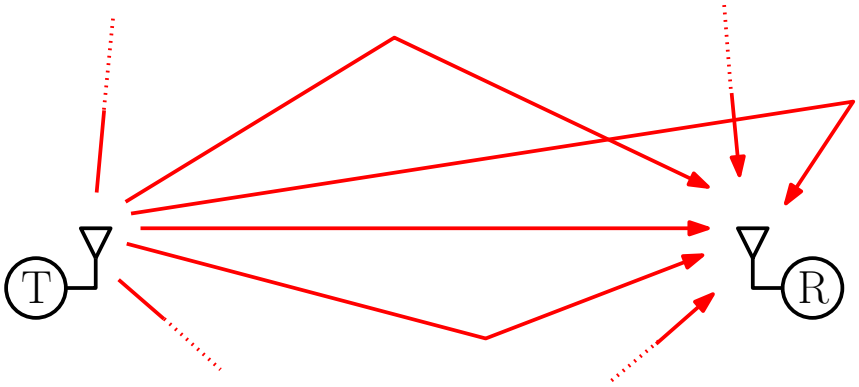

Figure 1.3: The ISI channel.

the channel is no longer separable, meaning that the optimal detection of a sequence of transmitted symbols must be done jointly. Methods of dealing with ISI include employing a RAKE receiver, which optimally combines the copies of the received signal that arrive at different times; and various equalization methods, which are typically filters that attempt to "undo" the ISI of the channel.

Perhaps the best technique for dealing with ISI is the signaling technique known as OFDM: orthogonal frequency-division multiplexing. The principle behind OFDM is that while ISI channels are not separable in time, they *are* still separable in frequency, so that data placed on separate frequencies remains orthogonal. For optimal detection with OFDM, the separate symbols can be independently detected, but unknown frequency-domain gains must first be estimated.

1.1.4 Motion through the Channel

The constructive and destructive interference seen by the receiver is different at each point in space. If the receiver is moving, then at each point in time it will see a different signal interference profile, and the apparent phase and gain of the channel will accordingly be different. The rate at which the channel gain and phase change depends on the velocity of the transmitter or receiver.

Another way to look at it is that the receiver sees an apparent Doppler shift that is different for each propagation path, depending on the angle of incidence of that path in relation to the direction of travel. For example, paths that the receiver is traveling directly into will experience an apparent positive Doppler shift (increase in frequency), and paths the receiver is traveling away from will experience a negative Doppler shift

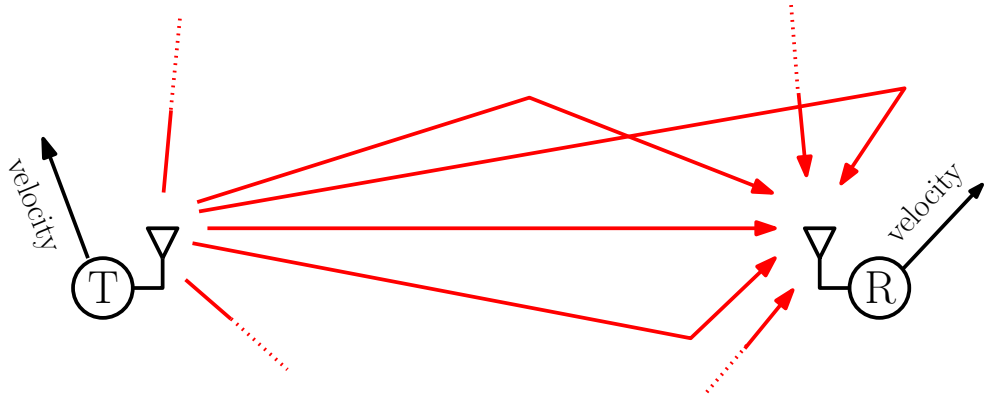

Figure 1.4: Node mobility creates channel time-variation.

(decrease in frequency). Propagation paths arriving at angles in between will experience a Doppler shift somewhere in between. The effect is that the transmitted signal appears to have been smeared in frequency.

The effect of time-variation is that the channel is no longer separable in frequency. Even using OFDM, due to the frequency-smearing effect of the various Doppler shifts, optimal symbol detection must be performed jointly. On top of that, all channel estimates must be updated as the channel changes, creating more effort for a receiver trying to keep accurate channel estimates and detect the transmitted data.

1.1.5 Carrier Frequency Offset

Carrier frequency offset (CFO) is a result of mismatch between the local oscillators of the transmitter and receiver of the communication system. After demodulation at the receiver, this mismatch results in the addition of a time varying phase of the received signal, on top of the varying phase and amplitude already present due to the channel time-variation. This phase must be estimated and then compensated for, adding an extra degree of coupling between all the transmitted symbols and the time-varying channel state.

1.1.6 Joint Estimation and Detection

We have described several channel scenarios in order of increasing complexity, and at each stage it becomes more clear that the role of the receiver to decide what data has been transmitted also becomes more complex. The receiver's two tasks are to

estimate the channel in some way, and then to detect the data, and both increase in complexity as the channel conditions degrade.

For the AWGN channel, no channel estimation is required. For the multipath fading channel, a single parameter (including both amplitude and phase) must be determined. For the ISI channel, channel estimates must be taken at each frequency. Adding transceiver motion into the equation, a full frequency-matrix must be known and finally, when CFO is present, an additional offset parameter must first be extracted from the system.

Similarly, detection algorithm complexity increases with the channel complexity. In the AWGN channel and the multipath fading channel, a sequence of transmitted data is optimally detected by deciding upon each symbol individually. In the ISI channel, this separability of signals is still possible–in the frequency domain. But in the time-varying channel or when CFO is present the sequence of transmitted symbols must be estimated jointly, which requires an exponential increase in detection algorithm computation power.

Estimation of the channel requires characterization of the channel, with an assumed probabilistic or deterministic model. The choice of model depends on the actual channel conditions, but also plays a significant role in determining the specific detection algorithm that can be used, and in its complexity. For this reason, the choice of channel model is even more critical in channels for which the transmitted data must be detected jointly.

As channels become more complex, a need arises for characterization methods and signaling techniques that allow efficient implementation of their corresponding detection algorithms. Thus, the goal of this dissertation is twofold: to determine channel characterization methods that both accurately describe the physical channel and lend themselves to simple detection techniques, and to develop signaling methods that are robust to difficult channel conditions and can also be efficiently detected.

1.2 Dissertation Overview

1.2.1 Time-Varying Channel Characterization

In Chapter 2 we start from basic principles and derive the mathematical models that represent the physical channel that we use throughout this dissertation. We begin with the lowpass representation of signals and systems and quickly arrive at the discrete-

time model of a time-varying frequency-selective channel.

We then show how the lowpass equivalent gain induced by a multipath-fading channel may be considered a complex Gaussian random variable. When the channel is time-varying, smearing in frequency results. As an example, Fig. 1.5 illustrates the spectrum of $s(t)$, a pure sinusoid transmitted at the carrier frequency f_c. Due to relative motion of the transmitter and receiver, the spectrum of the received signal is smeared in frequency, as illustrated in Fig. 1.6. We derive the associated time-correlation of the channel and show how it depends on the parameter f_dT, the normalized Doppler frequency.

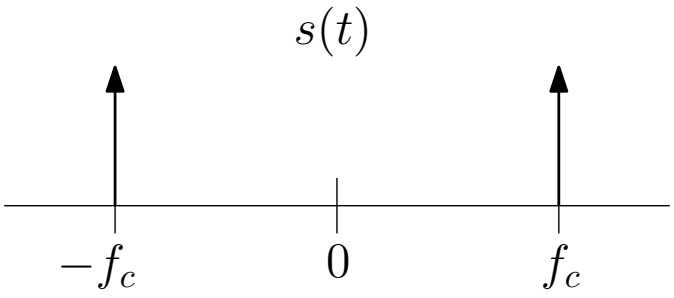

Figure 1.5: A transmitted signal $s(t)$.

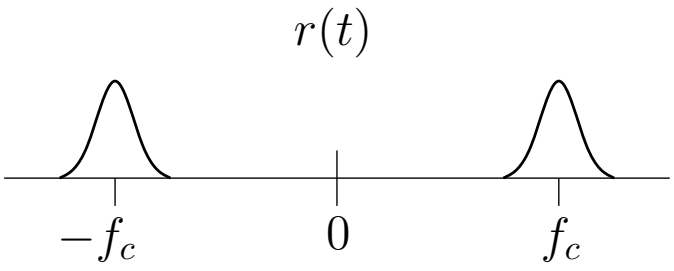

Figure 1.6: Received signal with Doppler broadening.

We then examine the normalized Doppler parameter, and illustrate how it depends on vehicle speed, carrier frequency, and symbol time-duration as well as give examples of specific combinations of these parameters which result in a certain value of f_dT.

Next, we show how carrier frequency offset (CFO), caused by a mismatch between the local oscillators of the transmitter and receiver, results in a multiplicative linearly-varying phase applied to the discrete-time signal at the receiver. Fig. 1.7 illustrates the spectrum of the received signal when CFO is present. The effects of CFO are an apparent

shift in carrier frequency of the received signal, and can also be seen as a special case of time-variation.

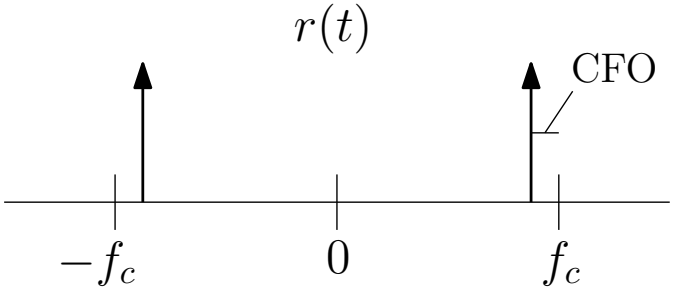

Figure 1.7: Received signal with Carrier Frequency Offset.

We then describe some deterministic channel modelling methods in relation to our preferred method, normalized Chebychev polynomials. Fig. 1.8 illustrates the basic idea behind our channel parameterization. The channel is approximated by a linear combination of a set number of increasingly varying basis functions defined by the Chebychev polynomials.

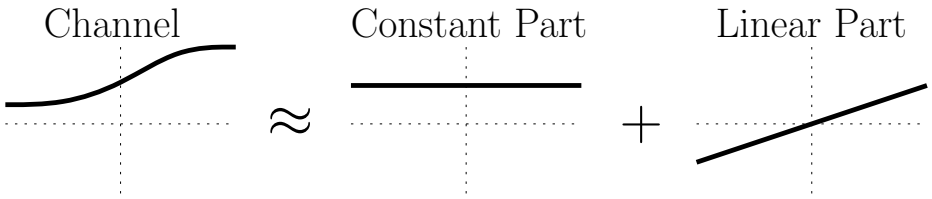

Figure 1.8: Parameterizing the time-varying channel.

Finally, we place OFDM in the discrete-time channel framework and observe the effect of channel time-variation. When the channel is time-varying, extra interference terms appear. These terms represent intercarrier interference: the output of the channel at one subcarrier depends on all the transmitted data symbols, not just the symbol on that specific subcarrier.

1.2.2 CFO Estimation

In Chapter 3 we address the issue of extraction of the CFO, itself a special case of time-variation, from a signal transmitted across a time-varying frequency-selective channel. If the CFO can be successfully estimated, a correction can be applied to the

received signal before any other processing, minimizing any performance degradation due to its presence.

We consider a pilot-aided approach: a pilot sequence perfectly known to the receiver is transmitted through the channel. This is then a first-order solution to the problem of CFO estimation when the data is not known, and serves as a bound on the achievable performance. We could also consider this technique to be part of a system that adapts and iterates over time, estimating data then using the estimated data to determine the CFO for a first attempt to estimate the next batch of data.

For the purpose of estimation, we consider the channel only in the time-domain. Thus we have NL unknown parameters, where N is the is the number of symbols over which we observe the channel output, and L is the number of channel taps, i.e. the length of the ISI of the channel.

Given the known sequence of transmitted pilots, we derive novel joint estimators of both the CFO and the time-varying frequency-selective channel (the NL channel parameters) for two channel models: probabilistic and deterministic.

The probabilistic model assumes that the channel is Rayleigh fading (thus complex Gaussian) and that the time-correlation is known. Bayesian estimation theory leads to an estimator that is shown to achieve the Cramer-Rao lower bound, which is the minimum variance attainable by any unbiased estimator, assuming that our channel model is correct.

If the probabilistic model cannot be assumed, either because the Rayleigh fading assumption does not fit reality or the channel statistics are not known, we resort to a deterministic modelling method based on parameterizing the channel with discrete Chebychev polynomials. The channel parameterization reduces the dimensionality of the problem from estimating NL unknown parameters to IL unknown parameters, where I is the number of basis functions that characterizes each of the L time-sequences that define the channel.

The performance of the deterministic estimator can also be good, but is sensitive to over or underparameterization of the channel. Fig. 1.9 illustrates CFO estimation performance for both schemes for a normalized Doppler of f_dT (the time duration T refers to the entire block of symbols). If too few polynomial functions are used in the channel parameterization, error floors result since the channel model cannot fully capture the actual channel time-variation. On the other hand, if too high a parameterization

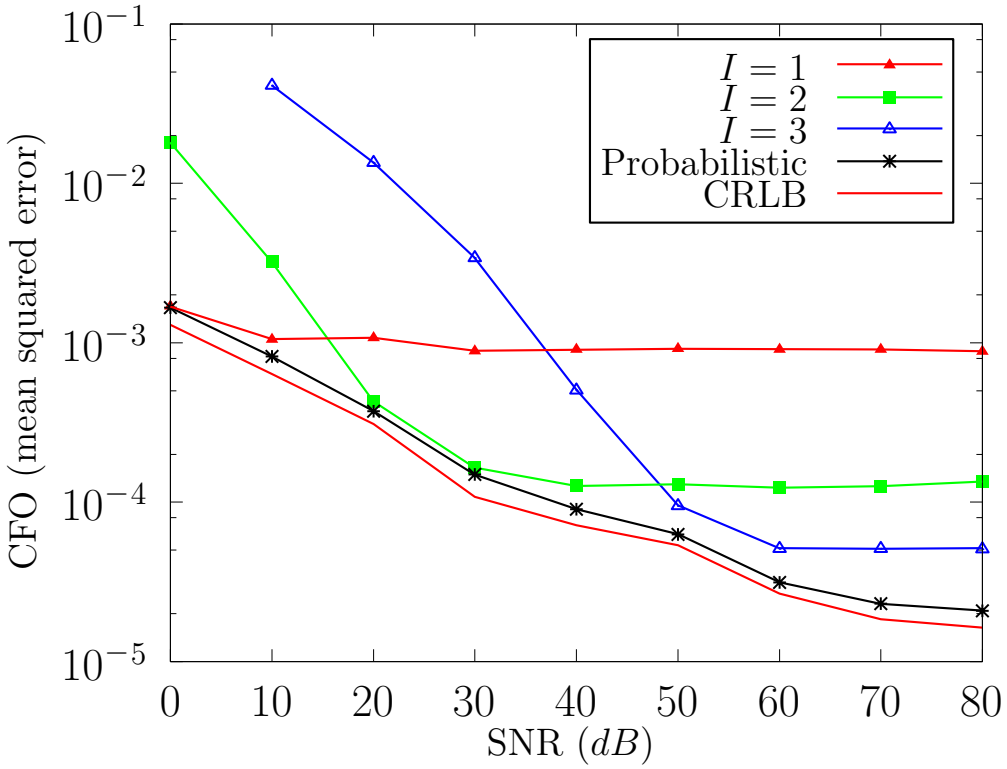

Figure 1.9: CFO Estimator performance vs. SNR for $f_dT = 0.1$

order is chosen, the error floor is indeed lower but performance is worse at low SNR. Obviously at some operating point the higher order basis functions are excited more by noise than by the actual time variation of the channel. The probabilistic model with Bayesian estimation techniques gives performance approaching the Cramer-Rao lower bound.

1.2.3 Blind Detection in Time-Varying Channels

In chapter 4 we focus our attention on symbol detection over a time-varying channel. In particular, we focus on the problem of transmission of a relatively short block of N symbols through a multiplicative varying channel. The specific scenario that we study has the following properties:

- The transmitter sends a block of N symbols through the channel.
- The detector is blind, i.e. the receiver does not know the state of the channel.

- The receiver has no information about past or future states of the channel or blocks of symbols.
- The channel varies with symbol index, i.e. the channel is time-varying and frequency-flat with symbols in the time-domain (this is similar and applicable to the channel being frequency-selective and static with symbols in the frequency domain).
- The receiver does not have access to the channel statistics.

In the scenario outlined above, we attempt to answer the questions of what detection schemes can be devised, and how will they perform? We begin by assuming that the channel is deterministic and unknown for the sake of estimation, and by parameterizing it with a set of basis functions based on discrete Chebychev polynomials. We then develop a cost function based on maximum likelihood sequence estimation (MLSE) for data detection which uses the channel parameterization, and doesn't require channel statistics.

The cost function is shown to have an inherent phase ambiguity in the transmitted data sequence, so to overcome that we stipulate that the data be differentially encoded, as illustrated in Fig. 1.10.

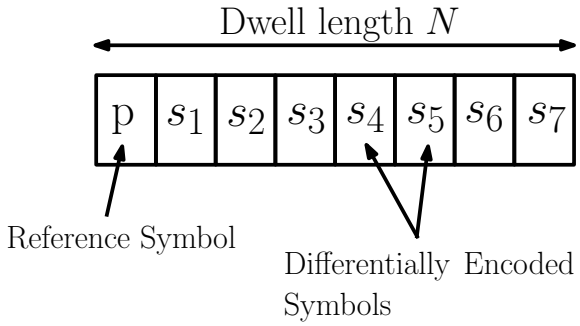

Figure 1.10: Block of symbols to be detected.

We then present a theoretical bit-error rate performance analysis of the proposed detector in Rician and Rayleigh fading channels, and develop expressions for the pairwise error probability in Rayleigh fading. To address the issue of detector complexity, we show how the cost function can be manipulated into a form that can be efficiently implemented with a sphere decoding algorithm.

The performance of the proposed detector with various channel parameterizations is shown to be very good, with performance near the differential-coherent lower bound

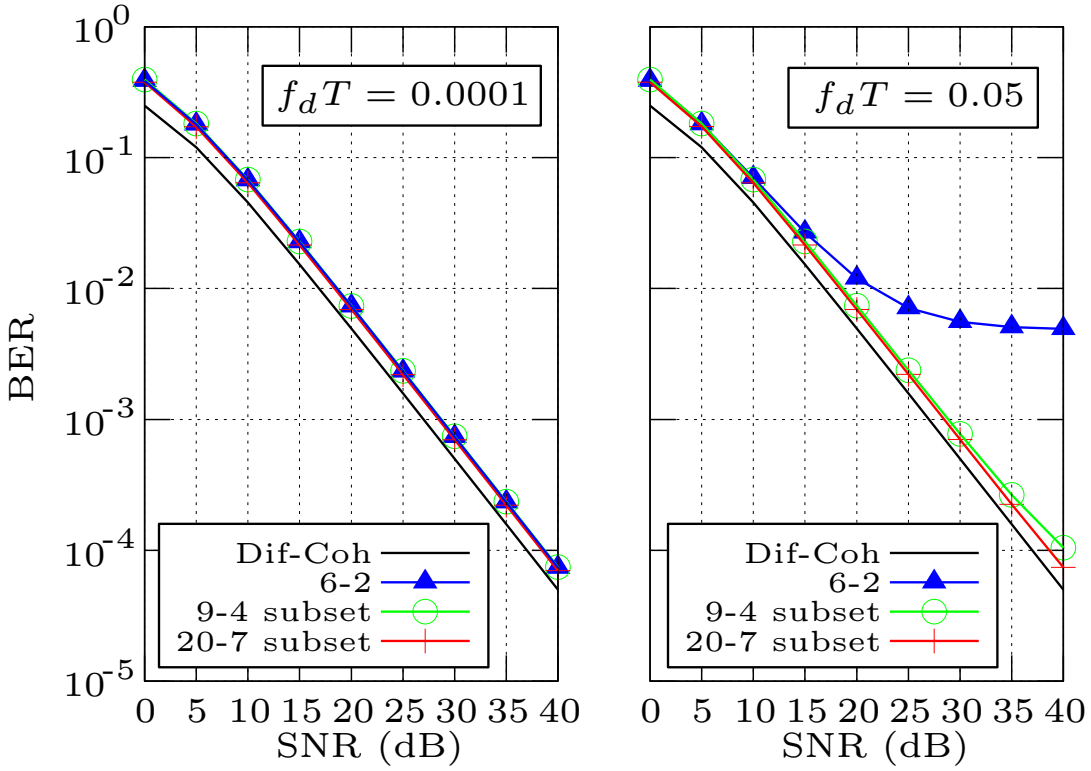

Figure 1.11: Theoretical performance of blind detectors in time-varying channels.

in some cases. However, similarly to the problems with channel parameterization in the context of CFO estimation, under or over-parameterizing the channel can result in error floors or poor low-SNR performance, respectively.

To combat this performance degradation, and to reduce the sensitivity of the detector to the choice of channel parameterization, we note that the two outside symbols in the transmitted block have the highest probability of error, and hurt overall performance the most. A "subset"-type detector is then proposed, which simply ignores the outer two symbols after detection, at a cost of slightly increased overhead and complexity.

Fig. 1.11 shows the theoretical bit-error rate performance of a detector with block-length 9 which parameterizes the channel with 2 basis functions, and two "subset"-type detectors with block lengths of 9 and 20 using 4 and 7 basis functions, respectively. The subset detectors achieve near-optimal performance over the SNR range 0 - 40 dB and over the normalized Doppler range $f_dT = 0.0001$ - 0.05.

1.2.4 ICI Reduction

In chapter 5 we increase the complexity of the problem addressed in the previous chapter. Instead of trying to recover a block of symbols transmitted over a time-varying channel, we consider a block of symbols in the frequency-domain transmitted via OFDM. The frequency-selectivity of the channel provides the multiplicative varying channel seen by the symbols, and is analogous to a block of time-domain symbols experiencing a multiplicative varying channel due to Doppler spread. Now we extend the problem and let the frequency-selective channel be time-varying, so that the symbols in frequency experience not only a varying channel from symbol to symbol, but intercarrier interference (ICI).

When ICI is present, the frequency-domain channel can be represented by a matrix, since each output symbol is affected by several (or all) input symbols.

There are various methods of dealing with intercarrier interference including frequency-domain equalization, which essentially tries to invert the channel matrix; time-domain windowing, which changes the frequency spectrum to reduce power outside the subcarrier of interest; and correlative coding of the frequency-domain data sequence, which can effectively and directly reduce ICI power.

We focus on correlative coding because it's separate and independent from any potential channel estimation, and it allows us to think of the channel as orthogonal, i.e. purely multiplicative and varying, which enables good possibilities for blind detection.

We begin by analyzing the ICI power due to time-variation/Doppler and show how correlation of the data sequence affects it. The conclusion is that symbols that are close to each other should be negatively correlated in order to reduce ICI power. The more negatively correlated they are, the more reduction in ICI power is possible.

With this in mind, we introduce a novel correlative coding method which we term "constant-magnitude correlative coding" (CMCC). CMCC is similar to DPSK in that the previous symbol is used as a reference and a phase is added depending on the input data sequence. The difference is that with DPSK the phase added is either 0 or π, while with CMCC the phase added is $+\theta$ or $-\theta$, with $\pi/2 \leq \theta \leq \pi$.

The CMCC scheme is shown to provide similar ICI reduction performance as other correlative coding schemes, is parameterized in a simple way that allows tweaking of ICI reduction capability, and provides full spectral efficiency.

Aside from the negative autocorrelation provided by CMCC which reduces ICI

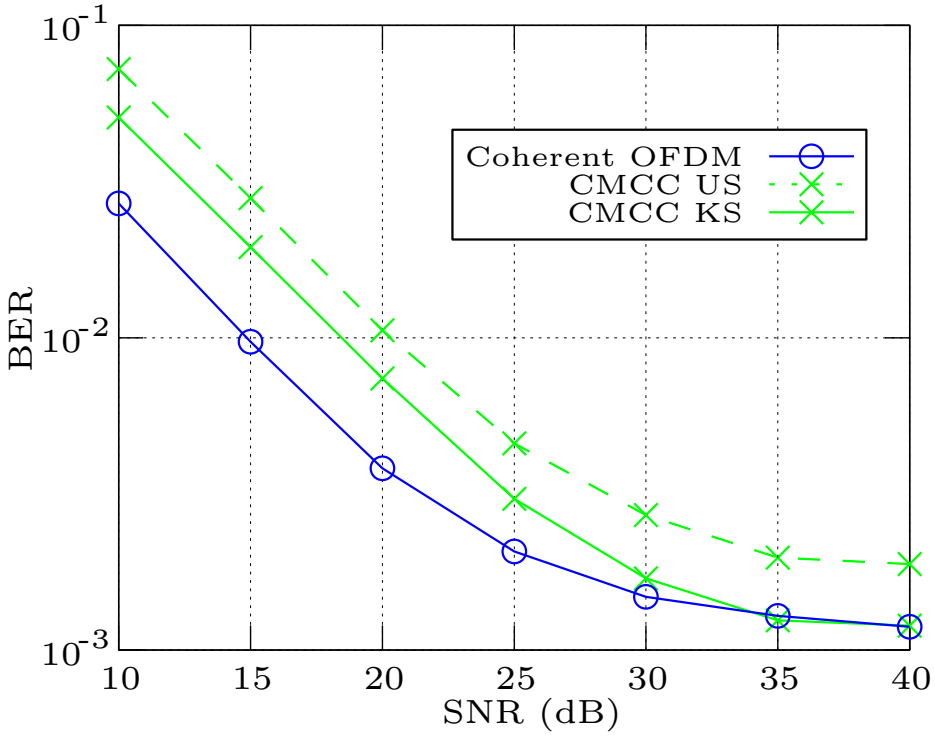

Figure 1.12: BER performance of constant-magnitude correlative coding with blind detection compared to coherent (known channel) OFDM in a time-varying frequency selective channel with $f_dT = 0.05$. Two detectors are shown, KS: known statistics and US: unknown statistics.

power, its most important property is that the coded symbols are constant in magnitude, a property that no other correlative coding scheme has. This property is shown to enable efficient and robust blind detection, and we develop blind detectors based on both a probabilistic (Rayleigh fading) and a deterministic (approximation with Chebyshev polynomials) channel model that are implemented with sphere decoding. The constant magnitude property is essential to the sphere decoding implementation.

Finally, simulation results show performance approaching and sometimes exceeding that of coherent OFDM that does not attempt to control the ICI. Fig 1.12 illustrates the performance of two detectors based on the CMCC signal precoding scheme. One assumes a probabilistic channel model (KS = known statistics), and one assumes a deterministic channel model (US = unknown statistics). The detectors are able to blindly estimate the frequency-selective channel in the face of ICI and provide performance near to that of uncoded OFDM which perfectly knows the channel, but does not attempt to reduce ICI.

Chapter 2

Time-Varying Channel Characterization

In this chapter we attempt to bridge the gap between physical reality and the abstract mathematical models that are intended to represent it. Throughout this dissertation signals and systems are represented by complex-valued discrete sequences. However, a quantity that exists in the real world cannot be complex. And similarly, practical signals do not exist at only discrete points in time; they are continuous.

Our intention is to derive from basic principles the discrete-time complex channel model framework, characterize it statistically, and then to present certain signalling structures and channel degradations within this framework. First, we show how the lowpass-equivalent channel model can be discretized. Then, we show how it can be characterized as a Gaussian random process for which we derive the time-correlation.

Next, we illustrate how real-world time-varying channel conditions can be condensed down into a number: the normalized Doppler f_dT, which is our basic parameter for describing the degree of time-variation. We then show the effect of carrier frequency offest on the discrete channel model. After that we present several ways of deterministically parameterizing the channel in time and our preferred method used throughout this dissertation. Finally, we show the effect of time-variation on the signalling technique OFDM within the framework of our discrete model, and show how it gives rise to intercarrier interference.

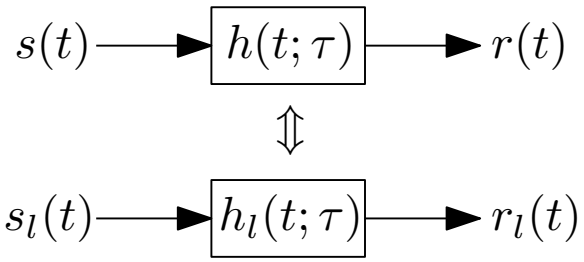

Figure 2.1: Equivalent lowpass system.

2.1 Lowpass-Equivalent Representation

A bandlimited signal may be represented by

$$s(t) = \mathrm{Re}\left\{s_l(t)e^{j2\pi f_c t}\right\} \tag{2.1}$$

as long as the carrier frequency f_c is much greater than the bandwidth W of the signal. The quantity $s_l(t)$ is the complex-valued "lowpass equivalent" or "complex baseband" representation of the signal $s(t)$. The real and imaginary parts of $s_l(t)$ conveniently carry information about the "in phase" and "quadrature" components of the signal, i.e. the components of the signal modulating $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$, respectively.

A useful property of the lowpass-equivalent representation is that the output of a signal through a linear filter has a lowpass representation equal to the convolution of the lowpass equivalent signal and the lowpass equivalent impulse response of the filter. In other words, if

$$r(t) = \int_{-\infty}^{\infty} h(t;\tau)s(t-\tau)d\tau \tag{2.2}$$

then

$$r_l(t) = \int_{-\infty}^{\infty} h_l(t;\tau)s_l(t-\tau)d\tau \tag{2.3}$$

where

$$r(t) = \mathrm{Re}\left\{r_l(t)e^{j2\pi f_c t}\right\} \tag{2.4}$$

is the lowpass equivalent representation of the output signal $r(t)$ and

$$h(t;\tau) = \mathrm{Re}\left\{h_l(t;\tau)e^{j2\pi f_c \tau}\right\} \tag{2.5}$$

is the lowpass equivalent representation of the linear filter $h_l(t;\tau)$.

2.2 Discrete Time Channel Model

Consider the communication system with linear, time-varying channel shown in Fig. 2.2.

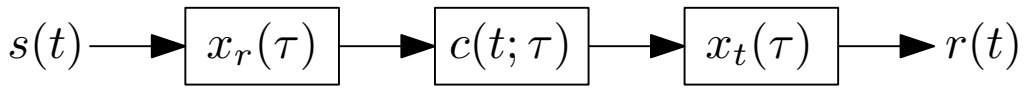

Figure 2.2: Basic communication system.

The various components of the system are defined as follows:

- $s(t)$: transmitted signal.
- $x_t(\tau)$: transmit shaping filter.
- $c(t;\tau)$: channel impulse response at time t.
- $x_r(\tau)$: receiver filter.
- $r(t)$: received signal.

We define the overall channel $h(t;\tau)$ as the cascade of the transmit filter, the channel impulse response, and the receiver filter

$$h(t;\tau) \equiv x_t(\tau) \star c(t;\tau) \star x_r(\tau) \tag{2.6}$$

and thus, the received signal is given by

$$r(t) = \int_{-\infty}^{\infty} h(t;\tau) s(t-\tau) d\tau. \tag{2.7}$$

Since the transmitted signal $s(t)$ is assumed to be bandlimited, its equivalent lowpass representation can be rewritten using the Nyquist Reconstruction Theorem as

$$s_l(t) = \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) \operatorname{sinc}\left(W\left(t - \frac{n}{W}\right)\right) \tag{2.8}$$

and its corresponding Fourier transform is

$$S_l(f) = \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) e^{-j2\pi f n/W}, \quad -\frac{1}{2W} \le f \le \frac{1}{2W} \tag{2.9}$$

where W is the total bandwidth occupied by the signal $s(t)$ and $\operatorname{sinc}(x) = \sin(\pi x)/\pi x$.

The lowpass equivalent representation of the received signal may be written as

$$r_l(t) = \int_{-\infty}^{\infty} S_l(f) H_l(t;f) e^{j2\pi ft} df \tag{2.10}$$

where $H_l(t;f)$ is the Fourier transform of $h_l(t;\tau)$. Substituting (2.9) into (2.10) results in

$$\begin{aligned} r_l(t) &= \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) \int_{-\infty}^{\infty} H_l(t;f) e^{j2\pi f\left(t-\frac{n}{W}\right)} df & (2.11)\\ &= \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) h_l\left(t; t-\frac{n}{W}\right) & (2.12) \end{aligned}$$

which can be rewritten as

$$r_l(t) = \sum_{k=1}^{L} h_l\left(t; \frac{k}{W}\right) s_l\left(t-\frac{k}{W}\right) \tag{2.13}$$

where the sum is truncated to L taps such that the multipath spread of the channel is covered. The expression for the received signal in (2.13) is known as the tapped-delay line channel model and is the basis for our ability to consider the transmitted signal, the channel, and the received signal as discrete sequences instead of continuous signals.

Finally, to fully discretize our channel model, we have

$$r_n = \sum_{k=1}^{L} h_{n,k} s_{n-k} \tag{2.14}$$

where $r_n = r_l(nT)$, $s_n = s_l(nT)$ and $h_{n,k} = h_l(nT, kT)$, $T = 1/W$ is the sampling interval of the discrete model. The discrete-time channel model of (2.14) is the linear time-varying channel model used throughout this dissertation.

2.3 Multipath Fading and Time Variation

Considering the discrete-time channel model in (2.14), we are interested in the statistical properties of the channel coefficients $h_{n,k}$ as they relate to multipath fading and Doppler spread. To do this we first make some assumptions about the nature of the underlying time-varying multipath channel $c(t;\tau)$.

We assume that there are N total propagation paths, where N is a very large number. Associated with each path is an attenuation factor α_n, a time delay τ_n, and an angle of incidence θ_n. We also assume that the receiver is moving at a constant velocity

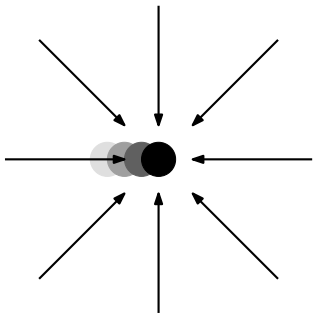

Figure 2.3: Multipath fading and receiver motion.

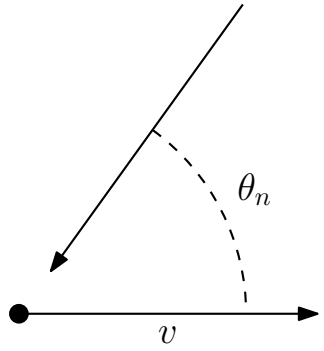

Figure 2.4: Receiver motion and the angle of incidence of the nth propagation path.

v, and that the angle of incidence is measured with respect to the direction of motion, as pictured in Fig. 2.4. The received signal due to the nth propagation path will undergo a Doppler shift such that the apparent received frequency will be given by

$$f_n = f_c + f_d \cos(\theta_n) \tag{2.15}$$

where $f_d = f_c v/c$ is the maximum Doppler shift.

A signal $s(t)$ propagating through the channel will be received as (ignoring the transmit and receive filters):

$$r(t) = \text{Re}\left\{\sum_{n=1}^{N} \alpha_n s_l(t-\tau_n) e^{j2\pi f_n (t-\tau_n)}\right\} \tag{2.16}$$

$$= \text{Re}\left\{\sum_{n=1}^{N} \alpha_n e^{-j2\pi(f_c\tau_n - f_d cos(\theta_n)(t-\tau_n)} s_l(t-\tau_n) e^{j2\pi f_c t}\right\} \tag{2.17}$$

which implies that the equivalent lowpass representation of $r(t)$ is

$$r_l(t) = \sum_{n=1}^{N} \alpha_n e^{-j2\pi(f_c\tau_n - f_d cos(\theta_n)(t-\tau_n)} s_l(t-\tau_n) \tag{2.18}$$

and that the equivalent lowpass representation of the channel is

$$c_l(t;\tau) = \sum_{n=1}^{N} \alpha_n e^{-j2\pi(f_c\tau_n - f_d cos(\theta_n)(t-\tau_n)} \delta(\tau - \tau_n). \tag{2.19}$$

Next, taking into account the effect of the transmit and receive filters, the overall lowpass equivalent channel impulse response is given by

$$h_l(t;\tau) = \int_{-\infty}^{\infty} c_l(t;\lambda) x_l(\tau - \lambda) d\lambda \tag{2.20}$$

where $x_l(\tau)$ is the equivalent lowpass representation of $x_t(\tau) \star x_r(\tau)$.

Substituting (2.19) into (2.20) results in

$$h_l(t;\tau) = \sum_{n=1}^{N} \alpha_n e^{-j2\pi(f_c\tau_n - f_d cos(\theta_n)(t-\tau_n)} x_l(\tau - \tau_n). \tag{2.21}$$

The effective channel $h_l(t;\tau)$ is the sum of a very large number of delayed, scaled, and phase rotated copies of $x_l(t)$. A very small change in the value of τ_n results in a large change in the overall phase, so that we can regard the phase as random and uniformly distributed. If we further assume that the different scatterers are independent, then application of the central limit theorem means that $h_l(t;\tau)$ is approximately a complex Gaussian random process.

To complete our statistical description of the channel, we find the time-correlation of the channel impulse response induced by receiver motion and Doppler. We can directly evaluate the correlation as follows:

$$E[h_l(t_1;\tau)h_l^*(t_2;\tau)] = E\left[\sum_{n=1}^{N} |\alpha_n|^2 |x_l(\tau - \tau_n)|^2 e^{j2\pi f_d(t_1 - t_2)\cos(\theta_n)}\right] \tag{2.22}$$

where we have made use of the fact that the attenuation of different scatterers is uncorrelated. We next make the assumptions that the path amplitudes and delays are independent from the phases, and that the phases are uniformly distrubuted between 0 and 2π [3], which results in

$$\begin{aligned} E[h_l(t_1;\tau)h_l^*(t_2;\tau)] &= \tilde{\sigma}_\tau^2 \frac{1}{2\pi} \int_0^{2\pi} e^{j2\pi f_d(t_1 - t_2)\cos(\theta)} d\theta & (2.23) \\ &= \tilde{\sigma}_\tau^2 J_0(2\pi f_d(t_1 - t_2)) & (2.24) \end{aligned}$$

where $\tilde{\sigma}_\tau^2 = E\left[\sum_{n=1}^{N} |\alpha_n|^2 |x_l(\tau - \tau_n)|^2\right]$ is the total received power of the scatterers influencing the channel at delay τ.

Finally, the time-correlation of the discrete channel sequence $h_{n,k}$ is

$$E\left[h_{n,k}h^*_{m,k}\right] = \sigma_k^2 J_0(2\pi f_d T(n-m)) \tag{2.25}$$

where $\sigma_k^2 = \tilde{\sigma}_{kT}^2$ is the power in the kth discrete channel tap.

2.4 Normalized Doppler

The quantity f_dT that appears in the derivation of channel time-correlation (2.25), known as the normalized Doppler frequency, is a measure of the Doppler shift relative to the carrier frequency. This measure fully encapsulates the degree of time-variation in the channel and is the fundamental parameter used to describe the degree of time-variation in this dissertation. In this section we describe how various communications system parameters affect the normalized Doppler frequency.

The normalized Doppler is

$$f_dT = f_c\frac{v}{c}T \tag{2.26}$$

where f_c is the carrier frequency, v is the velocity of the mobile receiver, c is the speed of light, and T is the time duration of a symbol. Clearly, when the carrier frequency or the mobile velocity increase, the normalized Doppler increases proportionally. We illustrate the interplay between these parameters in Figs. 2.5 and 2.6.

In Fig. 2.5 we fix the symbol duration at $T = 200\mu s$ and plot carrier frequency–velocity curves for normalized Dopplers of $fdT = 0.01$, 0.05, 0.1, and 0.2. Moving along a curve shows the values of carrier frequency and velocity required to achieve a certain normalized Doppler, for fixed T.

In Fig. 2.6 we fix the normalized Doppler at $f_dT = 0.1$ and plot carrier frequency–velocity curves for symbol durations of $T = 50\mu s$, $100\mu s$, $200\mu s$, and $400\mu s$. Moving along a curve shows the values of carrier frequency and velocity required to achieve the normalized Doppler $f_dT = 0.1$, for a fixed T corresponding to that curve.

These plots can be used to translate the values of normalized Doppler used throughout this dissertation into the physical parameters under which the communications system is operating.

It is less obvious how the bandwidth of the communications system and the number of subcarriers (as in the case of OFDM) affect the normalized Doppler.

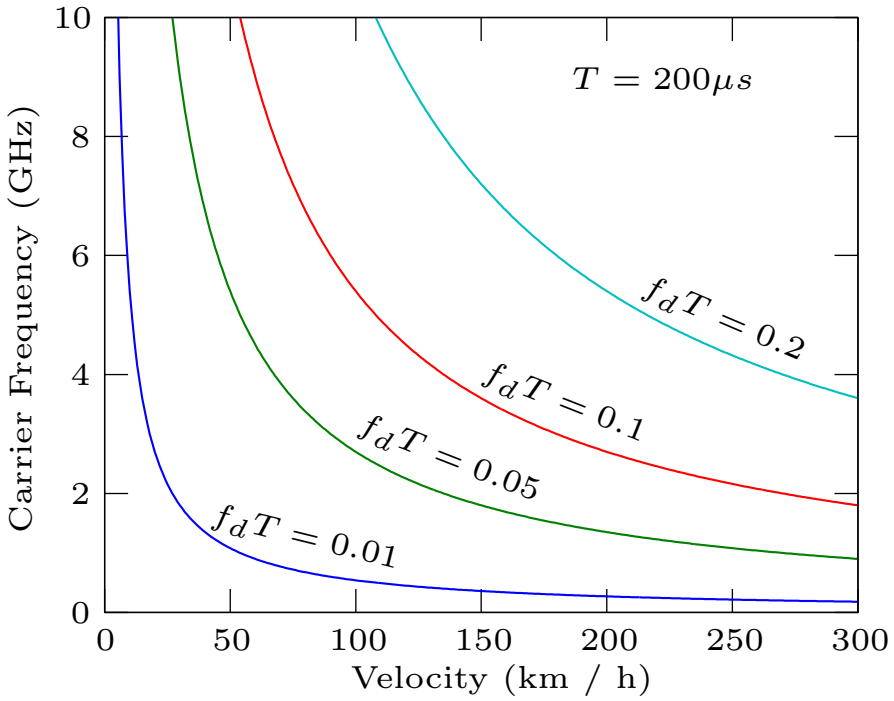

Figure 2.5: Carrier frequency - velocity curve for various levels of normalized Doppler and fixed symbol duration $T = 200\mu s$.

Fig. 2.7 illustrates the effect of increasing the bandwidth of the communications system. The solid line represents a system with bandwidth W_1 in the frequency domain on the top and corresponding symbol time-duration $T_1 = 1/W_1$ in the time-domain on the bottom. If the system bandwidth is increased to W_2, the corresponding symbol time-duration *decreases* to $T_2 = 1/W_2$. Since the symbol time-duration decreases, the normalized Doppler f_dT also decreases and the system is affected less by Doppler. A useful way to visualize this is that if the signal occupies more bandwidth overall, then the Doppler's effect of smearing the symbol in frequency has less of an impact.

In the case of OFDM it is interesting to note the effect on normalized Doppler of increasing the number of subcarriers while keeping the bandwidth fixed. Fig. 2.8 illustrates this effect. In the frequency domain, the black line represents the spectrum of an OFDM system with 5 subcarriers, and in the time domain the corresponding time duration is $T_1 = 5/W$. If the number of subcarriers is doubled to 10 within the same bandwidth, as represented by the gray line, the corresponding time-duration of one OFDM symbol also doubles to $T_2 = 10/W$. Essentially, a finer frequency resolution is achieved only by increasing the time duration of the OFDM symbol. This causes

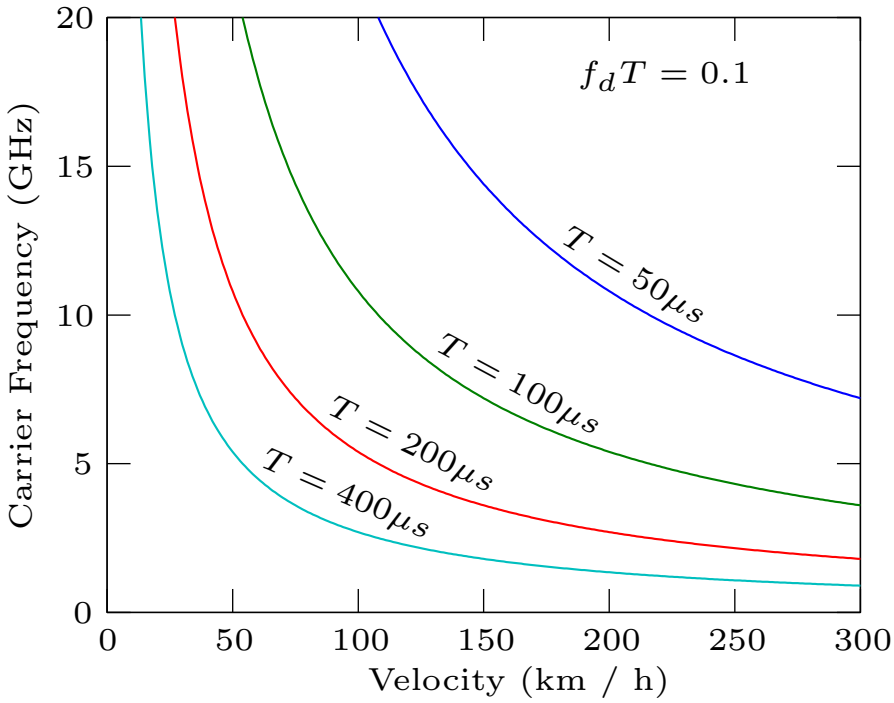

Figure 2.6: Carrier frequency - velocity curve for various levels of symbol duration and fixed normalized Doppler $f_dT = 0.1$.

Table 2.1: Effects of physical and system parameters on the normalized Doppler frequency f_dT.

Parameter	Effect on f_dT
Mobile velocity ↑	↑
Carrier frequency ↑	↑
Symbol time-duration ↑	↑
Bandwidth ↑	↓
# subcarriers ↑, bandwidth fixed	↑

the normalized Doppler f_dT to increase. Intuitively, if each subcarrier takes up less bandwidth, then smearing in frequency has more of an impact, and if the OFDM symbol time duration is larger, there is more variation between two adjacent OFDM symbols.

Table 2.1 summarizes the effects of changing system parameters on the normalized Doppler frequency.

2.5 Carrier Frequency Offset

Carrier frequency offset (CFO) arises as a mismatch between the local oscillators of the transmitter and receiver. Unless dealt with, CFO can cause performance

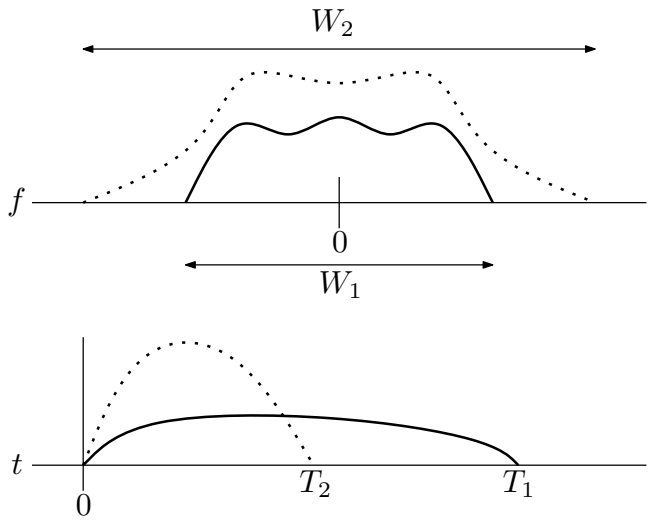

Figure 2.7: Increasing the bandwidth decreases the symbol duration.

degradation, depending upon the type of modulation used and the receiver design.

To determine the effect of CFO on the discrete channel input output relation (2.14), first consider the simple scenario where the transmitter transmits a signal $s(t)$ on carrier $f_c + \Delta_f$, and the receiver recovers the signal assuming that the carrier frequency is f_c.

The transmitted signal $s(t)$ can be written as

$$s(t) = \mathrm{Re}\left\{ s_l(t) e^{j2\pi(f_c+\Delta_f)t} \right\} \tag{2.27}$$

where $s_l(t)$ is the lowpass equivalent representation of $s(t)$, however from the perspective of the receiver the lowpass equivalent representation of $s(t)$ is $\tilde{s}_l(t) = s_l(t)e^{j2\pi\Delta_f t}$.

The received signal can then be written as

$$\begin{aligned} r(t) &= \mathrm{Re}\left\{ \int_{-\infty}^{\infty} h_l(t;\tau)\tilde{s}_l(t-\tau)d\tau e^{j2\pi f_c t} \right\} && (2.28) \\ &= \mathrm{Re}\left\{ e^{j2\pi\Delta_f t} \int_{-\infty}^{\infty} h_l(t;\tau) s_l(t-\tau)d\tau e^{j2\pi f_c t} \right\} && (2.29) \end{aligned}$$

and thus the lowpass received signal is,

$$r_l(t) = e^{j2\pi\Delta_f t} \int_{-\infty}^{\infty} h_l(t;\tau) s_l(t-\tau)d\tau. \tag{2.30}$$

Finally, bringing (2.30) into the discrete domain, we have the discrete-time input-output relation

$$r_n = e^{j2\pi\theta n} \sum_{k=1}^{L} h_{n,k} s_{n-k} \tag{2.31}$$

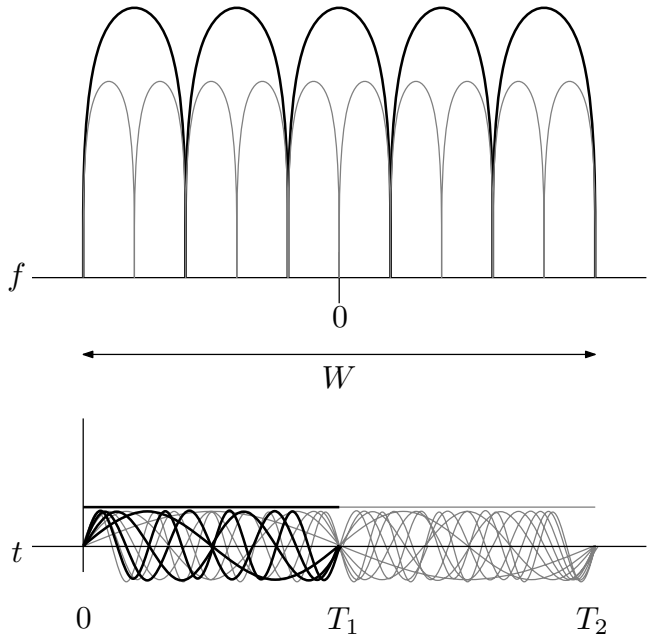

Figure 2.8: Increasing the number of subcarriers for a fixed bandwidth increases the symbol duration.

where $\theta = \Delta_f T$ is the CFO.

From (2.31) we can see that CFO manifests itself as a multiplicative time-varying phase applied to the received signal. The CFO is different from Doppler in that all propagation paths experience a frequency shift, not each separetely depending on angle of incidence. Thus at the receiver there is an overall shift in the frequency domain, as opposed to broadening or smearing due to Doppler.

It is interesting to note that the CFO can be thought of as a special case of channel time-variation. The input-output relation in (2.31) can be rewritten as

$$r_n = \sum_{k=1}^{L} \tilde{h}_{n,k} s_{n-k} \tag{2.32}$$

where $\tilde{h}_{n,k} = e^{j2\pi\theta n} h_{n,k}$. It is this time-variation ambiguity between CFO and channel that confounds CFO estimation in the presence of time-varying channels. We address this issue in Chapter 3 in our treatment and analysis of CFO in a time-varying frequency-selective channel.

2.6 Modeling the Time-Varying Channel Deterministically

Knowledge of the statistics of the time-varying channel is not always available at the receiver. For the purposes of estimation and detection, it is often advantageous to assume a deterministic model of the channel time-variation that is conveniently parameterized. To be useful in estimation or detection, a deterministic model should use as few parameters as possible to describe the time-variation of the channel as fully as possible.

Here we present several deterministic channel models including our own preferred model, which we use throughout the rest of this dissertation. Each of these models attempts to approximate the channel $h_{n,k}$ over a time-duration of interest $1 \leq n \leq N$ at a particular channel tap k through a parameterization in the time variable n.

The general deterministic channel model that we assume throughout this work is given by

$$h_{n,k} \approx \sum_{i=1}^{I} \alpha_{k,i} f_i(n) \tag{2.33}$$

where $f_i(n)$ are a set of orthonormal basis functions, i.e. they have the property

$$\sum_{n=1}^{N} f_{i_1}(n) f_{i_2}^*(n) = \begin{cases} 1, & i_1 = i_2 \\ 0, & i_1 \neq i_2 \end{cases} \tag{2.34}$$

and $\alpha_{k,i}$ unknown time-invariant coefficients that must be estimated. The following channel time-variation models all fit into this framework.

2.6.1 Basis Expansion Model

The basis expansion model [4] (BEM) assumes that the time-varying channel can be accurately approximated by a weighted sum of complex exponentials. The channel is approximated as

$$h_{n,k} \approx \sum_{q=1}^{Q} h_q(k) e^{j\omega_q n} \tag{2.35}$$

where the coefficients $h_q(k)$ and the frequencies ω_q are all unknown parameters to be estimated. The ω_q must be estimated first using estimated high-order moments of received data, which in our view runs counter to the purpose behind deterministic modeling, since second order statistics might as well be estimated and then it makes more sense to use a statistical channel model.

Additionally, using complex exponential basis functions has implications for CFO estimation, since the CFO manifests itself as a multiplicative complex exponential applied to the received signal. The received signal with CFO, as derived in (2.31),

$$\begin{aligned} r_n &= e^{j2\pi\theta n}\sum_{k=1}^{L} h_{n,k}s_{n-k} && (2.36)\\ &\approx e^{j2\pi\theta n}\sum_{k=1}^{L}\sum_{q=1}^{Q} h_q(k)e^{j\omega_q n}s_{n-k} && (2.37)\\ &= \sum_{k=1}^{L}\sum_{q=1}^{Q} h_q(k)e^{j\tilde{\omega}_q n}s_{n-k} && (2.38)\\ & && (2.39) \end{aligned}$$

where $\tilde{\omega}_q = \omega_q + 2\pi\theta$ which shows that there is an ambiguity between the CFO and the unknown BEM frequencies. Joint channel/CFO estimation will be unable to differentiate between the CFO and channel using the BEM channel model.

2.6.2 Discrete Prolate Spheroidal Sequences

It was proposed in [5] that a good deterministic channel model could be parameterized by basis functions that are perfectly bandlimited in frequency and that have the highest energy concentration in the time-duration of interest. Sequences that satisfy these criteria are known as the Discrete Prolate Spheroidal (DPS) sequences. These sequences are defined by the eigenvectors $\mathbf{u}_l$ of the real N by N matrix $\mathbf{C}$ which is defined as

$$[\mathbf{C}]_{i,j} = \frac{\sin\left(2\pi(i-j)f_dT\right)}{\pi(i-j)} \tag{2.40}$$

resulting in the channel approximation

$$h_{n,k} \approx \sum_{i=1}^{I} \alpha_{k,i}[\mathbf{u}_i]_n. \tag{2.41}$$

Similarly to the BEM channel model, in order to define the basis functions the parameter f_dT must be known a priori, however if f_dT is known then the channel correlation can be constructed as

$$E\left[h_{n,k}h^*_{m,k}\right] = \sigma_k^2 J_0(2\pi f_dT(n-m)) \tag{2.42}$$

and statistical estimation and detection techniques can be used.

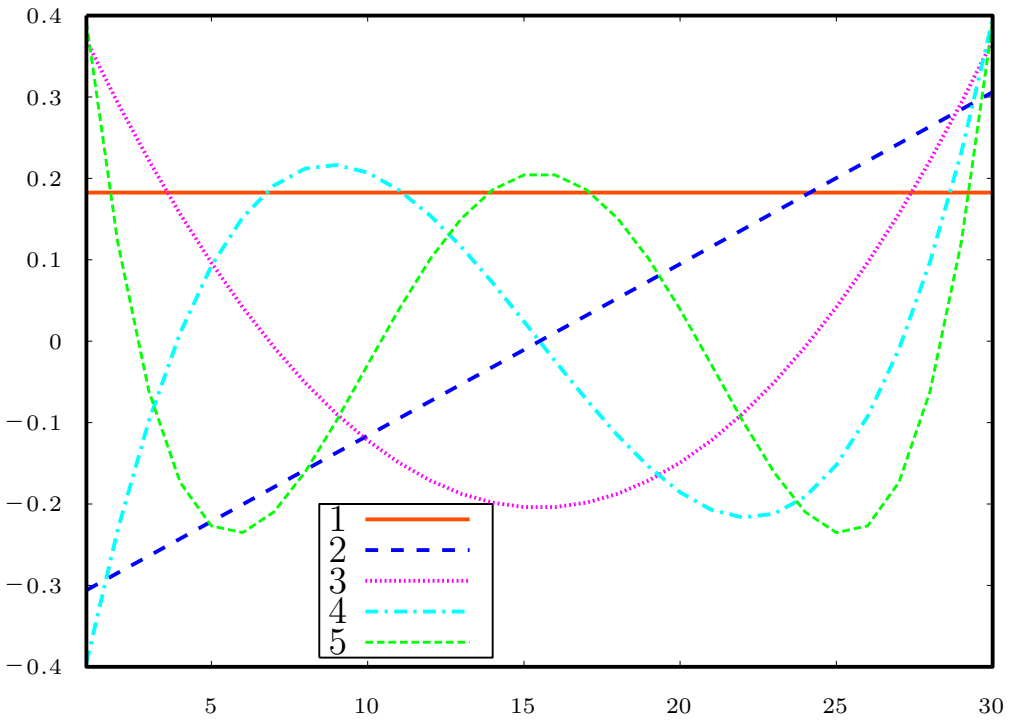

Figure 2.9: Discrete Chebyshev polynomials for parameterizing time-variation of the channel.

2.6.3 Discrete Chebyshev Polynomial

Throughout this work our deterministic channel model of choice is

$$h_{n,k} \approx \sum_{i=1}^{I} \alpha_{k,i} f_i(n) \tag{2.43}$$

where $f_i(n)$ are discrete, orthonormalized, Chebyshev polynomials. We choose this model for simplicity, namely no parameters need to be estimated or assumed in order to generate the basis functions $f_i(n)$. Fig. 2.9 displays the first five discrete Chebyshev polynomials for a sequence length of $N = 30$. If statistical parameters can be estimated or assumed, then statistical models should be employed for estimation and detection, and deterministic modeling is unnecessary.

2.7 OFDM Sensitivity to Time Variation

Orthogonal Frequency Division Multiplexing (OFDM) is a simple and efficient method of dealing with the frequency-selectivity of the channel, i.e. the intersymbol

interference (ISI) [6]. Instead of equalizing the ISI in the time-domain, the information bearing symbols are considered to be in the frequency domain. Through a simple trick of cyclically extending the time-domain transmitted sequence, the ISI channel is made to appear as a circular convolution instead of a linear convolution with the data sequence, and thus the ISI channel in the time-domain is transformed into a frequency-channel with multiplicative gains on each symbol.

When the channel is time-varying due to Doppler, or when CFO is present, the frequency-domain data symbols become smeared into each other, resulting in so-called intercarrier interference (ICI). In this section, we illustrate the degradation of OFDM due to Doppler spread and the origin of ICI.

The block of N data symbols x_k, $1 \leq k \leq N$ are considered to be in the frequency domain. The time-domain signal x_n to be transmitted is constructed by taking the inverse discrete Fourier transform (IDFT):

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} x_k e^{j2\pi nk/N}. \tag{2.44}$$

After passing through the time-varying frequency-selective channel, the received signal is

$$\begin{aligned} r_n &= \sum_{l=1}^{L} h_{n,l} x_{n-l} && (2.45) \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \sum_{l=1}^{L} h_{n,l} x_k e^{j2\pi(n-l)k/N}. && (2.46) \end{aligned}$$

Note that for values of $n \leq L$ the received time-domain signal r_n depends on values of x_n for $n \leq 0$, i.e. before we started transmitting x_n and during the previous OFDM block. Thus some of the x_k in (2.45) refer to the previous block of frequency-domain symbols. To make sure we are correctly referring to the symbols of only the current OFDM block, a cyclic prefix must be prepended to the time-domain sequence x_n.

The received sequence can be rewritten as

$$r_n = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} H_{n,k} x_k e^{j2\pi nk/N} \tag{2.47}$$

where $H_{n,k} = \sum_{l=1}^{L} h_{n,l} e^{-j2\pi lk/N}$ can be thought of as the frequency response of the channel at time n.

Next, the receiver applies the DFT to recover the transmitted symbols in the frequency domain

$$y_m = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} r_k e^{-j2\pi nm/N} \tag{2.48}$$

$$= \frac{1}{N} \sum_{k=1}^{N} x_k \sum_{n=1}^{N} H_{n,k} e^{j2\pi n(k-m)/N} \tag{2.49}$$

$$\equiv \sum_{k=1}^{N} x_k \tilde{H}_{k,m} \tag{2.50}$$

where we have defined

$$\tilde{H}_{k,m} \equiv \frac{1}{N} \sum_{n=1}^{N} H_{n,k} e^{j2\pi n(k-m)/N}. \tag{2.51}$$

Upon examining $\tilde{H}_{k,m}$ closer, we see that if the channel is time-invariant then $H_{n,k} = H_k$ is constant in n, and $\tilde{H}_{k,m} = H_k \delta_{k-m}$. Thus in the case of a time-invariant channel, we have the result that $y_m = H_m x_m$, or that the data modulated on different frequencies remains separable and may be detected independently. On the other hand, when the channel is varying in time, the received sequence after transformation to the frequency domain is

$$y_m = \tilde{H}_{m,m} x_m + \sum_{\substack{k=1 \\ k \neq m}}^{N} \tilde{H}_{k,m} x_k. \tag{2.52}$$

The second term on the right hand side of (2.52) is the ICI, a direct effect of the Doppler induced channel time-variation. The ICI leads to degraded system performance [7], and can be dealt with in a variety of ways. In chapter 5 we describe a novel data precoding method that can reduce the ICI power, while being well suited for effecient detection algorithms.

Chapter 3

Carrier Frequency Offset Estimation

Carrier frequency offset is a result of mismatch between the local oscillators of the transmitter and receiver of a communication system. The mismatch causes performance degradation depending upon the specific modulation format and receiver structure.

OFDM is particularly sensitive to carrier frequency offset because it relies upon the orthogonality of its subcarriers to separate its multiplexed data streams [8]. A slight offset in carrier frequency shifts the OFDM spectrum, causing a loss of orthogonality among the sub-carriers, which, upon applying the receiver FFT, results in so-called intercarrier interference. Hence, carrier frequency offset must be effectively estimated and corrected to achieve optimal system performance.

Communication systems frequently make use of a preamble block of known data as a training sequence before data is transmitted and detection is performed at the receiver. The training sequence can be used for channel impulse response estimation as well as CFO estimation.

Previous data-aided CFO estimation approaches derived for the time-invariant frequency-selective channel include [9], in which two consecutive identical symbols are transmitted, inducing a correlation structure in the received signal from which the CFO can be extracted, and [10], which presents a similar technique, but uses PN sequences to achieve a wider estimation range. In [11] an elegant framework is provided for joint ML estimation of CFO and time-invariant channel and it is shown that the joint estimation decouples to first estimating the CFO, then the channel. Extensions of [11] include [12]

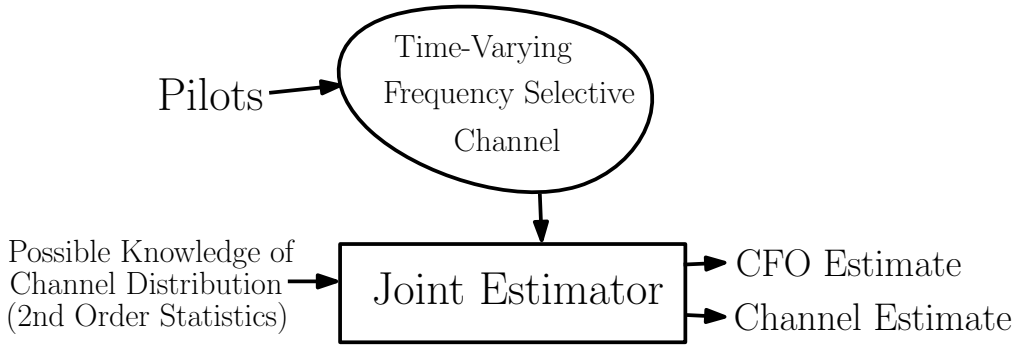

Figure 3.1: CFO and channel estimation objective.

and [13] which perform gradient descent on the joint likelihood function and a decision directed approach, respectively. The CFO estimator derived in [14] is also an extension of [11], with reduced complexity and a closed form approximation.

When the receiver and/or transmitter are mobile, a Doppler shift spreads the signal in frequency, and causes a time-variation of the channel taps. Because the CFO can be viewed as a special case of channel time-variation, it is more difficult to jointly estimate both the CFO and channel, and the previously mentioned estimators suffer a performance degradation.

Previous approaches to joint estimation of CFO and time-varying channel include [15], which presents a blind method and assumes channel time-variation statistics are known, and [16], which parameterizes the channel with an autoregressive model then uses the EM algorithm to determine the AR coefficients, from which a CFO estimate can be extracted. In [17], two consecutive OFDM preamble symbols comprised of CAZAC sequences are used, allowing a closed form CFO estimate.

In this work we propose a more direct approach to data-aided joint CFO and channel estimation in a frequency-selective time-varying channel. We use a framework very similar to [11], but that takes into account each channel tap at each point in time. We directly estimate the channel and the CFO via maximum likelihood when the channel statistics are not known, and via a Bayesian approach when they are.

When channel statistics are unknown, joint ML estimation faces the potential problem of having to estimate more parameters (all channel taps at all points in time plus CFO) than can be supported by the available data, also known as the "identifiability problem" [18]. This problem arises in the time-varying channel, in contrast to the time-

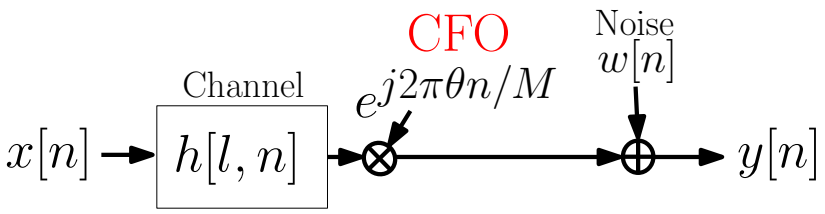

Figure 3.2: System model with carrier frequency offset.

invariant channel, because simply taking more data only leads to more points in time at which the channel needs to be estimated. Thus the ML estimator is found to require a channel parameterization.

This chapter is organized as follows: In section 4.1 we define the system model for the time-varying frequency-selective channel then in sections 3.2 and 3.3 we derive joint CFO/channel estimators for the cases of unknown and known channel statistics, respectively. We derive the CRLB and high SNR approximations to the mean and variance of the estimators in section 3.4, and finally present simulation results in 3.5.

3.1 System Model

We assume the system model pictured in Fig. 3.2. For a sequence $x[n], -L+1 \leq n \leq M-1$, transmitted through a frequency-selective time-varying channel, the sampled complex-baseband received sequence $y[n]$ is given by

$$y[n] = e^{j2\pi\theta n/M} \sum_{l=0}^{L-1} h[l,n]x[n-l] + w[n], \tag{3.1}$$

where θ is the carrier frequency offset normalized to the block symbol rate (i.e. measured in OFDM subcarrier spacings), M is the block size (number of OFDM subcarriers), L is the number of channel taps, $w[n]$ is zero-mean complex Gaussian circularly-symmetric noise and $h[l,n]$ is the lth channel tap at time n.

The values of $x[n]$ at times $n < 0$ affect the received sequence at times $n \geq 0$ due to the length of the channel, and in the case of OFDM, they would be part of the cyclic prefix. The received sequence $y[n]$ can be stacked and represented in vector form as

$$\mathbf{y} = \Omega \sum_{l=0}^{L-1} \mathbf{A}_l \mathbf{h}_l + \mathbf{w}, \tag{3.2}$$

where $\mathbf{y} = [y[0], y[1], \cdots, y[M-1]]^T$, $\mathbf{w} = [w[0], w[1], \cdots, w[M-1]]^T$, the CFO matrix Ω is M by M and diagonal with entries $\left[1, e^{j2\pi\theta/M}, \cdots, e^{j2\pi\theta(M-1)/M}\right]$,

$$\mathbf{h}_l = [h[l,0], h[l,1], \cdots, h[l,M-1]]^T$$

fully describes the time variation of the lth channel tap, and $\mathbf{A}_l$ are diagonal matrices containing the transmitted pilot sequence with entries $[x[0-l], x[1-l], \cdots, x[M-1-l]]$.

Intuitively, we can see that estimation of the carrier frequency offset is confounded by the fact that it is hard to tell tell it apart from the time-varying channel. Note that carrier frequency offset is in fact just a special case of channel time variation; we could consider the lth channel tap at time n to be $e^{j2\pi\theta n/M}h[l,n]$, or in the vector case, the lth channel tap vector could be $\Omega\mathbf{h}_l$.

3.2 CFO Estimation with Unknown Channel Statistics

We can further rewrite equation (3.2) as

$$\mathbf{y} \quad = \quad \Omega\mathbf{A}\mathbf{h} + \mathbf{w}, \tag{3.3}$$

where $\mathbf{A}$ is the stacked $\mathbf{A}_l$ matrices, $\mathbf{A} = [\mathbf{A}_0, \mathbf{A}_1, \cdots, \mathbf{A}_{L-1}]$, and $\mathbf{h}$ is the stacked channel tap time-variation vectors: $\mathbf{h} = [\mathbf{h}_0^T, \mathbf{h}_1^T, \cdots, \mathbf{h}_{L-1}^T]^T$.

When the channel statistics are not known, both the channel and CFO are assumed deterministic and unknown. Estimation takes the form of jointly maximizing the likelihood of the received data given the channel and CFO:

$$\{\hat{\theta}, \hat{\mathbf{h}}\} \quad = \quad \arg\max_{\theta,\mathbf{h}} p\left(\mathbf{y}|\theta,\mathbf{h}\right) \tag{3.4}$$

$$\quad = \quad \arg\min_{\theta,\mathbf{h}} \|\mathbf{y} - \Omega\mathbf{A}\mathbf{h}\|^2. \tag{3.5}$$

For any θ, the $\mathbf{h}$ that minimizes the cost function in (3.5) is

$$\hat{\mathbf{h}} \quad = \quad \mathbf{A}^\dagger\Omega^H\mathbf{y}, \tag{3.6}$$

where $(\cdot)^\dagger$ denotes the pseudoinverse. Plugging (3.6) into (3.5) it follows that

$$\hat{\theta} \quad = \quad \arg\max_{\theta} g_1(\theta), \tag{3.7}$$

where $g_1(\theta) = \mathbf{y}^H\Omega\mathbf{A}\mathbf{A}^\dagger\Omega^H\mathbf{y}$ is the cost function for estimating the CFO via joint ML in the time-varying channel.

But, since $\mathbf{h}$ contains ML elements, the matrix $\mathbf{A}$ is M by ML, and so $\mathbf{A}\mathbf{A}^{\dagger}$ is identity and our cost function reduces to $\mathbf{y}^H\mathbf{y}$, rendering CFO estimation impossible. This is the "identifiability problem" as discussed in [18], where conditions are given that guarantee the estimability of the CFO in a static channel. Note that if the channel were time-invariant, we could simply increase the length of the pilot sequence and thereby increase the number of observed data points to solve this problem. However, in the the time-varying channel, increasing the length of the transmitted pilot sequence does not mitigate the identifiability problem, because as the pilot sequence gets longer, so do the lengths of the channel tap time-variation vectors. We are forced to try an alternate method.

Since it is impossible to jointly estimate the value of each channel tap at each time instant along with the CFO, we reduce the number of parameters that need to be estimated by parameterizing the channel. To do this, we project $\mathbf{h}_l$ onto a set of I orthonormal basis vectors $\mathbf{f}_1, \cdots, \mathbf{f}_I$ that presumably can describe most of the time variation.

Grouping the basis vectors into a matrix $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_I]$, we can approximate the lth channel time variation vector as

$$\mathbf{h}_l \approx \mathbf{F}\alpha_l, \tag{3.8}$$

where α_l is a length I vector of coefficients corresponding to each basis vector. Using this channel parameterization, we must now estimate the LI channel coefficients $\alpha_0, \cdots, \alpha_{L-1}$ and the CFO, a total of $LI + 1$ parameters. This is possible if the received vector $\mathbf{y}$ contains at least this many entries.

Rewriting (3.2) to take into account the approximation in (4.7), we obtain

$$\mathbf{y} \;=\; \Omega \sum_{l=0}^{L-1} \mathbf{A}_l \mathbf{F} \alpha_l + \mathbf{w}, \tag{3.9}$$

which can be rewritten similarly to (3.3) as

$$\mathbf{y} \;=\; \Omega \mathbf{A}_F \alpha + \mathbf{w}, \tag{3.10}$$

where $\mathbf{A}_F = [\mathbf{A}_0\mathbf{F}, \mathbf{A}_1\mathbf{F}, \cdots, \mathbf{A}_{L-1}\mathbf{F}]$, and $\alpha = [\alpha_0^T, \alpha_1^T, \cdots, \alpha_{L-1}^T]^T$. The conditional PDF of the data according to the model in (3.10) is

$$p\left(\mathbf{y}|\theta, \alpha\right) = \frac{1}{\pi^M |\mathbf{R}_{\mathbf{w}}|} \exp\left[(\mathbf{y} - \Omega\mathbf{A}_F\alpha)^H \mathbf{R}_{\mathbf{w}}^{-1} (\mathbf{y} - \Omega\mathbf{A}_F\alpha)\right]. \tag{3.11}$$

Now, jointly maximizing this approximate likelihood of both CFO and channel parameters can be done as

$$\{\hat{\theta}, \hat{\alpha}\} = \arg\max_{\theta,\alpha} p\left(\mathbf{y}|\theta, \alpha\right) \tag{3.12}$$

$$= \arg\min_{\theta,\alpha} \|\mathbf{y} - \Omega\mathbf{A}_F\alpha\|^2 \tag{3.13}$$

and, paralleling the underparameterized case, the α that minimizes (3.13) for any θ is

$$\hat{\alpha} = (\mathbf{A}_F^H\mathbf{A}_F)^{-1}\mathbf{A}_F^H\Omega^H\mathbf{y}. \tag{3.14}$$

Note that if we choose a small enough number of basis functions, $\mathbf{A}_F$ is a tall matrix and $(\mathbf{A}_F^H\mathbf{A}_F)^{-1}$ exists. The CFO estimate then becomes

$$\hat{\theta} = \arg\max_{\theta} g_2(\theta), \tag{3.15}$$

where $g_2(\theta) = \mathbf{y}^H\Omega\mathbf{A}_F(\mathbf{A}_F^H\mathbf{A}_F)^{-1}\mathbf{A}_F^H\Omega^H\mathbf{y}$ is the cost function for estimating the CFO via joint ML after parameterizing the time-varying channel. Here we see that the CFO estimator decouples from the channel estimator, but it depends on the particular channel parameterization that we have chosen.

Note that if the channel is assumed constant, there is only one basis vector and $\mathbf{F}$ is simply the (scaled) all ones vector. Under this scenario $\mathbf{A}_F$ becomes an M by L static-channel convolution matrix and the above cost function in (3.15) reduces to that of Morelli and Mengali's MLE#1 estimator from [11] which is optimal (achieves the CRLB) for time-invariant channels.

3.3 CFO Estimation with Known Channel Statistics

When the channel statistics are known, i.e. we have available to us the channel correlation matrix $\mathbf{R_h} = E[\mathbf{h}\mathbf{h}^H]$ and the noise variance σ^2 with $E[\mathbf{w}\mathbf{w}^H] = \sigma^2\mathbf{I}$, we can apply Bayesian methods to estimate the CFO and channel. Here the channel is assumed to be a zero mean complex Gaussian circularly symmetric random variable in accordance with Rayleigh fading, but the CFO is still considered deterministic.

Assuming Rayleigh fading and uncorrelated channel taps, $\mathbf{R_h}$ is an ML by ML

block diagonal matrix

$$\mathbf{R_h} = \begin{bmatrix} \mathbf{R_h^{(0)}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R_h^{(1)}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R_h^{(L-1)}} \end{bmatrix} \tag{3.16}$$

where each block on the diagonal is equal to the time correlation of a particular channel tap $\mathbf{R_h^{(l)}} = E[\mathbf{h}_l \mathbf{h}_l^H]$. If we assume that each channel tap varies at the same rate, then each block in (3.16) is a scaled version of the same time-correlation matrix:

$$\mathbf{R_h^{(l)}} = p_l \mathcal{R}_\mathbf{h} \tag{3.17}$$

where p_l is the power in the lth channel tap and the $\{i,j\}$th entry of $\mathcal{R}_\mathbf{h}$ is $J_0(2\pi f_d T(i-j)/M)$, according to the Jakes model. f_d is the Doppler frequency and T is the duration of the length-M received data vector.

Maximum a posteriori (MAP) estimation yields

$$\{\hat{\theta}, \hat{\mathbf{h}}\} = \arg\max_{\theta,\mathbf{h}} p\left(\mathbf{y}|\theta,\mathbf{h}\right) p(\mathbf{h}) \tag{3.18}$$

$$= \arg\min_{\theta,\mathbf{h}} \frac{1}{\sigma^2}\|\mathbf{y} - \Omega\mathbf{A}\mathbf{h}\|^2 + \mathbf{h}^H \mathbf{R_h^{-1}} \mathbf{h}. \tag{3.19}$$

For any θ, the $\mathbf{h}$ that minimizes the cost function in (3.19) is

$$\hat{\mathbf{h}} = (\mathbf{A}^H\mathbf{A} + \sigma^2 \mathbf{R_h^{-1}})^{-1} \mathbf{A}^H \Omega^H \mathbf{y}, \tag{3.20}$$

which is also the MMSE estimate of $\mathbf{h}$. Plugging (3.20) into (3.19) we arrive at

$$\hat{\theta} = \arg\max_{\theta} g_3(\theta), \tag{3.21}$$

where $g_3(\theta) = \mathbf{y}^H \Omega \mathbf{A} (\mathbf{A}^H\mathbf{A} + \sigma^2 \mathbf{R_h^{-1}})^{-1} \mathbf{A}^H \Omega^H \mathbf{y}$ is the cost function for estimating the CFO when the time-varying channel statistics are known.

Notes:

- There is no identifiability issue due to regularization of the inverse in $g_3(\theta)$. Hence, the estimator with known channel statistics does not require a channel parameterization.
- Maximizing $g_3(\theta)$ is equivalent to minimizing $\mathbf{y}^H \mathbf{R_{yy}^{-1}} \mathbf{y}$ which maximizes the likelihood $p(\mathbf{y}|\theta)$. So this Bayesian approach decouples to first finding $\hat{\theta}$ via ML and then finding $\hat{\mathbf{h}}$ via MMSE or MAP.

As is shown analytically and in simulations below, performance of the estimator without statistics is highly dependent upon the basis functions and the number of parameters chosen to represent each channel tap, degrading performance as compared to the estimator with known statistics.

3.4 CFO Estimator Performance

3.4.1 Cramer-Rao Bound

The Cramer-Rao Lower Bound is the minimum variance attainable by any unbiased estimator for a given parameter and observations related by $p(\mathbf{y}|\theta)$. For a complex Gaussian data vector $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_{yy}}(\theta))$ with scalar parameter θ, the CRLB is given by [19]

$$\mathrm{Var}(\hat{\theta}) \quad \geq \quad \frac{1}{tr\left[\left(\mathbf{R_{yy}^{-1}}(\theta)\frac{\partial}{\partial\theta}\mathbf{R_{yy}}(\theta)\right)^2\right]}. \tag{3.22}$$

The correlation of the data is

$$\mathbf{R_{yy}}(\theta) = E\left[\mathbf{y}\mathbf{y}^H\right] = \Omega\tilde{\mathbf{R}}_{\mathbf{yy}}\Omega^H \tag{3.23}$$

where $\tilde{\mathbf{R}}_{\mathbf{yy}} = \sum_{l=0}^{L-1} p_l \mathbf{A}_l \mathcal{R}_h \mathbf{A}_l^H + \sigma^2\mathbf{I}$. Taking the derivative with respect to θ we have

$$\frac{\partial}{\partial\theta}\Omega\tilde{\mathbf{R}}_{\mathbf{yy}}\Omega^H = \Omega\left(\mathbf{D}\tilde{\mathbf{R}}_{\mathbf{yy}} + \tilde{\mathbf{R}}_{\mathbf{yy}}\mathbf{D}^H\right)\Omega^H \tag{3.24}$$

where $\mathbf{D}$ is a diagonal matrix with elements $[0, j2\pi/M, \cdots, j2\pi(M-1)/M]$. Using the results of (3.24) and the fact that $\mathbf{R_{yy}^{-1}}(\theta) = \Omega\tilde{\mathbf{R}}_{\mathbf{yy}}^{-1}\Omega^H$, we arrive at

$$\mathrm{Var}(\hat{\theta}) \quad \geq \quad \frac{1}{2tr\left(\mathbf{D}^2 + \tilde{\mathbf{R}}_{\mathbf{yy}}^{-1}\mathbf{D}\tilde{\mathbf{R}}_{\mathbf{yy}}\mathbf{D}^H\right)}. \tag{3.25}$$

This bound holds for any unbiased estimator of θ that makes use of the data $\mathbf{y}$ to form the estimate. For example, the bound holds even if the estimator does not take advantage of, or even know, the channel statistics and correspondingly the PDF $p(\mathbf{y}|\theta)$. As such, it is a lower bound on the variance for all the estimators presented in this chapter.

3.4.2 Known Channel Statistics

With the same technique used in [11] and originally developed in [20] and [21], when the SNR is high we can approximate the mean and variance of the CFO estimator

as

$$E[\hat{\theta}] \approx \theta - \frac{E[\frac{\partial}{\partial\theta}g(\theta)]}{E[\frac{\partial^2}{\partial\theta^2}g(\theta)]} \tag{3.26}$$

and

$$\mathrm{Var}(\hat{\theta}) \approx \frac{E[(\frac{\partial}{\partial\theta}g(\theta))^2]}{E[\frac{\partial^2}{\partial\theta^2}g(\theta)]^2}, \tag{3.27}$$

where $g(\theta)$ is the cost function optimized by the estimator. Inserting the known-channel-statistics cost function $g_3(\theta)$ into the mean in (3.27), the numerator is easily found to be zero, indicating that the estimator is unbiased.

When the cost function $g_3(\theta)$, which is equivalently $\log p(\mathbf{y}|\theta)$, is inserted into the variance in (3.27), it is easily manipulated into the definition of the CRLB for a single scalar parameter. Hence, the Bayesian joint estimator achieves the CFO CRLB, as expected, since it decouples to MAP or MMSE channel estimation and ML CFO estimation, which has the property of asymptotic efficiency.

3.4.3 Unknown Channel Statistics

To find the bias of the estimator that doesn't know the channel statistics we insert $g_2(\theta)$ into the mean in (3.27) to find

$$E[\hat{\theta}] - \theta = \frac{-tr\left[\left(\mathbf{DQ} + \mathbf{QD}^H\right)\mathbf{R}'_{\mathbf{yy}}\right]}{tr\left[\left(\mathbf{D}^2\mathbf{Q} + \mathbf{QD}^2 + 2\mathbf{DQD}^H\right)\mathbf{R}'_{\mathbf{yy}}\right]}, \tag{3.28}$$

where $\mathbf{Q} = \mathbf{A}_F(\mathbf{A}_F^H\mathbf{A}_F)^{-1}\mathbf{A}_F^H$ and $\mathbf{R}'_{\mathbf{yy}} = \sum_{l=0}^{L-1} p_l\mathbf{A}_l\mathbf{R}_h\mathbf{A}_l^H$. Equation (3.28) is not in general equal to zero, but it can be shown that if the pilots are purely real or imaginary, or if the channel is well parameterized, meaning that the modeling error inherent in the approximation (4.7) is zero, the estimator is unbiased.

To find the variance of the estimator we insert $g_2(\theta)$ into the variance in (3.27), and after some manipulation, we find

$$\mathrm{Var}(\hat{\theta}) = \frac{a + \sigma^2 b + \sigma^4 c}{d}, \tag{3.29}$$

where

$$a = tr^2\left[\left(\mathbf{DQ} + \mathbf{QD}^H\right)\mathbf{R}'_{\mathbf{yy}}\right] + tr\left[\left(\mathbf{DQ} + \mathbf{QD}^H\right)\mathbf{R}'_{\mathbf{yy}}\right]^2,$$

$$b = 2tr\left[\left(\mathbf{DQ} + \mathbf{QD}^H\right)^2\mathbf{R}'_{\mathbf{yy}}\right],$$

$$c = tr\left[\mathbf{DQ} + \mathbf{QD}^H\right]^2,$$

and

$$d = tr^2\left[\left(\mathbf{D}^2\mathbf{Q} + \mathbf{Q}\mathbf{D}^2 + 2\mathbf{D}\mathbf{Q}\mathbf{D}^H\right)\mathbf{R}'_{\mathbf{yy}}\right].$$

In (3.29) we see a term that is constant in the noise variance, leading to an error floor at high SNR with a value equal to a/d. If the approximation in (4.7) is perfect (i.e. $\mathbf{h}_l = \mathbf{F}\alpha_l$), it can be shown that $a = 0$, meaning there is no error floor. Thus, we conclude that the error floor is due to the modeling error in the channel parameterization.

3.5 Simulation Results

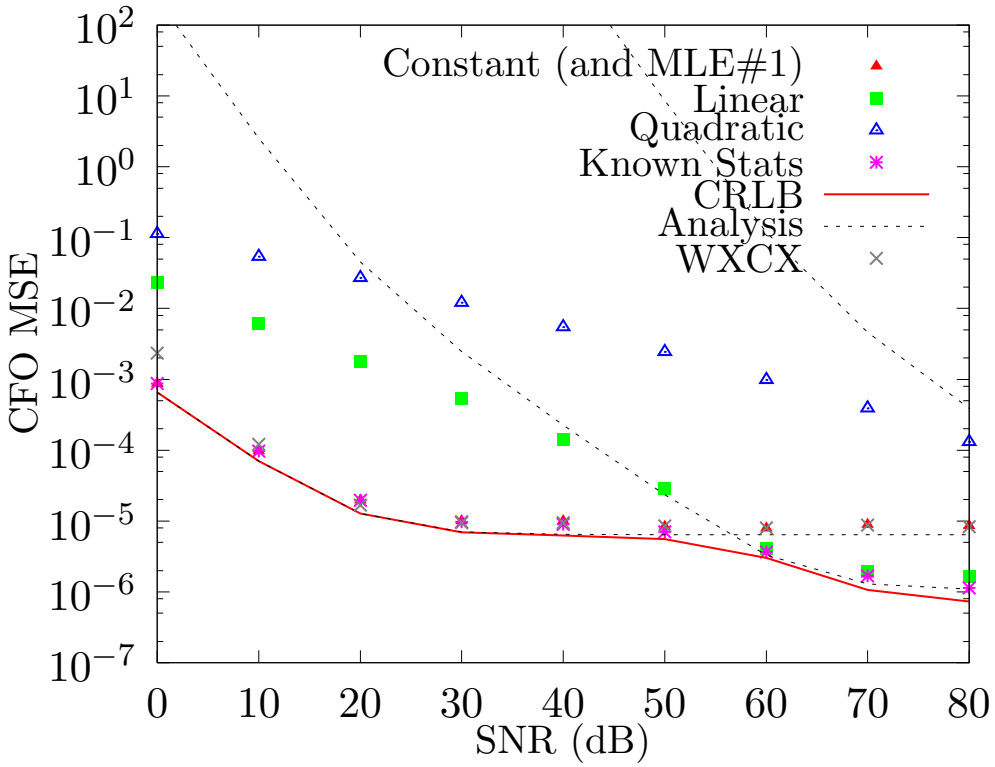

Figure 3.3: CFO Estimator performance vs. SNR for $f_dT = 0.01$

Here we present simulations of the derived estimators in a frequency-selective time-varying channel and we compare to Morelli and Mengali's MLE#1 from [11] and to Wei, Xu, Cai, and Xu's (WXCX) estimator from [17] (with parameters $r_1 = 2$ and $r_2 = 20$. The WXCX estimator was derived for a fast time-varying channel and actually requires two separate training sequences. In our simulations we therefore give it an

advantage and allow it a training sequence twice as long as that used for our estimators and the MLE#1.

For the estimator without channel statistics, we parameterize the channel with a varying number of orthogonal basis functions: from one to three. We choose these basis functions to be constant, linear, and quadratic in time. Note that the "optimal" basis functions (see [22]) cannot be chosen since they depend on the channel statistics, which are not known to the estimator. Additionally, what is optimal for channel estimation may not be optimal for CFO estimation.

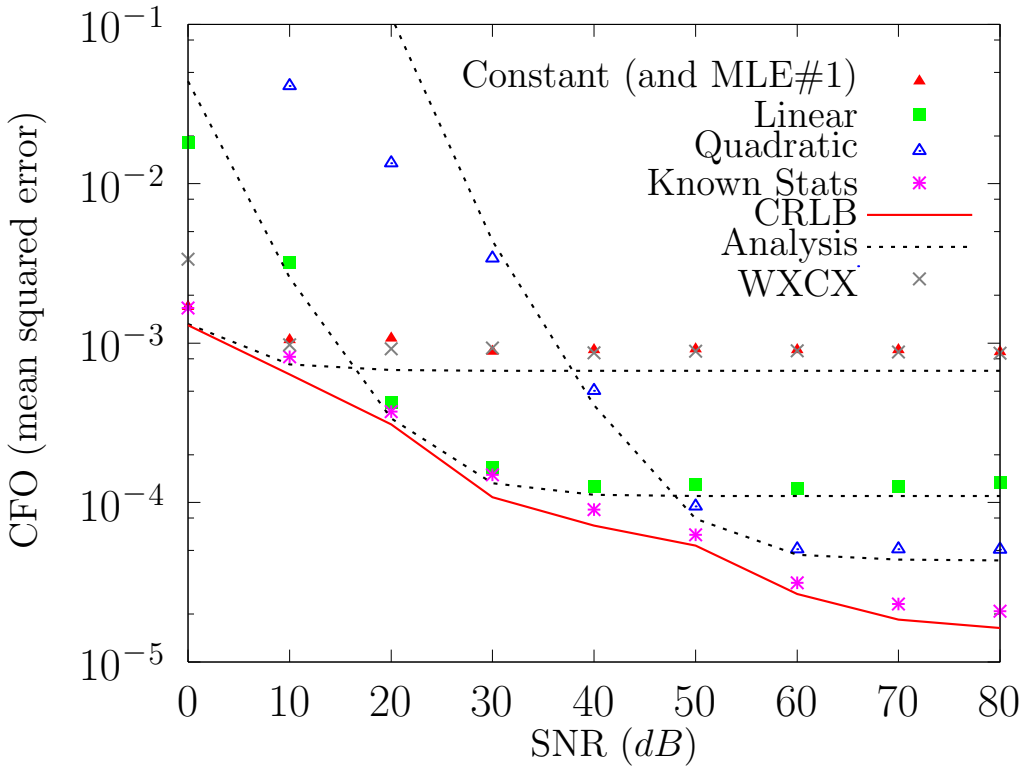

Figure 3.4: CFO Estimator performance vs. SNR for $f_dT = 0.1$

We fix the length of the transmitted pilot sequence to be the sum of $M = 256$ and the channel impulse response length $L = 4$. The channel taps exhibit Rayleigh fading and vary in time according to the Jakes model as in section 3.3, with an exponential decay power delay profile given by $p_l = c10^{-\beta l}$ where c is a normalizing constant and $\beta = 1/20$.

Figures 3.3-3.5 show simulations of CFO MSE vs. SNR for normalized Doppler from $f_dT = 0.01$ to $f_dT = 0.3$ (T refers to the duration of an entire block of length

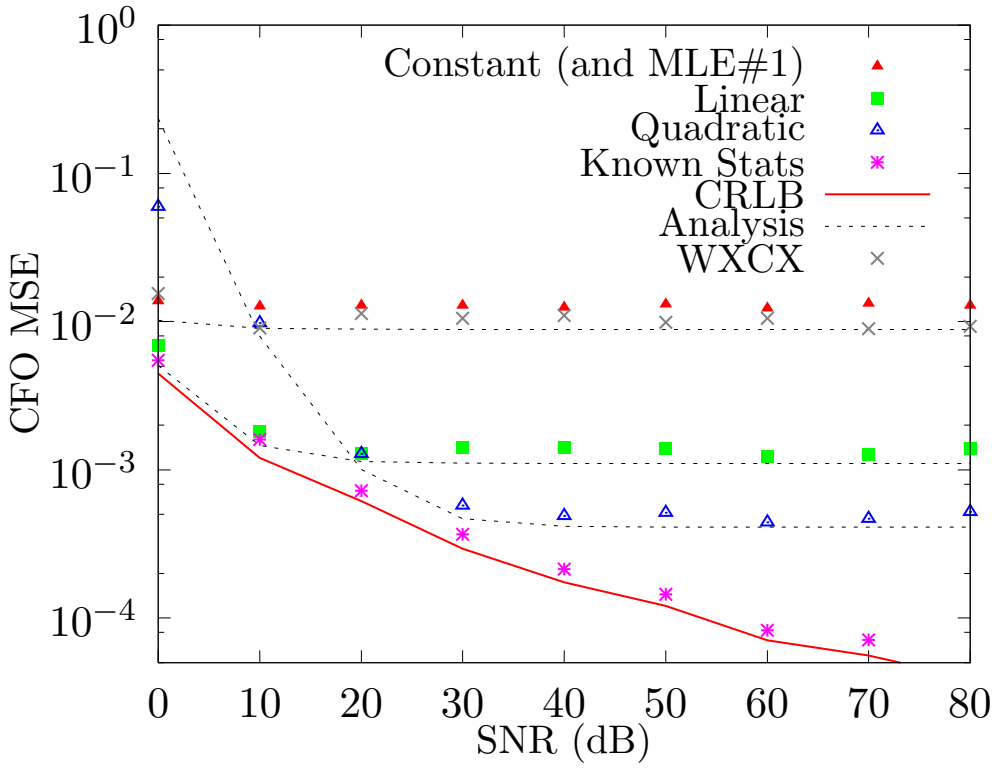

Figure 3.5: CFO Estimator performance vs. SNR for $f_dT = 0.3$

M). The lines show the CRLB and variance approximations while the points show simulated performance of the estimators. For all levels of Doppler, the CRLB is shown to underbound all estimators and it is achieved by the estimator with known channel statistics, as predicted in section 3.4.2. At higher Doppler, in figures 3.4 and 3.5, the error floors of the estimators become apparent, with a higher order model used for the parameterization corresponding to a lower error floor. However, the estimators based on higher order models experience poor performance at low SNR and low Doppler due to overparameterization, i.e., the channel parameters corresponding to the higher order basis functions contain very little energy, and at low SNR estimation error is more significant than modeling error.

When the channel statistics are unknown it is difficult to choose the proper model to use for channel parameterization. Using more parameters ensures a lower error floor, but results in poor performance at low SNR. Using fewer parameters yields better performance at low SNR, but results in a higher error floor.

The WXCX estimator, which was designed for a fast varying channel, performs the same as our unknown statics estimator using a constant basis function, even though

it is able to use twice the amount of training data. This is due to the fact that in its derivation, the channel is assumed constant over each of its two pilot sequences. The extra amount of training data does not improve its performance because the second sequence merely resolves ambiguity in the CFO estimate that arises due to the CAZAC structure of the first sequence.

3.6 Conclusion

We have derived joint CFO/channel estimators for use in a time-varying frequency-selective channel for both known and unknown channel statistics. When the statistics are known, the joint estimation reduces to sequentially estimating the CFO via ML and the channel via MAP or MMSE, and the CFO estimator achieves the CRLB. When the statistics are unknown, the channel must be parameterized and performance suffers due to over or under-parameterization. Thus knowledge of the channel statistics is extremely important to achieve the best performance in all scenarios.

Chapter 3, in part, is a reprint of the paper "Data-Aided Joint Estimation of Carrier Frequency Offset and Frequency-Selective Time-Varying Channel", Nathan D. Ricklin and James R. Zeidler, as it appeared in the proceedings of IEEE ICC, Beijing, 2008.

Chapter 4

Blind Detection in Time-Varying Channels

Differential phase shift keying (DPSK) is a simple and robust modulation technique that is useful in a wide range of communication channels, as it is simple to implement and does not require channel knowledge at the receiver. The well-known performance penalty of DPSK as compared to coherent detection can be partially overcome by jointly detecting multiple consecutive symbols, known as multiple-symbol differential-detection (MSDD) [23–25].

When the receiver and/or transmitter are mobile, a Doppler shift spreads the received signal in frequency, and causes the channel to vary in time. The noncoherent nature of DPSK makes it well suited to time-varying channels, as no channel estimation or tracking is required, and performance is good as long as the channel is approximately constant between two adjacent symbols. However, in rapidly varying channels, the performance of DPSK and multiple-symbol extensions suffers due to the incorrect assumption that the channel is constant, and error floors result. In this case, the performance of MSDD is actually worse than that of DPSK since it implicitly assumes that the channel is constant over more than two consecutive symbols.

If the time-varying channel statistics (e.g. channel autocorrelation and SNR for Rayleigh fading, K-factor and line-of-sight Doppler frequency for Rician fading) are known to the receiver, these error floors can be suppressed via a variety of methods. In [26], a maximum-likelihood sequence estimation (MLSE) block detection method was developed for use in Rayleigh fading channels, and a corresponding fast implementation

based on sphere decoding was described in [1]. In [27] a similar MLSE block detection method was developed for the Rician channel, which is more complicated to implement, and there is no known low-complexity implementation. Other known-statistics techniques include [28], in which the carrier phase between adjacent symbols is modelled as a uniform random variable; [29], a sliding-window Viterbi implementation of MLSE in Rayleigh fading; and [30], a sliding-window technique based on decision feedback. These methods all perform well if the assumed channel model is correct and the statistics of the chosen model are known, but obtaining channel statistics is not always practical, especially in rapidly time-varying channels. Additionally, there will always be errors in the estimated statistics, degrading the practical performance of such schemes.

If time-varying channel statistics are not available to the receiver, there are also a number of methods capable of reducing the performance error floors due to channel time-variation. In [31], a transformation on the data sequence is given such that the receiver is able to cancel out successively higher orders of carrier-phase time-variation, but is only applicable to channels with non-time-varying amplitude. In [32], a decision-feedback receiver which uses linear prediction to estimate the channel is developed. Two other unknown-statistics sliding-window type receivers rely upon parameterizing the time-varying channel with a polynomial (or Taylor series) model, using linear-prediction for channel estimation and per-survivor processing for data detection [33, 34].

There are limitations to all of the above approaches in in rapidly time-varying channels. Coherent detectors require accurate channel estimation, requiring pilot symbols and possible feedback–which compromises throughput. Also, when the channel is varying fast enough, its estimate is always out of date. When the communication dwell interval is very short, as with frequency-hopping or burst-mode systems, it may be difficult or impossible to obtain accurate channel estimates or channel statistics, and there is even less overhead available for pilot symbols. Additionally, the short communication interval renders decision feedback and sliding-window based receivers impractical.

In this chapter we develop a blind detector that jointly recovers a short block of symbols transmitted through a rapidly time-varying channel for which the above receivers either do not work or require too much overhead. The detector developed for this scenario does not require overhead for pilot symbols, nor feedback of any kind. Additionally, the channel statistics need not be known at the receiver.

To develop our detector we first introduce a framework for parameterizing the

unknown time-varying channel with an arbitrary set of orthonormal basis functions. This parameterization reduces the dimensionality of the joint maximum-likelihood channel estimation/data detection problem, and enables us to derive a data-detection cost function that is independent of channel statistics, and is thus applicable to any type of time-varying channel (e.g. Rayleigh, Rician, etc.). We then show how the cost function may be efficiently implemented via sphere decoding which can be viewed as an extension of [1] to our unknown-statistics framework.

The choice of channel parameterization has a direct impact on the performance of the detector. If the channel parameterization does not match the actual time-variation of the channel, the performance can degrade either with the appearance of error floors or by over fitting the noise.

To overcome the problem of how to select the channel parameterization, we show that using the so-called "subset" technique [35] can reduce the detector's sensitivity to the choice of channel parameterization, and yields near-optimal performance over a large SNR and Doppler range of interest.

This chapter is organized as follows: In section 4.1 we describe the system model, dwell structure, and specific communication scenario which we will investigate. We then derive the MSDD-US statistics-independent, data-detection cost function in section 4.2 for which we provide a theoretical performance analysis in section 4.3. In section 4.4 we describe a reduced-complexity implementation for the derived cost function based on sphere decoding. In section 4.5 we describe the optimal detectors (which rely on known channel statistics) which we then compare to our detector in section 4.6. We treat the issue of selection of the channel parameterization order in section 4.7 and discuss implementation complexity in section 4.8. Finally, in section 4.9, we conclude.

4.1 System Model

The length-N complex-baseband sequence to be transmitted is given by $x_n, 1 \leq n \leq N$, where x_n is the symbol transmitted in the nth signaling interval $(n-1)T \leq t \leq nT$, and T is the time-duration of a single symbol interval. As in Fig. 4.1, information bits are mapped to PSK symbols s_n, which are then differentially encoded to obtain x_n as

$$x_n = s_n x_{n-1}, \quad n = 2, \ldots, N \tag{4.1}$$

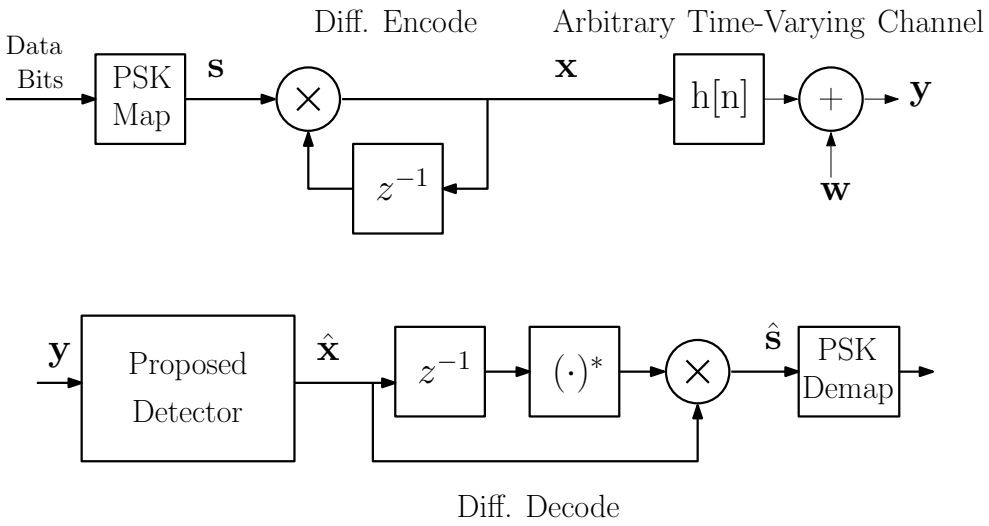

Figure 4.1: System model showing the transmit chain: PSK mapper and differential encoder, the frequency-flat time-varying channel with AWGN noise process, and receive chain: our proposed MLSE detector followed by differential decoder and PSK unmapper.

where $s_n = \exp(j\theta_n)$ is an M-ary PSK information symbol with $\theta_n \in \{2\pi m/M, m = 0, \ldots, M-1\}$ and x_1 is an arbitrary phase reference. For convenience, we set $x_1 = 1$ in the rest of this chapter. Each length-N sequence contains $N-1$ differentially encoded information symbols.

After transmission through a frequency-flat time-varying channel, the sampled complex-baseband received sequence y_n is given by

$$y_n = h_n x_n + w_n, \quad n = 2, \ldots, N \tag{4.2}$$

where w_n is zero-mean complex Gaussian circularly-symmetric noise with variance σ^2 and h_n is the time-varying channel at time n. We make no assumption on the channel model. We assume only that the underlying continuous-time channel process is continuous and relatively "smooth", since some degree of channel time-correlation is required for data detection.

4.2 MLSE Detection

Detection and recovery of the information bits can be done in three steps as illustrated in the receive chain of Fig. 4.1. The estimated transmitted data sequence $\hat{x}_n$ is first recovered from the data y_n, and then $\hat{x}_n$ is differentially decoded to yield

an estimate of the data symbols $\hat{s}_n$, which are then unmapped from PSK symbols to recover the information bits. In the rest of this chapter, we develop a detector which addresses the first step of recovering $\hat{x}_n$ from y_n for an unknown frequency-flat time-varying channel.

The received sequence y_n can be represented in vector form as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w} \tag{4.3}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$, $\mathbf{w} = [w_1, w_2, \cdots, w_N]^T$, $\mathbf{h} = [h_1, h_2, \cdots, h_N]^T$, and the diagonal data matrix $\mathbf{A} = \text{diag}\{\mathbf{x}\}$, with $\mathbf{x} = [x_1, x_2 \cdots, x_N]^T$.

We have no information on the statistics of the channel process h_n, so we consider it deterministic and unknown for the purpose of maximum likelihood sequence estimation. The PDF of the received data $\mathbf{y}$ in AWGN conditioned on both the channel and the transmitted data is

$$p(\mathbf{y}|\mathbf{x}, \mathbf{h}) = \frac{1}{\pi^N |\sigma^2 \mathbf{I}|} \exp\left\{-\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{h}\|^2\right\} \tag{4.4}$$

According to MLSE, the optimal data and channel estimates are given by

$$\{\hat{\mathbf{h}}, \hat{\mathbf{x}}\} = \arg\max_{\mathbf{h},\mathbf{x}} p(\mathbf{y}|\mathbf{x}, \mathbf{h})$$

which is equivalent to minimizing the cost function

$$\Lambda(\mathbf{x}, \mathbf{h}) = \|\mathbf{y} - \mathbf{A}\mathbf{h}\|^2 \tag{4.5}$$

over both $\mathbf{h}$ and $\mathbf{x}$. Unfortunately, minimizing (4.5) directly provides no information about the transmitted data $\mathbf{x}$, since for any fixed $\mathbf{x}$ the channel that minimizes (4.5) is $\hat{\mathbf{h}} = \mathbf{A}^{-1}\mathbf{y}$. Substitution of $\hat{\mathbf{h}}$ for $\mathbf{h}$ in (4.5) results in $\Lambda(\mathbf{x}, \hat{\mathbf{h}}) = 0$ for any value of $\mathbf{x}$. Thus, data detection is impossible with this cost function.

Intuitively, the problem is that in (4.5) the values of the channel $\mathbf{h}$ are completely unconstrained, i.e. there are enough degrees of freedom in the variability of the channel that the cost function can always be forced to zero, even though the resulting "optimal" channel could be disjoint and unrealistic.

To limit the variability of $\mathbf{h}$ we constrain allowable $\mathbf{h}$ vectors to a subset of the complex vector space:

$$\begin{aligned} \min_{\mathbf{h},\mathbf{x}} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{h}\|^2 \\ \text{s.t.} \quad & \mathbf{h} \in \mathcal{H} \end{aligned} \tag{4.6}$$

where $\mathcal{H}$ is the set of allowable $\mathbf{h}$ vectors with limited variation.

If we let $\mathcal{H}$ be a vector subspace, then the constraint $\mathbf{h} \in \mathcal{H}$ is equivalent to $\mathbf{h} = \mathbf{F}\mathbf{F}^H\mathbf{h}$ where $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_I]$ contains the orthonormal basis vectors that span the subspace, where I is the number of bases vectors used. This constraint is equivalent to approximating the channel as a linear combination of a few orthonormal basis vectors:

$$\mathbf{h} \approx \mathbf{F}\boldsymbol{\alpha}, \tag{4.7}$$

where $\boldsymbol{\alpha} = \mathbf{F}^H\mathbf{h}$ is a length I vector of coefficients corresponding to each basis vector. Thus, our channel approximation (equivalently, the subspace of allowed channel variability) allows time-variation of the type:

$$h_n = \sum_{i=1}^{I} \alpha_i f_i[n], \tag{4.8}$$

where $f_i[n]$ is the nth element of the ith basis vector.

The orthonormal bases f_i can be chosen to allow any sort of time variation, and thus can accommodate, for example, the basis expansion model (BEM) [4], discrete prolate spheroidal sequences [5], or a simple polynomial approximation:

$$h_n = \alpha_1 + \alpha_2 n + \alpha_3 n^2 + \ldots \tag{4.9}$$

Applying the above constraint to (4.6) (which is accomplished by simply replacing $\mathbf{h}$ with $\mathbf{F}\mathbf{F}^H\mathbf{h}$), we have

$$(\hat{\mathbf{h}}, \hat{\mathbf{x}}) = \min_{\mathbf{h},\mathbf{x}} \left\|\mathbf{y} - \mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{h}\right\|^2. \tag{4.10}$$

The $\hat{\mathbf{h}}$ that minimizes (4.10) for any particular $\mathbf{x}$ is given by

$$\hat{\mathbf{h}} = \mathbf{F}\mathbf{F}^H\mathbf{A}^H\mathbf{y}. \tag{4.11}$$

Substituting (4.11) into (4.10) gives us a cost function that is solely a function of the transmitted data:

$$\Lambda(\hat{\mathbf{h}}, \mathbf{x}) = \Lambda(\mathbf{x}) = \left\|\mathbf{y} - \mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H\mathbf{y}\right\|^2 \tag{4.12}$$

or, after dropping terms constant in $\mathbf{x}$,

$$\Lambda(\mathbf{x}) \doteq -\mathbf{y}^H\mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H\mathbf{y}. \tag{4.13}$$

This cost function is completely independent of the statistics of both the time-varying channel and the additive white Gaussian noise, and is therefore suitable for data detection in general time-varying channels of unknown distribution. Finally, in order to construct and implement this cost function the number of basis functions I that are used to parameterize the channel must be determined. The method used to select the value of I is treated in Section 4.7.

Note that the cost function in (4.13) contains a phase ambiguity, i.e. the cost of the data sequence $\mathbf{x}e^{j\theta}$ is the same for any value of θ. It is the requirement that this phase ambiguity be resolved (as well as the added robustness to time-variation) that motivates the use of a differentially encoded data sequence.

Since $\mathbf{F}$ is defined as a size N by I matrix composed of I orthonormal basis vectors, we have the condition that $I \leq N$. If $\mathbf{F}$ is a square matrix, or $I = N$, then we have $\mathbf{F}\mathbf{F}^H = \mathbf{I}$, and the cost function in (4.13) reduces to $\Lambda(\mathbf{x}) = -\mathbf{y}^H\mathbf{y}$, independent of the transmitted data $\mathbf{x}$, and maximization over the data symbols is impossible. Thus, we have the condition that $I \leq N-1$, i.e. we must use at most one less basis vector than the number of data symbols in $\mathbf{y}$.

Re-writing the cost function (4.13) in scalar form, we have

$$\Lambda(\mathbf{x}) = -\sum_{i=1}^{I}\left|\sum_{n=1}^{N} y_n x_n^* f_i^*[n]\right|^2. \tag{4.14}$$

To force the channel to be constant, we set $I = 1$ with $f_1[n] = 1/\sqrt{N}$. In this case the cost function reduces to

$$\Lambda(\mathbf{x}) = -\sum_{i=1}^{I}\left|\sum_{n=1}^{N} y_n x_n^* f_i^*[n]\right|^2 = -N\left|\sum_{n=1}^{N} y_n x_n^*\right|^2$$

which is equivalent to the MSDD cost function derived in [23] for multiple symbol DPSK in a constant channel.

4.3 Performance Analysis

For the MSDD-US system analyzed in this work, the probability of bit error can be approximated by

$$P_b \approx \frac{1}{\log_2(M)(N-1)}\sum_{\mathbf{x}'\in\chi} w(\mathbf{x}\rightarrow\mathbf{x}')P(\mathbf{x}\rightarrow\mathbf{x}') \tag{4.15}$$

where $N-1$ is the number of information symbols in the transmitted sequence $\mathbf{x}$, M is the size of the symbol alphabet, χ is the set of dominant error events, $w(\mathbf{x} \to \mathbf{x}')$ is the hamming weight or number of differing bits in the underlying data sequences s_n and s'_n, and $P(\mathbf{x} \to \mathbf{x}')$ is the pairwise error probability (PEP), the probability that the receiver chooses $\mathbf{x}'$ over $\mathbf{x}$ when $\mathbf{x}$ was transmitted.

We consider the same dominant error events as in [26], i.e. we define the set of dominant error events as those that maximize the correlation metric

$$\mu = \left| 1 + \sum_{k=2}^{N} x_n^* x'_n \right|. \tag{4.16}$$

When $N > 2$ and $M > 2$, there are $2N$ such error events: each x'_n being $\pm 2\pi/M$ away from the true x_n, and $\{x'_2, \ldots, x'_N\}$ all being $\pm 2\pi/M$ away from the true values. When $N = 2$ there are two dominant error events: x'_2 being $\pm 2\pi/M$ away from x_2.

Each of the dominant error events satisfying (4.16) differs from the transmitted sequence in exactly one symbol location. Since the underlying data is differentially encoded, if the bits are properly mapped to PSK symbols with Gray coding, then one dominant symbol error in $\mathbf{x}'$ corresponds to two bit errors: one bit error reaching the errant symbol, and one leaving it. If the symbol error is at the beginning or end of the block, only one bit is in error.

Thus, the hamming weight of a dominant error event is

$$w(\mathbf{x} \to \mathbf{x}') = \begin{cases} 1, & i = 1, i = N \\ 2, & 2 \le i \le N-1 \end{cases} \tag{4.17}$$

where i is the location of the symbol error.

The receiver decides upon $\hat{\mathbf{x}}$ if

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \Lambda(\mathbf{x}) \tag{4.18}$$

so to make the pairwise error $\mathbf{x} \to \mathbf{x}'$, we require $\Lambda(\mathbf{x}) > \Lambda(\mathbf{x}')$. The PEP is then $P(\mathbf{x} \to \mathbf{x}') = P(\Lambda(\mathbf{x}') - \Lambda(\mathbf{x}) \le 0)$ or

$$\begin{aligned} P(\mathbf{x} \to \mathbf{x}') &= P(\mathbf{y}^H(\mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H - \mathbf{A}'\mathbf{F}\mathbf{F}^H\mathbf{A}'^H)\mathbf{y} \le 0) \\ &= P(\Delta \le 0). \end{aligned}$$

where we have defined

$$\Delta \equiv \mathbf{y}^H(\mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H - \mathbf{A}'\mathbf{F}\mathbf{F}^H\mathbf{A}'^H)\mathbf{y}. \tag{4.19}$$

To evaluate the PEP we need to compute the integral

$$P(\Delta \leq 0) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\Phi_\Delta(s)}{s} ds \tag{4.20}$$

where $\Phi_\Delta(s)$ is the moment generating function of Δ and c is an arbitrary constant in the real part of the region of convergence of $\Phi_\Delta(s)$. As shown in [36],

$$\begin{aligned} P(\Delta \leq 0) \quad &\approx \quad \frac{1}{\nu} \sum_{k=1}^{\nu/2} \mathrm{Re}\{\Phi_\Delta(c + jc\tau_k)\} \\ &+ \quad \tau_k \mathrm{Im}\{\Phi_\Delta(c + jc\tau_k)\} \end{aligned} \tag{4.21}$$

where $\tau_k = \tan((2k-1)\pi/(2\nu))$, the parameter ν is even, and $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary parts, respectively. The approximation can be made tighter by increasing the value of ν. To ensure that c is within the real part of the region of convergence of $\Phi_\Delta(s)$, it is chosen as half the real part of the right-hand pole of $\Phi_\Delta(s)$ closest to the imaginary axis.

Using (4.21), all that is required to evaluate the error performance of the proposed detection scheme under various types of fading is to determine the moment generating function $\Phi_\Delta(s)$ for each particular fading process of interest.

4.3.1 Rician Fading

Under Rician fading the channel h_n is composed of the sum of a scattering component and a time-varying line-of-sight (LOS) component [37]. In other words h_n is distributed as $\mathcal{CN}(\bar{h}_n, \sigma_h^2)$ where $\bar{h}_n = E[h_n] = \mu \exp(j2\pi f_l T n)$ is the time-varying LOS component, σ_h^2 is the power of the scattering component, and μ is a complex constant that does not vary in time. The LOS Doppler shift f_l depends upon the maximum Doppler shift f_d as defined by $f_l = f_d \cos(\theta_0)$, where θ_0 is the angle of arrival of the LOS component relative to the difference in velocity between the receiver and transmitter.

The Rician K-factor is defined as the ratio of power in the LOS component to the scattering component: $K \equiv |\mu|^2/\sigma_h^2$. We normalize total channel power to one: $E[|h_n|^2] \equiv 1 = |\mu|^2 + \sigma_h^2$ and so $|\mu|^2 = K/(K+1)$ and $\sigma_h^2 = 1/(K+1)$.

The channel is therefore given by

$$h_n \quad = \quad \sqrt{\frac{K}{K+1}} e^{(j2\pi f_l T n)} e^{j\theta} + \sqrt{\frac{1}{K+1}} v_n \tag{4.22}$$

where θ is the phase of μ and v_n, the scattering component of the channel, is a complex Gaussian circularly-symmetric random variable distributed as $\mathcal{CN}(0,1)$.

The time-correlation of the channel scattering component v_n is defined by the Jakes' model [38] as $E[v_n v_m^*] = J_0(2\pi f_d T(n-m))$ where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind.

The mean of the vector form of the channel $\mathbf{h}$ is

$$\bar{\mathbf{h}} = E[\mathbf{h}] = \sqrt{\frac{K}{K+1}}\mathbf{u}e^{j\theta}$$

where the elements of $\mathbf{u}$ are $u_n = \exp(j2\pi f_l T n)$ and the covariance is

$$\mathbf{C}_h = E[(\mathbf{h}-\bar{\mathbf{h}})(\mathbf{h}-\bar{\mathbf{h}})^H] = \frac{1}{K+1}\mathbf{R}_J$$

where $\mathbf{R}_J$ is the time-correlation matrix of the scattering component of the channel with elements $\mathbf{R}_{J\{n,m\}} = E[v_n v_m^*]$.

Using the channel model given in (4.22), $\mathbf{y}$ is a complex Gaussian vector with mean

$$\bar{\mathbf{y}} = E[\mathbf{y}] = \sqrt{\frac{K}{K+1}}\mathbf{A}\mathbf{u}e^{j\theta} \tag{4.23}$$

and covariance

$$\begin{aligned}\mathbf{C}_y &= E[(\mathbf{y}-\bar{\mathbf{y}})(\mathbf{y}-\bar{\mathbf{y}})^H] \\ &= \frac{1}{K+1}\mathbf{A}\mathbf{R}_J\mathbf{A}^H + \sigma^2\mathbf{I}.\end{aligned} \tag{4.24}$$

Thus, the random variable Δ is a Hermitian quadratic form of a complex normal random variable. From [39] the moment generating function of Δ is given by

$$\Phi_\Delta(s) = \frac{\exp\{-\bar{\mathbf{y}}^H\mathbf{C}_y^{-1}\left(\mathbf{I}-(\mathbf{I}+s\mathbf{C}_y\mathbf{Q})^{-1}\right)\bar{\mathbf{y}}\}}{|\mathbf{I}+s\mathbf{C}_y\mathbf{Q}|} \tag{4.25}$$

where $\mathbf{Q} \equiv \mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H - \mathbf{A}'\mathbf{F}\mathbf{F}^H\mathbf{A}'^H$. Error performance in Rician fading can now be easily computed by using (4.21) and evaluating $\Phi_\Delta(s)$ at the specified points.

4.3.2 Rayleigh Fading

Rayleigh fading is equivalent to Rician fading with Rician K-factor equal to zero. There is no LOS component and the channel is therefore $h_n = v_n$ with the same time-correlation as in the previous section.

Under Rayleigh fading the received data vector $\mathbf{y}$ is a complex Gaussian vector with mean $\bar{\mathbf{y}} = 0$ and covariance

$$\mathbf{C}_y = E[\mathbf{y}\mathbf{y}^H] = \mathbf{A}\mathbf{R}_v\mathbf{A}^H + \sigma^2\mathbf{I}.$$

The moment generating function in this case is much simpler and is given by

$$\Phi_{\Delta}(s) = \frac{1}{|\mathbf{I} + s\mathbf{C}_y\mathbf{Q}|},$$

and the PEP can be numerically evaluated with (4.21).

Alternatively, it is possible to obtain the pairwise error probabilities for Rayleigh fading in closed form. If $\mathbf{x}$ and $\mathbf{x}'$ differ in the pth symbol position, then the PEP $P(\mathbf{x} \to \mathbf{x}') = P(\Delta_p \leq 0)$ where there random variable Δ_p can be written as

$$\Delta_p = \Re\{2\mathbf{XY}\}$$

with

$$\begin{aligned} \mathbf{X} &= y_p^*(x_p - x_p') \\ \mathbf{Y} &= \sum_{n=1, n\neq p}^{N} y_n^* x_n \sum_{i=1}^{I} f_i(n) f_i^*(p). \end{aligned}$$

Both $\mathbf{X}$ and $\mathbf{Y}$ are zero-mean complex Gaussian random variables, expressions developed in [40] reduce to

$$P(\Delta_p \leq 0) = \frac{1}{2}\left(1 - \frac{1}{\sqrt{1 + \frac{2}{(1-cos(2\pi/M))\rho_p}}}\right), \tag{4.26}$$

where

$$\rho_p = \frac{|\mu_{xy}|^2}{\sigma_x^2\sigma_y^2 - |\mu_{xy}|^2}, \tag{4.27}$$

$$\sigma_x^2 = |x_p - x_p'|^2(1 + \sigma_n^2), \tag{4.28}$$

$$\sigma_y^2 = \sum_{n=1, n\neq p}^{N} \sum_{m=1, m\neq p}^{N} \mathbf{C}_{y\{n,m\}} x_n x_m^* \sum_{i=1}^{I} f_i(n) f_i^*(p) \sum_{k=1}^{I} f_k^*(m) f_k(p), \tag{4.29}$$

$$\mu_{xy} = (x_p - x_p') \sum_{n=1, n\neq p}^{N} \mathbf{C}_{y\{n,p\}} x_n^* \sum_{i=1}^{I} f_i^*(n) f_i(p). \tag{4.30}$$

Thus, the pairwise error probabilities for the dominant error events in Rayleigh fading can be calculated using (4.26)-(4.30).

4.4 Complexity Reduction via Sphere Decoding

The complexity of a brute force search over the cost function in (4.13) is exponential in the block size N, since the cost function must be evaluated for each of M^{N-1}

possible transmitted sequences. Sphere decoding is a search-space reduction technique for finding the closest lattice point to a given point (see [41] and [42]).

To reduce the complexity of our detector, we employ sphere-decoding with radius update and the Schnorr-Euchner [43] search strategy as presented in [1]. The algorithm in [1] relies on known Rayleigh fading statistics and is capable of efficiently solving a shortest-vector problem of the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{U}\mathbf{x}\|^2 \tag{4.31}$$

where $\mathbf{U}$ is an upper triangular matrix.

To manipulate our cost function into the form of (4.31) we first note that

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \mathbf{y}^H \mathbf{A}(-\mathbf{F}\mathbf{F}^H)\mathbf{A}^H\mathbf{y} \\ &= \arg\min_{\mathbf{x}} \mathbf{y}^H \mathbf{A}(\epsilon\mathbf{I} - \mathbf{F}\mathbf{F}^H)\mathbf{A}^H\mathbf{y} \end{aligned} \tag{4.32}$$

for any value of ϵ since the additional term $\mathbf{y}^H\mathbf{A}(\epsilon\mathbf{I})\mathbf{A}^H\mathbf{y} = \epsilon\mathbf{y}^H\mathbf{y}$ is independent of the data vector $\mathbf{x}$.

We then rewrite (4.32) as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} (\mathrm{diag}\{\mathbf{y}\}\mathbf{x}^*)^H(\epsilon\mathbf{I} - \mathbf{F}\mathbf{F}^H)(\mathrm{diag}\{\mathbf{y}\}\mathbf{x}^*)$$

and apply the Cholesky factorization to obtain $(\epsilon\mathbf{I} - \mathbf{F}\mathbf{F}^H) = \mathbf{L}\mathbf{L}^H$ with the condition that $\epsilon > 1$ to ensure that $(\epsilon\mathbf{I} - \mathbf{F}\mathbf{F}^H)$ is positive definite (the nonzero eigenvalues of $\mathbf{F}\mathbf{F}^H$ are equal to one since $\mathbf{F}$ is composed of orthonormal vectors). Setting $\mathbf{U} = (\mathbf{L}^H\mathrm{diag}\{\mathbf{y}\})^*$ puts the problem into the form of (4.31), and implementation follows [1] closely.

Note on the choice of ϵ: The choice of ϵ greatly affects the complexity (number of visited nodes) of the sphere-decoding algorithm. For larger values of ϵ, a larger constant term is added to the cost function. The effect of this is a reduced relative spread between values of the cost function at different hypothesized sequence points. This requires the sphere-decoding algorithm to search farther down a particular path before learning that nodes beyond the current node are outside the sphere, resulting in a higher total number of visited nodes. For this reason, ϵ should be chosen as small as possible, yet still large enough to avoid problems with numerical precision. We have found that for a wide range of channel conditions and parameters, a value of $\epsilon = 1.001$ works well.

4.5 Optimal Known-Statistics Cost Functions

We would like to compare the performance of our MSDD-US detection scheme to detectors that are optimal in both the Rayleigh and Rician fading channel. In this section we review the optimal cost functions and discuss their complexity.

The optimal detectors are found by joint maximum likelihood detection of the differentially encoded information symbols and require full knowledge of the fading channel distributions.

For Rayleigh fading, the cost function to maximize for MLSE detection is [26]

$$\Lambda(\mathbf{x}) \quad = \quad \mathbf{y}^H \mathbf{A} \left(\mathbf{R}_J + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{A}^H \mathbf{y}. \tag{4.33}$$

This detector requires knowledge of both the SNR and the second-order statistics of the channel $E[\mathbf{h}\mathbf{h}^H]$.

For Rician fading, the cost function to maximize for MLSE detection was developed in [27]

$$\Lambda(\mathbf{x}) \quad = \quad \exp\left\{ -\mathbf{y}^H \mathbf{C}_y^{-1} \mathbf{y} \right\} I_0 \left(2 |\mathbf{z}^H \mathbf{C}_y^{-1} \mathbf{y}| \right) \tag{4.34}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. This detector requires knowledge of the SNR, the Rician K-factor, and the second-order statistics of the non-zero mean channel $E[(\mathbf{h} - \bar{\mathbf{h}})(\mathbf{h} - \bar{\mathbf{h}})^H]$.

It was shown in [1] that a reduced-complexity implementation exists for the optimal Rayleigh fading detector (4.33), however there is no fast way to implement the known-statistics Rician cost function given in (4.34). Consequently, the estimated transmitted data sequence must be found by a brute force search. This is in contrast with our detection strategy derived in section 4.2, which is implementable with a sphere-decoding algorithm. Thus, while the cost function of (4.34) achieves optimal performance in Rician fading, there is a huge penalty in terms of implementation complexity as compared to our sub-optimal detection scheme.

4.6 Performance Results

4.6.1 Channel Parameterization Vectors

We present four versions of our MSDD-US detector using between 2 and 5 basis vectors to parameterize the channel. For simplicity, we choose an orthonormalized

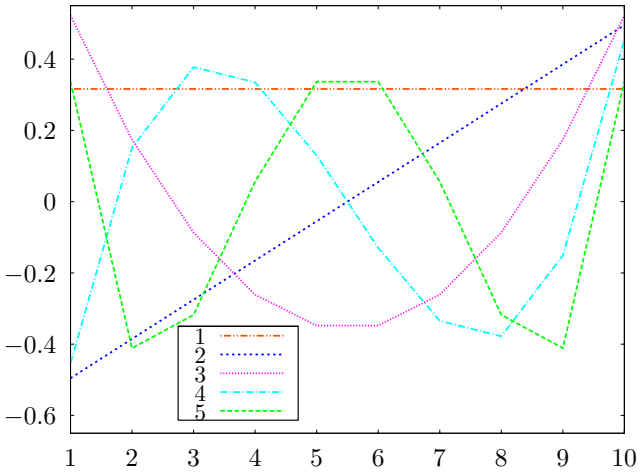

Figure 4.2: Orthonormalized Chebyshev polynomial-based basis vectors for $N = 10$.

Chebyshev polynomial basis in which to restrict the time-varying channel. Fig. 4.2 shows these basis vectors for a block length of $N = 10$. For example, if three basis vectors are to be used in the MSDD-US detector, the first three vectors (the constant, linear, and quadratic components) are used in the construction of the $\mathbf{F}$ matrix.

4.6.2 Rayleigh and Rician BER Performance

Here we present bit error rate (BER) performance results of our derived unknown-statistics multiple-symbol differential detector (MSDD-US) in frequency-flat Rayleigh and Rician fading channels, using a signal constellation of 4-PSK.

We compare the performance of our MSDD-US detector to:

- Standard DPSK
- Coherent detection with differential encoding ("Dif-Coh"), i.e. perfectly known channel state, which provides a lower bound on BER performance for any detection method when the information symbols are differentially encoded.
- Monte-Carlo simulations of the optimal known time-varying-statistics methods ((4.33) for Rayleigh fading, (4.34) for Rician fading), referred to here as the "Known Stats" method.

Fig. 4.3 shows BER performance in Rayleigh fading for a block length $N = 10$ and normalized Doppler shifts of $f_dT = 0.01$ and $f_dT = 0.03$. Fig. 4.4 shows BER

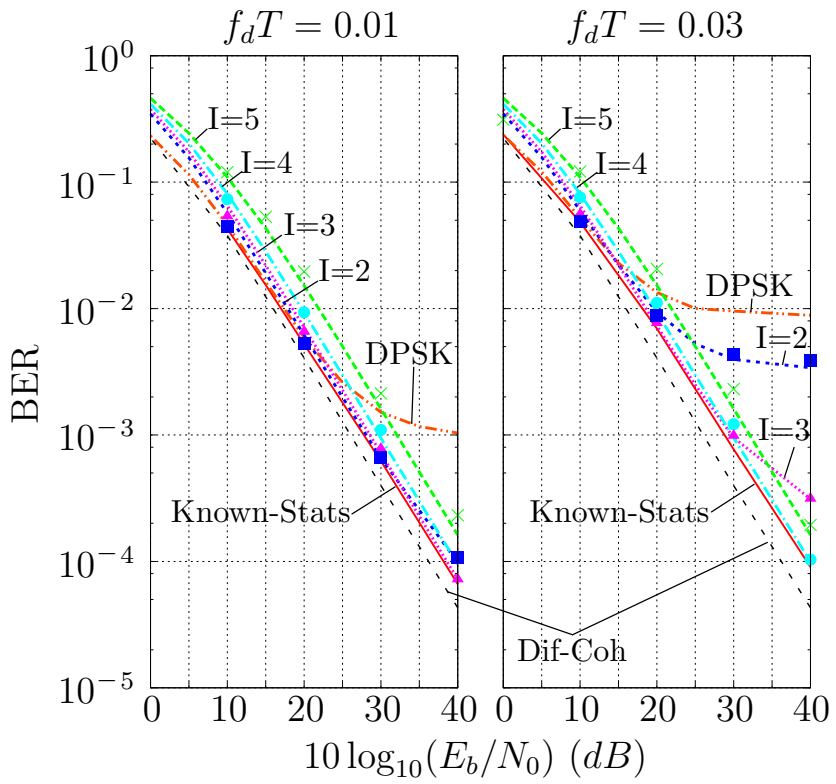

Figure 4.3: Bit error rate vs. SNR in Rayleigh fading for block size $N = 10$. Results are shown for parameterization orders of 2, 3, 4, and 5, and normalized Doppler values of $f_dT = 0.01$ and $f_dT = 0.03$. In both Doppler scenarios the performance is compared with 4-DPSK, the known-statistics detector, and differentially-encoded coherently-detected 4PSK. Curves are analysis and solid points are simulated.

performance in Rician fading for a block length of $N = 6$ with Rician K-factors of $K = 1$ and $K = 5$, and scattering component and LOS normalized Dopplers of $\{f_dT = 0.01, f_lT = 0\}$ and $\{f_dT = 0.03, f_lT = 0.01\}$. In both figures, lines show the theoretical performance as derived in section 4.3 and solid symbols show the simulated MSDD-US detection scheme using sphere decoding. Squares, triangles, circles, and "x"s correspond to parameterization orders of $I = 2, 3, 4$ and 5, respectively.

In each of these scenarios, all versions of the MSDD-US detector outperform ordinary DPSK above a certain SNR threshold, above which the DPSK detector's performance is degraded by being unable to account for the time-variation in the channel, evidenced by an error floor.

If too few basis vectors are used, error floors result, with worse performance for higher Doppler spreads. In these cases the chosen basis cannot fully describe all of the

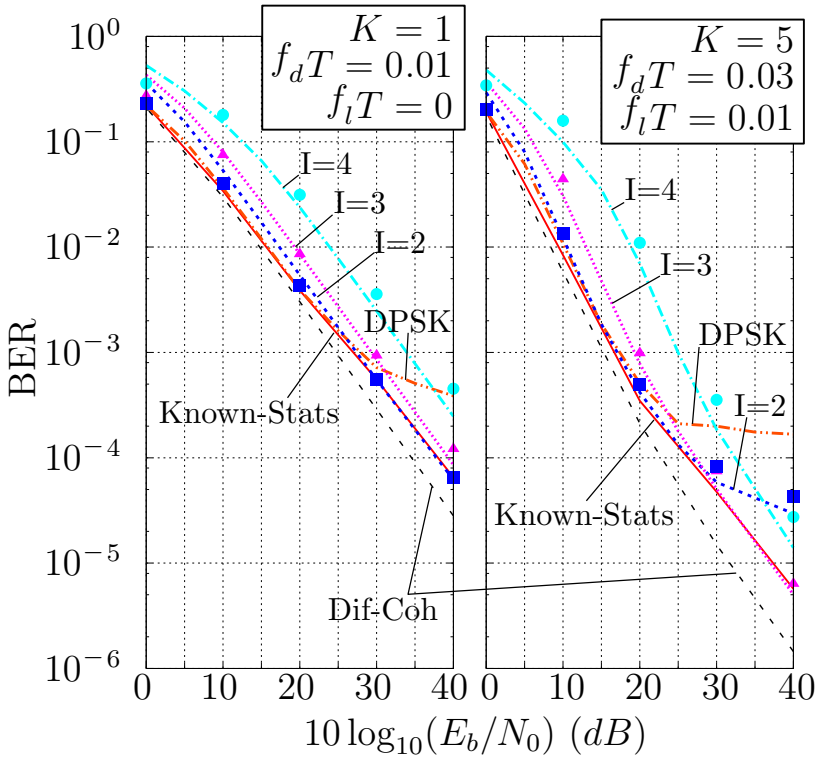

Figure 4.4: Bit error rate vs. SNR in Rician fading for block size $N = 6$. Results are shown for parameterization orders of 2, 3, 4, and 5, scattering Doppler values of $f_dT = 0.01$ and $f_dT = 0.03$, LOS Doppler values of $f_lT = 0$ and $f_lT = 0.01$, and Rician K-factors of $K = 1$ and $K = 5$. In both channel scenarios the performance is compared with 4-DPSK, the known-statistics detector, and differentially-encoded coherently-detected 4PSK. Curves are analysis and solid points are simulated.

channel time-variation, and channel estimation error plays a large role in overall BER performance. For example, in Fig. 4.3 for a normalized Doppler of $f_dT = 0.03$, choosing $I = 2$ overly constrains the allowable amount of time-variation as compared to the actual channel variation.

Conversely, if the channel order is overparameterized, there are no error floors, but there is a performance gap as compared to the optimal "known stats" detector that is most pronounced at lower levels of channel time-variation. This is due to the chosen basis allowing too much channel time-variation, or, in another sense, the channel estimate using the additional parameters to fit to the noise.

In each scenario, there is a parameterization order I, for which the MSDD-US detectors performance is very close to that of the optimal "known stats" detector. In this case, the actual time-variation of the channel is a close match to the chosen vector

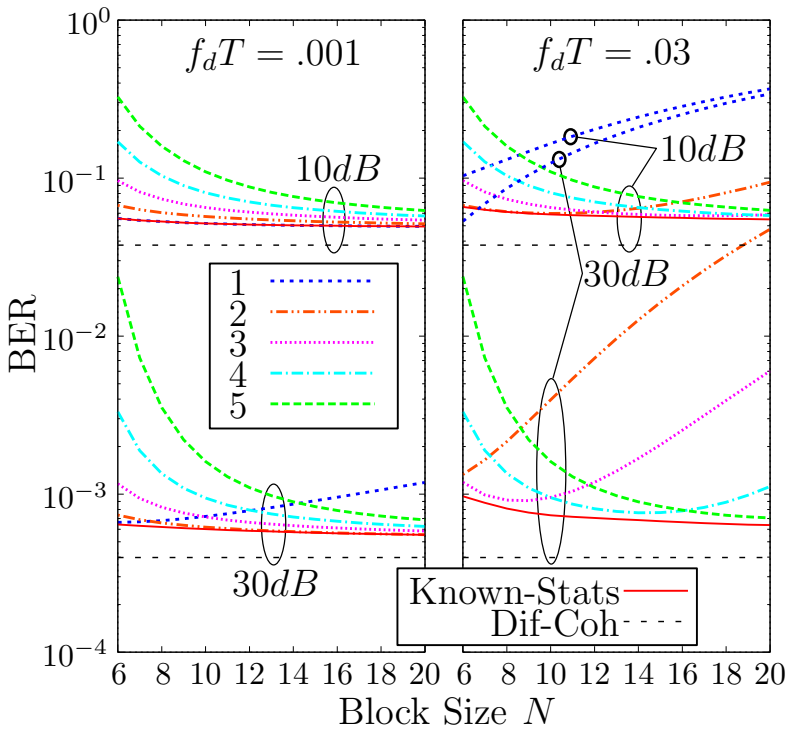

Figure 4.5: BER vs. block size N in Rayleigh fading for parameterization orders of 1, 2, 3, 4, and 5, SNR values of 10 dB and 30 dB, and Doppler values of $f_dT = 0.001$ and $f_dT = 0.03$.

space of allowable time variation.

The value of I is clearly closely linked to the normalized Doppler f_dT of the channel, with a larger I required as the Doppler increases. In order for the MSDD-US detector to be practical, it is necessary to find a priori combinations of block size N and parameterization order I that provide good performance over a wide range of channel conditions.

To accomplish this, in the next sections we examine performance vs. block size and performance vs. symbol location in order to draw conclusions about a proper choice of the parameterization order.

4.6.3 Performance vs. Block Size N

Fig. 4.5 shows Rayleigh BER performance vs. block size N for various parameterization orders, normalized Doppler shifts of $f_dT = 0.001$ and $f_dT = 0.03$, and SNR of $10dB$ and $30dB$.

Regions of negative slope correspond to overparameterization, i.e. there are too

many basis functions I as compared to the block size N and the channel estimate is over-fitting the noise. As N is decreased the performance degrades further because the space of allowable time-variation becomes larger relative to the full vector space. The performance degradations in these regions are most prominent at high SNR.

Regions of positive slope correspond to underparameterization, i.e. there are too few basis functions to describe the time-variation of the channel. The error floors that go along with underparameterization are indicated by the convergence of curves at differing SNRs, for example the parameterizations $I = 1$ and $I = 2$ at normalized Doppler $f_dT = 0.03$. As N is increased performance worsens because the channel is even more restricted relative to its length. The performance degradations in these regions are most prominent at high Doppler.

Points of zero slope correspond to an optimal choice of channel parameterization I, given the block size N.

Maximum sensitivity to the parameterization order I occurs when both the SNR and Doppler are high, which suggests that the choice of N and I should be based on performance in this regime. Following this logic, when $N = 10$, one should choose $I = 4$ or $I = 3$ for good performance up to an SNR of $30dB$ and Doppler up to $f_dT = 0.03$, which is validated in Fig. 4.4.

However, using this technique for selection of I, the performance is still subject to slight overparameterization within the range of interest, and to error floors outside this range.

4.6.4 Performance vs. Symbol Location

In Fig. 4.6 we examine the Rayleigh fading pairwise error probabilities of different symbols within the block for various parameterization orders, normalized Doppler shifts of $f_dT = 0.001$ and $f_dT = 0.03$, SNR of $10dB$ and $30dB$, and block size of $N = 10$.

It is clear that when the channel is overparameterized (e.g. $I = 5$ in Doppler of $f_dT = 0.001$), the PEPs of the two symbols on the outside of the block are significantly larger than the others, and contribute most to the degradation of the overall BER. Similarly, when the channel is underparameterized ($I = 1$ in Doppler of $f_dT = 0.03$), the PEPs of the two outside symbols are also the largest, but not as drastically so. Similar observations have been made for the known-statistics detectors ([40], [35]).

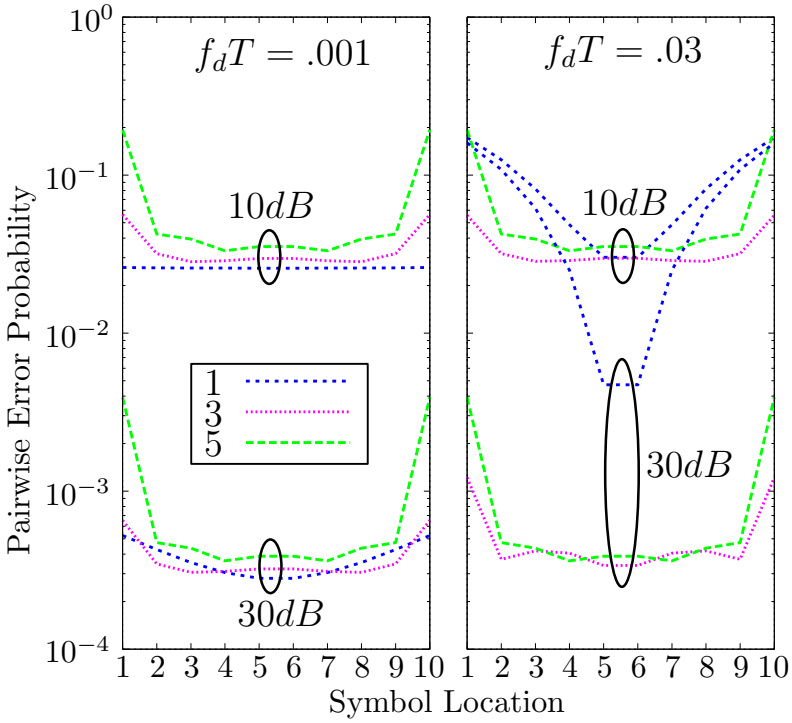

Figure 4.6: Pairwise error probability vs. location in the block of symbols for parameterization orders of 1, 3, and 5, SNR values of 10 dB and 30 dB, and Doppler values of $f_dT = 0.001$ and $f_dT = 0.03$.

4.6.5 Improving Performance via "Subset MSDD"

These observations suggest that using a subset method ([35]), i.e. using only the decisions on the innermost $N-2$ symbols and ignoring the two outside symbols can protect against performance degradation due to overparameterization as well as slightly improve the error floors resulting from underparameterization.

To illustrate the performance improvements of this concept, in Fig. 4.7 we plot BER of our detector in Rayleigh fading vs. normalized Doppler f_dT for various parameterization orders, an SNR of $30dB$, and block size $N = 10$, for both standard and subset MSDD.

Regions of positive slope correspond to an error floor which worsens as the Doppler increases, as a particular parameterization order cannot handle the time-variation of the channel. Curves that are constant across the whole Doppler range, such as the $I = 5$ detector, can fully describe the channel time-variation but suffer from overparameterization at lower Doppler. However, when the subset technique is used, the performance is greatly improved as the overparameterization effects have been suppressed.

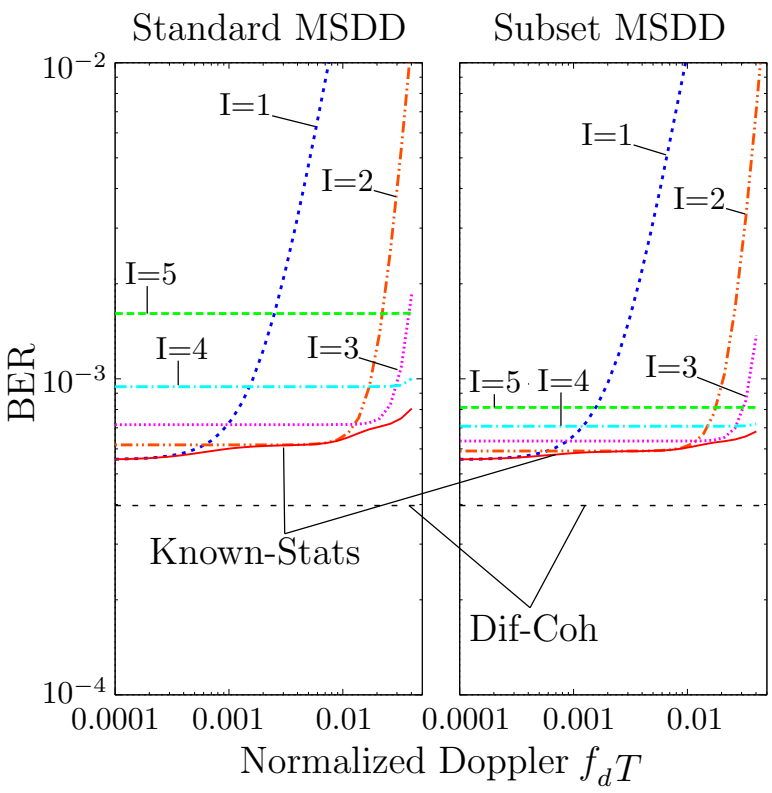

Figure 4.7: A comparison of standard MSDD-US and subset-MSDD-US: BER in Rayleigh fading vs. Normalized Doppler for parameterization orders of 1, 2, 3, 4, and 5, fixed SNR = $30dB$, and block size $N = 10$.

These results motivate the choice of a larger parameterization order I for a given block size N, as any possible overparameterization effects can be suppressed by using the subset-technique.

The trade-offs in using this technique are overhead and complexity for performance, as the two outside symbols must be decoded, but are not used.

4.7 Selection of the Channel Parameterization Order

The numerical results in the previous sections provide insight that allows us to select the channel parameterization order I for a given block size N to design detectors that perform well and are robust to high levels of channel time-variation.

In this section, we consider the block size N and choice of parameterization order I that provide good performance in a given SNR and Doppler range of interest. For example, we consider SNR and Doppler range of interest to be $0 - 40dB$ and 0.00 to

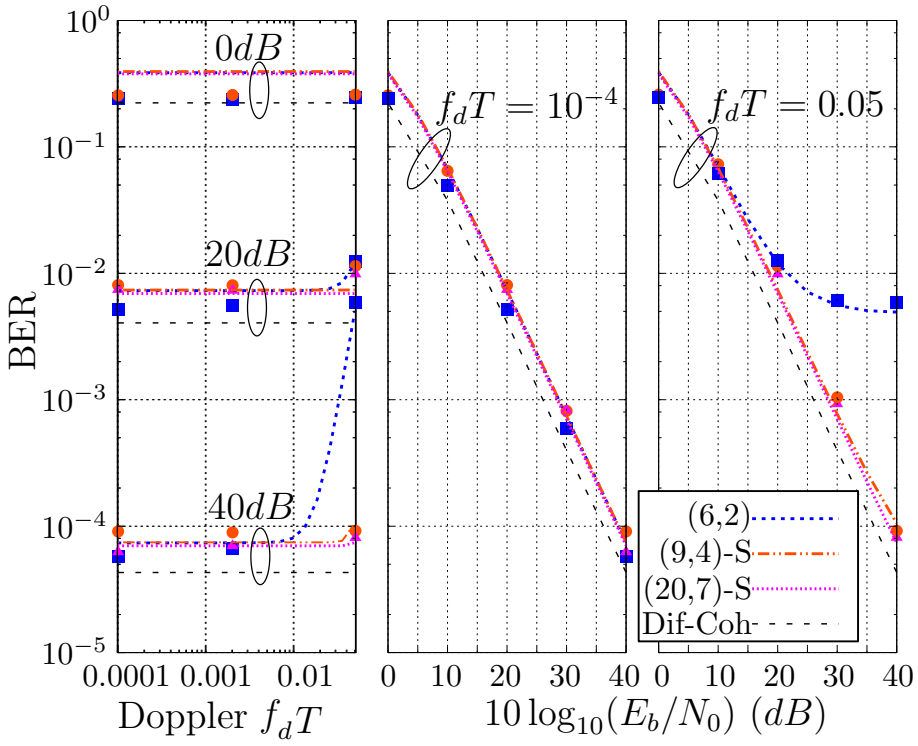

Figure 4.8: Bit error rate performance in Rayleigh fading of some particularly robust pairs of window size N and parameterization order I.

0.05, respectively. The block size N should be large enough to obtain joint-detection performance gains, but also small enough to keep complexity low. Additionally, using a small block size N with the subset technique requires much more overhead than a larger N. This means that for smaller N, we are more concerned with possible overparameterization since we cannot suppress it with the subset technique, hence we must choose I low and deal with error floors at the high end of our Doppler range.

For larger N we simply choose I large and to have a very low error floor outside our Doppler range of interest, and we rely on the subset technique to improve performance at lower Doppler and lower SNR.

In Fig. 4.8 we plot BER performance in Rayleigh fading vs. both Doppler and SNR for three well-performing combinations of block size and parameterization order $\{N, I\}$: $\{6, 2\}$, $\{9, 4\}$-subset, and $\{20, 7\}$-subset. Lines show theoretical performance from section 4.3 and solid symbols show simulated MSDD-US using sphere decoding. Squares, circles, and triangles correspond to $\{6, 2\}$, $\{9, 4\}$-subset, and $\{20, 7\}$-subset, respectively.

All three versions of the detector perform exceptionally well over the entire SNR and Doppler range of interest, with the exception of the $\{6, 2\}$ detector in high Doppler, for the reasons mentioned above. In particular the $\{20, 7\}$-subset detector performs within about $2.5dB$ of the differential-coherent lower bound over the entire range of interest.

4.8 Complexity

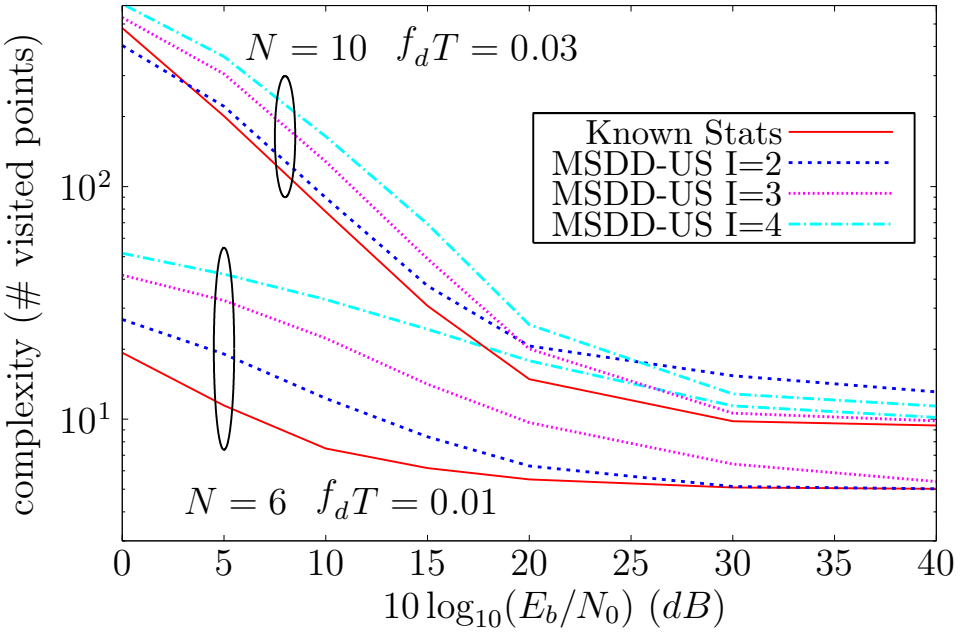

Figure 4.9: Complexity vs. $10 \log_{10}(E_b/N_0)$ of our detection method (MSDD-US) with sphere decoding compared to the known-statistics sphere decoding technique of [1] in Rayleigh fading.

In Fig. 4.9 we plot complexity vs. SNR of our detection method (MSDD-US) with sphere decoding compared to the known-statistics sphere decoding technique of [1] for a variety of parameterization orders in two different Rayleigh fading scenarios: $N = 10$ with $f_d T = 0.03$ and $N = 6$ with $f_d T = 0.01$. Our measure of complexity is the number of prospective sequence points visited by the algorithm (if a whole branch of sequence points is excluded, it is counted as one sequence point visited), which we can compare to the number of cost function evaluations required for a brute force search, which for $N = 6$ is $M^{(N-1)} = 1024$ and for $N = 10$ is 262,144. In all scenarios the complexity of MSDD-US with sphere-decoding is typically much less than an order of magnitude away

from that of the "known-stats" method [1], and is far superior to a brute-force search.

4.9 Conclusion

We have derived a block-detection method for multiple symbol DPSK for use in a statistically unknown time-varying channel. We provide a parameterization framework for the channel which allows us to derive a data-detection cost function that does not require channel statistics, and is thus suitable for use in a general statistically-unknown time-varying channel. Additionally, we have provided a reduced-complexity implementation based on sphere decoding, which is many orders of magnitude less complex than a brute-force cost function optimization. This is especially attractive for the case of Rician fading, for which there is no other reduced-complexity implementation.

We show that this truly blind detection method obtains near-optimal performance (within about $2.5dB$ of differential-coherent detection) over a wide range-of-interest of both SNR and Doppler.

Chapter 4, in part, is a reprint of the papers "Block Detection of Multiple Symbol DPSK in a Statistically Unknown Time-Varying Channel" Nathan D. Ricklin and James R. Zeidler, as it appeared in the proceedings of IEEE ICC, Dresden, 2009; "Multiple Symbol Differential Detection of MPSK over Statistically-Unknown, Frequency-Flat, Time-Varying Channels" Nathan D. Ricklin and James R. Zeidler, accepted for publication in the IEEE Transactions on Communications.

Chapter 5

ICI Reduction: Constant-Magnitude Correlative Coding

Orthogonal frequency division multiplexing (OFDM) and its multiple-access extension OFDMA, are widely used in wireless systems due to their high spectral efficiency, ease of implementation, and robustness to frequency selective channels. In particular, OFDMA has been used in commercial standards such as the LTE downlink [44], digital video broadcasting [45], and the wireless metropolitan-area network standard [46].

To maintain subcarrier orthogonality, the OFDM system must operate over a time-invariant channel. However, if the receiver or transmitter are mobile, a Doppler shift spreads the received signal in frequency, and causes the multipath channel to vary in time. This channel time-variation destroys the orthogonality of the subcarriers, resulting in so-called intercarrier interference (ICI), which degrades system performance [7].

Previous efforts to reduce the impact of ICI include frequency-domain equalization, which aims to invert the known or estimated frequency-domain channel matrix. To reduce the complexity of this task, [47] uses a short equalizer to undo the effect of only the significant nearby adjacent subcarriers, while [48] finds an efficient way to compute the inverse using FFTs. Such methods still entail significant complexity and require knowledge of the frequency channel matrix, which in a fast-fading environment can significantly change from OFDM symbol to OFDM symbol.

Other methods of reducing ICI include time-domain windowing [49], [50], but

these reduce spectral efficiency or rely on the guard interval being larger than the channel impulse response.

A third way to mitigate the effects of ICI involves coding the transmitted data sequence in frequency such that it has an autocorrelation that causes a reduction in ICI power. Such correlative coding techniques are attractive because they require no special modifications to the transmitter or receiver; the data is simply coded before transmission and decoded after it is received.

The so-called ICI self-cancellation technique [51] intersperses symbols with redundant 180 degree phase shifted copies to achieve the desired autocorrelation, at the expense of throughput. The ICI compression technique [52] takes a different approach: filtering the frequency-domain data with a 2-tap FIR filter with coefficients $[1, -1]$. Further analysis and performance bounds are provided in [53]. An extension to the ICI compression technique to FIR filters of length greater than 2 is derived in [2], and optimal coefficients are provided.

The drawback of the above correlative coding schemes are that they result in the transmission of non-constant magnitude symbols, and frequently null symbols. This has consequences for blind detection: non-constant magnitude symbols can increase the complexity of blind detection algorithms by requiring matrix inversions where they would otherwise be unnecessary, and null data symbols provide no information about the channel at their location.

In this chapter we first express ICI power in a form that makes clear the effect of the autocorrelation of the data sequence. After providing insight into how to design the autocorrelation, we present a simple constant-magnitude correlative coding (CMCC) scheme by which binary data is placed onto the unit circle in such a way that the resulting autocorrelation is favorable to reducing ICI power. The scheme is similar to both DPSK and symmetric DPSK [54], [55]. Also, a similar unit-circle modulation scheme for limiting adjacent-channel interference was outlined in [56].

We then present comparisons that show that the proposed CMCC scheme reduces ICI power similarly to that of the Partial Response Coding (PRC)-based schemes of [2] and [52], and that a parameter in the CMCC scheme, which we term the modulation angle, can be adjusted to increase or decrease the degree of ICI power reduction.

Next, we design a blind detector for a short block of contiguous CMCC-encoded symbols in the frequency-domain (e.g. the block of symbols could be an OFDMA subcar-

rier allocation) that have been transmitted through a time-varying frequency-selective channel. The constant-magnitude nature of the encoded data directly enables efficient optimization methods such as sphere decoding to be applied to the blind detection problem. This is not possible with varying-magnitude correlative coding schemes such as PRC, which must use a brute-force search in their associated blind detectors.

Performance simulations show that the CMCC encoding and blind detection combination is both effective at reducing ICI and can achieve bit error rates better than that of coherent uncoded OFDM and the PRC-based correlative coding methods, while maintining a low implementation complexity.

5.1 OFDM Data Transmission

A block of information to be transmitted across the channel is first encoded into the sequence of symbols $x_k, 1 \leq k \leq N$, which are considered to be in the frequency domain. The inverse DFT is applied to the sequence to bring it into the time-domain for transmission:

$$x_n^t = \sum_{k=1}^{N} x_k e^{j2\pi nk/N}.$$

The sequence is then cyclically extended (to account for the delay spread of the channel) and transmitted. The received time-domain sequence is given by

$$y_n^t = \sum_{l=1}^{L} h_{l,n}^t x_{n-l+1}^t + w_n^t \tag{5.1}$$

where $h_{l,n}^t$ is the time-varying multiplicative channel gain of the lth path at time n, w_n^t is additive white Gaussian noise with power $E|w_n^t|^2 = \sigma_n^2$, and L is the length of the channel impulse response. We assume that the time-domain channel is Rayleigh fading and that discrete channel paths are uncorrelated, i.e. $E[h_{l_1,n_1}^t h_{l_2,n_2}^{t*}] = 0$ when $l_1 \neq l_2$. We also assume that all paths have the same maximum Doppler frequency f_d, thus the time correlation of the lth channel gain is

$$E[h_{l,n_1}^t h_{l,n_2}^{t*}] = \sigma_l^2 r_h(n_1 - n_2) \tag{5.2}$$

where according to the Jakes' channel model [38] $r_h(n_1 - n_2) = J_0(2\pi f_d T_s(n_1 - n_2))$ and $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. The power in the lth channel

gain is $\sigma_l^2, 1 \leq l \leq L$, and the total power in the channel is normalized:

$$\sum_{l=1}^{L} \sigma_l^2 = 1. \tag{5.3}$$

At the receiver, the cyclic prefix is removed and then the DFT is applied to bring the received sequence back into the frequency domain.

$$y_k = \sum_{n=1}^{N} y_n^t e^{-j2\pi kn/N}.$$

After cyclic-prefix removal the length-N complex-baseband received sequence is stacked into the vector $\mathbf{y}$ and is given by

$$\mathbf{y} = \mathbf{F}\mathbf{H}_t\mathbf{F}^H\mathbf{x} + \mathbf{w} \tag{5.4}$$

where $\mathbf{x} = [x_1, \cdots, x_N]^T$ contains the frequency-domain transmitted symbols with average transmitted power $E|x_n|^2 = 1$, $\mathbf{F}$ is the discrete Fourier transform (DFT) matrix with $\{m,n\}$th entry $f_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi(m-1)(n-1)/N}$, $\mathbf{w} = [w_1, \cdots, w_N]^T$ is the DFT of the noise sequence w_n^t with covariance $E[\mathbf{w}\mathbf{w}^H] = \sigma_n^2\mathbf{I}$.

The structure of $\mathbf{H}_t$ takes into account the FIR filtering effect of the multipath channel as well as the cyclic prefix addition and removal. If the cyclic prefix is of sufficient length and the channel is time-invariant (i.e. no Doppler spread, no CFO, etc.) then $\mathbf{H}_t$ is a circulant matrix and can be diagonalized by the DFT matrix, and the effective frequency-domain channel, defined by

$$\mathbf{H} \equiv \mathbf{F}\mathbf{H}_t\mathbf{F}^H \tag{5.5}$$

is diagonal and there is no ICI. However when the channel is time-varying, $\mathbf{H}_t$ cannot be diagonalized and ICI is manifested as non-zero off-diagonal entries of $\mathbf{H}$.

Each element of $\mathbf{H}$ is given by

$$h_{k,n} = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{L} h_{l,i}^t e^{-j2\pi(i(k-n)+l(n-1)-k+1)/N} \tag{5.6}$$

and the correlation between any two elements is

$$\begin{aligned} E\left[h_{k_1,n_1} h_{k_2,n_2}^*\right] &= \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{l_1=1}^{L} \sum_{l_2=1}^{L} E\left[h_{l_1,i_1}^t h_{l_2,i_2}^{t*}\right] \\ &\quad \times e^{-j2\pi(i_1(k_1-n_1)+l_1(n_1-1)-k_1-i_2(k_2-n_2)-l_2(n_2-1)+k_2)/N} \qquad (5.7) \\ &= \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{l=1}^{L} \sigma_l^2 r_h(i_1 - i_2) \\ &\quad \times e^{-j2\pi(i_1(k_1-n_1)+l(n_1-n_2)-k_1-i_2(k_2-n_2)+k_2)/N} \qquad (5.8) \end{aligned}$$

where (5.8) follows from the uncorrelated nature of the different channel taps.

5.2 ICI and Signal Design Criteria

In this section we present the ICI power in terms of the autocorrelation function of the transmitted sequence and develop conditions that must be satisfied in order to reduce the ICI power.

5.2.1 ICI Analysis

Consider the intercarrier interference power onto the kth OFDM tone. The received data at the kth tone y_k is given by

$$y_k = \sum_{n=1}^{N} h_{k,n}x_n + w_k \tag{5.9}$$

$$= h_{k,k}x_k + \sum_{\substack{n=1 \\ n\neq k}}^{N} h_{k,n}x_n + w_k \tag{5.10}$$

where $h_{k,n}$ is the $\{k,n\}$th element of the frequency-domain channel matrix $\mathbf{H}$, i.e. the effect of the data transmitted on the nth subcarrier onto the kth subcarrier at the receiver. The first term of (5.10) is the signal component and the second term is the ICI component.

We first find the power in the signal component:

$$P_{SIGNAL} = E\left|h_{k,k}x_k\right|^2 \tag{5.11}$$

$$= \frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} r_h(i_1-i_2)\sum_{l=1}^{L}\sigma_l^2 \tag{5.12}$$

$$= \frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} r_h(i_1-i_2). \tag{5.13}$$

The ICI Power onto the kth tone is then given by

$$P_{ICI} = E\left|\sum_{\substack{n=1 \\ n\neq k}}^{N} h_{k,n}x_n\right|^2 \tag{5.14}$$

$$= \sum_{\substack{n=1 \\ n\neq k}}^{N} E|h_{k,n}|^2 + \sum_{\substack{n=1 \\ n\neq k}}^{N}\sum_{\substack{m=1 \\ m\neq k \\ m\neq n}}^{N} E\left[h_{k,n}h_{k,m}^*\right]E\left[x_n x_m^*\right] \tag{5.15}$$

$$\equiv P_{ICI\_OFDM} + P_{ICI\_MOD}. \tag{5.16}$$

The first term in (5.15) is the ICI due to uncoded OFDM and is given by

$$
\begin{aligned}
P_{ICI\_OFDM} &= \sum_{\substack{n=1 \\ n\neq k}}^{N} E|h_{k,n}|^2 && (5.17)\\
&= \sum_{\substack{n=1 \\ n\neq k}}^{N} \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} r_h(i_1-i_2) e^{-j2\pi((i_1-i_2)(k-n))/N} && (5.18)\\
&= \frac{N-1}{N} - \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{\substack{i_2=1 \\ i_2\neq i_1}}^{N} r_h(i_1-i_2) && (5.19)\\
&= 1 - \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} r_h(i_1-i_2) && (5.20)\\
&= 1 - P_{SIGNAL} && (5.21)
\end{aligned}
$$

where (5.19) follows from taking the sum over n which results in

$$
\sum_{\substack{n=1 \\ n\neq k}}^{N} e^{\frac{j2\pi}{N}(i_1-i_2)n} = \begin{cases} -e^{\frac{j2\pi}{N}(i_1-i_2)k}, & i_1 \neq i_2 \\ N-1, & i_1 = i_2 \end{cases} \tag{5.22}
$$

and (5.20) is evident because $r_h(0) = 1$.

The second term in (5.15) is a modification to the total ICI power and is directly impacted by the autocorrelation of the transmitted sequence x_k. To determine the dependence of total ICI power we analyze this term. We first note that

$$
E\left[h_{k,n}h_{k,m}^*\right] = \frac{1}{N^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} r_h(i_1-i_2) e^{\frac{-j2\pi}{N}(i_1(k-n)-i_2(k-m)+n-m)} Z_{n-m} \tag{5.23}
$$

where Z_τ, defined as

$$
Z_\tau \equiv \sum_{l=1}^{L} \sigma_l^2 e^{\frac{-j2\pi}{N}(l-1)\tau} \tag{5.24}
$$

contains the whole effect of the frequency selectivity of the channel. For example for a frequency-flat channel, we have $L = 1$ and $Z_\tau = 1$.

Developing the second term in (5.15), we have

$$P_{ICI\_MOD} = \sum_{\substack{n=1 \\ n\neq k}}^{N} \sum_{\substack{m=1 \\ m\neq k \\ m\neq n}}^{N} E\left[h_{k,n}h_{k,m}^{*}\right] E\left[x_n x_m^{*}\right] \tag{5.25}$$

$$= \frac{1}{N^2} \sum_{\tau=1}^{N-1} \sum_{\substack{n=\tau+1 \\ n\neq k \\ n\neq k+\tau}}^{N} r_x(\tau) \times \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} r_h(i_1-i_2) e^{\frac{-j2\pi}{N}(k(i_1-i_2)-n(i_1-i_2)-i_2\tau+\tau)} Z_\tau. \tag{5.26}$$

where we have substituted (5.23) into (5.25), we have made the change of variables $\tau = n - m$, and $r_x(\tau) = E[x_n x_{n-\tau}^{*}]$ is the autocorrelation of the data sequence.

For simplicity, we assume that the tone of interest k is not near the edge of the OFDM symbol, which has the side effect of removing the dependence on k. This is accurate if k is farther than about 3 symbols away from the edge. To achieve this we make the assumption that the data sequence is circularly correlated, that is $E[x_1 x_N^{*}] = r_x(1)$, $E[x_2 x_N^{*}] = r_x(2)$, etc. The result of this assumption is that the sum over n is extended to begin at 1 instead of $\tau + 1$.

Performing the sum over n in (5.26), we have

$$\sum_{\substack{n=1 \\ n\neq k \\ n\neq k+\tau}}^{N} e^{\frac{j2\pi}{N}(n(i_1-i_2))} = \begin{cases} -e^{\frac{j2\pi}{N}(i_1-i_2)k} - e^{\frac{j2\pi}{N}(i_1-i_2)(k+\tau)}, & i_1 \neq i_2 \\ N-2, & i_1 = i_2 \end{cases} \tag{5.27}$$

and then substituting (5.27) back into (5.26) results in

$$P_{ICI\_MOD} = \frac{-1}{N^2} \sum_{\tau=1}^{N-1} r_x(\tau) \sum_{i_1=1}^{N} \sum_{\substack{i_2=1 \\ i_2\neq i_1}}^{N} r_h(i_1-i_2) \left(e^{j\frac{2\pi}{N}(i_1-1)\tau} + e^{j\frac{2\pi}{N}(i_2-1)\tau}\right) Z_\tau$$

$$= \sum_{\tau=1}^{N-1} r_x(\tau) G(\tau) \tag{5.28}$$

where

$$G(\tau) \equiv -\frac{2}{N^2} \mathrm{Re}\left\{ Z_\tau \sum_{i_1=1}^{N} \sum_{\substack{i_2=1 \\ i_2\neq i_1}}^{N} r_h(i_1-i_2) \left(e^{j\frac{2\pi}{N}(i_1-1)\tau} + e^{j\frac{2\pi}{N}(i_2-1)\tau}\right) \right\} \tag{5.29}$$

is the explicit dependence of the ICI power on each delay of the data autocorrelation. The effect of the frequency selectivity of the channel is wholly contained in the term Z_τ which,

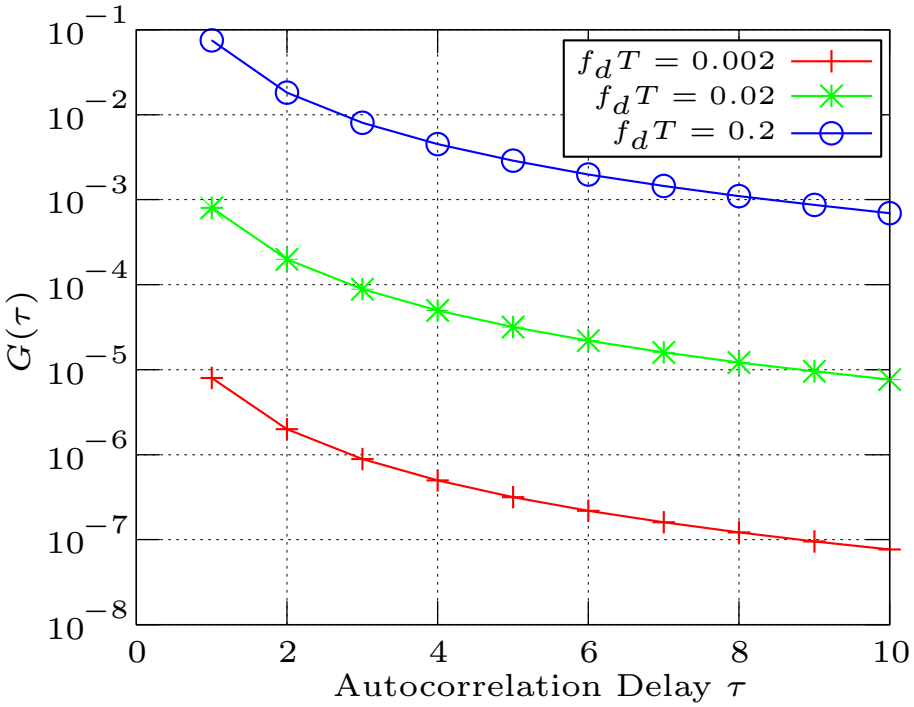

Figure 5.1: The first 10 values of $G(\tau)$ for a flat Rayleigh fading channel with $N = 128$ subcarriers and values of normalized Doppler f_dT equal to 0.2, 0.02, and 0.002.

for $L << N$ (as is typically the case in OFDM systems), we have the approximation that $Z_\tau \approx 1$, and so the frequency selectivity of the channel has only a very small effect on the ICI power and on the data autocorrelation's role in reducing it. Thus, we reach a conclusion also reached in [2], that correlative coding schemes designed to reduce ICI will work effectively in both flat and frequency-selective fading.

5.2.2 Signal Design Criteria

The total ICI power on the subcarrier of interest k is

$$P_{ICI} = P_{ICI\_OFDM} + P_{ICI\_MOD} \tag{5.30}$$

$$= P_{ICI\_OFDM} + \sum_{\tau=1}^{N-1} r_x(\tau)G(\tau). \tag{5.31}$$

P_{ICI_OFDM} is the ICI power of ordinary OFDM without correlative coding. The second term depends upon the autocorrelation of the transmitted sequence $\mathbf{x}$. It is through this term that we are able to modify the total ICI power by tweaking the autocorrelation via correlative coding, driving $\sum_{\tau=1}^{N-1} r_x(\tau)G(\tau)$ as negative as possible.

In Fig. 5.1 we plot the values of $G(\tau)$ for $1 \leq \tau \leq 10$ for a flat Rayleigh fading channel with $N = 128$ subcarriers and values of normalized Doppler f_dT equal to 0.2, 0.02, and 0.002. We make the following observations:

1. $G(\tau)$ decays rapidly as τ increases, and thus the sum $\sum_{\tau=1}^{N-1} r_x(\tau)G(\tau)$ is dominated by only the first few terms.

2. $G(\tau)$ is positive for the first few (dominant) terms.

3. $G(\tau)$ decays similarly for differing values of Doppler.

Taken together, these observations indicate that from a signal design perspective, we need only worry about the autocorrelation $r_x(\tau)$ for a small number of values. Additionally, we should design the sequence x_n such that $r_x(\tau)$ is large and negative for the first few values of τ. The third point above indicates that a signal design scheme will work similarly across a wide range of Doppler conditions.

5.3 Constant Magnitude Correlative Coding

In the previous section we determined the effect of the data autocorrelation on the ICI power and showed how the autocorrelation of the transmitted sequence can drive the ICI lower than that of ordinary OFDM.

In this section we introduce the concept of constant magnitude correlative coding (CMCC). We then determine the optimal degree of correlation to place on the transmitted symbols in order to balance the dual requirements of ICI reduction and detection capability in noise in terms of bit error probability.

From the analysis in the previous section we know that the data autocorrelation $r_x(\tau)$ should be designed to be large and negative for the first few values of τ. In order to achieve this while still maintaining a constant magnitude coded sequence, we proposing the following symbol encoding scheme:

Let the $d_k = \pm 1$ be the sequence of information bits slated for transmission via OFDM. These bits are encoded into the transmitted sequence x_k as follows:

$$x_k = x_{k-1} e^{j\theta d_k}. \tag{5.32}$$

Fig. 5.2 illustrates the encoding process. To transmit $+1$, a phase of θ is added to the previously transmitted symbol, and to transmit -1, a phase of $-\theta$ is added. Similarly

to differential encoding, a reference symbol must be transmitted at the beginning of the sequence, e.g. by setting $x_1 = 1$. In Fig. 5.3 the example information bit sequence $\{+1, +1, -1, -1, -1\}$ is encoded.

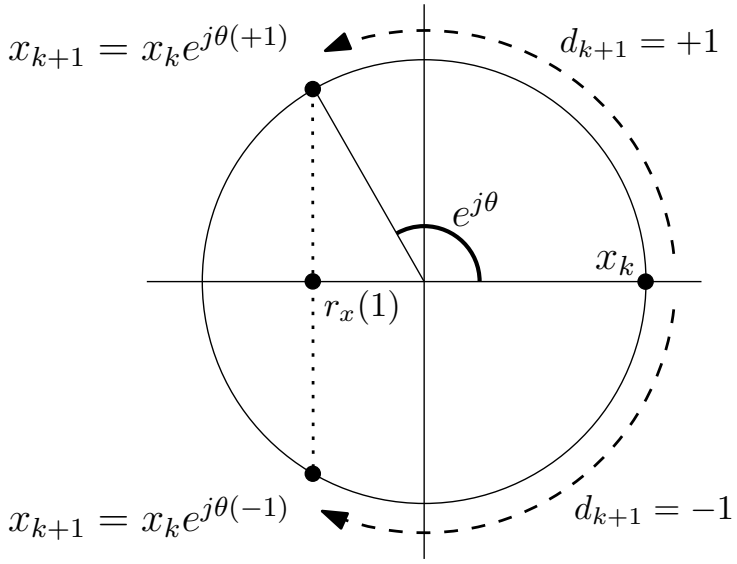

Figure 5.2: Diagram of the proposed constant-magnitude correlative encoding process.

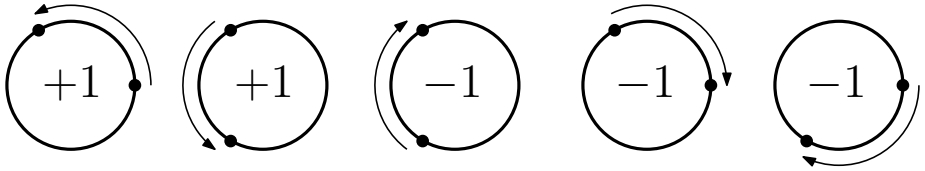

Figure 5.3: A toy example of the proposed correlative data encoding scheme: transmission of the data sequence $\{+1, +1, -1, -1, -1\}$.

The autocorrelation induced upon the sequence x_k by the correlative encoding

scheme in (5.32) is

$$\begin{aligned} r_x(\tau) &= E[x_k x_{k-\tau}^*] && (5.33)\\ &= E\left[x_{k-\tau}x_{k-\tau}^* \prod_{i=k-\tau+1}^{k} e^{j\theta d_i}\right] && (5.34)\\ &= \prod_{i=k-\tau+1}^{k} E\left[e^{j\theta d_i}\right] && (5.35)\\ &= \prod_{i=k-\tau+1}^{k} \frac{1}{2}\left(e^{j\theta} + e^{-j\theta}\right) && (5.36)\\ &= \cos^{\tau}(\theta) && (5.37) \end{aligned}$$

where (5.35) follows from the independence of the information bits and (5.36) follows from the fact that $d_i = \pm 1$ with equal probability.

In choosing θ we have direct control over the autocorrelation of the transmitted sequence. It is clear that for $r_x(1)$ to be negative, we want $\frac{\pi}{2} \leq \theta \leq \pi$. With such a choice of θ, $r_x(2)$, and all values of $r_x(\tau)$ for even τ will be positive, however $G(\tau)$ decreases rapidly enough that the ICI power is dominated by $G(1)$ and $r_x(1)$.

This coding scheme is similar to both DPSK (in which the additive phase is either 0 for $d_i = +1$, or π for $d_i = -1$), and symmetric DPSK [54] [55] (which is equivalent to CMCC with $\theta = \pi/2$).

Clearly, with the CMCC scheme there is a trade-off between ICI reduction and detection performance, as increasing θ to force $r_x(1)$ more negative results in the two possible values of the next symbol being closer together in Euclidean distance (in Fig. 5.2, the dashed line between potential values of x_{k+1} becomes shorter), degrading performance when additive noise is present and the detector must differentiate between the two. We address this trade-off and determine optimal values of θ in section 5.5.

5.4 Comparisons to Other Correlative Coding Schemes

In this section we compare the ICI reduction capability of the proposed constant-magnitude correlative coding (CMCC) scheme to that of other correlative coding schemes including Partial Response Coding (PRC) [2], the ICI self-cancellation technique [51], and to uncoded OFDM.

The values of the autocorrelation of the data sequence $r_x(\tau)$ can be used in (5.16), (5.20), and (5.28) to calculate the total ICI of any of the mentioned schemes. Since the

Table 5.1: Data Autocorrelation of different Correlative Coding Schemes

	$r_x(1)$	$r_x(2)$	$r_x(3)$	$r_x(4)$
OFDM	0	0	0	0
CMCC $\theta = \frac{\pi}{2}$	0	0	0	0
CMCC $\theta = \frac{2\pi}{3}$	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{8}$	$\frac{1}{16}\cdots$
CMCC $\theta = 2.35$	$-.704$	.495	$-.349$	$.246\cdots$
CMCC $\theta = 2.511$	$-.808$	.652	$-.527$	$.425\cdots$
CMCC $\theta = \pi$	-1	1	-1	$1\cdots$
PRC-2	$-\frac{1}{2}$	0	0	0
PRC-3	$-.704$	.228	0	0
PRC-4	$-.808$	.430	$-.123$	0
Self-Canc.	$-\frac{1}{2}$	0	0	0

ICI powers of these schemes only differ as a result of differing data autocorrelations, we compare the first four values of $r_x(\tau)$ for various schemes in table 5.1.

When θ is set to $\pi/2$, the CMCC scheme results in zero data correlation, thus there is no ICI reduction. In general, a parallel can be seen between the CMCC scheme and the PRC scheme, in that increasing the value of θ for CMCC is analogous to increasing the filter-length of PRC: $r_x(1)$ becomes more negative, and we expect improved ICI reduction. For CMCC with values of θ resulting in the same $r_x(1)$ as the corresponding PRC scheme (for example CMCC with $\theta = 2\pi/3$ and PRC-2), the values of $r_x(2)$ are comparatively higher, so we expect slightly better ICI reduction from the PRC scheme. When θ is set to π, the ICI is theoretically minimized, however no information is actually transmitted.

The ICI self-cancellation scheme has the same data autocorrelation and thus ICI reduction performance as PRC-2, however its throughput is only half of the other schemes.

In Fig. 5.4 we illustrate the reduction in ICI power of CMCC as compared to uncoded OFDM and the PRC-2 and PRC-3 correlative coding schemes. ICI self-cancellation is omitted with the understanding that its performance is identical to that of PRC-2. Solid lines are the theoretical ICI according to (5.16), and the points are simulated. In this simulation the ICI power is computed on one tone at the center of the spectrum with $N = 128$ subcarriers. The rest of the data in the OFDM symbol is correlative coded and then transmitted through a frequency-flat Rayleigh fading channel.

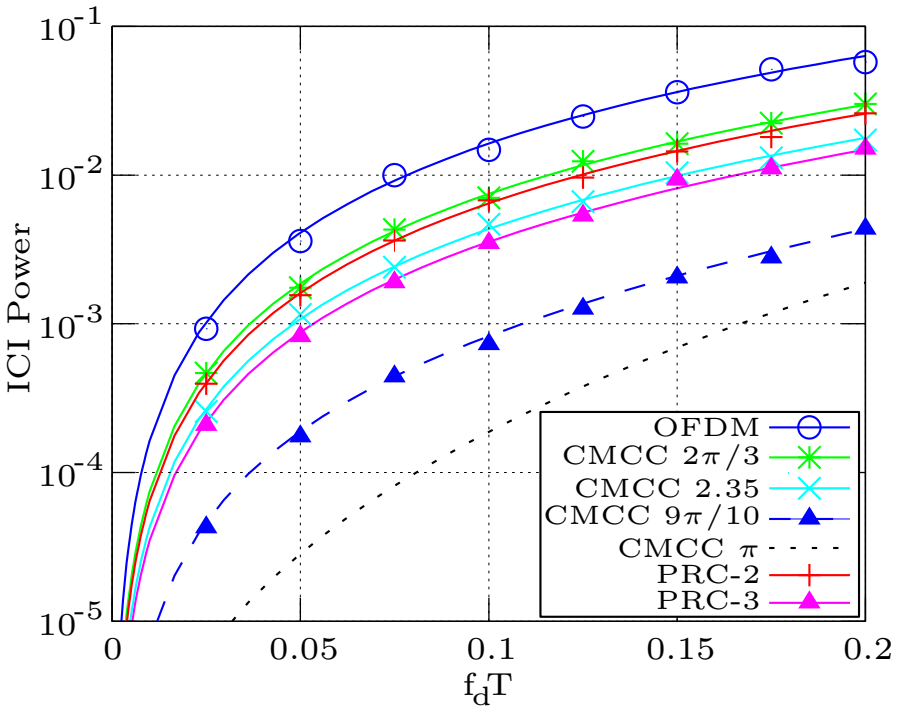

Figure 5.4: ICI Power vs. normalized Doppler f_dT for uncoded OFDM, the proposed CMCC scheme with various values of θ, and partial-response coding (PRC) of orders 2 and 3 [2]. Lines are theoretical ICI according to (5.16), and points are simulated.

The ICI power is measured on the tone of interest and averaged over 1000 trials.

All of the correlative coding schemes can effectively reduce the total ICI power as compared to that of uncoded OFDM. We can also observe the parallel between the CMCC scheme and the PRC scheme: both increasing θ and increasing the PRC filter length push the ICI power lower, to a lower bound defined by CMCC with $\theta = \pi$.

The total ICI power of CMCC with $\theta = 2\pi/3$ is very close to that of PRC-2 (and the ICI self-cancellation scheme). It is only slightly higher due to the fact that $r_x(2) > 0$. For the same reason, the total ICI power of CMCC with $\theta = 2.35$ slightly higher than that of PRC-3.

The self-cancellation scheme achieves its improved ICI performance at the expense of spectral efficiency, cutting in half the the transmitted data per tone. The PRC scheme achieves its performance at the cost of having non-constant magnitude symbols, with the possibility of some of them being nulls, the practical consequences of which are discussed in section 5.6.

5.5 Signal Design

In this section we examine the effect of the modulation angle θ on bit error rate, as it balances the requirements of ICI reduction and symbol detection, and we find the optimal θ for different noise and Doppler regimes.

5.5.1 The Trade-off Between ICI Reduction and Symbol Detection Performance

Consider a detector which assumes that the ICI has been mitigated by correlative coding and what remains is now simply noise (i.e. there is no attempt at interference cancellation). The received signal on the tone of interest k is given by

$$y_k = h_{k,k}x_k + \sum_{\substack{n=1 \\ n\neq k}}^{N} h_{k,n}x_n + w_k. \tag{5.38}$$

The job of the detector is to make a decision on x_k considering that $\sum_{\substack{n=1 \\ n\neq k}}^{N} h_{k,n}x_n$ and w_k are noise corrupting the signal, i.e. it must decide whether $x_k = x_{k-1}e^{j\theta}$ or $x_k = x_{k-1}e^{-j\theta}$ was transmitted.

The magnitude of the distance between the two received points that must be decided between is given by

$$|h_{k,k}x_{k-1}e^{j\theta} - h_{k,k}x_{k-1}e^{-j\theta}| = 2|h_{k,k}|\sin(\theta). \tag{5.39}$$

Thus, in the best case scenario of known channel coefficient $h_{k,k}$ and known symbol phase reference, the probability of the detector committing an error is

$$P_e = P\left(\sum_{\substack{n=1 \\ n\neq k}}^{N} h_{k,n}x_n + w_k > |h_{k,k}|\sin(\theta)\right). \tag{5.40}$$

In the Rayleigh fading channel, the signal component $|h_{k,k}|$ is a Rayleigh fading random variable with variance

$$P_{SIGNAL} = E\left[|h_{k,k}|^2\right] = \frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} r_h(i_1 - i_2). \tag{5.41}$$

If we make the simplifying assumption that the ICI term is complex Gaussian noise and is independent of the signal component, then the Rayleigh fading probability of error

is [57]

$$P_e = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma(cos(\theta))}{1+\gamma(cos(\theta))}}\right) \tag{5.42}$$

where

$$\gamma(cos(\theta)) = P_{SIGNAL}\frac{1-cos^2(\theta)}{P_{ICI\_OFDM} + \sigma_n^2 + \sum_{\tau=1}^{N-1} cos^{\tau}(\theta)G(\tau)} \tag{5.43}$$

is the signal to noise and interference ratio as a function of the modulation angle.

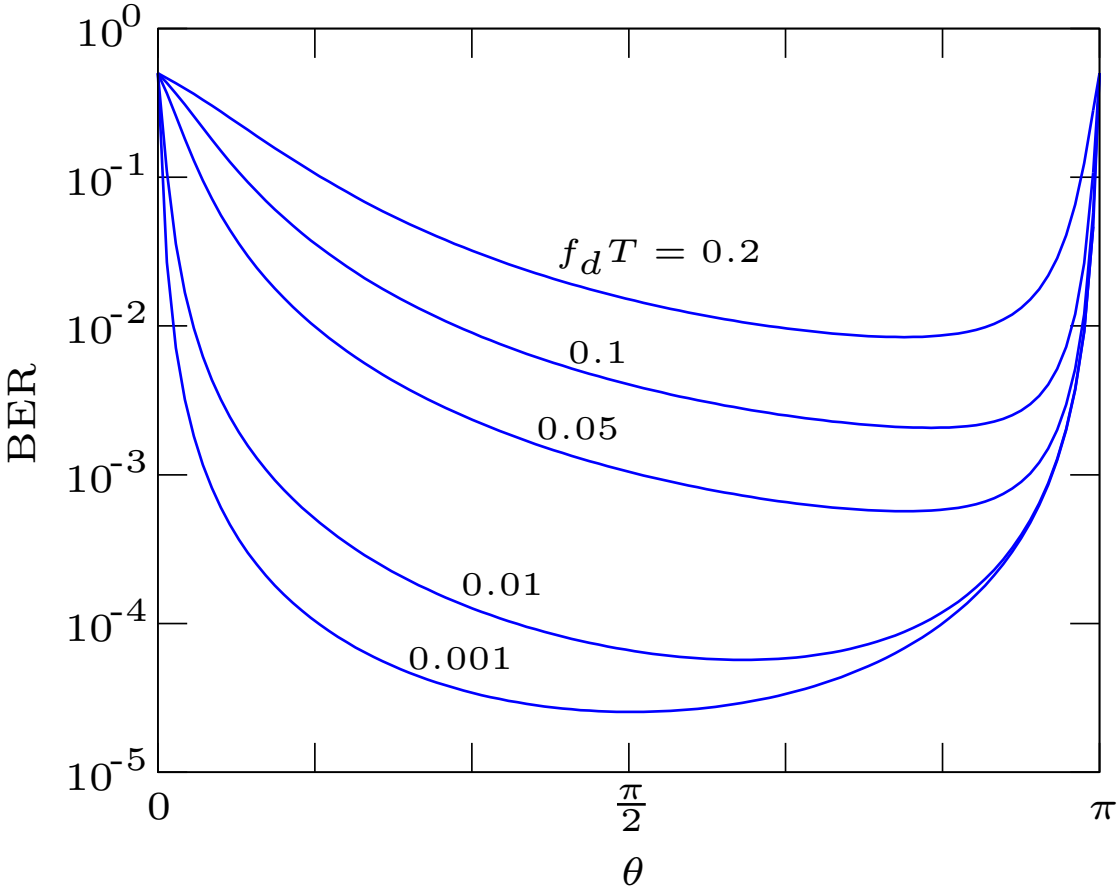

Figure 5.5: Bit error rate in Rayleigh fading of the CMCC scheme as a function of modulation angle θ for $SNR = 40dB$ and varying degrees of normalized Doppler f_dT.

From section 5.4 we know that a higher modulation angle (close to π) gives a larger negative symbol correlation and thus the best ICI reduction performance. Conversely, detection in noise is improved when the two possible symbol points are maximally separated, when the modulation angle is low (close to $\pi/2$). To illustrate the modulation angle's role in the trade-off between ICI reduction and symbol detection in noise, in Fig.

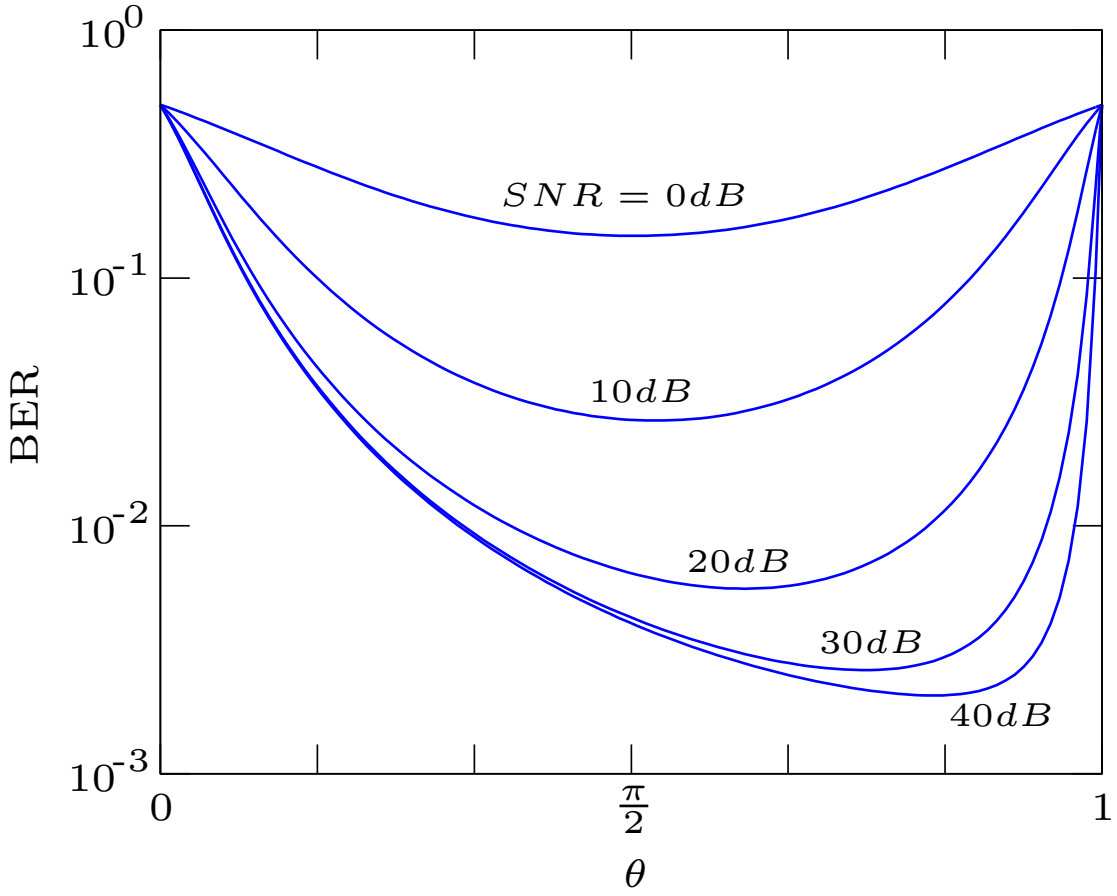

Figure 5.6: Bit error rate in Rayleigh fading of the CMCC scheme as a function of modulation angle θ for normalized Doppler $f_dT = 0.1$ and varying degrees of SNR.

5.5 we plot the bit error rate of (5.42) versus modulation angle θ for $SNR = 40dB$ and varying degrees of Doppler, and in Fig. 5.6 we plot the bit error rate for fixed normalized Doppler $f_dT = 0.1$ and varying degrees of SNR. In both figures the sum $\sum_{\tau=1}^{N-1} cos^{\tau}(\theta)G(\tau)$ was truncated to the first few non-negligible terms, and $N = 128$ subcarriers were used in the calculation of $G(\tau)$. We first note that in both figures, the angle θ that minimizes the BER is always between $\pi/2$ and π, indicating that some degree of ICI reduction is always beneficial. In Fig. 5.5 we see that for a fixed noise power, higher Doppler requires a higher modulation angle, approaching π, since in this regime the ICI dominates the BER and must be mitigated. In Fig.5.6 we see that for fixed Doppler, higher noise power requires a lower modulation angle, approaching $\pi/2$, since the noise dominates and symbol point separation is most important.

5.5.2 Finding the Optimal Modulation Angle

Minimizing the probability of the detector making an error is equivalent to maximizing $\gamma(cos(\theta))$. For tractability, we minimize its reciprocal:

$$\hat{\theta} = \min_{\theta} \frac{1}{\gamma(cos(\theta))}, \quad \frac{\pi}{2} \leq \theta \leq \pi. \tag{5.44}$$

We next make the change of variables $\alpha = cos(\theta)$ which results in the equivalent optimization problem

$$\hat{\alpha} = \min_{\alpha} \frac{1}{\gamma(\alpha)}, \quad -1 \leq \alpha \leq 0. \tag{5.45}$$

Next, taking the derivative with respect to α and setting to zero yields

$$\frac{\partial}{\partial \alpha} \frac{1}{\gamma(\alpha)} =$$

$$2(P_{ICI\_OFDM} + \sigma_n^2)\alpha + \sum_{\tau=1}^{N-1} G(\tau)\left((2-\tau)\alpha^{\tau+1} + \tau\alpha^{\tau-1}\right) = 0, \quad -1 \leq \alpha \leq 0 \tag{5.46}$$

which consists of a polynomial of degree $N+1$. The sum in (5.46) can be truncated to only the first few terms since $G(\tau)$ decays exponentially. After truncation, the solution to (5.46) can be found numerically by finding the roots of the reduced-degree polynomial and selecting the root within the range $-1 \leq \alpha \leq 0$. Finally, the optimal modulation angle θ can be found as

$$\hat{\theta} = \arccos(\hat{\alpha}) \tag{5.47}$$

and selecting the solution in the top half of the complex plane.

In Fig. 5.7 we plot bit error rate vs. normalized Doppler f_dT for SNRs of $20dB$ and $40dB$ vor various values of modulation angle θ. The channel is flat Rayleigh fading with $N = 128$ subcarriers. When θ is set to the optimal $\hat{\theta}$ from (5.47) for each specific SNR and Doppler, the performance serves as a lower bound on BER. When there is no ICI reduction, i.e. $\theta = \pi/2$, performance suffers at high SNR and at high Doppler due to the effects of the ICI. When θ is set to favor a high amount of ICI reduction, i.e. $\theta = 3\pi/4$, performance is excellent at high SNR, but sub-optimal at low SNR where the noise dominates the performance. We also plot the performance for $\theta = 2\pi/3$ which offers a middle ground between performance in noise and performance in ICI. Using this value of θ is perhaps a good trade-off between ICI reduction and symbol detection when the system operating point is a priori unknown.

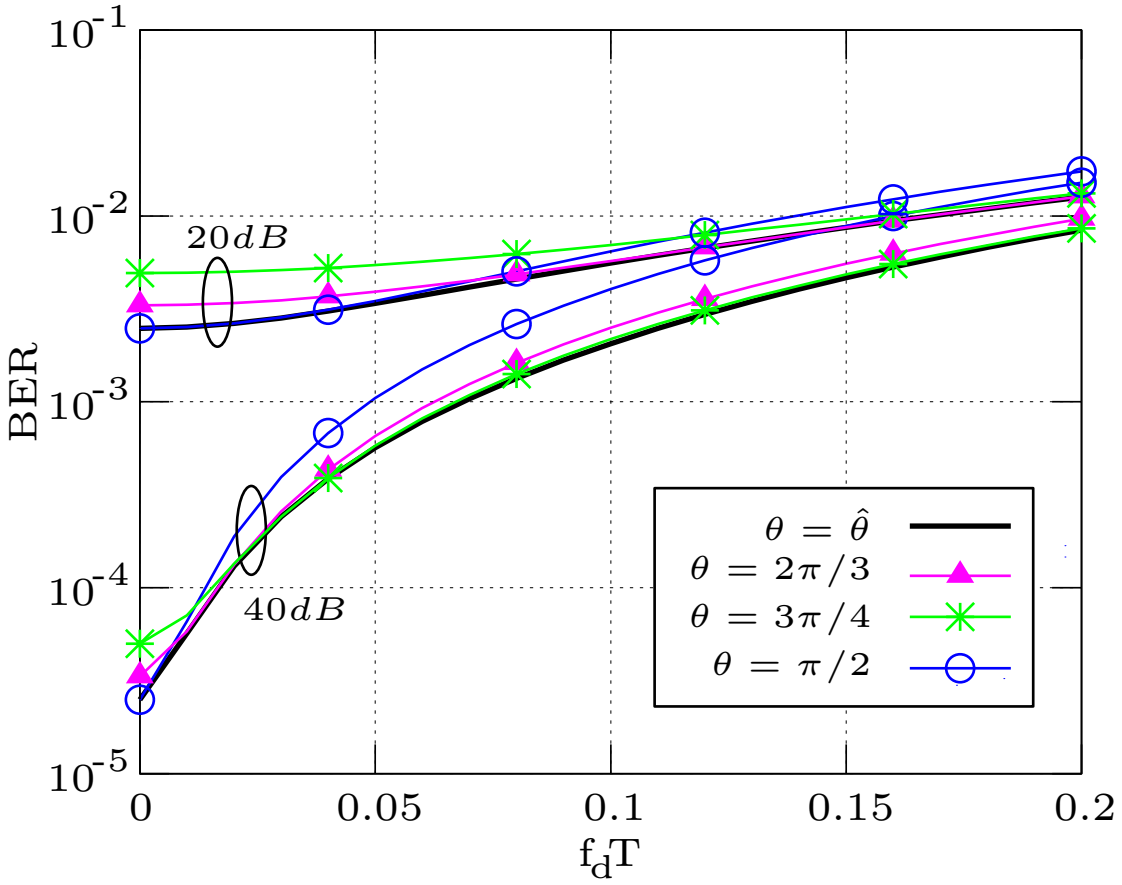

Figure 5.7: Bit error rate vs Doppler in Rayleigh fading for various values of modulation angle θ and SNRs of $20dB$ and $40dB$.

5.6 Considerations for Blind Detection: OFDMA

A class of detectors particularly applicable to OFDMA reception are blind sequence detectors that jointly decode all the symbols in a particular user's contiguous block of subcarriers. Blind detection is useful in the mobile environment because the channel, which varies in time and frequency and may be changing rapidly, need not be estimated before symbol detection takes place.

In evaluating the practical applicability of the constant-magnitude correlative coding ICI reduction scheme, we consider two classes of blind sequence detector: multiple symbol differential detection (MSDD) [26], which requires channel and noise statistics, and a variant on MSDD which does not require channel statistics and instead relies on a deterministic parameterization of the channel [58, 59]. Both classes of detector may be implemented via a reduced-complexity algorithm based on sphere decoding [1, 58, 59]

or other tree-search techniques [60]–provided that the transmitted symbols are constant magnitude.

In an OFDMA system, for any block of N contiguous subcarriers (which can include the whole OFDM symbol), the length N received signal vector $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{5.48}$$

where $\mathbf{x}$ is the vector of transmitted data applied to those subcarriers, $\mathbf{w}$ is complex Gaussian noise, and $\mathbf{H}$ is the associated frequency-domain channel matrix for those subcarriers, and is a sub-matrix on the diagonal of the whole OFDM symbol's frequency-domain channel matrix.

We assume that the data in the vector $\mathbf{x}$ has been correlatively coded, and thus the off-diagonal elements of $\mathbf{H}$, caused by ICI due to time-variation, can be ignored. Thus, we rewrite (5.48) as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w} + ICI \tag{5.49}$$

where $\mathbf{h}$ is a vector containing the diagonal elements of $\mathbf{H}$, $\mathbf{A}$ is a diagonal matrix containing the elements of $\mathbf{x}$ on the diagonal, and ICI is a noise term resulting from ignoring the diagonal of $\mathbf{H}$. Since the detector will consider the remaining ICI that has not been eliminated by correlative coding as noise, we consider the ICI an additive Gaussian noise term which is independent of both the channel and the data.

5.6.1 Known Statistics Blind Detection

When the channel and noise statistics are known to the receiver, the maximum likelihood sequence detector chooses the sequence $\mathbf{x}$ that maximizes the probability of the received data conditioned on $\mathbf{x}$ [26]:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) \tag{5.50}$$

$$= \arg\max_{\mathbf{x}} \frac{1}{|\mathbf{R}_y|} \exp\left\{-\mathbf{y}^H \mathbf{R}_y^{-1} \mathbf{y}\right\} \tag{5.51}$$

where $\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H]$ is the correlation matrix of the received data $\mathbf{y}$ equal to

$$\mathbf{R}_y = \mathbf{A}\mathbf{R}_h\mathbf{A}^H + \left(P_{ICI} + \sigma_n^2\right)\mathbf{I}, \tag{5.52}$$

$\mathbf{R}_h = E[\mathbf{h}\mathbf{h}^H]$ is the correlation matrix of the diagonal of the frequency-domain channel $\mathbf{h}$ whose $\{k_1, k_2\}$th entry is equal to

$$\{\mathbf{R}_h\}_{k_1,k_2} = \frac{1}{N^2}\left(\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} r_h(i_1 - i_2)\right)\left(\sum_{l=1}^{L}\sigma_l^2 e^{-j2\pi(k_1-k_2)(l-1)/N}\right), \tag{5.53}$$

and P_{ICI} is the ICI power after mitigation by correlative coding, defined in (xx).

If the symbols in $\mathbf{x}$ are constant in magnitude, then the determinant of $\mathbf{R}_y$ is independent of the data, and performing the maximization in (5.51) is equivalent to performing the minimization

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{y}^H \mathbf{A} \left(\mathbf{R}_h + (P_{ICI} + \sigma_n^2)\mathbf{I} \right)^{-1} \mathbf{A}^H \mathbf{y} \tag{5.54}$$

which can be efficiently implemented via sphere decoding [1].

The advantage of an ICI-mitigating correlative coding scheme that retains constant magnitude symbols, such as the proposed CMCC scheme, is that it can be used directly with the efficiently implementable detector of (5.54). The PRC correlative coding schemes, on the other hand, result in varying magnitude transmitted symbols so the detector of (5.54) cannot be used. In this case a brute force search of (5.51) must be completed at full computational complexity.

5.6.2 Unknown Statistics Blind Detection

When channel and noise statistics are not known to the receiver, we consider the channel to be deterministic and unknown. If the data has been correlatively coded, then again we can assume that $\mathbf{H}$ is a diagonal matrix and ignore the off-diagonal ICI contributing terms, whose effects have been mitigated by the correlative coding. Thus for joint data detection and channel estimation we must minimize

$$\|\mathbf{y} - \mathbf{A}\mathbf{h}\|^2 \tag{5.55}$$

over both the channel diagonal elements $\mathbf{h}$ and the data vector $\mathbf{x}$ on the diagonal of $\mathbf{A}$.

Following [59], we parameterize the channel vector $\mathbf{h}$ (which varies due to the frequency-selectivity of the channel) with a set of I orthonormal basis vectors $\mathbf{f}_i$ as

$$\mathbf{h} \approx \sum_{i=1}^{I} \alpha_i \mathbf{f}_i = \mathbf{F}\boldsymbol{\alpha} \tag{5.56}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_I]^T$ are the coefficients of the parameterization. The vectors $\mathbf{f}_i$ are chosen to be orthonormalized Chebyshev polynomials. Next, substituting $\mathbf{h} = \mathbf{F}\boldsymbol{\alpha}$ into (5.55) results in the optimization problem

$$\{\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}\} = \arg\min_{\mathbf{x},\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{A}\mathbf{F}\boldsymbol{\alpha}\|^2. \tag{5.57}$$

The value of $\boldsymbol{\alpha}$ that minimizes (5.57) is

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{F}^H\mathbf{A}^H\mathbf{A}\mathbf{F}\right)^{-1}\mathbf{F}^H\mathbf{A}^H\mathbf{y} \tag{5.58}$$

which, when substituted back into (5.57) yields the data detection optimization problem:

$$\hat{\mathbf{x}} \;=\; \arg\max_{\mathbf{x}} \quad \mathbf{y}^H\mathbf{A}\mathbf{F}\left(\mathbf{F}^H\mathbf{A}^H\mathbf{A}\mathbf{F}\right)^{-1}\mathbf{F}^H\mathbf{A}^H\mathbf{y}. \tag{5.59}$$

If the symbols in $\mathbf{x}$ are constant in magnitude, then $\mathbf{A}^H\mathbf{A} = \mathbf{I}$, so that the inverse in (5.59) reduces to

$$\left(\mathbf{F}^H\mathbf{A}^H\mathbf{A}\mathbf{F}\right)^{-1} \;=\; \left(\mathbf{F}^H\mathbf{F}\right)^{-1} \tag{5.60}$$

$$\;=\; \mathbf{I} \tag{5.61}$$

and the optimization problem (5.59) reduces to [59]

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \mathbf{y}^H\mathbf{A}\mathbf{F}\mathbf{F}^H\mathbf{A}^H\mathbf{y}. \tag{5.62}$$

As in the known-statistics detection framework, this optimization problem can be efficiently implemented via sphere decoding or other fast tree-search detection algorithms. The detector of (5.62) and associated fast implementations can be directly applied to the proposed CMCC scheme since the transmitted symbols are constant magnitude.

If the symbols are not constant magnitude, as is the case with the PRC correlative coding schemes, then again efficient algorithms cannot be applied and (5.59) must be evaluated separately for each potential sequence at full computational complexity.

A further disadvantage of PRC correlative coding schemes in this unknown-statistics detection framework stems from the possibility of transmitted null symbols. For example, with PRC-2 the probability of a particular symbol being null is 0.5, and with PRC-4 the probability is 0.25. Intuitively, null symbols are detrimental to blind detection since no information about the channel is provided to the receiver at their locations in the transmitted sequence. If the number of non-null symbols is less than the number of basis vectors, then the inverse in (5.59) is not possible. In practice, this means that PRC schemes with nulls must transmit extra overhead symbols to ensure implementability, at the cost of decreased throughput.

5.6.3 OFDMA Structure

Consider the downlink OFDMA scenario where each user is allocated a block of N contiguous subcarriers. All of the data at the receiver destined for separate users may

be CMCC encoded together as one block spanning all subcarriers. With this encoding structure, the receiver for each user takes as its input a block of $N+1$ subcarriers spanning the user's own subcarrier block and the previous user's last subcarrier, which is used as a reference. Fig. 5.8 illustrates this structure. With such an encoding structure, no symbol overhead is required, i.e. N bits may be transmitted over a block of N subcarriers.

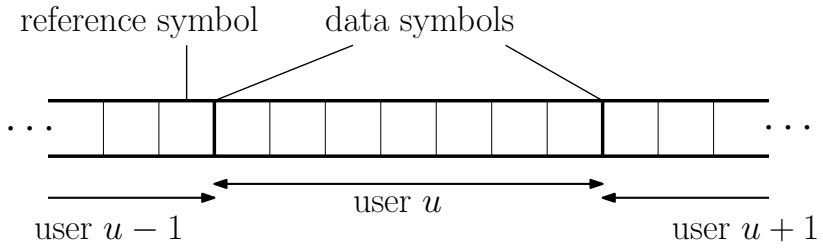

Figure 5.8: OFDMA downlink subcarrier allocation. With the proposed CMCC scheme, user u may use the last symbol of the previous adjacent user's subcarrier block as a reference symbol. No reference overhead is required.

With PRC encoding, on the other hand, the state of the encoder is not wholly contained in the previous coded symbol, so the data bits must be encoded separately in each user's subcarrier block. For fair comparisons with CMCC encoded blind detection, we design the PRC encoded subcarrier block to have the same length and same redundancy as a CMCC encoded block, i.e. we want to encode N bits into $N+1$ transmitted symbols. To achieve this, we include one reference symbol and one reference bit for PRC-2 encoded data, and one reference symbol and two reference bits for PRC-3 encoded data. The reference bits serve to initialize the state of the encoder, and the reference symbol provides phase information, similarly to the CMCC scheme.

5.7 Performance Results

5.7.1 Simulation Parameters

In this section we present simulated bit-error-rate performance of the CMCC scheme in a downlink OFDMA system over a frequency-flat and frequency-selective time-varying channel. The frequency-selective channel is comprised of six channel taps, with a power delay profile of $[0.21802, 0.19431, 0.17318, 0.15434, 0.13756, 0.12260]$ corresponding to the exponential decay power delay profile model, i.e. $\sigma_l^2 = \beta \exp(-\alpha(l-1))$ where $\alpha = 0.5$ and $\beta = 1/\sum_l \sigma_l^2$. The transmitter transmits an OFDM symbol with 128

subcarriers and the entire OFDM symbol is encoded according to the CMCC scheme, as described in section 5.6.3. The receiver of a particular user examines a block of $N = 9$ contiguous subcarriers at the center of the OFDM symbol and the sphere-decoding based blind detector decides upon the length $N - 1$ sequence of bits that was transmitted.

5.7.2 Comparisons to Other Detectors

We will now evaluate two versions of the blind CMCC-based detector: one for known, and one for unknown channel statistics, as presented in sections 5.6.1 and 5.6.2, respectively.

We compare the CMCC-based detector to similar blind detectors based on the PRC-2 and PRC-3 correlative coding schemes [2]. These detectors cannot be made more efficient by use of the sphere-decoding algorithm, and so are full brute-force complexity. To ensure that the comparison is fair, each of the PRC schemes is allowed to encode $N-1$ bits into N symbols which make up the frequency dwell of a particular user. To achieve this, the PRC-2 scheme is allowed to have one reference bit before correlative coding, and one reference transmitted symbol after correlative coding. The PRC-3 scheme is allowed two reference bits and one reference symbol. All of the other transmitted symbols on the subcarriers are encoded according to either PRC-2 or PRC-3.

We also compare to the performance of coherently detected uncoded OFDM. The coherent OFDM receiver does not identify or correct for the ICI. The receiver has perfect knowledge of the diagonal of the frequency-domain channel matrix $\mathbf{H}$, but the off-diagonal elements which contribute to the ICI are unknown, and the ICI is treated as noise.

5.7.3 Performance

In Fig. 5.9 we present the bit error rate performance of the proposed CMCC scheme and the other abovementioned detectors after transmission through a frequency-flat Rayleigh-fading channel with normalized Doppler $f_dT = 0.05$. Since the channel is flat, all of the unknown statistics detectors are built using only the first constant Chebychev polynomial as their deterministic channel approximation. The performance of coherently detected OFDM is seen to have an error floor due to the intercarrier interference. The CMCC scheme, on the other hand, can outperform coherent OFDM above an SNR of about 27 dB because of the robustness to ICI, even though the detector is

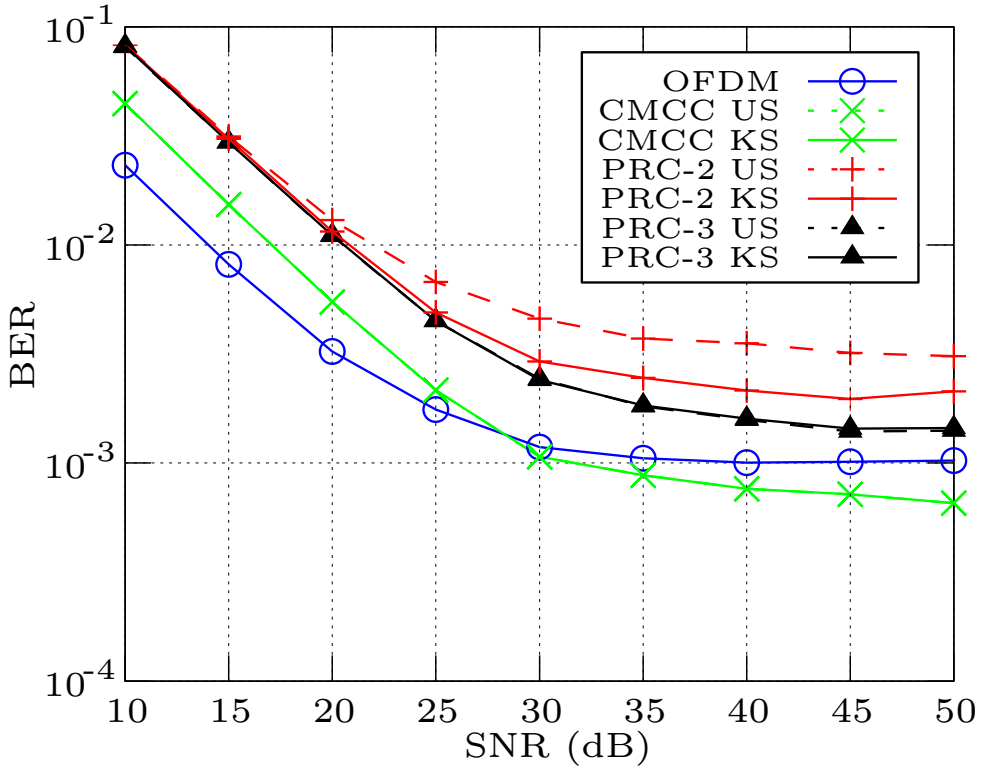

Figure 5.9: BER vs. SNR in an OFDMA system in a frequency-flat Rayleigh-fading channel with $f_dT = 0.05$. The proposed CMCC scheme is compared to PRC-2 and PRC-3, using both the known and unknown-statistics blind detectors, and to coherent OFDM (known diagonal of $\mathbf{H}$).

blind. The known and unknown-statistics CMCC detectors perform the same since they are equivalent in a flat channel. The PRC-3 scheme performs worse than CMCC even though it is better at mitigating ICI because the combination of noise and residual ICI is still large compared to the minimum distance of PRC-3, which is much smaller than CMCC, due to the many symbol amplitude levels inherent in the encoding scheme. Similarly, the PRC-2 scheme performs worse still because its minimum distance is small and the residual ICI is higher. The unknown-statistics version of PRC-2 performs relatively worse because the presence of data nulls means that less information is available for the joint channel estimation/data detection.

In Fig. 5.10 we plot the same curves at the same normalized Doppler of $f_dT = 0.05$ but for a frequency-selective channel instead of a flat channel. To be able to track the channel time-variation, the unknown statistics detectors use 3 Chebychev basis polynomials in their respective implementations, except for the PRC-2 scheme which uses 2

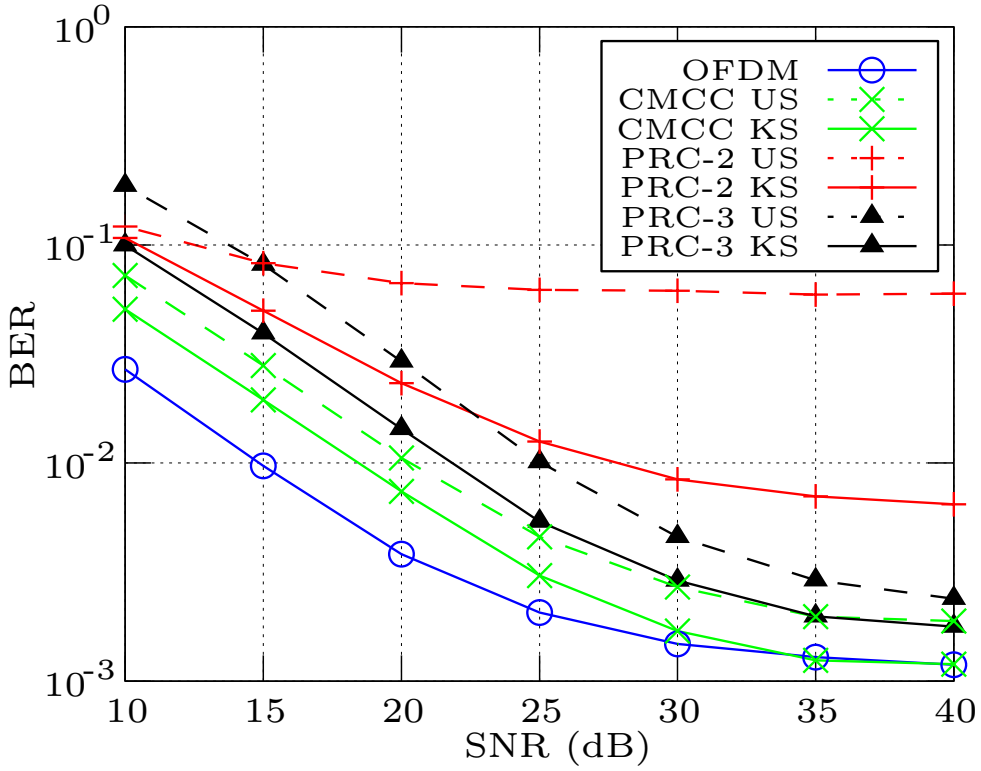

Figure 5.10: BER vs. SNR in an OFDMA system in a frequency-selective Rayleigh-fading channel with $f_dT = 0.05$. The proposed CMCC scheme is compared to PRC-2 and PRC-3, using both the known and unknown-statistics blind detectors, and to coherent OFDM (known diagonal of $\mathbf{H}$).

because it cannot support more due to the data nulls. The coherent OFDM detector outperforms all the other detectors because it does not have the task of determining the frequency-varying channel and data in the presence of ICI. Again, the CMCC scheme outperforms the PRC schemes because of the combination of ICI mitigation and large minimum distance for symbol discernibility. The unknown statistics schemes perform worse than their associated known-statistics counterparts because they are relying on the imperfect Chebychev polynomial channel approximation.

Next, in Fig. 5.11 we reduce the Doppler to $f_dT = 0.01$ and plot performance of each of the detectors in a flat-fading channel. Again, the blind CMCC detector performs very closely to coherent uncoded OFDM and only outperforms it at high SNR, where ICI plays a larger role in performance degradation than the noise, and again, the CMCC scheme outperforms both full-complexity PRC schemes.

Fig. 5.12 shows performance of the detectors in a frequency selective channel with

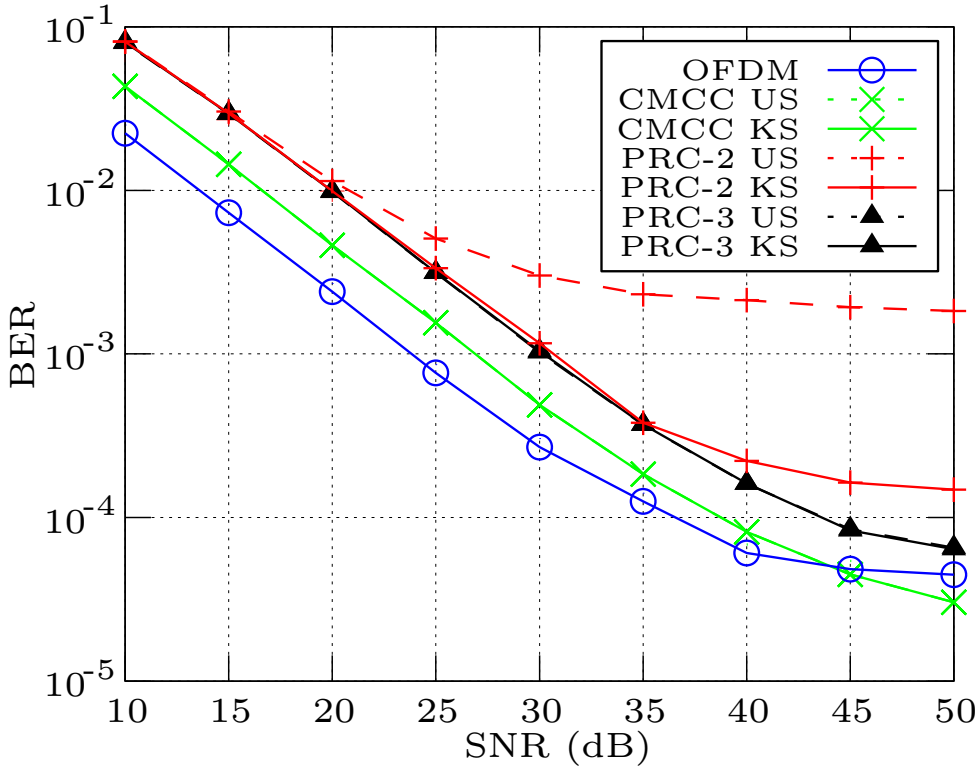

Figure 5.11: BER vs. SNR in an OFDMA system in a frequency-flat Rayleigh-fading channel with $f_dT = 0.01$. The proposed CMCC scheme is compared to PRC-2 and PRC-3, using both the known and unknown-statistics blind detectors, and to coherent OFDM (known diagonal of $\mathbf{H}$).

low normalized Doppler of $f_dT = 0.01$. The CMCC scheme outperforms all the other blind detectors at low SNR. In this case, the PRC-3 scheme can outperform CMCC and also coherent OFDM at a high SNR. The PRC-3 scheme actually performs better in a frequency-selective channel than in a flat channel, suggesting that the scheme takes advantage of the frequency diversity of the channel due to the larger degree of symbol spreading in the frequency domain. However, to achieve this blind-detection performance, cost function optimization must be implemented with a full complexity brute-force search.

Finally, in Fig. 5.13, we plot BER vs. SNR for the known-statistics CMCC detector for various values of the modulation angle θ in a frequency-selective channel with normalized Doppler $f_dT = 0.05$. Increasing the modulation angle should improve ICI reduction while decreasing the Euclidean distance between coded symbols. It is clear that increasing the modulation angle can improve performance at high SNR to better

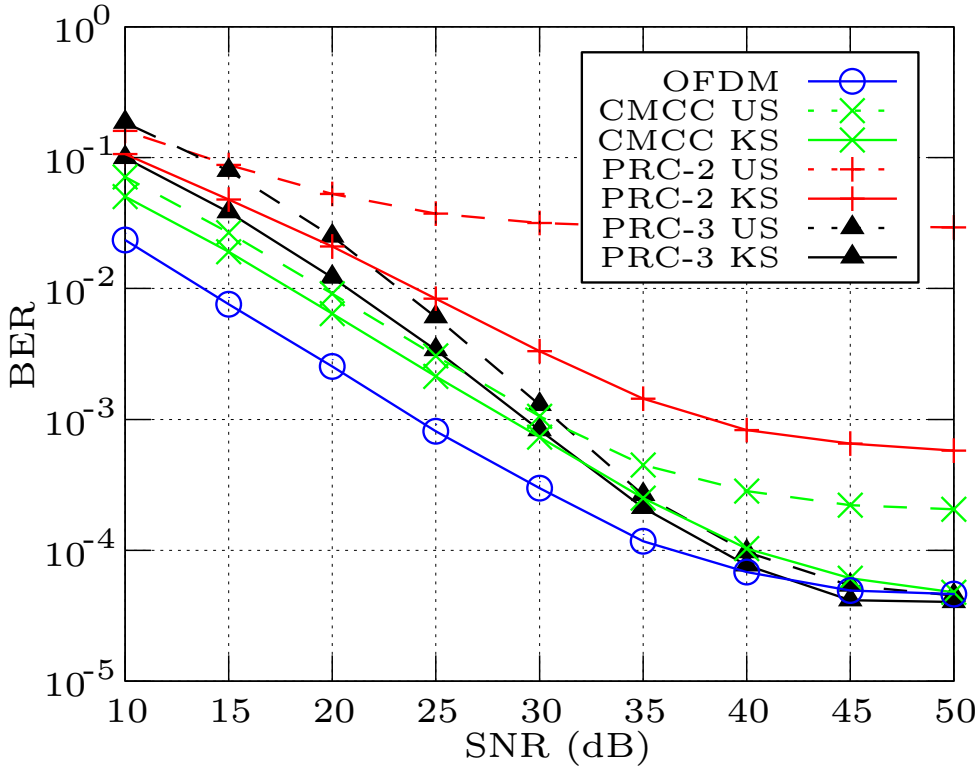

Figure 5.12: BER vs. SNR in an OFDMA system in a frequency-selective Rayleigh-fading channel with $f_dT = 0.01$. The proposed CMCC scheme is compared to PRC-2 and PRC-3, using both the known and unknown-statistics blind detectors, and to coherent OFDM (known diagonal of $\mathbf{H}$).

than that of uncoded coherent OFDM (due to the improved ICI reduction), but the trade-off is poorer performance at low SNR due to closer Euclidean distance between coded symbols.

5.8 Conclusion

We have introduced a new method of correlative coding which reduces ICI power in OFDM due to Doppler spread and consists of constant magnitude symbols. Comparative existing correlative coding methods reduce ICI power, but result in varying-magnitude coded symbols and transmitted null symbols.

The proposed constant-magnitude correlative-coding scheme is simple, effective at reducing ICI power, and–since the resulting coded symbols are constant in magnitude–enables low complexity blind detection. We have shown through bit error rate simulations that the blind detector based on the proposed CMCC scheme can outperform both

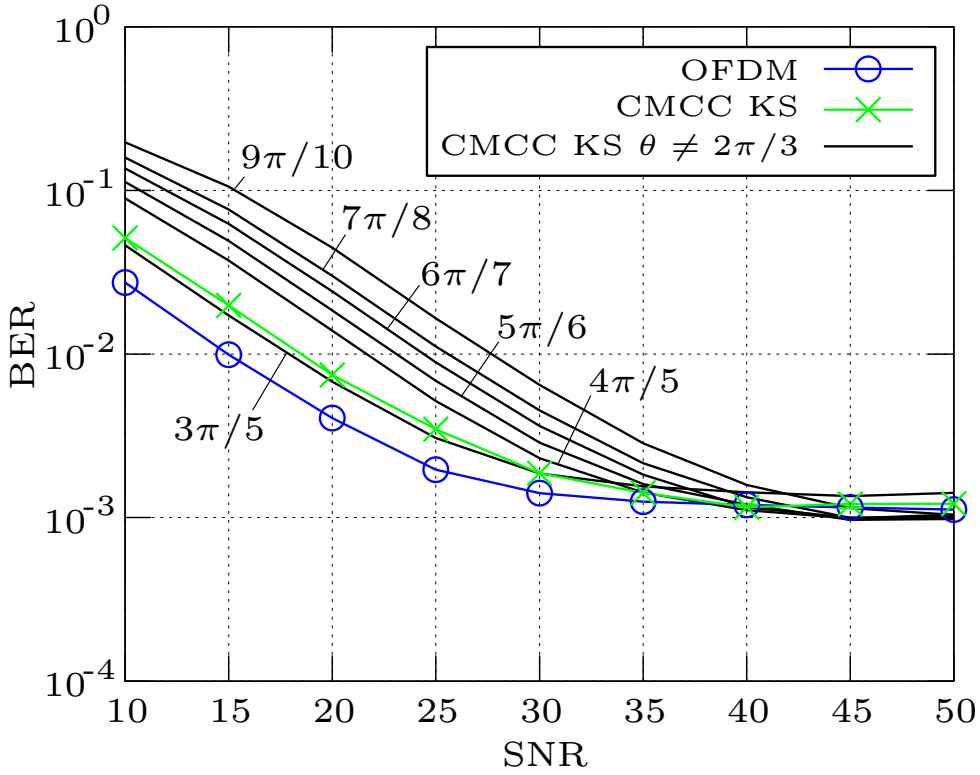

Figure 5.13: BER vs. SNR for various values of θ with normalized Doppler $f_d T = 0.05$.

uncoded OFDM and blind detectors based on previously proposed correlative coding schemes.

Chapter 5, in part, has been submitted for publication to IEEE Globecom 2010, under the title "Constant Magnitude Correlative Coding for ICI Reduction and Blind Detection in OFDM" Nathan D. Ricklin and James R. Zeidler, March 2010. Chapter 5, in part, has also been submitted for publication to the IEEE Transactions on Communications, under the title "Constant Magnitude Correlative Coding for OFDM: ICI Reduction and Low-Complexity Blind Detection" Nathan D. Ricklin and James R. Zeidler, May 2010.

Chapter 6

Conclusions

In this dissertation we considered the problem of digital wireless communications over channels that are rapidly time-varying due to velocity of the receiver or transmitter. In particular we have focused on the proper characterization of such channels and associated CFO estimation and blind detection algorithms. The basic problem that we have attempted to solve is to find a characterization of the channel that accurately represents the physical channel and that also enables low-complexity, well-performing algorithms for both CFO estimation and blind data detection.

In chapter 2 we reviewed the mathematical models that represent the physical channel. We first introduced the discrete-time model of the time-varying frequency-selective channel, a framework through which we performed all the analysis in this dissertation. We reviewed the relevant statistical representations of the channel, and introduced a deterministic channel approximation based on orthonormal polynomials that can accurately portray the channel and enable the derivation of estimators and detectors developed in the next chapters. We introduced the normalized Doppler parameter f_dT and illustrated how it relates to the physical channel parameters. Finally, we reviewed OFDM in the discrete-time channel framework and showed how channel time-variation leads to intercarrier interference (ICI).

In Chapter 3 we addressed the issue of joint CFO/channel estimation over a time-varying frequency-selective channel. Given a known sequence of transmitted pilot symbols, we derived novel joint estimators of both the CFO and the time-varying frequency-selective channel using both a probabilistic and a deterministic channel characterization. Using the probabilistic channel model, the CFO estimate is shown to achieve

the Cramer-Rao lower bound. The performance of the deterministic estimator can also be good, but is sensitive to over or underparameterization of the channel, depending on the number of basis polynomials used in the channel parameterization.

In Chapter 4 we focussed our attention on blind detection in time-varying channels. We developed a blind detector for a short block of data transmitted in the time-domain through a time-varying channel that does not require channel statistics. In developing the detector we considered the channel deterministic and unknown, and approximated it with a linear combination of discrete Chebyshev polynomials. We showed how the resulting cost function can be implemented via the sphere decoding algorithm. Additionally, we show that when the so-called "subset" method is used, the performance of the detector is within about 2.5 dB of the optimal differential-coherent detector over a wide range of SNR and Doppler.

Finally, in Chapter 5 we turned our attention to the problem of blind detection of a short block of symbols transmitted through a time-varying frequency-selective channel. The short block of symbols is placed in the frequency-domain via OFDM, however the data is then correupted by ICI. To both mitigate the ICI and enable low-complexity blind detection we introduced a novel data encoding scheme called constant-magnitude correlative-coding (CMCC). The correlative coding reduces ICI power, and the fact that the resulting symbols are constant in magnitude directly enables low-complexity implementation of a blind detector. Performance simulations show that the blind CMCC encoding and detection scheme is both effective at reducing ICI and can achieve bit error rates better than that of coherent uncoded OFDM and previously suggested correlative coding methods.

Appendix A

Nomenclature

A.1 Acronyms/Abbreviations

AWGN	Additive White Gaussian Noise
BEM	Basis Expansion Model
CAZAC	Constant Amplitude Zero AutoCorrelation sequence
CFO	Carrier Frequency Offset
CMCC	Constant Magnitude Correlative Coding
CRLB	Cramer-Rao Lower Bound
DFT	Discrete Fourier Transform
DPS	Discrete Prolate Spheroidal (a set of orthonormal sequences)
DPSK	Differential Phase-Shift Keying
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
ICI	Intercarrier Interference
IDFT	Inverse Discrete Fourier Transform
ISI	Intersymbol Interference
LOS	Line-of-Sight
MAP	Maximum a Posteriori
ML	Maximum Likelihood
MLSE	Maximum Likelihood Sequence Estimation
MSDD	Multiple Symbol Differential Detection
MSDD-US	Multiple Symbol Differential Detection - Unknown Statistics
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PDF	Probability Density Function
PEP	Pairwise Error Probability
PRC	Partial Response Coding
PSK	Phase Shift Keying Modulation Method
PN	Pseudo-Noise (sequence)
SNR	Signal to Noise Ratio

A.2 Notation

$\mathrm{Re}\{\cdot\}$	The real part of a complex number
$\mathrm{Im}\{\cdot\}$	The imaginary part of a complex number
$\{\cdot\}^*$	The conjugate of a complex number
$\{\cdot\}^T$	The transpose of a matrix or vector
$\{\cdot\}^H$	The conjugate transpose of a matrix or vector
$\{\cdot\}^\dagger$	The Moore-Penrose pseudoinverse of a matrix
$\lvert\cdot\rvert$	The determinant of a matrix, or the absolute value of a scalar
$\lVert\cdot\rVert$	The Euclidean norm of a vector
$\star$	The convolution operator
$\equiv$	Stands for "defined as"
$\approx$	Stands for "approximately equal to"
$\sim$	Stands for "follows the distribution"
$E[\cdot]$	The expectation operator
$e^{\{\cdot\}}$ or $\exp\{\cdot\}$	The exponential function
$\mathbf{x}$	A vector
$[\mathbf{x}]_n$	The nth element of a vector
$\mathbf{A}$	A matrix
$\mathbf{I}$	The identity matrix
$[\mathbf{A}]_{n,m}$	The $\{n, m\}$th element of a Matrix
$\delta(\cdot)$	The Dirac delta function
$J_0(\cdot)$	The zeroth order Bessel function of the first kind
$I_0(\cdot)$	The zeroth-order modified Bessel function of the first kind
$\mathcal{CN}(0, 1)$	The Complex Normal distribution with zero mean and unit variance

Bibliography

[1] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger, "Multiple-symbol differential sphere decoding," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 1981–1985, 2005.

[2] Hua Zhang and Ye Li, "Optimum frequency-domain partial response encoding in OFDM system," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1064–1068, 2003.

[3] R. Clarke, "A Statistical Theory of Mobile Radio Reception," *Bell System Tech. J.*, vol. 47, pp. 957–1000, 1968.

[4] G.B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1969–1986, 1998.

[5] T. Zemen and C.F. Mecklenbrauker, "Time-Variant Channel Estimation Using Discrete Prolate Spheroidal Sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, 2005.

[6] Taewon Hwang, Chenyang Yang, Gang Wu, Shaoqian Li, and G. Ye Li, "OFDM and Its Wireless Applications: A Survey," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 4, pp. 1673–1694, 2009.

[7] T. Wang, J. G. Proakis, E. Masry, and J. R. Zeidler, "Performance Degradation of OFDM Systems Due to Doppler Spreading," *IEEE Trans. Wireless Commun.*, vol. 5, June 2006.

[8] T. Pollet, M. Van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 191–193, 1995.

[9] P.H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, pp. 2908–2914, 1994.

[10] T.M. Schmidl and D.C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, 1997.

[11] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, 2000.

[12] Xiaoqiang Ma, H. Kobayashi, and S.C. Schwartz, "Joint frequency offset and channel estimation for OFDM," in *Proc. Globecom '03*, 1-5 Dec. 2003, pp. 15–19 Vol.1.

[13] Tao Cui and C. Tellambura, "Robust joint frequency offset and channel estimation for OFDM systems," in *Proc. 60th IEEE Veh. Technol. Conf.*, 2004, pp. 603–607 Vol. 1.

[14] J.H. Yu and Y.T. Su, "Pilot-assisted maximum-likelihood frequency-offset estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1997–2008, 2004.

[15] T. Lv, Hua Li, and Jie Chen, "Joint estimation of symbol timing and carrier frequency offset of OFDM signals over fast time-varying multipath channels," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4526–4535, 2005.

[16] M. Nissila and S. Pasupathy, "Joint estimation of carrier frequency offset and statistical parameters of the multipath fading channel," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1038–1048, 2006.

[17] L. Wei, Y.Y. Xu, Y.M. Cai, and X. Xu, "Robust frequency offset estimator for OFDM over fast varying multipath channel," *Electronics Letters*, vol. 43, no. 6, pp. 53–54, 2007.

[18] F. Gao and A. Nallanathan, "Identifiability of Data-Aided Carrier-Frequency Offset Estimation Over Frequency Selective Channels," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3653–3657, 2006.

[19] Steven M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[20] L. E. Franks, "Acquisition of carrier and timing data-I," in *NATO Advanced Study Inst. New Directions Signal Processing Commun. Contr.* Groningen, The Netherlands: Noordhoff, 1975, pp. 429–447.

[21] M. Meyers and L. Franks, "Joint Carrier Phase and Symbol Timing Recovery for PAM Systems," *IEEE Trans. Commun.*, vol. 28, no. 8, pp. 1121–1129, 1980.

[22] K.A.D. Teo and S. Ohno, "Optimal MMSE finite parameter model for doubly-selective channels," in *Proc. IEEE Globecom '05*, 28 Nov.-2 Dec. 2005, vol. 6.

[23] D. Divsalar and M.K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300–308, 1990.

[24] S.G. Wilson, J. Freebersyser, and C. Marshall, "Multi-symbol detection of m-dpsk," in *Global Telecommunications Conference, 1989, and Exhibition. Communications Technology for the 1990s and Beyond. GLOBECOM '89., IEEE*, 1989, pp. 1692–1697 vol.3.

[25] Jr. Mackenthun, K.M., "A fast algorithm for multiple-symbol differential detection of mpsk," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1471–1474, 1994.

[26] P. Ho and D. Fung, "Error performance of multiple-symbol differential detection of psk signals transmitted over correlated rayleigh fading channels," *IEEE Trans. Commun.*, vol. 40, no. 10, pp. 1566–1569, 1992.

[27] Dong Wu and Fuqin Xiong, "Multiple-symbol differential detection of $\pi/4$ QPSK over land mobile satellite communications channels," in *Communications, 1996. ICC 96, Conference Record, Converging Technologies for Tomorrow's Applications. 1996 IEEE International Conference on*, 1996, vol. 2, pp. 901–905 vol.2.

[28] C.R. Nassar and M.R. Soleymani, "Data detection of mpsk in the presence of rapidly changing carrier phase," *IEEE Trans. Veh. Technol.*, vol. 45, no. 3, pp. 484–490, 1996.

[29] G.M. Vitetta and D.P. Taylor, "Viterbi decoding of differentially encoded psk signals transmitted over rayleigh frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1256–1259, 1995.

[30] R. Schober, W.H. Gerstacker, and J.B. Huber, "Decision-feedback differential detection of mdpsk for flat rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1025–1035, 1999.

[31] F. Gini and G.B. Giannakis, "Generalized differential encoding: a nonlinear signal processing perspective," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2967–2974, 1998.

[32] Li Bin and Paul Ho, "Data-aided linear prediction receiver for coherent DPSK and CPM transmitted over Rayleigh flat-fading channels," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1229–1236, 1999.

[33] D.K. Borah and B.T. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, 1999.

[34] Wing Seng Leon and D.P. Taylor, "Generalized polynomial-based receiver for the flat fading channel," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 896–899, 2003.

[35] V.. Pauli and L.. Lampe, "Tree-search multiple-symbol differential decoding for unitary space-time modulation," *Communications, IEEE Transactions on*, vol. 55, no. 8, pp. 1567–1576, 2007.

[36] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Simple method for evaluating error probabilities," *Electronics Letters*, vol. 32, no. 3, pp. 191–192, 1996.

[37] L.H.-J. Lampe and R. Schober, "Bit-interleaved coded differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1429–1439, 2002.

[38] William C. Jakes, *Microwave Mobile Communications*, New York, Wiley, 1974.

[39] G. L. Turin, "The characteristic function of hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, no. 1/2, pp. 199–201, 1960.

[40] V. Pauli, R. Schober, and L. Lampe, "A unified performance analysis framework for differential detection in mimo rayleigh fading channels," *Communications, IEEE Transactions on*, vol. 56, no. 11, pp. 1972–1981, 2008.

[41] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.

[42] M.O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.

[43] C. P. Schnorr and M. Euchner, "Lattice basis reduction: improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, no. 2, pp. 181–199, 1994.

[44] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)," [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25814.htm.

[45] Digital Video Broadcasting (DVB), "Interaction channel for digital terrestrial television (RCT) incorporating multiple access OFDM," Mar. 2002, ETSI.

[46] "IEEE P802.16, Part 16: Air interface for fixed broadband wireless access systems," Revision of IEEE Std. 802.16-2004 as amended by IEEE Std. 802.16f-2005 and IEEE Std. 802.16e-2005, Mar. 2007. Draft.

[47] Xiaozhou Huang and Hsiao-Chun Wu, "Robust and Efficient Intercarrier Interference Mitigation for OFDM Systems in Time-Varying Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2517–2528, 2007.

[48] Chao-Yuan Hsu and Wen-Rong Wu, "Low-Complexity ICI Mitigation Methods for High-Mobility SISO/MIMO-OFDM Systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2755–2768, 2009.

[49] C. Muschallik, "Improving an ofdm reception using an adaptive nyquist windowing," *IEEE Trans. Consum. Electron.*, vol. 42, no. 3, pp. 259–269, 1996.

[50] J. Armstrong, "Analysis of new and existing methods of reducing intercarrier interference due to carrier frequency offset in ofdm," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 365–369, 1999.

[51] Yuping Zhao and S.-G. Haggman, "Intercarrier interference self-cancellation scheme for OFDM mobile communication systems," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1185–1191, 2001.

[52] Yuping Zhao, J.-D. Leclercq, and S.-G. Haggman, "Intercarrier interference compression in OFDM communication systems by using correlative coding," *IEEE Commun. Lett.*, vol. 2, no. 8, pp. 214–216, 1998.

[53] Wei Bai and Younghak Kim, "Effects of Correlative Coding on OFDM Systems for Intercarrier-Interference Suppression," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3732–3738, 2007.

[54] J. Winters, "Differential Detection with Intersymbol Interference and Frequency Uncertainty," *IEEE Trans. Commun.*, vol. 32, no. 1, pp. 25–33, 1984.

[55] R. Pawula, "Offset DPSK and a Comparison of Conventional and Symmetric DPSK with Noise Correlation and Power Imbalance," *IEEE Trans. Commun.*, vol. 32, no. 3, pp. 233–240, 1984.

[56] R. Ramesh, R.L. Toy, R.D. Koilpillai, and S. Chennakeshu, "A new family of modulation schemes for land mobile radio," in *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95. 'Wireless: Merging onto the Information Superhighway'., Sixth IEEE International Symposium on DOI - 10.1109/PIMRC.1995.480913*, 1995, vol. 2, pp. 471–476 vol.2.

[57] J. G. Proakis, *Digital Communications*, New York: McGraw-Hill, 4th edition, 2001.

[58] N. Ricklin and J.R. Zeidler, "Block Detection of Multiple Symbol DPSK in a Statistically Unknown Time-Varying Channel," in *Communications, 2009. ICC 2009. 2009 IEEE International Conference on*, 2009.

[59] N. Ricklin and J. R. Zeidler, "Multiple symbol differential detection of MPSK over statistically-unknown, frequency-flat, time-varying channels," *IEEE Trans. Commun.*, in submission.

[60] P. Pun and P. Ho, "The performance of fano-multiple symbol differential detection," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on DOI - 10.1109/ICC.2005.1494797*, 2005, vol. 4, pp. 2516–2521 Vol. 4.