UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Optimizing symbol timing, frequency spacing, and SNR estimation for communication systems**

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

in

Electrical Engineering (Communication Theory and Systems)

by

Aniruddha Das

Committee in charge:

Professor Bhaskar D. Rao, Chair
Professor William S. Hodgkiss
Professor Ryan Kastner
Professor Kenneth Kreutz-Delgado
Professor Lawrence B. Milstein

2013

The dissertation of Aniruddha Das is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2013

DEDICATION

*To Ma, Baba, and Lopa*

# EPIGRAPH

*To strive, to seek, to find, and not to yield.*

— Tennyson

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have been incredibly fortunate to have received the support and love of a large number of people throughout my professional and personal life and while it will be impossible to acknowledge each one individually, undoubtedly some deserve special mention.

Special thanks are due to my advisor, Prof. Bhaskar D. Rao. For a part time student with significant work, travel, and family commitments, it required a special kind of advisor and in Prof. Rao, I found just that person. Not only did I get excellent and creative ideas, advise, and timely feedback, but, and equally important, his patience and understanding of my other commitments were key to getting me to the finish line. I would also like to thank the rest of my committee members for their time and involvement and for teaching some of the excellent courses which taught me the background material required to make progress.

My connection to UCSD would never have taken off had it not been for the late Magnus Almgren who got me involved with CalIT2. Magnus' insight, creativity, and brilliance are sorely missed. Thanks also to Prof. Ramesh Rao who allowed me the freedom to explore and play around in CalIT2, to Arnold, Josh, and to Prof. Bill Hodgkiss and Prof. Pankaj Das for navigating through the bureaucracy and letting me teach at UCSD. I am indebted to all my teachers and my undergraduate professors and especially to Prof. John Molinder for opening the door to the world of digital communications for me.

I am very grateful to ViaSat Inc., for not only providing financial support, but also for allowing a very flexible schedule. I've been fortunate to be able to work with, and learn from, very talented engineers at ViaSat. In particular, the exemplary vision and technical acumen of Mark Miller, an inspiration for me, deserves special mention. Thanks are also due to MarkD, Steve Hart, Don Becker, Rich Lindstrom, Simon, Kristi, Dave, Tim, Nirmal, Girish, Marc, my many supervisors, to Kevin, Shameem, Cliff, Mark, Andrew, Clarice, Navdeep, Brian, Veena, Bill, Gina, Sandeep and everyone else on the incredible Surfbeam team.

The chapters of this dissertation have been adapted from works that have been published or have been submitted for publications and I am grateful for the

contribution of my co-authors and reviewers.

- Chapter 2 is adapted from "MIMO Systems with Intentional Timing Offset," Eurasip Journal on Advances in Signal Processing, 2011, and "Impact of Receiver Structure and Timing Offset on MIMO Spatial Multiplexing," IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2008, both papers by A. Das and B.D. Rao.
- Chapter 3 has been adapted from a) A. Das and M. Miller, "Remote nonlinearity detection via burst power dithering and EM based SNR Estimation," International Workshop on Satellite and Space Communications 2007, (IWSSC 2007), b) A. Das, "NDA SNR estimation: CRLBs and EM based estimators," IEEE Region 10 Conference, TENCON 2008 and c) from the work behind US Patent 7,894,510, issued July 05, 2011.
- Chapter 4 has been adapted from A. Das and B. D. Rao, "SNR and Noise Variance Estimation for MIMO Systems," IEEE Transactions on Signal Processing, August 2012.
- The contents of Chapter 5 have been submitted for consideration of publication to the IEEE Transactions of Signal Processing.

My friends have always enriched my life and I am grateful to be able to count many talented and wonderful people as my friends. Dhananjay, Kamesh, Mayur, Rohit, Muthu, Kaushik, Basundhara, Kiran, Venkat, Niraj, Devavrat, Matt, Deny, Cheryl, Allison, Tom, Ratul, Rajarshi, Rakhee, Arnab, Nilim, Antar, Esha, thanks for the memories and the friendship. The "desigurus", and especially Reshmi, Pushkar, Chhipli, Dhruba, Piu, Tuku, Atish, Malabika, Rajesh, Muniya, Bidhan, Swapna, Soumitra, Lali, Chumki, Bappa, Debra, Madhumita, and their talented kids provided the closest replacement one can find to an extended family – you are the best! The Saikat family of San Diego, and especially Mahasweta, Sandip-da, Pradip-da, Chandana-di, Aloke Raj, Sharmila, Pijush-da, Indrani-di, Muthya, Bishwaroop, Easeeta-di, Siddhartha-da, Avijit-da, Gauri-di, Gargi, Kaushik, Shweta, Badri, and LA's Rakhi and Arup never failed to open their hearts and homes, always brightening up a weekend. Thanks to Rajo, Bappa, Ana,

Sanjay, Arijit, Praveen, Sokuba, Imti, Shamik, Dhuju, Debdeep, Saugata, Baloo, Dadu, Debajyoti, Alfi, Sauvik, and the rest of SJC94, for still finding all *those* jokes funny.

Throughout my life, I had the fortune of having a very supportive family and I owe so much to so many. I am extremely grateful to my uncle Dev Purkayastha, without whose generosity I would probably never have come to the US; to my uncle Dipankar Purkayastha who invariably found help for me to finish my college applications in an era when I didn't know much about computers; to my aunts and uncles Jharna & Mohit Dutta Roy, Aparna & Ronen Dam, and Naina whose love and generosity were omnipresent. Some of them are sadly no longer around to see this day, but whose impact on my life is unquestionably profound. My late grandparents served as beacons of inspiration, one of whom was an engineer, and the other, a widowed single mother who was a successful high school principal in an era when her gender all but precluded her from success. I am incredibly thankful to many more: mamas, mashis, jethi, kaku, kaki, chhordamoni, chordimoni, dadamoni, didimoni, my cousins, Pratima-didi and so many others. Invariably I have left out some, and for that, I apologize. They formed the incredible tableau that is my family and while it is not often that I say thanks, from an early childhood, my entire family inspired me to strive for the best and I am very grateful for the role models, the sense of place and history, and support and love I've always received.

The PhD journey has been made that much nicer due to the support and love of my wife, Lopa. Paradoxically, she ensured that research wasn't the sole focus of my life. Her creative thinking, willingness to challenge the conventional, and sense of aesthetic have added new dimensions to my life. Hopefully the end of this journey is the beginning of many others, and we shall have more time to look up and wonder at M31, climb more pyramids, and snorkel in more exotic places. I also owe Lily a ton of thanks for never failing to stay up nights with me; she probably knows more about this material than I will ever know.

It will be futile even to try to enumerate the reasons for which I am grateful to my parents. This work is as much their achievement as it is mine.

## VITA

| | |
|---|---|
| 1998 | Bachelor of Science, Engineering, Harvey Mudd College, Claremont, California |
| 2000 | Master of Science, Electrical Engineering, Stanford University, Stanford, California |
| 2013 | Doctor of Philosophy, Electrical Engineering (Communications Theory and Systems), University of California, San Diego, La Jolla, California |

## PUBLICATIONS

**A. Das** and B. D. Rao, "SNR and Noise Variance Estimation for MIMO Systems," Signal Processing, IEEE Transactions on, August 2012.

**A. Das** and B. D. Rao, "MIMO Systems with Intentional Timing Offset," Eurasip Journal on Advances in Signal Processing, 2011.

**A. Das**, "NDA SNR estimation: CRLBs and EM based estimators," TENCON 2008. IEEE Region 10 Conference, 2008, pp. 1-6.

**A. Das** and B. D. Rao, "Impact of Receiver Structure and Timing Offset on MIMO Spatial Multiplexing," in Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on, 2008.

**A. Das** and J. Ng, "NDA SNR Estimation for 32APSK in AWGN," in Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on, 2008.

B. D. Rao and **A. Das**, "Multiple Antenna Enhancements via Symbol Timing Relative Offsets (MAESTRO)," in Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on, 2007, pp. 1-5.

**A. Das** and M. Miller, "Remote non-linearity detection via burst power dithering and EM based SNR Estimation," Satellite and Space Communications, 2007. IWSSC 2007. International Workshop on, Sept. 2007. **Best Paper Award Winner.**

**A. Das**, "Enhancing capacity of a satellite broadband system via adaptive coding and modulation," 24th AIAA International Communications Satellite Systems Conference, ICSSC, 2006, p 532-538

Z. Cao, **A. Das**, K. P. Liolis, M. Almgren, and B. D. Rao, "System level design methodology for wireless MIMO prototyping," in Radio and Wireless Symposium, 2006 IEEE, 2006, pp. 67-70.

## PATENTS

**A. Das**, B. Collins, "Adaptive satellite return link symbol rate determination," U.S. Patent number 8,493,877, July 23, 2013.

M. Miller, T. Eidenschink, **A. Das**, "MIMO satellite system," U.S. Patent number 8,130,693, March 6, 2012.

M. Dankberg, **A. Das**, "Scalable satellite deployment," U.S. Patent number 8,005,034, August 23, 2011.

M. Dankberg, **A. Das**, M. Miller, C. Pateros, "Multi-antenna satellite system with wireless interface to vehicle," U.S. Patent number 7,974,571, July 05, 2011.

M. Miller, **A. Das**, R. Fitting, "Remote non-linearity detection via burst power dithering," U.S. Patent number 7,894,510, February 22, 2011.

## PROFESSIONAL APPOINTMENTS

2003-Present, ViaSat Inc., Carlsbad, CA

2002-2003, Trellisware Technologies Inc., San Diego, CA

2000-2002, Siliconwave Inc., San Diego, CA

ABSTRACT OF THE DISSERTATION

**Optimizing symbol timing, frequency spacing, and SNR estimation for communication systems**

by

Aniruddha Das

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

University of California, San Diego, 2013

Professor Bhaskar D. Rao, Chair

This dissertation presents techniques to improve the performance of both coherent as well as non-coherent wireless communication systems via optimizing symbol timing, frequency spacing and by making efficient SNR estimations. We show that some of the design choices made in traditional systems are not optimal and demonstrate the gains that may be achieved by making unconventional, but judicious, choices for these parameters.

We start in the area of coherent multi-antenna communications where we introduce an offset between the symbol boundaries of the transmitted waveforms from the different antennas and show that this improves performance in comparison

to the traditional symbol aligned transmission. For this modified system, we derive various optimal receivers such as maximum likelihood (ML), best linear unbiased estimator (BLUE), minimum mean squared error (MMSE), and zero forcing (ZF) receivers and show that they outperform the equivalent receiver for the system with aligned symbol boundaries. In some system configurations, the performance gain is close to 2dB. The design methodology for a new symbol pulse shape that increases the performance even more is also presented.

Next, we extend the study of SNR estimation from the previously published results of a data aided (DA) single antenna system to the non-data aided (NDA) model and also to systems with multiple antennas (MIMO). In both these cases, we have derived the Cramér-Rao lower bound (CRLB) as well as ML estimators that achieve or perform very close to the CRLB. For MIMO systems we define the SNR and then derive the CRLB and the ML estimators for both the DA as well as the NDA data model. We show that previously published results for single antenna systems are a special case of our general solution. The proposed SNR estimation techniques are demonstrated in a patented algorithm to detect the onset of non-linearity in a remote transmitter by dithering the power of transmitted bursts and estimating the difference in the received SNR.

For non-coherent systems we show that the performance of multi-tone M-ary frequency shift keying (MT-MFSK) modulation may be significantly improved if, instead of the usual choice of mutually orthogonal tones, non-orthogonal tones are used. In some system configurations, the proposed system can lead to a **4-fold increase** in system capacity. The channel capacity, as well as the performance gains of systems using practical receivers such as ML, least squared (LS) error, and compressed sensing (CS) are demonstrated for both flat and frequency selective channels. Many more choices of spectral efficiency are achievable by the non-orthogonal system, thus enabling the system to adapt to changing link SNR and send data at the optimum spectral efficiency. In order to make this practical, we derive the CRLB and ML estimators for SNR estimation for non-orthogonal MT-MFSK in both the DA as well as the NDA data model.

# Chapter 1

# Introduction

The digital lifestyle of the 21st century, with ubiquitous computing, has fundamentally changed the way people think about Internet connectivity. People expect to be able to connect to cheap, reliable, high speed broadband networks not only from the comfort of their homes, but also from fast moving planes, trains, boats etc. This expectation of constant connectivity is no longer confined to traditional computers and laptops, but has also extended to include mobile phones, televisions, tablets, and other everyday home appliances. This "always on" lifestyle has brought about an explosion of data that is transmitted through our communications networks. Cisco, a major network equipment company, estimates that "*globally, mobile data traffic will grow 13-fold from 2012 to 2017, a compound annual growth rate of 66%*," and that "*annual global IP traffic will pass the zettabyte threshold by the end of 2016.*"[1]

With the growing scarcity and increasing expense of available spectrum, the challenge for the engineering community is to develop systems that will allow this hunger for data to grow unabated in the years ahead. At the physical layer, some key innovations that address this challenge include modulation and coding that operate close to the theoretical Shannon limits, higher dimension signals using multiple antennas, and high capacity satellite systems that solve the last-mile problem. This dissertation attempts to take another step towards this never ending

[1] Cisco Systems: Zettabyte = $10^{21}$ bytes. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf

goal of increasing the efficiency and performance of our communication systems.

We propose techniques that will significantly improve the performance of both coherent as well as non-coherent systems by judicious, but unconventional, use of symbol timing and frequency spacing. We also demonstrate theoretical lower limits for SNR estimation and derive efficient practical estimators that achieve these limits. The proposed SNR estimation techniques are demonstrated in an innovative application to detect the onset of compression in a transmitter.

## 1.1 Coherent versus Non-Coherent Systems

Communication systems may be classified into two broad groups based on the amount of information that the system uses about the channel through which the signal propagates. The wireless communication channel can impose gain and phase uncertainty to the transmitted signal amongst other possible distortions in the time and frequency domains [89], [74], [85]. At one end of the spectrum, receivers are assumed to know the channel perfectly. This is known as coherent communications with full channel knowledge as discussed in Chapter 3 of [89]. At the other end of the spectrum, receivers are assumed to operate blindly and make no attempt to learn about the channel. This is known as non-coherent communications. Within the class of non-coherent communications, often, it is easier to estimate the channel gain than the channel phase and so one sub-class of non-coherent communications assumes the knowledge of channel gain, but not that of phase as discussed in Chapter 5 of [85]. This is a reasonable assumption since channel phase can change quickly due to numerous reasons and hence, may be hard to estimate. First, the oscillators in the communication chain could have excessive phase noise – common especially in systems where it is required to use cheap oscillators. Second, as mobile terminals move, the propagation distance of the signals change and even very small changes in propagation distance or time, can lead to big changes in the phase of the signal. Consequently maintaining phase synchronization in high data rate systems with high mobility is a significant challenge and often some receivers choose to operate without trying to achieve

phase synchronization.

In the following chapters we shall be discussing both classes of communication systems. We start by discussing coherent communications in Chapter 2, and we end with non-coherent communcations in Chapter 5. In between, in Chapters 3 and 4 we deal with systems where the receivers attempt to estimate the channel (with and without phase and gain uncertainty) in order to estimate the SNR of the link.

## 1.2 MIMO Systems with Offset Symbol Boundaries

One of the most promising innovations at the physical layer of communication systems has been the use of multiple antennas at both the transmitter as well as at the receiver. Popularly referred to as multiple input, multiple output (MIMO) systems, under certain conditions, such systems are able to significantly increase the capacity of the channel [89], [71]. Even though, theoretically speaking, this allows for an infinite expansion of channel capacity, practical restrictions imposed by the operating environment and device size limitations ensure that the gains, though significant, are limited. MIMO systems are now a part of many wireless standards such as LTE.

In these MIMO systems, especially when a single carrier system is used, the usual design choice is to align the transmit symbol boundaries of the signals from the various transmit antennas. In order to limit the transmit signal energy to the bandwidth of interest while reducing inter symbol interference (ISI), typically the transmit waveform is pulse shaped with square root raised cosine (SRRC) filters with non-zero excess bandwidth. The frequency response of some raised cosine filters with excess bandwidth is shown in Fig. 1.1. The minimum bandwidth required to transmit signals while meeting the Nyquist criterion is from $-\frac{1}{2T_s}$ to $\frac{1}{2T_s}$, where $T_s$ is the symbol duration, however, practical systems use some "excess bandwidth".

Excess bandwidth is defined to be the additional bandwidth in excess of

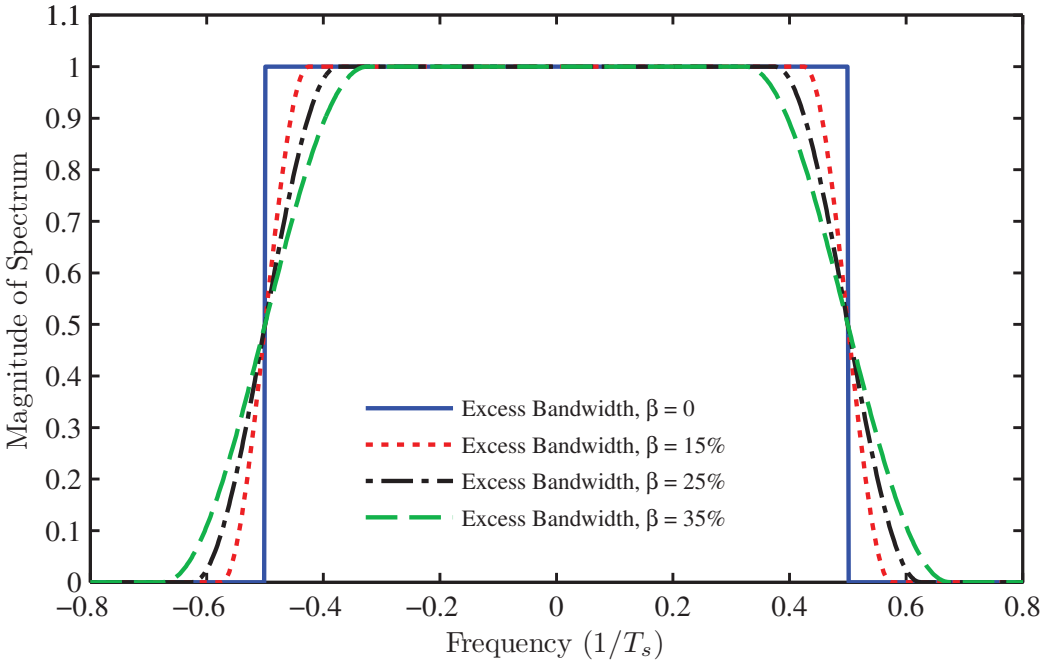


**Figure 1.1**: Raised cosine pulse shapes with excess bandwidth, $\beta$.

the theoretical minimum (the Nyquist bandwidth) that is required to transmit the waveform and is usually expressed as a fraction. This excess bandwidth leads to some exploitable inefficiencies. In Chapter 2 we show that by intentionally introducing a known non-zero relative offset between the transmit waveforms, performance gains may be obtained.

The signal at any receive antenna in a MIMO system is a linear combination of the transmit signals from all the antennas, each one weighted by its corresponding channel gain. Consequently, assuming equal average channel gains for the various paths, the receiver receives the various transmit signals, on average, at equal power. With aligned symbol boundaries, the optimal sampling time (at the receiver) for all the transmit signals is the same, thus the average received power of the signals from all the transmit antennas is the same. We refer to this interference power of the various transmit signals interfering with each other at the receiver as the inter antenna interference (IAI) power. Another source of degradation is inter symbol interference (ISI) where, due to imperfect timing recovery

or perhaps due to the presence of multiple propagation paths, the symbol energy is spread over multiple symbol intervals. Due to the use of Nyquist pulse shaping and symbol timing recovery loops that typically are close to perfect, ISI power is typically quite low.

By offsetting the transmit symbols relative to one another, we can reduce the IAI while increasing the ISI since the optimal sampling points of the various transmit symbols are no longer synchronized. This is shown in Fig. 2.1 in Chapter 2. We show that an advanced receiver can utilize this information to extract significant performance gains. The ISI converts a memoryless modulation like $M$-ary phase shift keying (MPSK) to one with memory. Thus, not only are single shot receivers (receivers which decide on a transmit symbol by only observing the corresponding received symbol) useful, but also sequence detection receivers (receivers which decide on a transmit symbol, by observing a sequence of received symbols) are also useful. We demonstrate the performance with different kinds of receivers including zero forcing (ZF), minimum mean squared error (MMSE), and sequence detection based maximum likelihood (ML) receivers.

The design of Nyquist pulse shapes has been studied for a long time but the focus has usually not been on reducing the interference power at specific symbol offsets. This is only to be expected since these pulse shapes are used either in point to point systems or in symbol aligned MIMO systems and these systems do not gain by reducing the signal power at specific offsets. However, in our case, if we include this additional constraint of reducing the signal power at specific symbol offsets we can improve the performance even further. In Chapter 2 we outline and demonstrate a methodology to design such pulse shapes and show additional improvements.

## 1.3 NDA SNR Estimation for SISO Systems

In most communication systems, certain parameters have to be periodically estimated in order to optimize or adapt the performance of the system. One of the key parameters of interest is the signal to noise ratio (SNR). For most point to

point systems, SNR is defined as the ratio of the received signal power to the noise power measured in the same bandwidth. For signal constellations normalized to unit transmit power, this is equivalent to the ratio of the channel gain to the noise power. One of the most common applications of SNR estimation is in adaptive coding and modulation. In this technique, the link's SNR is used to vary the coding and modulation of the signaling between the transmitter and the receiver. There are numerous other applications of SNR estimation such as turbo decoding, network monitoring and control etc.

The field of SNR estimation can be broadly split into two distinct areas – data aided (DA) and non-data aided (NDA) SNR estimation. In the DA system model, the transmitted symbols are assumed to be perfectly known while in the NDA model, only the symbol constellation is assumed to be known, but not the transmitted symbols. DA SNR estimation for single input single output (SISO) systems have been studied for a long time and both the theoretical lower bounds as well as estimators that achieve (or come very close to) the theoretical lower bounds are well established. In this regard, the Cramér-Rao lower bound (CRLB), which is the lowest variance that any unbiased estimator may achieve (subject to some regularity conditions) is a common metric of comparison.

Although DA SNR estimation has been well studied, NDA SNR estimation had not been studied in depth until recently. Although a few specific cases (e.g., BPSK and QPSK modulation) had been studied, until recently the general solution for different modulations was unknown. In chapter 3 we present the CRLB for NDA SNR estimation for single antenna SISO systems for various modulation schemes. The technique presented is very general and can be easily adapted to any digital modulation. We also develop an iterative ML estimator using the expectation-maximization (EM) algorithm that performs very close to the CRLB.

### 1.3.1 NDA SNR Estimation for Remote Non-Linearity Detection

The techniques developed for NDA SNR estimation are also applied to the problem of detecting the onset of non-linearity in a transmitter. Most communica-

tion schemes using a non-constant envelope modulation scheme need to operate in the linear region of their transmit power amplifiers in order to prevent distortion as well as to lower adjacent channel interference from spectral regrowth. Thus, a key task of many power control schemes is to determine when a terminal's power amplifier is entering or operating in the non-linear region.

In Chapter 3 we present an innovative patented scheme that dithers the transmit power of its transmitted bursts on a symbol by symbol basis. Thus e.g., the even numbered symbols could be transmitted at a higher power than the odd numbered symbols for a particular burst. An alternate dithering pattern could transmit the first half of the burst at a higher power than the second half. The key point is that a differential power profile is introduced in the digital constellation (before the power amplifier), while keeping the total transmit power of the burst the same. In a perfectly linear system, the received SNR difference between the high power region and the low power region would be identical to that of the transmitted power difference. However, as the system enters compression, the difference between the received SNR for the two different power regions is reduced. This can be used as a metric to predict whether the power amplifier is entering its non-linear region and is discussed in more depth in Chapter 3.

In order to ensure accurate estimation, iterative ML estimators that approach the CRLB bound are employed. We not only present the implementation of the iterative algorithm but we also present simpler to implement algorithms that can be efficiently implemented on digital hardware such as field programmable gate arrays (FPGA). The techniques described in Chapter 3 have been implemented and have been deployed on a real network with close to 1 million users.

## 1.4 SNR Estimation for MIMO Systems

After presenting the results for NDA SNR estimation for SISO systems and applying it to the problem of detecting transmitter non-linearity in Chapter 3, we generalize the solution to multiple input, multiple output (MIMO) systems in Chapter 4. The generalization is not only in expanding the number of antennas

in the transmitter and receiver, but in considering a complex valued channel and presenting an extension that may be applied to frequency selective channels.

For MIMO systems, SNR estimation is more complicated since there is no unique definition of SNR. The received signal at the receiver of any MIMO system is expected to contain the linear mixture of the multiple transmitted signals. Thus, there is no unique definition of signal power. One may define an SNR for each pair of transmit-receive antenna where the signal power is regarded as the received power of the signal transmitted *only* from the transmit antenna in consideration. This leads to as many SNR parameters as there are combinations of transmit and receive antennas (i.e., the product of the number of received antennas and the number of transmit antennas). Another definition could define signal power as the total power of *all* the signals received at a receive antenna. There can be other definitions of signal power as well.

In Chapter 4, we chose to define the SNR for MIMO systems as the ratio of the Frobenius norm of the channel to the noise power. One of the primary motivations for this definition is in its simplicity and aesthetic appeal since this general form resolves to the usual definition of SNR for single antenna systems if the number of antennas is set to one in the formulations. Thus, the formal expressions derived in Chapter 4 are generalized versions of the single antenna results and we show that the previously published results are special cases of our general formulation. The second motivation for this definition is that this form shows up in well known equations for calculating the MIMO capacity, analogous to the role of the single antenna SNR in the well known equation for Shannon's channel capacity formula.

Using this general definition, in Chapter 4, we derive the CRLB and ML estimators for both NDA and DA SNR estimation for MIMO systems. We derive closed form expressions in the DA model, closed form expressions that are valid over specific ranges of SNR in the NDA model, and also numerical results for the NDA model that are valid for all SNRs. Our derivations assume a channel with complex gain which is a distinction from previous works some of which worked in the less general domain of a real valued channel gain. The ML estimators derived

are shown to operate very close to the CRLB for all SNRs of practical interest.

## 1.5 Rate Adaptive Non-Orthogonal MT-MFSK

The improvements discussed thus far pertain to the arena of coherent communications. In coherent communications, the channel is either estimated or assumed to be known at the receiver. In many practical instances, however, channel estimation is futile. As discussed in Sec. 1.1, even if the channel gain changes slowly and can be estimated, maintaining phase coherence may not be possible. Such communication systems typically use non-coherent communication techniques where it is assumed that the channel phase and sometimes the entire channel is unknown at the receiver.

For non-coherent communications, orthogonal signaling, e.g., $M$-ary frequency shift keying (MFSK) has been the de facto standard for about 50 years. Apart from the simplicity of the form of the optimal receiver in non-coherent systems, MFSK has an additional advantage of having a constant envelope transmit signal. This leads to cheaper implementation costs. Although MFSK using orthogonal tones is very useful in non-coherent channels, its major disadvantage is its low spectral efficiency. To maintain orthogonality, the minimum separation of the tones is required to be equal to the symbol rate of the system. This leads to a maximum spectral efficiency of single tone MFSK of just over 0.5 bits/s/Hz. Modern communication systems typically have spectral efficiencies that are significantly higher. To overcome this deficiency, multi-tone (MT-MFSK) systems where each symbol consists of multiple simultaneous orthogonal tones have been recently proposed.

In Chapter 5 we show that if non-orthogonal tones are used instead of orthogonal tones in MT-MFSK systems, and if the mapping of tones to symbols is intelligently carried out, then non-orthogonal MT-MFSK outperforms orthogonal tone MT-MFSK. We demonstrate this performance gain for various practical configurations in both flat fading as well as frequency selective fading channels.

By calculating the channel capacity we demonstrate that the improvement is

a fundamental characteristic of the non-orthogonal transmission scheme proposed and at specific operating points the spectral efficiency can improve well over 3dB. This is usually well worth the incremental increase in implementation complexity and cost due to the use of a multi-tone waveform.

For practical configurations, we derive optimal maximum likelihood (ML), least squared error (LS), and compressed sensing (CS) receivers. For flat fading non-coherent channels with only an unknown phase, we show that the form of the ML receiver is very similar to that of the equivalent receiver for orthogonal MT-MFSK. These ML receivers perform poorly when the channel is frequency selective and so we develop the LS receiver which is more appropriate when each of the transmitted tones have a different phase uncertainty. Compressed sensing is becoming popular as a way to solve underdetermined systems that have a sparse solution with less complexity than the ML receiver. Since our system has numerous possible transmit tones, of which only a small minority are transmitted simultaneously, the solution we seek is sparse, thus falling directly into the ambit of compressed sensing.

In addition to outperforming the equivalent orthogonal tone MT-MFSK, the proposed non-orthogonal system offers the designer considerable more flexibility, thus allowing modulations of different spectral efficiencies to have the same symbol rate. For a system with fixed bandwidth and symbol rate, orthogonal tone MT-MFSK functions within the constraints of limited combinatorial choices which sometimes results in very few design choices. As an example, with 2 orthogonal tones, the designer has no flexibility to increase the spectral efficiency since there can at most be 2 symbols. Adding more orthogonal tones would either increase the bandwidth (with constant symbol rate) or decrease the symbol rate (with constant bandwidth). The same bandwidth could, however, be occupied by any arbitrary number of non-orthogonal tones (while keeping the symbol rate unchanged), and an appropriate subset of symbols chosen depending on the desired spectral efficiency. With 4 orthogonal tones, the designer has only 2 choices – to send 1 tone per symbol (which results in 4 symbols) or to send 2 tones per symbol (which results in 6 symbols). In contrast a non-orthogonal system operating in the same bandwidth,

could use a constellation with any desired number of symbols at the same symbol rate.

This flexibility allows us to introduce the concept of adaptive coding and modulation to the world of non-coherent MFSK modulation. Using this scheme, e.g., a MFSK modulated waveform shared between various users each with a different link SNR could have significantly higher overall spectral efficiency since the data for each user could be modulated at the spectral efficiency that is appropriate for each user. In order to make this achievable, we have also derived SNR estimation techniques for non-orthogonal MT-MFSK. We derive both the CRLB as well as ML estimators for both NDA and DA SNR estimation for MT-MFSK.

## 1.6 Contribution Summary

The contributions in this dissertation have been summarized in Table 1.1:

**Table 1.1**: Summary of contributions

| **Chapter** | **Contributions** |
|---|---|
| Chapter 2 | Proposed a MIMO system with offset symbol boundaries.<br>Derived optimal receivers for offset MIMO system.<br>Proposed new pulse shape design technique that is appropriate for an offset MIMO system. |
| Chapter 3 | Derived CRLB for NDA SNR estimation for SISO systems.<br>Derived iterative ML estimator based on the EM algorithm that achieves the CRLB for all practical SNRs.<br>Proposed technique to detect non-linearity at a remote transmitter.<br>Proposed computationally simple algorithm that may be easily implemented on an FPGA. |
| Chapter 4 | Proposed a new definition of SNR for MIMO systems. |

Continued on next page

**Table 1.1 – continued from previous page**

| Chapter | Contributions |
|---|---|
| | Derived NDA and DA CRLB for MIMO SNR estimation. |
| | Derived closed form ML estimators for DA MIMO SNR estimation. |
| | Derived CRLB and estimators for noise variance of MIMO systems. |
| | Derived NDA ML estimators for MIMO SNR estimation based on EM algorithm. |
| Chapter 5 | Proposed a new MT-MFSK scheme that utilizes non-uniformly spaced non-orthogonal tones. |
| | Derived optimal receivers for proposed MT-MFSK scheme for frequency selective and frequency flat fading channels. |
| | Derived channel capacity for proposed scheme for coherent and non-coherent channels. |
| | Derived expression for symbol error rate for proposed scheme for non-coherent channels. |
| | Proposed techniques for computationally tractable receivers using CS techniques. |
| | Derived CRLB for DA and NDA SNR estimation for non-coherent MT-MFSK. |
| | Derived ML estimators for SNR estimation for DA and NDA system models. |

# Chapter 2

# MIMO Systems with Offset Symbol Boundaries

## 2.1 Introduction

Ever since some of the pioneering research done at Bell Labs in the last decade of the 20th century, multi antenna systems have been one of the most exciting new frontiers of research at the physical layer of communication systems [71], [89]. The use of multiple antennas at the transmitter and/or receiver opened up the prospect of leveraging the additional dimension of space to increase the performance substantially in very fundamental ways. Most modern wireless communication standards such as 3GPP LTE, IEEE 802.16 (WiMax), IEEE 802.11n and others incorporate multi antenna communications. These multiple antenna systems are also known as space-time communication systems or multiple input, multiple output systems.

In multiple input multiple output (MIMO) communication systems, typically the transmitters are all collocated and the system is designed such that the symbol boundaries are aligned at the transmitters and, if the path between the transmitters and the receivers have the same path delay, also aligned at the receivers. It has been shown in [100] that, under the assumption of a richly scattered environment, such a system can lead to very high spectral efficiencies.

Practical communication systems typically use pulse shaping such as the square root raised cosine (SRRC) to limit the bandwidth occupied by the signal [74], [43], [55]. These pulses typically have an "excess bandwidth" which is usually denoted by a factor $0 \leq \beta \leq 1$. The presence of excess bandwidth was used to improve performance in a fractionally sampled orthogonal frequency division multiplexing (OFDM) system in [88] where the cause of gain is similar to that discussed here, even though the system under consideration is very different.

### 2.1.1 Overview of Prior Work

We showed some preliminary results and demonstrated that significant gains could be obtained via a system with intentionally offset transmissions in [77]. Independently, and at about the same time, Shao et al. also presented a similar MIMO scheme with sub-symbol timing offsets between the transmitted signals [84], [83] and Wang et al. presented a frequency domain equalization scheme for MIMO OFDM with intentional timing offsets in [94]. The contents of this chapter is a modified version of our works, [34], [33], where we showed that contrary to previously published results by Shao et al., the offset scheme is superior for all block sizes. More recently, the capacity of MIMO systems with asynchronous pulse amplitude modulation (PAM) was studied in [6] where the authors show that offset transmission schemes increase the capacity of MIMO systems.

Delay diversity schemes for transmission, proposed previously (see e.g., [99] and [87]), might appear to be similar to the proposed scheme since those schemes also involve offset transmission. However, there are a couple of significant differences. First, delay diversity transmit schemes aim to increase the spatial diversity by transmitting the same, or precoded data stream, whereas in our proposed scheme, independent streams are transmitted from the different antennas preserving maximum spatial multiplexing gain. Second, in delay diversity schemes, the delays introduced are typically of a symbol duration or longer, whereas, the inter transmitter timing offset here is of a sub-symbol duration. Recent standards such as the IEEE 802.11n as well as 3GPP LTE have included cyclic delay diversity (CDD), as a modification of delay diversity techniques proposed by [99]. These

are typically applied in conjunction with an OFDM scheme and so even though the delays could be a fraction of an OFDM symbol, these techniques are generally presented as a precoding scheme designed to increase the inherent diversity of the channel [25]. In our case, the intent of introducing the offset between the different transmit antennas in a single carrier system is to reduce the inter antenna interference (IAI) and introduce inter symbol interference (ISI) in the modulation while maintaining maximum spatial multiplexing gain.

In MIMO systems with multiple transmit antennas the signals transmitted from the different transmitters interfere with each other at each receive antenna resulting in IAI. In the absence of perfect Nyquist pulse shaping (or due to timing offset), ISI is introduced. Thus, there are two sources of impairment, ISI and IAI, that are distinct and each one leads to a degradation in performance. In traditional aligned systems with Nyquist pulse shaping, there is little to no ISI, but on average, the IAI power is the same as that of the desired signal. In this chapter we show that by offsetting the transmit symbols relative to each other the IAI power can be reduced. In addition, we show that by using a different pulse shape that trades off ISI with the IAI, gains may be achieved practically for free. Although there is a large volume of prior research in the design of quasi zero ISI practical pulse shapes that conform to various criterion such as spectral mask requirements, robustness to timing jitter, peak to average power ratio etc. (e.g., [9, 44, 51, 64, 70, 74] and references therein), to our knowledge, this is the first time that pulses have been designed with this criterion of lowering the IAI.

### 2.1.2 Our Contribution

To summarize, the contributions in this chapter are the following: We demonstrate the practical gains that may be achieved in a single carrier MIMO system by intentionally introducing a sub-symbol delay offset between the transmitted waveforms. We show the performance of zero forcing (ZF), minimum mean squared error (MMSE) and sequence detection based receivers with SRRC pulse shapes and show that the performance is always better than that of the corresponding traditional MIMO system with timing aligned transmission, contrary to

previously published research (for more details, see Section 2.5.1). We also introduce a novel new pulse shape that lowers the energy at half symbol offsets, thus reducing the IAI and improving performance.

This chapter is organized in the following sections. In Sec. 2.3 we present an intuitive rational behind the superior performance of MIMO systems with timing offsets. Then, in Sec. 2.4, we present the analytical system model. In Sec. 2.5 different receiver structures are discussed. A novel pulse shape design criterion is given in Sec. 2.7 following which simulation results are presented in Sec. 2.8 before concluding.

## 2.2 Notation

The notation adopted is as follows: Lower case boldface indicates a vector quantity, as in $\mathbf{a}$. A matrix quantity is indicated by upper-case boldface as in $\mathbf{A}$. Some of the most widely used symbols used throughout this chapter are tabulated below. The rest of the variables will be defined as and when they appear.

**Table 2.1**: Common symbols used

| **Symbol** | **Definition** | **Comments** |
|---|---|---|
| $\beta$ | Excess bandwidth of Nyquist pulses | Real scalar, $0 \leq \beta \leq 1$ |
| $T_s$ | Symbol duration | Real scalar |
| tx | Transmitter | |
| rx | Receiver | |
| $i$ | Time index | Discrete, positive integer |
| $\tau_k$ | Offset of symbol boundaries of tx $k$ relative to tx 0 | Real scalar, $0 \leq \tau_k < T_s$ |
| $M_T$ | Number of transmitters | Real scalar |
| $M_R$ | Number of receivers | Real scalar |

Continued on next page

**Table 2.1 – continued from previous page**

| Symbol | Definition | Comments |
|---|---|---|
| $h_{kj}$ | Complex channel gain between $j$-th tx and $k$-th rx | Complex scalar |
| $\mathbf{H}_k$ | $h_{kj}$ for entry in $j$-th row, $j$-th column. 0 otherwise | Complex diagonal $M_T \times M_T$ matrix |
| $\mathbf{y}_k[i]$ | $i$-th group of $M_T$ outputs at the $k$-th receiver | Complex, $M_T \times 1$ vector |
| $b_k[i]$ | $i$-th transmitted symbol from $k$-th tx | Complex, scalar |
| $\mathbf{n}_k[i]$ | $i$-th noise vector at $k$-th rx | Complex, $M_T \times 1$ vector |
| $\mathbf{R_{xy}}$ | $E[\mathbf{x}\mathbf{y}^H]$, $E[\mathbf{x}] = E[\mathbf{y}] = 0$ | Covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ |
| $E()$ | Expectation operator | n/a |
| $()^H$ | Hermitian operator | n/a |
| $()^T$ | Transpose operator | n/a |
| $I_k$ | Identity matrix of size $k \times k$ | n/a |
| † | Pseudoinverse operator | n/a |
| Quant() | Quantization operator | Thresholds soft decisions to nearest constellation point |

## 2.3 Motivation behind Timing Offset

In traditional single carrier MIMO systems, each receive chain downconverts the received signal to baseband, carries out analog to digital conversion and then employs matched filtering before down sampling the received signal to the system symbol rate. Assuming equal channel gain, the signals from the symbol aligned transmitters contribute equal power to the received signal at the output of the downsampled received matched filter. It may be shown that in a rich scattering environment, the channel gains are statistically independent and thus the

receiver can demodulate the independent streams in either successive interference cancellation mode or in joint detection mode.

In the offset scheme proposed, the symbol boundaries of the different transmitters are offset in time. Thus, when matched filtering is employed at the receiver, under equal gain channel conditions the signals from the two transmitters are not of the same power. This is shown in Fig. 2.1 for rectangular pulse shaping. Indeed the received signal power from the transmitter with the offset symbol is *lower* than that from the transmitter which has its symbol boundaries aligned to that used by the received matched filter. Thus, for the same channel, the offset scheme has lower IAI power in comparison to that in the aligned case.

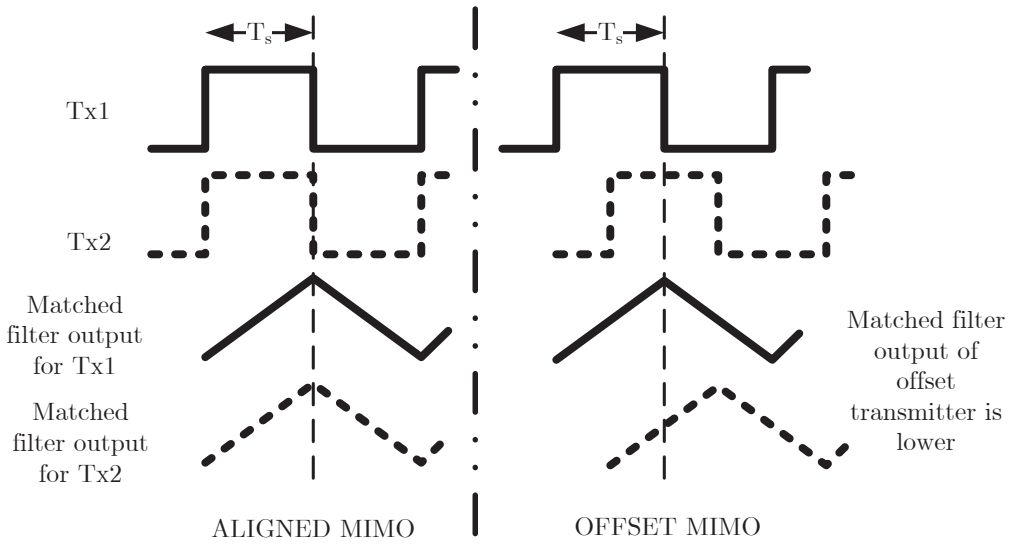


**Figure 2.1**: Reduction of interference power in offset MIMO

The amount of reduction in interference power depends on the pulse shape. While rectangular pulse shaping with half a symbol offset leads to a 3dB reduction in interference power, most practical systems use bandlimited pulse shaping schemes using Nyquist pulse shapes such as the SRRC pulse shape. The interference reduction for various pulse shapes is obtained by sampling the convolution of the two pulses shapes (one at the transmitter and one at the receiver) at the

various offsets. Since it is known that the convolution of two SRRC filters is the raised cosine filter, the IAI power at an offset $\tau_1$ for a SRRC transmit pulse shape with excess bandwidth $\beta$ and symbol duration $T_s$, is given by

$$\mathrm{IAI}(\tau_1) = \sum_{k=-\infty}^{k=\infty} \left( \frac{\sin(\pi(kT_s + \tau_1)/T_s)}{\pi(kT_s + \tau_1)/T_s} \frac{cos(\pi\beta(kT_s + \tau_1)/T_s)}{(1 - (2\beta(kT_s + \tau_1)/T_s)^2} \right)^2 \tag{2.1}$$

and is shown for various offsets and $\beta$ in Fig. 2.2 below. The above formula samples the raised cosine pulse (Eqn 3 of [8]), at symbol intervals as a function of the offset from the symbol boundary, $\tau_1$, and determines the power thus obtained. It may be seen that for a pulse with no excess bandwidth ($\beta = 0$), there is no reduction in interference power, and thus no gains. However as the excess bandwidth increases, the interference power reduces and thus gains increase.

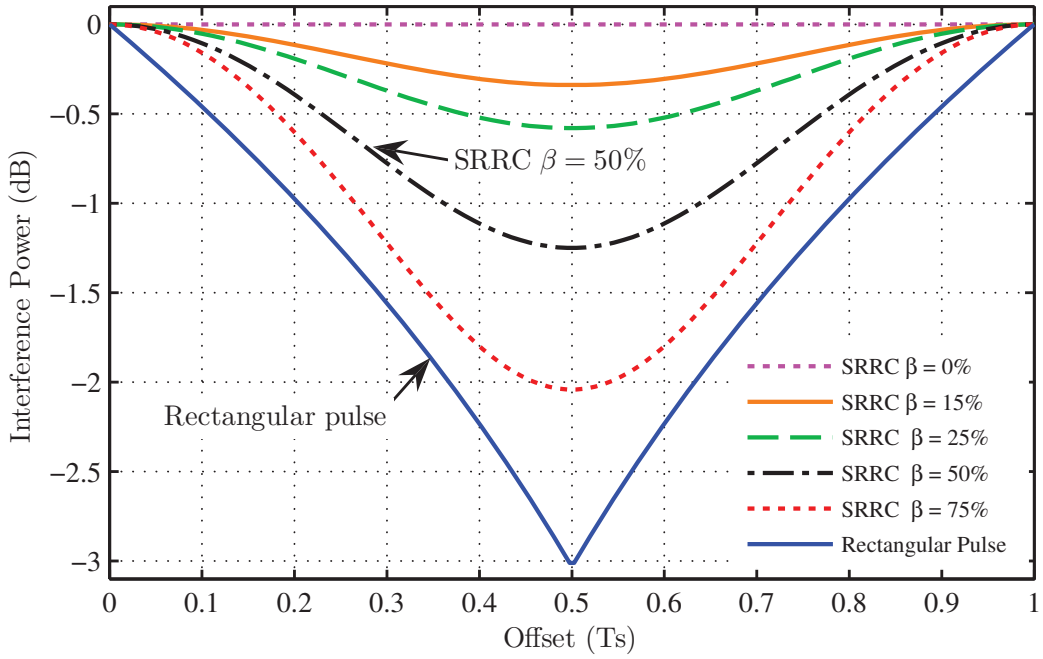


**Figure 2.2**: Interference power for various excess bandwidths and offsets. 0 excess BW leads to no gain

In addition to the lowering of interference power, the system performs better for one more reason. Offsetting the two transmit waveforms relative to each

other introduces ISI thus effectively converting memoryless modulation schemes into those with memory. Consequently, an intelligent receiver can use the ISI to predictively cancel the interference in subsequent symbols thus leading to an even greater suppression of interference.

These two effects combine to provide significant system gains to a MIMO system with intentional timing offset in comparison to an equivalent symbol synchronous MIMO system.

## 2.4 The Timing Offset MIMO System

Fig. 2.3 shows an offset of $\tau_1$ in a particular embodiment of the proposed system with 2 transmit antennas. The symbol duration is denoted by $T_s$ with $0 \leq \tau_1 < T_s$. Other embodiments of the proposed system using $M_T$ antennas would have different $\tau_k$s offsetting the signals from the different transmitters. For simplicity of illustration the transmit signals are depicted with a rectangular pulse shape in Fig. 2.3.

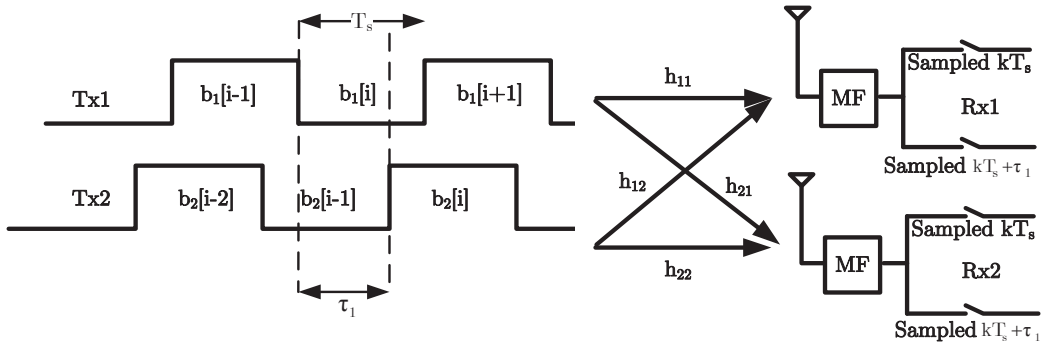


**Figure 2.3**: Sub symbol timing offset: 2 Tx antennas

### 2.4.1 $2 \times 2$ MIMO System with Timing Offset

For simplicity of presentation, a system with 2 antennas at either end, i.e., with $(M_T, M_R) = (2, 2)$, with a rectangular pulse shaping is considered first. The

signals transmitted from the 2nd transmitter is intentionally offset with respect to the first by $\tau_1$. In traditional symbol aligned MIMO the output of the matched filter downsampled to the symbol rate (at the optimal sampling points) are the sufficient statistics for estimating the transmitted symbols. However, in timing offset MIMO, the matched filter output of each receiver is sampled every $T_s$ as well as every $T_s+\tau_1$ thus collecting the output sampled optimally for both transmitters.

Let $h_{kj}$ be the complex path gain from the $j$-th transmitter to the $k$-th receiver. Then stacking the $i$-th output (in time) of the two matched filters, the received vector for each of the receive antennas is given by

$$\mathbf{y}_1[i] = \begin{bmatrix} 0 & 0 \\ h_{11}\rho_{21} & 0 \end{bmatrix} \begin{bmatrix} b_1[i+1] \\ b_2[i+1] \end{bmatrix} + \begin{bmatrix} h_{11} & h_{12}\rho_{12} \\ h_{11}\rho_{12} & h_{12} \end{bmatrix} \times \begin{bmatrix} b_1[i] \\ b_2[i] \end{bmatrix} + \begin{bmatrix} 0 & h_{12}\rho_{21} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1[i-1] \\ b_2[i-1] \end{bmatrix} + \mathbf{n}_1[i] \tag{2.2a}$$

$$\mathbf{y}_2[i] = \begin{bmatrix} 0 & 0 \\ h_{21}\rho_{21} & 0 \end{bmatrix} \begin{bmatrix} b_1[i+1] \\ b_2[i+1] \end{bmatrix} + \begin{bmatrix} h_{21} & h_{22}\rho_{12} \\ h_{21}\rho_{12} & h_{22} \end{bmatrix} \times \begin{bmatrix} b_1[i] \\ b_2[i] \end{bmatrix} + \begin{bmatrix} 0 & h_{22}\rho_{21} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1[i-1] \\ b_2[i-1] \end{bmatrix} + \mathbf{n}_2[i] \tag{2.2b}$$

where $\mathbf{y}_k[i]$ is the $i$-th pair of outputs of the matched filter in the $k$-th receiver, $b_k[i]$ is the $i$-th transmitted symbol from the $k$-th transmitter and $\mathbf{n}_k[i]$ is the additive white Gaussian noise (AWGN) noise vector at the $k$-th receiver. The first row of Eqn. 2.2 is the output of the matched filter sampled at the optimal sampling time for the first transmitter and the second row is the output of the matched filter sampled at the optimal sampling time for the symbols from the second transmitter. The crosscorrelations $\rho_{12}$ and $\rho_{21}$ are a function of the pulse shape and timing offset, with the detailed form given by Eqn. 2.7. For a rectangular pulse, $\rho_{12}$ and $\rho_{21}$ are shown in Fig. 2.4.

It is seen that when the received matched filter is aligned to the first transmitter, the $i$-th symbol of the first transmitter not only interferes with the $i$-th symbol of the second transmitter (as would be the case in standard aligned MIMO

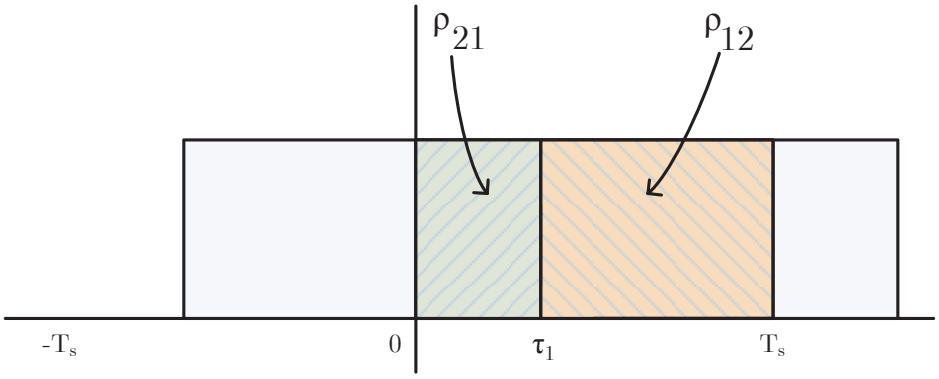


**Figure 2.4**: Cross correlations, $\rho_{12}$ and $\rho_{21}$

architectures), but also interferes with the $(i-1)$-th symbol of the second transmitter. However, the interference power is reduced due to the offset of the transmit pulses from the two transmitters.

Some simple algebraic manipulations of Eqn. 2.2 allow us to write the received samples of receiver $k$ as:

$$\mathbf{y}_k[i] = \begin{bmatrix} 0 & \rho_{21} \\ 0 & 0 \end{bmatrix}^T \begin{bmatrix} h_{k1} & 0 \\ 0 & h_{k2} \end{bmatrix} \begin{bmatrix} b_1[i+1] \\ b_2[i+1] \end{bmatrix} + \mathbf{n}_k[i] + \\ \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} h_{k1} & 0 \\ 0 & h_{k2} \end{bmatrix} \begin{bmatrix} b_1[i] \\ b_2[i] \end{bmatrix} + \begin{bmatrix} 0 & \rho_{21} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} h_{k1} & 0 \\ 0 & h_{k2} \end{bmatrix} \begin{bmatrix} b_1[i-1] \\ b_2[i-1] \end{bmatrix} \tag{2.3}$$

It will be seen later that Eqn. 2.3 is a special case of the more general formula derived for any arbitrary number of transmitters in Eqn. 2.6. The above equations for $\mathbf{y}_1[i]$ and $\mathbf{y}_2[i]$ may be combined and written more compactly in the following matrix format:

$$\mathbf{r}[i] = \begin{bmatrix} \mathbf{y}_1[i] \\ \mathbf{y}_2[i] \end{bmatrix} = \mathbf{P}_1\mathbf{b}[i+1] + \mathbf{P}_0\mathbf{b}[i] + \mathbf{P}_{-1}\mathbf{b}[i-1] + \mathbf{n}[i] \tag{2.4}$$

To elucidate further, $\mathbf{P}_1$, $\mathbf{P}_0$, and $\mathbf{P}_{-1}$ are all $4 \times 2$ matrices, $\mathbf{b}[i]$ is a $2 \times 1$ vector and $\mathbf{n}[i]$ and $\mathbf{r}[i]$ are both $4 \times 1$ vectors.

When practical pulse shapes of longer duration such as the SRRC pulse shaping is used, then the interference from the offset is not limited to the adjacent symbols but depends on the length of the filter used. Although the ideal SRRC pulse is infinite in duration, all practical schemes use finite length pulse shapes. This may be seen in Fig. 2.5 where a 11 symbol long raised cosine pulse shape is shown which is non-zero from $-5T_s$ to $5T_s$. In this case, in an offset transmission scheme, the interference arises from 10 symbols as shown in Fig. 2.5.

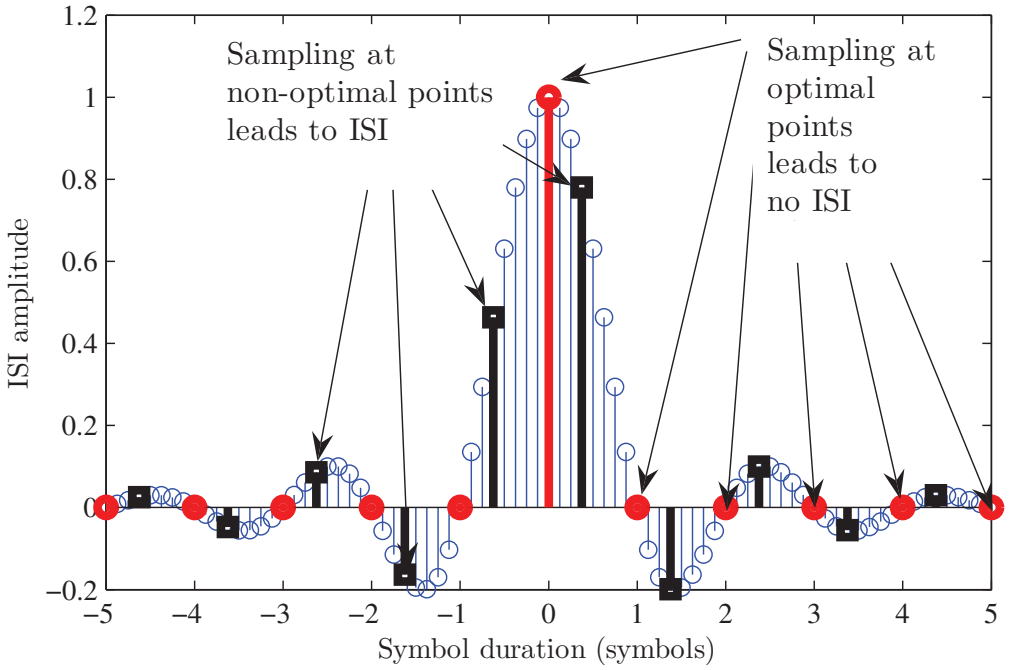


**Figure 2.5**: Raised cosine pulse: Impact of sampling on ISI

In this case, the expressions equivalent to Eqn. 2.3 get more complex. Let $d(t)$ denote the continuous time convolution of the pulse shapes at the receiver and at the transmitter. $d(t)$ is assumed to be $2L$ symbols in duration and thus is equal to zero for time, $t$, outside the interval $[-LT_s, LT_s]$.

As is the case for most practical pulse shapes, it is assumed that $d(t)$ is symmetric such that $d(-t) = d(t)$. Analogous to Eqn. 2.3, the received samples

at the k-th receiver matched to both the first and the second transmitter may be expressed as:

$$\begin{aligned}\mathbf{y}_k[i] = &\sum_{l=0}^{L} \begin{bmatrix} d(lT_s) & d(lT_s-\tau_1) \\ d(lT_s+\tau_1) & d(lT_s) \end{bmatrix}^T \begin{bmatrix} h_{k1} & 0 \\ 0 & h_{k2} \end{bmatrix} \begin{bmatrix} b_1[i+l] \\ b_2[i+l] \end{bmatrix} \\ &+ \sum_{l=1}^{L} \begin{bmatrix} d(lT_s) & d(lT_s-\tau_1) \\ d(lT_s+\tau_1) & d(lT_s) \end{bmatrix} \begin{bmatrix} h_{k1} & 0 \\ 0 & h_{k2} \end{bmatrix} \begin{bmatrix} b_1[i-l] \\ b_2[i-l] \end{bmatrix} + \mathbf{n}_k[i]\end{aligned} \tag{2.5}$$

### 2.4.2 $M_T \times M_R$ MIMO System with Timing Offset

The more general case with $M_T$ transmitters and $M_R$ receivers is now considered. In this setup, the relative timing offset between the first transmitter and k-th transmitter is $\tau_k$. Without loss of any generality, it is assumed that $0 = \tau_0 \leq \tau_1 \leq \tau_2 ... \leq \tau_{M_T-1} < T_s$ where $T_s$ is the symbol duration. Each receiver conceptually has $M_T$ matched filters, each one matched to the optimal sampling times of one of the transmitters (but in reality, would be implemented as a single matched filter sampled $M_T$ times a symbol). It should be mentioned that for excess bandwidth $0 \leq \beta \leq 1$, sampling each matched filter at 2 samples per symbol meets the Nyquist sampling criterion and thus an intelligent receiver should be able to operate with the 2samples/symbol out of the matched filter. In this analysis, we sample the output of the matched filer at $M_T$ samples per symbol only to keep the receiver structure conceptually simple.

For systems using pulse shapes $s_l(t)$ at the l-th transmitter such as the rectangular pulse that is zero outside $t \in [0, T_s]$ it may be shown that the samples received at the k-th receiver is a $M_T \times 1$ vector, $\mathbf{y}_k[i]$, that may be expressed as:

$$\mathbf{y}_k[i] = (\mathbf{R_1})^T \mathbf{H}_k \mathbf{b}[i+1] + \mathbf{R_0} \mathbf{H}_k \mathbf{b}[i] + \mathbf{R_1} \mathbf{H}_k \mathbf{b}[i-1] + \mathbf{n}_k[i] \tag{2.6}$$

where the $M_T \times M_T$ matrix $\mathbf{H}_k = \text{diag}(h_{k1}, h_{k2}, h_{k3} .... h_{kM_T})$ and the correlations $\rho_{kl}$ and $\rho_{lk}$ are given by:

$$\rho_{kl} = \int_{\tau}^{T_s} s_k(t) s_l(t-\tau) dt \tag{2.7a}$$

$$\rho_{lk} = \int_{0}^{\tau} s_k(t) s_l(t+T_s-\tau) dt \tag{2.7b}$$

The entries in the j-th row, k-th column of the $M_T \times M_T$ matrices, $\mathbf{R_0}$ and $\mathbf{R_1}$ are respectively given by

$$\mathbf{R_0}[j,k] = \begin{cases} 1, & if\ j = k \\ \rho_{jk}, & if\ j < k \\ \rho_{kj}, & if\ j > k \end{cases} \tag{2.8a}$$

$$\mathbf{R_1}[j,k] = \begin{cases} 0, & if\ j \geq k \\ \rho_{kj}, & if\ j < k \end{cases} \tag{2.8b}$$

It can be seen that Eqn. 2.3 is a special case of Eqn. 2.6 for $M_T = 2$. The zero-mean Gaussian noise process $\mathbf{n}_k[i]$ has the following autocorrelation matrix where $\sigma^2$ denotes the noise variance:

$$E[\mathbf{n_k}[i]\mathbf{n_l}^H[j]] = \sigma^2(\mathbf{R_1})^T, \qquad \text{if } j = i+1,\ k = l \tag{2.9a}$$

$$= \sigma^2(\mathbf{R_0})^T, \qquad \text{if } j = i,\ k = l \tag{2.9b}$$

$$= \sigma^2\mathbf{R_1}, \qquad \text{if } j = i-1,\ k = l \tag{2.9c}$$

$$= 0, \qquad \text{otherwise} \tag{2.9d}$$

It is noted that the expressions above are very similar to those in the derivation of the multiuser discrete time asynchronous model developed in Section 2.10 of [91]. Although the notation has been chosen to be consistent with [91], the application space is quite different. We also note that comparing Eqn. 2.6 with Eqn. 14 of [83], it may be concluded that the received samples are identical in both our model, and in the case of offset MIMO presented by Shao et al. This has been shown by us in more details in [33].

The derivations above can be extended for use with practical pulse shapes that extend beyond $t \in [0, T_s]$. Analogous to the derivation of Eqn. 2.5, Eqn. 2.6

can also be extended to the case where the convolution of the pulse shape at the transmit and the receive side ($d(t)$) is non zero for $t \in [-LT_s, LT_s]$ and is assumed to be zero for $t$ outside this interval. In that case, the received samples at the k-th receiver can be written as:

$$\mathbf{y}_k[i] = \sum_{l=0}^{L} (\mathbf{R_l})^T \mathbf{H}_k \mathbf{b}[i+l] + \sum_{l=1}^{L} \mathbf{R_l} \mathbf{H}_k \mathbf{b}[i-l] + \mathbf{n}_k[i] \tag{2.10}$$

where, like before, $\mathbf{H}_k$ is a $M_T \times M_T$ diagonal matrix given by $\mathbf{H}_k = \text{diag}(h_{k1}, h_{k2}, h_{k3}....h_{kM_T})$ and the i-th row, j-th column entry of the $M_T \times M_T$ matrix, $\mathbf{R_l}$ is given by

$$\mathbf{R_l}(i,j) = \begin{cases} d(lT_s - (\tau_i - \tau_j)), & \text{if } i < j \text{ where } i,j = [0,1,...(M_T-1)] \\ d(lT_s - (\tau_i - \tau_j)), & \text{if } i \geq j \text{ where } i,j = [0,1,...(M_T-1)] \end{cases} \tag{2.11}$$

## 2.5 Receiver Design

In this section, we develop 3 different forms of receivers for the proposed system: i) Zero forcing (ZF) receivers, ii) minimum mean squared error (MMSE) receivers and iii) Viterbi algorithm based sequence detection receivers.

All the receivers assume memoryless linear modulations such as M-ary phase shift keying (M-PSK) or M-ary quadrature amplitude modulation (M-QAM) with a block transmission scheme as shown in Fig. 2.6. It is assumed that there is no inter block interference (IBI). This condition can be satisfied by inserting an appropriate amount of idle time between the transmission of two blocks as shown in Fig. 2.6. Each block is assumed to contain S symbols long. As S increases, the overhead due to the inter block gap decreases. The transmitted symbols are assumed to be zero mean, unit energy and uncorrelated in time and space. It is assumed that the channel is flat fading and unchanged over the duration of the entire block and independent from block to block and that the channel is known perfectly at the receiver. The noise is assumed to be Gaussian and independent of the data symbols. Two different noise models are used below – the first where the

noise is spatially uncorrelated and the second where the noise has mutual coupling between the receivers.

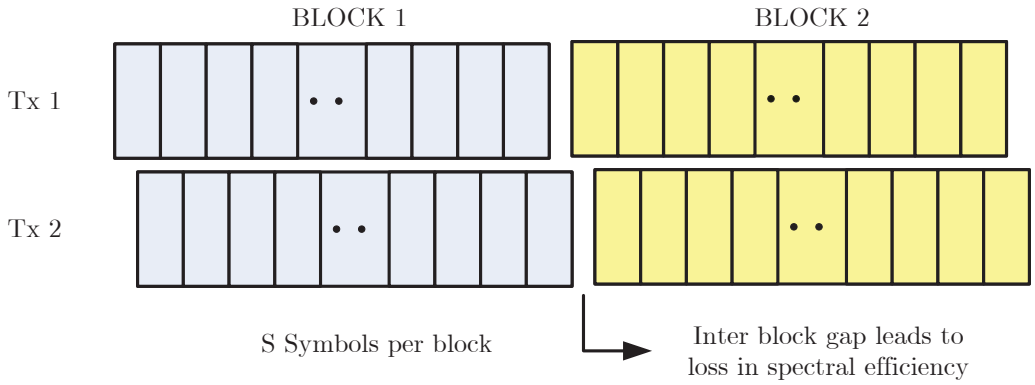


**Figure 2.6**: Block transmission scheme

### 2.5.1 ZF Receivers

In [83], the authors present a zero forcing (ZF) receiver whose performance is strongly dependent on the blocksize, $S$. They conclude that for large block sizes the performance of the offset transmission scheme is *worse* than that of the traditional MIMO schemes and thus, the offset scheme should be used only for very short block sizes. In their work the block sizes are typically 2, 4 or 10 symbols. This is a very severe restriction as such short block sizes lead to significant spectral efficiency reductions. With a block size of 2 symbols with 2 transmit antennas and offset $\tau_1 = 0.6T_s$, the system has a spectral efficiency that is 23% less than that of synchronized systems and with a block size of 10 symbols, the spectral efficiency is reduced by 5.7%. This reduction in spectral efficiency makes the offset MIMO scheme, proposed in [83], of limited use in practical systems.

A closer examination of the ZF receiver proposed by Shao et al. showed that it was not the optimal ZF receiver. This was first shown by us in [33]. The authors of [83] had mistakenly chosen a formulation that suffered a lot of noise enhancement as the block size, $S$, grew larger. The ZF receiver from [83] is given

by

$$\widehat{\mathbf{b}} = \left( \sum_{k=1}^{M_R} \mathbf{H}_k^H \mathbf{H}_k \right)^{-1} \sum_{k=1}^{M_R} \mathbf{H}_k^H \mathbf{R}^{-1} \mathbf{y}_k \tag{2.12}$$

where $\mathbf{y}_k$ and $\mathbf{H}_k$ are given by Eqn. 2.10 and $\mathbf{R}$ is a $SM_T \times SM_T$ matrix given by Eqn. 2.15. Utilizing this receiver, the authors show a degradation of about 5dB when the block size is increased from 2 to 10 (Fig. 6 of [83]). The offset VBLAST receiver outperforms the traditional scheme only for very small block sizes.

A problem with the above zero forcing receiver is that as the block size, $S$, increases the $SM_T \times SM_T$ sized correlation matrix, $\mathbf{R}$, gets progressively closer to being singular. This may be seen by observing from Eqn. 2.15 that if the block size increases from $S$ to $S+1$, the correlation matrices for the two cases, $\mathbf{R}_S$ and $\mathbf{R}_{S+1}$, respectively, are related to each other by the following equation where $\mathbf{R}_1$ and $\mathbf{R}_1$ are given by Eqn. 2.8:

$$\mathbf{R}_{S+1} = \begin{bmatrix} & & 0 \\ & \mathbf{R}_S & 0 \\ & & \mathbf{R}_1^T \\ 0 & \mathbf{R}_1 & \mathbf{R}_1 \end{bmatrix} \tag{2.13}$$

Thus, as the block size increases, the correlation matrix $\mathbf{R}$ also increases in size by adding $M_T$ columns and $M_T$ rows to the correlation matrix for the smaller block size. It was shown in [11] that the condition number of correlation matrices with such an embedding property keeps growing with matrix size. Fig. 2.7 shows the rapid increase in condition number for the correlation matrix with 2 transmit antennas as the block size increases. This ill-conditioning of $\mathbf{R}$ has a severe negative effect on the performance of ZF receivers proposed in [83] by Shao et al. Consequently, the performance of the ZF receiver proposed degrades when used with long block sizes. In contrast, Fig. 2.7 shows that the condition number of $\mathbf{H}_{tot}^H \mathbf{R}_{tot} \mathbf{H}_{tot}$ does not keep increasing with block size. We define these variables below where we show that the numerical stability of matrix $\mathbf{H}_{tot}^H \mathbf{R}_{tot} \mathbf{H}_{tot}$ shall be crucial to determining the optimal ZF receiver since the optimal ZF receiver takes an inverse of this quantity.

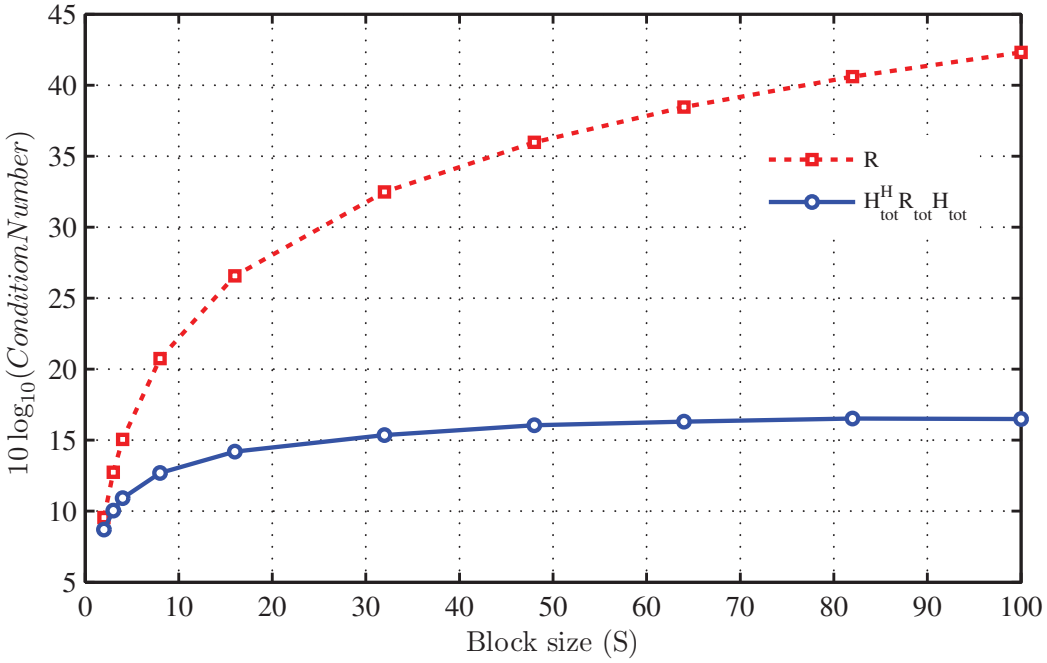


**Figure 2.7**: Condition number changes with block size

### Optimal ZF receiver

The optimal ZF receiver shall use all the received samples for the entire block of data simultaneously to make its decision. To obtain the optimal ZF receiver, we first stack all the outputs of each block for the k-th receiver from Eqn. 2.6 to obtain

$$\mathbf{z}_k = \mathbf{R}\widetilde{\mathbf{H}}_k\mathbf{b}_{\mathbf{block}} + \mathbf{n}_k \tag{2.14}$$

where $\mathbf{z}_k = [\mathbf{y}_k^T(0), \mathbf{y}_k^T(1), \mathbf{y}_k^T(2)...\mathbf{y}_k^T(S-1)]^T$, the transmitted symbols, $\mathbf{b}_{\mathbf{block}} = [\mathbf{b}^T(0), \mathbf{b}^T(1), ...\mathbf{b}^T(S-1)]^T$ and $\widetilde{\mathbf{H}}_k = \text{diag}(\mathbf{H}_k, \mathbf{H}_k, \mathbf{H}_k...)$. $\mathbf{y}_k(i)$ and $\mathbf{b}(i)$, both $M_T \times 1$ vectors, represent the received samples matched to each transmitter received at receiver $k$ at time $i$ and the transmitted symbols from all transmitters at time $i$ respectively. $\mathbf{H}_k$ is a diagonal matrix of channel gains of size $M_T \times M_T$. Thus in Eqn. 2.14, $\mathbf{z_k}$ is a $SM_T \times 1$ vector of all received samples in a block of S transmitted symbols per transmit antenna at receiver $k$. $\mathbf{b_{block}}$ is the $SM_T \times 1$ vector of all transmitted symbols in that block, $\widetilde{\mathbf{H}}_k$ is a diagonal matrix of $SM_T \times SM_T$ elements of channel gains from the transmitters to the $k$-th receiver (assumed con-

stant over the block). $\mathbf{R}$ is a $SM_T \times SM_T$ real symmetric correlation matrix given by Eqn. 2.15 where $\mathbf{R_0}$ and $\mathbf{R_1}$ are given by Eqn. 2.8

$$\mathbf{R} = \begin{bmatrix} \mathbf{R_0} & \mathbf{R_1}^T & 0 & ... & ... & 0 \\ \mathbf{R_1} & \mathbf{R_0} & \mathbf{R_1}^T & 0 & ... & 0 \\ 0 & \mathbf{R_1} & \mathbf{R_0} & \mathbf{R_1}^T & 0 & ... \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & 0 & \mathbf{R_1} & \mathbf{R_0} & \mathbf{R_1}^T \\ 0 & ... & ... & 0 & \mathbf{R_1} & \mathbf{R_0} \end{bmatrix} \tag{2.15}$$

Then all the $\mathbf{z_k}$ outputs of each receiver is stacked in the following manner:

$$\begin{bmatrix} \mathbf{z_1} \\ \mathbf{z_2} \\ .. \\ \mathbf{z}_{M_R} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & 0 & .. & 0 \\ 0 & \mathbf{R} & 0 & .. \\ .. & .. & .. & .. \\ 0 & .. & 0 & \mathbf{R} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{H}}_1 \\ \widetilde{\mathbf{H}}_2 \\ .. \\ \widetilde{\mathbf{H}}_{M_R} \end{bmatrix} \mathbf{b_{block}} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ .. \\ \mathbf{n}_{M_R} \end{bmatrix} \tag{2.16a}$$

$$\mathbf{z}_{tot} = \mathbf{R}_{tot}\mathbf{H}_{tot}\mathbf{b_{block}} + \mathbf{n}_{tot} \tag{2.16b}$$

and the optimum ZF receiver is given by

$$\hat{\mathbf{b}}_{ZFopt} = (\mathbf{H}_{tot}^H \mathbf{R}_{tot} \mathbf{H}_{tot})^{-1} \mathbf{H}_{tot}^H \mathbf{z}_{tot} \tag{2.17}$$

The above optimal ZF receiver not only cancels all the interference, but it minimizes the output noise variance. It can be readily derived by noting that the optimal ZF receiver is the well known best linear unbiased estimator (BLUE) (Ch. 6, [59]). This can be seen by noting that in the BLUE estimation, we seek an unbiased estimator which minimizes the estimator variances. The unbiased criterion ensures cancellation of interference while minimizing variance corresponds to maximizing signal to noise ratio.

It should be pointed out that the optimal ZF receiver is a batch receiver, i.e. it works on the received samples from the entire block at the same time. This increases complexity and introduces latency in the system (since the first transmitted symbols can only be decoded after the samples corresponding to the last transmitted symbol in the block have been received). The above receiver also

needs to calculate the pseudoinverse of a $SM_T \times SM_T$ matrix. The block sizes of practical systems often consists of hundreds (sometimes thousands) of symbols and thus the complexity of this step is non trivial and indeed could be impractical with current hardware.

In Sec. 2.8.5 we plot the performance of the optimal ZF receiver developed here and compare the performance to that in [83]. Fig. 2.17 demonstrates that the optimal ZF receiver does not suffer any significant performance degradation when the block size is increased.

### 2.5.2 MMSE Receivers

The linear MMSE (LMMSE) receiver is known [74] to outperform the ZF receiver and is considered in this section. The LMMSE estimate of $\mathbf{b}[i]$, given observation $\mathbf{r}[i]$, is given by $\mathbf{R}_{\mathbf{b}[i]\mathbf{r}[i]}\mathbf{R}^{\dagger}_{\mathbf{r}[i]\mathbf{r}[i]}\mathbf{r}[i]$ where $\dagger$ indicates the pseudoinverse, and $\mathbf{R}_{\mathbf{b}[i]\mathbf{r}[i]} = E[\mathbf{b}[i]\mathbf{r}[i]^H]$ and $\mathbf{R}_{\mathbf{r}[i]\mathbf{r}[i]} = E[\mathbf{r}[i]\mathbf{r}[i]^H]$ [59]. It is known that for Gaussian noise, the MMSE solution and the LMMSE solution are the same and so the terms are used interchangeably here. For notational convenience, the dependence on the time index, $i$, shall also be dropped for clarity.

Two classes of MMSE receivers are analyzed. The first class carries out joint detection of the symbols while the second carries out layered interference cancellation. For both these receiver types, one-shot receivers (i.e. those that estimate $\mathbf{b}[i]$, given $\mathbf{r}[i]$) and windowed receivers (i.e. those that estimate $\mathbf{b}[i]$ given $\mathbf{r}[i-W], ..\mathbf{r}[i], ..\mathbf{r}[i+W]$, thus implying a window length of $2W+1$ observations) are developed. We shall also develop a MMSE joint batch receiver, i.e. one that estimates all the transmitted symbols of the block, using all the received samples in that block.

#### One Shot LMMSE Receiver, ($W = 0$)

In this scenario, the observations, $\mathbf{r}[i]$, are given by Eqn. 2.4 and only one measurement vector is used to estimate the corresponding information carrying symbols. It is assumed that: a) $\mathbf{b}[i]$s are zero mean, unit energy and uncorrelated in time, b) $h_{kj}$s, the channel gains, are perfectly known at receiver and do not

change over the duration of a block of data, and c) the additive Gaussian noise is spatially uncorrelated and also uncorrelated with the information carrying signal. Under these assumptions, from Eqn. 2.4 we have

$$\mathbf{R_{rr}} = \sum_{i=-1}^{1} \mathbf{P}_i \mathbf{P}_i^H + \mathbf{R_{nn}} \tag{2.18a}$$

$$\mathbf{R_{br}} = \mathbf{Q}^H \tag{2.18b}$$

In the symbol aligned $2 \times 2$ model (traditional MIMO), $\mathbf{R_{nn}}$, the noise covariance matrix, is often modeled as $2 \times 2$ identity matrix scaled with the noise variance $\sigma^2$. This simple model assumes that the noise variance, $\sigma^2$ is the same for both the receive antennas and that there is no noise coupling between the antennas. In offset MIMO we have 2 sets of matched filters per receiver and so $\mathbf{R_{nn}}$ is a $4 \times 4$ matrix. By observing that the continuous time AWGN noise is zero mean and independent between the two receivers and by noting that part of the integration period for each symbol is the same between the two matched filters in the same receiver, it may be shown that $\mathbf{R_{nn}}$ for this noise model is no longer a scaled identity matrix, but is given by Eqn. 2.19 where $\sigma^2$ is the noise variance and $\rho_{12}$ is given by Eqn. 2.8.

$$\mathbf{R_{nn}} = \begin{bmatrix} \sigma^2 & \rho_{12}\sigma^2 & 0 & 0 \\ \rho_{12}\sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \rho_{12}\sigma^2 \\ 0 & 0 & \rho_{12}\sigma^2 & \sigma^2 \end{bmatrix} \tag{2.19}$$

In the more general case where the noise is not assumed to be independent between the two antenna, the noise covariance matrix in the traditional symbol aligned $2 \times 2$ system is given by:

$$\mathbf{R_{nn\text{-aligned}}} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix} \tag{2.20}$$

where $\sigma_{11}^2$ and $\sigma_{22}^2$ are, respectively, the noise variances of the 1st receive antenna and the 2nd receive antenna. $\sigma_{12}^2$ and $\sigma_{21}^2$ are, respectively, the covariance

of the noise on the first receive antenna with that of the 2nd receive antenna and vice-versa. In all these cases, the noise is assumed to be zero mean.

In this model for the noise, Eqn. 2.19 can also be more generalized and is determined to be:

$$\mathbf{R_{nn}} = \begin{bmatrix} \sigma_{11}^2 & \rho_{12}\sigma_{11}^2 & \sigma_{12}^2 & \rho_{12}\sigma_{12}^2 \\ \rho_{12}\sigma_{11}^2 & \sigma_{11}^2 & \rho_{12}\sigma_{12}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \rho_{12}\sigma_{21}^2 & \sigma_{22}^2 & \rho_{12}\sigma_{22}^2 \\ \rho_{12}\sigma_{21}^2 & \sigma_{21}^2 & \rho_{12}\sigma_{22}^2 & \sigma_{22}^2 \end{bmatrix} \tag{2.21}$$

Using Eqns. 2.18 and 2.19 or 2.21, the transmitted symbols are thus estimated at the receiver to be

$$\widehat{\mathbf{b}[i]} = \text{Quant}\left(\mathbf{R_{b[}}_{i\mathbf{]r}}\mathbf{R_{rr}}^{\dagger}\mathbf{r}[i]\right) \tag{2.22}$$

where $\mathbf{r}[i]$ is a vector of all observations being used for the estimate of $\mathbf{b}[i]$ and the Quant() function is used to make hard decisions on the processed samples i.e., it maps $\mathbf{R_{b[i]r}}\mathbf{R_{rr}^{\dagger}}(\mathbf{r}[i]$ to the nearest transmit constellation point.

**Adjacent symbol LMMSE Receiver, ($W = 1$)**

From the observation model, it is clear that because of correlation between adjacent measurements, a LMMSE receiver that estimates the information symbols using measurements that span more than one symbol duration can lead to improvements. In this section the adjacent symbol LMMSE receiver that utilizes three received vectors to decide $\mathbf{b}[i]$ will be considered. Using Eqn. 2.4, the received vectors used to determine $\mathbf{b}[i]$ are:

$$\mathbf{r}[i-1] = \mathbf{P}_1\mathbf{b}[i] + \mathbf{P}_0\mathbf{b}[i-1] + \mathbf{P}_{-1}\mathbf{b}[i-2] + \mathbf{n}[i-1] \tag{2.23a}$$

$$\mathbf{r}[i] = \mathbf{P}_1\mathbf{b}[i+1] + \mathbf{P}_0\mathbf{b}[i] + \mathbf{P}_{-1}\mathbf{b}[i-1] + \mathbf{n}[i] \tag{2.23b}$$

$$\mathbf{r}[i+1] = \mathbf{P}_1\mathbf{b}[i+2] + \mathbf{P}_0\mathbf{b}[i+1] + \mathbf{P}_{-1}\mathbf{b}[i] + \mathbf{n}[i+1] \tag{2.23c}$$

These three equations may be stacked and expressed more compactly as

$$\mathbf{y}[i] = \begin{bmatrix} \mathbf{P}_{-1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{b}[i-2] + \begin{bmatrix} \mathbf{P}_{0} \\ \mathbf{P}_{-1} \\ \mathbf{0} \end{bmatrix} \mathbf{b}[i-1] + \begin{bmatrix} \mathbf{P}_{1} \\ \mathbf{P}_{0} \\ \mathbf{P}_{-1} \end{bmatrix} \mathbf{b}[i]$$
$$+ \begin{bmatrix} \mathbf{0} \\ \mathbf{P}_{1} \\ \mathbf{P}_{0} \end{bmatrix} \mathbf{b}[i+1] + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{P}_{1} \end{bmatrix} \mathbf{b}[i+2] + \widetilde{\mathbf{n}}[i] \tag{2.24a}$$
$$= \mathbf{M}_{-2}\mathbf{b}[i-2] + \mathbf{M}_{-1}\mathbf{b}[i-1] + \mathbf{M}_{0}\mathbf{b}[i] + \mathbf{M}_{1}\mathbf{b}[i+1]$$
$$+ \mathbf{M}_{2}\mathbf{b}[i+2] + \widetilde{\mathbf{n}}[i] \tag{2.24b}$$

Note that $\mathbf{y}[i]$ and $\widetilde{\mathbf{n}}[i]$ are $12 \times 1$ vectors, each $\mathbf{M}_{-2}$, $\mathbf{M}_{-1}$, ..., $\mathbf{M}_{2}$ are $12 \times 2$ matrices that are used for notational convenience, and $\mathbf{b}[i]$ is a $2 \times 1$ vector. Thus the LMMSE receiver is given by

$$\widehat{\mathbf{b}}[i] = \text{Quant}\left(\mathbf{R}_{\mathbf{b}[i]\mathbf{y}}\mathbf{R}_{\mathbf{yy}}^{\dagger}\mathbf{y}[i]\right) \tag{2.25a}$$
$$= \text{Quant}\left(\mathbf{M}_{0}^{H}(\sum_{k=-2}^{2}\mathbf{M}_{k}\mathbf{M}_{k}^{H} + \mathbf{R}_{\widetilde{\mathbf{n}}\widetilde{\mathbf{n}}})^{\dagger}\mathbf{y}[i]\right) \tag{2.25b}$$

In this context, the covariance matrix of the noise vector $\widetilde{\mathbf{n}}[i]$ given by $\mathbf{R}_{\widetilde{\mathbf{n}}\widetilde{\mathbf{n}}}$ is a matrix with similar structure as in Eqn. 2.19 or Eqn. 2.21 except that it is a $12 \times 12$ matrix. This approach can be extended to more general receivers using a wider window of received samples to estimate the $i$-th transmitted symbol.

### MMSE Joint Batch Receivers

The above two MMSE receivers estimated the transmitted symbol vectors one at a time, i.e., $\mathbf{b}[0]$ is estimated, then $\mathbf{b}[1]$ is estimated and so on until all the transmitted symbols of the block are estimated. In this section, we present the joint batch MMSE receiver. This receiver estimates all the transmitted symbols of the block $\mathbf{b_{block}}$ based on all the received samples from that block, $\mathbf{z_{tot}}$, as shown in Eqn. 2.16.

Similar to the subsections above, the optimal estimate is derived below as:

$$\hat{\mathbf{b}}_{MMSE-block} = \text{Quant}\left[E\left[\mathbf{b_{block}}\mathbf{z_{tot}}^H\right]\left(E\left[\mathbf{z_{tot}}\mathbf{z_{tot}}^H\right]\right)^\dagger \mathbf{z_{tot}}\right] \quad (2.26a)$$

$$= \text{Quant}\left[\mathbf{H}_{tot}^H\mathbf{R}_{tot}^H\left(\mathbf{H}_{tot}\mathbf{R}_{tot}\mathbf{H}_{tot}^H\mathbf{R}_{tot}^H + \mathbf{R}_{\mathbf{n}_{tot}\mathbf{n}_{tot}}\right)^\dagger \mathbf{z_{tot}}\right] \quad (2.26b)$$

As discussed in Sec. 2.5.1, these batch receivers are significantly more complicated to implement and require taking the inverse of matrices of size $SM_T \times SM_T$. They also add latency to the system and are included here for the sake of completion.

**MMSE Receivers with Layered Detection and Interference Cancellation**

The receivers discussed above carry out joint decoding of symbols transmitted from the two transmitters. However, receivers using the vertical Bell labs layered space time (VBLAST) technique [100] where one transmitter is decoded (using a LMMSE receiver) and then the decoded symbols are used to carry out interference cancellation were also designed. As shown in [71] and [100], the layered approach achieves superior performance in the traditional symbol-aligned case and here it is expected that the layered detection will also improve performance in the proposed offset scheme.

It is well known (see e.g., [100], [14], [71]) that optimal ordering of the decoding layers leads to performance improvements. As [100] has shown, decoding the layer with the highest SINR (or the lowest error variance) yields the optimal ordering.

Using Eqn. 2.18, in the case of the one shot ($W = 0$) offset MIMO system, the error covariance matrix may be expressed as:

$$E[(\mathbf{b} - \hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})^H] = \mathbf{R}_{bb} - \mathbf{R}_{br}\mathbf{R}_{rr}^\dagger\mathbf{R}_{rb} \quad (2.27a)$$

$$= \mathbf{I}_2 - \mathbf{P}_0^H\left(\sum_{i=-1}^{1}\mathbf{P}_i\mathbf{P}_i^H + \mathbf{R}_{\mathbf{nn}}\right)^\dagger\mathbf{P}_0 \quad (2.27b)$$

Thus the error variance of decoding the symbol from the first transmitter is given by the magnitude of the (1,1) element and the error variance of decoding the symbol from the second transmitter is given by the magnitude of the (2,2) element

of the $2 \times 2$ error covariance matrix. The layer that has the lower error variance (and hence higher SINR) is decoded first.

### 2.5.3 Viterbi Algorithm based Receivers

Since ISI is inherently present in the proposed offset system, the optimal receiver is the maximum likelihood sequence detector (MLSD). The Viterbi algorithm [46] is a very well known algorithm for implementing the MLSD in a computationally tractable manner. As shown in [58] and implied by Section II of [46], the usual implementation of the Viterbi algorithm yields the MLSD only if the noise is memoryless and is independent from sample to sample. In our case, however, this is not true as the noise has temporal correlation as indicated by Eqn. 2.9.

In order to reduce the impact of the temporal noise correlation, we carried out noise whitening over different observation windows. Thus the Viterbi algorithm was run not on the received samples, but on $\mathbf{Rnn}^{-1/2}\mathbf{y}[i]$ where $\mathbf{Rnn}$ denotes the covariance of the noise vector and $\mathbf{y}[i]$ denotes the received vector as given by Eqn. 2.4 for the one shot case and by Eqn. 2.24 for the windowed case. Although this method whitens the noise locally, it does not whiten the noise over the entire received burst and thus is an approximation to the ML solution.

#### Rectangular Pulse

A cursory examination of Eqn. 2.4 reveals a channel memory of 3 symbols and with BPSK signaling with 2 transmit antenna this leads to a total of $(2^2)^3 = 64$ states in the trellis. However, a more careful inspection using the structure of matrices $\mathbf{P}_1$ and $\mathbf{P}_{-1}$ from Eqn. 2.4, indicates that the channel memory can be reduced to 4 bits and this results in 16 states as shown in Fig. 2.8.

#### Raised Cosine Pulse

When the SRRC pulse shape is employed the channel memory depends on the length of the filters employed. Our simulations employed a SRRC filter of

length 21 symbols with 25% excess bandwidth and thus the ISI extends over 20 symbol durations. This causes the trellis to grow unacceptably large for implementation purposes. The optimal trellis for a pulse with L symbol ISI and for a system using $M_T$ transmitters and an $M$-ary constellation is $(M^{M_T})^L$ long. This is usually impractical to implement and so sub-optimal trellis decoders are often employed. In our simulations, we have opted for a sub-optimal solution that uses a very similar 16 state trellis as is used for the rectangular pulse and pretends that the ISI is only from the adjacent symbols and ignores the ISI from the other interfering symbols. This is clearly sub-optimal. However, since most of the interference power comes from the adjacent symbols, this sub-optimal receiver captures most of the performance gain and the improvements by going to more complex receivers are likely to be marginal. In passing we note that the conventional scheme does not have ISI and so sequence detection does not improve its performance.

The 16 state Viterbi Trellis used for the sequence detection receivers is shown in Fig. 2.8.

## 2.6 Optimizing the timing offset, $\tau_1$

Thus far, we have presented the MAESTRO scheme for a $(M_T, M_R) = (2, 2)$ with an arbitrary offset of $\tau_1$ between the first and second transmitters. In this section, we shall present a scheme to optimize the value of $\tau_1$ for best error rate performance.

The error covariance matrix of MMSE receivers is well known (see [59]) and is given by Eqn. 2.27. This error covariance can be easily simulated as a function of the timing offset $\tau_1$ to determine the optimal $\tau_1$ at which the error variance is minimized. Below we present an analytical derivation of the optimal $\tau_1$ for a $(M_T, M_R) = (2, 1)$ system.

### 2.6.1 Optimal $\tau_1$ for $2 \times 1$ system

For a $(M_T, M_R) = (2, 1)$, the received vector $\mathbf{r}[\mathbf{i}]$ is given by the first two rows of Eqn. 2.4. Then, it may be shown that, given the channel coefficients, and

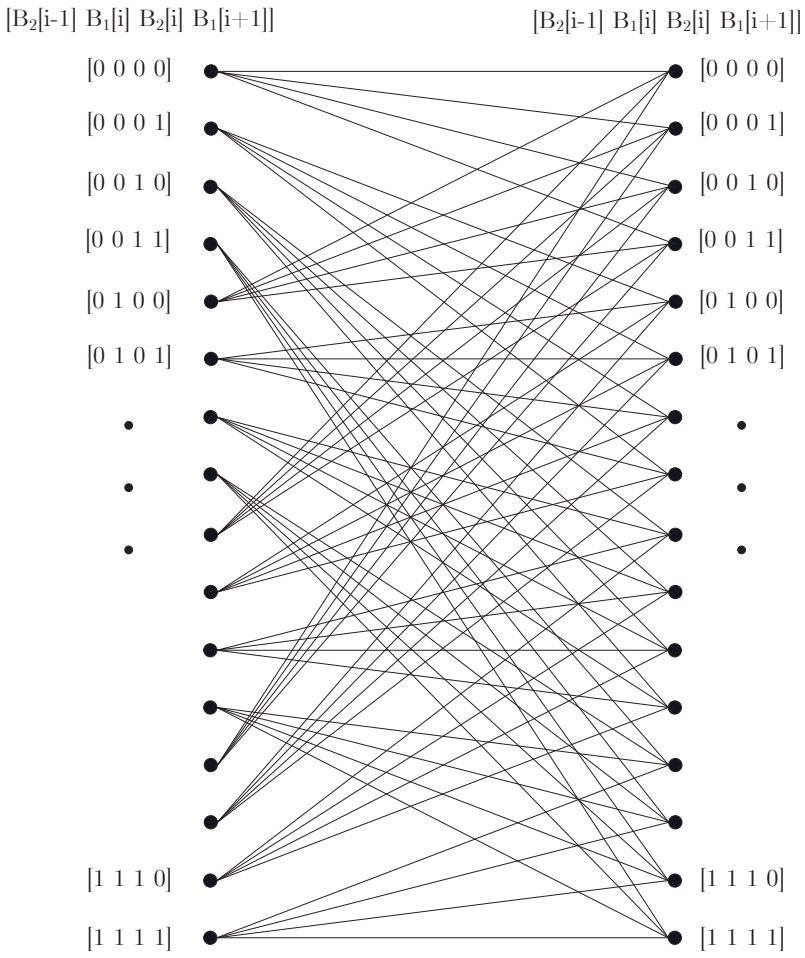


**Figure 2.8**: Trellis connectivity

substituting $\rho$ for $\rho_{12}$, $(1-\rho)$ for $\rho_{21}$, and $\alpha$ for $\rho^2+(1-\rho)^2$ the (1,1)th entry of the error covariance matrix, **E**, is given by

$$E_{11}(\rho) = 1 - \frac{|h_{11}|^2}{K}[|h_{11}|^2(1-\rho)^2 + |h_{12}|^2(1+\rho^2\alpha) + \sigma^2(1-\rho^2)] \tag{2.28a}$$

where

$$K = (|h_{11}|^2 + \alpha|h_{12}|^2 + \sigma^2)(\alpha|h_{11}|^2 + |h_{12}|^2 + \sigma^2) - \rho^2(|h_{11}|^2 + |h_{12}|^2 + \sigma^2)^2 \tag{2.28b}$$

For notational convenience the following substitutions are made, $a = |h_{11}|^2$ and $b = |h_{12}|^2$. It can be shown that in order to minimize the error variance, $\rho$ must satisfy the following quartic equation:

$$A\rho^4 + B\rho^3 + C\rho^2 + D\rho + E = 0 \qquad \text{where} \tag{2.29a}$$

$$A = -4ab \tag{2.29b}$$

$$B = -2\sigma^4 + 4a\sigma^2 + 2b^2 - 2a^2 + 8ab \tag{2.29c}$$

$$C = -3\sigma^4 - (2a + 6b)\sigma^2 + 5a^2 - 2ab - 3b^2 \tag{2.29d}$$

$$D = -4a\sigma^2 - 4a^2 - 4ab \tag{2.29e}$$

$$E = \sigma^4 + (2a + 2b)\sigma^2 + (a + b)^2 \tag{2.29f}$$

At very low SNRs, when $\sigma^4 \gg \sigma^2, a, b$, we can make the following approximation to Eqn. 2.29.

$$-2\sigma^4\rho^3 - 3\sigma^4\rho^2 + \sigma^4 = 0 \tag{2.30a}$$

$$\Rightarrow (\rho + 1)^2(2\rho - 1) = 0 \tag{2.30b}$$

This yields $\rho = 0.5$ (the only solution between 0 and 1) as the optimal solution at very low SNRs. This corresponds to a timing offset of half a symbol between the first and second transmitters. For other values of SNR, there is no simple closed form expression for the optimal $\rho$ or $\tau_1$. The error variance averaged over many channel instances, can be easily plotted as a function of offset $\tau_1$, by using Eqn. 2.28 or Eqn. 2.27. The error variance of the first stream for a $(M_T, M_R) = (2, 2)$ system is shown in Fig.2.14. Although not shown, the error variance of a $(M_T, M_R) = (2, 1)$ is qualitatively very similar.

As expected from Eqn. 2.30, $\tau_1$ gets progressively closer to $T_s/2$ as the SNR gets lower. At high SNRs the offset tends to 0. This has an interesting interpretation. Note that the interference levels are lowest when $\tau_1 = T_s/2$. Yet, the optimal offset at high SNRs is not $T_s/2$, but tends to 0 as SNRs increase. This seems to indicate that the receiver yields a better BER performance in a higher interference environment. This seemingly counter intuitive result may be

explained by observing that higher the SNR, the easier it is for the MMSE receiver to predictively cancel the interference. Thus, the higher interference levels do not adversely affect the receiver since the interference gets canceled.

## 2.7 Pulse Shape Design for MIMO with Timing Offset

In this section, we propose robustness to IAI (defined in Eqn. 2.1) as a new criterion for pulse shape design. The key idea is the following: Once the transmitters are offset from each other, the IAI is controlled by the correlation of the transmit pulse shape with the received pulse shape at an offset equal to the offset of the symbol boundaries. Without an offset, this criterion is no longer valid since the IAI is given by the correlation of the two pulses at zero offset (which is unity for all normalized pulse shapes).
Similar to the formulation of Eqn. 3 in [44] we minimize the cost function

$$\xi = \xi_s + \sum_{n \in S_{ISI}} \gamma(g[n] - d[n])^2 + \sum_{n \in S_{IAI}} \eta g^2[n] \tag{2.31}$$

where $\xi_s$ is the stop band energy of the square root Nyquist ($M$) discrete-time filter given by $h[n]$ which runs at $M$ samples/symbol where $n$ is the discrete time index for the samples of the waveform. $d[n]$ is the response of the convolution of the two square root Nyquist filters being designed with the target response given by $g[n]$. $S_{ISI}$ and $S_{IAI}$ respectively identify different subsets of samples of $n$ as shown below. $\gamma$ and $\eta$ are weighting functions that allow us to tradeoff one constraint with another. In an ideal square root Nyquist filter, $g[n] = h[n] * h[-n]$ where $*$ denotes convolution and $g[n]$ satisfies the no-ISI Nyquist criterion given by

$$g[n] = \begin{cases} 1, & if\ n = 0 \\ 0, & if\ n = mM, m \neq 0 \\ \text{arbitrary}, & if\ n \neq mM \end{cases} \tag{2.32a}$$

Thus $S_{ISI} = \{0, \pm M, \pm 2M ...\}$ is the subset of n where constraints are placed to minimize the ISI.

In order to reduce the IAI, we need to lower the energy of $g[n]$ at the offset points. Thus, e.g., for an offset of $T_s/2$, the sum of the squares of the samples of $g[n]$ at $\pm M/2$, $\pm M(1+1/2)$, $\pm M(2+1/2)$ and so on need to be lowered. By choosing $S_{IAI}$ to be the set $\{\pm M/2, \pm M(1+1/2), \pm M(2+1/2)...\}$ and by choosing appropriate weights, $\gamma$ and $\eta$, we can perform a tradeoff between the reduction of ISI and IAI. In [44], an iterative method for designing a filter confirming to such a cost function is described in detail and is used by us.

Using this method of pulse shape generation we can create a family of pulses that have various tradeoffs of ISI, IAI and stop-band attenuation. Here we show an example of such a pulse, by choosing an excess bandwidth of 25% and $\gamma = 1$ and $\eta = 0.6$. The key properties of this pulse in comparison to the square root raised cosine pulse shape is summarized in Table 2.2 below.

**Table 2.2**: Square root raised cosine (SRRC) compared to new pulse

| | **SRRC $*$ SRRC** | **New Pulse $*$ New Pulse** |
|---|---|---|
| Residual ISI (dB) | -74 | -19 |
| $T_s/2$ IAI (dB) | -0.58 | -1.02 |

The residual ISI goes up from -74dB (practically zero) in the case of two SRRC pulses convolved with each other to -19dB (still pretty low) in the case of the two proposed pulses convolved with each other. The IAI power caused by an offset of half a symbol time, $T_s/2$, however has been improved from about -0.58dB to about -1.02dB.

The frequency response of 3 different filters are plotted in Fig. 2.9 below. It may be seen that compared to the frequency response of a SRRC filter of same length, the proposed pulse has worse stop band attenuation. The peak sidelobe level is still close to -30dB below the main lobe and is thus considered acceptable. The time domain response is shown in Fig. 2.10 where it may be seen that the two pulse shapes are similar, though ISI has increased for the proposed pulse at the benefit of a lower IAI at $T_s/2$ offset.

Although we are showing only a single pulse shape here, different designers

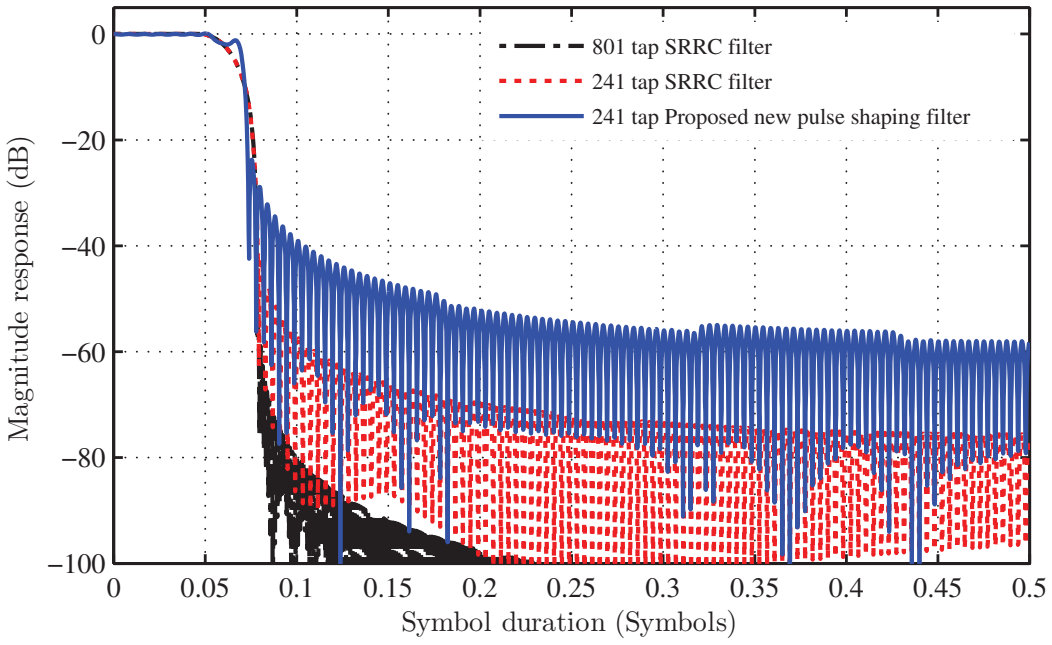


**Figure 2.9**: Frequency response of proposed new pulse compared with SRRC filter

could come up with a different pulse shape depending on different weights imposed in Eqn. 2.31 depending on the designers preference for various system parameters. Our emphasis here is on the importance of minimization of IAI as a filter design parameter for offset MIMO systems not so much on the exact choice of the parameters which might vary from system to system.

## 2.8 Simulation Results

The simulations have been done as a set of experiments where, in each case, comparisons have been made to similar aligned systems. In all cases, the channel is assumed to be known perfectly at the receiver. Each simulation also assumes a block fading model where the channel is independent from block to block and is assumed to be constant over the duration of each block. The channel coefficients have been generated as samples from a mean zero, unit variance complex Gaussian random variable. To obtain statistically reliable results, each datapoint is obtained

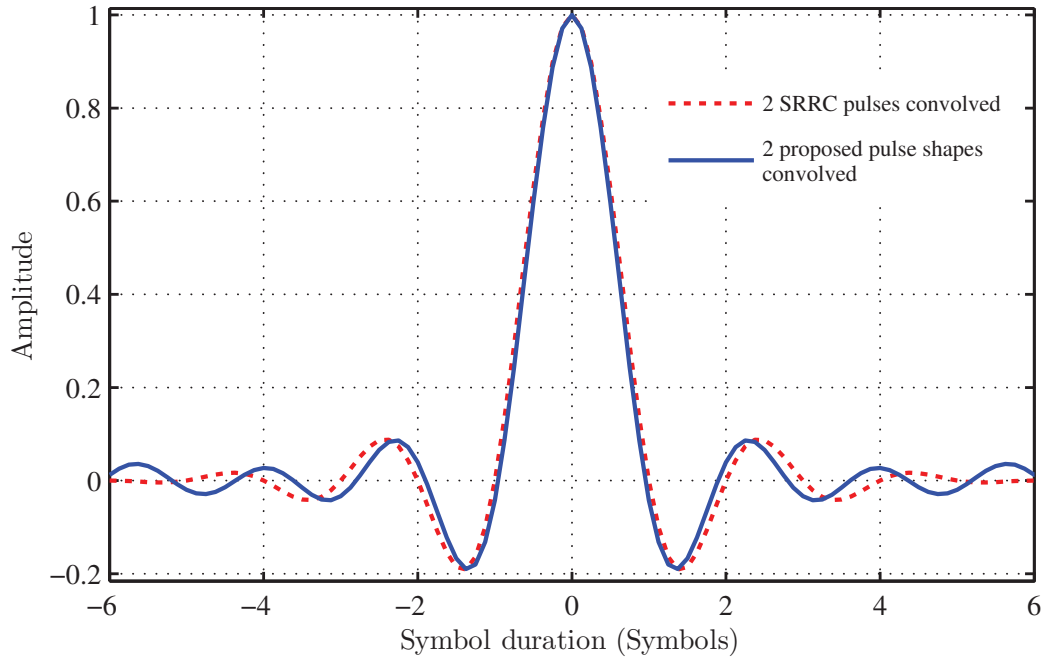


**Figure 2.10**: Impulse response of proposed new pulse compared with SRRC filter

by simulating at least 10000 blocks. The total transmit power is held constant irrespective of the number of transmitters by normalizing the output power from each transmitter by the number of transmitters, $M_T$. The performance metric of choice is symbol error rate (SER) or bit error rate (BER) which is plotted in the following graphs as a function of $E_s/N_0$, the ratio of the symbol energy ($E_s$) to the noise power spectral density ($N_0$). The performance is compared at a SER equal to $10^{-2}$.

### 2.8.1 Comparison with OSIC VBLAST

In Fig. 2.11 and 2.12, the performance of the proposed system with MMSE receivers is compared to that of a traditional aligned VBLAST with ordered successive interference cancellation (OSIC). A $(M_T, M_R) = (2, 2)$ system with quadrature phase shift keying (QPSK) modulation is simulated with blocks containing 128 symbols. The performance of systems with rectangular pulse shaping is shown

in Fig. 2.11 and that of systems with raised cosine pulse shaping with 25% excess bandwidth is shown in Fig. 2.12. The square root raised cosine (SRRC) filters on the transmitter and receiver sides have both been truncated to 13 symbols.

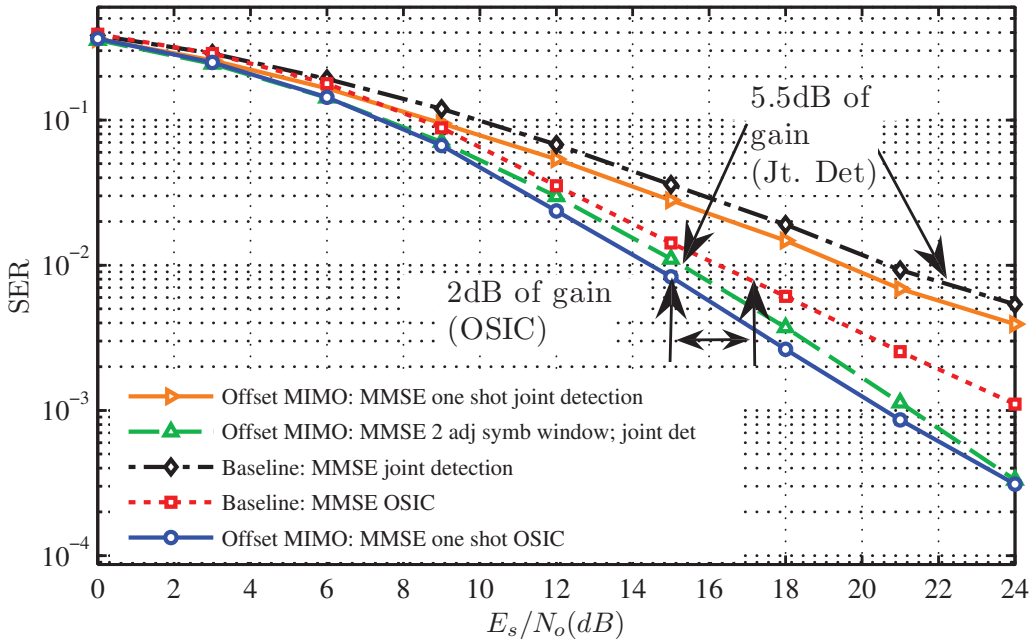


**Figure 2.11**: Offset MIMO with MMSE Rx compared OSIC VBLAST. $(M_T, M_R) = (2, 2)$. Modulation: QPSK. Rectangular pulse shaping.

In either case, the comparison has been made to the "best" aligned scheme which is when the VBLAST receivers employ OSIC [71]. It may be seen that the proposed system outperforms the VBLAST scheme both when rectangular pulse shaping is employed as well as when the raised cosine pulse shape is employed. In the latter, more practical case, the gain is about 1.8dB (at a SER of $10^{-2}$) when OSIC is employed on both the proposed system as well as on aligned traditional VBLAST.

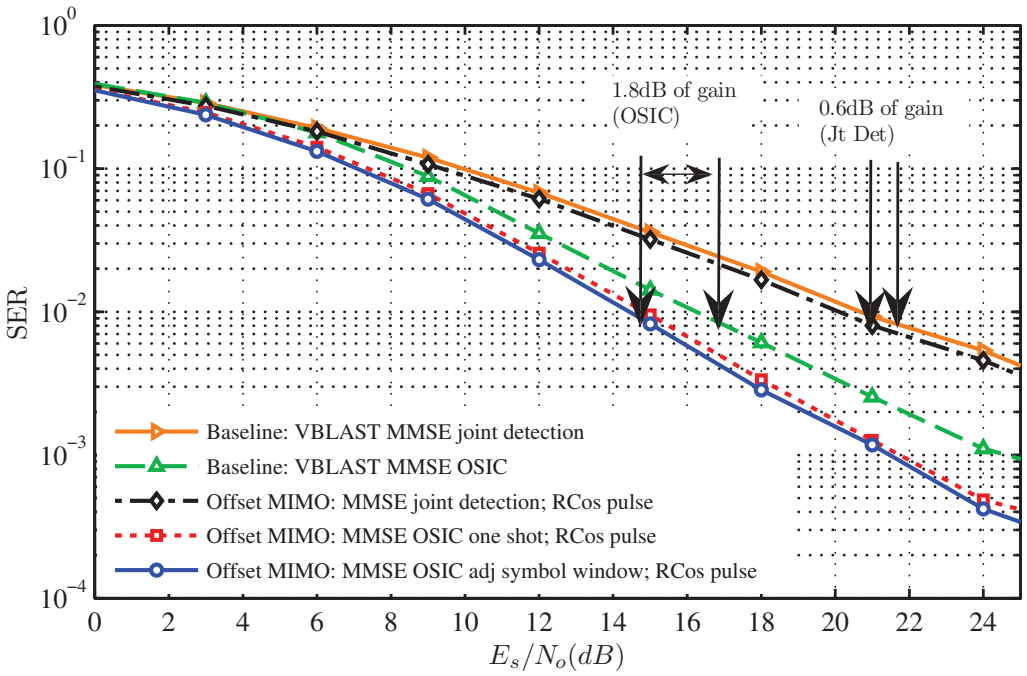


**Figure 2.12**: Offset MIMO with SRRC pulse shaping vs OSIC VBLAST. $(M_T, M_R) = (2, 2)$. Modulation: QPSK. Raised cosine pulse shaping.

### 2.8.2 Performance for Various Offsets

Fig. 2.13 shows the performance of a $(M_T, M_R) = (2, 2)$ system with BPSK modulation for various offsets between the first and second transmitters. A rectangular pulse shape is used. The performance of both an one-shot as well as a windowed receiver are shown. It may be seen that the MMSE windowed receiver achieves a lower BER with offset of $0.5T_s$, whereas when the one-shot receiver is employed, an offset of $0.2T_s$ is better at higher SNRs. More details on the performance at various offsets as well as an analytical derivation of an optimal offset for a $(M_T, M_R) = (2, 1)$ may be found in our prior work [33] and is also given in Section 2.6.

Fig. 2.14 also shows the error covariance of the first stream of a joint windowed MMSE receiver. The optimal offset for this receiver is $T_s/2$ regardless of SNR.

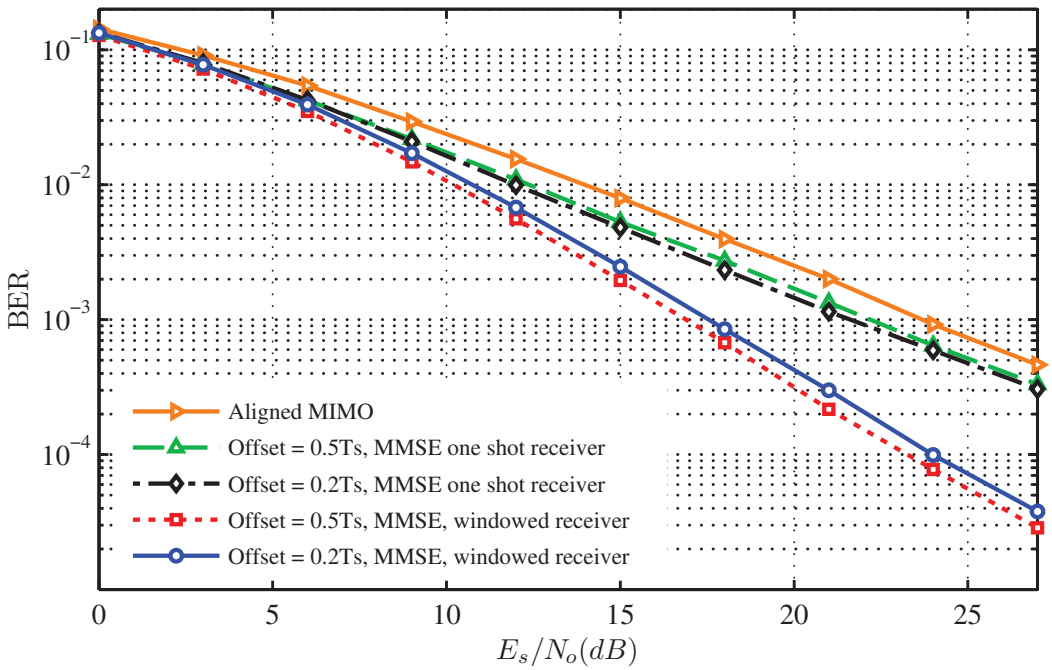


**Figure 2.13**: BER performance with various offsets, $(M_T, M_R) = (2, 2)$. Modulation: BPSK

### 2.8.3 Performance of Sequence Detection based Receivers

In Fig. 2.15, the performance of Viterbi algorithm based receivers is shown in comparison to that for a traditional MIMO system employing symbol by symbol ML detection. BPSK modulation with rectangular pulse shaping was used in a $(M_T, M_R) = (2, 2)$ system. Three curves are shown for offset MIMO: i) without employing noise whitening, ii) using one shot noise whitening ($W = 0$), iii) using extended window noise whitening ($W = 2$).

It may be seen that without noise whitening, the performance of the Viterbi algorithm based receiver is approximately equal to that of the traditional symbol aligned system with ML detection. However, when noise whitening is employed we pick up a gain of about 0.5dB at a BER of $10^{-2}$. While the gains in this case are admittedly smaller, in some systems even a 0.5dB gain in performance might be worth the additional complexity.

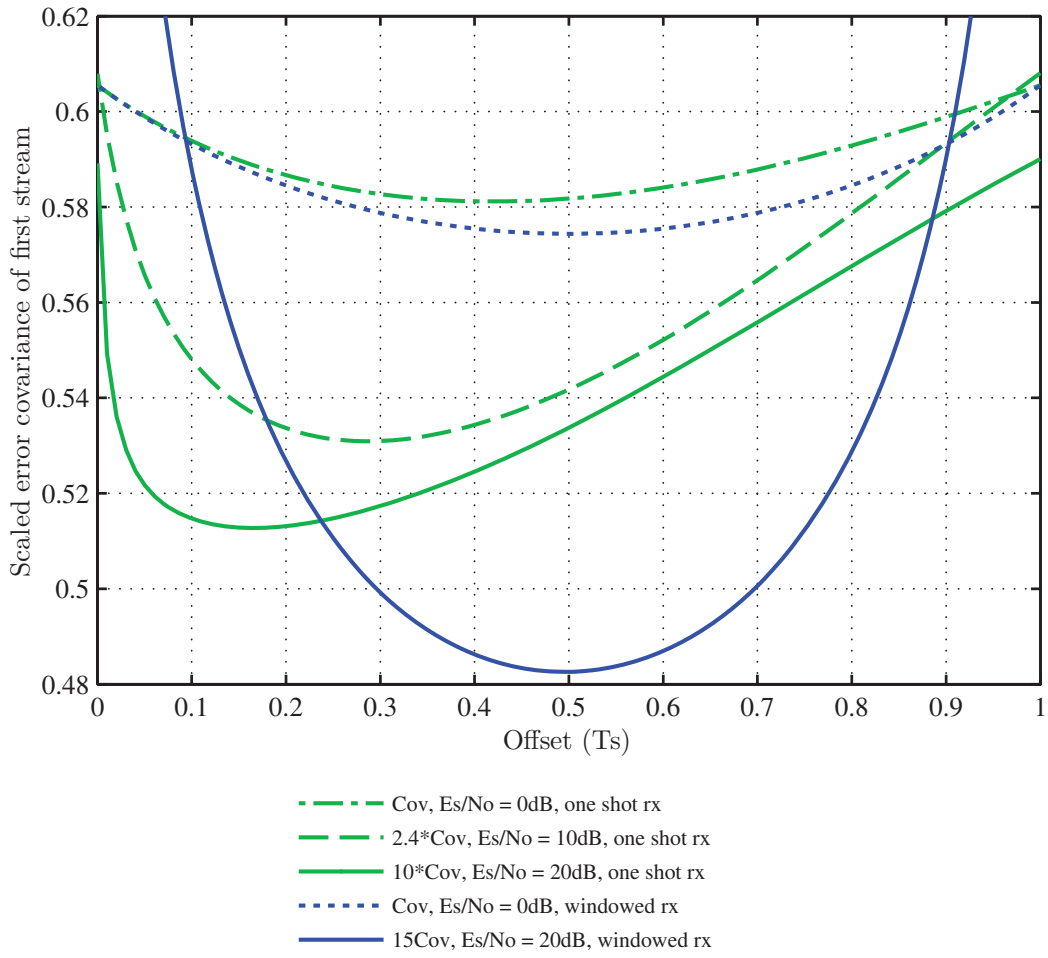


**Figure 2.14**: Error covariance for one shot and windowed MMSE Rx as a function of offset. $(M_T, M_R) = (2, 2)$

### 2.8.4 Performance of a $3 \times 3$ system

In Fig. 2.16, we present the results of a $3 \times 3$ MIMO system with offset transmission with MMSE joint detection receivers. In this case, there are two offsets and they have been set to $T_s/3$ and $2T_s/3$. It may be seen that the performance gains are over 6dB (at $SER = 10^{-2}$) when used with a rectangular pulse shape.

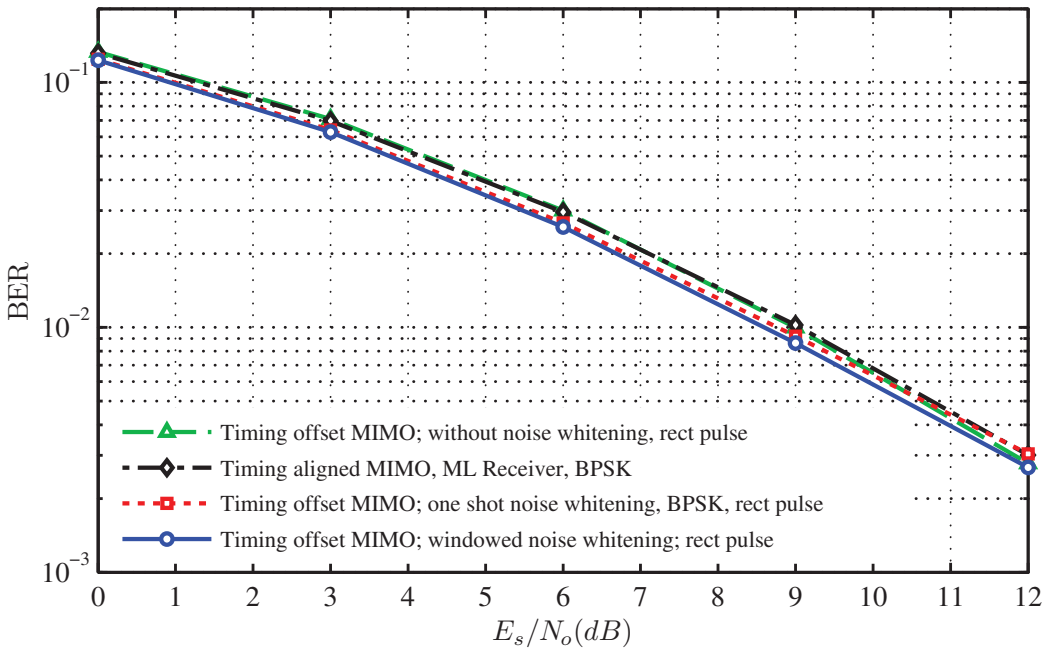


**Figure 2.15**: Impact of noise whitening on trellis based receivers. $(M_T, M_R) = (2, 2)$. Offset = 0.5Ts. Modulation: BPSK

### 2.8.5 ZF Receivers

The performance of the optimal ZF receiver is plotted against the performance of the ZF receiver presented by Shao et al. in Fig. 2.17. It may be seen that while the Shao et al. receiver degrades significantly with increasing block size $S$, the optimal ZF receiver has a very weak dependence on block size. In Fig. 2.17, the x-axis has been plotted in terms of $E_t/N_0 = \frac{ST_s+\tau_1}{ST_s} E_s/N_0$, where $S$ is the block size, $T_s$ the symbol duration and $\tau_1$ the offset. As shown in [83], this ensures that the data rate across all the systems is the same. We emphasize, however, that normalizing the data rate does not imply that all the block sizes are equally efficient. Very short block sizes, lead to considerably less spectral efficiency due to the inter gap idle time representing a higher overhead.

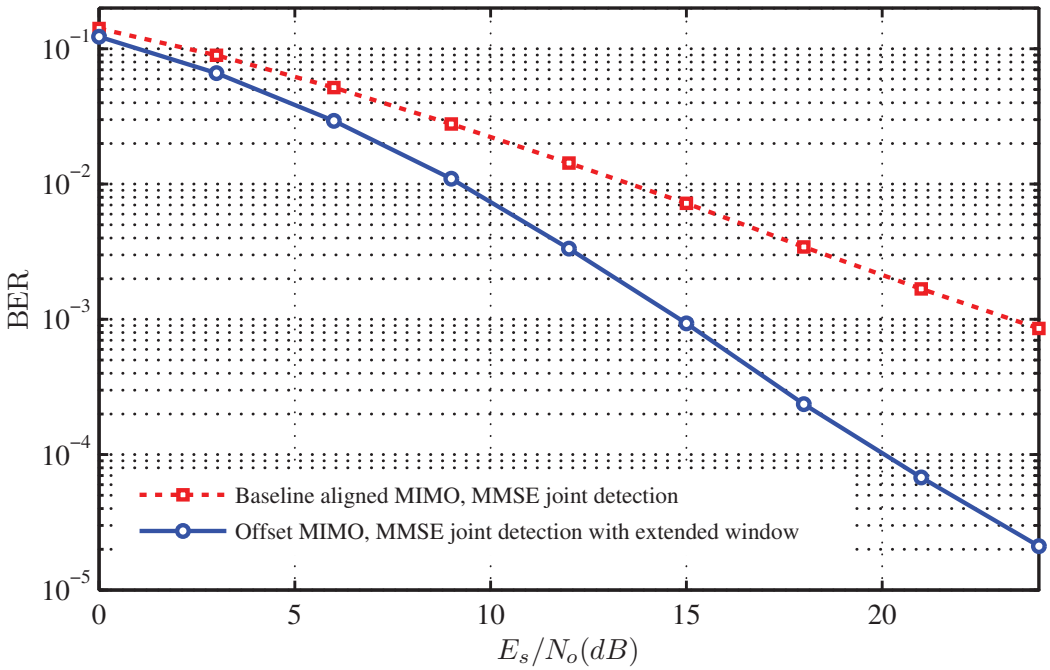


**Figure 2.16**: Performance of a 3x3 system. $(M_T, M_R) = (3, 3)$. Modulation:BPSK

### 2.8.6 Performance of New Pulse Shaping

To show the benefits of the proposed pulse shaping, the performance of a system using a member of the new proposed pulse family is compared in Fig. 2.18 to the performance of a system using a SRRC pulse. Both systems were simulated using a MMSE joint detection receiver. The additional performance improvement of approximately 0.25dB comes with no additional system complexity and can thus be regarded of as "free". Although not shown, the new pulse could be used with the trellis based receivers or zero forcing receivers as well.

## 2.9 Conclusions

A novel MIMO transmission scheme using transmitters that are intentionally offset in time from each other, has been presented in this chapter. A non-zero (but known) symbol timing offset is introduced between the signals transmitted

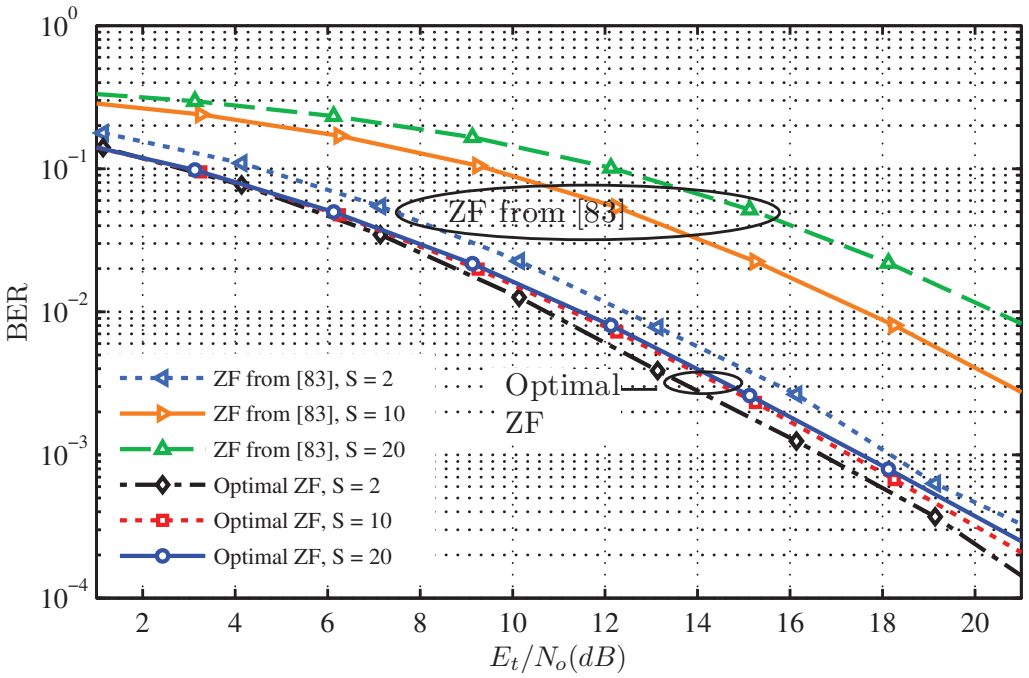


**Figure 2.17**: Optimal ZF receiver outperforms ZF receiver from [83]

from the different transmitters to take advantage of the inefficiencies in practical signalling systems. It is shown that a suitably designed receiver can utilize this information to extract significant performance gains. This transmission scheme is studied in conjunction with different kinds of receivers: ZF, MMSE receivers, as well as MIMO MMSE receivers with OSIC and trellis based sequence detection based receivers. A new pulse shape design that lowers IAI has also been introduced and is shown to increase the gains of such offset transmission schemes.

A summary of highlights of the comparison between an aligned scheme like VBLAST with the proposed scheme is shown in Table 2.3. The main source of complexity increase is shown along with the performance gain. The performance gain is shown for a $(M_T, M_R) = (2, 2)$ system with an offset of $T_s/2$ using BPSK at a BER of $2 \times 10^{-3}$ in comparison to an aligned system.

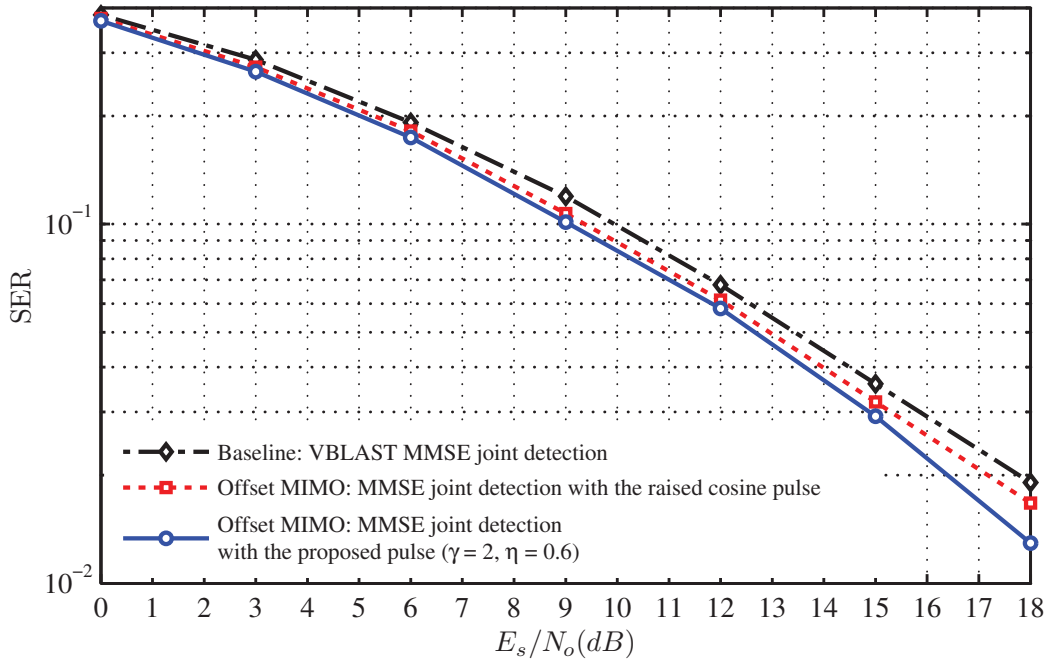


**Figure 2.18**: Performance of new pulse shaping, $(M_T, M_R) = (2, 2)$. $\beta = 0.25$

**Table 2.3**: Timing aligned MIMO compared to timing offset MIMO

| | **Offset MIMO** | **Aligned MIMO** |
|---|---|---|
| Matched filter rate | $M_T$ samp/symb | 1 samp/symb |
| ZF Rx | Needs inverse(or pseudoinverse) of $SM_T \times SM_T$ matrix. Performance gain $\sim$ 5dB | Needs inverse (or pseudoinverse) of $M_R \times M_T$ matrix |
| One-Shot MMSE Rx | Needs inverse (or pseudoinverse) of $M_T M_R \times M_T M_R$ matrix. Performance gain $\sim$ 1.5dB | Needs inverse (or pseudoinverse) of $M_R \times M_T$ matrix |

Continued on next page

**Table 2.3 – continued from previous page**

| | **Offset MIMO** | **Aligned MIMO** |
|---|---|---|
| Windowed MMSE Rx | More gains from more complexity. Complexity grows with window size. Performance gain $\sim$ 6dB | No gains over MMSE |
| Trellis based Rx | Trellis size (and thus complexity) can be traded for performance. Performance gain $\sim$ 0.5dB | Symbol by symbol ML receivers are optimal |
| IAI lowering pulse shapes | Performance gain $\sim$ 0.5dB | No gains from new pulse shapes |

The contents of this chapter have been adapted from "MIMO Systems with Intentional Timing Offset," Eurasip Journal on Advances in Signal Processing, 2011, and "Impact of Receiver Structure and Timing Offset on MIMO Spatial Multiplexing," in IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2008., both papers by A. Das and B.D. Rao.

# Chapter 3

# NDA SNR Estimation for SISO Systems

## 3.1 Introduction

Many communication systems (see e.g. [31, 75]) require the accurate estimation of signal to noise ratio (SNR) in the absence of any known data or pilot symbols in a non data aided (NDA) manner. SNR estimation in additive white Gaussian noise (AWGN) channels may be divided into 2 categories depending on whether the receiver carries out coherent demodulation or whether the channel has an unknown phase offset (non-coherent). In this chapter, we consider the case of NDA SNR estimation with coherent demodulation in AWGN. This implies that the channel gain considered in this chapter is a positive scalar quantity. In Chapter 4, we generalize this result to that of non-coherent demodulation and in that chapter we consider the channel gain to be a complex number.

The Cramér-Rao lower bound (CRLB) gives the minimum variance of unbiased estimators subject to some regularity conditions (See Chapter 3 of [59]), and is a very useful tool for evaluating the performance of an estimator. In [3], the CRLBs for NDA SNR estimation with quadrature/binary phase shift keying (QPSK/BPSK) modulation schemes have been presented for coherent demodulation. However, with the advent of high data rate communication systems, many

systems are using higher order modulation schemes such as 8-ary phase shift keying (8PSK), 16-ary or 32-ary quadrature amplitude modulation (16QAM or 32QAM) schemes. The satellite standard DVB-S2 [43] uses 16-ary and 32-ary amplitude and phase shift keying (16APSK and 32APSK respectively). The results in [3] are not extendable to other constellations. Thus, there is a need to evaluate the NDA SNR estimation CRLBs for these constellations and that is one of the objectives of this chapter. We present a methodology to compute the CRLB for any symmetric constellation and we evaluate and present the CRLBs for QPSK, 8PSK, 16QAM, 32QAM, 16APSK and 32APSK modulation schemes.

### 3.1.1 Overview of Prior Work

Gao and Tepedelenlioğlu presented the CRLB for SNR estimators for non-constant envelope modulation in [47]. The primary difference between their work and ours, is that in [47], the authors work with envelope based NDA SNR estimators and they determine the CRLB for these envelope based estimators. This chapter deals with the baseband, symbol-rate sampled, I and Q samples of the data. As the authors note in Sec III of [47], since the I/Q samples have more information than the envelope, CRLBs of the envelope based detectors have a worse performance than those evaluated in this chapter.

This chapter is adapted from our works [31] and [28]. Contemporaneous to our publication [28], Gappmair, in [48], also independently presented the CRLB for M-PSK and 16QAM. The difference between our work and [48] is that the CRLB for SNR estimation for more constellation schemes such as 16APSK, 32APSK and 32QAM was presented by us in [28] for the first time. Secondly, we evaluate the CRLB in units of dBs, instead of in normalized units. Since a systems engineer usually measures SNR in units of dBs, the evaluation of the CRLB in more natural units of dBs is generally more useful. Third, the estimators presented by [48] all significantly deviate from the CRLB at low SNRs whereas the estimator presented here will perform very close to the CRLB. Moreover, in Chapter 4, we generalize the findings of this chapter to a system model using a complex channel gain.

There has been considerable interest in estimators for SNR estimation. The

vast majority of prior work is applicable only to MPSK modulations [10, 26, 61, 62, 72, 97]. Some of the estimators developed for higher order modulations such as QAM, perform significantly above the CRLB even at high SNRs [102], [101] or are based on envelope based SNR estimators [47], [66]. NDA SNR estimators for 16APSK were recently proposed in [65] but the estimator deviates from the CRLB at higher SNRs. To our knowledge, our earlier conference work in [31] was the first where an expectation-maximization (EM) algorithm based NDA SNR estimation algorithm was presented in a framework that is easily generalizable to any arbitrary constellation. Using the 2 exemplary constellations of QPSK and 16APSK, experiments show that this estimator performs very close to the CRLB for all SNRs of practical interest even for burst sizes as small as 100 symbols.

An EM Algorithm based NDA SNR estimator was first presented for BPSK by Wiesel in [97]. Recently more general forms applicable to other constellations were presented, almost simultaneously, in [66] and by us in [31]. In [66], the authors consider an envelope based SNR estimator in a non-coherent (flat fading) environment while this current work considers a baseband I/Q sample based SNR estimator with coherent reception. Unlike the estimator presented here, the estimator in [66] is not applicable to all constellations (notably 32APSK with code rate 3/4 from [43]) and has a non-zero probability of failing completely. This latter issue is addressed by the authors in Sec. 4 of [66]. For very similar reasons, the envelope based estimator presented in [47] also has a non-zero probability of failing completely due to the argument of the square-root in Eqn. 7 of [47] evaluating to a negative number.

### 3.1.2 Application for Detection of Non-Linearity

Most wireless communications systems have a radio frequency (RF) power amplifier (PA) at the end of their transmit chain. These PAs are typically approximately linear only over a certain range of input power beyond which they eventually saturate [78]. Communications systems that use non-constant envelope modulation schemes such as square root raised cosine (SRRC) pulse shaped M-PSK or M-QAM, need to operate in the linear region of the PA in order to minimize

signal distortion as well as to reduce spectral emissions in adjacent bands [2].

Such systems rely on various methods for estimating the onset of non-linearity. Typical methods include current sensing techniques [19] or correlation methods [63], or subsampling techniques [54], all of which require specialized hardware at the amplifier. Once the onset of saturation is detected, the system usually takes remedial measures (e.g., reducing the symbol rate or changing to a more robust coding and/or modulation scheme) that allow it to reduce the required transmit power and thus stay in the linear range of the amplifier. In commercial satellite communications, the power amplifier is typically on the roof or outside the user's home and the satellite modem (SM) is located inside. Thus non-linearity estimation at the output of the power amplifier and sending the information back to the SM adds to the cost of each unit and this is undesirable.

The method proposed in this chapter estimates the onset of saturation without the addition of any hardware cost at the terminal by taking advantage of the ranging messages that each registered SM periodically transmits to the gateway. The *input* power of these bursts to the PA is dithered by a known amount and the *received* dither amount is estimated at the gateway and is used as a metric to determine whether the PA is operating in the linear region.

Since the ranging bursts (see Section 3.2) are relatively short bursts, this algorithm hinges on the ability to accurately estimate the received SNR. The proposed algorithm estimates the received dither amount by estimating the SNR using the EM algorithm. Although Wiesel et al., have demonstrated in [97] the use of the EM algorithm for SNR estimation for BPSK, the algorithm proposed in this chapter, unlike the one in [97], does not assume that all points from constellation are equally likely. Moreover, our algorithm is suitable for use with any digital constellation.

### 3.1.3 Our Contribution

Thus, the contribution of this chapter can be summarized to be the following: a) The presentation of the CRLBs for SNR estimation using baseband I-Q samples for any symmetric constellation and b) the presentation of an EM based

estimator that may be used with any arbitrary constellation for NDA SNR estimation with coherent reception in AWGN. The EM based SNR estimator is seen to perform very close to the CRLB at all SNRs of interest and c) the application of the NDA SNR estimation techniques presented to detect the onset of non-linearity at an amplifier.

The structure of the rest of this chapter is as follows. First, in Section 3.2 we describe the satellite system in which we shall apply the NDA SNR estimation techniques and outline the key principles behind the non-linearity estimation technique. In Section 3.3 we present the system model under consideration and pose the estimation problem. In Section 3.4 the CRLB is derived. The EM based SNR estimator is developed in Section 3.5 and the results are presented in Section 3.7 before concluding.

## 3.2 Surfbeam System Description

*SurfBeam*® is a broadband satellite system developed by ViaSat Inc., based on the data over cable service interface specification (DOCSIS®)[1]. *SurfBeam* is designed for use with geostationary spot-beam based satellites in the Ka (20/30 GHz) or Ku (12/14 GHz) band and provides the user with a broadband experience comparable to that currently provided by cable modems at similar costs. The network is currently in operation internationally supporting hundreds of thousands of users.

Although the upper layer protocols are all based on DOCSIS, the physical layer (PHY) has been modified to accommodate the unique challenges of the satellite-land channel. The modulation scheme used by the *Surfbeam* user SM terminals on the return link back to the satellite is QPSK with a forward error correction (FEC) rate of 1/2 or 3/4. The system utilizes various return link symbol rates including 2560ksps, 1280ksps, 640ksps etc. The return link is a multi-frequency time division multiple access (MF-TDMA) system.

[1]For more information, please visit www.viasat.com

### 3.2.1 Ranging Bursts

The DOCSIS standard requires each registered SM to periodically transmit a short RNG-REQ burst (also known as ranging burst) regardless of whether the SM is sending user data or not [55]. The ranging bursts of the Surfbeam system consist of 36-bytes of data and a 110-symbol preamble. The periodicity of transmission of a ranging burst is programmable, but is typically on the order of a second. While the preamble is uncoded, the data portion is coded by the FEC of the channel. Thus at QPSK r-1/2, the ranging burst consists of the 110-symbol preamble and 288 symbols while at QPSK r-3/4 the ranging bursts consists of 110-symbol preamble and 192symbols of data. The preamble overhead is large because the system has been optimized for data-bursts that are typically much larger than the 36-byte RNG-REQ bursts (See Section 8.3.5 of [55]).

Although we are presenting this algorithm in the context of a DOCSIS based satellite communication system, it is a very general algorithm and works for any system where there is a periodic burst (say $N$ byte messages) every $T$ seconds. These type of ranging bursts are typically used for synchronization, power control, link maintenance and channel estimation purposes and are fairly common in communication systems.

### 3.2.2 Power Control

For the exemplary satellite system described, the return link, or upstream, is the path from the user terminal back to the gateway. In order to ensure reliable communication, typically power control algorithms are employed at the gateway. Upstream power control is maintained by a power control loop operating at the gateway that utilizes these ranging bursts to estimate the received SNR. This loop then appropriately instructs the SM to increase or decrease its signal power in an attempt to always maintain a desired SNR based on the link budget and quality of service (QoS) requirements.

When the signal strength is attenuated (e.g., during a rain event), the upstream power control loop instructs the SM to increase its signal power until the upstream SNR target is achieved. It should be mentioned that the SM can

only increase or decrease the power at intermediate frequencies (IF), as shown in Fig. 3.2. The SM has no knowledge of the actual increase or decrease in signal power at RF. Typically the SM (inside the user's house) and the RF PA (typically on the roof), are connected with a coaxial cable whose length varies from house to house depending on the separation between the SM and the RF PA. Consequently, the same output power from two different SMs, could result in two different input powers at the RF PA since the attenuation through the cable could be quite different.

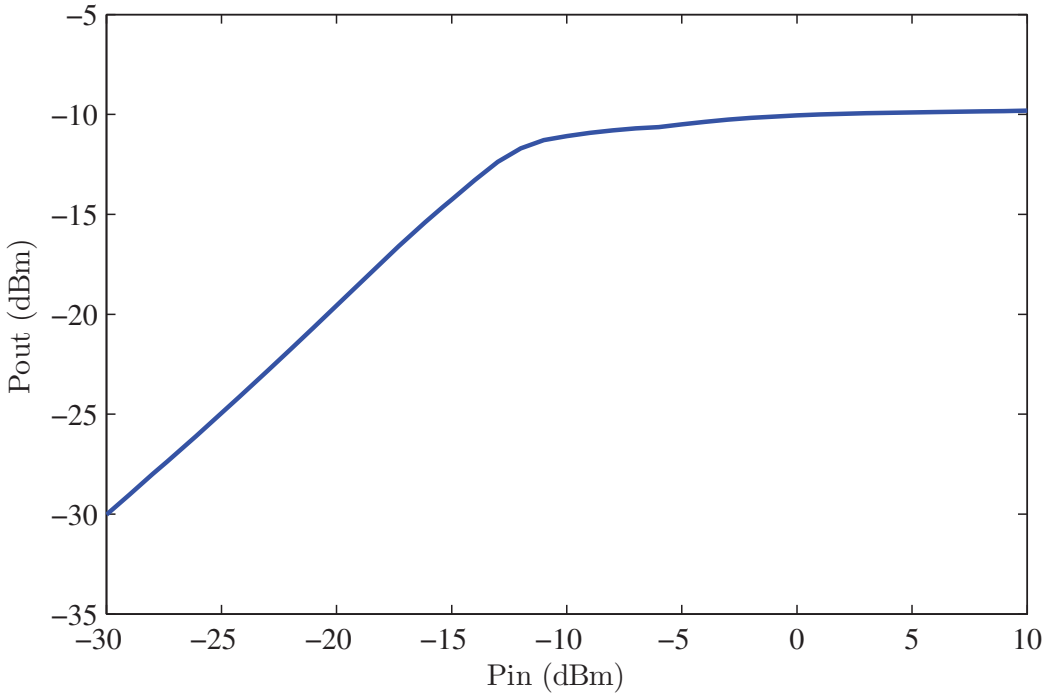


**Figure 3.1**: A typical power amplifier curve of output power as a function of input power

The relative increase of signal power at RF depends on the operating point of the PA. A typical power profile of a PA is shown in Fig. 3.1. There is a linear region and then a non-linear region where the amplifier eventually satures. Thus during a fade event, as a terminal gradually increases its transmit power, the operating point of a particular SM could gradually move from the linear region to the non-linear region of the power amplifier and it is this transition that the

proposed algorithm is designed to detect. Once this is detected, the algorithm switches the SM from a high symbol rate channel to a lower symbol rate channel (e.g., from 2560ksps to 1280ksps), thus enabling the SM to achieve the same error rate for 3dB less power (albiet at the cost of a lower symbol rate).

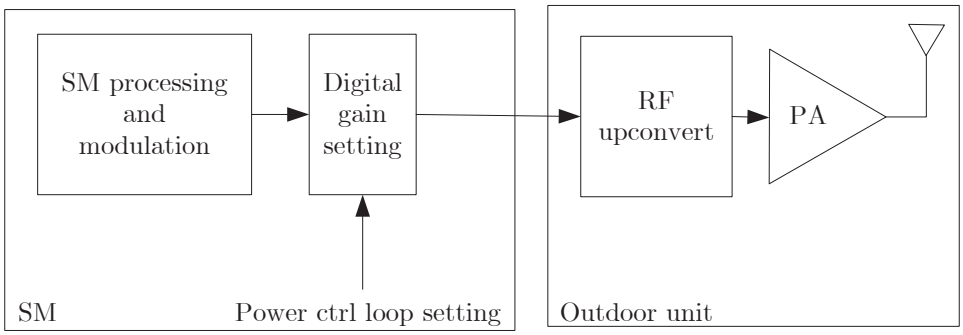


**Figure 3.2**: Remote terminal signal gain setting

### 3.2.3 Dithering Ranging Bursts

The proposed algorithm dithers the IF transmit power of the ranging bursts by a known amount, $\Delta$dB. There are different ways in which this dither amount may be applied. In one embodiment, alternate RNG-REQ bursts are transmitted at $P$ and $P + \Delta$, respectively, where $P$ is the power set by the power control loop. In this method, all the N symbols of a particular burst are transmitted at the same power. One of the drawbacks with this approach is that in the presence of an unknown amount of fade (perhaps due to rain) the RNG-REQ burst arriving later will be attenuated (relative to the one arriving first). Thus, even with perfect estimation, the dither amount estimated at the receiver (which equals to the difference in received SNR between the two bursts) will be less than $\Delta$ even when the PA is operating in the linear region. A second limitation of this approach is that half the bursts are transmitted $\Delta$dB higher than what is required by the power control loop and, depending on the system, this could violate the off-axis emissions limit imposed by regulatory bodies like the FCC.

In order to overcome these drawbacks, it is proposed that symbols within each RNG-REQ burst be dithered. Thus, for example, the even symbols could be transmitted at $P + \Delta_1$ and the odd symbols could be transmitted at $P - \Delta_2$ such that $\Delta = \Delta_1 + \Delta_2$. Since the symbols at the higher and lower power are now temporally contiguous, this approach is unaffected by practical rain fades. Moreover, $\Delta_1$ and $\Delta_2$ may be chosen such that the average burst power is unchanged from $P$ thus ensuring that the off-axis emissions limit is not violated. If $\Delta$ is expressed in dBs, then in order to ensure that the average power of the burst is unchanged, $\Delta_1$ and $\Delta_2$ (also in dBs) are given by Eqn. 3.1.

$$\Delta_1 = 10 \log 10 \left( 2 \frac{10^{\Delta/10}}{10^{\Delta/10} + 1} \right) \tag{3.1a}$$

$$\Delta_2 = 10 \log 10 \left( \frac{2}{10^{\Delta/10} + 1} \right) \tag{3.1b}$$

If a 3dB dither is desired, then $\Delta = 3dB$ and using Eqn. 3.1, $\Delta_1 = 1.25dB$ and $\Delta_2 = -1.75dB$. So half the burst will be transmitted at 1.25dB higher than the power setting of the control loop and the other half will be transmitted at 1.75dB lower power.

Other dither patterns, e.g., where the first half the burst is sent at the higher power and the second half is sent at the lower power are also potential candidates. The different dither schemes are shown in Fig. 3.3

With this approach to dithering, half the symbols of all ranging bursts are transmitted at a power that is *less* than that set by the power control loop for the desired system performance and the other half of the symbols are transmitted at a power *higher* than that required. Since the symbols are encoded with a strong turbo code, simulations show that the performance degradation is less than 0.5dB ($\Delta = 2dB$, FEC: $r = 3/4$, Burst Error Rate = $10^{-4}$). Secondly, when the SM is not idle the vast majority of the return link traffic is data packets (which are not dithered). Thus, return link performance averaged over all the bursts from the SM suffers no measurable degradation.

At the gateway, the difference in received SNR is estimated and is compared to a threshold. Based on this metric, a determination is made as to whether the

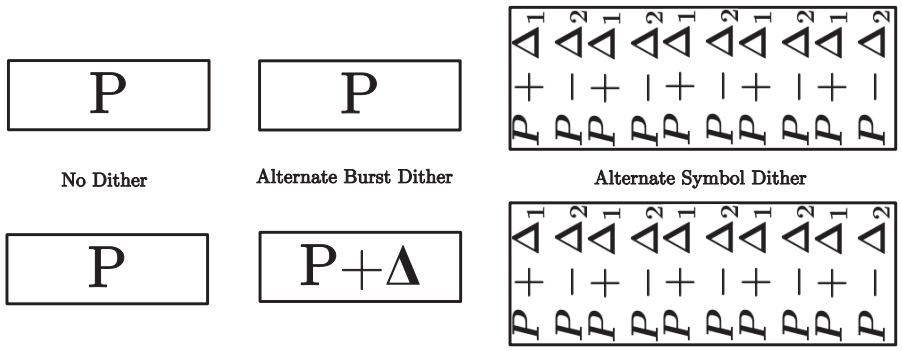


**Figure 3.3**: Power settings with and without dither

PA is operating in the linear region or in the compression region. The difference in received power could be averaged over a few pairs of bursts in order to have a more accurate estimate if so desired. There is the usual tradeoff, common in most detection problems, between probability of false alarm and probability of miss in setting the threshold.

At the gateway, the ranging bursts from each SM are received in preassigned timeslots. After processing through the receiver front end, the signal is matched filtered and downsampled to symbol rate. These soft decision I, Q samples are then made available for SNR estimation. This occurs before the turbo-decoding has taken place and thus the algorithm cannot take advantage of known data to aid the SNR estimation process. Although the preamble of the RNG-REQ bursts is known, for the purposes of this section, it is assumed that none of the symbols are known. In Sec. 3.6, we use the known preamble symbols to simplify the algorithm.

This algorithm, thus falls in the realm of non-data aided (NDA) SNR estimation. In the following we shall derive the appropriate system model and then solve the estimation problem.

## 3.3 System Model

In the absence of frequency, phase or timing estimation errors, the output of a matched filter at symbol rate (assuming the use of Nyquist pulse shaping at both transmitter and receiver), may be expressed as

$$y_i = hx_i + n_i \quad i = 0...N \tag{3.2}$$

where $h$ is a positive scalar channel gain (assumed constant over the estimation block of $N$ symbols), $n_i$ are samples of an independent and identically distributed (i.i.d) complex Gaussian noise process with the variance of its real part and complex part each equaling $\sigma^2/2$. Thus the pdf of $n_i$ is given by $\mathcal{N}(0, \sigma^2)$. The assumption of a positive scalar channel gain assumes a coherent receiver. In Eqn. 3.2, $i$ is the time index and we assume a block fading channel where the channel gain, $h$, and the noise variance, $\sigma^2$, is constant for the whole block of $N$ symbols over which the estimation will be carried out. For a non-coherent receiver, the channel gain should be a complex quantity and this is addressed in Chapter 4. The transmitted symbols, $x_i$, are picked randomly from a normalized unit energy constellation with $M$ constellation points. Thus, e.g., for QPSK, $x_i \in \{\pm 1/\sqrt{2} \pm j/\sqrt{2}\}$ and $M = 4$. We show the values of $x_i$ for the other constellations studied in Appendix 3.A.

Using this formulation, the SNR, $\rho$, assumed to be constant over the block of $N$ symbols over which the estimation is to be carried out, is given by:

$$\rho(\mathbf{\Theta}) = \frac{h^2}{\sigma^2} \tag{3.3}$$

or in dB scale by:

$$\widetilde{\rho}(\mathbf{\Theta}) = 10\log_{10}(\rho) = 10\log_{10}\left(\frac{h^2}{\sigma^2}\right) \tag{3.4}$$

where the parameter vector to be estimated is given by:

$$\mathbf{\Theta} = \begin{bmatrix} h & \sigma^2 \end{bmatrix}^T \tag{3.5}$$

For symmetric constellations, assuming i.i.d. received samples, and equiprobable transmit symbols, the pdf of the received vector $\mathbf{y} = [y_0, y_1, .., y_i, ...y_N]^T$ is given by Eqn. 3.6.

$$p(\mathbf{y};\mathbf{\Theta}) = \prod_{i=1}^{N} p(y_i \mid \mathbf{\Theta}) \tag{3.6a}$$

$$= \prod_{i=1}^{N}\sum_{l=1}^{M} p(y_i|x_l;\mathbf{\Theta})p(x_l) \tag{3.6b}$$

$$= \frac{1}{M}\prod_{i=1}^{N}\sum_{l=1}^{M} \frac{1}{\pi\sigma^2}\exp\left(\frac{-(y_i - hx_l)^*(y_i - hx_l)}{\sigma^2}\right) \tag{3.6c}$$

The log likelihood function, $g$ is easily derived from Eqn. 3.6 to be

$$g = \ln(p(\mathbf{y};\sigma^2)) = -N\ln(M\pi) - N\ln(\sigma^2) - \sum_{i=1}^{N}\left[\ln[\sum_{l=1}^{M}\exp(\frac{\phi_{i,l}}{\sigma^2})]\right] \tag{3.7}$$

where

$$\phi_{i,l} = (y_i - hx_l)^*(y_i - hx_l) \tag{3.8}$$

## 3.4 CRLB Evaluation

As shown in [59] and using Eqn. 3.3, the CRLB is given by

$$CRLB = \frac{\partial\rho(\mathbf{\Theta})}{\partial\mathbf{\Theta}^T}{\mathbf{F}_{\mathbf{\Theta}}}^{-1}\left(\frac{\partial\rho(\mathbf{\Theta})}{\partial\mathbf{\Theta}^T}\right)^T \tag{3.9}$$

where $\mathbf{F}_{\mathbf{\Theta}}$ is the $4\times 4$ Fisher Information Matrix (FIM), given by:

$$\mathbf{F}_{\mathbf{\Theta}} = E\left[\left(\frac{\partial g}{\partial\mathbf{\Theta}}\right)\left(\frac{\partial g}{\partial\mathbf{\Theta}}\right)^T\right] \tag{3.10}$$

$g$ is the log likelihood function of Eqn. 3.7, $E()$ is the expectation operator and the expectation is taken with respect to the distribution of $\mathbf{y}$. If it is desired to calculate CRLB in units of $(dB)^2$, then instead of $\rho$ in Eqn. 3.9, we shall use $\widetilde{\rho}$ from Eqn. 3.4.

From Eqn. 3.3 and Eqn. 3.4 we can derive the following two equations:

$$\frac{\partial\rho(\mathbf{\Theta})}{\partial\mathbf{\Theta}^T} = \begin{bmatrix} 2h/\sigma^2 & -h^2/\sigma^4 \end{bmatrix} = \begin{bmatrix} 2\rho/h & -\rho/\sigma^2 \end{bmatrix} \tag{3.11a}$$

$$\frac{\partial\widetilde{\rho}(\mathbf{\Theta})}{\partial\mathbf{\Theta}^T} = \begin{bmatrix} \frac{20}{\ln(10)h} & \frac{-10}{\ln(10)\sigma^2} \end{bmatrix} \tag{3.11b}$$

Thus, in order to evaluate the CRLB of Eqn. 3.9, only $\mathbf{F_\Theta}$ remains to be evaluated.

From Eqn. 3.7, the partial derivatives of the log likelihood function, $g$, with respect to the two parameters, $h$ and $\sigma^2$, are given by:

$$\frac{\partial g}{\partial \sigma^2} = \frac{-N}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{N} \frac{\sum_{l=1}^{M}\left(\phi_{i,l}\exp(\frac{-\phi_{i,l}}{\sigma^2})\right)}{\sum_{l=1}^{M}\exp(\frac{-\phi_{i,l}}{\sigma^2})} \tag{3.12a}$$

$$\frac{\partial g}{\partial h} = \frac{-2}{\sigma^2}\sum_{i=1}^{N} \frac{\sum_{l=1}^{M}\left(\exp(\frac{-\phi_{i,l}}{\sigma^2})\,(-h|x_l|^2 + Re(y_i^* x_l))\right)}{\sum_{l=1}^{M}\exp(\frac{-\phi_{i,l}}{\sigma^2})} \tag{3.12b}$$

In order to determine the 4 terms of the FIM of Eqn. 3.10, we have to determine the following three terms: $E(\frac{\partial g}{\partial \sigma^2}\frac{\partial g}{\partial h})$, $E(\frac{\partial g}{\partial h}\frac{\partial g}{\partial h})$, and $E(\frac{\partial g}{\partial \sigma^2}\frac{\partial g}{\partial \sigma^2})$. The expectations of these three terms cannot be determine analytically.

We evaluated the terms numerically using Monte Carlo simulations. In Chapter 4, an alternate method of numerical evaluation, using the Gauss Hermite quadrature has been presented. Both the methods were verified to yield the same result.

It may be shown [59], [90], that an alternate equivalent expression of the FIM is given by

$$\mathbf{F_\Theta} = \begin{bmatrix} -E(\frac{\partial^2 \ln p(\mathbf{y}|\theta)}{\partial h^2}) & -E(\frac{\partial^2 \ln p(\mathbf{y}|\theta)}{\partial h \partial \sigma^2}) \\ -E(\frac{\partial^2 \ln p(\mathbf{y}|\theta)}{\partial \sigma^2 \partial h}) & -E(\frac{\partial^2 \ln p(\mathbf{y}|\theta)}{\partial (\sigma^2)^2}) \end{bmatrix} \tag{3.13}$$

A slightly different, but equivalent, approach which involves finding the second derivatives directly and then doing a numerical evaluation of the expectation was shown by us in [28].

## 3.5 EM algorithm based Iterative SNR Estimation

The EM algorithm is known (for regular exponential family distributions) to iteratively attain the maximum likelihood(ML) estimate [35]. We use this algorithm is used to estimate ML estimates of $h$ and $\sigma^2$ and then the ML estimate of

the SNR is found using the invariance property of the ML estimate [59]. Thus, if $\hat{h}$ is the ML estimate of $h$ and $\hat{\sigma^2}$ is the ML estimate of $\sigma^2$, then the ML estimate of the SNR, $\rho$, is given by $\hat{\rho} = \frac{\hat{h}^2}{\hat{\sigma}^2}$

As stated in [35], the EM algorithm works on a set of *incomplete* data. In this case, the incomplete data is the sequence of unknown transmitted symbols, $x_i$. Thus the complete data set at time $i$ is $z_i = [y_i; x_i]$. If $M$ is the constellation order, $x_i$ can take one of $M$ different values. Consequently, in the presence of Gaussian noise, the probability density function of the observed samples $y_i$ is a mixture density of $M$ Gaussian mixtures. We denote the $M$ values of the symbols of the transmit constellation $c_1, c_2, ...c_M$.

Thus, the probability density function (pdf) of $y_i$ may be written as

$$f_Y(y_i|\mathbf{\Theta}) = \sum_{l=1}^{M} \alpha_l p(y_i|x_i = c_l, \mathbf{\Theta}) = \sum_{l=1}^{M} \alpha_l \frac{1}{\pi\sigma^2} \exp\left(\frac{-\|y_k - hc_l\|)^2}{\sigma^2}\right) \tag{3.14a}$$

where $\alpha_l$ is the proportion of the $l$-th mixture in the overall distribution and if all symbols are equally likely is equal to $1/M$. For the derivation of the SNR estimator, we shall maintain the possibility that the symbols may not be equiprobable. The deterministic, but unknown parameter vector to be estimated is given by $\mathbf{\Theta}$ as shown in Eqn. 3.5.

As is outlined in Equation 2.17 of [35], using the log likelihood function of the parameter conditioned on the complete data, $\mathcal{L}(\mathbf{\Theta}|Z)$, the $k$-th step of the iterative EM algorithm carries out the following two steps. Here the observation vector is $\mathbf{y}$ and the estimate of the parameter vector from the previous iteration is $\mathbf{\Theta}_{k-1}$:

$$\text{E-Step} \qquad Q(\mathbf{\Theta}; \mathbf{\Theta}_{k-1}) = E[\mathcal{L}(\mathbf{\Theta}|Z)|\mathbf{y}, \mathbf{\Theta}_{k-1}] \tag{3.15a}$$

$$\text{M-Step} \qquad \mathbf{\Theta}_k = \arg\max Q(\mathbf{\Theta}; \mathbf{\Theta}_{k-1}) \tag{3.15b}$$

Assuming independent data, the likelihood function may be computed as:

$$L(\mathbf{\Theta}|\mathbf{z}) = \prod_{i=1}^{N} \prod_{l=1}^{M} p(x_i = c_l)^{\gamma_{i,l}} p(y_i|x_i = c_l, \mathbf{\Theta})^{\gamma_{i,l}} \tag{3.16}$$

where $\gamma_{i,l}$ is the indicator function that equals 1 if $x_i = c_l$ and is 0 otherwise.

Since $p(x_i = c_l) = \alpha_l$, the log likelihood function can then be written as

$$\log[L(\mathbf{\Theta}|\mathbf{z})] = \sum_{i=1}^{N}\sum_{l=1}^{M}\gamma_{i,l}(\log(\alpha_l) + \log(p(y_i|x_i = c_l, \mathbf{\Theta}))) \tag{3.17a}$$

$$\mathcal{L}(\mathbf{\Theta}|\mathbf{z}) = \sum_{i=1}^{N}\gamma_{\mathbf{k}}^{T}[\mathbf{V}(\alpha) + \mathbf{U}_i(\mathbf{\Theta})] \tag{3.17b}$$

where $\gamma_i$ is a $M \times 1$ vector of $\gamma_{i,l}$, $\mathbf{V}(\alpha)$ and $\mathbf{U}_i(\mathbf{\Theta})$ are $M \times 1$ vectors of $\log(\alpha_l)$ and $\log(p(y_i|x_i = c_l, \mathbf{\Theta}))$ respectively. As an example, for QPSK, with knowledge of the complete data, $\gamma_i$ is one of the following 4 vectors: $[1,0,0,0]^T$, $[0,1,0,0]^T$, $[0,0,1,0]^T$, $[0,0,0,1]^T$. In the absence of perfect knowledge of what was transmitted, $\gamma_i$, is a $M \times 1$ vector of soft decisions, e.g., $[0.7, 0.1, 0.1, 0.1]^T$ with each entry giving the probabilities of each of the transmitted symbols based on the observation. Thus, in this example, $\gamma_i$ would indicate that based on the received symbol, the probability that the first symbol was transmitted was 0.7, while the probabilities that the other 3 symbols were transmitted were each 0.1.

In performing the E-step, conditioned on the observation vector, $\mathbf{y}$, and estimates from the previous iteration, $\Theta_{k-1}$, the only random part of the log likelihood function is $\mathbf{x_i}^T$. Thus the E-step updates $\mathbf{x_i}^T$ using Eqn. 3.18. After the E-step $\mathbf{x_i}^T$ are *soft* decision vectors and not the hard decision vectors as discussed above.

$${\gamma_{i,l}}^{+} = E[x_{i,l}|\mathbf{y}, \Theta_{k-1}] \tag{3.18a}$$

$$= \frac{exp\left(-\|y_i - h_{k-1}c_l\|^2/\sigma_{k-1}^2\right)}{\sum_{l=1}^{M} exp\left(-\|y_k - h_{k-1}c_l\|^2/\sigma_{k-1}^2\right)} \tag{3.18b}$$

The M-step is used to update the estimates. Since the two parameters, $h$ and $\sigma^2$ are independent, the following optimization equations follow:

$$\frac{\partial Q}{\partial h} = \frac{\partial}{\partial h}\sum_{i=1}^{N}\sum_{l=1}^{M}\gamma_{i,l}^{+}(\log\sigma^2 + \frac{\parallel y_i - hc_l \parallel^2}{\sigma^2}) = 0 \tag{3.19a}$$

$$\Longrightarrow \sum_{i=1}^{N}\sum_{l=1}^{M}\gamma_{i,l}^{+}(2h\|c_l\|^2 - (y_k c_l^* + c_l y_k^*) = 0 \tag{3.19b}$$

$$\frac{\partial Q}{\partial \sigma^2} = \sum_{i=1}^{N}\sum_{l=1}^{M}x_{k,l}^{+}(\frac{1}{\sigma^2} - \frac{\parallel y_i - hc_l \parallel^2)}{\sigma^4}) = 0 \tag{3.19c}$$

These yield the following solutions where the + sign indicates the updated parameters of the i-th iteration, * indicates conjugation and Re indicates the real part of a complex number.

$$h^{+} = \frac{\sum_{i=1}^{N}(\gamma_{\mathbf{i}}^{T})^{+}\mathrm{Re}(y_i\mathbf{c}^{*})}{\sum_{i=1}^{N}(\gamma_{\mathbf{i}}^{T})^{+}\|\mathbf{c}\|^{2}} \tag{3.20a}$$

$$\sigma^{2+} = \frac{1}{2N}\sum_{i=1}^{N}(\gamma_{\mathbf{i}}^{T})^{+} \parallel y_i - h^{+}\mathbf{c} \parallel^{2} \tag{3.20b}$$

$$\widehat{\rho}^{+} = \frac{(S^{+})^{2}}{\sigma^{2+}} \tag{3.20c}$$

## 3.6 Implementation Issues

The SNR estimator using the EM algorithm has been targeted for design on field programmable gate arrays (FPGA). The equations that need to be implemented in hardware are Eqns. 3.18 and 3.20. Of these equations, Eqn. 3.18 is much more challenging to implement in digital hardware due to two reasons. First, there is the need to carry out $N$ divisions for the normalization of each $\gamma_i$ and a division by $\sigma^2$, thus requiring $(N+1)$ divisions for each iteration. It is assumed that since FPGAs typically have a fairly large number of dedicated multipliers, first the quantity $1/\sigma^2$ will be calculated (1 division) and then multipliers will be used to carry out the rest of eqn. 3.18. Secondly, there is the need to update all $4N$ entries of $\gamma_{i,l}$ every iteration. For efficient implementation, the exponential functions could be implemented as look up tables in FPGA block memory. Eqn. 3.20 is far simpler to implement assuming dedicated multipliers and adders are available in the FPGA. Even with modern FPGAs, an iterative algorithm like the one outlined in this chapter is challenging to implement considering that the design is required to support hundreds of SMs on a particular channel. A simpler soft decision based SNR estimation algorithm is outlined below.

### 3.6.1 Simpler SNR estimation algorithm

#### Using known preamble

This algorithm takes advantage of the presence of some known data symbols in the bursts being used for SNR estimation. Recall (from Sec. 3.2.1) that in the system of interest, 110 symbols of each burst is a known preamble. In general, the target system is assumed to know $N_p$ of the $N$ symbols in the burst. This is a reasonable assumption as the vast majority of burst-based systems will have a preamble in the burst for acquisition purposes. Thus in generating the $N \times 4$ entries of Eqn. 3.18, the first $M$ rows are hard decision values and do not require any iteration.

#### Quantizing probability

Further simplifications may be made by quantizing the probability of a particular observation coming from one of the transmitted symbols into one of $P$ levels. A particularly efficient scheme uses only four levels and thus the ${x_{k,l}}^+$s of eqn. 3.18 are quantized to one of four values: 0, $\frac{1}{4}$, $\frac{1}{2}$ and 1. The case where the probabilities are quantized to only two levels (0 and 1) is the degenerate case of using only hard decisions instead of using soft-decisions.

Quantizing the probabilities divides up the received I-Q space into what we call *zones of uncertainties*. This is shown in Fig. 3.4. Thus for instance, if the received symbol is detected in the top right hand corner of the I-Q map near $c_1$, $\gamma_i$ for that symbol is quantized to $[1, 0, 0, 0]^T$.

This greatly simplifies the algorithm since it obviates the need for lookup tables and divisions in the implementation of Eqn. 3.18. Moreover, calculating $\gamma_i$ does not depend on the current estimate of the parameter set to be estimated, $\mathbf{\Theta}$. Thus the algorithm is no longer an iterative one. This simplification comes at the cost of some degradation in performance in comparison to the full soft-decision iterative EM algorithm. The actual performance degradation is however small (in the SNR ranges of interest) and is discussed in Sec. 3.7.

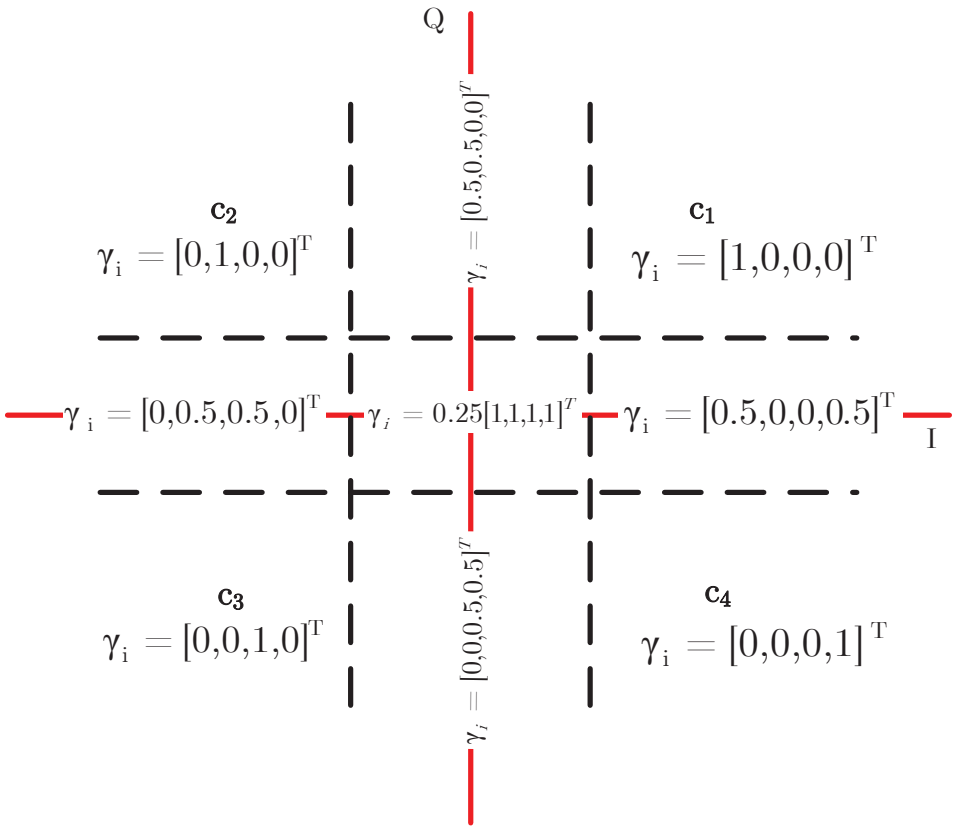


**Figure 3.4**: Quantized soft decisions: Zones of uncertainties. I/Q axis shown in red.

## 3.7 Simulation Results

### 3.7.1 Evaluation of CRLB for SNR Estimation

The CRLB was computed for various constellations. As mentioned above, Monte Carlo simulations were carried out by generating samples of $y_k$ in computer simulations. In each simulation, one million samples of $y_k$ were generated for determining the elements of the FIM from Eqn. 3.12. Each constellation point was assumed to be equally likely. The estimation was carried out with blocks of length $N = 100$ symbols. The CRLB is inversely proportional to $N$ and thus the CRLB for other block sizes can be easily found.

As a comparison, the CRLB for QPSK modulated signals presented by Alagha in [3] have also been calculated. They are seen to be the same as the values calculated using our method.

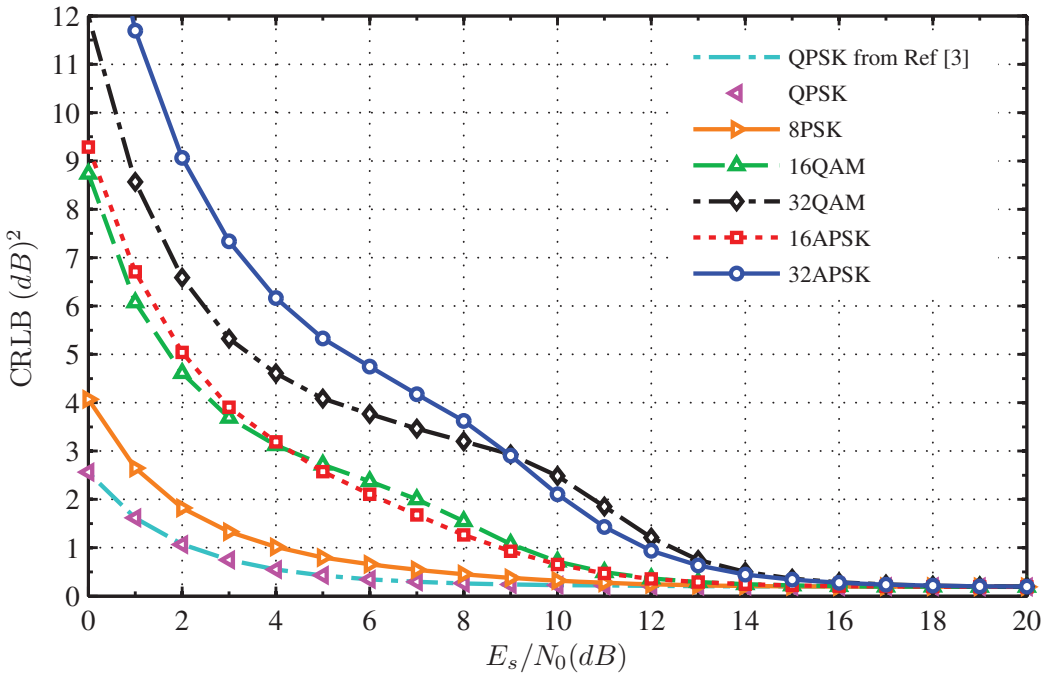


**Figure 3.5**: CRLB for NDA SNR estimation. $N = 100$

The results are shown in Fig. 3.5 and also presented in tabular form for ease of readability in Table 3.2 in Appendix 3.B.

### 3.7.2 EM based estimator

The EM based estimator was evaluated for two exemplary constellations: QPSK and 16APSK. The mean squared error (MSE) was calculated by running 10,000 bursts (each with $N = 100$ symbols) through a Monte Carlo simulation. The initial condition of the channel gain, $h$, and the noise variance $\sigma^2$ were both set to 1. We have found that, although the choice of the initial condition affects the convergence time, it does not affect estimator accuracy since the likelihood function has one maximum as discussed later.

Fig. 3.6 shows the MSE of the EM algorithm based NDA SNR estimator. As can be seen, the estimator performance is very close to the CRLB at all SNRs of interest. It should be pointed out that at SNRs of 2dB or less, the ML solution (that the EM algorithm converges to) is sometimes an outlier. In other words, the ML solution yields an SNR that is far lower than the true SNR. In order to prevent this, the estimator has been clamped to a lower limit of $-5$dB. This is unlikely to have much practical significance since 16APSK and QPSK modulations are typically not used at such low SNRs.

For comparison purposes, in Fig. 3.6, we have also shown the performance of four NDA SNR estimators available in the literature: a) for QPSK: i) $M_2M_4$ NDA SNR estimator (Eqn. 7 of [47] or Eqn. 39 of [72]) ii) $\rho_{12}$ estimator from [47] and iii) the iterative estimator from [5] and b) for 16APSK: 6th order estimator from [65]. As can be seen the iterative EM based estimator performs significantly better than all these estimators. In addition, some of these other estimators work only for MPSK modulation schemes. This performance gain does not, of course, come for free. The proposed EM based estimator is more computationally intensive. Whether the additional complexity is worth the performance gains will depend on the application at hand.

**Dependence on Initial Condition**

One of the known drawbacks of the EM algorithm is that it converges to the local maxima in the neighborhood of the initial condition. Thus, if the log likelihood function has multiple peaks, the EM algorithm is not guaranteed to converge to the maximum likelihood (ML) solution.

Fig. 3.7 shows the contour plot of the log likelihood function for 16APSK at SNR = 2dB together with a trajectory of estimates of $h$ and $\sigma^2$. It can be seen that there is only one maximum which ensures that the EM algorithm converges to the ML solution. This is also true for QPSK. Our results show that the terrain of the contour plot is flatter for 16APSK in comparison to QPSK. This explains why 16APSK is slower to converge than QPSK as shown in Fig. 3.8. The log likelihood is also seen to have rather high values in the region of low $h$ and high $\sigma^2$, which

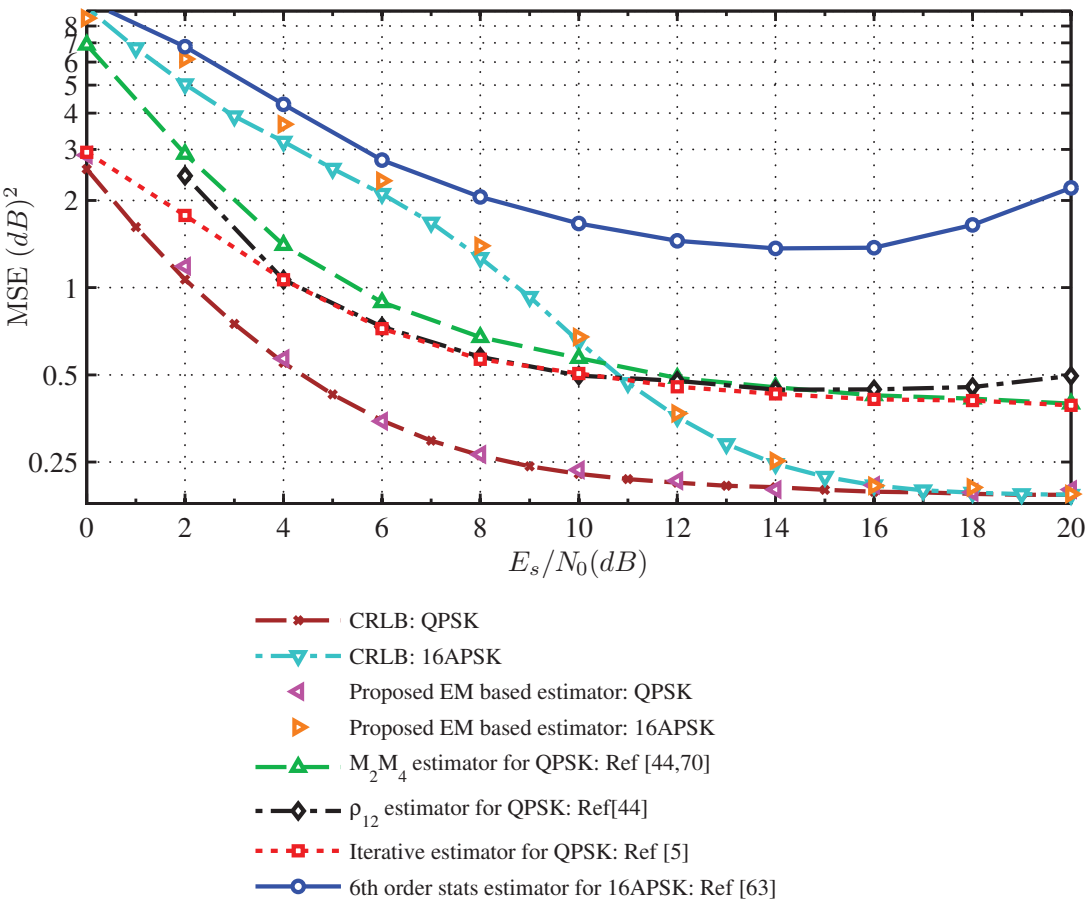


**Figure 3.6**: MSE performance of EM based NDA SNR estimator compared to previously published results

corresponds to low SNRs. This explains why at low SNRs (2dB or less), the ML estimate is occasionally an outlier (an estimate that is significantly less than the actual SNR), as discussed above.

### Convergence Speed

Fig. 3.8 shows the mean squared error of the EM based SNR estimator as a function of iteration number for both QPSK and 16APSK at SNRs of 2dB and 20dB. It can be observed that the convergence rate of 16APSK is slower than that of QPSK. At high SNRs, the algorithm converges faster than at low SNRs. Variants of the EM algorithm and better initialization might be used to speed up

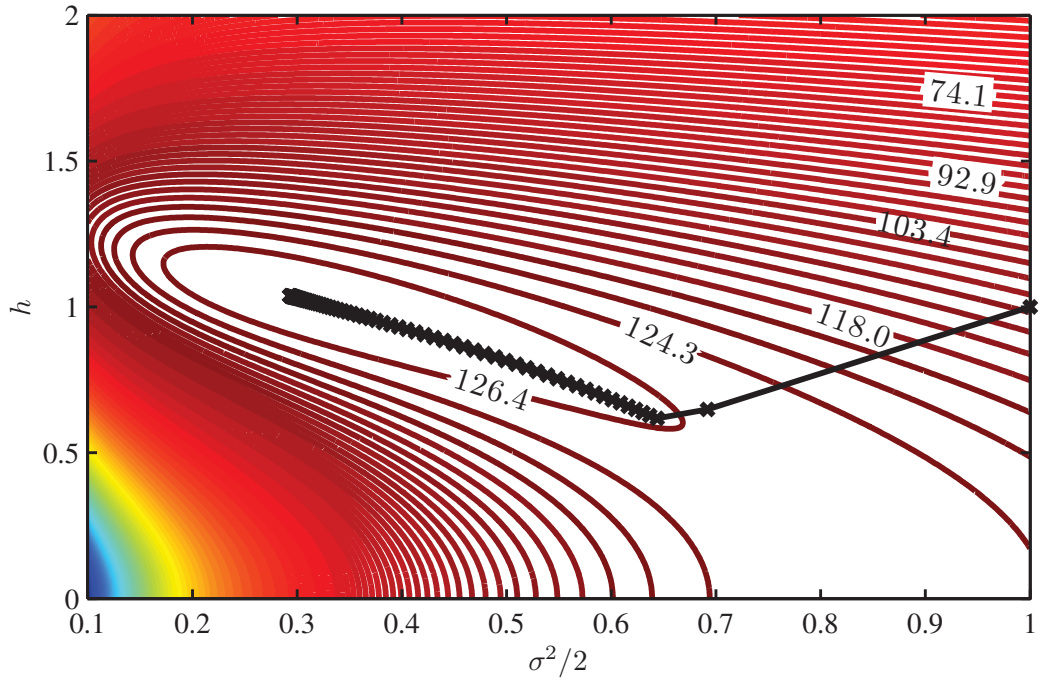


**Figure 3.7**: Contours of log likelihood function with estimator trajectory. Modulation: 16APSK. SNR = 2dB. $h = 1$. $\sigma^2 = 0.63$.

the convergence rates, but this is outside the scope of this chapter.

### 3.7.3 Performance of Non-Linearity Estimation Techniques

Fig. 3.9 shows the root mean square error (rmse) of different SNR estimators (for $N = 398$ symbols) compared to the CRLB (evaluated in [3]) for the two cases: i) where all the transmit data symbols are known by the SNR estimator (the data aided (DA) model) as well as ii) the NDA model where the SNR estimator does not know the transmitted data symbols. The initial condition for $\mathbf{\Theta}$ in the EM algorithm is set to $h = 0.5$, $\alpha_l = 0.25$, $\sigma^2 = 0.5$. It is seen that the EM algorithm achieves the CRLB for the NDA model. Fig. 3.9 shows the results after 10 iterations, however, 4 or 5 iterations are sufficient to achieve good results (for $E_s/N_0 > 5dB$, just 2 iterations are sufficient).

The performance of three reduced complexity estimators are also plotted. These estimators differ only in the way Eqn. 3.18 is updated. It is seen that

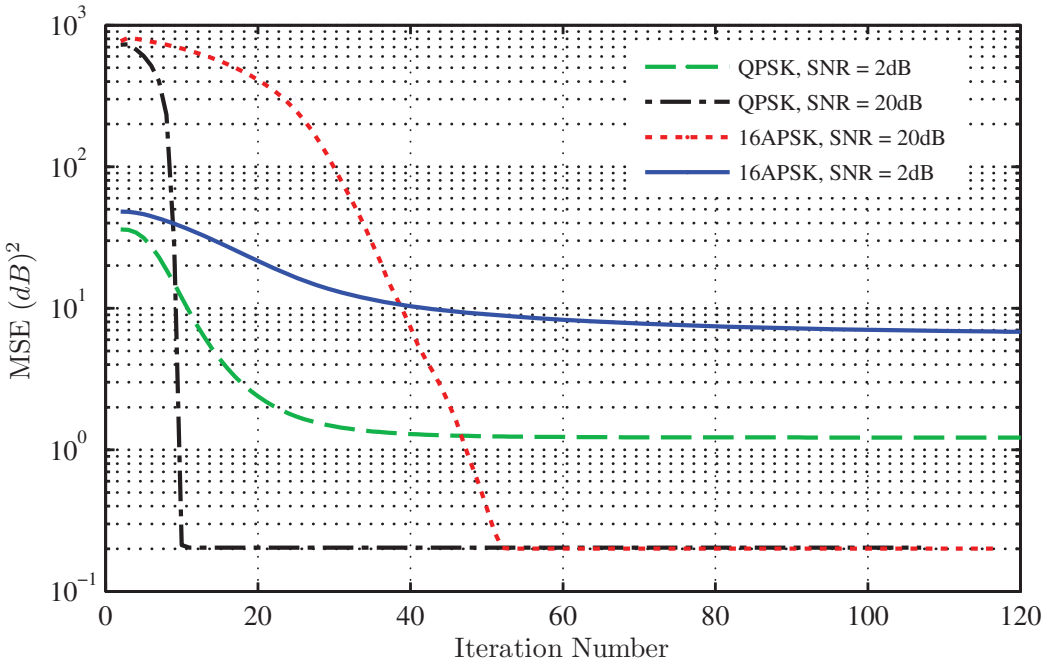


**Figure 3.8**: Convergence speed of EM based estimator

an iterative EM algorithm that takes advantage of 110 known symbols (See Sec. 3.2.1) in the preamble performs better than the CRLB of the NDA model. This is expected, since now some of the symbols are known. As expected, the 4-level soft-quantized ${\gamma_i}^+$ based non-iterative algorithm performs better than the 2-level soft-quantized algorithm (See Sec. 3.6.1). It is interesting to note that the 4-level quantized soft decision based model actually beats the CRLB for the NDA model for $E_s/N_0$ between 7 and 10dB. Again, this is due to the fact that the 110 of the symbols are assumed to be known perfectly.

Fig. 3.10 shows the cumulative distribution function (CDF) of the detection of the P1dB point relative to the actual P1dB point after adjusting for a bias in the switchpoint. It may be seen that around 80% of the time, the algorithm detects the onset of non-linearity within within $\pm 1$ dB of the actual P1dB point of the PA.

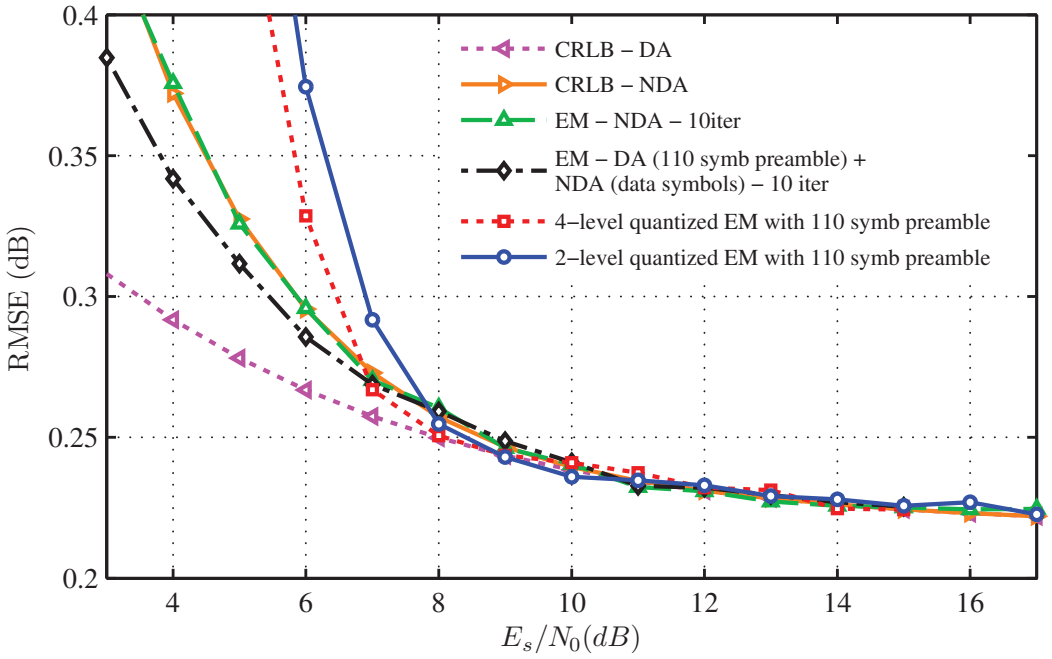


**Figure 3.9**: Performance of different QPSK SNR estimators. $N = 398$ (110 symbol preamble).

## 3.8 Conclusions

In this chapter we have evaluated the CRLB for NDA SNR estimation in an AWGN channel for the following QPSK, 8PSK, 16QAM, 32QAM, 16APSK and 32APSK constellations for a block size of 100 symbols. The CRLB is inversely proportional to the number of symbols, thus enabling the CRLB to be easily determined for any other block size from our results. We have also presented an estimator based on the EM algorithm that may be used for any arbitrary constellation for NDA SNR estimation. Using QPSK and 16APSK as two exemplary constellations, the proposed estimator is shown to perform very close to the CRLB for all SNRs of practical interest.

We also presented a novel scheme for the detection of PA non-linearity and have demonstrated that this scheme can be used to detect the P1dB of an amplifier within 1dB of the actual P1dB in the presence of realistic fade rates. An iterative EM based algorithm (as well as low-complexity non-iterative variants) to estimate

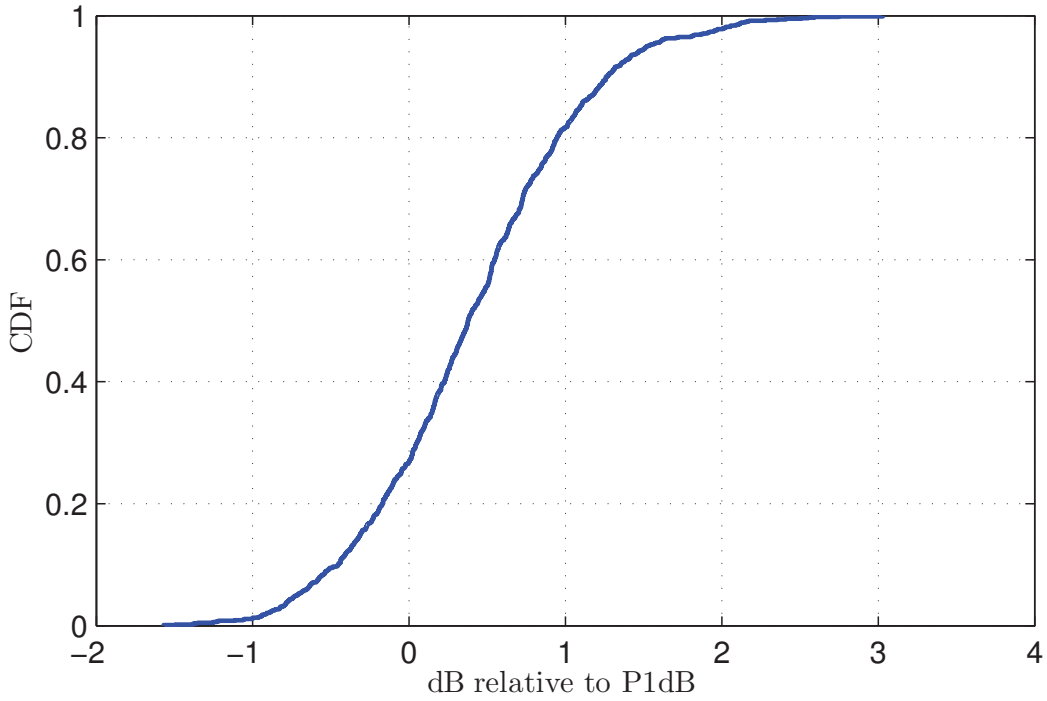


**Figure 3.10**: CDF of transmit power when compression is detected (relative to P1dB)

the SNR of QPSK modulated data has also been developed and is shown to attain the CRLB.

The algorithm has been applied in a DOCSIS based satellite broadband system, but the algorithm presented is independent of DOCSIS and can be used to detect the onset of non-linearity due to any reason at a remote terminal without adding to the hardware cost per terminal of the system.

Although this chapter dealt with the common case of single antenna systems with coherent reception, modern communication techniques commonly use multiple antennas both at the transmitter as well as at the receiver and the channel gain is often complex. The results of this chapter are more generalized to the multi antenna systems with complex channel gain in Chapter 4.

The contents of this chapter have been adapted from A. Das and M. Miller, "Remote non-linearity detection via burst power dithering and EM based SNR Es-

timation," Satellite and Space Communications, 2007. IWSSC 2007. International Workshop on, Sept. 2007 and from the work which lead to US Patent 7,894,510, issued July 05, 2011. Sections of this chapter have also been adapted from A. Das, "NDA SNR estimation: CRLBs and EM based estimators," IEEE Region 10 Conference, TENCON 2008.

## 3.A Appendix: Constellation Points

The constellation points used for various constellations are shown in Table 3.1. All the constellations are normalized to have unit energy.

**Table 3.1**: Constellation points

| **Constellation** | **Symbol Values** |
|---|---|
| QPSK | $\exp(j\pi m/4)$ where m = 1,3,5,7 |
| 8PSK | $\exp(j\pi m/8)$ where m = 1,3,5...15 |
| 16QAM | $\frac{1}{\sqrt{10}}\{\pm 1 \pm j, \pm 3 \pm 3j, \pm 1 \pm 3j, \pm 3 \pm j\}$ |
| 32QAM | $\frac{1}{\sqrt{20}}[\pm 1 \pm j, \pm 3 \pm 3j, \pm 1 \pm 3j, \pm 3 \pm j]$, $\frac{1}{\sqrt{20}}[\pm 5 \pm 3j, \pm 5 \pm j, \pm 3 \pm 3j, \pm 1 \pm 5j]$ |
| 16APSK | $\{R_1 \exp(j\pi m/4), R_2 \exp(j\pi n/12)\}$ m = 1,3,5,7. n = 1,3,...23 $R_1 = 0.4109, R_2 = 1.1301$ |
| 32APSK | $\{R_1 \exp(j\pi p/4), R_2 \exp(j\pi q/12), R_3 \exp(j\pi r/8)\}$ p = 1,3,5,7. q = 1,3,..23, r = 1,2,3..16, $\gamma_1 = 2.64, \gamma_2 = 4.64, R_2 = \gamma_1 R_1, R_3 = \gamma_2 R_1,\ R_1 = \sqrt{(\frac{8}{1+3\gamma_1^2+4\gamma_2^2})}$ |

The 16-APSK and 32-APSK constellations have been used as defined by the DVB-S2 standard [43].

## 3.B Appendix: CRLB for Various Constellations

Table 3.2 shows the CRLB for various constellations for SNR estimation with an NDA model for a block size of 100 symbols. Some researchers have published the normalized CRLB (NCRLB) where NCRLB = $\frac{\text{CRLB}}{\text{SNR}^2}$ where the CRLB

**Table 3.2**: CRLB ($dB^2$) for N=100

| SNR (dB) | QPSK | 8PSK | 16QAM | 32QAM | 16-APSK | 32-APSK |
|---|---|---|---|---|---|---|
| 0 | 2.56 | 4.07 | 8.73 | 12.02 | 9.29 | 16.37 |
| 2 | 1.07 | 1.82 | 4.61 | 6.58 | 5.04 | 9.06 |
| 4 | 0.55 | 1.02 | 3.12 | 4.61 | 3.19 | 6.16 |
| 6 | 0.35 | 0.66 | 2.38 | 3.77 | 2.11 | 4.75 |
| 8 | 0.26 | 0.45 | 1.55 | 3.20 | 1.27 | 3.62 |
| 10 | 0.23 | 0.32 | 0.71 | 2.48 | 0.65 | 2.11 |
| 12 | 0.21 | 0.24 | 0.36 | 1.21 | 0.36 | 0.94 |
| 14 | 0.20 | 0.21 | 0.25 | 0.50 | 0.25 | 0.45 |
| 16 | 0.20 | 0.20 | 0.21 | 0.28 | 0.21 | 0.28 |
| 18 | 0.19 | 0.19 | 0.20 | 0.21 | 0.20 | 0.22 |
| 20 | 0.19 | 0.19 | 0.19 | 0.20 | 0.19 | 0.20 |

and SNR are both expressed in a linear scale. For convenience, we also present a table of NCRLBs in Table 3.3. The CRLBs are inversely proportional to the block size and thus the CRLBs for any other block size can be easily determined.

**Table 3.3**: NCRLB = $\frac{\text{CRLB}}{\text{SNR}^2}$ for N=100

| SNR (dB) | QPSK | 8PSK | 16QAM | 32QAM | 16-APSK | 32-APSK |
|---|---|---|---|---|---|---|
| 0 | 0.135 | 0.214 | 0.460 | 0.638 | 0.494 | 0.862 |
| 2 | 0.057 | 0.096 | 0.246 | 0.348 | 0.263 | 0.483 |
| 4 | 0.029 | 0.054 | 0.164 | 0.245 | 0.168 | 0.326 |
| 6 | 0.018 | 0.035 | 0.127 | 0.199 | 0.111 | 0.251 |
| 8 | 0.014 | 0.024 | 0.082 | 0.171 | 0.068 | 0.190 |
| 10 | 0.012 | 0.017 | 0.038 | 0.130 | 0.035 | 0.113 |
| 12 | 0.011 | 0.013 | 0.019 | 0.064 | 0.019 | 0.050 |
| 14 | 0.011 | 0.011 | 0.013 | 0.026 | 0.013 | 0.024 |
| 16 | 0.011 | 0.011 | 0.011 | 0.015 | 0.011 | 0.015 |
| 18 | 0.010 | 0.010 | 0.010 | 0.011 | 0.010 | 0.011 |
| 20 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |

# Chapter 4

# SNR Estimation for MIMO Systems

## 4.1 Introduction

As described in Chapter 3, many modern communication systems often need accurate estimates of signal to noise ratios (SNR) to carry out important functions such as adaptive coding and modulation [22], [30]. In Chapter 3 we analyzed single antenna systems with coherent reception. However, as multiple input, multiple output (MIMO) systems gain popularity, adaptive MIMO modulations are becoming more popular [38] and these systems also require accurate SNR feedback. In this chapter, we shall generalize the results of Chapter 3 to multiple antenna systems and to systems where the channel gain is a complex quantity.

MIMO systems are often used in conjunction with orthogonal space time block codes (OSTBC) and in [60] the authors show how the appropriate constellation for transmission is chosen based on estimating the ratio of the Frobenius norm of the channel to the noise variance. This ratio, which is directly analogous to the SNR in single antenna systems, plays a direct role in determining the maximum rate that may be transmitted over the channel since it plays a role in the formula for the capacity of the channel [81]. This definition of the received SNR for MIMO systems is also used by [21] and by [104] and thus must be estimated in practical

systems.

Typically, the MIMO SNR has to be determined at the receiver by estimating both the channel and the noise variance. However, sometimes the receiver needs to estimate only the noise variance since the channel estimates are made by the transmitter as in MIMO time division duplexing (TDD) systems. In these TDD systems with channel reciprocity, feedback from the transmitter is used to send the receiver estimates of the channel. Initially it was thought that the difference in the transmit and receive chains was enough to ensure that the overall measured channel was not reciprocal, but recent experimental results have showed that this difference may be accounted for by doing relative calibration [57], [50]. Even when the channel estimates are fed back from the transmitter (and thus may be assumed to be known at the receiver), the receiver needs to estimate the received noise variance accurately.

Traditionally, both the noise variance as well as the SNR are estimated by using the pilot (or preamble) sequences typically embedded in the signal. This, of course, leads to a loss of spectral efficiency and thus there is interest in estimates obtained using the transmitted data in a non-data aided (NDA) manner. In our work, we shall address both the the DA as well as the NDA approaches.

### 4.1.1 Overview of Prior Work

CRLBs of NDA SNR estimation and efficient estimators for various constellations in single input, single output (SISO) systems have been evaluated in [28] and discussed in detail in Chapter 3. These have been extended to single input multiple output (SIMO) systems in [16] for both data aided (DA) as well as NDA models. In [16], the authors assume that the non-random phase shifts are perfectly known, thus assuming that the SIMO channels are all real valued. In this current work, we present the CRLB for SNR estimation for MIMO systems for complex valued channels. We shall show that the SISO and SIMO results previously published are special cases of this more general result. Unlike [16], our determination of the CRLB for SNR uses complex valued parameters (See Section 4.7.3 of [90]).

Prior to our work in [29] there had been relatively little published work

in the area of NDA MIMO SNR estimation. This chapter is adapted from this work [29]. Noise variance and SNR estimation algorithms using training sequences for MIMO orthogonal frequency division multiplexing (OFDM) systems have been published in [17] and [45] and references therein. These prior works are based on known training symbols and are thus in a DA framework. Moreover, they present algorithms for estimation without presenting the Cramér-Rao bound which is one of the contributions of this current work.

The CRLB for channel estimation for MIMO systems has been presented in [92] and [13]. The differences between these and similar works and ours are the following: a) we assume that the noise variance is also unknown whereas [92] assumes that the noise variance is known, b) we estimate the CRLB for the ratio of the Frobenius norm of the channel to the noise variance (the MIMO SNR) while the cited works evaluate the channel and c) we evaluate the NDA model where the unknown transmitted symbols are assumed to be from a finite digital constellation while the above works evaluate the CRLB in the DA model.

The estimation of noise variance has attracted some recent interest in the literature. Jiang et al., develop a polynomial based noise variance estimator for known pilots in [56]. Benjebbour and Yoshida develop accurate noise variance estimators for ordered successive interference cancellation MIMO receivers in [12] but this depends on training symbols as well. In [12] the authors first determine a noise variance by using the training symbols in a DA model and then update that estimate using the estimates of the data symbol obtained from the initial estimate of the noise variance.

Blind noise estimation without using any preambles (NDA model) was presented in [86] for an OFDMA system utilizing the time frequency sparsity of the OFDMA signal as well as the redundancy introduced by the cyclic prefix. In the OFDM framework, considerable work has taken place in the area of training based noise variance estimation, (see [18, 45, 103] and references therein). Unique word (or preamble) based noise variance estimation for single carrier systems was also presented by Coon and Beach in [23]. In the ultra wideband (UWB) channel model, noise variance estimation has been studied recently in [93]. In these prior

works, however, the CRLB bound for NDA MIMO noise variance estimation has not been evaluated.

In Chapter 3, we discussed the expectation maximization (EM) algorithm, which is known (for regular exponential family distributions) to iteratively attain the maximum likelihood (ML) estimate [35]. The EM algorithm has been used in similar NDA parameter estimation problem for SNR estimation in SISO and SIMO systems in [15, 28, 32]. In this chapter, we apply the EM algorithm to a similar, but different problem of estimating the MIMO SNR and the noise variance at each received antenna. We show that the obtained estimator achieves the CRLB for BPSK, QPSK, 8PSK and 16QAM modulation schemes over a wide range of $E_s/N_0$.

The work that comes closest to ours is [95], where the authors present EM based estimators for the NDA model for joint channel and noise variance estimation for MIMO systems. However, the authors do not present the CRLB, nor do they present closed form approximations for the ML estimator in the NDA case. To the best of our knowledge, our work, [29], was the first that presented bounds and estimators for NDA SNR estimation for a MIMO single carrier system with block fading employing M-ary modulation scheme.

### 4.1.2 Our Contribution

Our contribution can thus be summarized to be the following. We present CRLBs for both MIMO SNR estimation as well as MIMO noise variance estimation for both DA and NDA system models. We also present the CRLB for SNR estimation for the special case of when the transmitters use the widely used Alamouti code. These bounds are derived for a couple of different noise models – one that assumes that the noise variance is the same across all received antennas and the other assumes that the noise variance of each received antenna is different. We show ML estimators for estimating both the MIMO SNR as well as the noise variance. We show closed form answers for the DA case and closed form approximations for the NDA case. The derivations are generalized for any M-ary constellation in a frequency flat block fading environment and simulations

are shown for BPSK, QPSK, 8-PSK and 16QAM modulation schemes. Although the focus of our work is on flat fading channels, we discuss how the results can be extended to frequency selective channels.

This chapter is organized in the following sections. In Section 4.2 we introduce the notation used in this chapter before developing the system model in Section 4.3. The CRLB for SNR estimation for both the DA and the NDA models are developed in Section 4.4 and the corresponding CRLBs for noise variance estimation are in Section 4.5. The estimators are presented in Section 4.7. Simulation results are presented in Section 4.8 before concluding.

## 4.2 Notation

Some of the common notation used throughout this chapter are introduced in this section. Vectors and matrices are represented respectively in lower case bold font and upper case bold font such as $\mathbf{h}$ and $\mathbf{H}$. $x_{i,l}$ is the l-th entry of the vector $\mathbf{x_i}$ where $i$ indicates the time index. $\mathbf{0}_{x,y}$ is an all-zeros matrix of size $x \times y$. $\mathbf{0}_x$ is an all-zeros matrix of size $x \times x$. $\mathbf{I}_x$ represents an identity matrix of size $x \times x$. $vec(\mathbf{H})$ represents all the elements of a matrix $\mathbf{H}$ rearranged in a column vector. $()^*$, $()^T$ and $()^H$ denote the conjugate, transpose and Hermitian operators respectively. $\mathcal{N}(\mu, \mathbf{R_{nn}})$ denotes the pdf of a circular Gaussian random variable of mean $\mu$ and covariance $\mathbf{R_{nn}}$. $\|\mathbf{H}\|_F^2$ denotes the Frobenius norm of matrix $\mathbf{H}$. The notation diag($\mathbf{x}$) is used to represent a $N \times N$ diagonal matrix whose entries are the N elements of the vector $\mathbf{x}$. diag($[\mathbf{A_1}, \mathbf{A_2}, ...\mathbf{A_N}]$) is a block diagonal matrix whose component matrices, $\mathbf{A_1}, \mathbf{A_2}, ..., \mathbf{A_N}$, form the successive diagonal elements.

## 4.3 System Model

For ease of exposition, the system model used is the block flat fading model where the channel, $\mathbf{H}$, is constant over a block of symbols of length N. The received vector, at time $i$, $\mathbf{y_i}$ is a $M_R \times 1$ vector, where $M_R$ is the number of received antennas. The transmit symbols at time $i$ are represented by a $M_T \times 1$ vector

$\mathbf{x_i}$ where $M_T$ is the number of transmit antennas. Each element of $\mathbf{x_i}$, i.e. each transmitted symbol is assumed to come from a digital constellation such as M-PSK or M-QAM. $\mathbf{H}$ is the $M_R \times M_T$ channel matrix where each element is assumed to be independent and is drawn from a zero mean, unit variance complex Gaussian distribution. $\mathbf{n_i}$ is the $M_R \times 1$ noise vector, also assumed to be complex Gaussian and spatially and temporally uncorrelated. Assuming perfect synchronization and no differential path delays between the various antennas, the received samples after matched filtering and downsampling are given by Eqn. 4.1.

$$\mathbf{y_i} = \mathbf{H}\mathbf{x_i} + \mathbf{n_i} \quad i = 1...N \tag{4.1}$$

The $N$ transmitted symbol vectors of the block, may be collected together for notational convenience in a matrix as shown in Eqn. 4.2.

$$\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}] \tag{4.2}$$

Two models for the noise vector will be analyzed. In the first model, the noise variance $\sigma^2$ is the same across all received antennas and thus the noise probability density function (pdf) is given by $\mathcal{N}(\mathbf{0}_{M_R,1}, \sigma^2\mathbf{I}_{M_R})$. In the second model, each received antenna has a different noise variance, but the noise is assumed to be spatially and temporally uncorrelated thus resulting in the pdf $\mathcal{N}(\mathbf{0}_{M_R,1}, \mathrm{diag}(\widetilde{\sigma}^2))$ where $\widetilde{\sigma}^{\mathbf{2}} = [\sigma_1^2, \sigma_2^2, ..., \sigma_{M_R}^2]^T$.

### 4.3.1 Definition of SNR

SNR is nominally defined as the ratio of the signal power to the noise power (measured in the same bandwidth). In SISO and SIMO systems the definition is straightforward, since at each receiver there is only one desired signal. For MIMO systems, however, the various transmit signals linearly add at each receiver, and thus there are multiple desired signals at each receive antenna. Thus, we can define a pairwise SNR between transmitter $y$ and receiver $x$ as $\rho_{xy} = \frac{\|h_{xy}\|^2}{\sigma^2}$ where $h_{xy}$ is the complex path gain between the $x$-th receiver and the $y$-th transmitter. We can also define a SNR based on the Frobenius norm of the channel as $\rho = \frac{\|\mathbf{H}\|_F^2}{\sigma^2}$. The Frobenius norm of the $M_R \times M_T$ channel matrix is defined in Eqn. 4.3.

$$\|\mathbf{H}\|_F^2 = \sum_{p=1}^{M_R}\sum_{q=1}^{M_T}\|h_{pq}\|^2 = \text{trace}(\mathbf{H}^H\mathbf{H}) \tag{4.3}$$

In this work we shall evaluate the CRLB for both these definitions of SNR though our emphasis will be on the latter definition. It should be noted that the second definition of SNR is compatible only with the first model of the noise distribution as described above. For the pairwise definition of SNR, if the second noise model is used, then $\sigma^2$ is replaced with $\sigma_x^2$ which is the noise variance seen on receiver $x$.

**Identifiability of SNR in NDA Estimation**

Certain constellations, e.g., 16-QAM or 32-QAM, contain constellation points that are scalar multiples of each other. In 16-QAM for example, the 4 inner constellation points are a scalar multiple of the 4 outer corner points. In a NDA estimation framework, where none of the transmitted symbols are assumed known, there exists a possible ambiguity in determining the channel gain since it cannot be determined whether the transmitted symbol was $x$ or $kx$ (where $k$ is the multiplicative scale factor between two constellation points that share this multiplicative relationship), and thus, the SNR cannot be uniquely identified. In [80], the authors refer to this as the *divisor ambiguity*. The authors show that when all the symbols of the constellation are equally likely, the probability of error due to this *divisor ambiguity* drops very quickly with increasing block size and even for blocks with only 10 symbols, the probability of error for 16-QAM is less than $10^{-5}$ (Fig 4 of [80]). Thus the practical limitation of this phenomenon in blind SNR estimation where block sizes are typically at least a few bytes long is expected to be negligible.

## 4.4 CRLB for SNR Estimation

### 4.4.1 Data Aided Model

In the DA model, we assume that the $N$ transmitted symbols, $\mathbf{x_i}$, $i = [1, 2, ..., N]$ come from a normalized constellation of unit power, are all known by the receiver and independent of the noise samples and the channel coefficients. The channel coefficients are assumed to be unknown, but complex valued deterministic constants. The parameter vector of interest in the estimation problem is then the $2M_RM_T + 1$ element vector:

$$\mathbf{\Theta} = \begin{bmatrix} (vec(\mathbf{H}))^T & (vec(\mathbf{H}^*))^T & \sigma^2 \end{bmatrix}^T \tag{4.4}$$

From Eqn. 4.1, the logarithm of the likelihood function may be expressed as:

$$g = \ln[p(Y; \mathbf{\Theta})] = -NM_R \ln(\pi) - NM_R \ln(\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^{N} \phi_i \tag{4.5}$$

where $\phi_i = (\mathbf{y_i} - \mathbf{Hx_i})^H(\mathbf{y_i} - \mathbf{Hx_i})$. In order to determine the CRLB, we shall determine the individual elements of the Fisher information matrix (FIM) which is given by $\mathbf{F_\Theta} = \mathbf{E}\left[\left(\frac{\partial g}{\partial \mathbf{\Theta}^*}\right)\left(\frac{\partial g}{\partial \mathbf{\Theta}^*}\right)^H\right]$ [90]:

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial \sigma^2}\right)\right] = \frac{NM_R}{\sigma^4} \tag{4.6a}$$

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial h_{xy}}\right)\right] = \mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial h^*_{xy}}\right)\right] = \mathbf{E}\left[\left(\frac{\partial g}{\partial h_{xy}}\right)\left(\frac{\partial g}{\partial h_{pq}}\right)\right] = 0 \tag{4.6b}$$

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial h_{xy}}\right)\left(\frac{\partial g}{\partial h^*_{pq}}\right)\right] = \begin{cases} \frac{1}{\sigma^2}\sum_{i=1}^{N} x_{i,y}x^*_{i,q} & \text{if } x = p \\ 0 & \text{if } x \neq p. \end{cases} \tag{4.6c}$$

Using Eqn. 4.2, and defining the $M_T \times M_T$ sized matrix, $\mathbf{A} = \mathbf{XX}^H$, we can write the resulting $(2M_TM_R + 1) \times (2M_TM_R + 1)$ sized block diagonal FIM:

$$\mathbf{F}_{\mathbf{\Theta}} = \mathbf{E}\left[\left(\frac{\partial g}{\partial \Theta^*}\right)\left(\frac{\partial g}{\partial \Theta^*}\right)^H\right] = \frac{1}{\sigma^2}\left[\begin{array}{c|c} \mathrm{diag}([\mathbf{A}^*, \mathbf{A}^*, ..., \mathbf{A}, \mathbf{A}, ...]) & \mathbf{0}_{2M_TM_R,1} \\ \hline \mathbf{0}_{1,2M_TM_R} & \frac{NM_R}{\sigma^2} \end{array}\right] \tag{4.7}$$

In order for the CRLB to exist, $\mathbf{A}$ has to be invertible. For $x_{i,k}$s selected from a zero mean, unit energy constellation with equiprobable points, and independently chosen at each time instant, $i$, for large block size, $N$, we can show that $\sum_{i=1}^{N} x_{i,k}x_{i,k}^* \approx N$ and $\sum_{i=1}^{N} x_{i,j}x_{i,k}^* \approx 0$ for $j \neq k$. Thus, under these conditions, $\mathbf{A} \approx N\mathbf{I}_{M_T}$ where $\mathbf{I}_{M_T}$ is the identity matrix of size $M_T \times M_T$.

Since the CRLB is given by $\left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right)\mathbf{F}_{\mathbf{\Theta}}^{-1}\left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right)^H$ (see Section 4.7.3 of [90]) where $\rho$ is the function of the various parameters being estimated, if $\alpha$ is the y-yth entry of the inverse of $\mathbf{A}$, then the CRLB for the pairwise SNR $\rho_{xy}$ is given by:

$$\text{CRLB of } \rho_{xy} = 2\alpha\rho_{xy} + \frac{\rho_{xy}^2}{NM_R} \tag{4.8}$$

**DA CRLB for Pairwise SNR with different Noise Variance on each Received Antenna**

In this case, the parameter vector to be estimated is the $(2M_T + 1)M_R$ element vector:

$$\mathbf{\Theta} = \left[\begin{array}{ccccccc} (vec(\mathbf{H}))^T & (vec(\mathbf{H}^*))^T & \sigma_1^2 & \sigma_2^2 & ... & \sigma_{M_R}^2 \end{array}\right]^T \tag{4.9}$$

Using Eqn. 4.29 and 4.30, it may be shown that the FIM of Eqn. 4.7 becomes the $(2M_T + 1)M_R \times (2M_T + 1)M_R$ matrix

$$\mathbf{F}_{\mathbf{\Theta}} = \left[\begin{array}{c|c} \mathrm{diag}([\frac{\mathbf{1}}{\sigma_{\mathbf{1}}^{\mathbf{2}}}\mathbf{A}^*, \frac{\mathbf{1}}{\sigma_{\mathbf{2}}^{\mathbf{2}}}\mathbf{A}^*, ..., \frac{\mathbf{1}}{\sigma_{\mathbf{M_R-1}}^{\mathbf{2}}}\mathbf{A}, \frac{\mathbf{1}}{\sigma_{\mathbf{M_R}}^{\mathbf{2}}}\mathbf{A}]) & \mathbf{0}_{2M_TM_R,M_R} \\ \hline \mathbf{0}_{M_R,2M_TM_R} & \mathrm{diag}([\frac{\mathbf{N}}{\sigma_{\mathbf{1}}^{\mathbf{2}}}, \frac{\mathbf{N}}{\sigma_{\mathbf{2}}^{\mathbf{2}}}, ..., \frac{\mathbf{N}}{\sigma_{\mathbf{M_R-1}}^{\mathbf{2}}}, \frac{\mathbf{N}}{\sigma_{\mathbf{M_R}}^{\mathbf{2}}}]) \end{array}\right] \tag{4.10}$$

Thus the pairwise CRLB is given by:

$$\text{CRLB of } \rho_{xy} = 2\alpha\rho_{xy} + \frac{\rho_{xy}^2}{N} \tag{4.11}$$

If the parameter of interest is $\rho = \frac{\|\mathbf{H}\|_F^2}{\sigma^2}$, the CRLB is obtained in a straightforward manner, but the final form is dependent on the individual channel coefficients. For a $(M_T, M_R) = (1, 2)$ system, the parameter vector $\mathbf{\Theta}$ has 5 elements: 4 for the two channel coefficients and their complex conjugates and 1 for the noise variance. Then

$$\frac{\partial \rho}{\partial \mathbf{\Theta}^T} = \frac{1}{\sigma^2} \begin{bmatrix} h_{11}^* & h_{21}^* & h_{11} & h_{21} & -\frac{\|H\|_F^2}{\sigma^2} \end{bmatrix} \tag{4.12}$$

The matrix $\mathbf{A} = \mathbf{X}\mathbf{X}^H$ becomes the real scalar, $\sum_{i=1}^{N} x_{i,1} x_{i,1}^* = \alpha$. Then it may be shown that the CRLB is given by:

$$\left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right) \mathbf{F}_{\mathbf{\Theta}}^{-1} \left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right)^H = \tag{4.13a}$$

$$\begin{bmatrix} h_{11}^* \\ h_{21}^* \\ h_{11} \\ h_{21} \\ -\frac{\|\mathbf{H}\|_F^2}{\sigma^2} \end{bmatrix}^T \begin{bmatrix} 1/\alpha & 0 & 0 & 0 & 0 \\ 0 & 1/\alpha & 0 & 0 & 0 \\ 0 & 0 & 1/\alpha & 0 & 0 \\ 0 & 0 & 0 & 1/\alpha & 0 \\ 0 & 0 & 0 & 0 & \sigma^2/2N \end{bmatrix} \begin{bmatrix} h_{11}/\sigma^2 \\ h_{21}/\sigma^2 \\ h_{11}^*/\sigma^2 \\ h_{21}^*/\sigma^2 \\ -\frac{\|\mathbf{H}\|_F^2}{\sigma^4} \end{bmatrix}$$

$$= \frac{\rho^2}{2N} + \frac{2\rho}{\alpha} \tag{4.13b}$$

For a $(M_T, M_R) = (2, 2)$, $\mathbf{A} = \mathbf{X}\mathbf{X}^H$ is a $2 \times 2$ matrix and if $\mathbf{A}^{-1} = \begin{bmatrix} \alpha & \beta \\ \beta^* & \alpha \end{bmatrix}$, then the CRLB can be shown to be equal to $2\alpha\rho + \frac{\rho^2}{2N} + 4\mathbf{Re}\beta^*[h_{11}^* h_{12} + h_{21} h_{22}^*]$. As discussed above, for large N, $\mathbf{A} \approx N\mathbf{I}_{M_T}$ and thus, under those conditions, the CRLB is given by Eqn. 4.14.

$$\text{CRLB of DA} = 2\rho/N + \frac{\rho^2}{2N} \tag{4.14}$$

From the above equation and from Eqn. 4.8, it may be seen that at high SNRs, the normalized CRLB ($= \frac{CRLB}{\rho^2}$) approaches $\frac{1}{NM_R}$.

### 4.4.2 Non Data Aided Model

In the NDA model, the transmitted symbols, $\mathbf{x_i}$ are unknown. Assuming $\mathbf{x_i}$ comes from a constellation of $M$ distinct equally likely symbols, then the pdf of the observed sample $\mathbf{y_i}$ is given by Eqn. 4.15. Note that since $\mathbf{x_i}$ is a $M_T \times 1$ vector, if a QPSK constellation (4 possible symbols) is used with 2 transmit antenna, $M = 16$. In general, for a constellation with $\mathcal{P}$ discrete points, $M = \mathcal{P}^{M_T}$.

$$p(\mathbf{y_i}; \mathbf{\Theta}) = \sum_{l=1}^{M} p(\mathbf{y_i}|\mathbf{x_l}; \mathbf{\Theta}) p(\mathbf{x_l}) \tag{4.15a}$$

$$= \frac{1}{M} \sum_{l=1}^{M} \frac{1}{\pi^{M_R} \sigma^{2M_R}} \exp \frac{-(\mathbf{y_i} - \mathbf{H}\mathbf{x_l})^H (\mathbf{y_i} - \mathbf{H}\mathbf{x_l})}{\sigma^2} \tag{4.15b}$$

Thus, assuming the $\mathbf{y_i}$ are all independent and identically distributed, the logarithm of the likelihood function is given by

$$g = \ln(p(\mathbf{Y}; \sigma^2)) = -N \ln(M\pi^{M_R}) - NM_R \ln(\sigma^2) - \sum_{i=1}^{N} \left[ \ln[\sum_{l=1}^{M} \exp(\frac{\phi_{i,l}}{\sigma^2})] \right] \tag{4.16}$$

where

$$\phi_{i,l} = (\mathbf{y_i} - \mathbf{H}\mathbf{x_l})^H (\mathbf{y_i} - \mathbf{H}\mathbf{x_l}) \tag{4.17}$$

From Eqn. 4.16, the following partial derivatives may be evaluated:

$$\frac{\partial g}{\partial \sigma^2} = \frac{-NM_R}{\sigma^2} + \frac{1}{\sigma^4} \sum_{i=1}^{N} \frac{\sum_{l=1}^{M} \left( \phi_{i,l} \exp(\frac{-\phi_{i,l}}{\sigma^2}) \right)}{\sum_{l=1}^{M} \exp(\frac{-\phi_{i,l}}{\sigma^2})} \tag{4.18a}$$

$$\frac{\partial g}{\partial h_{xy}} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{\sum_{l=1}^{M} \left( x_{l,y} (y_{i,x} - (\mathbf{H}\mathbf{x_l})_x)^* \exp(\frac{-\phi_{i,l}}{\sigma^2}) \right)}{\sum_{l=1}^{M} \exp(\frac{-\phi_{i,l}}{\sigma^2})} \tag{4.18b}$$

$$\frac{\partial g}{\partial h^*_{xy}} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{\sum_{l=1}^{M} \left( x^*_{l,y} (y_{i,x} - (\mathbf{H}\mathbf{x_l})_x) \exp(\frac{-\phi_{i,l}}{\sigma^2}) \right)}{\sum_{l=1}^{M} \exp(\frac{-\phi_{i,l}}{\sigma^2})} \tag{4.18c}$$

In order to determine the CRLB, we need to evaluate the FIM given by $\mathbf{E}\left[ \left(\frac{\partial g}{\partial \mathbf{\Theta}^*}\right) \left(\frac{\partial g}{\partial \mathbf{\Theta}^*}\right)^H \right]$ where $E()$ is the expectation operator that averages over the noise vector and transmitted symbols. Unfortunately from Eqn. 4.18, there is no

easy way to evaluate analytically the terms of the FIM, however, this expectation may be written in the following form assuming that the transmitted symbols are all equally likely:

$$E[f(\mathbf{x_i}, \mathbf{n_i}, \sigma^2)] = \frac{1}{M} \sum_{x_l \in \mathcal{C}} \int_{-\infty}^{\infty} f(\mathbf{x_l}, \mathbf{n_i}, \sigma^2) p(\mathbf{n_i}) d\mathbf{n_i} \tag{4.19}$$

where $\mathcal{C}$ is the set of all possible transmitted symbol vectors of cardinality M. $p(\mathbf{n_i})$ is the probability density function of the $M_R \times 1$ noise vector (assumed to be complex Gaussian). Since the noise vector is assumed to be complex Gaussian, this integral can be decomposed into $2M_R$ multiple integrals each one integrating over the real or imaginary component of the noise vectors of each of the $M_R$ received antennas.

Integrals of the form $\int_{-\infty}^{\infty} f(x) \exp(-x^2) dx$ are well known to be numerically evaluated using the Gauss-Hermite quadrature. (See Section 25.4.46 of [1]). They are approximated as $\sum_{i=1}^{N} w_i f(x_i)$ where N is the order of the approximation and $x_i$ are i-th root of the Hermite polynomial of n-th order and $w_i$ are the appropriate weights. Using this approximation, the above integral in Eqn. 4.19 has been approximated as $2M_R$ finite summations. These integrals may also be evaluated using Monte Carlo methods. Both methods have been used and simulated and the results are shown in Sec. 4.8.

### High SNR Approximation

Although there is no closed form solution for the SNR estimation for MIMO NDA, it may be seen that at high SNRs, the NDA solution becomes equal to the DA solution derived above in Sec. 4.4.1. For high SNRs, $p(\mathbf{y_i}; \mathbf{\Theta} | \mathbf{x_k}) \neq 0$ implies that $p(\mathbf{y_i}; \mathbf{\Theta} | \mathbf{x_j}) \approx 0$ for $k \neq j$ since, for high SNRs, the received samples can be modeled to lie in non-overlapping noise spheres around each of the transmitted symbols. This approximation, also used in [82] for a similar approximation, albeit in a different context, is the same as assuming that the data symbols are known exactly – which is the DA system model.

### 4.4.3 Comparison to SISO and SIMO Results

Eqn. 4.8 can be compared to Eqn. 27 of [16] ($\text{CRLB}_{DA}(\rho) = \frac{\rho^2+2N_a\rho_i}{N_aN}$) where the authors evaluate the CRLB for DA SNR estimation for SIMO systems. In [16], the authors use $N_a$ to represent the number of receive antennas (which we call $M_R$). The authors also assume $\sum_{i=1}^{N}|x_i|^2 = N$, thus implicitly assuming an unit norm constellation with large $N$ where all symbols are equally likely and symbols are chosen independently. It may be seen that for SIMO systems $\mathbf{A}$, the $M_T \times M_T$ matrix is a $1\times 1$ scalar $\sum_{i=1}^{N} x_{i,1}x_{i,1}^*$. Thus, the SIMO CRLB for DA SNR estimation in [16] is a special case of the MIMO CRLB for DA SNR estimation as evaluated in Eqn. 4.8. Comparisons can also be made to the SISO DA SNR CRLB evaluated in Section III A of [48] and it can easily be seen that the SISO case is also a special case of the MIMO results derived above.

For the NDA case, since there is no closed form solution, no direct comparison can be made. However, by comparing the results in Table III of [28] to the results in the simulation section it may be seen that the results are identical. In Fig. 4.6, exemplary results from [28] are plotted alongside the MIMO results to show their similarity.

### 4.4.4 CRLB for SNR Estimation with Alamouti Coding

Probably the most common use of OSTBC is the use of the famous Alamouti code in a $(M_T, M_R) = (2, 2)$ system [89]. In this section we derive the CRLB for SNR estimation for such a system. The received samples in two consecutive time-slots, $j$ and $j+1$ may be combined together and put in the following form

$$\begin{bmatrix} y_1[j] \\ y_2[j] \\ y_1[j+1] \\ y_2[j+1] \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{H} \end{bmatrix} \begin{bmatrix} x_1[j] \\ x_2[j] \\ -x_2^*[j] \\ x_1^*[j] \end{bmatrix} + \begin{bmatrix} n_1[j] \\ n_2[j] \\ n_1[j+1] \\ n_2[j+1] \end{bmatrix} \tag{4.20a}$$

$$\widetilde{\mathbf{y}_\mathbf{i}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{x}_\mathbf{i}} + \widetilde{\mathbf{n}_\mathbf{i}} \tag{4.20b}$$

The received symbols from each group of two consecutive time slots, $j$ and

$j+1$, are collected into $\widetilde{\mathbf{y}}_\mathbf{i}$, where $i$ represents an index for each group of 2 time slots. Although similar in form to Eqn. 4.1, the number of parameters here is fewer. The unknown parameter vector has 9 unknowns:

$$\mathbf{\Theta} = \begin{bmatrix} h_{11} & h_{12} & h_{21} & h_{22} & h_{11}^* & h_{12}^* & h_{21}^* & h_{22}^* & \sigma^2 \end{bmatrix}^T \tag{4.21}$$

Using Eqn. 4.20 and Eqn. 4.5 with $M_R = 4$ since $\mathbf{y}_\mathbf{i}$ in Eqn. 4.20 is a $4 \times 1$ vector (even though the true number of received antenna is 2) we can write the following partial derivatives of the log likelihood function, $g$:

$$\frac{\partial g}{\partial \sigma^2} = \frac{-4N}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{N}\phi_i \tag{4.22a}$$

$$\frac{\partial g}{\partial h_{11}} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left[n_{i,1}^* x_{i,1} - n_{i,3}^* x_{i,2}^*\right] \qquad \frac{\partial g}{\partial h_{12}} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left[n_{i,1}^* x_{i,2}^* + n_{i,3}^* x_{i,1}^*\right] \tag{4.22b}$$

$$\frac{\partial g}{\partial h_{21}} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left[n_{i,2}^* x_{i,1} - n_{i,4}^* x_{i,2}^*\right] \qquad \frac{\partial g}{\partial h_{22}} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left[n_{i,2}^* x_{i,2} + n_{i,4}^* x_{i,1}^*\right] \tag{4.22c}$$

Assuming that the noise is zero mean and independent from sample to sample in both space and time, the following may be shown

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial \sigma^2}\right)\right] = \frac{NM_R}{\sigma^4} \tag{4.23a}$$

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial h_{xy}}\right)\right] = \mathbf{E}\left[\left(\frac{\partial g}{\partial \sigma^2}\right)\left(\frac{\partial g}{\partial h_{xy}^*}\right)\right] = \mathbf{E}\left[\left(\frac{\partial g}{\partial h_{xy}}\right)\left(\frac{\partial g}{\partial h_{pq}}\right)\right] = 0 \tag{4.23b}$$

$$\mathbf{E}\left[\left(\frac{\partial g}{\partial h_{xy}}\right)\left(\frac{\partial g}{\partial h_{pq}^*}\right)\right] = \begin{cases} \frac{1}{\sigma^2}\sum_{i=1}^{N}\left[|x_{i,1}|^2 + |x_{i,2}|^2\right] & \text{if } (x,y) = (p,q) \\ 0 & \text{if } (x,y) \neq (p,q). \end{cases} \tag{4.23c}$$

For the DA case, where the transmitted symbols are assumed to be known, it can be seen from Eqn. 4.23 that the FIM is a diagonal $9 \times 9$ matrix with zeros for all the terms except the main diagonal where the first eight terms are $\alpha = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(|x_{i1}|^2 + |x_{i2}|^2\right)$ and the last term is $\frac{4N}{\sigma^4}$ since $M_R = 4$.

If $\rho = \frac{\|\mathbf{H}\|_F^2}{\sigma^2}$, then it may be shown that:

$$\text{CRLB of } \rho = \frac{\|\mathbf{H}\|_F^2}{\sigma^2} = \left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right) \mathbf{F}_{\Theta}^{-1} \left(\frac{\partial \rho}{\partial \mathbf{\Theta}^T}\right)^H = \frac{\rho}{\alpha} + \frac{\rho^2}{4N} \tag{4.24}$$

## 4.5 CRLB for Noise Variance Estimation

In most of the following sections, it is assumed that the channel matrix, $\mathbf{H}$, is constant and known perfectly at the receiver. In Section 4.5.6, we deal with the case where the channel is random (but still known at the receiver).

### 4.5.1 DA Model with One Parameter

In the noise model with one parameter, $\sigma^2$, the CRLB for the estimation of $\sigma^2$ is shown in Eqn. 4.28. This is determined in a straightforward manner by taking the derivative of the log likelihood function as shown below

$$\ln(p(\mathbf{Y}; \sigma^2)) = \sum_{i=1}^{N} \ln(p(\mathbf{y_i}; \sigma^2)) \tag{4.25a}$$

$$= -N \ln(\pi^{M_R}) - N \ln \sigma^{2M_R} - \frac{1}{\sigma^2} \sum_{i=1}^{N} \underbrace{(\mathbf{y_i} - \mathbf{H}\mathbf{x_i})^H (\mathbf{y_i} - \mathbf{H}\mathbf{x_i})}_{\phi} \tag{4.25b}$$

$$= -N \ln(\pi^{M_R}) - N \ln \sigma^{2M_R} - \frac{1}{\sigma^2}\phi \tag{4.25c}$$

$$\frac{\partial^2 \ln p(\mathbf{Y}; \sigma^2)}{\partial (\sigma^2)^2} = \frac{NM_R}{\sigma^4} - \frac{2\phi}{\sigma^6} \tag{4.26}$$

Using

$$\text{CRLB} = \frac{1}{-E[\frac{\partial^2 \ln p(\mathbf{Y}; \sigma^2)}{\partial (\sigma^2)^2}]} \tag{4.27}$$

and the fact that $E[\phi] = \sigma^2$ we get

$$\text{CRLB}_{DA}[\sigma^2] = \frac{\sigma^4}{NM_R} \tag{4.28}$$

### 4.5.2 DA Model with $M_R$ Parameters

When using the noise model where the noise variance on each receive antenna is different, there are $M_R$ parameters to be estimated: $\mathbf{\Theta} = [\sigma_1^2, \sigma_2^2, ..., \sigma_{M_R}^2]^T$. In this case, it may be shown that the log-likelihood function is given by

$$\ln(p(\mathbf{y_i}; \boldsymbol{\sigma^2})) = -\ln(\pi^{M_R}) - \sum_{i=1}^{M_R} \ln \sigma_i^2 - (\mathbf{y_i} - \mathbf{Hx_i})^H \mathbf{R_{nn}}^{-1} (\mathbf{y_i} - \mathbf{Hx_i}) \quad (4.29)$$

where **Rnn** is given by $\mathcal{N}(\mathbf{0}, \mathrm{diag}(\widetilde{\sigma}^{\mathbf{2}}))$ as defined in Section 4.3. The CRLB is given by the inverse of the Fisher Information Matrix (FIM), whose j-kth entry is given by $-E[\frac{\partial^2 \ln p(\mathbf{y_i}; \boldsymbol{\sigma^2})}{\partial \sigma_j^2 \partial \sigma_k^2}]$. This may be evaluated using the following:

$$-E[\frac{\partial^2 \ln p(\mathbf{y_i}; \sigma^2)}{\partial (\sigma_k^2)^2}] = -E[\frac{1}{\sigma_k^4} - \frac{2(y_{i,k} - (\mathbf{Hx_i})_k)^H (y_{i,k} - (\mathbf{Hx_i})_k)}{\sigma_k^6}] = \frac{1}{\sigma_k^4} \quad (4.30)$$

In this notation, $y_{i,k}$ is the scalar quantity that is the k-th entry of the vector $\mathbf{y_i}$ and $(\mathbf{Hx_i})_k$ is similarly the k-th entry of $\mathbf{Hx_i}$. Also, noting that $\frac{\partial^2 \ln p(\mathbf{Y}; \sigma^2)}{\partial \sigma_k^2 \partial \sigma_j^2} = 0$ for $j \neq k$, we get the FIM as shown in Eqn. 4.31.

$$I_{DA} = \mathrm{diag}([N/\sigma_1^4, N/\sigma_2^4, ...N/\sigma_{M_R}^4]^T) \quad (4.31)$$

and thus the CRLB for the variance of the estimate of the k-th parameter, given by $kk$-th entry of the inverse of the Fisher Information Matrix is given by

$$CRLB_{DA_{M_R Parms}}[\sigma_k^2] = \frac{\sigma_k^4}{N} \quad (4.32)$$

### 4.5.3 NDA Model with One Parameter

In the following, we derive the CRLB for the non-data aided model where one parameter, $\sigma^2$, is to be estimated. Defining the scalar $\delta_{i,l} = 2\mathcal{R}(\mathbf{y_i Hx_l}) - (\mathbf{Hx_l})^{\mathbf{H}}(\mathbf{Hx_l})$ (where $\mathcal{R}(x)$ is the real part of $x$) and taking the derivative of the log-likelihood function, we get

$$\frac{\partial \ln p(\mathbf{y_i}; \sigma^2)}{\partial (\sigma^2)} = -\frac{M_R}{\sigma^2} - \frac{\mathbf{y_i^H y_i}}{\sigma^4} - \frac{\sum_{l=1}^{M} \frac{\delta_{i,l}}{\sigma^4} \exp(\frac{\delta_{i,l}}{\sigma^2})}{\sum_{l=1}^{M} \exp(\frac{\delta_{i,l}}{\sigma^2})} = f(\mathbf{x_i}, \mathbf{n_i}, \sigma^2) \quad (4.33)$$

The CRLB is given by the following equation,

$$CRLB_{NDA} = \frac{1}{E[(\frac{\partial \ln p(\mathbf{y_i};\sigma^2)}{\partial(\sigma^2)})^2]} \tag{4.34}$$

where the expectation in Eqn. 4.34 may also be evaluated using the Gauss Hermite Quadrature or by Monte Carlo techniques in computer simulation.

### 4.5.4 NDA Model with $M_R$ Parameters

Similar to the data-aided case, in this model we have the $M_R$ parameter vector $\mathbf{\Theta}$. To obtain the CRLB, we have to determine the Fisher information matrix, whose j-k-th entry is given by:

$$I[\Theta]_{j,k} = -E[\frac{\partial^2 \ln p(\mathbf{y_i};\sigma^2)}{\partial\sigma_j^2\partial\sigma_k^2}] = E[\frac{\partial \ln p(\mathbf{y_i};\sigma^2)}{\partial\sigma_j^2}\frac{\partial \ln p(\mathbf{y_i};\sigma^2)}{\partial\sigma_k^2}] \tag{4.35}$$

It may be shown that the log-likelihood function of the i-th received sample is given by

$$\ln(p(\mathbf{y_i};\sigma^2)) = -\ln(M\pi^{M_R}) - \sum_{k=1}^{M_R}\ln(\sigma_k^2) + \ln[\sum_{l=1}^{M}\exp(\sum_{k=1}^{M_R}\frac{\delta_{i,l,k}}{\sigma_k^2})] \tag{4.36}$$

where

$$\delta_{i,l,k} = (y_{i,k} - (\mathbf{Hx_l})_k)^*(y_{i,k} - (\mathbf{Hx_l})_k) \tag{4.37}$$

In this notation, $y_{i,k}$ is the scalar quantity that is the k-th entry of the vector $\mathbf{y_i}$ and $(\mathbf{Hx_l})_k$ is similarly the k-th entry of $\mathbf{Hx_l}$.

The derivative of Eqn. 4.36 with respect to $\sigma_k^2$ is given by

$$\frac{\partial \ln p(\mathbf{y_i};\sigma^2)}{\partial\sigma_k^2} = \frac{-1}{\sigma_k^2} + \frac{\sum_{l=1}^{M}\left[\exp(\sum_{k=1}^{M_R}\frac{\delta_{i,l,k}}{\sigma_k^2})(-\frac{\delta_{i,l,k}}{\sigma_k^4})\right]}{\sum_{l=1}^{M}\exp(\sum_{k=1}^{M_R}\frac{\delta_{i,l,k}}{\sigma_k^2})} \tag{4.38}$$

The expectations may be carried out either using Monte Carlo simulations or using Gauss-Hermite approximations as discussed in the preceding section.

### 4.5.5 Mixed DA and NDA Model

In most practical systems, in every block of N symbols, there are $N_p$ known pilot or preamble symbols and $N_d$ unknown data symbols such that $N = N_p + N_d$. For this model, the log-likelihood function may be expressed as:

$$\ln(p(\mathbf{Y};\sigma^{\mathbf{2}})) = \sum_{i=1}^{N_p} \ln(p(\mathbf{y_i};\sigma^{\mathbf{2}})) + \sum_{i=1}^{N_d} \ln(p(\mathbf{y_i};\sigma^{\mathbf{2}})) \tag{4.39}$$

Since derivatives and expectations are linear operations, we can obtain following from which Eqn. 4.41 is derived.

$$-E\left[\frac{\partial^2 \ln(p(\mathbf{Y};\sigma^{\mathbf{2}}))}{\partial(\sigma^2)^2}\right] = N_p E[-\frac{\partial^2 \ln(p(\mathbf{y_i};\sigma^{\mathbf{2}}))}{\partial(\sigma^2)^2}] + N_d E[-\frac{\partial^2 \ln(p(\mathbf{y_i};\sigma^{\mathbf{2}}))}{\partial(\sigma^2)^2}] \tag{4.40}$$

$$\frac{1}{CRLB[N]_{DA\_NDA}} = \frac{1}{CRLB[N_p]_{DA}} + \frac{1}{CRLB[N_d]_{NDA}} \tag{4.41}$$

### 4.5.6 With Random Channel

In this section we address the case where the channel matrix is independent from block to block (but known at the receiver). The determination of the CRLB would require the determination of

$$p(\mathbf{Y};\sigma^2) = \int p(\mathbf{Y}|H;\sigma^2)p(H)dH \tag{4.42}$$

Since this is difficult to evaluate in closed form, we resort to the modified Cramér-Rao lower bound (MCRLB). The MCRLB was introduced [27], [49] as a lower bound to the variance of an unbiased estimator provided the usual regularity conditions are satisfied. In general the MCRLB is lower than the CRLB and thus is a looser bound, but the authors show that it is tight enough to be useful in most situations.

As shown in Eqn. 5 of [27], the MCRLB is given by:

$$MCRLB(\sigma^2) = \left[E_{\mathbf{H}}\left(E_{\mathbf{Y}|\mathbf{H}}\left[\frac{\partial \ln p(\mathbf{Y};\sigma^2)}{\partial\sigma^2}\right]^2\right)\right]^{-1} \tag{4.43}$$

For the DA model, since Eqn. 4.28 and Eqn. 4.32 do not depend on the channel matrix, **H**, the MCRLB for the case of the random channel is identical to the CRLB when the channel is constant. For the NDA model, the MCRLB is determined by averaging the CRLB over many channel instantiations.

## 4.6 Extension to Frequency Selective Channels

Although the above derivations have all been carried out using flat fading channels, the extension to frequency selective channels in the orthogonal frequency division multiplexing (OFDM) framework is straight forward. It is well known that after the removal of the cyclic prefix (and assuming that the delay spread is shorter than the length of the cyclic prefix) each subcarrier faces flat fading and thus the received samples at time $i$, from each subcarrier $k$, $\mathbf{y_i^{(k)}}$ may be stacked in the following manner:

$$\begin{bmatrix} \mathbf{y_i^{(1)}} \\ \mathbf{y_i^{(2)}} \\ ... \\ ... \\ \mathbf{y_i^{(N_s)}} \end{bmatrix} = \begin{bmatrix} \mathbf{H^{(1)}} & 0 & ... & ... & 0 \\ 0 & \mathbf{H^{(2)}} & ... & ... & 0 \\ 0 & 0 & \mathbf{H^{(3)}} & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & \mathbf{H^{(N_s)}} \end{bmatrix} \begin{bmatrix} \mathbf{x_i^{(1)}} \\ \mathbf{x_i^{(2)}} \\ ... \\ ... \\ \mathbf{x_i^{(N_s)}} \end{bmatrix} + \begin{bmatrix} \mathbf{n_i^{(1)}} \\ \mathbf{n_i^{(2)}} \\ ... \\ ... \\ \mathbf{n_i^{(3)}} \end{bmatrix} \tag{4.44}$$

In the above equation, $N_s$ is the total number of subcarriers, $\mathbf{x_i^{(k)}}$ is the $M_T \times 1$ vector representing the transmit symbols on subcarrier $k$ at time $i$, $\mathbf{n_i^{(k)}}$ and $\mathbf{y_i^{(k)}}$ represent the $M_R \times 1$ vectors representing respectively, the AWGN noise samples and the received samples on subcarrier $k$ at time $i$. $\mathbf{H^{(k)}}$ represents the $M_R \times M_T$ channel matrix experienced by the $k$-th subcarrier. This channel is assumed to be constant for a particular block of data. Each subcarrier is assumed to undergo flat fading. The above equation may be seen to be in the similar form as Eqn. 4.1 and thus all the tools developed thus far may be used for the frequency selective channel case. The problem, however, could be significantly more computationally intensive due to the presence of a large number of potential subcarriers. Even in a non OFDM environment, a similar matrix equation may

be written. This has been shown in the system model of [95] and is not repeated here.

## 4.7 Estimators for SNR

In this section we derive maximum likelihood (ML) estimators for MIMO SNR. Since it is known from the invariance property of ML estimators that the ML estimate of the function of a set of parameters is the same function of the ML estimates of each of the parameters (Chapter 7, [59]), we shall evaluate the ML estimates of the noise variance and the channel coefficients and use those to find the ML estimates of the SNR.

We shall summarize some of the results already developed by Wautelet et al. in [95] where the authors develop ML estimators for both the channel coefficients and noise variance for both the DA and NDA models. We use these estimators to estimate the MIMO SNR and show that they achieve the CRLB calculated in preceding sections. We shall also present new results for closed form approximations of the ML estimator for the noise variance with the NDA model.

### 4.7.1 DA SNR Estimation

The ML estimator for the MIMO SNR is given by $\frac{\|\widehat{\mathbf{H}}\|_F^2}{\widehat{\sigma^2}}$ where $\widehat{\sigma^2}$ and $\|\widehat{\mathbf{H}}\|_F^2$ are respectively the ML estimate of the noise variance and the ML estimate of the Frobenius norm of the channel.

#### ML Estimator for Noise Variance with DA model

The maximum likelihood (ML) estimator for the DA model with one parameter is determined easily by taking the derivative of the log-likelihood function and setting it equal to zero and is given by Eqn. 4.45.

$$\hat{\sigma}^2_{DA\_ML} = \frac{1}{NM_R}\sum_{i=1}^{N}(\mathbf{y_i} - \mathbf{Hx_i})^H(\mathbf{y_i} - \mathbf{Hx_i}) \tag{4.45}$$

When jointly estimating the channel, $\mathbf{H}$, and the noise variance, $\sigma^2$, the $\mathbf{H}$ in the above equation is replaced with the ML estimates of the channel, $\widehat{\mathbf{H}}$. Luckily (see below), the ML estimate of $\mathbf{H}$ does not depend on the variance and thus, the ML estimator of the SNR may be easily determined. It was shown by Wautelet et al. [95] that this estimator of the noise variance is biased and that an unbiased estimator is given by $\hat{\sigma}^2_{DA\_ML\_unbiased} = \frac{N}{N-M_T}\hat{\sigma}^2_{DA\_ML}$.

**ML Estimator for Channel Coefficients with DA Model**

The ML estimate in this case is identical to the least squared error (LS) solution, since the ML estimate maximizes $-\sum_{i=1}^{N} \|\mathbf{y_i} - \mathbf{H}\mathbf{x_i}\|^2$ which is the same as minimizing the least squares error. This has also been shown before in [95].

An alternate derivation including simplified expression under certain conditions are presented in Appendix 4.A.

### 4.7.2 NDA SNR Estimation

**ML Estimator for Noise Variance with NDA model**

Unfortunately in the NDA case, there is no nice closed form solution for the ML estimator like in the DA case. Using Eqn. 4.33, the log likelihood function is given by

$$\frac{\partial \ln p(\mathbf{Y};\sigma^2)}{\partial(\sigma^2)} = \frac{NM_R}{\sigma^2} - \frac{\sum_{i=1}^{N} \mathbf{y_i^H y_i}}{\sigma^4} - \frac{1}{\sigma^4}\sum_{i=1}^{N} \frac{\sum_{l=1}^{M} \delta_{i,l} \exp(\frac{\delta_{i,l}}{\sigma^2})}{\sum_{l=1}^{M} \exp(\frac{\delta_{i,l}}{\sigma^2})} \tag{4.46}$$

where the scalar $\delta_{i,l} = 2\mathcal{R}(\mathbf{y_i}\mathbf{H}\mathbf{x_l}) - (\mathbf{H}\mathbf{x_l})^{\mathbf{H}}\mathbf{H}\mathbf{x_l}$ and $\mathcal{R}(x)$ is the real part of $x$. By setting the derivative equal to 0, we recognize that the ML estimate is the solution to the following equation which does not have a closed form solution

$$\sigma^2 NM_R = \sum_{i=1}^{N} \mathbf{y_i^H y_i} - \sum_{i=1}^{N} \frac{\sum_{l=1}^{M} \delta_{i,l} \exp(\frac{\delta_{i,l}}{\sigma^2})}{\sum_{l=1}^{M} \exp(\frac{\delta_{i,l}}{\sigma^2})} \tag{4.47}$$

However, we can make approximations in the low SNR and high SNR regions that do yield closed form solutions. In the low SNR regime, $\sigma^2 \gg \delta_{i,l}$ and

thus $\exp(\frac{\delta_{i,l}}{\sigma^2}) \simeq 1$. In this case, the ML estimate of the noise variance may be approximated by:

$$\sigma^2_{ML\_NDA\_lowSNR} \simeq \frac{1}{NM_R}\sum_{i=1}^{N}\mathbf{y_i^H y_i} - \frac{1}{NMM_R}\sum_{i=1}^{N}\sum_{l=1}^{M}\delta_{i,l} \tag{4.48}$$

Other approximations may also be made using $\exp(\frac{\delta_{i,l}}{\sigma^2}) \simeq 1 + \frac{\delta_{i,l}}{\sigma^2}$ or using $\exp(\frac{\delta_{i,l}}{\sigma^2}) \simeq 1 + \frac{\delta_{i,l}}{\sigma^2} + \frac{\delta^2_{i,l}}{2(\sigma^2)^2}$ which lead to closed form expressions for the estimate of $\sigma^2$ as the solutions of quadratic or cubic equations. These formulations are shown below.

**Derivation of Higher Order Approximations of NDA ML Estimator for Noise Variance**

Using the first three terms of the Taylor series expansion, we approximate $\exp(\frac{\delta_{i,l}}{\sigma^2}) \simeq 1 + \frac{\delta_{i,l}}{\sigma^2} + \frac{\delta^2_{i,l}}{2(\sigma^2)^2}$. For low SNRs, $\sigma^2$ is large, thus allowing this approximation.

$$\frac{\sum_{l=1}^{M}\delta_{i,l}\exp(\frac{\delta_{i,l}}{\sigma^2})}{\sum_{l=1}^{M}\exp(\frac{\delta_{i,l}}{\sigma^2})} \simeq \frac{\sum_{l=1}^{M}\delta_{i,l} + \frac{1}{\sigma^2}\sum_{l=1}^{M}\delta^2_{i,l} + \frac{1}{2(\sigma^2)^2}\sum_{l=1}^{M}\delta^3_{i,l}}{M + \frac{1}{\sigma^2}\sum_{l=1}^{M}\delta_{i,l}} \tag{4.49a}$$

$$\simeq \frac{1}{M}\left[\underbrace{\sum_{l=1}^{M}\delta_{i,l}}_{\alpha_i} + \frac{1}{\sigma^2}\underbrace{\sum_{l=1}^{M}\delta^2_{i,l}}_{\beta_i} + \frac{1}{2(\sigma^2)^2}\underbrace{\sum_{l=1}^{M}\delta^3_{i,l}}_{\gamma_i}\right] \tag{4.49b}$$

$$= \frac{1}{M}\left[\alpha_i + \frac{1}{\sigma^2}\beta_i + \frac{1}{2(\sigma^2)^2}\gamma_i\right] \tag{4.49c}$$

In Eqn. 4.49, the second approximation is done using the fact that for low SNR, $M \gg \frac{1}{\sigma^2}\sum_{l=1}^{M}\delta_{i,l}$ and $\alpha_i$, $\beta_i$ and $\gamma_i$ are used for notational convenience. Thus, Eqn. 4.47, may be put in the following form:

$$(\sigma^2)^3 NM_R - (\sigma^2)^2\sum_{i=1}^{N}\left[\mathbf{y_i^H y_i} - \frac{\alpha_i}{M}\right] + \sigma^2\sum_{i=1}^{N}\frac{\beta_i}{M} + \sum_{i=1}^{N}\frac{\gamma_i}{2M} = 0 \tag{4.50}$$

This is a cubic equation in $\sigma^2$ and may be solved in closed form. An easier approximation is to ignore the 2nd order term in the Taylor series expansion of

the exponential, thus setting $\gamma_i = 0$ and converting Eqn. 4.50 into a quadratic equation that has a more compact closed form solution.

At the other end of the spectrum, in the high SNR regime, $\sigma^2 \ll \delta_{i,l}$ and the $M$ values of $\exp(\frac{\delta_{i,l}}{\sigma^2})$ (for any particular $i$) are dominated by the maximum term. Ignoring all but the maximum term, the ML estimate of the noise variance may be approximated by

$$\sigma^2_{ML\_NDA\_highSNR} \simeq \frac{1}{NM_R}\sum_{i=1}^{N}\mathbf{y_i^H y_i} - \frac{1}{NM_R}\sum_{i=1}^{N}\max(\delta_{i,l}) \tag{4.51}$$

Although no closed form solution exists for the general case, the EM algorithm is known (for regular exponential family distributions) to iteratively attain the ML estimate [35] and is thus proposed here for determining the ML estimate (See [95] fore more details). In this case, the estimate of the noise variance at the i-th iteration is

$$\hat{\sigma}_i^2 = \frac{1}{NM_R}\sum_{l=1}^{M}\left[p(\mathbf{x_l}|\widehat{\sigma}^2_{i-1},\mathbf{Y})\sum_{i=1}^{N}(\mathbf{y_i}-\mathbf{Hx_l})^H(\mathbf{y_i}-\mathbf{Hx_l})\right] \tag{4.52}$$

Note that formally this estimator is very similar to the ML estimator, except that here the $M$ estimates (one for each of the $M$ constellation points) are weighted by $p(\mathbf{x_l}|\widehat{\sigma}^2_{i-1},\mathbf{Y})$ which is the posterior probability of each of the constellation points given the observations and the current estimate of the noise variance.

Initialization is always a key issue with the EM algorithm. If there are some known pilots, then they should be used to form a ML DA estimate and that is used as the starting point of the EM algorithm. If, there are no known pilots, then the EM algorithm is initialized to the variance corresponding to a low SNR.

### ML Estimator for Channel Coefficients with NDA model

The EM algorithm for estimating the channel coefficients have been developed in [95] and are not repeated here. This algorithm is used in our simulations in Sec. 4.8.

## 4.8 Simulation Results

The following simulation results will show that the estimators described above perform very close to the CRLB. The metric used in all cases is the normalized mean squared error ($NMSE = \frac{MSE}{\rho^2} = \frac{E[(\rho-\widehat{\rho})^2]}{\rho^2}$), where $\rho$ is the true SNR and $\widehat{\rho}$ is the estimate of the SNR. The simulations are carried out for BPSK, QPSK, 8PSK and 16QAM modulations but could be easily generalized for any M-ary digital modulation. Each point on the curves is generated by running at least 10,000 independent bursts each of size N symbols (N will be specified from experiment to experiment). The data symbols are assumed to be chosen from constellation points that are equiprobable and independent from the channel and noise samples. The Gaussian noise samples are generated to be spatially and temporally independent. Some of the experiments below were conducted with the same value for the channel from burst to burst. For a $(M_T, M_R) = (2, 2)$ system, the channel was a constant channel set to $H = \begin{bmatrix} -0.3059 - 0.8107i & 0.0886 + 0.8409i \\ -1.1777 + 0.8421i & 0.2034 - 0.0266i \end{bmatrix}$. This channel is a particular instantiation of the random channel that is generated by Matlab when the seed for the randn function is set to 0. For other system configurations, the channel used was that generated by Matlab, using the command $\mathbf{H} = \frac{1}{\sqrt{2}}(\text{randn}(Mr, Mt) + \sqrt{-1} \times \text{randn}(Mr, Mt))$ after setting the seed for the randn function to 0.

### 4.8.1 Estimating MIMO SNR

#### CRLB and NMSE for NDA and DA SNR Estimation

Fig. 4.1 compares the normalized CRLB ($= \frac{CRLB}{\rho^2}$ where $\rho = \frac{\|\mathbf{H}\|_F^2}{\sigma^2}$ is the true SNR) to the NMSE for various constellations for NDA MIMO SNR estimation. It also compares the DA normalized CRLB to the DA ML estimator and shows that both the NDA and DA estimators perform very close to their respective CRLBs. Note that the DA result is independent of the constellation (Sec. 4.4.1) for large values of N, the number of symbols in the block. In Fig. 4.1, the number of symbols in the block, N, was chosen to be 1000. The system under consideration

was a $(M_T, M_R) = (2, 2)$ system. Since the CRLB is proportional to $1/N$, the CRLB for other values of N may be easily determined from this figure. For the NDA CRLBs, Gauss Hermite Quadratures of order 6 was used to evaluate the FIM. It may be seen from Fig. 4.7 that approximations of this order are very close to the values obtained from Monte Carlo simulations. The NDA estimator used is the iterative EM algorithm based estimator. The EM algorithm has been initialized by assuming that the 1000 symbol unknown data block is preceded by a 20 symbol preamble. These known preamble symbols are used in a DA manner to initialize the channel in the EM algorithm. The noise variance was initialized to 10 corresponding to a low SNR. The stopping criterion used was $\frac{\widehat{\rho}_i - \widehat{\rho}_{i-1}}{\widehat{\rho}_{i-1}} \leq 0.1\%$ where $\widehat{\rho}_i$ is the estimate of the SNR from the $i$-th iteration.

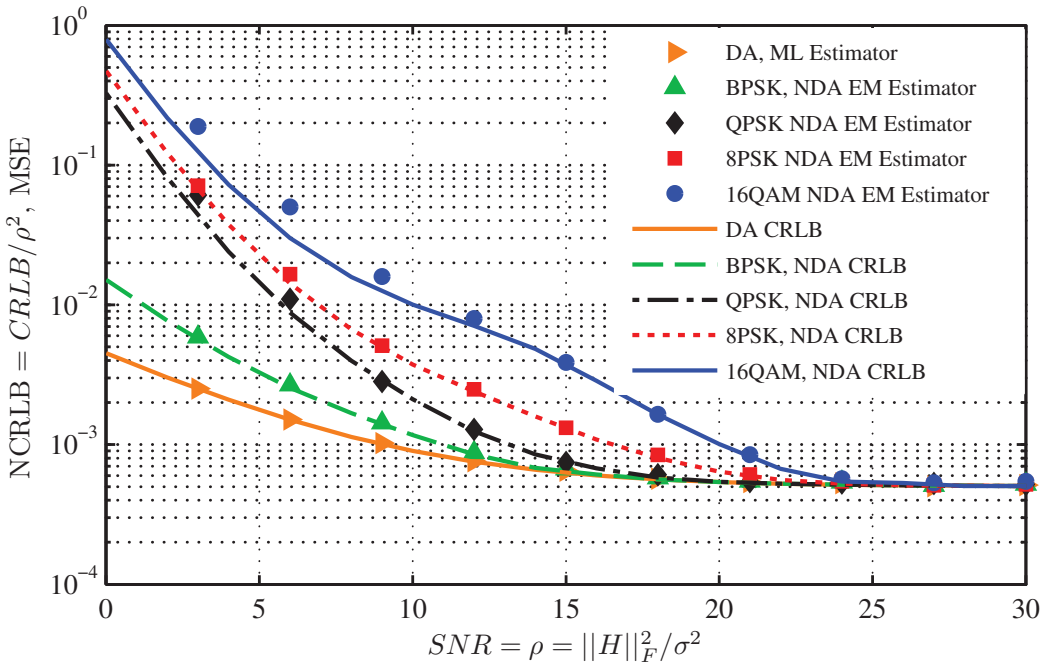


**Figure 4.1**: NCRLB and NMSE for MIMO SNR: NDA and DA. $(M_T, M_R) = (2, 2)$. $N = 100$.

Fig. 4.2 shows the cumulative distribution function (CDF) of convergence times for the various blocks simulated for a $(M_T, M_R) = (2, 2)$ system with QPSK modulation at a true SNR of 0dB. At such a low SNR, it can be seen that the EM

algorithm takes a large number of iterations to converge. In this simulation, the convergence criterion is $\frac{\widehat{\rho}_i - \widehat{\rho}_{i-1}}{\widehat{\rho}_{i-1}} \leq 0.05\%$. It may be seen that better initializations can significantly speed up the EM convergence speeds. At one extreme, the channel matrix $\mathbf{H}$, has been initialized with a random matrix (normalized to have the same power as the channel matrix), while at the other extreme, the channel matrix has been initialized to the true value. It can be seen that the median convergence speed drops by more than 100 iterations from more than 340 to about 240. If some of the transmitted symbols are known, they may be used in a data aided manner for initialization and that helps speed the convergence up as well.

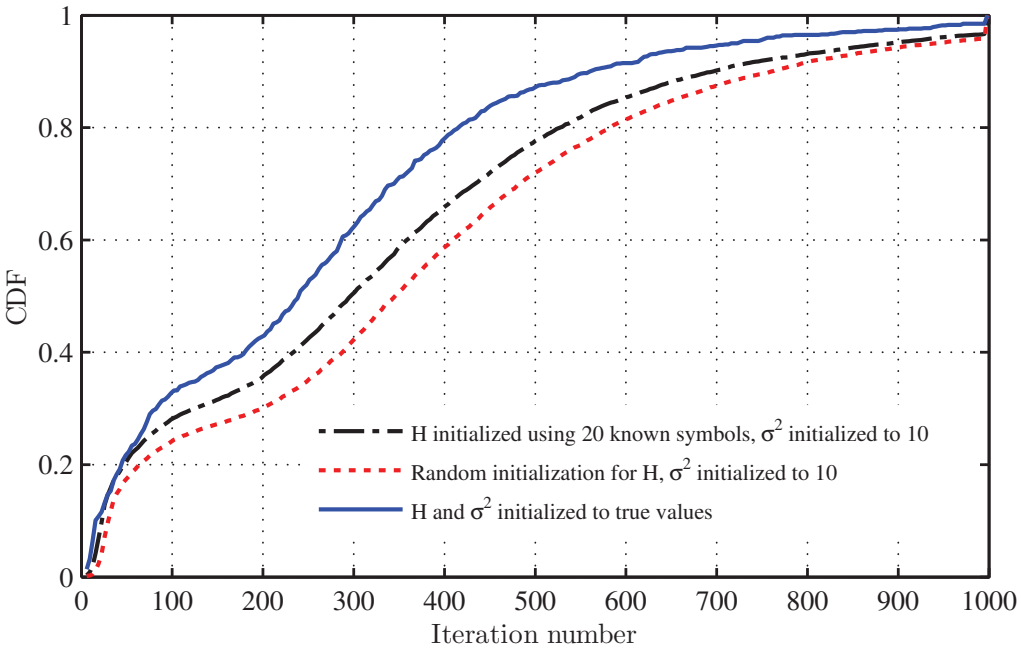


**Figure 4.2**: Better initializations speed up EM convergence. Modulation: QPSK. SNR $= \frac{\|\mathbf{H}\|_F^2}{\sigma^2} = 0$dB. $(M_T, M_R) = (2, 2)$. $N = 1000$. Stopping criterion $= \frac{\widehat{\rho}_i - \widehat{\rho}_{i-1}}{\widehat{\rho}_{i-1}} \leq 0.05\%$

### Impact of Varying Number of Antennas

Figs. 4.3-4.5 shows the impact to the normalized CRLB when the number of antennas are varied. It can be seen (Fig. 4.3) that increasing the number of

received antennas decreases the normalized CRLB and, as shown in Sec.4.7.1, at high SNRs, the normalized CRLB can be seen to tend to $1/(NM_R)$. Fig. 4.4 and 4.5 show the CRLB for BPSK and QPSK modulation schemes respectively for 2, 3, 4 and 5 transmit antennas while keeping the number of receive antennas to be constant at 2. Once again, it can be seen that at high SNRs the the normalized CRLB can be seen to tend to $1/(NM_R)$. In these simulations, the total transmit power of the signal waveform is constant regardless of the number of transmit antennas.

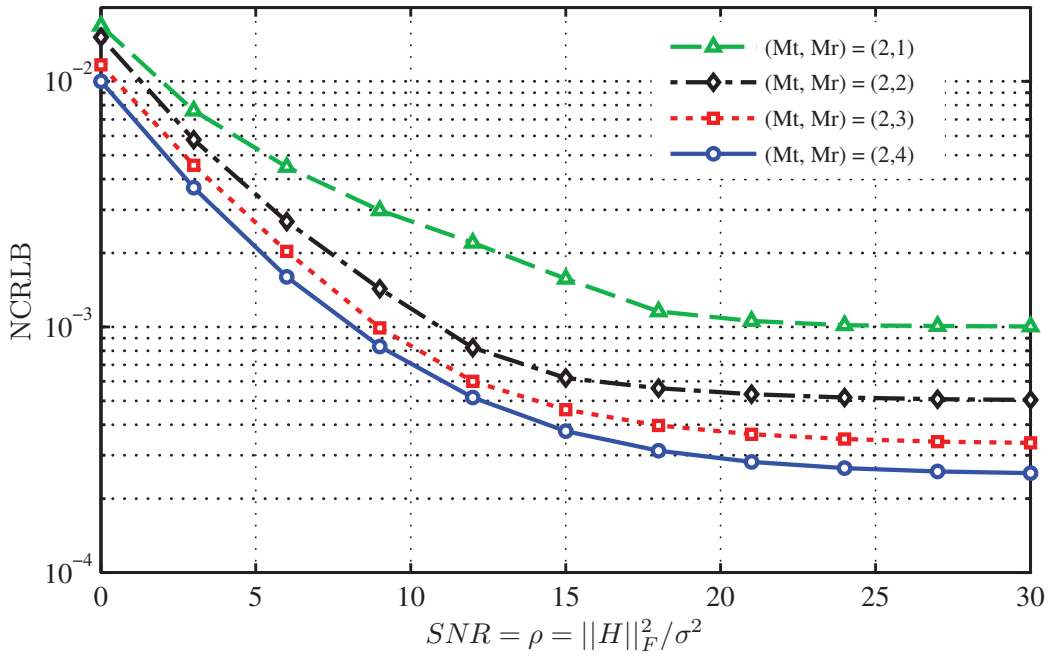


**Figure 4.3**: Increasing $M_R$ reduces NCRLB. Modulation: BPSK. N = 1000. NDA SNR CRLB.

In Fig. 4.6, the CRLB for SNR estimation is shown for square systems (i.e., $M_T = M_R$) with 1, 2 or 3 antennas on either end. Fig. 4.6 also shows previously published SISO results from [28] and it can be seen that the SISO results are the same as those obtained in our current work here for $M_T = M_R = 1$. In order to compare with the previously published results, the number of symbols, $N$, was set to 100.

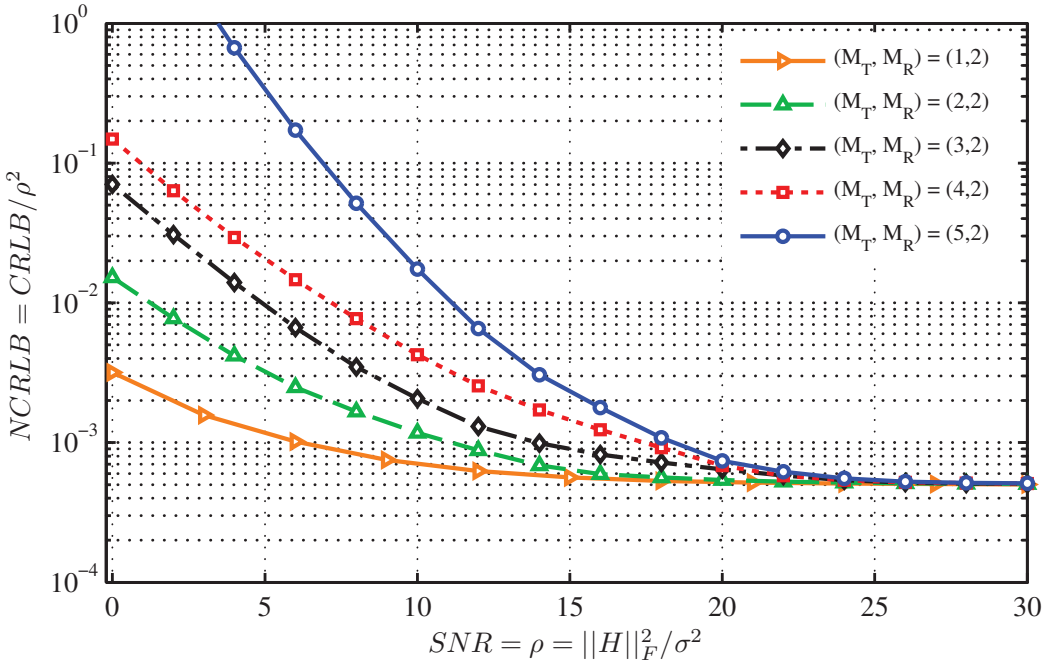


**Figure 4.4**: Increasing $M_T$ increases NCRLB at low SNRs. Modulation: BPSK. $N = 1000$. NDA CRLB.

### 4.8.2 Estimating MIMO Noise Variance

In this section we present the simulation results for estimating only the noise variance (assuming the channel is known perfectly). In these experiments, a block size of 100 symbols was simulated.

#### NDA Estimation

**CRLB and EM based NDA Estimators** Fig. 4.8 shows that the EM based estimators perform very close to the CRLB for the 4 modulations simulated (BPSK, QPSK, 8PSK and 16QAM). 10 iterations of the EM algorithm were seen to be sufficient. The initial value of the variance was initialized to 5. It is seen from Fig. 4.8 that the EM algorithm achieves the CRLB over a wide range of Es/No.

**Approximations to NDA ML Estimation** Fig. 4.9 shows the performance of the closed form approximations to the ML solution for the NDA model. As can be

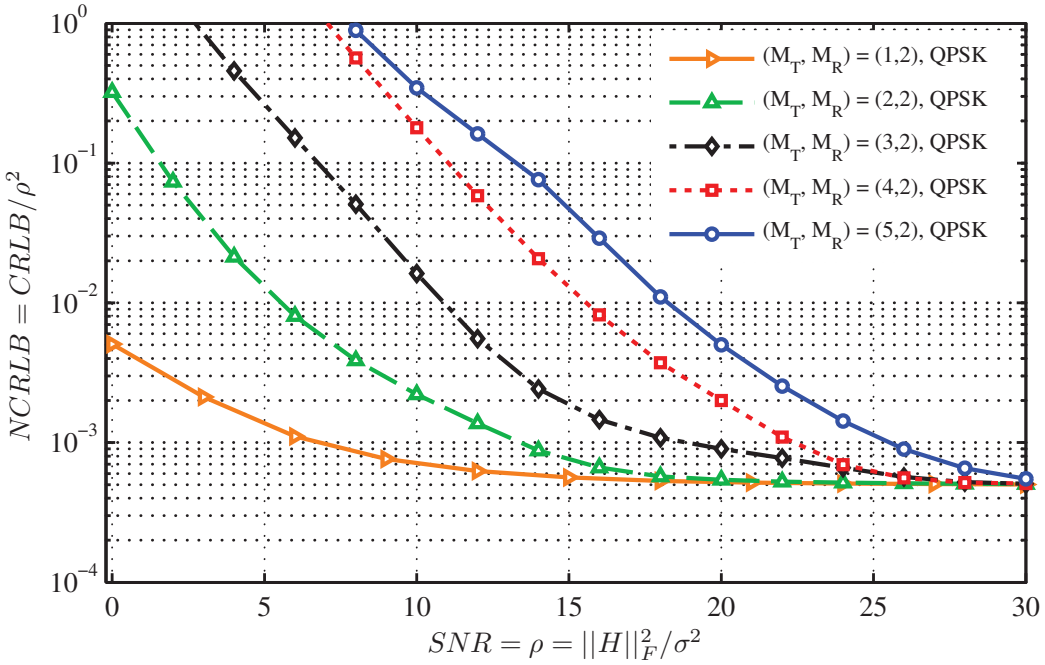


**Figure 4.5**: NDA CRLB with varying $M_T$. Modulation: QPSK. $N = 1000$.

seen, the high SNR approximation achieves the CRLB at SNRs greater than 5dB. 3 different approximations to the ML solution are shown for the low SNR case. Each one is progressively better (and more complex to implement). In Fig. 4.9, the 0-th order approximation refers to the one given by Eqn. 4.48, the 1-st order approximation is given by setting $\gamma_i = 0$ in Eqn. 4.50 and thus solving a closed form quadratic equation and the 2nd-order approximation is given by solving the cubic equation given by Eqn. 4.50. While the cubic equation is guaranteed to provide at least one real root (since the coefficients are real), the estimator based on the quadratic equation has the unfortunate characteristic that as the SNR increases, occasionally the estimator fails completely and provides complex answers. This indicates that the approximation made is no longer valid at that particular SNR. Interestingly it can be seen that the high SNR approximation comes close to the CRLB for very low SNRs as well. This is due to the fact that at very low SNRs, the first term in Eqn. 4.51 and Eqn. 4.48, $\frac{1}{NM_R}\sum_{i=1}^{N}\mathbf{y_i^H y_i}$, dominates the equation.

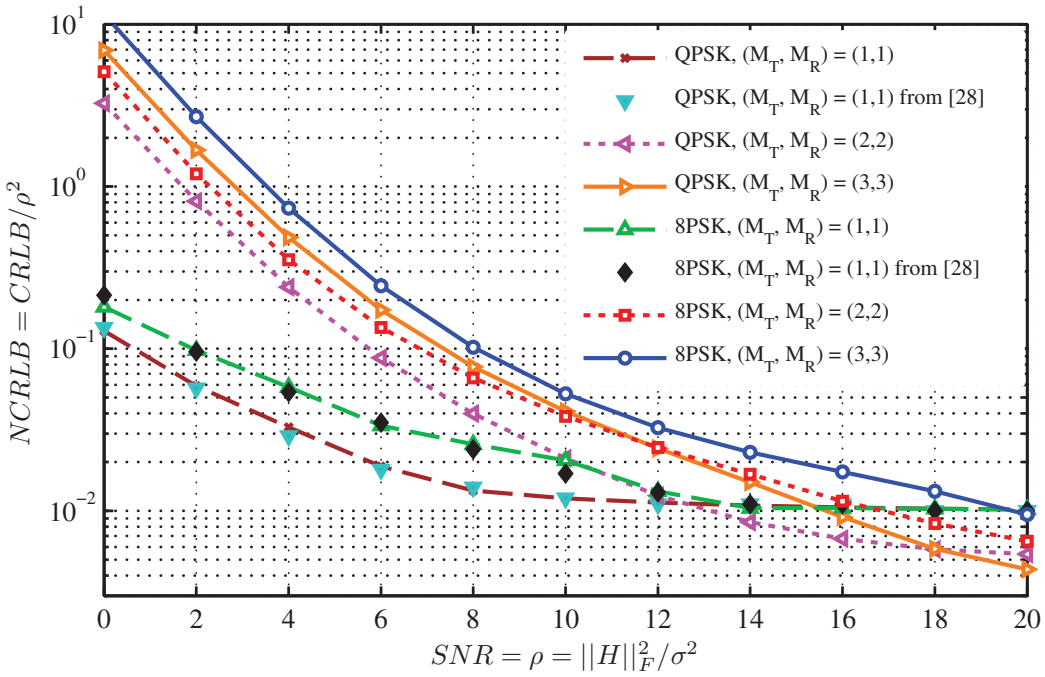


**Figure 4.6**: Previously published SISO results are a special case of MIMO results, $N = 100$

### DA Estimation

The DA MCRLB was plotted along with the variance of the ML estimator. BPSK modulation was used (though in the DA case, the results are independent of which modulation scheme is used). For comparison, the NDA CRLB is also plotted. It may be seen from Fig. 4.10 that the ML estimator achieves the MCRLB over the entire range of Es/No plotted.

The EM algorithm is not guaranteed to converge to the global maxima of the log likelihood function – it could converge to a local maxima. Luckily, the log likelihood function in our case is well behaved and has one unique maxima as shown in Fig. 4.11. This ensures that the EM algorithm converges to the global maxima.

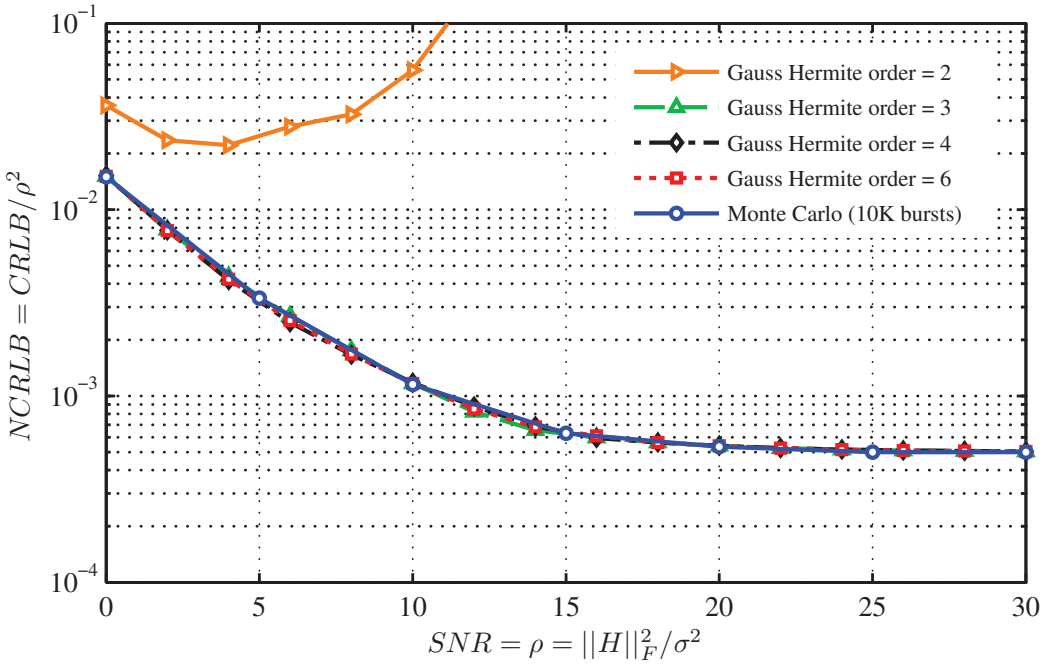


**Figure 4.7**: Gauss Hermite approximations compared to Monte Carlo methods. Modulation: BPSK. NDA CRLB. $N = 1000$. $(M_T, M_R) = (2, 2)$.

### 4.8.3 DA + NDA Mixed Estimation

The advantage to using all the available symbols in a block and not just the $N_p$ known pilot/training symbols may be seen in Fig. 4.12 which shows how much improvement may be attained by using all the $N$ symbols in the block, rather than just the known pilot/training symbols. The advantage is quite significant in those communication systems that operate with less than 10% overhead but there are reasonable gains to be made even when $20\% - 30\%$ of the data consists of known pilot symbols. Fig. 4.12 was calculated for a 100 symbol block in a $2 \times 2$ system with BPSK modulation scheme at an $E_s/N_0 = 0dB$.

## 4.9 Conclusion

In this chapter, we have presented the CRLB for SNR estimation and noise variance estimation for MIMO systems in a data aided (DA), a non-data aided

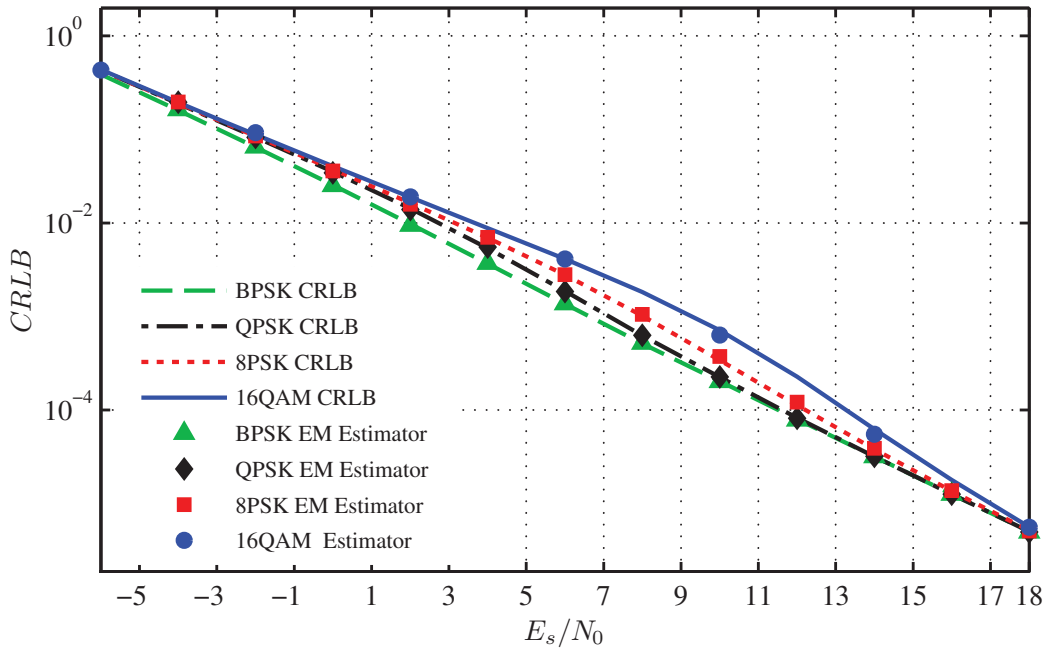


**Figure 4.8**: Noise variance estimation: CRLB and EM based estimators. $(M_T, M_R) = (2, 2)$.

(NDA) and a mixed DA-NDA system. We have also presented closed form expressions for the ML estimator for the DA case, closed form approximations to the ML estimator for the NDA case as well as iterative EM based estimators for the NDA case. Although most of the results have been presented for a frequency flat block fading model, we have discussed extensions to the frequency selective case using the OFDM framework. Since SNR and noise variance estimation are needed by most receivers, we believe this chapter presents results that computes the limits of this important problem and demonstrates practical estimators that achieve these limits.

The contents of this chapter have been adapted from A. Das and B. D. Rao, "SNR and Noise Variance Estimation for MIMO Systems," IEEE Transactions on Signal Processing, August 2012 and from A. Das, "NDA SNR estimation: CRLBs and EM based estimators," IEEE Region 10 Conference, TENCON 2008.

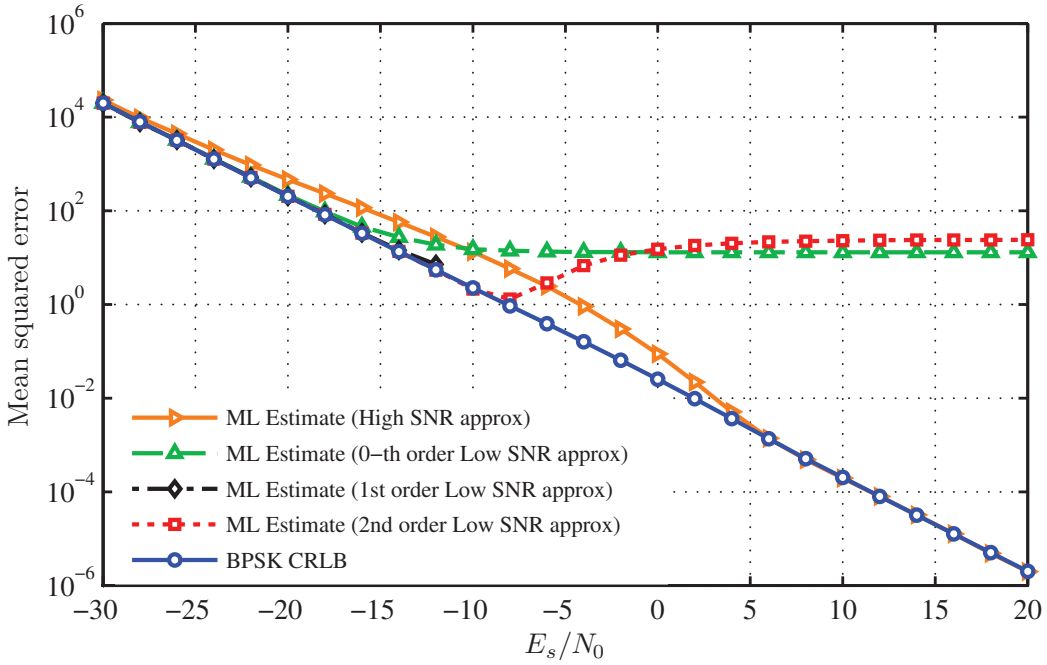


**Figure 4.9**: NDA noise variance estimation: Performance of closed form approximations to ML solution. Modulation: BPSK. $(M_T, M_R) = (2, 2)$. $N = 100$.

## 4.A Appendix: Alternate Derivation of ML DA Channel Estimation

The derivative of the log-likelihood function, $g$, with respect to the channel gain between the $x$-th receiver and the $y$-th transmitter, $h_{xy}$ is given by

$$\frac{\partial g}{\partial h_{xy}} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \left[ (y_{i,x} - (\mathbf{H}\mathbf{x_i})_x)^* x_{i,y} \right] \tag{4.53}$$

Setting this equal to zero, the $M_T M_R$ equations for the ML estimates of $h_{xy}$ are:

$$\sum_{i=1}^{N} y_{i,x}^* x_{i,y} = \sum_{j=1}^{M_T} \left[ h_{xj}^* \left( \sum_{i=1}^{N} x_{i,j}^* x_{i,y} \right) \right] \tag{4.54}$$

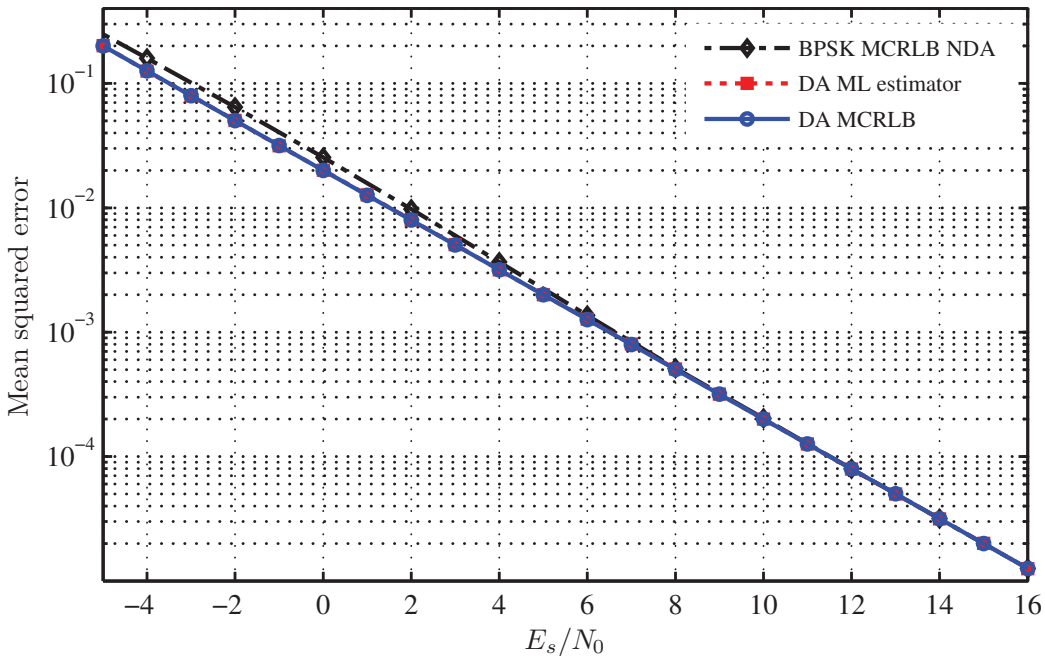


**Figure 4.10**: DA noise variance estimation: ML estimators achieve MCRLB for DA model with random channel. Modulation: BPSK. $N = 100$.

which in matrix form is given by Eqn. 4.55, where $\mathbf{A}$ is given by 4.7.

$$\begin{bmatrix} \sum_{i=1}^{N} y_{i,1}^* x_{i,1} \\ \sum_{i=1}^{N} y_{i,1}^* x_{i,2} \\ \dots \\ \sum_{i=1}^{N} y_{i,1}^* x_{i,M_T} \\ \sum_{i=1}^{N} y_{i,2}^* x_{i,1} \\ \sum_{i=1}^{N} y_{i,2}^* x_{i,2} \\ \dots \\ \sum_{i=1}^{N} y_{i,M_R}^* x_{i,M_T} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^* & \mathbf{0}_{M_T} & \dots & \dots & \dots & \mathbf{0}_{M_T} \\ \mathbf{0}_{M_T} & \mathbf{A}^* & \dots & \dots & \dots & \mathbf{0}_{M_T} \\ \dots & \dots & \dots & \dots & \dots & \\ \mathbf{0}_{M_T} & \mathbf{0}_{M_T} & \dots & \dots & \mathbf{0}_{M_T} & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} h_{11}^* \\ h_{12}^* \\ \dots \\ h_{1M_T}^* \\ h_{21}^* \\ h_{22}^* \\ \dots \\ h_{M_R M_T}^* \end{bmatrix} \tag{4.55}$$

Writing the above equation as $\mathbf{G} = \mathbf{B}\mathbf{H}^*$ for notational convenience, if $\mathbf{A}$ is non-singular, the ML estimates of the channel coefficients are given by $\mathbf{H} = (\mathbf{B^{-1}G})^*$. As discussed in Section 4.4.1, under certain assumptions, $\mathbf{A} \approx \mathbf{N}\mathbf{I}_{\mathbf{M_T}}$ and in that case, the ML estimate of $h_{xy}$ is given by $\widehat{h_{xy}} = \frac{1}{N}\sum_{i=1}^{N} y_{i,x} x_{i,y}^*$

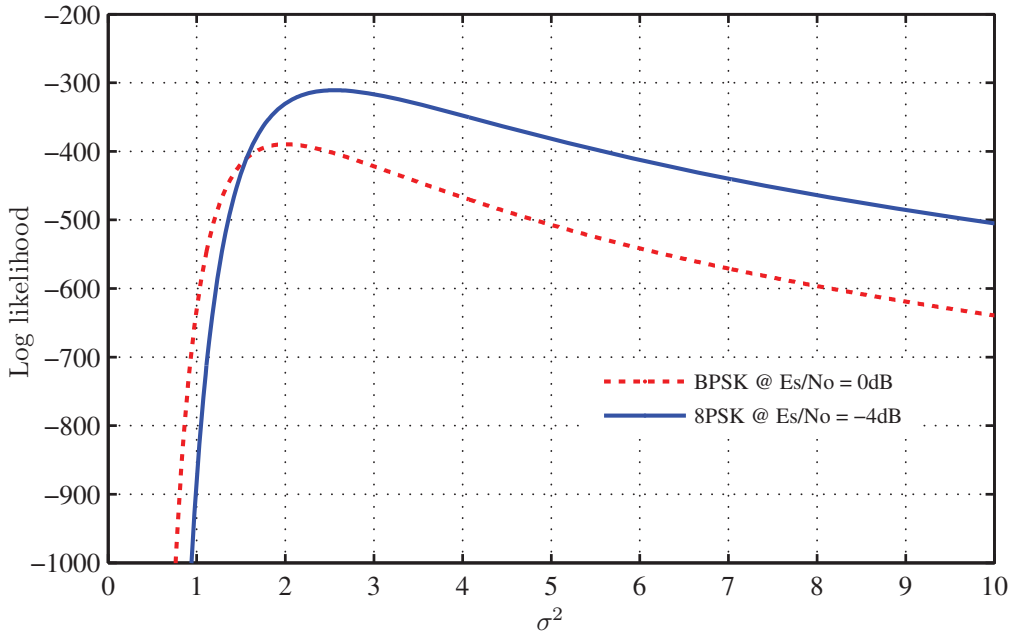


**Figure 4.11**: Log likelihood function has unique maxima

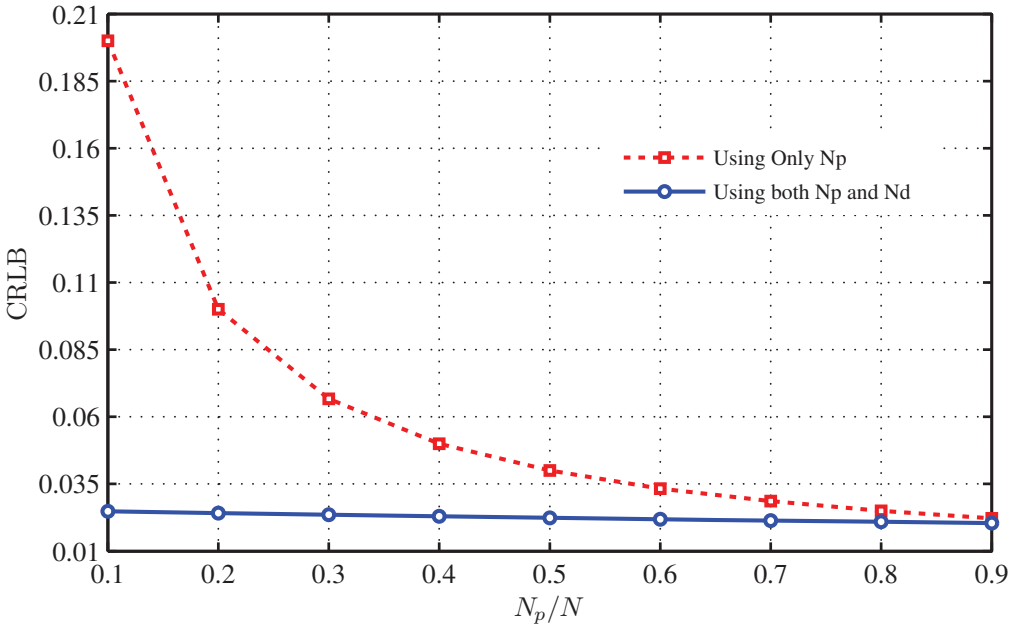


**Figure 4.12**: Using all the symbols in a block leads to improvements. Modulation: BPSK. $N = 100$. $(M_T, M_R) = (2, 2)$. $E_s/N_0 = 0dB$.

# Chapter 5

# Rate Adaptive Non-Orthogonal MT-MFSK

## 5.1 Introduction

Thus far, in this dissertation, in Chapters 2 and Chapters 3 we have discussed system improvements for coherent systems with no phase uncertainty. In Chapter 4 we analyzed NDA SNR estimation for complex valued channels with an implicit phase uncertainty and thus started investigating non-coherent systems. Although much of modern wireless communication takes place with coherent receivers, in some cases, non-coherent receivers may be more appropriate. Although in Chapter 4, we try to estimate the channel, some non-coherent receivers don't estimate the channel since the operating conditions may make it futile. In some systems even if the channel gain can be estimated accurately, it may not be possible to maintain phase coherence. This is often the case for systems with very high dynamics where channel estimation is futile as shown in [96], [73]. In these two works, the authors propose a non-coherent system for communications over channels associated with high speed trains. Power limited systems with low duty cycle is another example where non-coherent systems may be more appropriate. In this chapter, we discuss techniques for improving non-coherent communication which need to be able to operate without channel estimation.

M-ary frequency shift keying (MFSK) has been the de facto standard for non-coherent communications for about 50 years. One of the first systems to use a 32-ary orthogonal tone MFSK was the Piccolo system, [7, 79], developed for diplomatic communication by the British Foreign & Commonwealth Office in 1962. A key advantage of MFSK is the ability to carry out non-coherent reception as shown in [4, 74, 85] (and references therein), thus obviating the need for any overhead for channel estimation.

An MFSK system has $M$ distinct symbols. In traditional MFSK systems, each "symbol" consists of a single frequency tone orthogonal to all the other tones representing the other symbols and thus the system has a total of $N$ orthogonal tones where $N = M$. Thus each symbol can transmit $\log_2(M)$ channel bits. The Piccolo system [7], had 32 orthogonal tones and so each symbol could send 5 bits of information. A second key advantage of single tone MFSK is its constant envelope transmit signal. This eliminates the need for linear amplifiers in the transmit chain and generally results in cheaper systems. This may be contrasted to the high peak to average power ratio (PAPR) of multi-tone systems such as orthogonal frequency division multiplexing (OFDM) [37].

In spite of these advantages, the principal disadvantage of single tone orthogonal MFSK is low spectral efficiency. In a non-coherent orthogonal MFSK system, the tones need to be separated in frequency by the symbol rate, $R_s$, and thus the total bandwidth $W = NR_s$ (see [85], Chapter 5). Hence, the spectral efficiency, $n$, which is the ratio of the bit rate, $R_b$, to the total bandwidth, $W$, is given by Eqn. 5.1.

$$n = \frac{R_b}{W} = \frac{log_2(M)R_s}{W} = \frac{log_2(M)}{N} bits/s/Hz \tag{5.1}$$

For integer $N$, and for $M = N$ (i.e., single tone MFSK), this equation has a maximum at $N = 3$ with $n = 0.528 bits/s/Hz$. This is the maximum spectral efficiency for single tone MFSK. For modern communication schemes, a higher spectral efficiency is desirable.

Most practical systems work under the constraint of a fixed bandwidth. Many systems also need to work with the constraint of a fixed symbol rate. This is

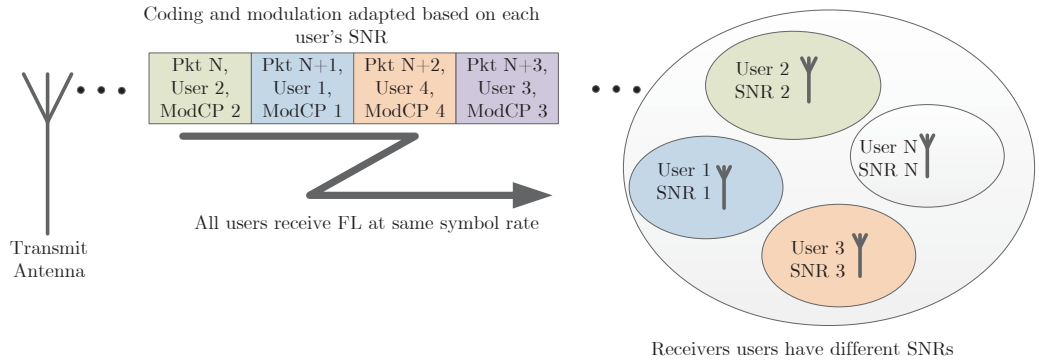


**Figure 5.1**: Adaptive coding and modulation (ACM) on shared forward link with constant symbol rate

typically the case, as shown in Fig. 5.1, where many users share the same channel with the same symbol rate, but the modulation varies based on the link quality of the different users [43], [30]. For fixed bandwidth and symbol rate, it is not possible to increase the bits per symbol in a single tone orthogonal MFSK system as the number of orthogonal tones (the symbols), are fixed. This is in contrast to M-ary phase shift keying (MPSK) and M-ary quadrature amplitude modulation (MQAM) systems where the number of possible symbols (and hence the number of bits per symbol), can be adapted keeping the bandwidth and symbol rate the same. In orthogonal MFSK systems, increasing the number of symbols, $M$, leads to either an increase in bandwidth or a reduction in symbol rate.

The only choice for increasing system spectral efficiency for an orthogonal MFSK system is to move from a single tone system to a multi-tone MFSK system (MT-MFSK) where each symbol consists of multiple simultaneous orthogonal tones. The price to pay for the increased spectral efficiency is that the same bit error rate can be achieved at a higher signal to noise ratio. Moreover, like all multi-carrier waveforms, MT-MFSK does not have a constant transmit signal envelope and thus requires linear (and more expensive) amplifiers. Multi-tone orthogonal MFSK, was shown in [67], [68] to achieve capacity in the wideband regime.

We show in this chapter, that the traditional orthogonal tone based MT-

MFSK system is not optimal and we propose an alternate MT-MFSK system using *non-orthogonal* signaling tones that outperforms an equivalent orthogonal MFSK system even in the finite bandwidth regime. We show that, in certain operating points, the proposed system has **four times** the capacity of an orthogonal tone based MT-MFSK system.

In addition to outperforming the equivalent orthogonal tone based MT-MFSK system, the proposed non-orthogonal MT-MFSK system has the additional benefit of providing the system designer with an infinite number of choices for system spectral efficiency – something that is lacking to systems using orthogonal MT-MFSK. An orthogonal MT-MFSK system can only increase the spectral efficiency within some combinatorial limits since the maximum number of symbols possible is limited by $\binom{N}{N/2}$. For small $N$, this choice may be fairly limiting. In a binary FSK system MT-MFSK is not possible and so there is no way to increase the spectral efficiency beyond 0.5 bits/s/Hz. In a system with 4 orthogonal tones, a MT-MFSK can at most send $\binom{4}{2} = 6$ symbols and in a system with 8 orthogonal tones one can send at most $\binom{8}{4} = 70$ symbols. A non-orthogonal system, however, can fill up the same bandwidth with a larger number of tones and send as many symbols as is desired thus providing flexibility to the systems designer.

The ability to have different spectral efficiencies in the same bandwidth and symbol rate opens up the prospect of employing adaptive coding and modulation(ACM) techniques where the signaling is adapted based on a particular user's link quality (SNR). Such schemes are common in coherent modulation schemes such as M-PSK and M-QAM [43], [30] and with our proposed system the same advantages can be enjoyed by systems using non-coherent receivers and MT-MFSK. Such adaptive MT-MFSK system requires estimation of SNR so that the system can adapt appropriately and another aspect of our work is in determining estimators and the Cramér-Rao lower bound (CRLB) for SNR estimation of non-orthogonal MT-MFSK.

### 5.1.1 Overview of Prior Work

The use of orthogonal multi-tone MFSK in systems with high dynamics has been proposed by Wetz et al. in [96], and in [73]. In their proposed system, orthogonal MFSK is used in conjunction with OFDM. Blocks of OFDM subcarriers are grouped together and MFSK modulation is carried out on each group of OFDM carrier and the authors show that this is very useful for channels such as those offered by high velocity trains.

Even though there is a large body of literature in the area of orthogonal FSK, to our knowledge, there exists relatively little prior work done in the area of non-orthogonal MT-MFSK. One of the earliest works in the area of non-orthogonal non-coherent multitone FSK was the work by Wittke, Lam and Schefter in 1995 [98]. In their work, the authors look at the performance of trellis coded modulation in non-orthogonal FSK. In their work, the non-orthogonal FSK tones are spaced (uniformly in frequency domain) at spacings of $R_s$ (orthogonal) to $R_s/3$ (non-orthogonal), where $R_s$ is the symbol rate. In contrast, our work will position the frequency tones over the same bandwidth in a non-uniform and non-orthogonal manner and a key component of our work is in selection of the tone positions.

Subsequently, around 2000, Durand, Bejjani and Boutros in [39–42] also contributed in this area of non-orthogonal MFSK. In their work, the authors defined a uniform grid of non-orthogonal tones over the bandwidth of interest. In order to overcome the performance degradation that occurs due to the correlation of these tones, the authors propose the use of multidimensional signals, i.e., signals that exist over multiple time periods. In essence the authors propose a code that helps to overcome the inherent drawback of starting with a highly correlated signal set. In [40], the authors analyzed these multidimensional symbols using non-orthogonal MFSK from an information theoretic point of view and concluded that non-orthogonal signal sets outperform orthogonal signal sets. In [42] and [39], the authors propose specific multidimensional non-orthogonal signal sets that outperform equivalent orthogonal ones.

The key differences from these prior works and ours are that in our system the tones are not uniformly spaced in frequency and that we transmit a symbol in

a single timeslot. However, since our work uses multi-tone signals, our transmit signals do not have a constant envelope.

MT-MFSK (called "Q-tone FSK" by the authors) was also presented in [67] and [68], however, in these works the authors focussed on proving that such schemes achieve the capacity of fading channels at the wideband limit and did not focus on finding optimal receivers and transmission schemes which is our goal. Hosman et al. have implemented a multi-tone non-orthogonal FSK with non-uniform tone spacing in [53]. However, in their work the tone positions were experimentally chosen based on their particular channel of interest (closed shipping containers). The receiver used in [53] was suboptimal and was a proof of concept of the feasibility of a non-uniform spaced non-orthogonal multi-tone FSK system.

SNR estimation has been studied by various authors. We presented the CRLBs and estimators for data aided (DA) and non-data aided (NDA) SNR estimation for M-ary constellations such as M-PSK and M-QAM in a very general multi-antenna framework in Chapter 4 [28,29]. Our present work leverages some of the results in [29] and generalizes the work of Hassan and Ingram in [52] where the authors have presented the results for SNR estimation for orthogonal non-coherent single tone MFSK. We present both CRLBs and estimators for the more general case of non-orthogonal MT-FSK. Moreover, we show that better results are obtained when working with the output of the received matched filters, rather than (as in [52]), the amplitude of the output of the received matched filters. We shall elaborate on this in Section 5.5.

Adaptive variable-rate non-coherent M-FSK modulation was studied in [36]. In this work, the authors have analyzed the gains from switching between different modulation levels at different SNR thresholds. Since they work with orthogonal single tone MT-MFSK, the system discussed is limited to, from Eqn. 5.1, a maximum spectral efficiency of about 0.528 bits/s/Hz. In our work, the maximum spectral efficiency can get significantly higher and we present results for systems with spectral efficiencies up to 1.5bits/s/Hz (although the proposed system is not limited to this).

### 5.1.2 Our Contribution

Our contribution may be summarized as: We present a non-orthogonal multi-tone MFSK communication technique that outperforms the equivalent orthogonal tone MT-MFSK with the same spectral efficiency. The proposed system also offers the systems engineer unlimited choices of spectral efficiency. We show that these systems can adapt its spectral efficiency to the prevailing SNR.

First, we define the proposed system in Section 5.3 and then we evaluate the theoretical maximum performance gain by comparing the capacity of the proposed scheme with that of orthogonal MT-MFSK. Then we present the performance of practical systems using maximum likelihood (ML) receivers, least squared (LS) error receivers as well as compressed sensing (CS) based receivers. Closed form expressions for ML and LS receivers and expressions for the symbol error rate for ML receivers in non-coherent channels are also presented. We evaluate the performance with and without coding, in frequency flat and frequency selective channels.

We also present closed form results for the CRLB and ML estimators for data aided (DA) SNR estimation for non-orthogonal, MT-MFSK as well as closed form approximations for the non-data aided (NDA) model.

## 5.2 Notation

Some of the common notation used throughout this chapter is introduced in this section. Vectors and matrices are represented respectively in lower case bold font and upper case bold font such as $\mathbf{x}$ and $\mathbf{X}$. $x_{i,l}$ is the l-th entry of the vector $\mathbf{x_i}$. $\mathbf{0}_{x,y}$ is an all-zeros matrix of size $x \times y$. $\mathbf{0}_x$ is an all-zeros matrix of size $x \times x$. $\mathbf{I}_x$ represents an identity matrix of size $x \times x$. $vec(\mathbf{H})$ represents all the elements of a matrix $\mathbf{H}$ rearranged in a column vector. $()^*$, $()^T$ and $()^H$ denote the conjugate, transpose and Hermitian operators respectively. $\mathcal{N}(\mu, \mathbf{R_{nn}})$ denotes the pdf of a circular complex Gaussian random variable of mean $\mu$ and covariance $\mathbf{R_{nn}}$. $P(X|Y)$ is the conditional probability density function for the random variable $X$ given random variable $Y$. $P_X(x)$ is the probability density function of random

variable $X$.

Table 5.1 gives a list of some of the common symbols used throughout this chapter.

**Table 5.1**: Common symbols used

| **Symbol** | **Definition** |
|---|---|
| $R_s$ | Symbol rate in Hz. |
| $T_s$ | Symbol duration in seconds. Equals $\frac{1}{R_s}$s. |
| $W$ | Total bandwidth of system in Hz. Equals $NR_s$Hz |
| N | Number of orthogonal tones that span bandwidth $W$. Equals $\frac{W}{R_s}$. |
| K | Tone packing factor. |
| NK | Total number of tones available for use. |
| M | Number of symbols. |
| Q | Number of tones transmitted per symbol. |
| $\binom{NK}{Q}$ | Maximum number of symbols possible. |
| $\nu$ | Signal to noise ratio (SNR). |
| $\mathbf{s_i^{(m)}}$ | Transmit symbol vector at time $i$ of size $N \times 1$. m-th symbol from a set of M symbols. |
| $\mathbf{v^{(m)}}$ | Binary vector of size $NK \times 1$ corresponding to $\mathbf{s_i^{(m)}}$. 1 in positions corresponding to transmit tones of symbol $\mathbf{m}$. 0 elsewhere. |
| $\mathbf{r_i}$ | Output of received matched filters at time $i$. Size $N \times 1$. |
| $\mathbf{n_i}$ | Noise output of received matched filters at time $i$. Size $N \times 1$. |
| $\rho_\mathbf{i}$ | Correlation between $i$-th tx symbol and orthonormal receive basis set. Size $N \times 1$. |

## 5.3 System Model

The overall system block diagram is shown in Fig.5.2. As is common to most systems, information bits first undergo forward error correction (FEC) encoding after which they are mapped to symbols. Since an adaptive system is proposed, the constellation is chosen by a block that uses the link quality (in the form of SNR information sent by the receiver) as a metric to choose the appropriate constellation which would maximize the spectral efficiency while achieving the bit error rate desired. The transmitter maps symbols to transmit samples that are then transmitted at the sampling rate over the channel to the receiver. On the receiver, the reverse process takes place. The received signal is first projected into an orthonormal basis that spans the transmit symbol set and then the receiver and FEC decoder is jointly carried out. SNR estimation is also conducted and the signal quality is sent back to the transmitter over the reverse link. Fig. 5.2 shows a functional block diagram and leaves out the analog-to-digital (A/D) and digital-to-analog (D/A) converters and the RF blocks that are needed by all wireless systems.

### 5.3.1 Transmitter

The system model uses a point to point transmitter-receiver pair. The system transmits symbols at a symbol rate of $R_s$ Hz. Each symbol has $Q$ tones simultaneously transmitted in the same symbol period, $T_s$ where $T_s = 1/R_s$. These tones come from a set of $NK$ non-orthogonal uniformly spaced tones, each separated in frequency by $R_s/K$ Hz. The total system bandwidth is thus $[0, NR_s]$ Hz which is the same as would be required by an orthogonal system with $N$ tones and a symbol rate of $R_s$. Not all these $NK$ tones may be used by all the symbols in the constellation. $K$ is the tone packing factor (also known as the non-orthogonality factor). For an orthogonal system with $N$ orthogonal tones, $K = 1$. For a system with tones packed twice as dense as in an orthogonal system, $K = 2$. All the systems described assume a total available bandwidth of unity (or 1 Hz) and for fair comparison, each symbol is normalized to have the same power and all the

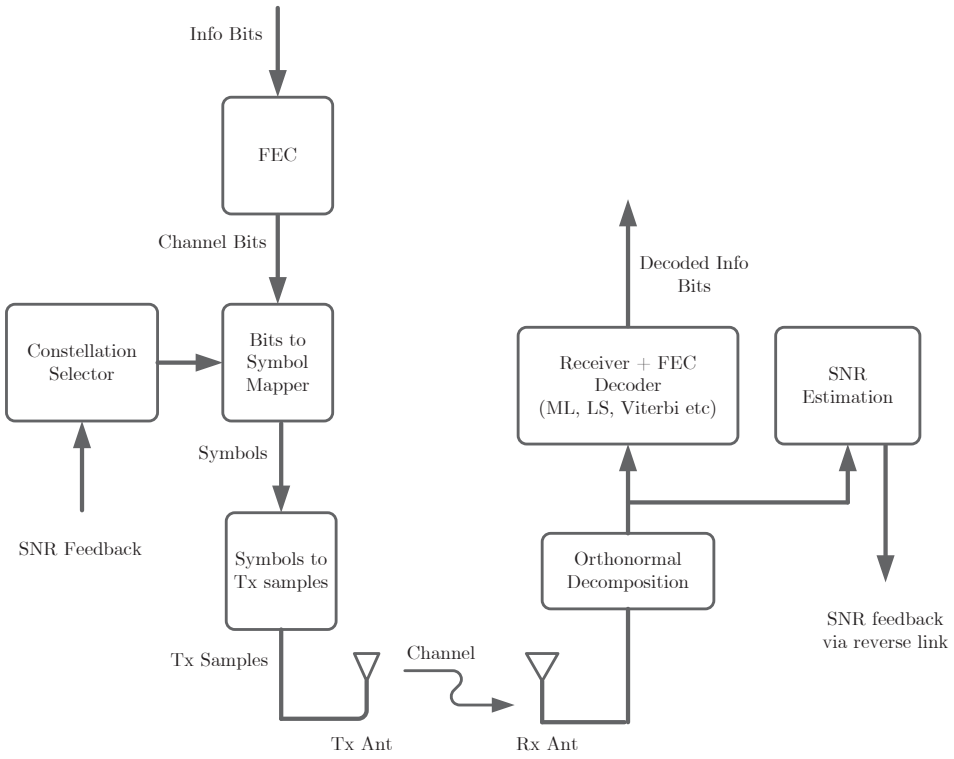


**Figure 5.2**: System block diagram for rate adaptive MT-MFSK

constellations compared here are normalized to have unit power.

The system has $M$ symbols, $\mathbf{s^{(0)}}, \mathbf{s^{(1)}}, ..., \mathbf{s^{(M-1)}}$ and for each symbol, $\mathbf{s^{(m)}}$, we associate an unique transmit binary vector of size $NK \times 1$, $\mathbf{v^{(m)}}$. This vector contains 1s in those positions corresponding to the tones that need to be transmitted for that particular symbol, $\mathbf{s^{(m)}}$ and 0s elsewhere. For a system with $NK$ tones, and $Q$ tones per symbol, there are $\binom{NK}{Q}$ possible symbols. However, in systems we shall consider, only a subset of these possible symbols are used.

In discrete time the Nyquist sampling rate is $N$ samples per symbol and thus, the $n$-th element of the $N \times 1$ vector $\mathbf{s^{(m)}}$ is given by:

$$s_n^{(m)} = \frac{1}{E_m} \sum_{p=0}^{NK-1} v_p^{(m)} \exp\left(\frac{j2\pi pn}{NK}\right) \tag{5.2}$$

where $E_m$ is a normalizing term which ensures that each of the transmit symbols

has unit power, i.e. $\mathbf{s^{(m)}}^H\mathbf{s^{(m)}} = 1$ for all $m = [0, 1, ..., M-1]$.

In vector notation the $N \times 1$ elements of the $m$-th transmitted symbol can be written as

$$\mathbf{s^{(m)}} = \frac{1}{E_m}\mathbf{\Phi}\mathbf{v^{(m)}} \tag{5.3}$$

where $\mathbf{\Phi}$ is a $N \times NK$ matrix whose $m$-th row, $n$-th column entry is $\exp\left(\frac{j2\pi mn}{NK}\right)$.

**Symbol Set Design**

A key design problem for MT-MFSK is to determine the set of symbols and the assignment of tones to symbols. Each symbol is transmitted at unit power, so there are no unfair advantages stemming from coherent addition of correlated transmitted tones. The key to minimizing probability of error is to minimize the worst case correlation between any two symbols of the constellation. In the orthogonal framework when multiple (orthogonal) tones are sent per symbol, the symbols themselves are not orthogonal. As an example, if an $N$-tone orthogonal MT-MFSK system uses all combinations of $Q = 2$ tones per symbol to create $\binom{N}{2}$ symbols, then even though the tones are orthogonal, the worst case correlation between the symbols is 0.5 (since 2 symbols may have a common tone). If a symbol consists of 3 orthogonal tones simultaneously transmitted, then the worst case correlation is 0.667 (since multiple symbols may have 2 common tones).

Thus, in order to have a better symbol error probability than the equivalent orthogonal 2-tone per symbol system, the non-orthogonal system with the same (or better) spectral efficiency needs to have a worst case correlation lower than 0.5. To outperform an equivalent orthogonal 3-tone per symbol system, the non-orthogonal system needs to have a worst case correlation lower than 0.667 and so on.

Our signal set design algorithm sets various values of $K$ and then searches for a symbol set of a specified size ($M$) with a specified number of tones per symbol to have a worst case correlation lower than some threshold. To give a specific design example, we designed a system to outperform a $N = 6$, $K = 1$, $Q = 2$ orthogonal

system which has $M = \binom{6}{2} = 15$ symbols. In our non-orthogonal design we chose $K = 20$, thus allowing for a total of $NK = 120$ tones in the system. Each symbol still used $Q = 2$ tones/symbol, thus there are a total of $\binom{120}{2} = 7140$ possible symbols. The design problem, is thus to choose a set of 15 of these 7140 possible symbols such that the worst case correlation will be less than 0.5. To do this exhaustively would require a search over $\binom{7140}{15} = 4.8 \times 10^{45}$ possible sets which is clearly prohibitive even for this relatively small system design so we employ smart search techniques that intelligently prune the set of possible sets to search over. The search is, however, sub-optimal and is not guaranteed to find the best set.

Fig. 5.3 shows a solution set of 15 symbols with a worst case correlation of 0.48. The red squares in the figure show the two tones that are activated for each symbol. Symbol number 1, e.g., uses tone numbers 19 and 75, while symbol number 2 uses tone numbers 1 and 41. The right hand subplot of Fig. 5.3 shows all the tones activated in this system. It can be seen that not all the 120 tones are used and that the set of tones used in the system are not uniformly spaced. Unlike in orthogonal MT-MFSK, no two symbols have a common tone. In Section 5.6 we shall discuss the performance gains stemming from this reduced correlation.

### 5.3.2 Channel Model

For most of the simulations, a non-coherent flat fading channel is assumed. i.e. the received signal has an unknown phase that is constant across the symbol and is independent from symbol to symbol. In some of the simulations, a frequency selective channel, where the channel is independent at each of the transmitted tones is also considered. Flat and frequency selective Rayleigh fading channels – where both the phase and gain of the received signal are distorted by the channel are also used. The output of the channel if the $m$-th symbol from the constellation was transmitted is given by Eqn. 5.4 where $c_p$ is a complex valued channel gain affecting tone $p$, $\theta_p$ is the channel phase assumed to be uniformly distributed between 0 and $2\pi$ radians and $n$ is the additive white Gaussian noise.

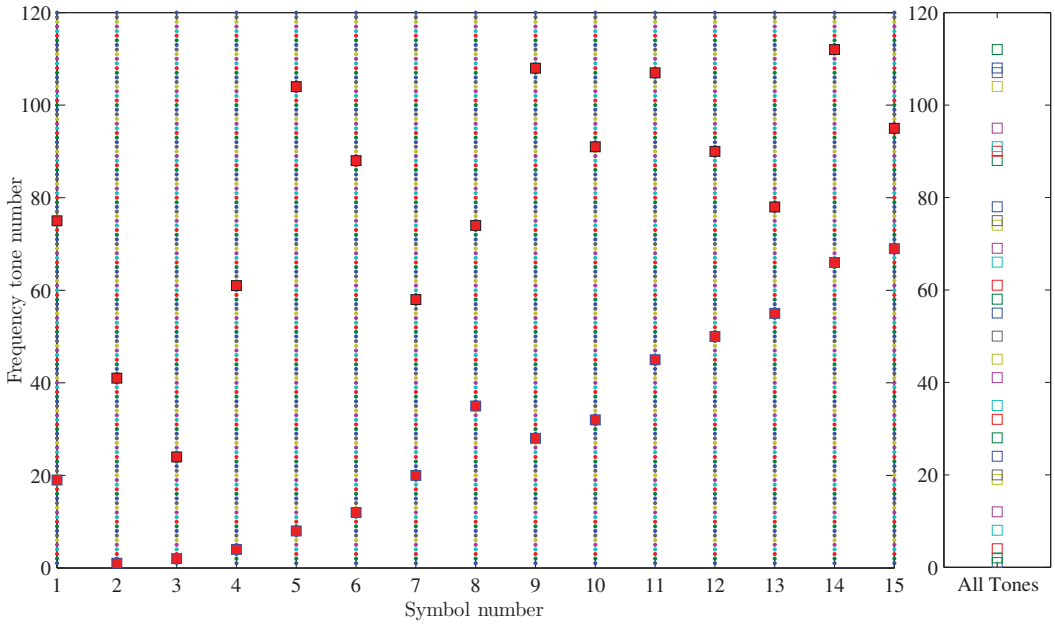


**Figure 5.3**: Assignment of 2 tones per symbol in an exemplary non-orthogonal system [red squares indicate tones that are used]. N = 6, Q = 2, K = 20, M = 15

$$y_n^{(m)} = \frac{1}{E_m} \sum_{p=0}^{NK-1} v_p^{(m)} c_p \exp\left(\frac{j2\pi pn}{NK} + \theta_p\right) + n \tag{5.4}$$

For the flat fading model, $c_p$ and $\theta_p$ are the same for all $p$ and so the dependence on $p$ in the notation will be omitted. For the frequency selective channel model, it is assumed that $c_p$ and $\theta_p$ are independent and identically distributed (i.i.d.) for all $p$. If only phase is unknown at the receiver, then $c_p = 1$ for all $p$.

If the $m$-th symbol was transmitted, the output of the channel can be expressed in vector notation as:

$$\mathbf{y^{(m)}} = \frac{1}{E_m}\mathbf{\Phi G v^{(m)}} + \mathbf{n} = \mathbf{\Phi \tilde{v}^{(m)}} + \mathbf{n} \tag{5.5}$$

where $\mathbf{G}$ is a $NK \times NK$ diagonal matrix with the complex channel coefficients along its diagonal and $\mathbf{\tilde{v}^{(m)}} = \frac{1}{E_m}\mathbf{G v^{(m)}}$.

Although $\mathbf{\tilde{v}^{(m)}}$ is a vector of size $NK \times 1$, only $Q$ elements are non-zero. Typically $NK$ will be much larger than $Q$ and thus, the vector $\mathbf{\tilde{v}^{(m)}}$ is sparse. Moreover the system as described in Eqn. 5.5 is an underdetermined linear system.

This fact will be utilized later in the development of the compressed sensing (CS) based receiver described in Section 5.3.3.

### 5.3.3 Receivers

Three general classes of receivers will be analyzed in this section. We shall first present the formulation of the maximum likelihood (ML) receiver and then present a minimum least squared error (LS) receiver [59]. Finally we shall present a compressed sensing (CS) based receiver.

#### Projection on Orthonormal Basis

The received signal is first received by a bank of filters which form an orthonormal basis for the signal space. Since the $N$ orthogonal tones that span this space form a basis, the received signal is projected on this $N$ dimensional basis set by passing it through a filter bank (which may be efficiently implemented via a fast Fourier transform (FFT)) and then downsampling to the symbol rate. This $N \times 1$ vector, $\mathbf{r_i} = [r_{0,i}, r_{1,i}, ..., r_{N-1,i}]^T$ is the input to the detector at time $i$ as shown in Fig. 5.4. Since the channel is assumed to be flat fading and the modulation doesn't have memory, symbol by symbol detection is optimal. When coding is used (See Section 5.3.4), the entire block of received vectors for the entire code block is passed to the decoder simultaneously.

#### Maximum Likelihood(ML) Receivers

In the non-coherent channel with unknown phase, the ML receiver determines the likelihood of each of the transmitted symbols averaged over all possibilities of the unknown phase. The ML detector at the $i$-th time slot will calculate each of $p(\mathbf{r_i}|\mathbf{s^{(0)}})$, $p(\mathbf{r_i}|\mathbf{s^{(1)}})$, ..., $p(\mathbf{r_i}|\mathbf{s^{(M-1)}})$ and choose the symbol corresponding to the maximum of these $M$ values. Thus, the ML estimate is given by:

$$\widehat{\mathbf{s}}_{ML} = \arg \max_{m=0..M-1} p(\mathbf{r_i}|\mathbf{s^{(m)}}) \tag{5.6}$$

Note that even though the symbols themselves are not orthogonal, since $\mathbf{r_i}$ is the output of an orthonormal basis, the *noise* components in $\mathbf{r_i}$ are orthogonal

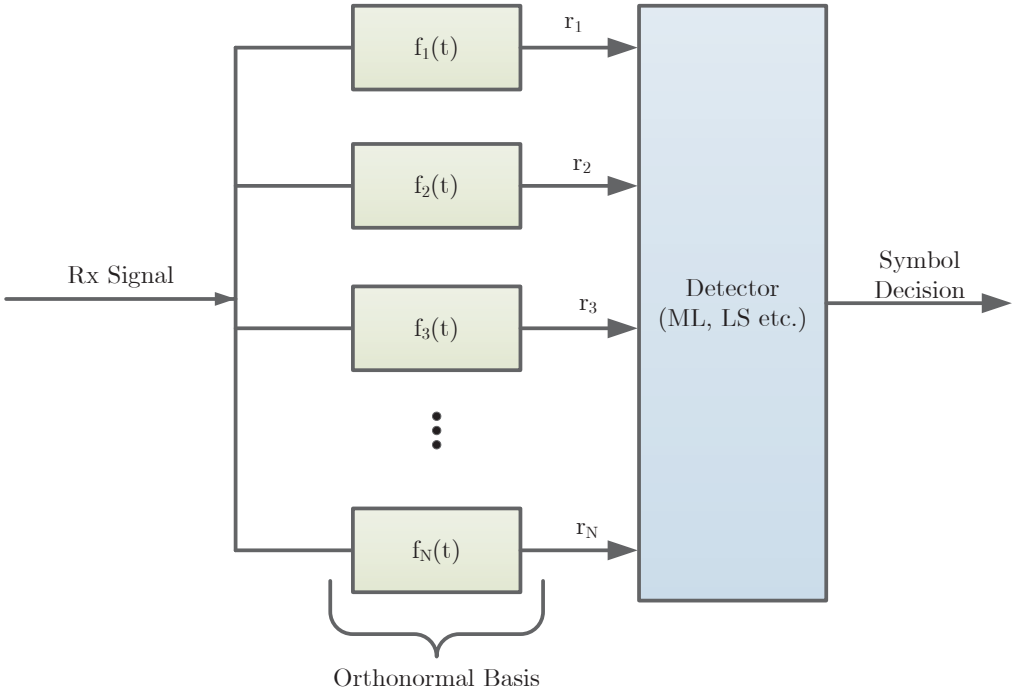


**Figure 5.4**: ML detector structure

and thus the noise covariance matrix, $\mathbf{R_{nn}}$ is diagonal and invertible. The expected value of the received vector, $\mathbf{r_i}$ for a particular transmitted symbol, $\mathbf{s^{(m)}}$, and an unknown phase shift imposed by the channel, $\phi$, is given by

$$E\left[\mathbf{r_i}|\mathbf{s^{(m)}}\right] = \begin{bmatrix} \rho_{m0} \\ \rho_{m1} \\ \cdots \\ \rho_{m(N-1)} \end{bmatrix} e^{j\phi} \tag{5.7}$$

where the $N \times 1$ vector $[\rho_{m0}, \rho_{m1}, ...\rho_{m(N-1)}]^T$ is the decomposition of the transmitted symbol $\mathbf{s^{(m)}}$ in the orthonormal basis set spanning the space. Note that for usual orthogonal FSK, only one of $[\rho_{m0}, \rho_{m1}, ...\rho_{m(N-1)}]^T$ is non-zero while for our system design with non-orthogonal tones, all the elements are likely to be non-zero. Since one of the key advantages of using MFSK is in the domain of non-coherent receivers, none of the receivers presented will rely on the knowledge

of the unknown phase offset $\phi$. It is assumed that no training symbols are used and while a fast automatic gain correction (AGC) keeps the gain of the system constant, there is no compensation for the unknown phase.

The likelihood function for a non-coherent channel with an unknown phase offset, $\phi$, can be calculated using the following approach where $\mu_{\mathbf{m}} = \rho_{\mathbf{m}} \exp^{j\phi}$:

$$p(\mathbf{r}|\mathbf{s^{(m)}}) \tag{5.8a}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} p(\mathbf{r}|\mathbf{s^{(m)}}, \phi) d\phi \tag{5.8b}$$

$$= \underbrace{\frac{1}{2\pi^{m+1}|\mathbf{R_{nn}}|}}_{\alpha} \int_0^{2\pi} \exp\left(-\left(\mathbf{r^H R_{nn}^{-1} r} + \mu_{\mathbf{m}}{}^{\mathbf{H}} \mathbf{R_{nn}^{-1}} \mu_{\mathbf{m}} - \mu_{\mathbf{m}}{}^{\mathbf{H}} \mathbf{R_{nn}^{-1}} \mathbf{r} - \mathbf{r^H R_{nn}^{-1}} \mu_{\mathbf{m}}\right)\right) d\phi \tag{5.8c}$$

$$= \frac{1}{\alpha} \exp\left(\underbrace{-\mathbf{r^H R_{nn}^{-1} r} - \rho_{\mathbf{m}}^{\mathbf{H}} \mathbf{R_{nn}^{-1}} \rho_{\mathbf{m}}}_{\beta}\right) \int_0^{2\pi} \exp\left(e^{-j\phi} \underbrace{\rho_{\mathbf{m}}^{\mathbf{H}} \mathbf{R_{nn}^{-1}} \mathbf{r}}_{\lambda_m = \lambda_c + j\lambda_s} + \mathbf{r^H R_{nn}^{-1}} \rho_{\mathbf{m}} e^{j\phi}\right) d\phi \tag{5.8d}$$

$$= \frac{2\pi}{\alpha} \exp(\beta) I_0(2|\lambda_m|^2) \tag{5.8e}$$

Since $I_0$, the modified Bessel function of the first kind, is a monotonically increasing function, in order to determine $\arg\max_{m=0..M-1} p(\mathbf{r_i}|\mathbf{s^{(m)}})$, it is sufficient to determine $\arg\max(|\lambda_m|^2)$. Thus the ML receiver calculates each of the $M$-values of $|\lambda_m|^2$, one for each symbol and then chooses the symbol corresponding to the maximum value of $|\lambda_m|^2$. This is a more general version of the well known envelope detector (see [74]) which is the optimal receiver for the single tone orthogonal MFSK. The expression derived above is exactly identical to the envelope detector when $\rho_{\mathbf{m}}$ has a 1 only in the $m$-th position and 0s everywhere else, as would be the case for orthogonal single tone MFSK.

Although there is no closed form expression for the symbol error rate (SER) for the ML receiver can derive the SER in the following manner. First we note that for equiprobable symbols the probability of a correct decision $P(C) = \sum_{m=0}^{M-1} P(C|\mathbf{s^{(m)}}) P(\mathbf{s^{(m)}}) = \frac{1}{M} \sum_{m=0}^{M-1} P(C|\mathbf{s^{(m)}})$. In turn $P(C|\mathbf{s^{(m)}}) = \int_{-\infty}^{\infty} P(C|\mathbf{s^{(m)}}, \lambda_m = \gamma) P_{\lambda_m}(\gamma) d\gamma$ where $P(C|\mathbf{s^{(m)}}, \lambda_m = \gamma)$ is the probability of a correct decision given

a particular value of $\lambda_m = \gamma$. This implies that for all $k \neq m$, $|\lambda_k|^2 < |\gamma|^2$. Thus SER may be expressed as:

$$SER = 1 - P(C) = 1 - \frac{1}{M}\sum_{m=0}^{M-1} P(C|\mathbf{s^{(m)}}) \tag{5.9a}$$

$$= 1 - \int_{-\infty}^{\infty}\int_{R_0}\int_{R_1}\int_{R_{m-1}}\int_{R_{m+1}} ... \int_{R_{M-1}} P_{\lambda_m}(\gamma) d\lambda_0 ... d\lambda_{M-1} d\gamma \tag{5.9b}$$

$\lambda_k$ is a scalar complex Gaussian random variable whose distribution is given by $\mathcal{N}(\mu, \sigma^2)$ where the mean, $\mu = \rho_{\mathbf{k}}{}^H \rho_{\mathbf{m}}$ and $\sigma^2$ is the noise variance at the output of the matched filters. Each of the regions of integration $R_k$ is given by the region where $|\lambda_k|^2 < |\gamma|^2$. Although the above expression cannot be evaluated in closed form, numerical evaluations or Monte Carlo evaluation may be carried out.

**LS Receivers**

The above expressions have been carried out for a non-coherent flat fading channel. In these channels, there is a constant and unknown phase offset, $\phi$, that is constant across all the tones. For frequency selective channels with unknown phase, the phase offset, is not constant across the signal bandwidth and in the extreme case is independent from tone to tone. There is no closed form expression for the ML receiver for this channel model and indeed if the above receiver is applied to a signal that goes through a frequency selective channel, then the performance is poor. However, the LS receivers described next are more useful in those channels and we derive those next.

The objective of the LS receiver is to select the symbol that minimizes the least squared error. Thus, if $e_m$ is the LS error for transmitted symbol $\mathbf{s^{(m)}}$ then LS solution is given by:

$$\widehat{\mathbf{s}}_{LS} = \arg \min_{m=0..M-1} e_m \tag{5.10}$$

If the indices of the $Q$ transmit tones of the $i$-th symbol are $j_1, j_2, ..., j_Q$ then after being subjected to a frequency selective channel the $N \times 1$ input to the LS detector at time $i$, $\mathbf{r_i}$ may be expressed as

$$\mathbf{r_i} = \frac{1}{EN}\mathbf{H}\begin{bmatrix}\alpha_1\\ \alpha_2\\ ...\\ \alpha_Q\end{bmatrix} + \mathbf{n_i} \tag{5.11}$$

where $E$ is a normalizing constant and $\mathbf{H}$ is given by

$$\begin{bmatrix} \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_1}{K}-0)\frac{n}{N}} & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_2}{K}-0)\frac{n}{N}} & ... & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_Q}{K}-0)\frac{n}{N}} \\ \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_1}{K}-1)\frac{n}{N}} & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_2}{K}-1)\frac{n}{N}} & ... & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_Q}{K}-1)\frac{n}{N}} \\ ... & ... & ... & ... \\ \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_1}{K}-\tilde{N})\frac{n}{N}} & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_2}{K}-\tilde{N})\frac{n}{N}} & ... & \sum_{n=0}^{N-1} e^{j2\pi(\frac{j_Q}{K}-\tilde{N})\frac{n}{N}} \end{bmatrix} \tag{5.12}$$

and $\tilde{N} = N-1$ and each of $\alpha_1$, $\alpha_2$, ... $\alpha_Q$ are the $Q$ channel coefficients corresponding to each of the transmitted tones. If phase is the only part of the channel that is unknown, then the channel coefficients have a known constant gain but unknown phase. If the channel is assumed to be flat in frequency, then the $Q$ coefficients are identical to each other.

It may be observed that the $N \times Q$ matrix $\mathbf{H}$ is constant for each transmit symbol. In fact, one can compute a different $\mathbf{H_m}$ corresponding to each of the $M$ transmitted symbols in the constellation. Assuming symbol $m$ was transmitted, the scalar least squared error $e_m$ is given by [59]:

$$e_m = \mathbf{r_i}^H \left(\mathbf{I_N} - \mathbf{H_m}(\mathbf{H_m}^H\mathbf{H_m})^{-1}\mathbf{H_m}^H\right)\mathbf{r_i} \tag{5.13}$$

The LS receiver chooses the transmit symbol that minimizes $e_m$ over all $m$. In the above equation, $\left(\mathbf{I_N} - \mathbf{H_m}(\mathbf{H_m}^H\mathbf{H_m})^{-1}\mathbf{H_m}^H\right)$ is a $N \times N$ matrix that can be precomputed for each $m = [0, 1, ..., (M-1)]$ and thus the computational burden on the LS receiver during operation is much less than what would be implied by the full evaluation of Eqn. 5.13.

For the flat-fading channel with unknown phase, the LS receiver and the ML receiver yield the same result. However, the LS receiver is more appropriate during a frequency selective channel, especially when both the gain and phase

of each tone are unknown. Even though the LS receiver attempts to "fit" the observation to the model of Eqn. 5.11, and thus indirectly estimates the channel coefficients, no system resources are expended in this effort.

**CS based receiver**

One of the disadvantages of the ML and LS receivers is that their complexity grows with the size of the constellation. Both the ML and LS receivers compute a metric (respectively, the likelihood and the squared error) for each of the possible $M$ symbols and then determines the symbol that either maximizes the likelihood or minimizes the squared error. As $M$ increases, the computational complexity of this approach also increases. Even though for practical SNRs, the number of symbols in the constellation of choice is unlikely to get too big, we present in this section receivers that have a reduced complexity of implementation.

The receiver presented here is a compressed sensing based receiver that is motivated by the subset selection approach for compressed sensing discussed in [76]. In this work, as in our present problem, the authors work with an underdetermined linear system as may be seen by inspecting Eqn 5.5. With the a priori knowledge that only $Q$ elements of the vector $\tilde{\mathbf{v}}^{(\mathbf{m})}$ are non-zero, we can apply many of the techniques from the compressed sensing literature to the problem at hand.

CS techniques assume that the solution to the problem is sparse. In our context, this would imply that $Q \ll NK$. Since most practical systems would use a handful of tones per symbol and $NK$ is likely to be hundred (or more), this assumption is indeed reasonable in this context. As shown in [76], the CS receiver determines $\arg\min_{\mathbf{v}} J(\widetilde{\mathbf{v}})$ where

$$J(\widetilde{\mathbf{v}}) = ||\mathbf{\Phi}\widetilde{\mathbf{v}} - \mathbf{y}_i||^2 + \gamma \sum_{i=1}^{NK} |\widetilde{\mathbf{v}}[i]|^p \tag{5.14}$$

where, $\mathbf{\Phi}$, $\widetilde{\mathbf{v}}$ are given by Eqn. 5.5, and $\gamma$ is a weighting factor that controls the tradeoff between the quality of the fit and the sparsity of the solution. For $p = 1$, the solution approaches the $l1$ norm minimization problem. In general, the CS approach is a suboptimal computationally simpler way of estimating the LS solution without doing an exhaustive search.

Of the many CS techniques available, we propose using the regularized version of the FOCal underdetermined system solver (FOCUSS) algorithm presented by Rao et al. in [76]. As will be discussed in Section 5.6.3, the CS receiver can be further improved if a ML or LS search is done in the immediate neighborhood of the solution of the CS receiver. The main advantage of the CS receiver is that the computational complexity does not increase as the number of symbols grows. As shown in [69], the complexity of various CS algorithms is considerably less than $\mathcal{O}(\binom{NK}{Q})$. For instance, two popular algorithms, $l_1$ norm minimization has a complexity of $\mathcal{O}(N^2(NK)^{(3/2)}))$, and orthogonal matching pursuit (OMP) has a complexity of $\mathcal{O}(QKN^2)$ As the problem scales to a large number of symbols, a CS based receiver may provide options for a lower complexity receiver with relatively little performance loss compared to a LS receiver. The performance of a CS receiver will be discussed in Section 5.6.3.

### 5.3.4 FEC Coding

To demonstrate the performance of the proposed system with coding, we have used the famous Cassini convolution code from NASA which is a rate $\frac{1}{6}$ code with a constraint length of 15. The octal representation of each of the 6 encoders is $\{042631, 047245, 073363, 056507, 077267, 064537\}$. The code is decoded using a soft decision Viterbi decoder [74] which is known to be the optimal receiver for such systems.

## 5.4 System Capacity

In this section, we investigate the system capacity of the transmission schemes proposed above. As is well known, the capacity of a channel is defined as the maximum mutual information, $I(\mathbf{X};\mathbf{Y})$, between the input, $\mathbf{X}$, and output, $\mathbf{Y}$, of a channel maximized over all possible input distributions [24]. In cases where the the input constellation is known, i.e., the input distribution is known, the capacity is simply the mutual information $I(\mathbf{X};\mathbf{Y})$.

Since, in all cases, the received signal shall initially be received by a bank

of $N$ matched filters (equivalent to projecting the received signal on a orthonormal basis that spans the received signal space), we shall evaluate the mutual information between the output of the matched filters, $\mathbf{r}$, and the input waveform, $\mathbf{s}$, which is assumed to come from a $M$-ary constellation with equiprobable symbols.

For this model, the capacity is given by:

$$C = \log_2(M) + E\left[\sum_{m=0}^{M-1} p(\mathbf{s^{(m)}}|\mathbf{r}) \log_2 p(\mathbf{s^{(m)}}|\mathbf{r})\right] \tag{5.15}$$

The expectation shown in Eqn. 5.15 can be evaluated using Monte Carlo simulations [40]. In order to carry this out, it is simplest to apply Bayes rule and express, $p(\mathbf{s^{(m)}}|\mathbf{r})$ as:

$$p(\mathbf{s^{(m)}}|\mathbf{r}) = p(\mathbf{r}|\mathbf{s^{(m)}}) / \sum_{j=0}^{M-1} p(\mathbf{r}|\mathbf{s^{(j)}}) \tag{5.16}$$

The capacity of various input constellations may be thus be easily compared by calculating the capacity for each set of $\mathbf{s^{(m)}}$.

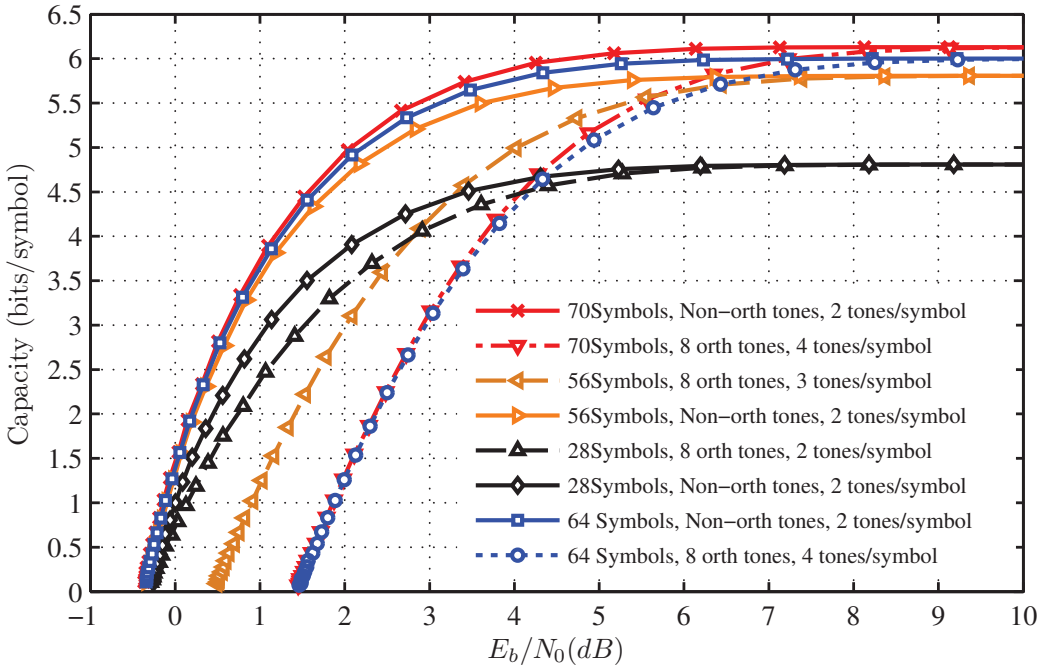


**Figure 5.5**: Non-orthogonal MT-MFSK increases system capacity in AWGN

Fig. 5.5 shows the capacity in an AWGN channel for both traditional multi-tone MFSK systems as well as the proposed non-orthogonal tone FSK system. Fig. 5.5 compares an orthogonal system with 8 orthogonal carriers (sending 2, 3 or 4 tones at a time, resulting in systems with 28, 56 and 70 symbols respectively), to three proposed systems with the same number of symbols, but using non-orthogonal tones. Since the number of symbols in most practical systems is typically a power of 2, two systems with 64 symbols are also compared. The 64-symbol system is obtained by selectively not using 6 of the worst symbols from the 70-symbol system. The exact mapping of the non-orthogonal tones to symbols for the proposed system is given in Appendix 5.A. As can be seen, the performance gains are significant, with some non orthogonal configurations achieving the same bit rate as the orthogonal systems with as much as 2.5dB less transmit power. Looking in the other dimension, at an $E_b/N_0 = 2dB$, the capacity with an orthogonal tone 64 symbol system is about 1.25bits/symbol whereas the capacity for a 64-symbol non-orthogonal system in the same bandwidth and symbol rate is just under 5bits/symbol – **a 4 fold increase** in system capacity.

Although Fig. 5.5 shows the gains that arise in an AWGN channel, we are typically more interested in the non-coherent case where there is an unknown phase offset across the whole symbols. In that case, $p(\mathbf{r}|\mathbf{s^{(m)}})$ needs to be obtained by averaging over the distribution of the unknown phase. This was evaluated in Eqn. 24 of [20] and was used by us for the simulations in Sec. 5.6.1 where we show the increased capacity of non-orthogonal MT-MFSK in a non-coherent system model. A similar approach was also taken by Céline Durand in [42] in evaluating the capacity of non-orthogonal signaling that was spread over multiple symbol durations.

## 5.5 SNR Estimation

In this section we shall derive the Cramér-Rao lower bound (CRLB) and maximum likelihood (ML) estimators for general MFSK. The expressions derived below are applicable to orthogonal and non-orthogonal FSK, to single tone and

multitone FSK. As described in Chapter 3, the estimation of signal to noise ratios (SNR), can be broadly divided into two different approaches – data aided (DA) and non-data aided (NDA). The DA approach assumes that the transmitted symbols are known perfectly while the NDA approach assumes that the transmitted symbols are unknown, but that they belong to a known discrete constellation.

### 5.5.1 DA SNR Estimation

The DA SNR estimation model used assumes a flat block fading model where a block of $L$ known MT-MFSK symbols are assumed to have the identical channel and noise characteristics. The noise and channel are independent of each other and independent of the transmitted symbols. The noise and channel are also assumed to be i.i.d from symbol to symbol. The received samples at the output of the receiver may be written as

$$\mathbf{r}_i = \alpha\rho_i + \mathbf{n}_i \tag{5.17}$$

where $\mathbf{r}_i$ and $\mathbf{n}_i$ are, respectively, the $N \times 1$ received vector and noise vector at time $i$, $\alpha$ is the scalar complex channel and $\rho_i$ is the correlation of the transmit symbol with the orthonormal receive basis as described in Sec. 5.3.3. Since the transmitted symbols are of unit power, the SNR is then $\nu = \frac{|\alpha|^2}{\sigma^2}$ where $\sigma^2$ is the noise variance (which also needs to be estimated). Since $\alpha$ is a complex quantity, the parameter vector of interest that needs to be evaluated is $\mathbf{\Theta} = [\alpha, \alpha^*, \sigma^2]^T$.

Taking an approach very similar to what we demonstrated in [28] and [29] and presented in Chapter 4, we can show that the normalized Cramér-Rao lower bound (NCRLB) for SNR estimation is given by

$$\text{NCRLB} = \frac{\text{CRLB}}{\nu^2} = \frac{2}{\nu L} + \frac{1}{NL} \tag{5.18}$$

The maximum likelihood (ML) estimator for MFSK SNR estimation is also straightforward. If all the received symbols in the block of length $L$ are stacked in a $NL \times 1$ sized vector $\tilde{\mathbf{r}} = \alpha\tilde{\rho} + \tilde{\mathbf{n}}$ where $\tilde{\mathbf{r}}$, $\tilde{\rho}$ and $\tilde{\mathbf{n}}$ are respectively $[\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_L]^T$, $[\rho_1, \rho_2, ...\rho_L]^T$ and $[\mathbf{n}_1, \mathbf{n}_2, ...\mathbf{n}_L]^T$, then the ML estimates of $\sigma^2$, $\alpha$ and $\nu$ are given

by:

$$\widehat{\alpha_{ML}} = (\tilde{\rho}^H \tilde{\rho})^{-1} \tilde{\rho}^H \tilde{\mathbf{r}} \tag{5.19a}$$

$$\widehat{\sigma^2_{ML}} = \frac{1}{NL} \sum_{i=1}^{L} \left(\mathbf{r}_i - \widehat{\alpha_{ML}} \rho_i\right)^H \left(\mathbf{r}_i - \widehat{\alpha_{ML}} \rho_i\right) \tag{5.19b}$$

$$\widehat{\nu_{ML}} = \frac{\widehat{\alpha_{ML}}}{\widehat{\sigma^2_{ML}}} \tag{5.19c}$$

In [52], the authors also determine the DA SNR for non-orthogonal MFSK. However the authors work with the magnitude squared of the output of the received filters, i.e., they work with $\mathbf{r}_i^2$, while we work with $\mathbf{r}_i$ and consequently we obtain a better (lower) limit for the CRLB.

### 5.5.2 NDA SNR Estimation

In the non-data aided case, there is no closed form solution for the CRLB. The received signal is represented by a Gaussian mixture density. The log-likelihood function for each of the i.i.d. received samples is given by:

$$q = \ln[p(\mathbf{r}_i; \Theta)] = -\ln(M) + \ln\left[\sum_{l=0}^{M-1} p(\mathbf{r}_i|\rho_l, \Theta)\right] \tag{5.20}$$

where $p(\mathbf{r}_i|\rho_l, \Theta) = \frac{1}{\pi^N \sigma^{2N}} \exp(-\frac{\phi_{i,l}}{\sigma^2})$ and $\phi_{i,l} = \left(\mathbf{r}_i - \alpha\rho_l\right)^H \left(\mathbf{r}_i - \alpha\rho_l\right)$ Then following the steps outlined in Chapters 3 and 4 and carrying out the final expectation necessary to evaluate the Fisher information matrix using a Gauss-Hermite quadrature, we evaluate the CRLB for NDA SNR estimation non-orthogonal MT-MFSK. The results are shown in Fig. 5.22.

For low SNR and high SNR regions we derived closed form approximations for the normalized CRLB as shown in Eqns. 5.21 and 5.22.

$$\text{NCRLB\_Low\_SNR} = \frac{2N}{\nu L} + \frac{1}{NL} \tag{5.21}$$

$$\text{NCRLB\_High\_SNR} = \frac{2}{\nu L} + \frac{1}{NL} \rightarrow \frac{1}{NL} \tag{5.22}$$

## 5.6 Simulation Results

### 5.6.1 Capacity of non-orthogonal MT-MFSK systems

The capacity of the proposed MT-MFSK systems in a non-coherent environment with an unknown phase (assumed to be constant over a whole symbol) was evaluated using the method of Monte-Carlo simulations as described in Sec. 5.4. As two illustrative examples, we evaluate the capacity of two unit bandwidth (1Hz) systems: a) with symbol rate $R_s = 1/2$ Hz, and b) with symbol rate $R_s = 1/8$ Hz. In the first case, if only orthogonal tones were used, the bandwidth would be occupied by $N = 2$ orthogonal tones while in the second, the bandwidth would be occupied by $N = 8$ orthogonal tones.

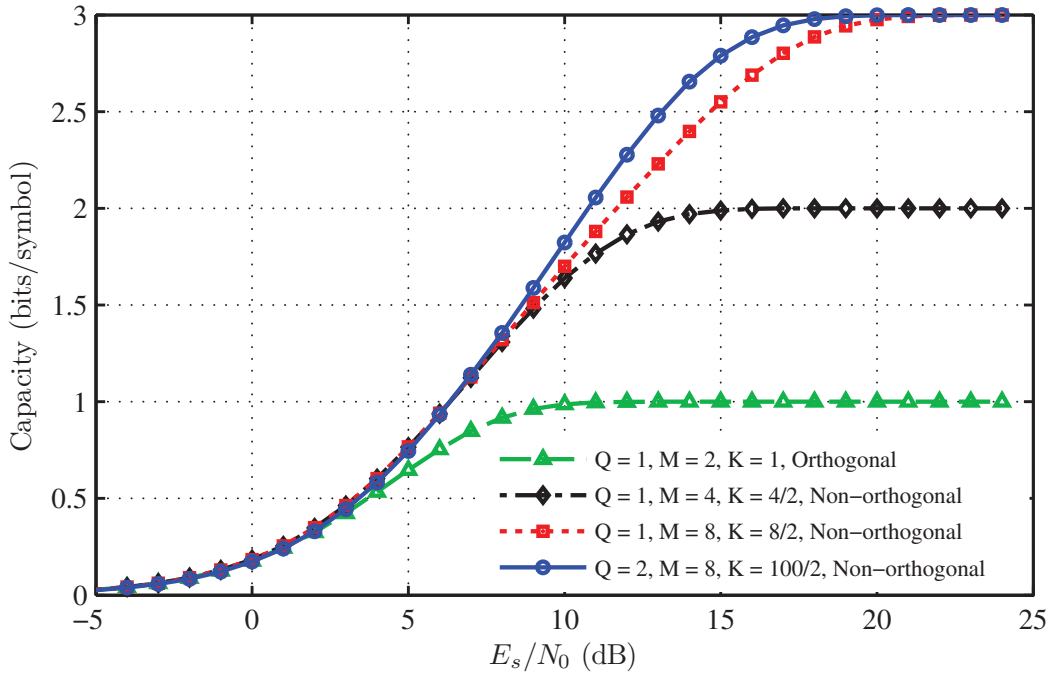


**Figure 5.6**: $N = 2$; Higher capacities for non-orthogonal MT-MFSK. Non-coherent channel with unknown phase.

In both cases, we evaluate 3 metrics of comparison. First we present the capacity in information bits per symbol as a function of $E_s/N_0$ in dB where $E_s$ is the energy per symbol and $N_0$ is the noise power spectral density. Second,

we show the minimum $E_b/N_0$ (in dB) required to obtain the capacity expressed in information bits per symbol where $E_b$ is the energy per information bit, and finally, we present the minimum $E_b/N_0$ required as a function of the capacity, this time expressed in bits per second per Hz (bits/s/Hz). These three equivalent views present slightly different interpretations that are useful in different contexts.

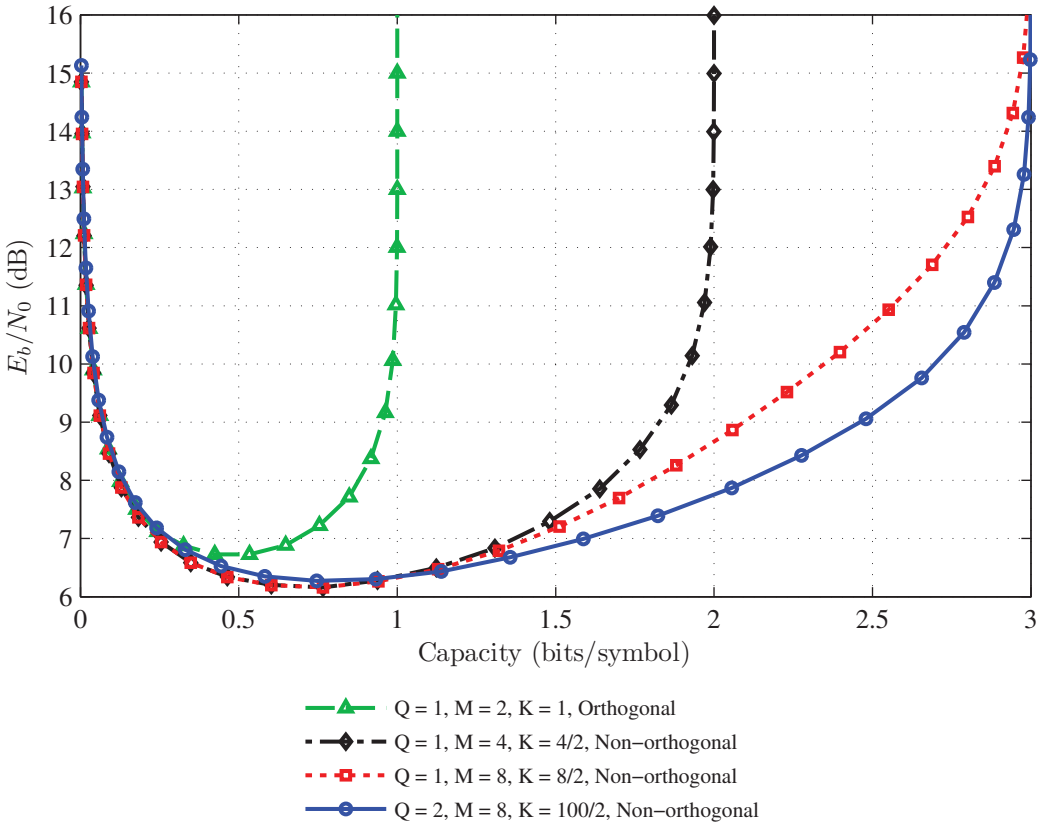


**Figure 5.7**: $N = 2$; Non-orthogonal MT-MFSK achieves 1 bits/symbol with $\approx$ 3.6dB less $E_b/N_0$. Non-coherent channel with unknown phase.

Figs. 5.6, 5.7 and 5.8 show the capacity of a non-coherent system with unit bandwidth (1Hz) and $R_s = 1/2$Hz. With that symbol rate and bandwidth, the only orthogonal tone system possible uses $N = 2$ tones and sends $Q = 1$ tone per symbol for a total of $M = 2$ symbols. The orthogonal system has a maximum capacity of 1 bits/symbol (Figs. 5.6 and 5.7) or 0.5 bits/s/Hz (Fig. 5.8) and it achieves this at $E_s/N_0 = E_b/N_0 \approx 10dB$. In contrast, non-orthogonal

systems achieve this same capacity at $E_s/N_0 = E_b/N_0 \approx 6.4dB$ – a significant improvement.

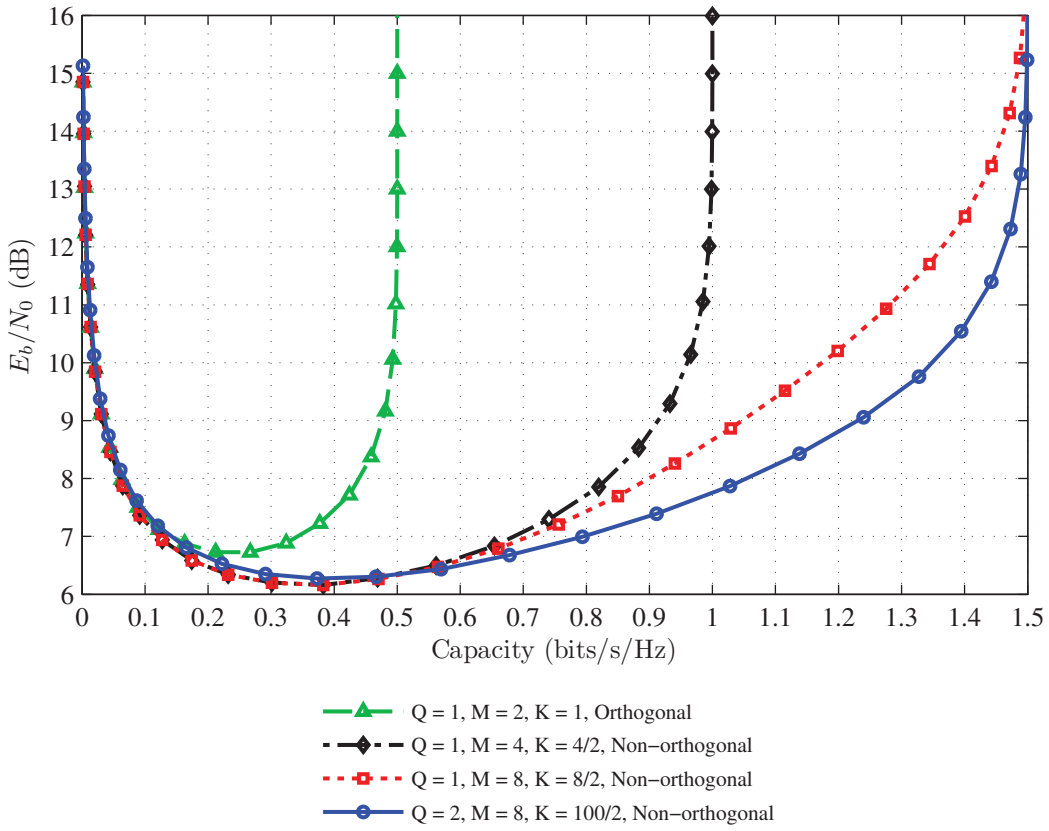


**Figure 5.8**: $N = 2$; Non-uniformly spaced MT-MFSK enables higher spectral efficiencies. Non-coherent channel with unknown phase.

Three different non-orthogonal systems are presented in Figs. 5.6, 5.7 and 5.8. The first two, are traditional non-orthogonal systems where the non-orthogonal tones are spread uniformly throughout the bandwidth. We show systems with 4 and 8 tones spread in the bandwidth occupied by the 2 orthogonal tones. For these two systems, the two values of $K$ (see Table. 5.1) are respectively given by $4/2 = 2$ and $8/2 = 4$ and they both transmit $Q = 1$ tone per symbol. The third non-orthogonal system also has $M = 8$ symbols, but sends $Q = 2$ tones per symbol and chooses these 8 symbols from a set of 100 non-orthogonal tones. Thus, $K = 100/2 = 50$ for this system. It can be seen from Figs. 5.6, 5.7 and

5.8 that the multi-tone non-orthogonal system out performs the orthogonal tone system.

Using the non-orthogonal MT-MFSK is advantageous on two fronts. First, the orthogonal system is incapable of achieving a capacity higher than 0.5 bits/s/Hz, whereas the non-orthogonal system can achieve much higher capacities and secondly, there is a significant performance improvement.

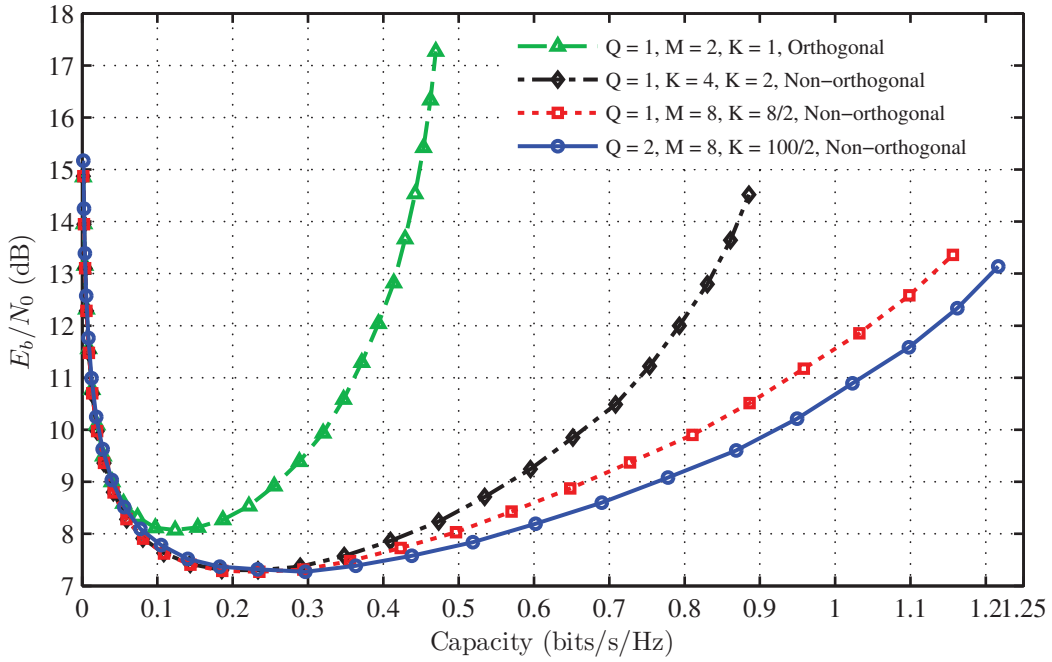


**Figure 5.9**: $N = 2$; Spectral efficiency in Rayleigh fading channel improves with non-orthogonal MT-MFSK

In Fig. 5.9 we show the performance of the same systems in frequency flat Rayleigh fading i.e., both the gain and the phase are assumed unknown (but constant for a symbol). As usual, the phase is assumed to be distributed uniformly from 0 to $2\pi$ radians and the gain is assumed to be distributed from a complex circular Gaussian random variable with mean and variance both equal to unity. It may be seen that at a spectral efficiency of 0.5 bits/s/Hz, the best non-orthogonal system outperforms the 2-orthogonal tone system by a remarkable 10dB! At 0.4 bits/s/Hz, the performance advantage is about 4.5dB.

Figs. 5.11, 5.10, 5.12 show the capacity of a unit bandwidth system with $R_s = 1/8$Hz. At this symbol rate, if only orthogonal tones were used, this bandwidth would have been occupied by $N = 8$ orthogonal tones.

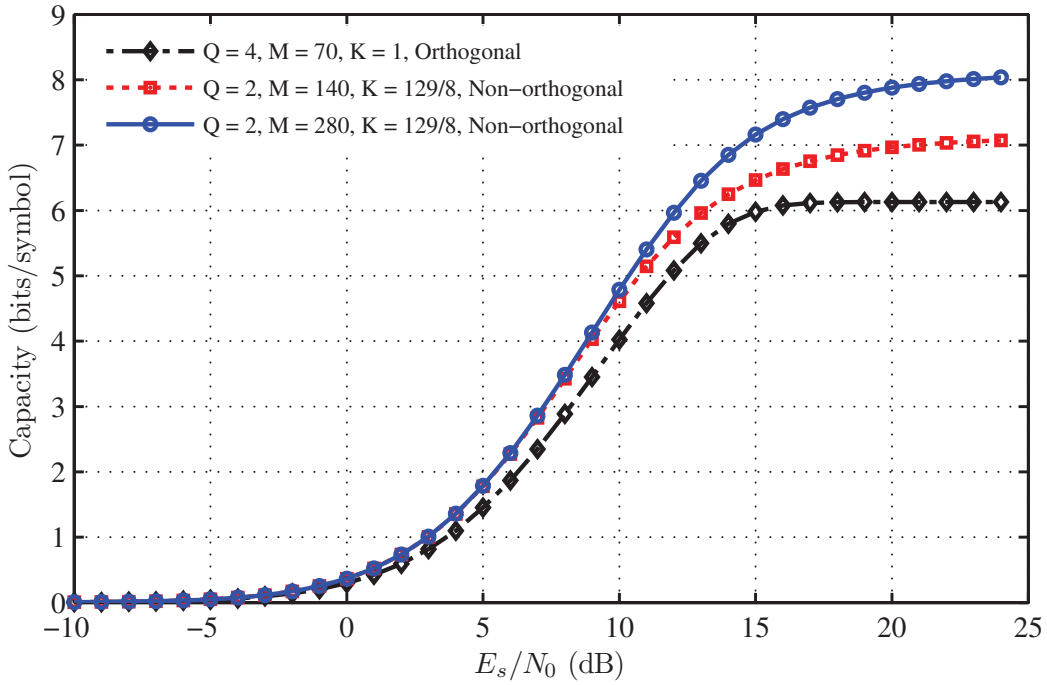


**Figure 5.10**: $N = 8$; Increased capacity with non-orthogonal MT-MFSK

As may be seen from Figs. 5.10 and 5.11, such a system achieves its maximum spectral efficiency when $Q = 4$ tones per symbol, resulting in $M = \binom{8}{4} = 70$ symbols. In that case, the spectral efficiency is 6.13 bits/symbol or 0.77 bits/s/Hz which may be achieved by the orthogonal system at $E_s/N_0 \approx 15$dB or $E_b/N_0 \approx 8$dB.

However, the best non-orthogonal system simulated can achieve the same spectral efficiency at an $E_s/N_0 \approx 12$dB a performance improvement of about 3dB. In this case, the non-orthogonal system used had a total of $M = 280$ symbols with each symbol consisting of $Q = 2$ tones chosen from a total of 129 tones spread in the same bandwidth. An additional advantage of this system over the orthogonal tone system is that in this case, the non-orthogonal MT-MFSK symbols use only 2 tones per symbol whereas the orthogonal system has 4 tones per symbol. This

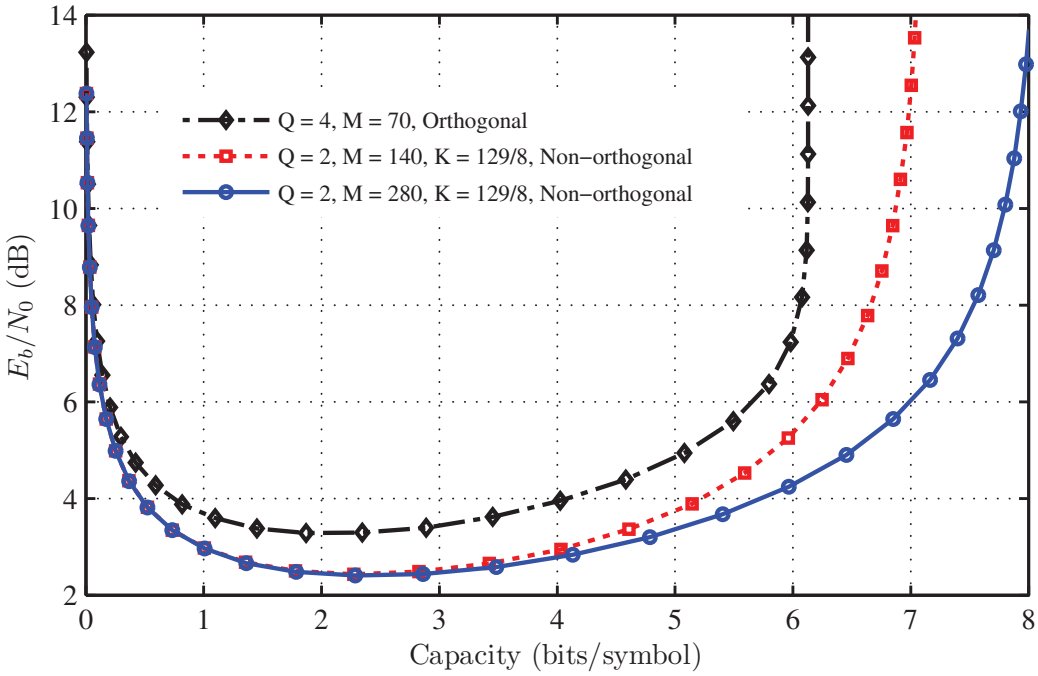


**Figure 5.11**: $N = 8$; Performance gain is $\approx$ 2dB at 6 bits/symbol

reduces the peak to average power ratio (PAPR) of the transmit waveform.

It is worth emphasizing that in both the systems simulated in this section, the orthogonal tone system has an upper limit of spectral efficiency. For the two tone example, it was 0.5 bits/s/Hz, and in the 8-tone example the maximum was 0.77 bits/s/Hz. However, there are no such limits when we use non-orthogonal tones. Even though, in the example with 2 orthogonal tones, Figs. 5.11-5.12 show non-orthogonal systems that go up to only 3 bits/symbol, there is nothing that prevents the design of a non-orthogonal systems with higher spectral efficiencies.

### 5.6.2 Non-orthogonal MT-MFSK with Optimal Receivers

In the following we present the results of simulations of various system configurations and compare it to the performance of traditional FSK. In all cases, the systems simulated have been assumed to have unit bandwidth and the symbol rate ($R_s$) is varied between the different experiments. For example, in Fig. 5.13, $R_s = 1/2$ Hz and consequently the bandwidth is occupied by 2 orthogonal tones.

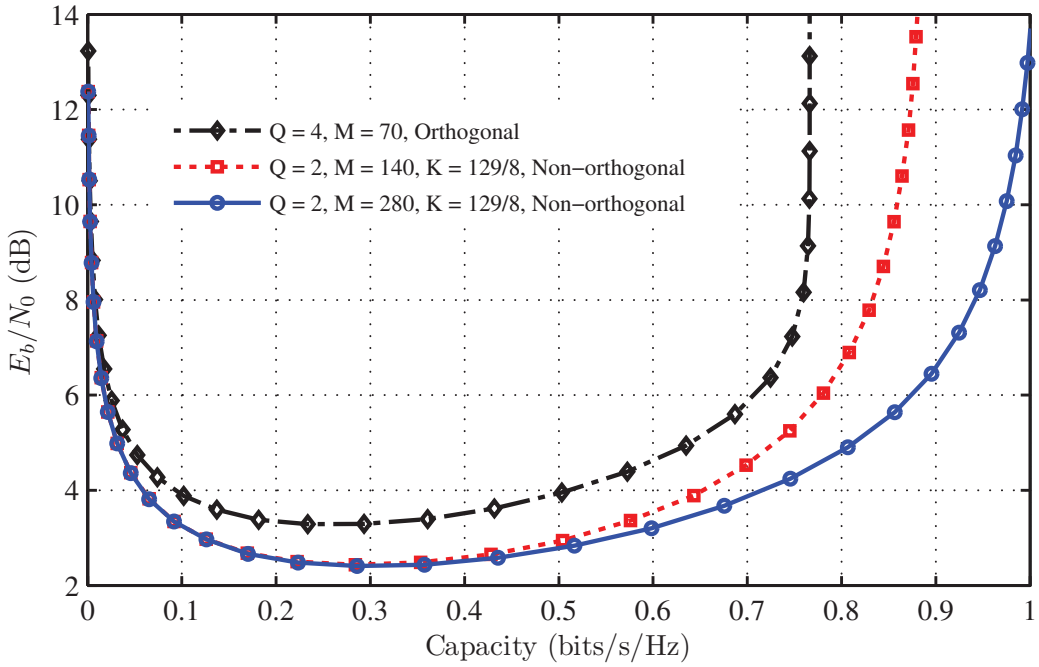


**Figure 5.12**: $N = 8$; Spectral efficiency in non-coherent channel with unknown phase

In Fig. 5.14, $R_s = 1/4$ Hz and so the bandwidth is occupied by 4 orthogonal tones. Each data point is the result of a Monte Carlo simulation with at least 300 errors. Fig. 5.13, 5.14, 5.15, 5.16, 5.17 all show the symbol error rate (SER) vs SNR curves of a non-coherent receiver where each symbol has a constant (and unknown) phase and, respectively, represent systems with $R_s = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{16}$ Hz. In each case, the simulations are conducted by generating a symbol randomly from the symbol set where each symbol is assumed to be equiprobable. Then the channel output is obtained by adding a zero mean random additive white Gaussian noise (AWGN) with variance commensurate to the SNR of interest and distorting the signal with a random phase distributed uniformly between 0 and $2\pi$ radians. For a fair comparison each symbol in all the simulations is normalized to unit power. The symbol to tone mapping of each constellation is given in Appendix 5.A.

In Fig. 5.13 the performance of traditional MFSK system with $N = 2$

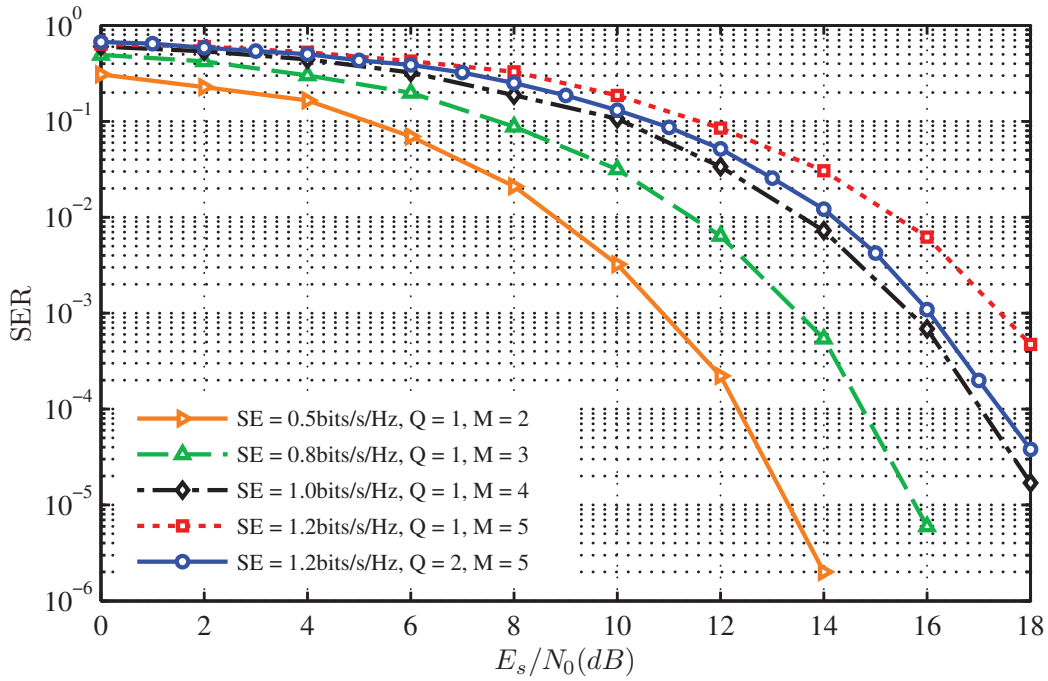


**Figure 5.13**: $N = 2$; Non-coherent channel with ML receiver. As SNR increases, spectral efficiency can be increased.

orthogonal tones and a spectral efficiency of $0.5bits/s/Hz$ is shown ($Q = 1$, $M = 2$). In the traditional system, it is not possible to improve the spectral efficiency above 0.5 and thus even if $E_s/N_0$ were to improve beyond that required by the orthogonal system, no further improvements in spectral efficiency could be made. Fig. 5.13 shows that the spectral efficiency could be increased to $1.2bits/s/Hz$ with a commensurate increase in the required $E_s/N_0$. Like the traditional system, three of the four non-orthogonal systems shown also sends a single tone per symbol, thus retaining the constant envelope modulation that leads to cheaper amplifier design with no linearity requirements. If, however, the cost of a linear amplifier can be tolerated by the system, a further performance gain of about 1dB (at a SER of $10^{-3}$) can be obtained by using a system that transmits $Q = 2$ tones instead of one.

Fig. 5.14 shows the performance for a system with $R_s = 1/4$. The traditional orthogonal MFSK could send either a single tone per symbol ($Q = 1$) (re-

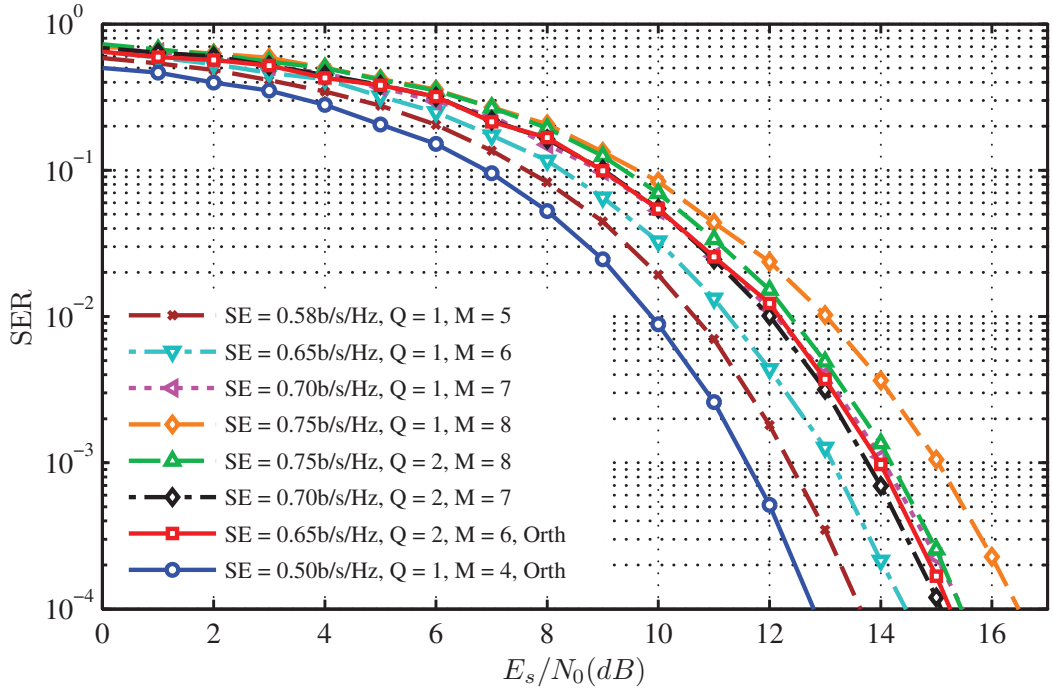


**Figure 5.14**: $N = 4$; Non-coherent channel with ML receiver. 2 solid lines indicate orthogonal MT-MFSK. Non-orthogonal MT-MFSK offers more choices.

sulting in $M = 4$ symbol constellation with spectral efficiency of $0.50 bits/s/Hz$) or $Q = 2$ orthogonal tones per symbol (resulting in a $M = \binom{4}{2} = 6$ symbol constellation with spectral efficiency of $0.65 bits/s/Hz$). It may be seen that the non-orthogonal system with a spectral efficiency of $0.7 bits/s/Hz$ has approximately the same performance as the traditional system with a spectral efficiency of $0.65 bits/s/Hz$. Moreover, the non-orthogonal system uses only a single tone per symbol and thus has a constant envelope modulation whereas the orthogonal MT-MFSK system uses 2 tones per symbol and thus has a higher PAPR. We show the performance non-orthogonal systems with a single tone per symbol as well as $Q = 2$ tones per symbol demonstrating that further gains can be made if the additional expense of a linear amplifier can be tolerated.

Fig. 5.15 shows the similar performance gains for a system with $R_s = \frac{1}{8}$. Orthogonal MT-MFSK can have systems with 1, 2, 3 or 4 tones per symbol, resulting in systems with 8, 28, 56, and 70 symbols respectively. It can be seen

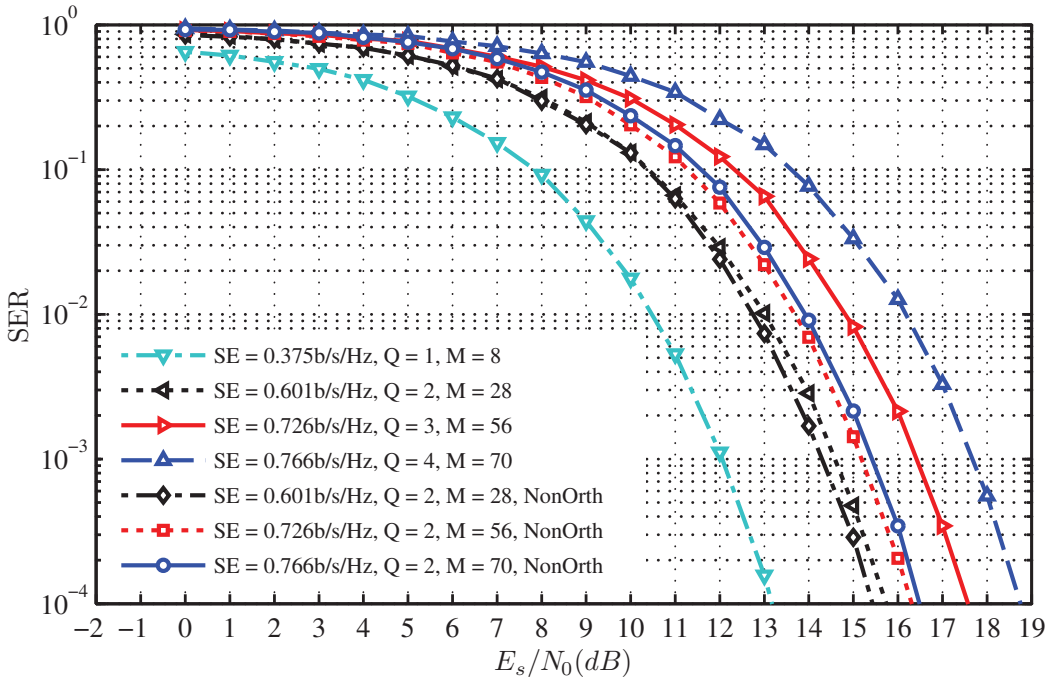


**Figure 5.15**: $N = 4$; Non-coherent channel with ML receiver. Performance gap between orthogonal and non-orthogonal MT-MFSK increases with $M$.

that each of the 28, 56 and 70 symbol orthogonal tone FSK systems is outperformed by a corresponding non-orthogonal tone system. Moreover, the proposed system sends at most 2 tones per symbol, whereas the orthogonal system would need to send as many as 4. Thus, the proposed system has a lower PAPR in addition to better performance. Fig. 5.16 and 5.17 show the corresponding result for $R_s = \frac{1}{16}$. For ease of display not all possible orthogonal systems are shown. Nearly all the orthogonal systems are outperformed by a corresponding non-orthogonal system. A notable exception is the orthogonal system that sends $Q = 4$ of the 16 orthogonal tones per symbol resulting in a system with $M = \binom{16}{4} = 1820$ symbols. In this case, an equivalent non-orthogonal system with better performance could not be found. However, this is likely to be of limited practical utility since most practical systems do not use over 1000 symbols.

Fig. 5.18 revisits the system with $R_s = \frac{1}{8}$, but now subjects the system with more harsh channels. We show the performance of systems where a) the phase is

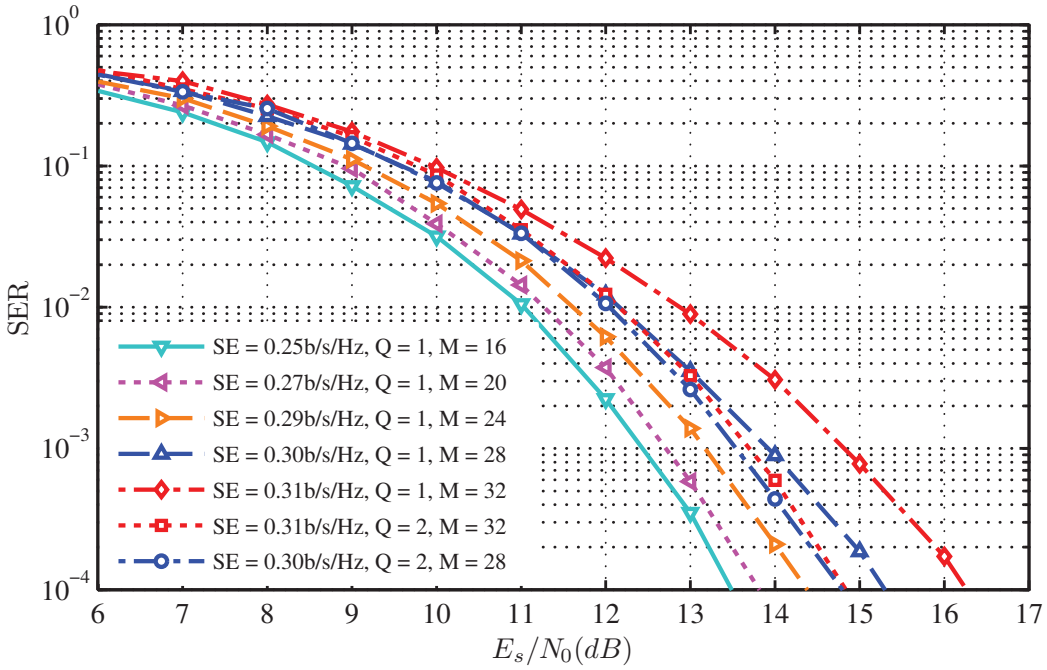


**Figure 5.16**: $N = 16$; Non-coherent channel with ML receiver. Shows increased granularity of spectral efficiency with MT-MFSK and the performance gains from using multitone symbols.

unknown but constant across the entire symbol b) where the phase is unknown and is independent from tone to tone and c) Rayleigh fading with each tone undergoing independent phase and gain distortion while going through the channel.

Fig. 5.19 shows the performance gains with the Cassini convolutional code described in Section 5.3.4. It can be seen that the non-orthogonal system has close to 2 dB of gain. The mapping of bits to symbols used is a simple binary representation. It is possible that with a better bit mapping, a lower BER could be obtained at the same $E_b/N_0$.

### 5.6.3 Reduced complexity receivers

Fig. 5.20 shows the performance of two reduced complexity receivers in two types of channels a) a flat fading channel with only an unknown phase offset across the symbol and b) a frequency selective channel with both the gain and

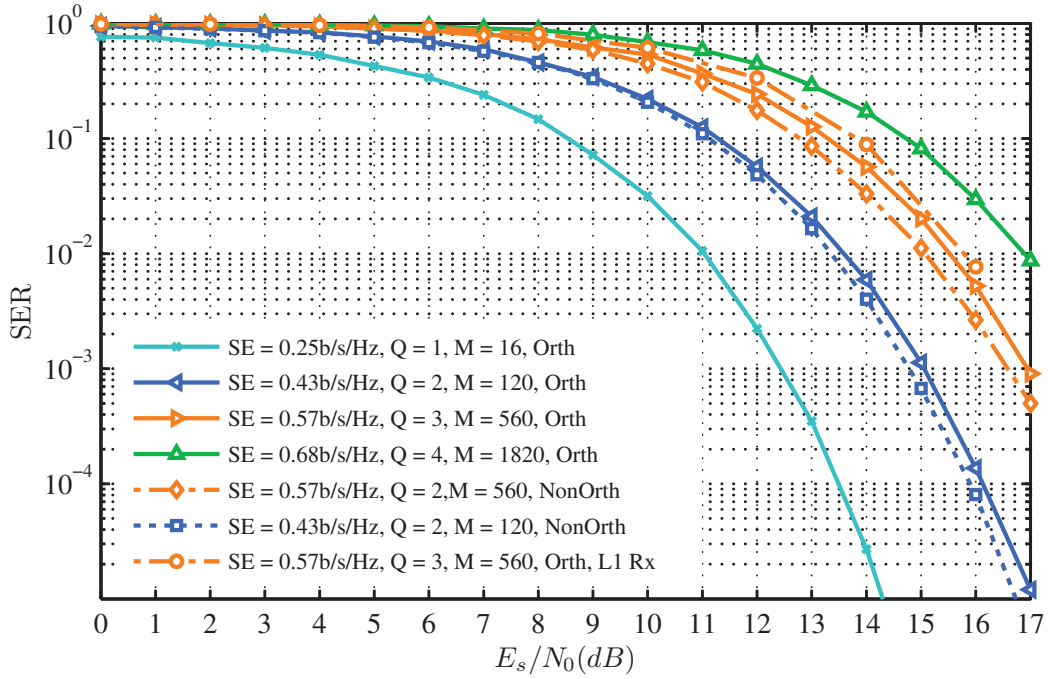


**Figure 5.17**: $N = 16$; Non-coherent channel with ML receiver and L1-norm minimization receiver. Performance of multitone symbols. Non-orthogonal systems outperform all except $M = 1820$ system.

phase being i.i.d. at each transmit tone. In both cases they are compared to the optimal ML or LS receivers which give the best performance. The system analyzed is a $R_s = 1/8$ system with the bandwidth occupied by $N = 8$ orthogonal tones packed with $NK = 100$ non-orthogonal tones, and each of the $M = 64$ symbols consisting of two transmit tones. The simplest receiver merely carries out a $NK = 100$ point FFT of the received signal, determines the two highest peaks and then does a LS search in the immediate neighborhood of the peaks.

The next receiver employs CS based techniques and employs the FOCUSS algorithm described in [76] and then does a LS search in the immediate neighborhood of the two peaks determined by the FOCUSS algorithm.

It can be seen that the performance of the simplest receiver is only slightly worse than that of the CS based receiver for the flat fading channel, however, it fails for the frequency selective channel. This is expected since the channel

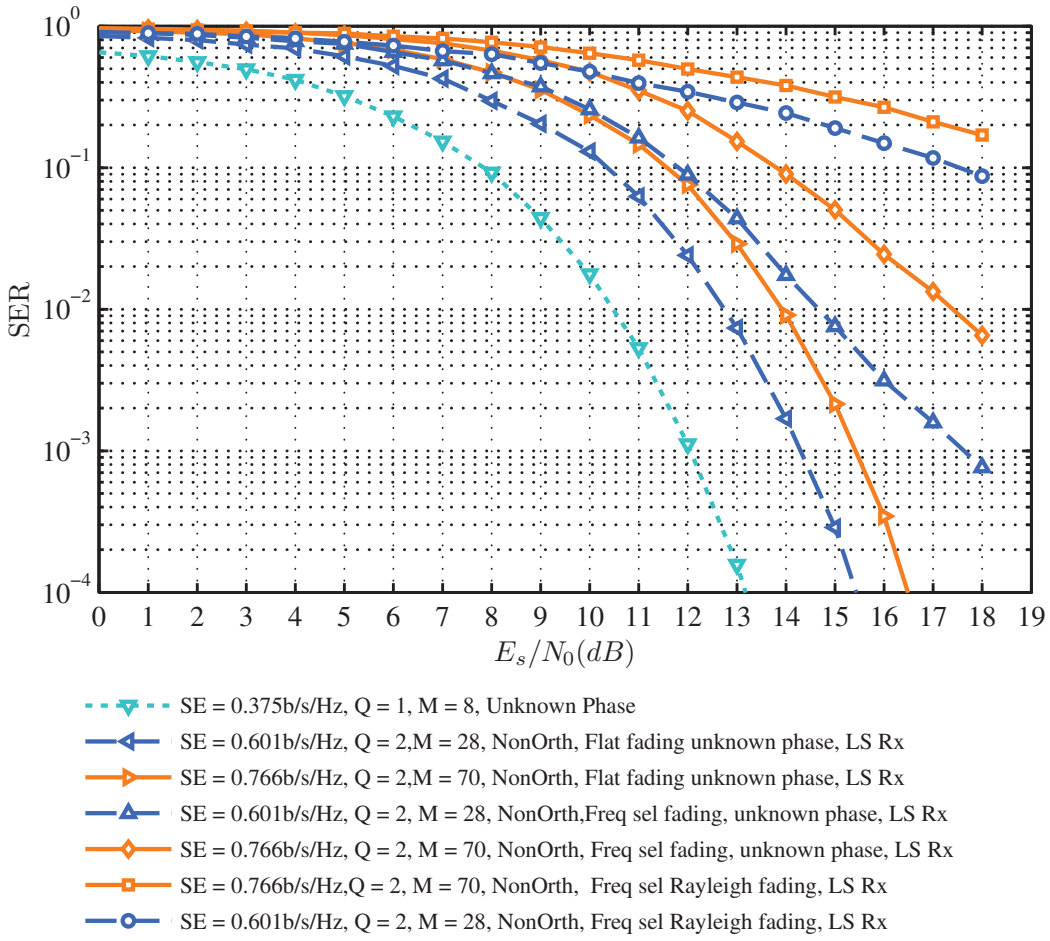


**Figure 5.18**: $N = 8$; Performance with 3 different channels.

gains of the two tones in the frequency selective fading can be quite different and thus the approach of determining the peaks after only a FFT at the receiver leads to erroneous results that cannot be corrected by looking in the immediate neighborhood of the two peaks. The CS technique however, performs significantly better and has a performance penalty of about 2dB at a BER of $10^{-2}$ for the frequency selective channel. For the flat fading channel, the CS receiver penalty is about 1dB.

For the 64 symbol system analyzed the computational complexity advantage is not significant and the performance penalty incurred eliminates the need for a CS based receiver, however, as the system scales up to potentially thousands of

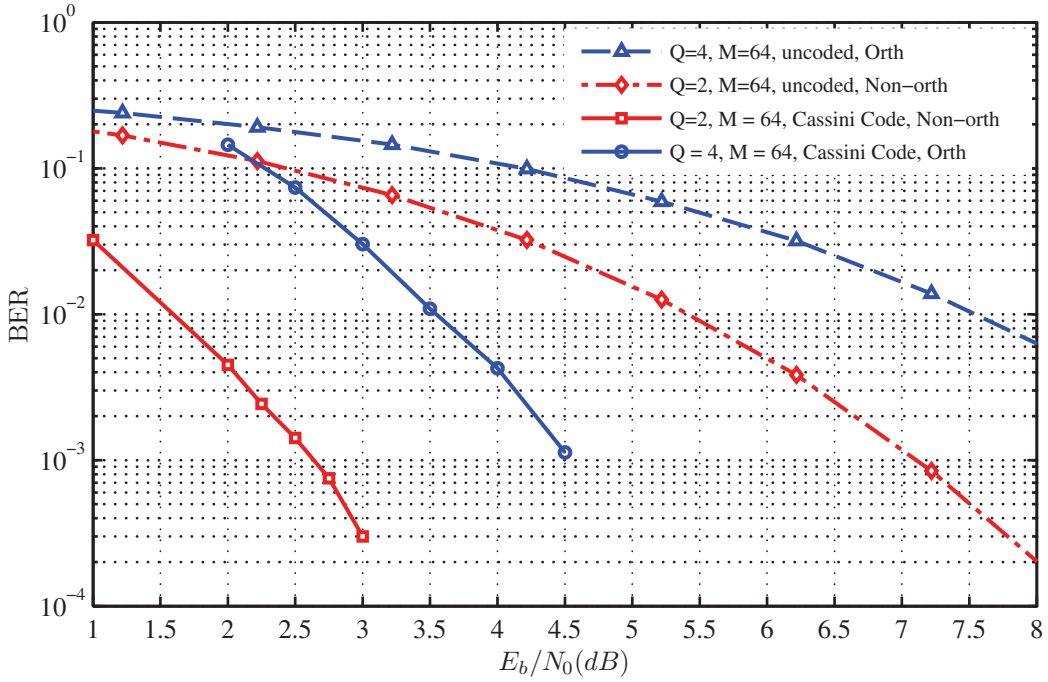


**Figure 5.19**: $N = 8$; Performance gain of $\approx$ 2dB with convolutional code and soft decision Viterbi decoding

symbols these receivers become much more important. The high SNR required for such large scale systems may make them unrealistic for practical systems, but we present this receiver for the sake of completion and to indicate the application of CS in a new arena.

### 5.6.4 SNR Estimation Results

Fig. 5.21 and 5.22 show the normalized CRLB and the normalized mean squared error (NMSE) plotted versus $E_s/N_0$ for various system configurations. Each simulation point was derived by keeping the channel gain fixed and generating random and i.i.d. samples of noise with the specified noise power. Each of the curves show the analytical normalized CRLB together with the NMSE for the ML estimator. The system simulated used a 100 symbol block fading model where the channel is assumed to be constant across 100 symbols and then varying independently from block to block. It can be seen that the ML estimator attains the

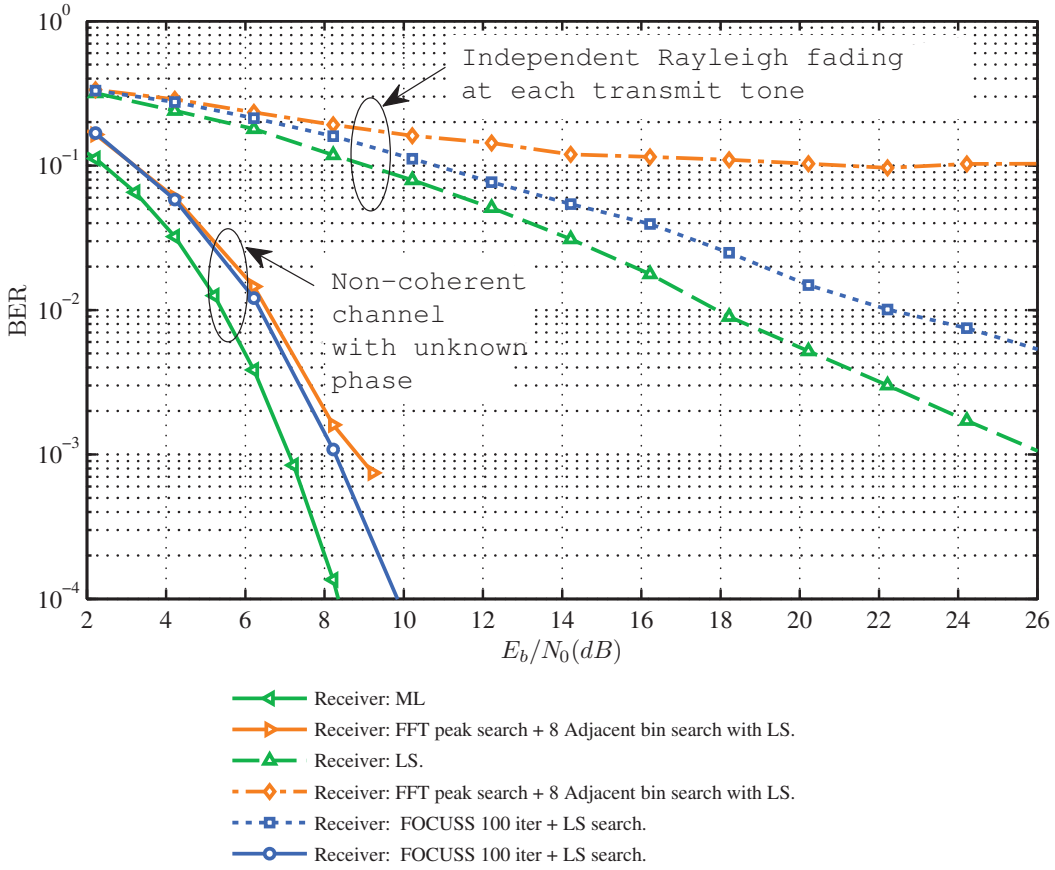


**Figure 5.20**: $N = 8$; Receiver complexity vs performance tradeoff

CRLB over the entire range of SNRs simulated. It should be pointed out that the CRLB is independent of the number of symbols in the system and only dependent on the number of tones in the orthogonal basis for the received signal space. Thus, e.g., a system with $R_s = 1/2$, which is spanned by $N = 2$ orthogonal carriers has the same CRLB for SNR estimation regardless of whether we are considering a system with 2 symbols or 5 symbols. This makes intuitive sense since the ability to estimate SNR is dependent on the number of independent noise paths through the system which is dependent on the number of elements in the orthonormal basis. In Fig. 5.21 we reproduce one of the curves from [52] where it may be seen that the NMSE of the estimator used is higher that what we can achieve. Since the estimator used in [52] uses the amplitude square of the receiver output while

we work with the complex output of the receiver, the estimator of [52] does not use all the information available and thus the performance is worse. We have also corrected a small error in the simulations done in [52] and we acknowledge the authors for their help in correcting this issue ( [52], Fig. 9 the x-axis should be labeled "average" SNR. Carrying out the simulation with a constant SNR results in slightly better results than what was published as shown in Fig. 5.21).

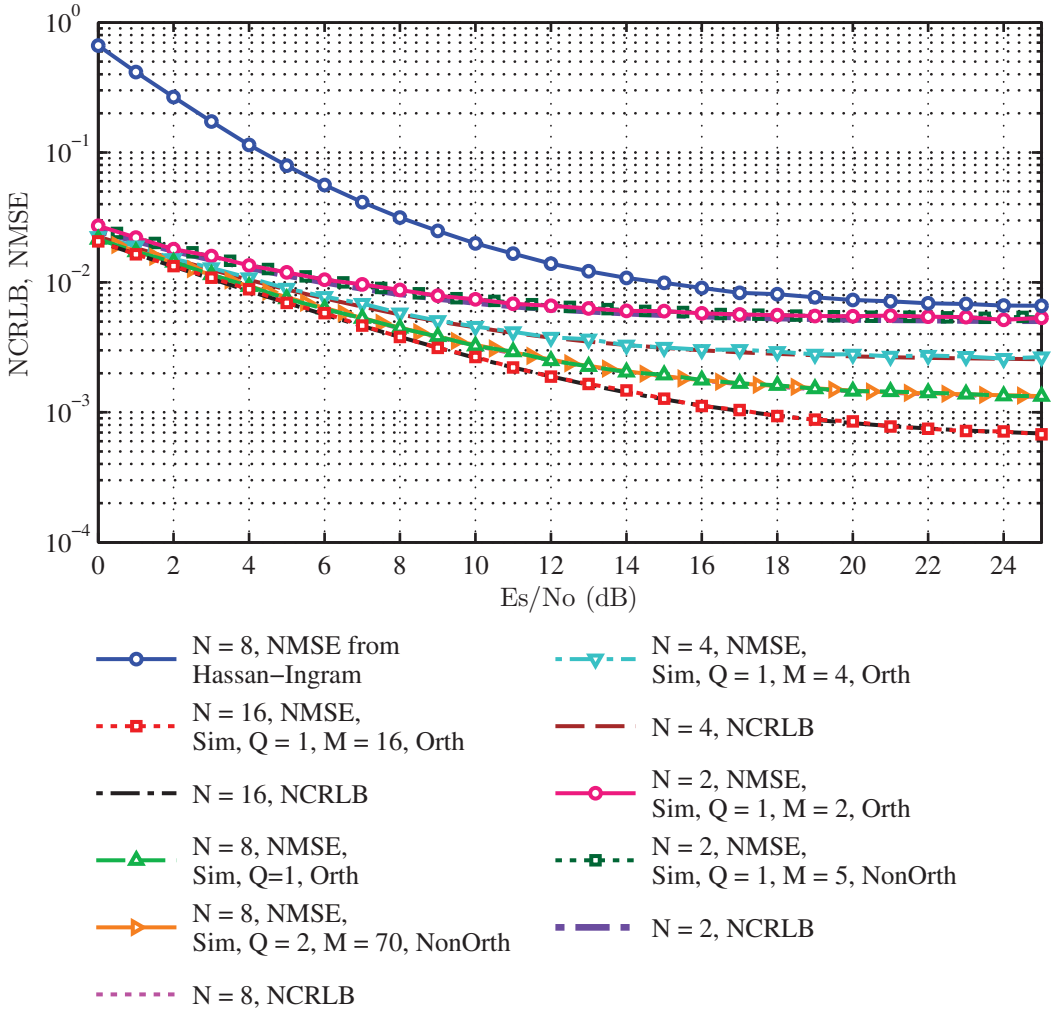


**Figure 5.21**: Data aided SNR estimation. Estimators achieve CRLB.

Fig. 5.22 compares the NCRLB vs $E_s/N_0$ performance for NDA and DA SNR estimation. It is interesting to note that at low SNRs the NCRLB for a system

with N = 4 (i.e., $R_s = \frac{1}{4}$ with 4 orthogonal carriers spanning the bandwidth) is higher than that for a system whose bandwidth is spanned by 2 orthogonal carriers (i.e., $R_s = \frac{1}{2}$) but at high SNRs their relative performance is swapped. Both the exact values as well as the closed form low SNR approximations are shown.

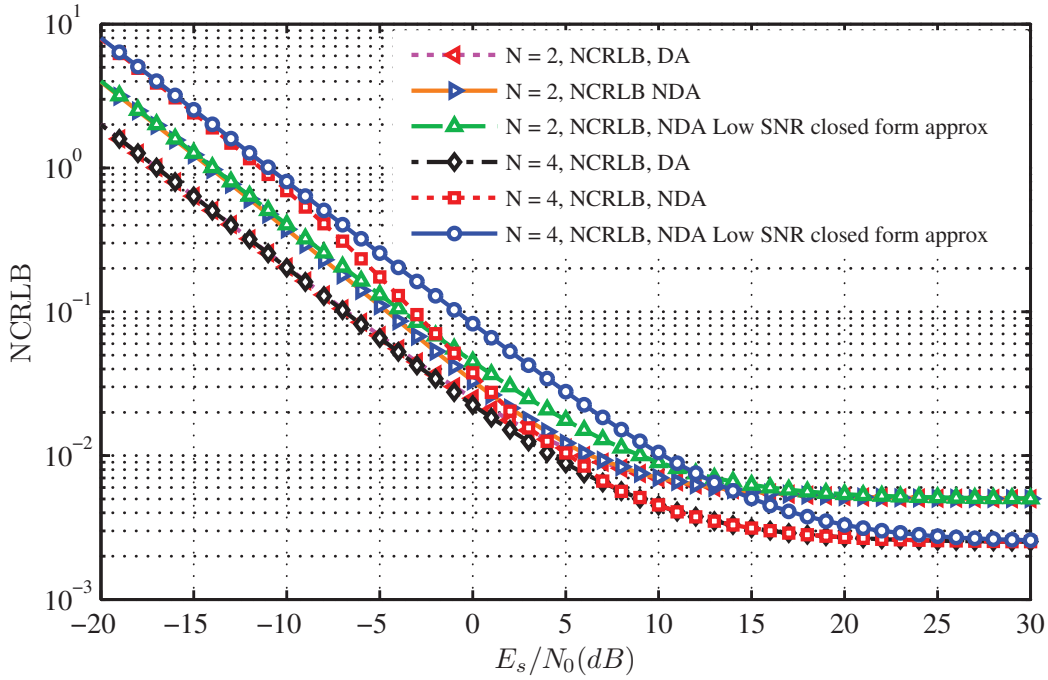


**Figure 5.22**: Non-data aided SNR estimation

## 5.7 Conclusion

In this chapter, we evaluate the advantages of a non-orthogonal multi-tone MFSK communication scheme. We show that the fundamental limits of performance of such schemes are better than that of orthogonal MT-MFSK by comparing the capacity of the channel. We demonstrate the gains that can be realized in practical systems by showing about a 2dB performance gain in specific configurations. We have presented the ML receiver for a flat fading channel with unknown phase. For frequency selective channels, we have derived the LS and CS based receivers. We also present the CRLB and ML estimators for the SNR estimation of DA non-

orthogonal FSK and show that the ML estimators achieve the CRLB. For non-data aided SNR estimation we present a method for calculating the exact CRLB and present closed form expressions for approximations of the NDA SNR estimate. We show that such a non-orthogonal MT-MFSK, coupled with accurate SNR estimation, could be used in a rate adaptive manner and different users could share the channel, but receive their data at a rate commensurate with their link conditions. This significantly improves the overall system spectral efficiency.

The contents of this chapter has been submitted for consideration of publication to the IEEE Transactions of Signal Processing.

## 5.A Appendix: Configurations for Non-orthogonal FSK

In this appendix we shall present the configurations for the various non-orthogonal systems presented. For each setup we shall denote the quantity $NK$, which represents the total number of tones in the grid to be considered and then for each of the $M$ symbols in the constellation we shall give the indices of the tones in the grid. In this convention, the first tone is numbered 1. Thus, as an example, if $NK = 16$, it implies that there are 16 equally spaced, but non-orthogonal tones in the available grid and if the system uses 2 symbols with tones $(1, 2, 3, 4)$ and $(13, 14, 15, 16)$ then it implies that the system sends 4 tones per symbol, that the 1st symbol sends tones numbered 1,2,3 and 4 from the grid, the second symbol sends tone numbers 13-16 and that some tones are left unused. The tones in the grid are uniformly placed across the unit bandwidth, but the tones actually transmitted per symbol and thus the tones actually used in the constellation may not be uniformly spaced.

**Table 5.2**: Configurations for $N = 2, N = 4$

| $R_s$ | $NK$ | M | Tones/Symb | Symbol to Tone Map |
|---|---|---|---|---|
| $\frac{1}{2}$ | 3 | 3 | 1 | (1), (2), (3) |
| $\frac{1}{2}$ | 4 | 4 | 1 | (1), (2), (3), (4) |
| $\frac{1}{2}$ | 5 | 5 | 1 | (1), (2), (3), (4), (5) |
| $\frac{1}{2}$ | 140 | 5 | 2 | (2,3), (1,71), (36,39), (71,74),(106,109) |
| $\frac{1}{4}$ | 5 | 5 | 1 | (1), (2), (3), (4), (5) |
| $\frac{1}{4}$ | 6 | 6 | 1 | (1), (2), (3), (4), (5), (6) |
| $\frac{1}{4}$ | 7 | 7 | 1 | (1), (2), (3), (4), (5), (6), (7) |
| $\frac{1}{4}$ | 140 | 7 | 2 | (1,38), (2,68), (10,105), (26,31), (45,112), (75,113), (81,89) |
| $\frac{1}{4}$ | 140 | 7 | 2 | (1,38), (2,68), (10,105), (26,31), (45,112), (75,113), (81,89) |
| $\frac{1}{4}$ | 120 | 8 | 2 | (1,94), (1,29), (2,62), (74,78), (16,22), (34,95), (45,50), (64,98) |

**Table 5.3**: Tone to symbol map for $N = 16$

| $R_s = \frac{1}{16}$, $NK = 140$, # of Symb = 28, Tones/Symb = 2 |
|---|
| (1,2) (4,135) (6,128) (7,121) (9,113) (10,106) (12,91) (13,84) (15,69) (16,62) (19,90) (22,68) (23,61) (27,30) (20,83) (24,54) (33,137) (35,131) (36,124) (38,109) (39,102) (41,87) (42,80) (43,73) (45,58) (46,51) (65,96) (76,116) |
| $R_s = \frac{1}{16}$, $NK = 140$, # of Symb = 32, Tones/Symb = 2 |
| (1,2) (4,135) (5,128) (7,120) (8,113) (10,97) (11,90) (12,83) (14,67) (15,60) (16,53) (17,46) (18,39) (19,32) (22,103) (25,88) (28,73) (36,78) (43,70) (56,124) (22,110) (24,95) (26,81) (36,71) (43,77) (49,100) (49,107) (57,117) (63,123) (64,116) (86,130) (92,136) |

**Table 5.4**: Tone to symbol map for $N = 8$

| $R_s = \frac{1}{8}$, $NK = 140$, # of Symb = 28, Tones/Symb = 2 |
|---|
| (1,32) (1,17) (3,124) (5,109) (8,95) (10,79) (11,62) (13,45) (16,30) (19,124) (21,108) (23,91) (26,78) (27,61) (35,125) (38,90) (39,73) (42,57) (47,136) (50,101) (50,118) (53,86) (64,134) (67,100) (67,117) (82,115) (97,131) (113,128) |
| $R_s = \frac{1}{8}$, $NK = 100$, # of Symb = 56, Tones/Symb = 2 |
| (1,24) (1,9) (1,16) (1,36) (1,49) (1,56) (2,87) (3,75) (5,64) (7,93) (9,54) (9,81) (10,41) (10,74) (11,29) (12,63) (12,89) (13,22) (13,49) (15,37) (17,81) (19,61) (19,70) (20,93) (21,45) (21,53) (23,78) (24,34) (24,88) (25,68) (26,61) (28,52) (30,75) (31,83) (33,43) (33,68) (33,91) (35,59) (38,83) (39,76) (40,97) (42,52) (42,66) (42,91) (46,83) (49,65) (49,73) (51,94) (54,89) (56,66) (56,80) (61,96) (64,86) (65,75) (70,94) (76,86) |
| $R_s = \frac{1}{8}$, $NK = 100$, # of Symb = 70, Tones/Symb = 2 |
| (1,24) (1,9) (1,15) (1,34) (1,41) (1,49) (1,60) (1,68) (1,87) (2,76) (6,21) (7,67) (7,92) (8,39) (9,52) (11,29) (8,59) (9,83) (10,75) (13,21) (13,47) (15,58) (15,69) (16,36) (16,88) (18,77) (20,46) (20,65) (21,54) (21,85) (23,41) (23,73) (24,33) (24,94) (25,62) (27,81) (28,48) (28,70) (29,56) (29,89) (31,78) (33,64) (34,44) (35,53) (35,86) (36,72) (36,94) (38,61) (39,80) (41,69) (43,88) (44,54) (44,77) (46,65) (48,85) (49,73) (49,94) (52,81) (53,63) (54,70) (55,92) (57,78) (61,72) (62,86) (63,94) (68,79) (70,91) (75,86) (78,96) (84,94) |

# Chapter 6

# Conclusions

In this dissertation, we have attempted to outline a few techniques that we believe will further the cause of more efficient communications. As the data exchanged daily through our communication networks continues to grow, every attempt should be made to convey information more efficiently and our work is a small step towards that goal.

Both in the arena of coherent communications as well as in the area of non-coherent communications the techniques proposed in this dissertation help to increase the spectral efficiency, or to have the same spectral efficiency at a lower transmit power. We have achieved these gains by challenging conventional design choices and by showing that conventional design choices are not necessarily optimal. Some of these choices have been made in the past due to simplicity of implementation, but with the incredible advances of computing power and technology, the time is right to question some of these choices.

Symbols aligned in time from different transmitters have been the hallmark of MIMO communication for quite sometime. However, in Chapter 2, we show that transmitting from the different transmitters offset in time leads to better performance. Similarly, orthogonal FSK has been the basis of non coherent communications for over 50 years, but again we show in Chapter 5 that by judiciously choosing to have non-orthogonal tones, not only is the performance improved but also the designer is presented with many more choices in system design. In both these areas, our unorthodox choices of symbol timing and frequency spacing leads

to significant performance improvements. We do not claim, however, that the last word has been said on these topics. There are quite a few areas open for additional research and we hope future researchers shall explore these boundaries.

In the arena of SNR estimation, we have derived the fundamental lower bounds that should form the basis of the evaluation of any practical estimator. Even if the iterative ML estimator presented is deemed to be too computational complex, other estimators should be compared against the fundamental lower bounds to see if there is a reasonable tradeoff of simplicity of implementation relative to performance degradation. The technique for detecting the onset of non-linearity described in Chapter 3 has been implemented and is currently operational in a network with about 1million customers.

## 6.1 Potential Future Research Areas

Some of the potential future areas of research are the following:

- Our work in the area of timing offset MIMO has been primarily limited to single carrier MIMO. Most modern MIMO systems have been proposed in the multi-carrier or OFDM framework and timing offset MIMO should be extended to such systems. In the area of timing offset MIMO techniques, although we have presented some techniques for a better pulse design that is explicitly designed for such systems, we have not designed an optimal pulse shape and this is one potential area of further research.

- In the area of non-orthogonal MFSK, research could go in quite a few different areas. First, a better search technique (or a constructive approach) to symbol set design could be researched. The problem could be stated as the following: Given a set of non-orthogonal tones, to determine the best symbol set with a specified number of symbols. Secondly, in the area of the design of compressed sensing based receivers, researchers could investigate the optimal receiver taking advantage of the structure in the codebook. Thirdly, practical systems have to deal with timing and frequency uncertainty and thus the

system gains of the proposed system in the presence of practical timing and frequency uncertainty could be investigated.

In conclusion, we hope our work will prove helpful to other researchers who share our passion of improving modern human communications. We believe that there is much yet to be discovered even in these specific areas and our research is merely "an arch wherethro' gleams the untravell'd world, whose margin fades for ever and for ever"[1] as we move the boundaries of human knowledge.

[1]Tennyson, "Ulysses", pub. 1842

# Appendix A

# Abbreviations

| Abbreviation | Full Form |
|---|---|
| 16-QAM | 16-ary Quadrature Amplitude Modulation |
| 32-APSK | 32-ary Amplitude and Phase Shift Keying |
| 3GPP | 3rd Generation Partnership Project |
| 8PSK | 8-ary Phase Shift Keying |
| AGC | Automatic Gain Correction |
| AWGN | Additive White Gaussian Noise |
| BER | Bit Error Rate |
| BPSK | Binary Phase Shift Keying |
| CDD | Cyclic Delay Diversity |
| CDF | Cumulative Distribution Function |
| CRLB | Cramer-Rao Lower Bound |
| CS | Compressed Sensing |
| DA | Data Aided |
| DOCSIS | Data over Cable Service Interface Specification |
| DVB-S2 | Digital Video Broadcasting - Satellite (2nd Generation) |
| EM | Expectation Maximization |
| FCC | Federal Communications Commission |
| FEC | Forward Error Correction |

... Continued

| Abbreviation | Full Form |
|---|---|
| FFT | Fast Fourier Transform |
| FIM | Fisher Information Matrix |
| FOCUSS | FOCal Underdetermined System Solver |
| FPGA | Field Programmable Gate Array |
| FSK | Frequency Shift Keying |
| GEO | Geosynchronous Satellite |
| GMM | Gaussian Mixture Density |
| IAI | Inter Antenna Interference |
| IBI | Inter Block Interference |
| IF | Intermedia Frequency |
| IID | Independent and Identically Distributed |
| ISI | Inter Symbol Interference |
| LLR | Log Likelihood Ratio |
| LMMSE | Linear Minimum Mean Squared Error |
| LS | Least Squared |
| LTE | Long Term Evolution |
| MAESTRO | Multiple Antenna Enhancement via Symbol Timing Relative Offset |
| MC-FSK | Multi Carrier Frequency Shift Keying |
| MCRLB | Modified Cramer Rao Lower Bound |
| MFSK | M-ary Frequency Shift Keying |
| MF-TDMA | Multi Frequency Time Division Multiple Access |
| MIMO | Multiple Input, Multiple Output |
| MISO | Multiple Input, Single Output |
| ML | Maximum Likelihood |
| MLSD | Maximum Likelihood Sequence Detector |
| MMSE | Minimum Mean Squared Error |
| MPSK | M-ary Phase Shift Keying |

... Continued

| Abbreviation | Full Form |
|---|---|
| MQAM | M-ary Quadrature Amplitude Modulation |
| MT-MFSK | Multi-Tone M-ary Frequency Shift Keying |
| NASA | National Aeronautics and Space Administration |
| NCRLB | Normalized Cramer Rao Lower Bound |
| NDA | Non Data Aided |
| NMSE | Normalized Mean Squared Error |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| OSIC | Ordered Successive Interference Cancellation |
| OSTBC | Orthogonal Space Time Block Codes |
| P1dB | 1dB compression point |
| PA | Power Amplifier |
| PAPR | Peak to Average Power Ratio |
| PDF | Probability Density Function |
| PER | Packet Error Rate |
| PHY | Physical Layer |
| QoS | Quality of Service |
| QPSK | Quadrature Phase Shift Keying |
| RF | Radio Frequency |
| RNG-REQ | Range Request |
| SER | Symbol Error Rate |
| SIC | Successive Interference Cancellation |
| SIMO | Single Input, Multiple Output |
| SINR | Signal to Interference plus Noise Ratio |
| SISO | Single Input, Single Output |
| SM | Satellite Modem |
| SNR | Signal to Noise Ratio |
| SRRC | Square Root Raised Cosine |

... Continued

| **Abbreviation** | **Full Form** |
|---|---|
| TDD | Time Division Duplexing |
| TDMA | Time Division Multiple Access |
| UWB | Ultra Wide Band |
| VBLAST | Vertical Bell Labs Layerd Space Time |
| ZF | Zero Forcing |

# Bibliography


[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Ninth Dover printing ed. New York: Dover, 1970.

[2] Y. Akaiwa and Y. Nagata, "Highly Efficient Digital Mobile Communications with a Linear Modulation Method," *Selected Areas in Communications, IEEE Journal on*, vol. 5, no. 5, pp. 890–895, 1987.

[3] N. S. Alagha, "Cramer-Rao bounds of SNR estimates for BPSK and QPSK modulated signals," *Communications Letters, IEEE*, vol. 5, no. 1, pp. 10–12, 2001.

[4] G. Atkin and H. Corrales, "A bandwidth efficient signaling for MFSK systems," in *Vehicular Technology Conference, 1988, IEEE 38th*, Jun 1988, pp. 596 –603.

[5] M. Bakkali, A. Stephenne, and S. Affes, "Iterative SNR Estimation for MPSK Modulation Over AWGN Channels," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, 2006, pp. 1–5.

[6] K. Barman and O. Dabeer, "Improving capacity in MIMO systems with asynchronous PAM," in *Information Theory and Its Applications, 2008. ISITA 2008. International Symposium on*, 2008, pp. 1–6.

[7] D. Bayley and J. Ralphs, "Piccolo 32-tone telegraph system in diplomatic communication," *Electrical Engineers, Proceedings of the Institution of*, vol. 119, no. 9, pp. 1229 –1236, september 1972.

[8] N. C. Beaulieu and M. O. Damen, "Parametric construction of Nyquist-I pulses," *Communications, IEEE Transactions on*, vol. 52, no. 12, pp. 2134–2142, 2004.

[9] N. C. Beaulieu, C. C. Tan, and M. O. Damen, "A "better than" Nyquist pulse," *Communications Letters, IEEE*, vol. 5, no. 9, pp. 367–368, 2001.

[10] N. C. Beaulieu, A. S. Toms, and D. R. Pauluzzi, "Comparison of four SNR estimators for QPSK modulations," *Communications Letters, IEEE*, vol. 4, no. 2, pp. 43–45, 2000.

[11] J. Benesty and T. Gansler, "Computation of the condition number of a nonsingular symmetric toeplitz matrix with the Levinson-Durbin algorithm," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 54, no. 6, pp. 2362–2364, 2006.

[12] A. Benjebbour and S. Yoshida, "Accurate noise variance estimation scheme for ordered successive detection in MIMO systems," *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, vol. 2, pp. 1023–1027 vol.2, May 2004.

[13] L. Berriche, K. Abed-Meraim, and J. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–397 – iv–400 vol.4.

[14] R. Bohnke, D. Wubben, V. Kuhn, and K. D. Kammeyer, "Reduced complexity MMSE detection for BLAST architectures," *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 4, pp. 2258–2262 vol.4, 2003.

[15] M. A. Boujelben, F. Bellili, S. Affes, and A. Stephenne, "EM Algorithm for Non-Data-Aided SNR Estimation of Linearly-Modulated Signals over SIMO Channels," *Global Telecommunications Conference, 2009. Globecom 2009. IEEE*, November 2009.

[16] ——, "SNR estimation over SIMO channels from linearly modulated signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6017–6028, Dec. 2010.

[17] S. Boumard, "Novel noise variance and SNR estimation algorithm for wireless MIMO OFDM systems," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 3, Jan. 2004, pp. 1330–1334.

[18] ——, "Novel noise variance and SNR estimation algorithm for wireless MIMO OFDM systems," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, vol. 3, January 2004, pp. 1330–1334 vol.3.

[19] K. Buer, *System and method for uplink power control by detecting amplifier compression point using dc current detection.* United States Patent, 2006, no. 7,006,791.

[20] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 927–946, 1998.

[21] R. Chen, J. Andrews, and R. Heath, "Multiuser space-time block coded MIMO with downlink precoding," in *Communications, 2004 IEEE International Conference on*, vol. 5, june 2004, pp. 2689 – 2693 Vol.5.

[22] S. Cioni, R. De Gaudenzi, and R. Rinaldo, "Channel estimation and physical layer adaptation techniques for satellite networks exploiting adaptive coding and modulation," *Int. J. Satell. Commun. Network.*, vol. 26, no. 2, pp. 157–188, 2008.

[23] J. Coon, M. Sandell, M. Beach, and J. McGeehan, "Channel and noise variance estimation and tracking algorithms for unique-word based single-carrier systems," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 6, pp. 1488–1496, June 2006.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 1991.

[25] A. Dammann, S. Plass, and S. Sand, "Cyclic delay diversity - A simple, flexible and effective multi-antenna technology for OFDM," in *Spread Spectrum Techniques and Applications, 2008. ISSSTA '08. IEEE 10th International Symposium on*, 2008, pp. 550–554.

[26] S. Dan and G. E. Lindong, "A Blind SNR Estimation Algorithm for MPSK Signals," in *Communications, Circuits and Systems Proceedings, 2006 International Conference on*, vol. 1, 2006, pp. 375–379.

[27] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *Communications, IEEE Transactions on*, vol. 42, no. 234, pp. 1391–1399, August 2002.

[28] A. Das, "NDA SNR estimation: CRLBs and EM based estimators," in *TENCON 2008 - 2008, TENCON 2008. IEEE Region 10 Conference*, 2008, pp. 1–6.

[29] A. Das and B. Rao, "SNR and Noise Variance Estimation for MIMO Systems," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 3929 –3941, Aug. 2012.

[30] A. Das, "Enhancing capacity of a satellite broadband system via adaptive coding and modulation," in *24th AIAA International Communications Satellite Systems Conference*. AIAA, Jun. 2006.

[31] A. Das and M. Miller, "Remote non-linearity detection via burst power dithering and EM based SNR Estimation," *Satellite and Space Communications, 2007. IWSSC 2007. International Workshop on*, Sept. 2007.

[32] A. Das and J. Ng, "NDA SNR Estimation for 32APSK in AWGN," in *9th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), July 2008*, July 2008.

[33] A. Das and B. D. Rao, "Impact of Receiver Structure and Timing Offset on MIMO Spatial Mulitplexing," in *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 8th Workshop on*, 2008.

[34] A. Das (Nandan) and B. Rao, "MIMO Systems with Intentional Timing Offset," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 267641, 2011. [Online]. Available: http://asp.eurasipjournals.com/content/2011/1/267641

[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B*, no. 1, pp. 1–38, 1977.

[36] F. Digham, M.-S. Alouini, and S. Arora, "Variable-rate variable-power non-coherent M-FSK scheme for power limited systems," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 6, pp. 1306 –1312, june 2006.

[37] N. Dinur and D. Wulich, "Peak-to-average power ratio in high-order OFDM," *Communications, IEEE Transactions on*, vol. 49, no. 6, pp. 1063 –1072, jun 2001.

[38] M. Dorrance and I. Marsland, "Adaptive Discrete-Rate MIMO Communications with Rate-Compatible LDPC Codes," *Communications, IEEE Transactions on*, vol. 58, no. 11, pp. 3115 –3125, 2010.

[39] C. Durand, E. Bejjani, and B. J, "Frequency space lattice encoding for non-coherent detection with correlated signals," in *6th Canadian Workshop on Information Theory, 1999*, June 1999.

[40] ——, "Non-coherent detection with nonorthogonal signals from an information theory perspective," in *International Symposium of Information Theory and Its Applications, 2000*, November 2000.

[41] C. Durand, J. Boutros, and E. Bejjani, "Forward error correction of FSK alphabets for noncoherent transmissions over AWGN channel," *Communications Letters, IEEE*, vol. 4, no. 10, pp. 318 –320, oct. 2000.

[42] C. Durand, "Modulations à haute efficacité spectrale pour les transmissions non cohérentes," Ph.D. dissertation, de l'Ecole Nationale Supérieure des Télécommunications, December 2000.

[43] ETSI, "Digital Video Broadcasting (DVB), ETSI EN 302 307 V1.1.2 (2006-06)," 2006.

[44] B. Farhang-Boroujeny, "A square-root Nyquist (M) filter design for digital communication systems," *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 2127–2132, 2008.

[45] J. Fei, R. Guangliang, and Z. Zhe, "A new noise variance and post detection SNR estimation method for MIMO OFDM systems," *Communication Technology, 2008. ICCT 2008. 11th IEEE International Conference on*, pp. 179–182, December 2008.

[46] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[47] P. Gao and C. Tepedelenlioglu, "Snr estimation for nonconstant modulus constellations," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 53, no. 3, pp. 865–870, 2005.

[48] W. Gappmair, "Cramér-Rao lower bound for Non-Data-aided SNR estimation of linear modulation schemes," *IEEE Transactions on Communications*, vol. 56, no. 5, pp. 689–693, May 2008.

[49] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramér-Rao bound in vector parameter estimation," *Communications, IEEE Transactions on*, vol. 46, no. 1, pp. 52–60, August 2002.

[50] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *ISSPA 2005, 8th International Symposium on Signal Processing and Its Applications, August 29-September 1, 2005, Sydney, Australia*, 08 2005.

[51] F. J. Harris, *Multirate Signal Processing for Communication Systems*. Prentice Hall PTR, May 2004.

[52] S. Hassan and M. Ingram, "SNR Estimation for M-ARY Non-Coherent Frequency Shift Keying Systems," *Communications, IEEE Transactions on*, vol. 59, no. 10, pp. 2786 –2795, october 2011.

[53] T. Hosman, M. Yeary, J. Antonio, and B. Hobbs, "Multi-tone FSK for ultrasonic communication," in *Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE*, may 2010, pp. 1424 –1429.

[54] J. Ibanez-Diaz, C. Pantaleon, I. Santamaria, T. Fernandez, and D. Martinez, "Nonlinearity estimation in power amplifiers based on subsampled temporal data," *Instrumentation and Measurement, IEEE Transactions on*, vol. 50, no. 4, pp. 882–887, 2001.

[55] C. T. L. Inc, "Data-Over-Cable Service Interface Specifications DOCSIS 2.0, Radio Frequency Interface Specification, CM-SP-RFIv2.0-I11-060602."

[56] B. Jiang, W. Wang, and X. Gao, "Polynomial-Based Noise Variance Estimation for MIMO-SCBT Systems," *Signal Processing Letters, IEEE*, vol. 16, no. 6, pp. 497–500, March 2009.

[57] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Future Network and Mobile Summit, 2010*, 2010, pp. 1 –10.

[58] A. Kavcic and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory," *Information Theory, IEEE Transactions on*, vol. 46, no. 1, pp. 291–301, 2000.

[59] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Prentice Hall, 1993.

[60] Y. Ko and C. Tepedelenlioglu, "Orthogonal space-time block coded rate-adaptive modulation with outdated feedback," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 2, pp. 290 – 295, 2006.

[61] B. Li, R. Difazio, and A. Zeira, "A low bias algorithm to estimate negative SNRs in an AWGN channel," *Communications Letters, IEEE*, vol. 6, no. 11, pp. 469–471, 2002.

[62] B. Li, R. A. Difazio, A. Zeira, and P. J. Pietraski, "New results on SNR estimation of MPSK modulated signals," in *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, vol. 3, 2003, pp. 2373–2377 vol.3.

[63] M. Y. Li, I. Galton, L. E. Larson, and P. M. Asbeck, "Correlation techniques for estimation of amplifier nonlinearity," *Radio and Wireless Conference, 2004 IEEE*, pp. 179–182, 2004.

[64] J.-K. Liang, R. de Figueiredo, and F. Lu, "Design of optimal Nyquist, partial response, Nth band, and nonuniform tap spacing FIR digital filters using linear programming techniques," *Circuits and Systems, IEEE Transactions on*, vol. 32, no. 4, pp. 386–392, 1985.

[65] R. Lopez-Valcarce and C. Mosquera, "Sixth-Order Statistics-Based Non-Data-Aided SNR Estimation," *Communications Letters, IEEE*, vol. 11, no. 4, pp. 351–353, 2007.

[66] R. Lopez-Valcarce, C. Mosquera, and W. Gappmair, "Iterative Envelope-Based SNR Estimation for Nonconstant Modulus Constellations," in *VIII IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2007*, 2007.

[67] C. Luo, M. Medard, and L. Zheng, "On approaching wideband capacity using multitone FSK," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 9, pp. 1830 – 1838, sept. 2005.

[68] C. Luo, M. Medard, L. Zheng, and D. Lun, "Multi-tone FSK with feedback," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, sept. 2005, pp. 112 –116.

[69] O. Milenkovic, "Sublinear Compressive Sensing (CS) and Support Weight Enumerators of Codes: A Matroid Theory Approach," http://temple.birs.ca/~09w5103/milenkovic_09w5103_talk.pdf, 2009, [Online; accessed 17-April-2013].

[70] S. S. Mneina and G. O. Martens, "Maximally flat delay Nyquist pulse design," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 51, no. 6, pp. 294–298, 2004.

[71] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[72] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *Communications, IEEE Transactions on*, vol. 48, no. 10, pp. 1681–1691, 2000.

[73] E. Peiker-Feil, M. Wetz, W. G. Teich, and J. Lindner, "OFDM-MFSK as a Special Case of Noncoherent Communication Based on Subspaces," in *OFDM 2012, 17th International OFDM Workshop 2012 (InOWo'12); Proceedings of*, 2012, pp. 1–5.

[74] J. Proakis, *Digital Communications*. McGraw-Hill Science Engineering Math, August 2000.

[75] A. Ramesh, A. Chockaligam, and L. B. Milstein, "SNR estimation in generalized fading channels and its application to turbo decoding," in *Communications, 2001. ICC 2001. IEEE International Conference on*, vol. 4, 2001, pp. 1094–1098 vol.4.

[76] B. Rao, K. Engan, S. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *Signal Processing, IEEE Transactions on*, vol. 51, no. 3, pp. 760–770, 2003.

[77] B. D. Rao and A. Das, "Multiple antenna enhancements via symbol timing relative offsets (MAESTRO)," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, 2007, pp. 1–5.

[78] B. Razavi, *RF Microelectronics*. Prentice Hall, 1998.

[79] H. Robin, D. Bayley, T. Murray, and J. Ralphs, "Multitone signalling system employing quenched resonators for use on noisy radio-teleprinter circuits," *Electrical Engineers, Proceedings of the Institution of*, vol. 110, no. 9, pp. 1554 –1568, September 1963.

[80] D. Ryan, I. Clarkson, and I. Collings, "Blind detection of PAM and QAM in fading channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1197 –1206, march 2006.

[81] S. Sandhu and A. Paulraj, "Space-time block codes: a capacity perspective," *Communications Letters, IEEE*, vol. 4, no. 12, pp. 384 –386, Dec. 2000.

[82] A. Scherb, V. Kuhn, and K.-D. Kammeyer, "Cramér-Rao lower bound for semiblind channel estimation with respect to coded and uncoded finite-alphabet signals," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, vol. 2, 2004, pp. 2193 – 2197 Vol.2.

[83] S. Shao, Y. Tang, T. Kong, K. Deng, and Y. Shen, "Performance analysis of a modified V-BLAST system with delay offsets using zero-forcing detection," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 6, pp. 3827–3837, 2007.

[84] S. Shao, Y. Tang, J. Liang, X. Li, and S. Li, "A modified V-BLAST system for performance improvement through introducing different delay offsets to each spatially multiplexed data streams," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, 2007, pp. 1062–1067.

[85] M. K. Simon, W. C. Lindsey, and S. M. Hinedi, *Digital Communication Techniques: Signal Design and Detection*, US ed. Prentice Hall PTR, Sep. 1994.

[86] F. X. Socheleau, D. Pastor, A. Aissa-El-Bey, and S. Houcke, "Blind noise variance estimation for OFDMA signals," *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 2581–2584, May 2009.

[87] J. Tan and G. L. Stuber, "Multicarrier delay diversity modulation for MIMO systems," *Wireless Communications, IEEE Transactions on*, vol. 3, no. 5, pp. 1756–1763, 2004.

[88] C. Tepedelenlioglu and R. Challagulla, "Low-complexity multipath diversity through fractional sampling in OFDM," *Signal Processing, IEEE Transactions on*, vol. 52, no. 11, pp. 3104–3116, 2004.

[89] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[90] A. van den Bos, *Parameter Estimation for Scientists and Engineers*. Wiley-Interscience, 2007.

[91] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[92] A. Vosoughi and A. Scaglione, "Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the CRB," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 940 – 954, 2006.

[93] D. Wang, L.-G. Jiang, and C. He, "Robust Noise Variance and Channel Estimation for SC-FDE UWB Systems under Narrowband Interference," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 6, pp. 3249–3259, June 2009.

[94] Q. Wang, Y. Chang, and D. Yang, "Deliberately designed asynchronous transmission scheme for MIMO systems," *Signal Processing Letters, IEEE*, vol. 14, no. 12, pp. 920–923, 2007.

[95] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparison of EM-Based Algorithms for MIMO Channel Estimation," *Communications, IEEE Transactions on*, vol. 55, no. 1, pp. 216–226, January 2007.

[96] M. Wetz, I. Perisa, W. Teich, and J. Lindner, "Robust Transmission Over Fast Fading Channels on the Basis of OFDM-MFSK," *Wireless Personal Communications*, vol. 47, no. 1, pp. 113–123, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11277-007-9395-8

[97] A. Wiesel, J. Goldberg, and H. Messer, "Non-data-aided signal-to-noise-ratio estimation," *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 1, pp. 197–201, 2002.

[98] P. Wittke, Y. M. Lam, and M. Schefter, "The performance of trellis-coded nonorthogonal noncoherent FSK in noise and jamming," *Communications, IEEE Transactions on*, vol. 43, no. 234, pp. 635 –645, feb/mar/apr 1995.

[99] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," *Communications, 1993. ICC 93. Geneva. Technical Program, Conference Record, IEEE International Conference on*, vol. 3, pp. 1630–1634 vol.3, 1993.

[100] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*, pp. 295–300, 1998.

[101] H. Xu, Z. Li, and H. Zheng, "A non-data-aided SNR estimation algorithm for QAM signals," in *Communications, Circuits and Systems, 2004. ICCCAS 2004. 2004 International Conference on*, vol. 1, 2004, pp. 103–107 Vol.1.

[102] H. Xu, D. Zhang, and J. Guo, "New Results of SNR Estimation for High-Order Modulation," in *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, 2007, pp. 1212–1215.

[103] W. Yi, L. Lihua, Z. Ping, L. Zemin, and Z. M. Yu, "A New Noise Variance Estimation Algorithm for Multiuser OFDM Systems," *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, pp. 1–4, December 2007.

[104] X. Yu, S. Leung, and X. Chen, "Performance Analysis of Space-time Block Coded MIMO Systems with Imperfect Channel Information over Rician Fading Channels," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2011.