Efficient Lattice Decoders for the Linear Gaussian Vector Channel:
Performance & Complexity Analysis

by

Walid Abediseid

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The theory of *lattices* — a mathematical approach for representing infinite discrete points in Euclidean space, has become a powerful tool to analyze many point-to-point digital and wireless communication systems, particularly, communication systems that can be well-described by the *linear Gaussian vector channel* model. This is mainly due to the three facts about channel codes constructed using lattices: they have simple structure, their ability to achieve the fundamental limits (the capacity) of the channel, and most importantly, they can be decoded using efficient decoders called *lattice decoders*.

Since its introduction to multiple-input multiple-output (MIMO) wireless communication systems, sphere decoders has become an attractive efficient implementation of lattice decoders, especially for small signal dimensions and/or moderate to large signal-to-noise ratios (SNRs). In the first part of this dissertation, we consider sphere decoding algorithms that describe lattice decoding. The exact complexity analysis of the basic sphere decoder for general space-time codes applied to MIMO wireless channel is known to be difficult. Characterizing and understanding the complexity distribution is important, especially when the sphere decoder is used under practically relevant runtime constraints. In this work, we shed the light on the (average) computational complexity of sphere decoding for the quasi-static, LAttice Space-Time (LAST) coded MIMO channel.

Sphere decoders are only efficient in the high SNR regime and low signal dimensions, and exhibits exponential (average) complexity for low-to-moderate SNR and large signal dimensions. On the other extreme, linear and non-linear receivers such as minimum mean-square error (MMSE), and MMSE decision-feedback equalization (DFE) are considered attractive alternatives to sphere decoders in MIMO channels. Unfortunately, the very low decoding complexity advantage that these decoders can provide comes at the expense of poor performance, especially for large signal dimensions. The problem of designing low complexity receivers for the MIMO channel that achieve *near-optimal* performance is considered a challenging problem and has driven much research in the past years. The problem can be solved through the use of *lattice sequential decoding* that is capable of bridging the *gap* between sphere decoders and low complexity linear decoders (e.g., MMSE-DFE decoder).

In the second part of this thesis, the asymptotic performance of the lattice sequential decoder for LAST coded MIMO channel is analyzed. We determine the rates achievable by lattice coding and sequential decoding applied to such a channel. The diversity-multiplexing tradeoff under such a decoder is derived as a function of its parameter—*the bias term*. In this work, we analyze both the computational complexity distribution and

the average complexity of such a decoder in the high SNR regime. We show that there exists a *cut-off* multiplexing gain for which the average computational complexity of the decoder remains bounded. Our analysis reveals that there exists a finite probability that the number of computations performed by the decoder may become excessive, even at high SNR, during high channel noise. This probability is usually referred to as the probability of a decoding failure. Such probability limits the performance of the lattice sequential decoder, especially for a one-way communication system. For a two-way communication system, such as in MIMO Automatic Repeat reQuest (ARQ) system, the *feedback channel* can be used to eliminate the decoding failure probability.

In this work, we modify the lattice sequential decoder for the MIMO ARQ channel, to predict in advance the occurrence of decoding failure to avoid wasting the time trying to decode the message. This would result in a huge saving in decoding complexity. In particular, we will study the throughput-performance-complexity tradeoffs in sequential decoding algorithms and the effect of preprocessing and termination strategies. We show, analytically and via simulation, that using the lattice sequential decoder that implements a simple yet efficient *time-out* algorithm for joint error detection and correction, the optimal tradeoff of the MIMO ARQ channel can be achieved with significant reduction in decoding complexity.

Acknowledgements

First and foremost, I would like to take this opportunity to thank my Thesis advisor Dr. Mohamed Oussama Damen for his support and guidance over the past several years. I would like to thank him for introducing to me the interesting topic of lattices which I really enjoyed working on some communication problems related to it during my research. I gratefully thank Dr. Tim Davidson, Dr. Patrick Mitran, Dr. Liang Xie, and Dr. David Jao, for their constructive comments on this thesis. I am thankful that in the midst of all their activity, they accepted to be members of the reading committee.

I owe an eternal debt to my lovely wife. I am indebted to her for the endless support to pursue and continuing my graduate studies. Nothing could have been achieved without her support, thank you is not enough.

There are a number of people who should be recognized as well. The first person to be acknowledged is my best friend Oussama Amin for his unceasing support, encouragement and the many pep talkes he has given me whenever I was frustrated. To my friends that I have met in Waterloo, Salama Ikki, Suhail Al-Dharrab, Mohamed Fathy Feteiha, Hussain Attia, Khalid Ali, thanks for the fun times and for the shoulders that only friends can give.

I would like to express my sincere thanks to my colleagues Mehdi Torbatian, Ahmad Al Shami, and Hussain Najafi for the interesting discussions and talk during my research work. I would also like to place on record my gratitude towards my bother Wajdi, and my lovely sisters Lubna, Deema, Ayat and Amani.

I convey special acknowledgement to Wendy Boles, and Susan King, for their indispensable help dealing with administration matters during my PhD work so that I optimally carry out my research.

Finally, the financial support provided by the Government of Canada (NSERC Alexander Graham Bell Canada Graduate Scholarship), and the Government of Ontario (OGSST) are gratefully acknowledged. I thank also the University of Waterloo for providing me with financial assistance throughout this work.

Dedication

إلى رمز الرجولة والتضحية

إلى من دفعني إلى العلم و به أزداد إفتخارا

أبي

الأستاذ الدكتور محمد توهيل أبو هنطش

To the hand which always gives. To the shoulder that has never let me down. To dad: Prof. Mohammad Tawhil Abediseid

إلى من يسعد قلبي بلقيها

إلى روضة الحب التي تنبت أزكى الأزهار

أمي

السيدة الفاضلة هناء أبو هنطش

To whom my heart dances when he sees. To the eternal source of warmth and patience. To Mom: Hana Iqab Abu Hantash

إلى من آنسني في دراستي وشاركني همومي

تذكاراً وتقديراً

زوجتي

فاطمة مقصود

To the one who shared with me laughs and tears. To whom my heart yearns and loves. To my wife: Fatima Maksoud

إلى قرة عيني

ابنتي نورة

To the one who made my life full of meanings. To my soul my daughter: Noura

Table of Contents

List of Tables

List of Figures

Chapter 1

The Road to Channel Capacity: An Introduction

THE main layering of a point-to-point communication system can be simply described by the following block diagram

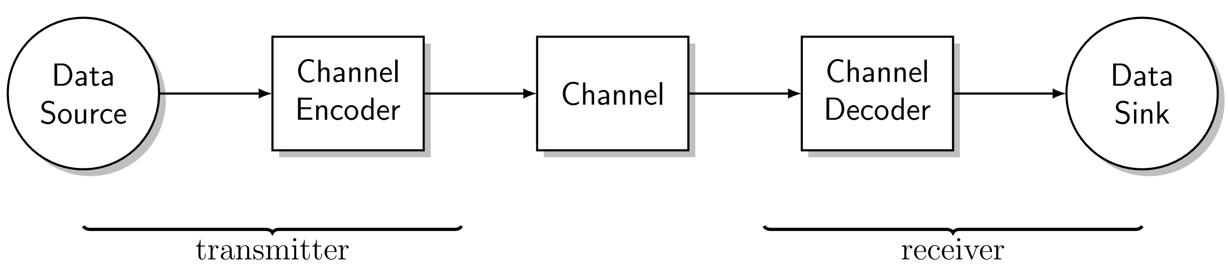


Figure 1.1: A simple communication system block diagram.

At the **transmitter** side, a message W, drawn from the index set $\{1, 2, \cdots, \mathcal{M}\}$, is first mapped into m-tuple sequence (word) $\boldsymbol{x}(W) = [x_1(W), x_2(W), \cdots, x_m(W)]$ using the **channel encoder**. The channel encoder introduces information redundancy in order to ensure reliable communication and serves as a protection to the data from channel disruption. The set of all m-tuples or codewords $\mathcal{C} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \cdots, \boldsymbol{x}(\mathcal{M})\}$, is called the *codebook*. The rate of the code, R, is given by

$$R = \frac{1}{m} \log |\mathcal{C}|,$$

where $|\mathcal{C}| = \mathcal{M}$ represents the cardinality of the code.

The communication **channel** is the physical medium that is used to send the data from the transmitter to the receiver. The channel introduces a source of noise that may cause some error in the transmitted signal. As such, the channel noise sets up a fundamental constant on how much data can be transmitted through the channel.

The **channel decoder** is designed to reconstruct the original information data from the received noisy signal with as low error as possible.

The metric that is commonly used to evaluate the performance of such a communication system is the **probability of decoding error**, P_e. Assuming that each codeword $\boldsymbol{x}(\mathcal{W}) \in \mathcal{C}$, for $1 \leq W \leq \mathcal{M}$, is equally likely to be transmitted, then the (average) probability of error can be expressed as

$$P_e = \Pr(\hat{\boldsymbol{x}} \neq \boldsymbol{x}), \tag{1.1}$$

where $\hat{\boldsymbol{x}}$ is the estimated word at the receiver.[1]

It is the dream of researchers to design low complexity encoding and decoding schemes that are capable of achieving the fundamental-limits of the communication channel. However, the improvement in the error performance usually comes at the expense of increasing the encoding/decoding complexity and/or reducing the transmission rate R. Therefore, a **performance-rate-complexity tradeoff** exists in any communication system. Our goal in this work is to introduce some interesting low encoding and decoding schemes that have excellent performance–rate-complexity tradeoff when applied to many communication systems of interest. Before doing that, we provide a brief introduction about the history of the design of coding and decoding schemes that achieve the fundamental limits of some important communication channels.

1.1 Shannon Capacity: 1948–2004

Many of today's communication systems would have not been made available without the existence of the field of *information theory*. This field has had a powerful impact on system

[1]As will be shown in the sequel, for some efficient decoders the output may not always correspond to a valid codeword in the code, i.e., $\hat{\boldsymbol{x}} \notin \mathcal{C}$. However, in some special cases, these decoders may achieve near-optimal performance.

design in both digital and wireless communications. Its basis was developed in 1948 by Claude-Shannon [1] to characterize the limits of reliable communication.

In earlier times, it was believed that the only way to achieve zero-error probability (i.e., $P_e \to 0$) is to let the transmission rate $R \to 0$. However, it was not until 1948, when Shannon showed in his landmark paper [1] that:

> "For every memoryless channel, there exists a parameter C, the channel capacity, such that, P_e, can be made arbitrary small for any transmission rate R less than C."

However, Shannon did not specify how to (practically) encode and decode the information data to achieve the channel capacity, which kept researchers occupied nearly 50 years. In particular, his proof was based on the use of **random codes** — codes that have *no structure*. Also, he assumed exhaustive **maximum-likelihood** (ML) decoding, whose complexity is proportional to the number of words in the code $\mathcal{M} = 2^{mR}$. Based on this random coding technique, and the ML decoding, Robert Gallager [2] showed that the average probability of error can be upper bounded as

$$P_e(R) \leq e^{-mE_r(R)}, \tag{1.2}$$

where $E_r(R)$ is called the random coding error exponent and is shown to be a non-zero, monotonically decreasing, positive function for all $R < C$ (see [2] for more details about this subject).

It is clear from the above bound that large codes would be required to approach capacity and therefore more practical decoding methods would be needed. Good (large) codes exist, and in fact a randomly chosen code will, with high probability, turns out to be good. However, as we mentioned before, such codes have no structure and may be difficult to implement. Therefore, the central objective now is to find a practical coding and decoding schemes that could approach channel capacity. Ever since Shannon's original paper, information theorists have attempted to construct structured codes and low complexity decoders that achieve the channel capacity.

There is a large body of work on capacity-achieving codes and their applications in communications. In order to introduce the major results on such a topic, we provide the diagram depicted in Figure 1.2 to summarize the corresponding works. We point the interested reader to an excellent survey by Forney and Costello [3].

In this work, our main interest focuses on the capacity-achieving codes that are based on *lattice theory* (see the black connection in Figure 1.2). These *structured* codes have recently found their way to many applications in both digital and wireless communication

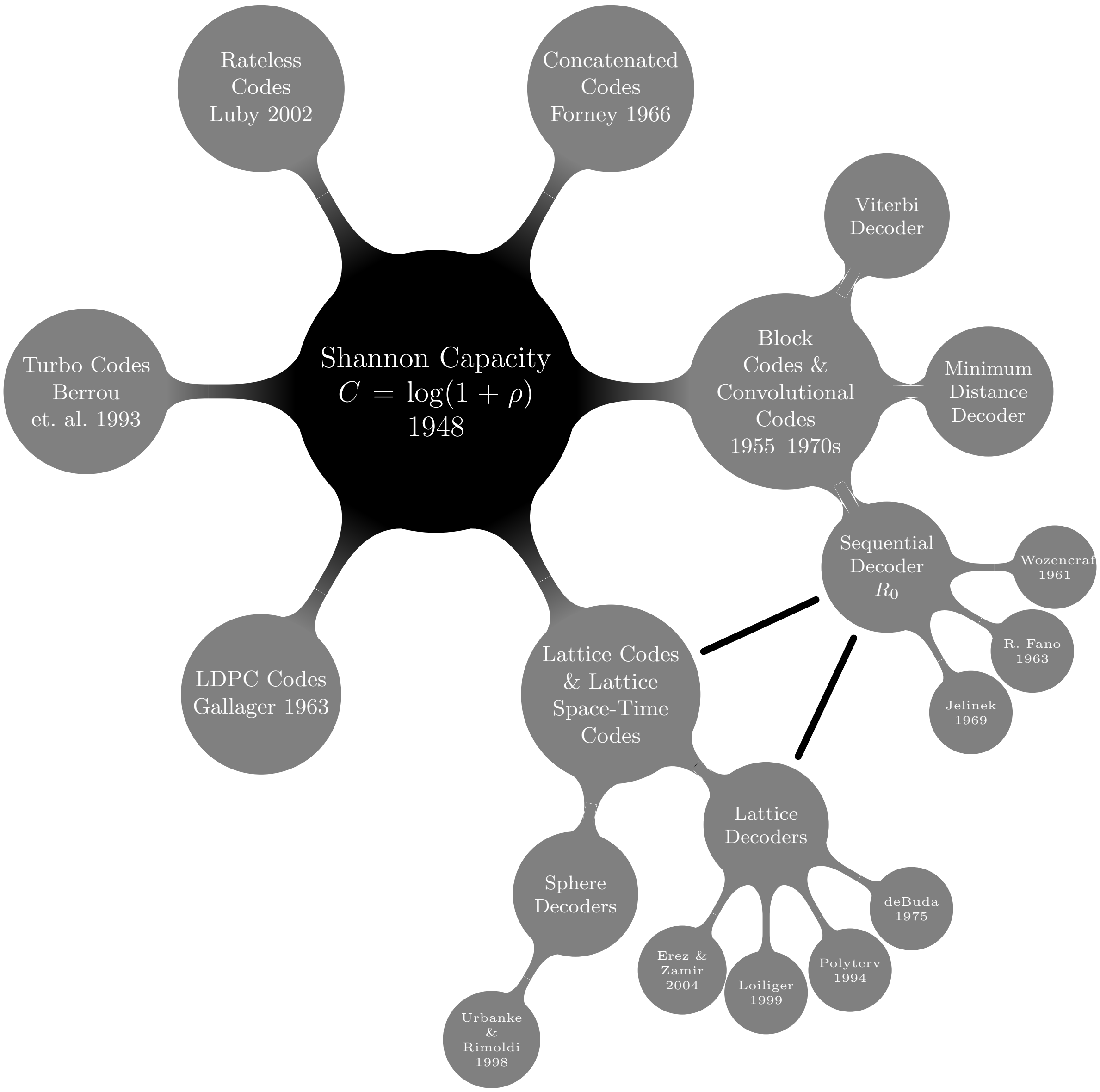


Figure 1.2: A summary of the major results about designing coding and decoding schemes that achieve the Shannon capacity.

systems due to their low complexity encoding/decoding and its excellent performance. The topic of lattice theory and its application to communication systems will be provided in details in Chapter 2.

We review the capacity of some important channels that provide the main motivation for the rest of the work.

1.2 The Linear Gaussian Vector Channel Model

Many today's point-to-point, digital and wireless communication systems can be well described by the *linear Gaussian vector channel model*. For a total of T channel uses, this channel model can be mathematically described by

$$\boldsymbol{y}_t^c = \boldsymbol{M}_t^c \boldsymbol{x}_t^c + \boldsymbol{e}_t^c, \quad t = 1, 2, \cdots T, \tag{1.3}$$

where $\boldsymbol{x}_t^c \in \mathbb{C}^{M \times 1}$ is the M-dimensional vector input to the channel at time t, $\boldsymbol{y}_t^c \in \mathbb{C}^{N \times 1}$ is the N-dimensional output vector of the channel, $\boldsymbol{e}_t^c \in \mathbb{C}^{N \times 1}$ is the additive complex Gaussian noise vector where each entry is zero-mean with independent real and imaginary parts with variance 1/2, and $\boldsymbol{M}_t^c \in \mathbb{C}^{N \times M}$ is a matrix representing the channel linear mapping.

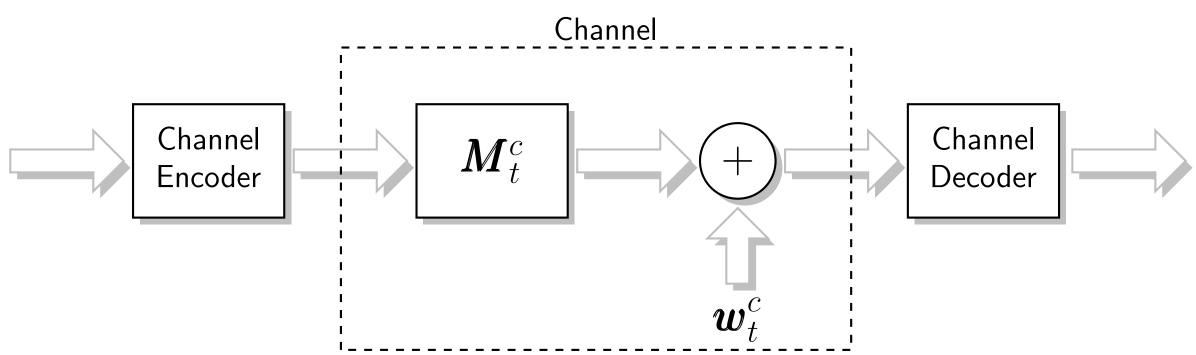


Figure 1.3: A general linear Gaussian vector channel model.

A well-known model that falls into such class of channels that is widely used in the literature is the so-called the power-constrained Additive White Gaussian Noise (AWGN) channel.

1.2.1 The Power-Constrained AWGN Channel

This model corresponds to the case where $M = N = 1$, and $\boldsymbol{M}_t^c = \sqrt{\rho}$, where ρ is defined as the signa-to-noise ratio (SNR) at the receiver. After T channel uses, the channel model may be expressed as

$$\boldsymbol{y}^c = \sqrt{\rho}\boldsymbol{x}^c + \boldsymbol{e}^c, \tag{1.4}$$

where $\boldsymbol{x}^c \in \mathbb{C}^{T\times 1}$ is a codeword that is selected from a code $\mathcal{C}$ satisfying the following input constrained

$$\frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{x}^c\in\mathcal{C}} \|\boldsymbol{x}^c\|^2 \leq T, \tag{1.5}$$

and $\boldsymbol{e}^c \in \mathbb{C}^{T\times 1}$ is the complex AWGN vector.

The *optimal* decoder that minimizes the word error probability (see (1.1)) is the ML decoder which can be described by

$$\hat{\boldsymbol{x}}^c = \arg\min_{\boldsymbol{x}^c\in\mathcal{C}} \|\boldsymbol{y}^c - \sqrt{\rho}\boldsymbol{x}^c\|^2.$$

Claude-Shannon [4] discovered that information data symbols can be transmitted through the AWGN channel with very low decoding error probability as long as the transmission rate is below capacity,

$$C = \log(1+\rho). \tag{1.6}$$

However, in his mathematical proof, Shannon assumed that the input codewords are drawn from the ensemble of Gaussian codes — a code that has no structure. He also, assumed that the received signal is detected using the ML receiver — a receiver that is considered practically to be infeasible.

Since then, there has been tremendous effort put towards the search for low complexity encoders and decoders for the AWGN channel that achieve near optimal (ML) performance. In his work, de Buda [5] was among the first to discover the existence of *structured* codes that achieve, for high SNR, rates very close to capacity. These codes are constructed based on the theory of *lattices* — a mathematical approach for representing infinite discrete points in Euclidean space. It is worth mentioning that most of traditional block and convolutional (trellis) codes can be constructed from lattices. The book by Conway and Sloane [6] is a good introduction about lattices and lattice codes construction.

One important feature of lattice codes is that they can be decoded by a class of efficient decoders known as *lattice decoders* — a decoder that decodes to the nearest lattice point, whether or not this point belongs to the code. This may significantly reduce decoding complexity, especially for small signal dimensions and/or moderate-to-high SNR without sacrificing performance. More details about lattice decoders will be discussed in Chapter 2.

Polytrev [7] studied the problem of coding for the "unconstrained" AWGN channel where the channel input being an infinite lattice. In his setting, the notion of capacity becomes meaningless as infinite rates of transmission are possible. Therefore, another significant measurement was defined that characterizes the performance limits of such coding scheme when decoded using lattice decoders which is the normalized density of the lattice or equivalently the information density rate of the lattice. It has also been shown that if only a *finite* number of lattice points are to be transmitted as codewords which satisfy a certain power constraint, then rates up to $\log \rho$ is achievable.

In contrast to Polytrev, where random coding has been used to show the above result, Loeliger [8] proved that rates up to $\log \rho$ can be achieved using ensembles of *linear* lattices. These codes are constructed using linear codes over the ring of p-prime integer numbers, i.e., $\mathbb{Z}_P$, which is usually referred to as Construction A [6]. An important aspect of both Polytrev's or Loeliger's proof is based on an important theorem in number theory that is referred to as *Minkowski-Hlawka* theorem [27], [28].

It is clear from the above discussion that, although we may achieve low encoding/decoding complexity, the penalty of using lattice decoding is a significant degradation in performance at low SNR and zero rate for $\rho < 1$. Therefore, since the capacity of the AWGN channel is $\log(1 + \rho)$, at this point one may ask whether the loss of "one" in the rate formula is due to the structured lattice codes or the sub-optimal (low complexity) lattice decoding. This question was addressed in [9] by Urbanke and Rimoldi, where they showed that (spherical) lattice codes can achieve the capacity of the AWGN channel under the high complexity (optimal) ML decoder.

It was not until 2004, when Erez and Zamir [24] proved that rates up to $\log(1 + \rho)$, can be achieved using lattice coding and decoding. The use of the minimum-mean square error (MMSE) estimator is essential to achieve the capacity. Their construction is based on the so-called *nested* lattices to generate nested codes or Voronoi codes. Such codes enjoy an important advantage over other capacity-achieving lattice codes (e.g., spherical lattice codes) due to their low encoding complexity (more details on this subject is provided in Chapter 2).

Next, we will discuss how the results of Erez and Zamir in [24] could be extended to a more general channel to achieve its capacity.

1.2.2 The Multiple-Input Multiple-Output Wireless Channel

Broadband wireless communication technologies are of great interest as mobile applications demand high data rates and quality service to support sophisticated real time services. This has motivated researchers to seek ways to increase the capacity of such systems.

Multiple-Input Multiple-Output (MIMO) technology, a mathematical model for communication systems with antenna arrays, promises significant increases in system performance and capacity. MIMO wireless communication systems have become an active research area for the past decade after the seminal work of Foschini [10] and Telatar [11]. Such systems has been shown to provide many advantageous over single-antenna-to-single-antenna systems and are mainly used to enhance the communication capabilities and overcome signal's transmission limitations that are caused by *fading* — a phenomena that is caused by multi-path propagation. The sensitivity to fading can be substantially reduced by the *spatial diversity*[2] provided by multiple spatial paths. Figure 1.4 shows a block diagram for a typical MIMO wireless communication system.

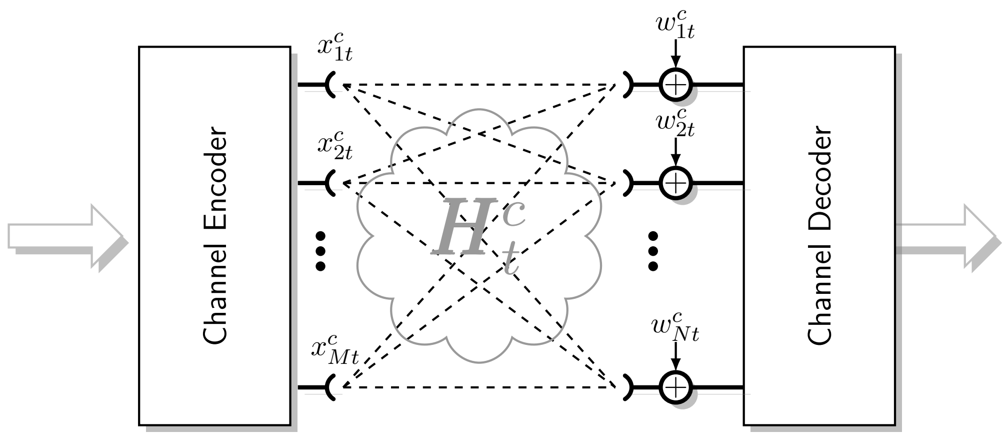


Figure 1.4: A Multi-antenna wireless communication system with M transmit and N receive antennas.

We consider a **frequency-flat fading**[3] MIMO channel with M-transmit, N-receive antennas. The complex base-band model of the received signal can be mathematically

[2]Diversity techniques collectively refer to methods of improving error performance by effectively transmitting the same information data multiple times, where each replica sees a different (ideally, independent) channel. There are many methods by which diversity can be achieved. Examples include: **time diversity**, frequency diversity, **space** (spatial) **diversity**, **channel coding** (as an efficient means of time diversity).

[3]In flat fading, the coherence bandwidth of the channel is larger than the bandwidth of the signal.

described by (1.3) with $\boldsymbol{M}_t^c = \sqrt{\rho}\boldsymbol{H}_t^c$, where $\boldsymbol{H}_t^c \in \mathbb{C}^{N\times M}$ is the channel matrix that is independent of $\boldsymbol{x}_t^c$ and $\boldsymbol{w}_t^c$, and $\rho = \mathrm{SNR}/M$ is the normalized average SNR with respect to M at each receive antenna. The elements of the channel fading gain matrix are assumed to be independent identically distributed (i.i.d) zero mean complex Gaussian random variables with variance per dimension 1/2. Equivalently, each entry of $\boldsymbol{H}_t^c$ has uniformly distributed phase and Rayleigh distributed magnitude with average $\mathsf{E}\{|[\boldsymbol{H}_t^c]_{ji}|\} = 1$ for all $1 \leq i \leq M$ and $1 \leq j \leq N$. This is intended to model a **Rayleigh fading channel** with enough physical separation within the transmitting and the receiving antennas to achieve independence in the entries of $\boldsymbol{H}_t^c$.

When the channel coefficients vary over time, the channel is considered random and hence the information rate[4] associated with it is also random. In that case, the "ergodic" capacity of a MIMO channel is the ensemble average of the information rate over the distribution of the channel matrix $\boldsymbol{H}^c$. The ergodic capacity, achieved by coding over an infinitely long interval, is given by [11]

$$C(\rho) = \mathsf{E}_{\boldsymbol{H}^c}\left\{\log\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right\}. \tag{1.7}$$

It is not difficult to show that capacity-achieving random Gaussian codes constructed for the AWGN channel can also be used to achieve the capacity of the ergodic MIMO channel [11].

The capacity formula in (1.7) is very difficult to evaluate in general. However, for high-SNR values, the ergodic capacity can be shown to be given by [10]

$$C(\rho) = \min\{M, N\}\log\rho + o(1).$$

In other words, at high SNR the channel capacity of the $M \times N$ MIMO channel increases with SNR as $\min\{M, N\}\log\rho$ as opposed to $\log\rho$ for the single-input single-output channel. The term $\min\{M, N\}$ is usually referred to as the number of spatial *degrees of freedom* in the channel in which *independent* information symbols may be transmitted through them. This is the so called spatial multiplexing [10].

On the other hand, if the channel matrix is chosen randomly but held fixed for the whole duration of channel uses, i.e., $\boldsymbol{H}_t^c = \boldsymbol{H}^c$, then the capacity of the channel is not

Therefore, all frequency components of the signal will experience the same magnitude of fading. In frequency-selective fading, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Different frequency components of the signal therefore experience decorrelated fading.

[4]The information rate is defined by the mutual information, $I(\boldsymbol{x}^c, (\boldsymbol{y}^c, \boldsymbol{H}^c))$, between the input to the channel $\boldsymbol{x}^c$, and the output of the channel $(\boldsymbol{y}^c, \boldsymbol{H}^c)$, assuming the channel matrix can be perfectly estimated at the receiver side.

given by (1.7) anymore. This type of a channel is called the **quasi-static** (non-ergodic) channel and in fact, its capacity in the strict sense is zero. This is due to the fact that there is a non-zero probability that the channel is in a deep fade, making it impossible to transmit information at positive rates while at the same time achieving arbitrary small decoding error probability. Therefore, a different criterion has been defined to characterize the performance limits of such a channel. This is the so-called *outage capacity* [12] — the largest rate of reliable communication for a fixed error probability. The tradeoff between data rate and error probability can be captured by the outage capacity which can be evaluated using the *outage probability*, $P_{\text{out}}(\rho, R)$. We say that the channel is in outage if the transmission rate cannot be supported by the channel. We define the outage event by

$$\mathcal{O}(\rho) \triangleq \{\boldsymbol{H}^c : R_{\text{achiv}}(\rho, \boldsymbol{H}^c) < R\}, \tag{1.8}$$

where R is the transmission rate and $R_{\text{achiv}}(\rho, \boldsymbol{H}^c)$ is the rate achievable by the coding and decoding schemes used in the system for a given channel realization. In this case, we have

$$P_{\text{out}}(\rho, R) = \Pr(\mathcal{O}(\rho)).$$

In this work, we will only consider the quasi-static, Rayleigh fading $M \times N$ MIMO channel. In this scenario, after T channel uses, the received signals $\{\boldsymbol{y}_t^c : 1 \leq t \leq T\}$ may be jointly combined into a matrix $\boldsymbol{Y}^c \in \mathbb{C}^{N \times T}$ as

$$\boldsymbol{Y}^c = \sqrt{\rho}\boldsymbol{H}^c\boldsymbol{X}^c + \boldsymbol{E}^c, \tag{1.9}$$

where $\boldsymbol{X}^c = [\boldsymbol{x}_1^c, \boldsymbol{x}_2^c, \cdots, \boldsymbol{x}_T^c] \in \mathbb{C}^{M \times T}$ is the transmitted code matrix, and $\boldsymbol{E}^c = [\boldsymbol{e}_1^c, \boldsymbol{e}_2^c, \cdots, \boldsymbol{e}_T^c] \in \mathbb{C}^{N \times T}$ is the noise matrix. The following average transmit power constraint is enforced:

$$\mathsf{E}\{\|\boldsymbol{X}^c\|_F\} \leq MT.$$

In this model, we assume perfect Channel State Information (CSI) at the receiver side, and no CSI at the transmitter. This is representative of systems based on "coherent" detection under the simplifying assumption that the channel matrix can be estimated very accurately at the receiver side.

As has been discussed earlier, exploiting multiple antennas at the transmitter and the receiver side provide substantial benefits in both increasing system capacity and improving the immunity to deep fading in the channel. To take advantage of these benefits, special *Space-Time Coding* (STC) techniques are employed.

Space-Time Coding

Space-time coding, a powerful coding technique that was developed by Tarokh et. al. in [13], is used to achieve spatial diversity. At the time of its invention, intensive research [13]–[15] had been conducted for the design of space-time codes that achieve high order diversity. After that, much attention was paid to the design of space-time codes that achieve high data rates [16],[17]. One approach, which attempts to achieve high data rates, is the Vertical Bell Labs Layered Space-Time (V-BLAST) scheme [10]. In such a scheme, the data stream is divided into independent substreams to be transmitted on the individual antennas. The V-BLAST receiver decodes the substreams using a sequence of nulling and canceling steps, usually referred to as zero-forcing decision feed-back equalization (ZF-DFE) [18]. Although V-BLAST offers full symbol rate in data transmission with low decoding complexity, it suffers from very poor performance in fading channels.

Since then, there has been considerable work on a variety of space-time transmission schemes such as space-time trellis codes [13] and space-time block codes [15]. For example, orthogonal space-time block codes [15] can provide full diversity with a linear complexity ML decoding, however, they suffer from having a limited transmission rate, and thus do not achieve the full capacity in MIMO channels [19]. Therefore, Hassibi and Hochwald [16] proposed linear dispersion (LD) codes, in which the transmitted codewrod is a linear combination of certain weighted matrices. The key to LD code design is that the basis matrices are optimized such that the resulting codes maximize the mutual information of the LD coded MIMO system. Unfortunately, for the LD codes proposed in [16], good error probability performance is not strictly guaranteed. The design of linear space-time block codes based on number theory were constructed in [20] and [21] to provide full rate and full diversity without information loss.

It was until the discovery made by Zheng and Tse [22] that shows a rigorous fundamental trade-off between the data rate increase possible via *multiplexing* versus the channel error probability reduction possible via *diversity*. Since then, the *diversity-multiplexing trade-off* has become the standard tool to compare different STC schemes and has been used as the main performance metric to evaluate any STC scheme. In the following, we summarize their main results:

Diversity-Multiplexing Tradeoff

As it is difficult to analyze the outage probability, $P_{\text{out}}(\rho, R)$, for all SNR, the evaluation is usually performed for the high SNR regime. For the random Gaussian code ensemble and the ML decoder, it has been shown in [22] that the *asymptotic* outage probability is given by

$$\begin{aligned} P_{\text{out}}(\rho, R) &= \Pr(\mathcal{O}(\rho)) \\ &\overset{\rho\to\infty}{=} \Pr\left(R > \log\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right). \end{aligned}$$

Definition 1. *Consider a family of space-time codes $\{\mathcal{C}_\rho\}$ for a given block length T and M-transmit, N-receive antennas with rate $R(\rho)$ and average error probability $P_e(\rho)$ (averaged over the random channel matrix $\boldsymbol{H}^c$). We say that the family achieves multiplexing gain r and diversity gain d if*

$$r = \lim_{\rho\to\infty} \frac{R(\rho)}{\log\rho}, \quad d = \lim_{\rho\to\infty} \frac{-\log P_e(\rho)}{\log\rho}.$$

Using the above definition, it has been shown in [22], through rigorous mathematical derivations, that the asymptotic outage probability

$$P_{\text{out}}(\rho, r\log\rho) \overset{\rho\to\infty}{=} \rho^{-d^*_{\text{out}}(r)},$$

where $d^*_{\text{out}}(r)$ is the *best* achievable outage SNR exponent which is defined in the following theorem:

Theorem 1. *(see [22]) For any $T \geq N + M - 1$, the optimal diversity-multiplexing tradeoff curve is the piecewise linear function $d^*_{\text{out}}(r)$ interpolating the points*

$$(r = k, d = (M-k)(N-k)), \quad \text{for} \quad k = 0, 1, \cdots, \min\{M, N\}.$$

An example of the optimal tradeoff in a 3×3 MIMO system is depicted in Figure 1.5. The tradeoff curve characterizes the performance capability of the quasi-static MIMO channel. At one extreme where $r = 0$, the diversity gain $d = MN$ which is the maximum diversity order achievable by the channel. At the other extreme, where $r = \min\{M, N\}$, the full degree of freedom is attained at the expense of poor performance (diversity order of 0!). The tradeoff curve bridges between the two extremes.

The analysis in [22] showed that the average error probability $P_e(\rho)$, defined in (1.1), for the random Gaussian coding and ML decoding schemes is dominated by the asymptotic outage probability, i.e.,

$$P_e(\rho) \overset{\rho\to\infty}{=} P_{\text{out}}(\rho, r\log\rho) = \rho^{-d^*_{\text{out}}(r)} \tag{1.10}$$

The key ideas of arriving to the above result are summarized by the following steps:

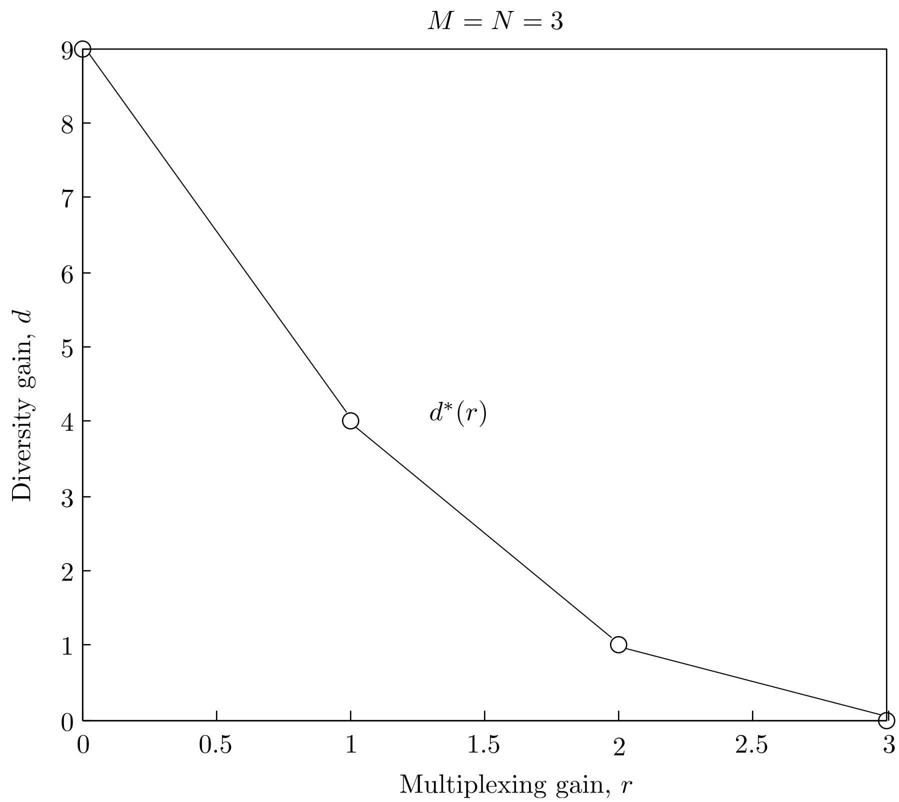


Figure 1.5: The diversity-multiplexing tradeoff of a quasi-static, flat-fading MIMO channel with $M = N = 3$ antennas.

- Separate the outage event from the non-outage event in the error probability, i.e.,

$$P_e(\rho) = \Pr(\text{error}|\boldsymbol{H}^c \in \mathcal{O}) \Pr(\mathcal{O}) + \Pr(\text{error}, \boldsymbol{H}^c \notin \mathcal{O}). \tag{1.11}$$

- When the channel is in outage, the decoder output is highly likely to be in error. Hence, the term $\Pr(\text{error}|\boldsymbol{H}^c \in \mathcal{O})$ can be upper bounded by 1.

- At high SNR, the outage probability $\Pr(\mathcal{O}(\rho))$ can be calculated as follows: denote the transmission rate $R(\rho) = r \log \rho$. Let $0 \leq \lambda_1 \leq \cdots \leq \lambda_M$ be the ordered eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c$, and define $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_M)$, $\alpha_i \triangleq -\log \lambda_i / \log \rho$. As discussed in [22], at high SNR, the non-negative values of $\boldsymbol{\alpha}$ only contributes to the outage event. Therefore, one can show that the outage event may be expressed as

$$\mathcal{O} = \lim_{\rho \to \infty} \mathcal{O}(\rho) = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^M : \sum_{i=1}^{M} (1 - \alpha_i)^+ < r \right\}. \tag{1.12}$$

In this case, the outage probability can be evaluated as follows:

$$P_{\text{out}}(\rho, r\log\rho) = \int_{\alpha\in\mathcal{O}} f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\, d\boldsymbol{\alpha},$$

where $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ is the joint probability density function of $\boldsymbol{\alpha}$ which, for all $\boldsymbol{\alpha} \in \mathcal{O}$, is asymptotically (as $\rho \to \infty$) given by [23]

$$f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) = \exp\left(-\log(\rho)\sum_{i=1}^{M}(2i-1+N-M)\alpha_i\right). \tag{1.13}$$

Applying Varadhan's lemma as in [22], as $\rho \to \infty$ we obtain

$$P_{\text{out}}(\rho, r\log\rho) = \rho^{-d^*_{\text{out}}(r)}, \tag{1.14}$$

where

$$d^*_{\text{out}}(r) = \inf_{\boldsymbol{\nu}\in\mathcal{O}} \sum_{i=1}^{M}(2i-1+N-M)\alpha_i. \tag{1.15}$$

It is not so difficult to see that the optimal channel coefficients that maximize (1.15) are

For the case that r is an integer, say $r = k$, we have:

$$\alpha_i^* = 1, \quad \text{for } i = 1, \cdots, M-k,$$

and

$$\alpha_i^* = 0, \quad \text{for } i = M-k+1, \cdots, M,$$

For the case that r is not an integer, say $r \in (k, k+1)$, we have:

$$\alpha_i^* = 1, \quad \text{for } i = 1, \cdots, M-k-1,$$

$$\alpha_i^* = 0, \quad \text{for } i = M-k+1, \cdots, M,$$

and

$$\alpha_i^* = k+1-r.$$

Substituting $\boldsymbol{\alpha}^*$ in (1.15), we get

$$d^*_{\text{out}}(r) = (M-r)(N-r), \quad \text{for } 0 \le r \le \min\{M, N\}. \tag{1.16}$$

- Using the random Gaussian code ensemble and the ML decoder, one can show [22] that the average error probability when the channel is *not* in outage is bounded from above by the outage probability, i.e., at high SNR we have

$$\Pr(\text{error}, \boldsymbol{H}^c \notin \mathcal{O}) \leq \rho^{-d^*_{\text{out}}(r)}.$$

- By using the Fano inequality, one can show [22] that the error probability is lower bounded by the outage probability. Therefore, at high SNR we have

$$P_e(\rho) \geq \rho^{-d^*_{\text{out}}(r)}. \tag{1.17}$$

Again, the drawbacks of such approach center at following: First, the lack of structure of the random Gaussian codebook that would allow for practical codeword enumeration. Secondly, the prohibitively complex ML receiver for which complexity increases exponentially with the number of transmit antennas, making it impossible to implement for large array sizes and high order digital modulation schemes.

Many works have focused on the design of low complexity encoders and decoders that are capable of achieving the optimal diversity-multiplexing tradeoff of the MIMO channel. In their paper [23], El-Gamal *et. al* have constructed a very efficient coding and decoding schemes that are able to achieve the optimal diversity-multiplexing tradeoff. The coding scheme is called *LAttice Space-Time* (LAST) coding which is developed from lattice codes designed for the AWGN channel in [8],[25]. An important factor of using LAST codes is that they can be also decoded using lattice decoders. An important ingredients for achieving the optimal tradeoff is through the use of *minimum mean-square error decision-feedback equalization* (MMSE-DFE) [29],[30].

Lattice decoding can be performed using *sphere decoders* [31]–[52]. Sphere decoders based on Ficnke-Pohst and Schnorr-Euchner (SE) enumeration [32] are known to provide ML performance with lower decoding complexity, especially for moderate-to-high SNR and small signal dimensions. Previous work on the complexity of sphere decoding focused on characterizing the mean and the variance of the decoder's complexity, particularly for the uncoded MIMO channel (e.g., V-BLAST) [52]–[54]. Seethaler *et. al.* [55] considered the derivation of the computational distribution of the sphere decoder for the $M \times N$ *uncoded* MIMO channel. It has been shown that the computational tail distribution follows a Pareto-type with tail exponent given by $N - M + 1$. However, the exact complexity analysis of the basic sphere decoder for general space-time codes applied to MIMO wireless channel is known to be difficult. In this work, **we shed the light on the computational complexity of sphere decoding for the quasi-static, LAST coded MIMO channel.**. This topic is discussed in Chapter 3.

Unfortunately , sphere decoding suffers from two drawbacks: the initial sphere-radius that must be chosen to ensure the existing of at least one lattice point inside the sphere (otherwise the search is reset with a larger sphere radius), and second is that for low-to-moderate SNR and/or large signal dimensionality, the computational complexity becomes increasingly prohibitive [52]. As such, many attempts have been made to further reduce the computational complexity of the sphere decoder .

On the other extreme, linear and non-linear receivers such as zero-forcing, minimum mean-square error (MMSE), and MMSE-DFE decoders, are considered attractive alternatives to lattice decoders in MIMO channels and have been widely used in many practical communication systems [10], [34], [35]. Unfortunately, the very low decoding complexity advantage that these decoders can provide comes at the expense of poor performance, especially for large signal dimensions. The problem of designing low complexity receivers for the MIMO channel that achieve *near-optimal* performance is considered a challenging problem and has driven much research in the past years. In this work, we analyze the performance of *lattice sequential decoding* that is capable of bridging the *gap* between sphere decoders and low complexity linear decoders (e.g., MMSE-DFE decoder).

Two main advantages result in using such an efficient decoder: we avoid the problem of selecting the appropriate sphere radius (i.e., avoid resetting the search), and for low-to-moderate SNR it achieves very low decoding complexity. It is well-known that sphere decoders can be viewed as a search in a tree for the closest lattice point to the received signal. The search can be efficiently performed using sequential decoding algorithms. Fano and Stack algorithms are two well-known algorithms that are widely used in the literature to describe the operation of the sequential decoder [36], [38]. Both algorithms were originally constructed as an alternative approach to the ML decoder for detecting convolutional codes transmitted via discrete memoryless channels. It has been shown [2] that as long as we operate below the cutoff rate, the decoder can achieve near-ML performance with complexity that scales linearly with the constraint length of the code. For the uncoded MIMO channel, specifically the V-BLAST, it has been shown in [39] that maximum receive diversity can be achieved with decoding complexity that scales linearly with the dimension of the transmitted signal. **However, no mathematical analysis has been provided for the coded MIMO channel, and the optimal diversity-multiplexing tradeoff of the coded MIMO channel that can be achieved under the use of lattice sequential decoders has not yet been studied**. This topic is fully addressed in Chapter 4.

1.3 The MIMO ARQ Channel

Another interesting channel model that will be considered in this work, which falls in to the class of linear Gaussian vector channel model is the so-called “The MIMO Automatic Repeat reQuest (ARQ) Channel”. ARQ is an important error-control mechanism that is widely used in high-speed wireless mobile networks (e.g., wireless LAN, WiMAX, and Wi-Fi) due to its very low error probability detection and generally low decoding complexity. Recently, there has been a great interest in MIMO-ARQ wireless communication systems.

In the MIMO ARQ system, the *delay* introduced by the channel provides a third dimension in the tradeoff region. The tradeoff between the diversity, the multiplexing, and the delay provided by the quasi-static MIMO ARQ channel has been established in the paper by El-Gamal et. al. in [41]. It has been shown in [41] that LAST codes achieve the optimal *diversity-multiplexing-delay* tradeoff (see Chapter 5 for more details about this result).

A powerful coding technique referred to as Incremental Redundancy LAttice Space-Time (IR-LAST) coding scheme has been constructed in [41] to achieve the optimal tradeoff of the MIMO ARQ channel. This scheme is the ARQ version of the LAST codes that is used to achieve the optimal tradeoff of the quasi-static, Rayleigh-fading MIMO channel. These lattice-based construction of space-time codes were designed using linear random coding techniques [41]. The problem of constructing explicit optimal IR-LAST codes for the above mentioned MIMO ARQ channel was discussed in [42].

However, in both papers, a *list lattice decoder*, for joint error correction and detection, implemented via sphere decoder is an essential part for achieving the optimal tradeoff. The optimality of such joint decoder is limited only to the high SNR regime and small system dimensions, and for low-to-moderate SNR and large system dimensions, the size of the candidate list could become large. This motivates us to search for a more efficient joint decoding technique that is capable of achieving the optimal tradeoff with a fairly low decoding complexity.

This property makes sequential decoding of lattice space-time codes a very attractive option in MIMO ARQ systems requiring very low undetected error probabilities. Many sequential decoding algorithms (e.g., the stack algorithm) were modified for the use of signal detection and decoding in ARQ systems. Among those algorithms that is considered simple but efficient is the so-called *time-out* sequential decoding. In this algorithm, the decoder simply tracks the number of computations performed by the decoder and asks for retransmission if the computations become excessive and exceed a certain pre-determined time limit. This results in reducing the decoding complexity by terminating the search during high channel noise. For the case of single-input single-output discrete memoryless channel, it was shown (see [43]) that there exists an optimal time-out limit value that maximizes both performance and throughput while achieving low decoding complexity. Here, we would like to extend the work in [43] to the MIMO ARQ channel, particularly, to the quasi-static, Rayleigh-fading MIMO ARQ channel. We will study the throughput-performance-complexity tradeoffs in sequential decoding algorithms and the effect of preprocessing and termination strategies such as the time-out algorithm.

We propose an efficient approach for joint detection and decoding based on the Fano/Stack sequential decoder [39] developed for the closest lattice point search problem. We show (via simulation and analysis) that the optimal tradeoff can be achieved using such a decoder with significant reduction in average decoding complexity.

1.4 Outline

The thesis is organized as follows: In **Chapter 2**, we briefly introduce an important topic in mathematics, specifically in number theory, which is the so called *lattices*. The application of lattices in various communication systems is presented. Basically we examine capacity-achieving lattice codes under lattice decoding for the linear Gaussian vector channel.

Our novel analysis is mainly concentrated in Chapter 3, Chapter 4, and Chapter 5. In **Chapter 3**, we review some previous results about the performance of the lattice decoder implemented via sphere decoding algorithms in the quasi-static LAST coded MIMO channel. We provide a complete analysis on the sphere decoding complexity at the high SNR regime.

In **Chapter 4**, we introduce lattice sequential decoders that are used as an alternative to sphere decoders to solve the closest lattice point search problem. We investigate the achievable rates of lattice sequential decoders for the outage-limited MIMO channel, and we derive the *general* diversity-multiplexing tradeoff achieved by the decoder as a function of its parameter — *the bias term*. We show how this parameter plays a fundamental role in determining the diversity-multiplexing tradeoff achieved by sequential decoding of lattice codes. This bias term is critical for controlling the amount of computations required at the decoding stage and is responsible for the excellent performance-rate-complexity tradeoff achieved by the decoder. We also provide a complete analysis for both the computational complexity tail distribution and the average complexity of the lattice sequential decoder in the high SNR regime.

In **Chapter 5**, the application of the lattice sequential decoder in the MIMO ARQ channel is introduced. Particularly, we investigate the asymptotic performance limits of the lattice sequential decoder when a time-out limit is imposed. We prove its optimality in terms of achieving the optimal tradeoff of the channel. We also provide complexity analysis of such a decoder and compare it the complexity of the optimal (sphere) decoder to demonstrate the complexity saving advantage achieved by the former decoder. Finally, we verify all our theoretical results using simulations. A summary of the main findings is presented in **Chapter 6**, including an outlook on future work.

1.5 Notations

Through out the thesis, we use the following notations:

- ◇ Bold lowercase letters $\boldsymbol{a}$ denote vectors, whose l_2-norm is denoted by $\|\boldsymbol{a}\|$. The notation $\boldsymbol{a}_1^k$ refers to the vector that contains the last k components of $\boldsymbol{a}$.
- ◇ Bold uppercase letters $\boldsymbol{A}$ denote matrices. The ith column vector is denoted by $\boldsymbol{a}_i$. In the definition of matrices and vectors the convention $\boldsymbol{A} \in \mathbb{A}^{n\times m}$ is used to denote a matrix with n rows and m columns whose components are taken from the set $\mathbb{A}$. Then notation $\boldsymbol{A}_{kk}$ denotes the lower $k \times k$ part of the square matrix $\boldsymbol{A}$.
- • The notations $\Re\{\cdot\}$ and $\Im\{\cdot\}$ represent the real part and the imaginary part of a complex number, respectively.
- ◇ $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix and $\otimes$ denotes the Kronecker product.
- ◇ The complement of a set $\mathcal{B}$ is $\overline{\mathcal{B}}$.
- ◇ The superscript ^{c} denotes complex quantities, $^\mathsf{T}$ denotes transpose, and $^\mathsf{H}$ denotes Hermitian transpose.
- ◇ The notation $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$ indicates that $\boldsymbol{v}$ is a real Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{K}$.
- ◇ For a bounded Jordan-measurable region $\mathcal{R} \subset \mathbb{R}^m$, $V(\mathcal{R})$ denotes the volume of $\mathcal{R}$. We denote

$$\mathcal{S}_m(r_s) = \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\| \leq r_s\}$$

by the m-dimensional hypersphere of radius r_s with volume

$$V(\mathcal{S}_m(r_s)) = \frac{(\pi r_s^2)^{m/2}}{\Gamma(m/2+1)}$$

where $\Gamma(n) = \int_0^\infty t^{n-1}e^{-t}\,dt$ is the *Gamma* function.

◇ We use $\doteq$ to denote exponential equality, i.e., we refer to $g(z) \doteq z^a$ as $\lim_{z\to\infty} g(z)/\log(z) = a$, $\dot{\geq}$ and $\dot{\leq}$ are used similarly.

- The positive part of a real variable x is denoted by $(x)^+ = \max\{0, x\}$.

- The notation $\mathsf{E}\{\cdot\}$ represents the statistical average.

Chapter 2

Capacity-Achieving Coding and Decoding Schemes from A Lattice Theory Prospective

THE present chapter is devoted to the theory of (real) lattices. Our treatment is short and focused on those aspects that are of direct importance for the construction of lattice codes which are appropriate for the linear Gaussian vector channel model, and for the design of efficient algorithms that solve the closest lattice point search (CLPS) problem [44] (referred here to as lattice decoding). For a full account of the general theory of lattices we refer the reader to the specialized texts, in particular to the rich text by Conway and Sloane [6].

2.1 Lattices: An Introduction

A *lattice* is a discrete pointset Λ in a Euclidean space $\mathbb{R}^m$ that is closed under vector addition, i.e., any translate $\Lambda + \boldsymbol{x}$ by a lattice point $\boldsymbol{x} \in \Lambda$ is just Λ again. Let $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m\}$ be a set of linearly independent vectors in $\mathbb{R}^m$. The set Λ of all linear combinations $\boldsymbol{x} = z_1\boldsymbol{g}_1 + z_2\boldsymbol{g}_2 + \cdots + z_m\boldsymbol{g}_m$ with integer coefficients z_i is a lattice, i.e.,

$$\Lambda = \{\boldsymbol{x} = \boldsymbol{G}\boldsymbol{z} : \boldsymbol{z} \in \mathbb{Z}^m\},$$

where $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m]$ is an $m \times m$ full-rank generator matrix. Thus, any lattice Λ in $\mathbb{R}^m$ can be seen as a linear transformation of the integer lattice $\mathbb{Z}^m$. Figure 2.1 shows some examples of lattices in $\mathbb{R}^m$.

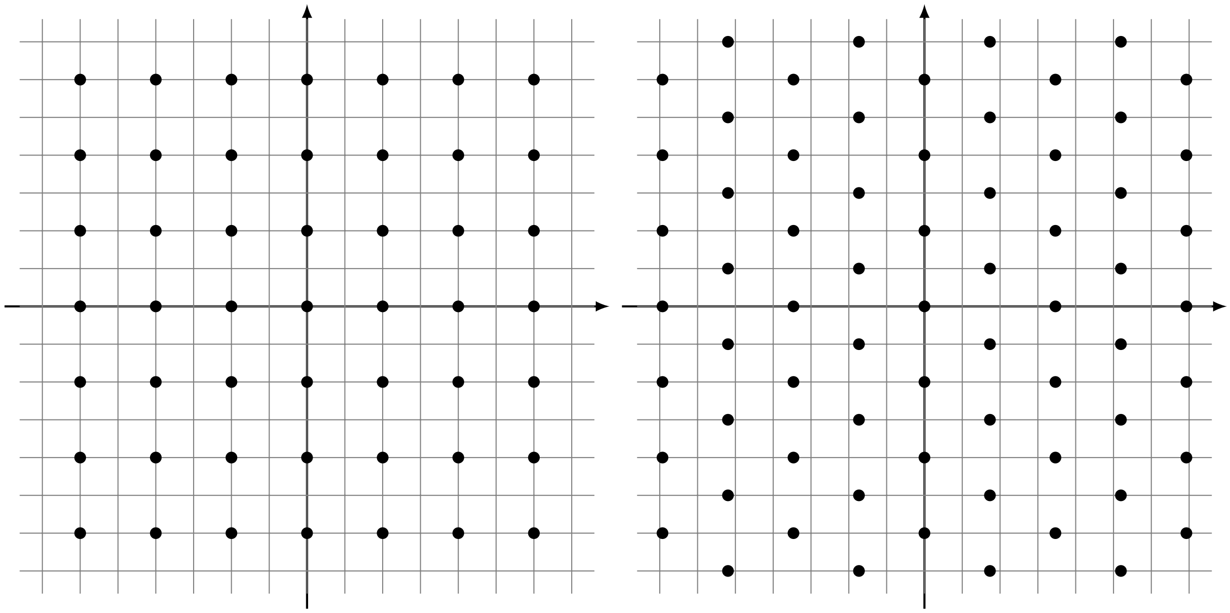


Figure 2.1: Examples of lattices in $\mathbb{R}^2$: The left figure corresponds to the rectangular lattice with $\boldsymbol{g}_1 = [1\ 0]^\mathsf{T}$, and $\boldsymbol{g}_2 = [0\ 1]^\mathsf{T}$. The right figure is the hexagonal lattice with $\boldsymbol{g}_1 = [0\ 1]^\mathsf{T}$, and $\boldsymbol{g}_2 = [\sqrt{3}/2\ 1/2]^\mathsf{T}$.

One can associate with every lattice point $\boldsymbol{x} \in \Lambda$ a bounded region $\mathcal{E}_{\boldsymbol{x}} \in \mathbb{R}^m$. The region $\mathcal{E}$ is called a *fundamental region* of the lattice Λ if every element in $\mathbb{R}^m$ can be uniquely written as the sum of an element in $\mathcal{E}$ and a lattice point in Λ. In other words, the translates of a fundamental region by the lattice points tile $\mathbb{R}^m$. The fundamental paralleltope $\mathcal{P}$ is an example of a fundamental region for the lattice Λ which is defined as

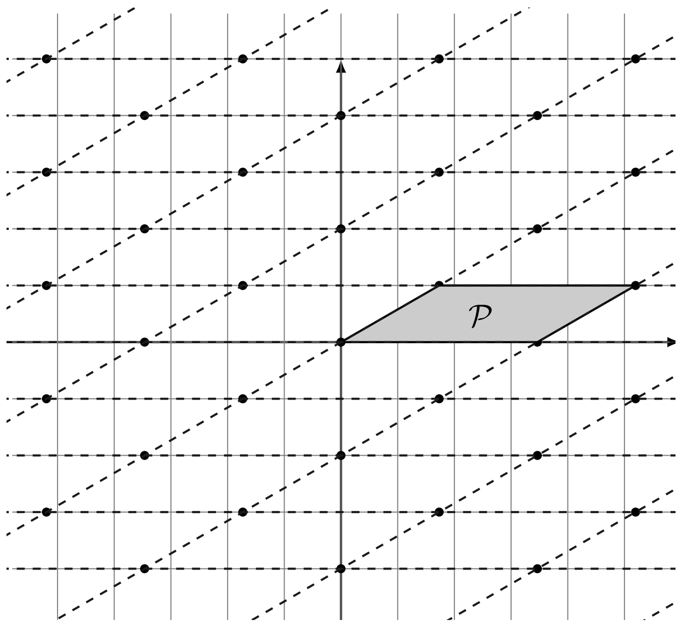
$$\mathcal{P} = \{u_1\boldsymbol{g}_1 + u_2\boldsymbol{g}_2 + \cdots + u_m\boldsymbol{g}_m : 0 \leq u_i \leq 1\}.$$

Another important fundamental region that will be widely used in this work is the so-called the *Voronoi cell.* For a lattice Λ generated by a matrix $\boldsymbol{G}$, the Voronoi (cell) region, $\mathcal{V}_{\boldsymbol{x}}(\boldsymbol{G})$, associated with a lattice point $\boldsymbol{x} \in \Lambda$ is defined as the set of points in $\mathbb{R}^m$ closest to $\boldsymbol{x}$, i.e.,

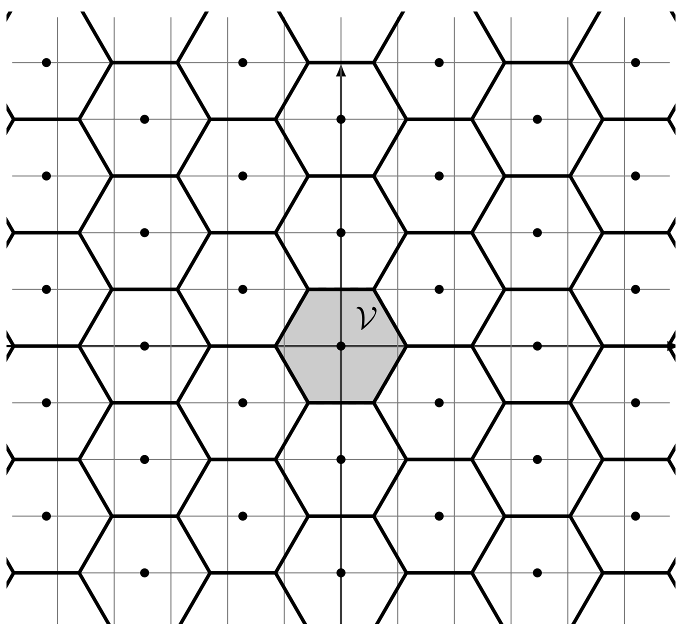
$$\mathcal{V}_{\boldsymbol{x}}(\boldsymbol{G}) = \{\boldsymbol{u} \in \mathbb{R}^m : \|\boldsymbol{x} - \boldsymbol{u}\| \leq \|\boldsymbol{\lambda} - \boldsymbol{u}\|,\ \boldsymbol{x} \neq \boldsymbol{\lambda} \in \Lambda\}$$

Figure 2.2 shows two fundamental regions associated with the hexagonal lattice. It must be noted that all fundamental regions of a lattice Λ have equal volumes, say $V_f(\Lambda)$, which is given by

$$V_f(\Lambda) = \sqrt{\det(\boldsymbol{G}^\mathsf{T}\boldsymbol{G})}.$$



(a) The fundamental paralleltope region.



(b) The fundamental Voronoi region.

Figure 2.2: Some fundamental regions of the hexagonal lattice.

Next, we introduce some properties associated with a lattice Λ having the Voronoi cell as its fundamental region, which are of great importance to our analysis:

1. The **nearest neighbor quantizer** $Q_\Lambda(\cdot)$ associated with Λ is defined by

$$Q_\Lambda(\boldsymbol{u}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \|\boldsymbol{\lambda} - \boldsymbol{u}\|.$$

2. We define the **modulo-lattice function** of $\boldsymbol{u} \in \mathbb{R}^m$

$$[\boldsymbol{u}] \mod \Lambda = \boldsymbol{u} - Q_\Lambda(\boldsymbol{u}).$$

3. The **second moment per dimension** of Λ is defined by

$$\sigma^2(\Lambda) \triangleq \frac{1}{m}\mathsf{E}\{\|\boldsymbol{u}\|^2\} = \frac{1}{mV_f(\Lambda)}\int_{\boldsymbol{\nu}\in\mathcal{V}_0(\boldsymbol{G})} \|\boldsymbol{\nu}\|^2 \, d\boldsymbol{\nu},$$

where $\boldsymbol{u}$ is a random vector uniformly distributed over $\mathcal{V}_{\mathbf{0}}(\boldsymbol{G})$.

4. The **normalized second moment** associated with Λ is given by

$$G(\Lambda) \triangleq \frac{\sigma^2(\Lambda)}{V_f(\Lambda)^{2/m}}.$$

One can show that as $m \to \infty$, the asymptotic normalized second moment $G(\Lambda) \to 1/2\pi e$. In fact, $G(\Lambda)$ is always greater than $1/2\pi e$, and for sufficiently large m, there exist lattices whose Voronoi region approaches a sphere. This is equivalent to saying that a random vector $\boldsymbol{u}$ uniformly distributed over $\mathcal{V}_{\mathbf{0}}$ converges in distribution (in the sense of divergence) to a Gaussian i.i.d random vector with per component variance equal to $\sigma^2(\Lambda)$, i.e., $\frac{1}{m}h(\boldsymbol{u})$ is close to $\frac{1}{2}\log(2\pi e\sigma^2(\Lambda))$.

In this case, a sequence of lattices $\{\Lambda_m\}$ of increasing dimension is said to be **good for mean-square error quantization** [26] if $G(\Lambda_m) \to 1/2\pi e$.

5. For a given radius r_s, the set $\Lambda + \mathcal{S}_m(r_s)$ is a *packing* in the Euclidean space if for all lattice points $\boldsymbol{x}, \boldsymbol{y} \in \Lambda$, $\boldsymbol{x} \neq \boldsymbol{y}$, we have $(\boldsymbol{x} + \mathcal{S}_m(r_s)) \cap (\boldsymbol{y} + \mathcal{S}_m(r_s)) = \emptyset$. The **packing radius** (see Figure 2.3) $r_{\text{pack}}(\Lambda)$ is defined by

$$r_{\text{pack}}(\Lambda) = \sup\{r_s : \Lambda + \mathcal{S}_m(r_s) \text{ is a packing}\}.$$

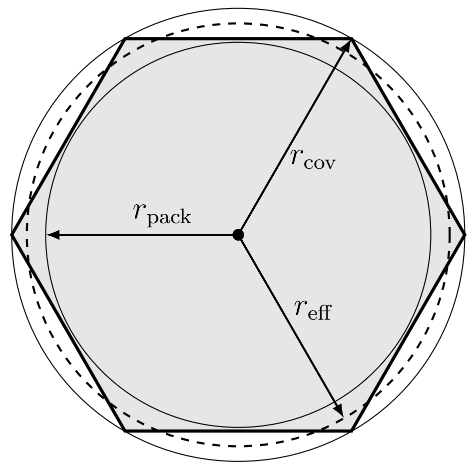


Figure 2.3: The packing radius, the effective radius, and the covering radius of the hexagonal lattice.

6. The **effective radius** $r_{\text{eff}}(\Lambda)$ is the radius of the sphere centered at the origin with volume equal to the volume of the Voronoi region (see Figure 2.3), i.e.,

$$V(\mathcal{S}_m(r_{\text{eff}})) = V(\mathcal{V}_{\mathbf{0}}(\boldsymbol{G})).$$

7. The **covering radius** $r_{\text{cov}}(\Lambda)$ is the radius of the smallest sphere centered at the origin that contains $\mathcal{V}_0$ (see Figure 2.3).

8. A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is **good for covering** [25] if their **covering efficiency**, $\eta_{\text{cov}}(\Lambda_m)$, satisfies

$$\eta_{\text{cov}}(\Lambda_m) \triangleq \frac{r_{\text{cov}}(\Lambda_m)}{r_{\text{eff}}(\Lambda_m)} \to 1.$$

9. A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is **good for packing** [25] if their **packing efficiency**, $\eta_{\text{pack}}(\Lambda_m)$, satisfies

$$\eta_{\text{pack}}(\Lambda_m) \triangleq \frac{r_{\text{pack}}(\Lambda_m)}{r_{\text{eff}}(\Lambda_m)} \geq \frac{1}{2}$$

10. **Minkowski-Hlawka Theorem**[27]: Let $f : \mathbb{R}^m \to \mathbb{R}$ be a Riemann integrable function of bounded support (i.e., $f(\boldsymbol{x}) = 0$ if $\|\boldsymbol{x}\|$ exceeds some bound). For any $\delta >$

0, there exist ensembles $\{\Lambda\}$ of lattices with fundamental volume V_f and dimension m such that

$$\mathsf{E}_{\Lambda}\left\{\sum_{\boldsymbol{x}\in\Lambda^*} f(\boldsymbol{x})\right\} \leq (1+\delta)\frac{1}{V_f(\Lambda)}\int_{\mathbb{R}^m} f(\boldsymbol{x})\, d\boldsymbol{x}. \tag{2.1}$$

where $\Lambda^* = \Lambda\backslash\{\mathbf{0}\}$, and $\delta \to 0$ as $m \to \infty$.

The above important theorem is sometimes regarded as a pre-Shannon result in information theory. In fact, the Mikowski-Hlawka theorem was originally used for packing lattices to solve the well-known *sphere packing* problem.

11. A pair of m-dimensional lattices $\{\Lambda, \Lambda'\}$ is **nested**, i.e., $\Lambda' \subset \Lambda$, if there exists corresponding generator matrices $\boldsymbol{G}$ and $\boldsymbol{G}'$ such that $\boldsymbol{G}' = \boldsymbol{N}\boldsymbol{G}$. Where $\boldsymbol{N}$ is an $m \times m$ integral matrix whose determinant is greater than one [6].

    A special case of nested lattices, which will be frequently used in our analysis and simulation, is the **self-similar** lattices. In such lattices Λ' is just a scaled version of Λ, i.e., $\Lambda' = \mathcal{Q}\Lambda$, where $\mathcal{Q} \in \mathbb{Z}^+$ is called the *nesting ratio*.

At this stage, we would like to point out to the reader that most of the results that have been introduced above are asymptotic in the dimension m. However, in this work we will be mostly dealing with communication systems with finite dimensionality. Therefore, the following result is of a great importance to our analysis (see the proof of Theorem 5 in [24]):

12. Suppose that a random vector $\boldsymbol{u}$ of (finite) dimension m is uniformly distributed over the Voronoi fundamental region $\mathcal{V}_{\mathbf{0}}$ of a lattice Λ with second order moment $\sigma^2(\Lambda) = \sigma^2$. Then, the probability density function of $\boldsymbol{u}$, say $f_{\boldsymbol{u}}(\boldsymbol{\nu})$, can be shown to be upper bounded by

$$f_{\boldsymbol{u}}(\boldsymbol{\nu}) \leq \beta_m f_{\boldsymbol{g}}(\boldsymbol{\nu}), \tag{2.2}$$

    where β_m is a constant that approaches unity as $m \to \infty$, and $f_{\boldsymbol{g}}(\nu)$ is the probability density function of a zero-mean Gaussian i.i.d random vector with per component variance σ^2, i.e., $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \sigma^2)$.

Next, we would like to introduce lattice codes, particulary (random) lattice codes that are based on Construction A. These codes will be intensively used in this work.

2.2 Lattice Codes: Construction A

Recently, there has been a tremendous amount of work in designing "good" lattice codes for the general linear Gaussian vector channel such as, point-to-point quasi-static MIMO channel, multiple access channel, interference channel, and more. This is due to the main three facts about lattice codes: first, its **low encoding complexity** as opposed to the unstructured Gaussian random code ensemble that are very difficult to construct in practice, its ability to **achieve the fundamental limits** of the standard random coding arguments, and most importantly, these codes can be **efficiently decoded** (at least more efficiently than random Gaussian codes).

It must be noted here that, although we use structured (lattice) codes, we resort to a random ensemble of lattices to derive our main results. For many communication channels (including the ones that are presented in this work), *explicit* lattice codes that achieve optimal performance limits can be found in the literature (see for example [42], [48]). However, in this work, our main concern is not in the encoder construction rather it is in the decoder design, particularly in the design of low complexity decoders that achieve *near*-optimal performance (more about this topic will be discussed in the subsequent chapters).

Lattices have become a standard tool for the construction of both block and convolutional trellis codes for the linear Gaussian vector channel [6], [49], [50]. In this work, lattice codes based on *linear* block codes will be considered. Such lattice codes consists of the intersection of a lattice Λ_c, termed as the *coding lattice* (or a translate of a lattice $\Lambda_c + \boldsymbol{u}_0$) with a bounded region. The bounded region is chosen to satisfy the input power constrained and the transmission rate.

Definition 2. *An m-dimensional lattice code $\mathcal{C}(\Lambda_c, \boldsymbol{u}_o, \mathcal{R})$ is the finite subset of the lattice translate $\Lambda_c + \boldsymbol{u}_0$ inside the shaping region $\mathcal{R}$, i.e.,*

$$\mathcal{C}(\Lambda_c, \boldsymbol{u}_o, \mathcal{R}) = \{\Lambda_c + \boldsymbol{u}_0\} \cap \mathcal{R},$$

where $\mathcal{R}$ is a bounded measurable region of $\mathbb{R}^m$.

The vector $\boldsymbol{u}_0$ is chosen so that the number of lattice "codewords" in the code $\mathcal{C}$ is maximized[1] by ensuring the boundary of the shaping region $\mathcal{R}$ does not contain any lattice point $\boldsymbol{x} \in \Lambda_c$. If this is the case, for a self-similar lattice, where $\boldsymbol{G}_s = \mathcal{Q}\boldsymbol{G}_c$, one can show

[1]There are infinitely many choices of $\boldsymbol{u}_0$, however, this vector is usually selected so that the average power of the code is minimized.

that the total number of lattice codewords in the code $\mathcal{C}$ is given by

$$|\mathcal{C}| = \frac{V(\mathcal{V}_s)}{V_f(\Lambda_c)} = \mathcal{Q}^m,$$

or equivalently, the transmission rate is given by $R = \frac{1}{m}\log|\mathcal{C}| = \log\mathcal{Q}$ (see Figure 2.4). In general, this is a very difficult problem for any dimension m, and therefore, we resort

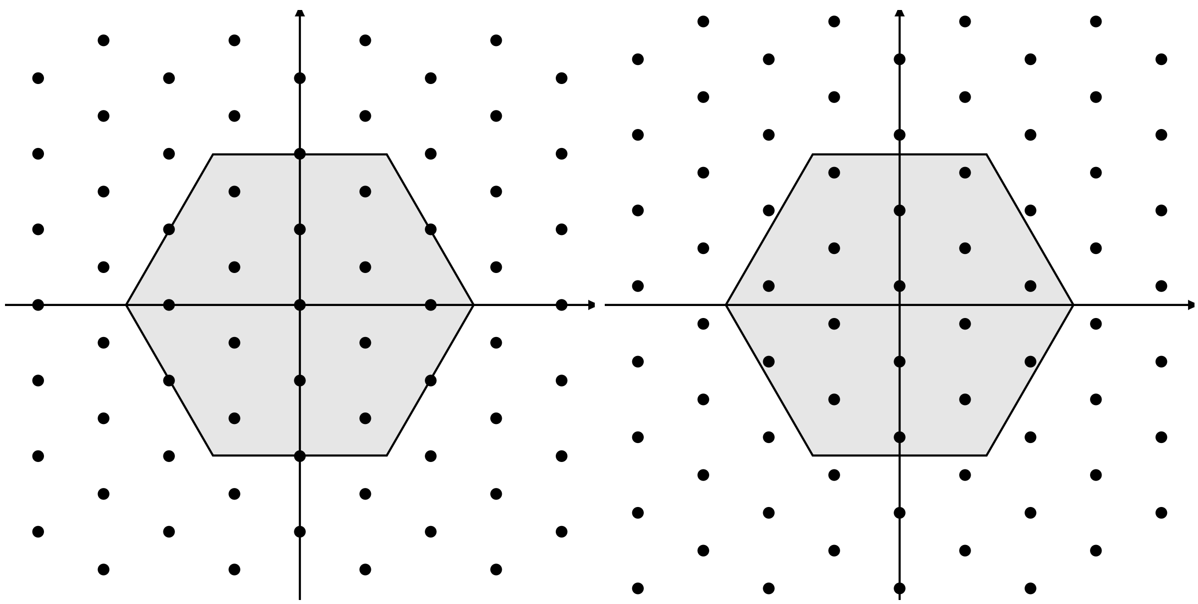


Figure 2.4: On the left, a nested lattice code with codewords located on the boundary of the Voronoi region. The right figure shows a nested code $\mathcal{C} = \{\Lambda_c + \boldsymbol{u}_0\} \cap \mathcal{V}_s$ where $\boldsymbol{u}_0 = (0, -0.25)$ consisting of all 16 points inside the Voronoi region. This code has the lowest average power of any known set of 16 points in the plane.

to a random technique that shows (see Lemma 2 in [8]).

Lemma 1. *For any Λ and $\mathcal{R}$, there exists $\boldsymbol{u}_0$ such that*

$$|\mathcal{C}(\Lambda, \boldsymbol{u}_o, \mathcal{R})| \geq \frac{V(\mathcal{R})}{V_f(\Lambda)}. \tag{2.3}$$

Now, depending on the characteristic of the shaping region $\mathcal{R}$, different lattice codes can be constructed. For example, if the shaping region $\mathcal{R}$ is chosen to be an m-dimensional hypersphere of radius R_s (where R_s is chosen to satisfy the power constraint), then the generated lattice code is called *spherical* lattice code [51]. If $\mathcal{R}$ is chosen to be the Voronoi region of another lattice, say Λ_s, where $\Lambda_s \subset \Lambda_c$, the generated lattice code is called the *Voronoi* code or *nested* code [6]. In the latter case, Λ_s is usually referred to as the *shaping lattice*.

Definition 3. *Let Λ_c be a lattice in $\mathbb{R}^m$ and Λ_s be a sublattice of Λ_c. The nested lattice code defined by the partitioning Λ_c/Λ_s is given by*

$$\mathcal{C} = \Lambda_c \cap \mathcal{V}_s,$$

where $\mathcal{V}_s$ is the fundamental Voronoi region of Λ_s. In other words, $\mathcal{C}$ is formed by the coset leaders of the cosets of Λ_s in Λ_c.

As mentioned previously, in this work we focus our analysis on *nested* lattice codes, specifically lattice codes that are generated using **construction A** which is described below (see [6], [8]).

We consider the Loeliger ensemble of mod-p lattices, where p is a prime. First, we generate the set of all lattices given by

$$\Lambda_p = \kappa(\mathsf{C} + p\mathbb{Z}^m),$$

where $\mathbb{Z}_p$ denotes the field of mod-p integers, $\mathsf{C} \subset \mathbb{Z}_p^m$ is a linear code over $\mathbb{Z}_p$ with generator matrix in systematic form $[\boldsymbol{I}\ \boldsymbol{P}^{\mathsf{T}}]^{\mathsf{T}}$, where $\boldsymbol{P}$ is a $m-k \times k$ parity check matrix, and $p \to \infty$, $\kappa \to 0$ is a scaling coefficient chosen such that the fundamental volume $V_f = \kappa^{2MT} p^{2MT-k} = 1,$. We use a pair of self-similar lattices for nesting. We take the shaping lattice to be $\Lambda_s = \phi\Lambda_p$, where ϕ is chosen carefully in order to satisfy the input power constraint. Finally, the coding lattice is obtained as $\Lambda_c = \zeta\Lambda_s$, where ζ is chosen appropriately to satisfy the transmission rate constraint.

Interestingly, one can construct a generator matrix of Λ_p as (see [6])

$$\boldsymbol{G}_p = \kappa \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{P} & p\boldsymbol{I} \end{pmatrix}, \tag{2.4}$$

which has a lower triangular form.

In order to be able to apply the above lattice coding scheme to the linear Gaussian vector channel, we will appeal to the following real channel model equivalent to (1.3). After T channel uses, the received signal is given by

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{e}, \tag{2.5}$$

where

$$\boldsymbol{x} = (\boldsymbol{x}_1^\mathsf{T}, \boldsymbol{x}_2^\mathsf{T}, \cdots, \boldsymbol{x}_T^\mathsf{T})^\mathsf{T},$$

with $\boldsymbol{x}_t = [\Re\{\boldsymbol{x}_t^c\}, \Im\{\boldsymbol{x}_t^c\}]$, and

$$\boldsymbol{e} = (\boldsymbol{e}_1^\mathsf{T}, \boldsymbol{e}_2^\mathsf{T}, \cdots, \boldsymbol{e}_T^\mathsf{T})^\mathsf{T},$$

with $\boldsymbol{e}_t = [\Re\{\boldsymbol{e}_t^c\}, \Im\{\boldsymbol{e}_t^c\}]$, and the (real) channel matrix is given by

$$\boldsymbol{M} = \begin{pmatrix} \Re\{\boldsymbol{M}^c\} & -\Im\{\boldsymbol{M}^c\} \\ \Im\{\boldsymbol{M}^c\} & \Re\{\boldsymbol{M}^c\} \end{pmatrix}.$$

Here, we have $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{e} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^n$, and $\boldsymbol{M} \in \mathbb{R}^{n\times m}$, where $n = 2NT$ and $m = 2MT$.

Note that the above configuration applies to the "non-ergodic" channel model (e.g., the quasi-static MIMO channel), where the channel matrix is random but remains fixed during all channel uses (i.e., $\boldsymbol{M}_t^c = \boldsymbol{M}^c$).

We say that an $M \times T$ space-time coding scheme is a full-dimensional LAttice Space-Time (LAST) code if its vectorized (real) codebook (corresponding to the channel model (2.5)) is a lattice code with dimension $m = 2MT$. As discussed in [23], the design of space-time signals reduces to the construction of a codebook $\mathcal{C} \subseteq \mathbb{R}^{2MT}$ with code rate $R = \frac{1}{T}\log|\mathcal{C}|$, satisfying the input averaging power constraint

$$\frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}\|^2 \leq MT. \tag{2.6}$$

2.3 Lattice Decoder: Near Optimal Decoding

The search for low complexity decoders that achieve *near* optimal performance is an ongoing research for information and coding theorist. For the linear Gaussian vector channel that is considered in this work, it is well-known that the ML decoder is the *optimal* solution that minimizes the word error probability. In such a decoder, the received signal is decoded to the nearest codeword or lattice point inside $\mathcal{R}$. For a random Gaussian code ensemble, the ML decoder achieves the capacity of the channel.

However, the ML decoder performs an exhaustive search over all codewords that belongs to the code $\mathcal{C}$, and hence it is considered very complex to implement in practice. Figure 2.5 shows an example of the ML decoder, where the Euclidean space is divided into subregions based on the distance between each codeword in the code. These decision regions are optimal in terms of minimizing the probability of decoding error.

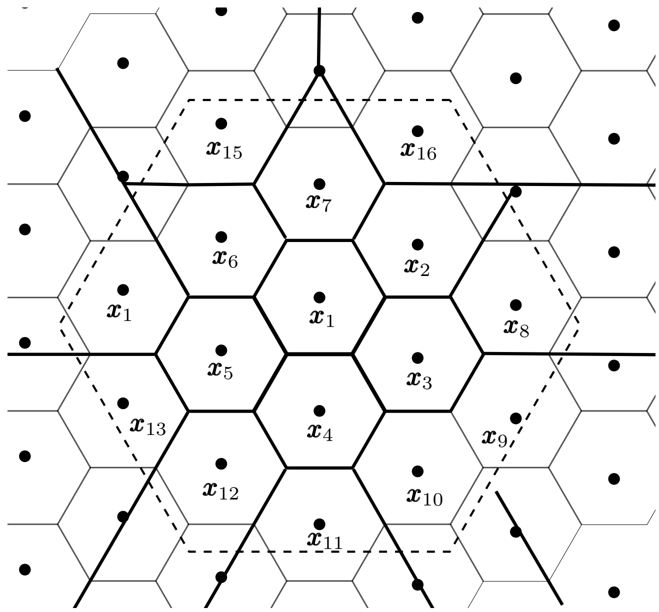


Figure 2.5: The decision regions that corresponds to the ML decoder for a hexagonal nested lattice code with nesting ratio 4.

For the real linear Gaussian vector channel model, the ML decoder can be expressed as

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|^2. \tag{2.7}$$

Searching over the codebook $\mathcal{C}$ is performed by a search algorithm (e.g., the sphere decoder) that takes into account the shaping region $\mathcal{R}$ which is referred to as *boundary control*. Due to its exponential complexity, the implementation of such optimal decoder is practically unfeasible and the design of low complex receivers that achieve near optimal performance is considered a challenging problem.

Relaxing the boundary control, or *lattice decoding*, is believed to reduce complexity at the expense of introducing some error performance degradation. Lattice decoder algorithms reduce complexity by relaxing the code boundary constraint and find the point of the underlying (infinite) lattice closest to the received point (which may or may not be a code point). In lattice theory, this is usually referred to as the closest lattice point search problem (CLPS) [44], which can be described by

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\Lambda_c} \|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|^2. \tag{2.8}$$

Many researchers have studied the information-theoretic limits of lattice coding and decoding schemes for the linear Gaussian vector channel model [5], [7], [8], [24], [23], [41]. A very interesting approach that may be used to prove the rate achievability of lattice coding and decoding schemes for such a channel is through the so-called the *ambiguity decoder*. Lattice ambiguity decoder was originally developed by Loeliger in [8] for the AWGN channel (see (2.5), where $\boldsymbol{M} = \sqrt{\rho}\boldsymbol{I}$) and was used in [23] to prove the achievability rate of the lattice decoder for the quasi-static, Rayleigh fading MIMO channel. The same technique will be used in this work to analyze the achievable rate of other efficient lattice decoders for the quasi-static MIMO channel. Therefore, for convenience, we introduce the ambiguity decoder and we extend it to the real linear Gaussian vector channel model:

Assume the received vector can be written as $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{x} \in \Lambda_c$ and $\boldsymbol{w} = \boldsymbol{M}^{-1}\boldsymbol{e}$ is an m-dimensional noise vector independent of $\boldsymbol{x}$, for which $\boldsymbol{M} \in \mathbb{R}^{m\times m}$ is an arbitrary full-rank matrix and $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, 0.5\boldsymbol{I})$. The ambiguity decoder is defined by a decision region $\mathcal{E} \subset \mathbb{R}^m$ and outputs $\boldsymbol{x} \in \Lambda_c$ if $\boldsymbol{y} \in \mathcal{E} + \boldsymbol{x}$ and there exists no other point $\boldsymbol{x}' \in \Lambda_c$ such that $\boldsymbol{y} \in \mathcal{E} + \boldsymbol{x}'$. An ambiguity occurs if the received vector $\boldsymbol{y} \in \{\mathcal{E} + \boldsymbol{x}\} \cap \{\mathcal{E} + \boldsymbol{x}'\}$ for some $\boldsymbol{x} \neq \boldsymbol{x}'$. If we define $\mathcal{A}(\mathcal{E})$ to be the ambiguity event for the decision region $\mathcal{E}$, then for a given Λ_c and $\mathcal{E}$, the probability of error can be upper bounded as

$$P_e(\mathcal{E}|\Lambda_c) \leq \Pr(\boldsymbol{e} \notin \mathcal{E}) + \Pr(\mathcal{A}(\mathcal{E})). \tag{2.9}$$

As mentioned in [8], the upper bound (2.9) holds for any Jordan measurable bounded subset $\mathcal{E}$ of $\mathbb{R}^m$. Consider now the following lemma:

Lemma 2. There exists an $m = 2MT$-dimensional lattice code $\mathcal{C}(\Lambda_c, \boldsymbol{u}_0, \mathcal{R})$ with fundamental volume V_c that satisfies (2.6), for some fixed translation vector $\boldsymbol{u}_0$, and $\mathcal{R}$ is the $m/2$-dimensional hypersphere with radius $\sqrt{MT}$ centered at the origin such that the error probability is upper bounded as

$$P_e(\Lambda_c, \mathcal{E}_{T,\gamma}) \leq (1+\epsilon')2^{-T[\log\det(\boldsymbol{M}^\mathsf{T}\boldsymbol{M})^{1/2T} - M\log(r_e^2/MT) - R]} + \Pr(\boldsymbol{e} \notin \mathcal{E}_{T,\gamma}), \tag{2.10}$$

where $\mathcal{E}_{T,\gamma} \triangleq \{\boldsymbol{z} \in \mathbb{R}^{2MT} : \boldsymbol{z}^\mathsf{T}\boldsymbol{M}^\mathsf{T}\boldsymbol{M}\boldsymbol{z} \leq r_e^2(1+\gamma)\}$, $r_e > 0$, $\gamma > 0$, and $\epsilon' > 0$.

Proof. See [23]. □

We consider the achievable rate of some special cases:

For the power-constrained AWGN channel:

the achievable rate under lattice decoding provided in (2.8) follows easily by letting $\boldsymbol{M} = \sqrt{\rho}\boldsymbol{I}$, $M = 1$, and $r_e^2 = T$ in the above lemma. In that case, from the standard typicality arguments it follows that for any $\epsilon > 0$ and $\gamma > 0$, there exists $T_{\gamma,\epsilon}$ such that for all $T > T_{\gamma,\epsilon}$ we have that $\Pr(\boldsymbol{e} \notin \mathcal{E}_{T,\gamma}) < \epsilon/2$. The second term in the upper bound (5.59) can be made smaller than $\epsilon/2$ for sufficiently large T if

$$R < \log\det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M})^{1/2T} = \log\rho. \tag{2.11}$$

Since the capacity of the AWGN channel is $\log(1+\rho)$, lattice coding and decoding as defined by deBuda, Polytrev and Loeliger achieve near capacity only for high SNR. The loss in "one" in the rate formula results in significant degradation in performance at low SNR and zero rate for $\rho < 1$.

Capacity-achieving lattice codes under lattice decoding has been made possible after the work of Erez and Zamir [24]. In their work, the power constrained AWGN channel is first transformed into a modulo-lattice additive noise channel (known as the MLAN-channel). The input alphabet of this channel is the Voronoi region ($\mathcal{V}_s$) of the shaping lattice Λ_s, i.e., $\boldsymbol{c} \in \mathcal{C} = \Lambda_c \cap \mathcal{V}_s$ added to a dither signal $\boldsymbol{u} \sim \mathcal{U}(\mathcal{V}_s)$. The dither signal assures that the power constrained of the channel is satisfied all the time. Another role of it is to de-correlate the estimation error from the channel input. The output of the transmitter is given by the modulo lattice operator as

$$\boldsymbol{x} = [\boldsymbol{c} - \boldsymbol{u}] \bmod \Lambda_s. \tag{2.12}$$

The received signal, $\boldsymbol{x} + \boldsymbol{e}$, is then multiplied by the MMSE factor $\alpha = P_X/(P_X + \sigma^2)$, where P_X is the average transmitted power, and the dither signal is then added. The result is reduced modulo-Λ_s, i.e.,

$$\begin{aligned} \boldsymbol{y} &= [\alpha(\boldsymbol{x} + \boldsymbol{e}) + \boldsymbol{u}] \mod \Lambda_s \\ &= [\boldsymbol{c} + \boldsymbol{e}'] \mod \Lambda_s, \end{aligned} \tag{2.13}$$

where $\boldsymbol{e}' = [\alpha\boldsymbol{e} + (1-\alpha)\boldsymbol{u}] \mod \Lambda_s$.

The analysis of Erez and Zamir revealed that using lattice coding and MMSE lattice decoding, reliable communication is possible as long as

$$R < \log(1+\rho) = C.$$

For the outage-limited $M \times N$ MIMO channel:

with channel gain matrix $\boldsymbol{H}^c$, the achievable rate follows by letting $\boldsymbol{M} = \sqrt{\rho}\boldsymbol{H}$, where

$$\boldsymbol{H} = \sqrt{\rho}\boldsymbol{I}_T \otimes \begin{pmatrix} \Re\{\boldsymbol{H}^c\} & -\Im\{\boldsymbol{H}^c\} \\ \Im\{\boldsymbol{H}^c\} & \Re\{\boldsymbol{H}^c\} \end{pmatrix}, \tag{2.14}$$

and $r_e^2 = MT$ in the Lemma 2. In that case, from the standard typicality arguments it follows that for any $\epsilon > 0$ and $\gamma > 0$, there exists $T_{\gamma,\epsilon}$ such that for all $T > T_{\gamma,\epsilon}$ we have that $\Pr(\boldsymbol{e} \notin \mathcal{E}_{T,\gamma}) < \epsilon/2$. The second term in the upper bound (5.59) can be made smaller than $\epsilon/2$ for sufficiently large T if

$$R < \log\det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M})^{1/2T} = \log(\rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c). \tag{2.15}$$

For such a channel, the lattice decoder that is given in (2.8) is usually referred to in literature as *naive* lattice decoding. The loss in the identity matrix $\boldsymbol{I}_M$ in the rate formula results in significant degradation in performance not just at low SNR, but at high SNR as well, due to the channel being in deep fading or near outage (more about this topic will be discussed in the subsequent chapters). It has been shown in [23] that such decoder cannot achieve the optimal diversity-multiplexing tradeoff[2] of the MIMO channel for any values of M, N, and $T \geq N + M - 1$.

It has been shown in [23] that a minimum-mean square error (MMSE) preprocessing can dramatically improve the performance of the lattice decoding in MIMO systems. The use of the so called MMSE decision feedback equalization (MMSE-DFE) prior lattice decoding led to achieving the optimal diversity-multiplexing tradeoff of the MIMO channel. The operation of the MMSE-DFE at the receiver followed by lattice decoding is summarized in the following (see Appendix I in [23] for more details).

MMSE-DFE Pre-Processing

First, we perform the QR-decomposition on the *augmented* channel matrix

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} \boldsymbol{H} \\ \boldsymbol{I} \end{pmatrix} = \tilde{\boldsymbol{Q}}\boldsymbol{R},$$

[2]The naive lattice decoder achieves the optimal tradeoff $d^*(r) = (M-r)(N-r)$ and $r \in [0, \min\{M, N\})$, only for the special case of $T = 1$ and squared channel matrix $N = M$, or for any $N \geq M$, $T \geq N - M + 1$, in the high rate segment $r \in [M-1, M)$

where $\boldsymbol{H}$ is given in (2.14), $\tilde{\boldsymbol{Q}} \in \mathbb{R}^{(n+m)\times m}$ has orthonormal columns, and $\boldsymbol{R} \in \mathbb{R}^{m\times m}$ is an upper triangular with positive diagonal elements. If we let $\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{R}^{-1}$ the upper $n \times m$ part of $\boldsymbol{Q}$, then the matrices $\boldsymbol{F} = \boldsymbol{Q}^{\mathsf{T}}$ and $\boldsymbol{B} = \boldsymbol{R}$ are called the MMSE-DFE *forward* and *backward* filters, respectively.

At the receiver, the received signal, $\boldsymbol{y}$, is multiplied by the forward filter matrix $\boldsymbol{F}$ of the MMSE-DFE. Moreover, we add the dither signal filtered by the upper triangular feedback filter matrix $\boldsymbol{B}$ of the MMSE-DFE. This is a nontrivial generalization of the Erez-Zamir lattice coding scheme for the AWGN channel, i.e., a generalization of the MLAN-channel, which is introduced next.

2.4 The mode-Λ Scheme

We define nested lattice codes for the quasi-static MIMO channel. We say that a LAST code is nested if the underlying lattice code is nested. Here, the information message is effectively encoded into the cosets Λ_s in Λ_c. As defined in [23], we shall call such codes the mod-Λ scheme. The proposed mod-Λ scheme works as follows:

Consider the nested LAST code $\mathcal{C}$ defined by Λ_c (the coding lattice) and by its sublattice Λ_s (the shaping lattice) in $\mathbb{R}^m$. Assume that Λ_s has a second-order moment $\sigma^2(\Lambda_s) = 1/2$ (so that $\boldsymbol{u}$ uniformly distributed over $\mathcal{V}_s$ satisfies $\mathsf{E}\{|\boldsymbol{u}|^2\} = MT$). The transmitter selects a codeword $\boldsymbol{c} \in \mathcal{C}$, generates a dither signal $\boldsymbol{u}$ with uniform distribution over $\mathcal{V}_s$, and computes $\boldsymbol{x} = [\boldsymbol{c} - \boldsymbol{u}] \bmod \Lambda_s$. The signal $\boldsymbol{x}$ is then transmitted on the MIMO channel.

At the receiver, the received signal, $\boldsymbol{y}$, is multiplied by the forward filter matrix $\boldsymbol{F}$ of the MMSE-DFE. Moreover, we add the dither signal filtered by the upper triangular feedback filter matrix $\boldsymbol{B}$ of the MMSE-DFE.

By construction, we have $\boldsymbol{x} = \boldsymbol{c} - \boldsymbol{u} + \boldsymbol{\lambda}$ with $\lambda = -Q_{\Lambda_s}(\boldsymbol{c} - \boldsymbol{u})$. Then, we can write

$$\boldsymbol{y}' = \boldsymbol{F}\boldsymbol{y} + \boldsymbol{B}\boldsymbol{u} = \boldsymbol{B}\boldsymbol{c}' + \boldsymbol{e}', \tag{2.16}$$

where $\boldsymbol{c}' = (\boldsymbol{c}+\boldsymbol{\lambda})$ and $\boldsymbol{e}' = -[\boldsymbol{B} - \boldsymbol{F}\boldsymbol{H}]\boldsymbol{x} + \boldsymbol{F}\boldsymbol{e}$. Since $\boldsymbol{x}$ is uniformly distributed over $\mathcal{V}_s$ and is independent of $\boldsymbol{c}$, it can be shown [23] that $\mathsf{E}\{\boldsymbol{e}'\boldsymbol{e}'^{\mathsf{T}}\} = 1/2\boldsymbol{I}_m$ and if the shaping lattice Λ_s is good for MSE quantization, then $\boldsymbol{e}' \to \mathcal{N}(\boldsymbol{0}, 1/2\boldsymbol{I}_m)$ as $T \to \infty$. The desired signal $\boldsymbol{c}$ is now translated by an unknown lattice point $\lambda \in \Lambda_s$. However, since $\boldsymbol{c}$ and $\boldsymbol{c} + \boldsymbol{\lambda}$ belong to the same coset of Λ_s in Λ_c, this translation does not involve any loss of information. It follows that in order to recover the information message, the decoder must identify the coset $\Lambda_s + \boldsymbol{c}$ that contains $\boldsymbol{c} + \boldsymbol{\lambda}$. The decoder first finds

$$\hat{\boldsymbol{z}} = \arg \min_{\boldsymbol{z}\in\mathbb{Z}^{2MT}} \|\boldsymbol{y}' - \boldsymbol{B}\boldsymbol{G}\boldsymbol{z}\|^2, \tag{2.17}$$

then the decoded codeword is given by $\hat{\boldsymbol{c}} = [\boldsymbol{G}\hat{\boldsymbol{z}}] \bmod \Lambda_s$.

The optimality of LAST codes with the mod-Λ scheme is given by the following results (see [23] for more details):

- For a fixed, nonrandom channel matrix $\boldsymbol{H}^c$, the rate

$$R_{\text{mod}}(\boldsymbol{H}^c, \rho) = \log\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right), \tag{2.18}$$

  is achievable by LAST codes with lattice coding and MMSE-DFE lattice decoding (i.e., the mod-Λ scheme).

- There exists a sequence of nested LAST codes with block length $T \geq N + M - 1$ that achieves the optimal diversity-multiplexing tradeoff curve $d^*(r)$ for all $r \in [0, \min\{M, N\})$ under the mod-Λ scheme.

It is well-known that lattice decoding can be realized efficiently by sphere decoding, whose average complexity grows exponentially with m for any fixed SNR. This limits sphere decoding to low signal dimensions. For instance, to decode a 3×3 LAST code that achieves the optimal tradeoff, one has to search in a 30-dimensional lattice. The sphere decoding is slow for this dimension. Thus, we often have to resort to an approximate solution. In this work, the approximation is performed through the use of sequential decoding algorithms to search for the closest lattice point. These type of decoders can achieve near-optimal performance in many communication channels with lower (average) decoding complexity.

Chapter 3

Asymptotic Sphere Decoding Complexity Analysis for the LAST Coded MIMO Channel

SINCE its introduction to MIMO wireless communication systems, the sphere decoder has become an attractive efficient implementation of the ML decoder, especially for small signal dimensions and/or moderate to large SNRs. Such a decoder allows for significant reduction in decoding complexity as opposed to the ML decoder without sacrificing performance. In general, the sphere decoder is commonly used in communication systems that can be well-described by the *linear Gaussian vector channel* model that was introduced in Chapter 1.

It is well-known that sphere decoding based on Fincke-Pohst and Schnorr-Euchner enumerations are efficient strategies to perform lattice decoding and have been widely considered for signal detection in MIMO systems of small dimensions. The outage performance analysis of the lattice decoder, implemented via sphere decoding algorithms, was considered in the work by El-Gamal, Caire, and Damen in [23]. In their analysis, it has been shown that the optimal diversity-multiplexing tradeoff of the quasi-static MIMO channel can be achieved using LAST coding and lattice or sphere decoding. However, the computational complexity of the sphere decoder was not considered. In fact, the exact complexity analysis of the basic sphere decoder for general space-time codes applied to multiple-input multiple-output (MIMO) wireless channel is known to be difficult.

Previous work on the complexity of sphere decoding focused on characterizing the mean and the variance of the decoder's complexity, particularly for the *uncoded* MIMO channel (e.g., V-BLAST) [52]–[54]. Seethaler *et. al.* [55] considered the derivation of the computational distribution of the sphere decoder for the $M \times N$ uncoded MIMO channel. Characterizing and understanding the complexity distribution is important, especially when the sphere decoder is used under practically relevant runtime constraints. It has been shown in [55] that the computational tail distribution follows a Pareto-type with tail exponent given by $N - M + 1$. However, the main drawback of their work is that they consider the decoder's complexity analysis when the number of computations performed by the decoder increases without bound. In other words, although the behavior of the tail distribution is characterized, when the search becomes excessive they do not specify at when the decoder must terminate the search and declare an error. As a result, the *exact* average complexity of the sphere decoder when applied to the uncoded MIMO channel was not studied.

Achieving higher diversity and multiplexing gains require incorporating error control coding (across antenna and time) at the transmitter. Several works have considered the computational complexity analysis of optimal and sub-optimal decoders for the LAST coded MIMO channel [32], [39], [57]. A first step toward specifying the exact complexity required by the decoder to achieve the optimal diversity-multiplexing tradeoff (DMT) of the quasi-static LAST coded MIMO channel was considered in [57]. It was shown that the optimal tradeoff can be achieved using lattice reduction aided linear decoders at a worst-case complexity $O(\log \rho)$, where ρ is the average SNR. This corresponds to a linear increase in complexity as a function of the code rate at high SNR. However, this very low decoding complexity comes at the expense of a large performance gap from the sphere decoder's error performance. In order to close the gap between the sphere decoder and linear decoders, lattice sequential decoding algorithms [39] are considered efficient decoders that achieve near-optimal performance with much lower decoding complexity compared to sphere decoders. In [**?**], we have analyzed in details the decoder's computational tail distribution and the average decoding complexity. Specifically, we have shown that when the computational complexity exceeds a certain limit, the tail distribution becomes upper bounded by the outage probability achieved by LAST coding and sequential decoding schemes. As a result, one may save on decoding complexity while still achieving near-optimal performance by setting a *time-out* limit at the decoder so that when the computational complexity exceeds this limit the decoder terminates the search. Moreover, we have shown analytically how the minimum-mean square-error decision feed-back equalization (MMSE-DFE) can significantly improve the tail exponent and as a consequence reduces (average) computational complexity. However, it would be interesting to study the complexity behavior of the optimal sphere decoder in the quasi-static MIMO channel. This would allow us to compare the complexity of the sphere decoder with other low complexity decoders and see whether it is worth sacrificing performance for complexity when using such decoders (e.g., lattice sequential decoders [?]).

In this work, we shed the light on the computational complexity of sphere decoding for the quasi-static, LAttice Space-Time (LAST) coded MIMO channel. We extend the results in [23] to show that a tradeoff exists between the computational complexity of lattice decoding (implemented via sphere decoding) and the multiplexing gain. Specifically, we drive an upper bound of the tail distribution of the decoder's computational complexity. We show that, when the computational complexity exceeds a certain limit, this upper bound becomes dominated by the outage probability achieved by LAST coding and lattice (sphere) decoding schemes. We show analytically how minimum-mean square-error decision feedback equalization can significantly improve the tail exponent and as a consequence reduces computational complexity. In particular, we show that there exists a *cut-off* multiplexing gain for which the average computational complexity of the decoder remains bounded.

3.1 LAST Coding and Lattice Decoding

We consider a quasi-static, Rayleigh fading MIMO channel with M-transmit, N-receive antennas, and no channel state information (CSI) at the transmitter and perfect CSI at the receiver. The complex base-band model of the received signal can be mathematically described by (for T channel usages)

$$\boldsymbol{Y}^c = \sqrt{\rho}\boldsymbol{H}^c\boldsymbol{X}^c + \boldsymbol{W}^c, \tag{3.1}$$

where $\boldsymbol{X}^c \in \mathbb{C}^{M\times T}$ is the transmitted space-time code matrix, $\boldsymbol{Y}^c \in \mathbb{C}^{N\times T}$ is the received signal matrix, $\boldsymbol{W}^c \in \mathbb{C}^{N\times T}$ is the noise matrix, $\boldsymbol{H}^c \in \mathbb{C}^{N\times M}$ is the channel matrix, and $\rho = \mathsf{SNR}/M$ is the normalized SNR at each receive antenna with respect to M. The elements of both the noise matrix and the channel fading gain matrix are assumed to be i.i.d. zero mean circularly symmetric complex Gaussian random variables with variance $\sigma^2 = 1$.

An $M \times T$ space-time coding scheme is a full-dimensional LAttice Space-Time (LAST) code if its vectorized (real) codebook (corresponding to the channel model (2.5)) is a lattice code with dimension $m = 2MT$. As discussed in Chapter 2, Section 2.2, the design of space-time signals reduces to the construction of a codebook $\mathcal{C} \subseteq \mathbb{R}^{2MT}$ with code rate $R = \frac{1}{T}\log|\mathcal{C}|$, satisfying the input averaging power constraint

$$\frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{x}\in\mathcal{C}}\|\boldsymbol{x}\|^2 \leq MT.$$

Depending whether lattice decoding is pre-processed by MMSE-DFE filtering or not, the equivalent real model of the above channel can be easily shown to be given by (2.5) with $\boldsymbol{M}$ that satisfies

$$\det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M}) = \left[\det\left(\rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right]^{2T}, \tag{3.2}$$

for the case of *naive* lattice decoding, and

$$\det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M}) = \left[\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right]^{2T}, \tag{3.3}$$

for *MMSE-DFE* lattice decoding (see [23] for more details).

It has been shown in [23] that LAST coding and lattice decoding (for both naive and MMSE-DFE decoding) can achieve rates up to

$$R_{\text{LAST}}(\rho, \boldsymbol{H}^c) = \det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M})^{1/2T}. \tag{3.4}$$

As has been discussed in Chapter 1, the asymptotic error performance of any coding and decoding schemes in the outage-limited MIMO channel is dominated by the *outage probability*, $P_{\text{out}}(\rho, R)$, i.e., $P_e(\rho) \doteq P_{\text{out}}(\rho, R)$. For LAST coding and lattice decoding schemes, the outage probability is defined by

$$P_{\text{out}}(\rho, R) = \Pr(R \geq R_{\text{LAST}}(\rho, \boldsymbol{H}_c)) \doteq \rho^{-d_{\text{out}}(r)}, \tag{3.5}$$

where $d_{\text{out}}(r) \leq (M-r)(N-r) \triangleq d^*(r)$, $\forall\, r \in [0, \min\{M, N\})$, is defined as the diversity-multiplexing tradeoff achieved by such coding and decoding schemes [22], and $d^*(r)$ is the *optimal* diversity-multiplexing tradeoff of the channel.

Define the outage event $\mathcal{O}(\rho)$ as

$$\mathcal{O}(\rho) = \{\boldsymbol{H}^c : R(\rho) \geq R_{\text{LAST}}(\rho, \boldsymbol{H}^c)\},$$

and denote the transmission rate $R(\rho) = r\log\rho$. Let $0 \leq \lambda_1 \leq \cdots \leq \lambda_M$ be the ordered eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c$, and define $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_M)$, $\alpha_i \triangleq -\log\lambda_i/\log\rho$. As discussed in [22], at high SNR, the non-negative values of $\boldsymbol{\alpha}$ only contributes to the outage event. Therefore, the outage event can be expressed as

$$\mathcal{O} \doteq \left\{\boldsymbol{\alpha} \in \mathbb{R}_+^M : \sum_{i=1}^{M} \alpha_i > M - r\right\}, \tag{3.6}$$

for *naive* lattice decoding. For *MMSE-DFE* lattice decoding,

$$\mathcal{O} \doteq \left\{\boldsymbol{\alpha} \in \mathbb{R}_+^M : \sum_{i=1}^{M} (1-\alpha_i)^+ < r\right\}. \tag{3.7}$$

In what follows, we summarize the results derived in [23]. For the lattice decoding, there exists a sequence of full-dimensional LAST codes that achieves diversity-multiplexing tradeoff (assuming $N \geq M$)

$$d(r) = \min\{T, N - M + 1\}(M - r), \quad \forall r \in [0, M], \tag{3.8}$$

for any block length $T \geq 1$. If the decoder is pre-processed by MMSE-DFE filtering, then lattice decoding achieves the optimal diversity-multiplexing tradeoff of the channel:

$$d^*(r) = (M - r)(N - r), \quad \text{for all } r \in [0, \min\{M, N\}], \tag{3.9}$$

under the constraint $T \geq M + N - 1$ (see [23] for more details).

Next, we show that the diversity-multiplexing tradeoff can be naturally extended to include the decoding complexity. In other words, we derive the *diversity-multiplexing-complexity* tradeoff of the MIMO channel under sphere decoding.

3.2 Lattice Decoding via Sphere Decoding

While ML decoding performs exhaustive search over all codewords $\boldsymbol{c} \in \mathcal{C}(\Lambda_c, \mathcal{R})$, sphere decoding algorithms find the closest lattice point $\boldsymbol{x} \in \Lambda_c$ to the received signal $\boldsymbol{y}$ within a sphere radius R_s centered at the received signal (see Figure 3.1).

It is well-known that the sphere decoder allows for significant reduction in decoding complexity for small dimensions and average to large values of SNR. Depending whether the sphere decoding incorporate the boundaries of the lattice code (i.e., $\mathcal{R}$) into the search algorithm or not, one can achieve ML or *near*-ML performance. Here, we consider sphere decoding algorithms that describe lattice decoding, i.e., the class of decoding algorithms that do not take into account the shaping region $\mathcal{R}$.

In general, the sphere decoder, after QR decomposition of the channel-code matrix $\boldsymbol{MG}$, finds all integer lattice points $\boldsymbol{z} \in \mathbb{Z}^m$ that satisfy the sphere constraint

$$\|\boldsymbol{y}' - \boldsymbol{Rz}\|^2 \leq R_s^2, \tag{3.10}$$

where $\boldsymbol{y}' = \boldsymbol{Q}^\mathsf{T}\boldsymbol{y}$, $\boldsymbol{Q}$ is an orthogonal matrix that corresponds to the QR decomposition of the channel-code matrix $\boldsymbol{MG} = \boldsymbol{QR}$, and $\boldsymbol{R}$ is an $m \times m$ upper triangular matrix with positive diagonal elements that is given by

$$\boldsymbol{R} = \begin{pmatrix} R_{m,m} & R_{m,m-1} & \cdots & R_{m,m} \\ 0 & R_{m-1,m-1} & \cdots & R_{m-1,m} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & R_{1,1} \end{pmatrix}. \tag{3.11}$$

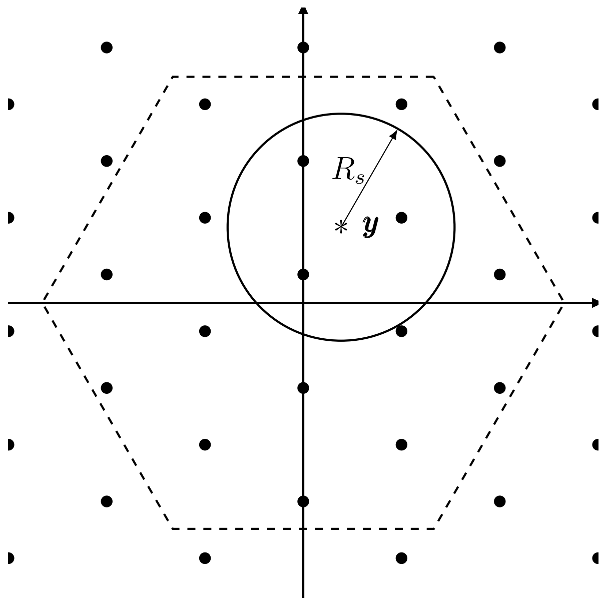


Figure 3.1: The operation of the sphere decoder. The sphere decoder searches for the closest lattice point to $\boldsymbol{y}$ among the points that are *only* inside the sphere (3 points). However, the ML decoder has to search over the 16 points inside the shaping region.

It is more convenient to look at the sphere decoder as a search in a *tree* with m layers. The k-th layer, where $1 \leq k \leq m$, contains nodes that correspond to the partial integer lattice points $\boldsymbol{z}_1^k \in \mathbb{Z}^k$ (the last k components of the integer vector $\boldsymbol{z}$). In this case, nodes $(\boldsymbol{z}_1^k)$ that satisfy the following constraint

$$\|\boldsymbol{y'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq R_s^2,$$

are allowed to be visited by the decoder, where $\boldsymbol{R}_{kk}$ is the lower $k \times k$ part of the matrix $\boldsymbol{R}$, $\boldsymbol{y'}_1^k$ is the last k components of the vector $\boldsymbol{y'}$. The structure of $\boldsymbol{R}$ allows one to perform a backward sequential search from layer (dimension) 1 (corresponds to the last vector coordinate) to layer m (corresponds to the first vector coordinate). Several algorithms were developed to efficiently perform the search (cf. [32]). Once all points are listed, one can find the point that is closest in distance to $\boldsymbol{y}$.

In this case, one may define the computational complexity of the sphere decoder as the total number of nodes that have been visited (or extended) by the decoder during the search. Define the indicator function $\phi(\boldsymbol{z}_1^k)$ by

$$\phi(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \boldsymbol{z}_1^k \text{ is extended;} \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

Then, the total number of partial integer lattice points $\boldsymbol{z}_1^k \in \mathbb{Z}^k$ found by the decoder at layer k can be expressed as

$$C_k = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k). \tag{3.13}$$

In this case, the total computational complexity of the sphere decoder C that is required to find the closest lattice point to the received signal is given by $C = \sum_{k=1}^{m} C_k$.

3.2.1 Sphere Radius Selection

The selection of the initial radius R_s at the beginning of the search is of crucial importance in the computational complexity analysis. Choosing a very small value R_s may result in finding no lattice points inside the sphere (i.e., $C_k = 0$ for some $1 \leq k \leq m$). On the other hand, choosing a very large value of R_s results in finding too many lattice points inside the sphere that leads to very large computational complexity. As such, the sphere radius R_s must be chosen sufficiently large for the search sphere to contain at least one lattice point.

Selecting $R_s = r_{\text{cov}}(\boldsymbol{MG})$, i.e., the covering radius[1] of the lattice generated by $\boldsymbol{MG}$, guarantees the existing of at least one lattice point inside the sphere. Unfortunately, the computation of r_{cov} for a general lattice is very difficult. Although, an upper bound of r_{cov} can be obtained, $r_{\text{cov}} \leq \frac{1}{2}\sum_{i=1}^{m}[\boldsymbol{R}]_{ii}$, it is random. This requires finding its distribution (which is difficult in general) and may not lead to finding the complexity distribution of the decoder. Another choice of R_s is the distance between the Babai estimate and the vector $\boldsymbol{y}$. As mentioned in [52], although this choice guarantees the existence of at least one lattice point (the Babai estimate) inside the sphere, it not clear in general whether it leads to too many lattice points inside the sphere.

In this work, we follow a different approach to find a fixed sphere radius that guarantees the existing of at least one lattice point inside the sphere (making the selection totally independent of the lattice and channel statistics). This particular choice of sphere radius is shown to simplify the analysis of deriving an upper bound on the decoder's computational

[1] The covering radius $r_{\text{cov}}(\boldsymbol{G})$ of a lattice $\Lambda(\boldsymbol{G})$ is the radius of the smallest sphere centered at the origin that contains $\mathcal{V}_{\boldsymbol{0}}(\boldsymbol{G})$.

complexity. The basic idea of this approach (as will be shown next) is to separate the typical noise events from the non-typical ones. This allows the separation of the "typical" lattice points (lattice points that are highly likely to be generated by the sphere decoder) from the atypical ones.

3.2.2 The k-th Layer Complexity

In this section, we would like to provide some insight about the computational complexity of the sphere decoder at the k-th layer. This may assist us in the derivation of an upper bound on the computational complexity distribution as will be shown in the sequel.

As mentioned previously, the computational complexity of the sphere decoder at the k-th layer is determined by the total number of partial lattice points $\boldsymbol{z}_1^k \in \mathbb{Z}^k$ that satisfy the k-th layer sphere constraint

$$\|\boldsymbol{y}'^k_1 - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\| \leq R_s.$$

We assume that R_s is chosen sufficiently large enough so that at least one lattice point is found inside the sphere (details on how R_s is selected will be introduced next). It is clear that the computational complexity of the decoder depends on the distributions of $\boldsymbol{y}_1^k$ and $\boldsymbol{R}_{kk}$. Since those two quantities are random, the computational complexity analysis of the sphere decoder is considered difficult.

A first step toward establishing the upper bound for the total computational complexity, i.e., $C = \sum_{k=1}^{m} C_k$, is using a well-known bound on C_k (see [22]) which is given by

$$C_k \leq \frac{V(\mathcal{S}_k(R_s + r_{\text{cov}}(\boldsymbol{R}_{kk})))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}. \tag{3.14}$$

where $r_{\text{cov}}(\boldsymbol{R}_{kk})$ is the covering radius of the lattice generated by the partial matrix $\boldsymbol{R}_{kk}$. However, as mentioned previously, finding the exact value of r_{cov} is very difficult in general. Therefore, most of the work on sphere decoding complexity (see [44], [55], and [56]), rely on approximating C_k by

$$C_k \approx \frac{V(\mathcal{S}_k(R_s))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}, \tag{3.15}$$

If $\boldsymbol{y}_1^k$ is assumed to be uniformly distributed over $\mathcal{V}_{\mathbf{0}}(\boldsymbol{R}_{kk})$ (which is not the case here) the above approximation becomes exact if averaging C_k is performed over $\boldsymbol{y}_1^k$. However, it is not yet clear how close this approximation is to the exact value for any $\boldsymbol{y}_1^k$. To overcome these problems, by bounding the noise from above, we establish an upper bound on C_k that is independent of $r_{\text{cov}}(\boldsymbol{R}_{kk})$ and $\boldsymbol{y}_1^k$, as shown in the following lemma:

Lemma 3. *The k-th layer complexity C_k of the sphere decoder with radius R_s, when the magnitude of the noise $|\boldsymbol{e}| \leq R_s$, can be upper bounded by*

$$C_k \leq \frac{V(\mathcal{S}_k(\sqrt{7}R_s))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}} = C_k'. \tag{3.16}$$

Proof. See Appendix A. □

It should not be so surprising that the k-th layer complexity of the sphere decoder is inversely proportional to the volume of the Voronoi region of the lattice generated by the partial upper triangular matrix $\boldsymbol{R}_{kk}$. Since $\boldsymbol{R}_{kk}$ is related to the channel matrix $\boldsymbol{H}^c$, it is to be expected that the computational complexity depends critically on the channel conditions, i.e., depends on whether the channel is *ill* or *well* conditioned. We are now ready to establish our upper bound on the decoder's complexity.

3.3 Computational Complexity: Tail Distribution in the High SNR Regime

In this section, we consider a *fixed* sphere radius $R_s^2 = MT(1 + \zeta \log \rho)$, where $\zeta > 0$. The reason for that choice will become evident as we further analyze the complexity of the decoder. We are interested in finding an upper bound to the tail distribution of the decoder's computational complexity at high SNR. The main result is summarized in the following theorem:

Theorem 2. ***The asymptotic computational complexity distribution of the sphere decoder in an*** $M \times N$ ***LAST coded MIMO channel with codeword length*** T***, is upper bounded by***

$$\Pr(C \geq L) \dot{\leq} \rho^{-\eta(r)}, \tag{3.17}$$

under the condition that

$$L \geq m + V(\mathcal{S}_m(2R_s)) \sum_{k=1}^{m} \frac{V(\mathcal{S}_k(\sqrt{7}R_s))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}, \tag{3.18}$$

where the SNR exponent $\eta(r) = \min\{T, N - M + 1\}(M - r)$ ***for naive decoding and*** $T \geq 1$***, and*** $\eta(r) = (M - r)(N - r)$ ***for MMSE-DFE decoding and*** $T \geq N + M - 1$***. The matrix*** $\boldsymbol{R}_{kk}$ ***is the lower*** $k \times k$ ***part of*** $\boldsymbol{R} = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{MG}$***.***

Proof. We follow the approach that is commonly used to derive an upper bound for the decoding error probability in quasi-static MIMO channel (see (1.11)). By separating the outage event from the non-outage event, we obtain:

$$\Pr(C \geq L) \leq \Pr(\boldsymbol{\alpha} \in \mathcal{O}) + \Pr(C \geq L, \boldsymbol{\alpha} \in \overline{\mathcal{O}}). \tag{3.19}$$

Let us concentrate on bounding the second term in the RHS of (3.19). As mentioned in section 3.1.2, bounding this term is considered difficult in general. However, as will be shown in the sequel, the analysis can be simplified by bounding the noise vector $\boldsymbol{e}$. In this case, one can upper bound the second term in the RHS of (3.19) as follows:

$$\Pr(C \geq L|\overline{\mathcal{O}}) \leq \Pr(C \geq L, \|\boldsymbol{e}\|^2 \leq R_s^2|\overline{\mathcal{O}}) + \Pr(\|\boldsymbol{e}\|^2 \geq R_s^2). \tag{3.20}$$

The problem now is to find R_s such that the above upper bound is tight. In order to do so, we consider the behavior of some of the parameters that corresponds to the channel-code lattice $\Lambda(\boldsymbol{MG})$ when the channel is not in outage, which may lead to finding the appropriate R_s.

First, there exists a shifted lattice code $\Lambda_c + \boldsymbol{u}_0^*$ with (see Chapter 2, Lemma 1)

$$|\mathcal{C}(\Lambda_c, \boldsymbol{u}_0^*, \mathcal{R})| = 2^{RT} = \rho^{rT} \geq \frac{V(\mathcal{R})}{V_c}.$$

When the channel is not in outage, one can verify that the asymptotic effective radius of the channel-code matrix, $r_{\text{eff}}(\boldsymbol{MG})$, can be lower bounded by

$$\begin{aligned} r_{\text{eff}}(\boldsymbol{MG}) &= \left[\frac{V_c \det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M})^{1/2}}{V(\mathcal{S}_{\mathbf{0}}^{m}(1))}\right]^{1/m} \\ &\dot{\geq} MT\left[\rho^{-rT}\det(\boldsymbol{M}^{\mathsf{T}}\boldsymbol{M})^{1/2}\right]^{1/m} \\ &\doteq MT\rho^{\gamma}, \end{aligned} \tag{3.21}$$

where $\gamma = [\nu(\boldsymbol{\alpha}) - r]/2M > 0$, when the channel is not in outage, and $\nu(\boldsymbol{\alpha}) = M - \sum_{j=1}^{M}\alpha_j$ or $\nu(\boldsymbol{\alpha}) = \sum_{j=1}^{M}(1-\alpha_j)^{+}$ for the naive or the MMSE-DFE lattice decoding, respectively.

It is clear from (3.21) that, when the channel is not in outage, as $\rho \to \infty$, the volume of the Voronoi region $\mathcal{V}_{\mathbf{0}}(\boldsymbol{MG})$ (corresponds to the channel-code matrix) as well as $r_{\text{eff}}(\boldsymbol{MG})$ grow quickly with SNR as ρ^{κ}, where $\kappa > 0$. According to this, the decoder's sphere radius is required to increase with SNR as well in order to ensure the existing of at least one lattice point inside the decoder's search sphere.

However, choosing $R_s = r_{\text{eff}}(\boldsymbol{MG}) \doteq \rho^{\gamma}$ results into too many points inside the sphere and may not lead to a tight upper bound. Therefore, R_s is required to grow with SNR at slower rate than ρ^{κ}. For that reason, we chose the search radius to be $R_s^2 = MT(1+\zeta\log\rho)$, where $\zeta > 0$ (asymptotically less than ρ^{κ}, for all $\zeta > 0$) and show that for sufficiently large ζ, such fixed radius sphere decoder guarantees (with high probability) the existing of at least one lattice point inside the sphere. Interestingly, such choice of R_s makes it totally independent of the lattice and the channel statistics.

Now, suppose that spheres of squared radius $R_s^2 = MT(1+\zeta\log\rho)$, $\mathcal{S}_m(R_s)$, are placed around each lattice point $\boldsymbol{x}$ that belongs to the (infinite) channel-code lattice (see Figure 3.2). There is still a non-zero probability that no lattice point will be found inside the decoder's sphere as depicted in Figure 3.2.

This may happen when $\boldsymbol{e} \in \mathcal{V}_{\mathbf{0}}(\boldsymbol{MG})\backslash\mathcal{S}_m(R_s)$. This event occurs with probability

$$\begin{aligned} \Pr(\text{no lattice point}) &\leq \Pr(\boldsymbol{e} \notin \mathcal{S}_m(R_s)) \\ &= \Pr(\|\boldsymbol{e}\|^2 > MT(1+\zeta\log\rho)) \\ &\leq \rho^{-MT\zeta}, \end{aligned} \tag{3.22}$$

where the last inequality follows from applying Chernoff bound. For sufficiently large ζ, the above probability becomes negligible. In other words, asymptotically, one can expect that the received signal is highly likely to be located inside a sphere of square radius $R_s^2 = MT(1+\zeta\log\rho)$. Therefore, we may neglect the output of the search (or declare an

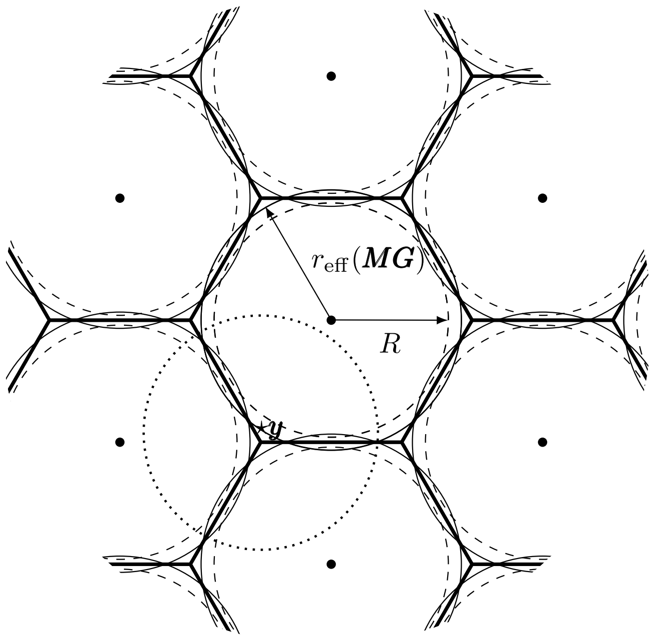


Figure 3.2: A geometrical approach for upper bounding the complexity distribution. Spheres of radius R_s centered at the lattice points $\boldsymbol{x} \in \Lambda(\boldsymbol{MG})$ are presented in dashed lines. The doted line represents the decoder's search sphere centered at the received signal $\boldsymbol{y}$ of radius R_s.

error) if the received signal is located outside $\mathcal{S}_m(R_s)$. It turns out that this modification on sphere decoding does not affect the asymptotic performance achieved by such decoding scheme.

Next, consider bounding the first term in the RHS of (3.20) from above. By viewing the decoder as a search on the tree one can interpret C as the total number of nodes in the tree visited by the decoder. Therefore, assuming the received vector $\boldsymbol{y} \in \mathcal{S}_m(R_s)$, one can rewrite C as $C = m + \tilde{C}$, where

$$\tilde{C} = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k \backslash \{\boldsymbol{0}\}} \phi(\boldsymbol{z}_1^k).$$

Now, let $\tilde{\phi}_k(\boldsymbol{z})$ be the indicator function defined by

$$\tilde{\phi}_k(\boldsymbol{x}) = \begin{cases} C'_k, & \text{if } \|\boldsymbol{e} - \boldsymbol{M}\boldsymbol{x}\|^2 \leq R_s^2; \\ 0, & \text{otherwise,} \end{cases}$$

where C'_k is as defined in Lemma 3. Then, one can easily verify that

$$\tilde{C} \leq \sum_{k=1}^{m} C'_k \sum_{\boldsymbol{x} \in \Lambda_c^*} \phi_k(\boldsymbol{x}),$$

where $\Lambda_c^* = \Lambda_c \backslash \{\boldsymbol{0}\}$. For a given lattice Λ_c, using Markov inequality, we have

$$\Pr(C \geq L|\Lambda_c) = \Pr(\tilde{C} \geq L - m|\Lambda_c) \leq \frac{\mathsf{E}_{\boldsymbol{e}'}\{\tilde{C}|\Lambda_c\}}{L - m}, \tag{3.23}$$

for $L > m$. Taking the expectation of $\tilde{C}$ with respect to the noise, one can easily show that[2]

$$\begin{aligned} &\Pr(C \geq L, \|\boldsymbol{e}\|^2 \leq R_s^2|\Lambda_c, \overline{\mathcal{O}}) \\ &\quad \leq \frac{\sum_{k=1}^{m} C'_k}{L - m} \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{e} - \boldsymbol{M}\boldsymbol{x}\|^2 \leq R_s^2, \|\boldsymbol{e}\|^2 \leq R_s^2|\overline{\mathcal{O}}) \\ &\quad \stackrel{(a)}{\leq} \frac{\sum_{k=1}^{m} C'_k}{L - m} \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{M}\boldsymbol{x}\|^2 \leq 4R_s^2|\overline{\mathcal{O}}) \\ &\quad = \frac{\sum_{k=1}^{m} C'_k}{L - m} \mathsf{E}_{\boldsymbol{M}} \left\{ \sum_{\boldsymbol{x} \in \Lambda_c^*} \boldsymbol{1}\{\|\boldsymbol{M}\boldsymbol{x}\|^2 \leq 4R_s^2\} \middle| \overline{\mathcal{O}} \right\}. \end{aligned} \tag{3.24}$$

where (a) follows from the fact that in general one can show that for any random vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, and $R_s > 0$, it holds $\{\|\boldsymbol{u} - \boldsymbol{v}\|^2 \leq R_s^2, \|\boldsymbol{v}\|^2 \leq R_s^2\} \subseteq \{\|\boldsymbol{v}\|^2 \leq 4R_s^2\}$, and $\boldsymbol{1}\{\mathcal{A}\}$ denotes the indicator function of the event $\mathcal{A}$. By taking the expectation of (3.24) over the ensemble of random lattices (see [8], Theorem 4)

$$\begin{aligned} \Pr(C \geq L, &\|\boldsymbol{e}\|^2 \leq R_s^2|\overline{\mathcal{O}}) \\ &\leq \frac{\sum_{k=1}^{m} C'_k}{L - m} \mathsf{E}_{\boldsymbol{M}} \left\{ \frac{V(\mathcal{S}_m(2R_s))}{V_c \det(\boldsymbol{M}^\mathsf{T}\boldsymbol{M})^{1/2}} \middle| \overline{\mathcal{O}} \right\} \\ &= \mathsf{E}_{\boldsymbol{M}} \left\{ \rho^{-T[\nu(\boldsymbol{\alpha}) - r]} |\overline{\mathcal{O}} \right\} \end{aligned} \tag{3.25}$$

[2] At this point, we would like to remind the reader that for the case of MMSE-DFE lattice decoding, the additive noise vector is non-Gaussian for finite T. However, one can show (see Chapter 2, Section 2.1) that for a well-constructed lattice the probability density function of the noise vector $\boldsymbol{e}$, $f_{\boldsymbol{e}}(\boldsymbol{\nu}) \leq \beta_m f_{\tilde{\boldsymbol{e}}}(\boldsymbol{\nu})$, where $\tilde{\boldsymbol{e}} \sim \mathcal{N}(\boldsymbol{0}, 0.5\boldsymbol{I})$, and β_m is a constant (has no effect at high SNR).

for $L \geq m + V(\mathcal{S}_m(2R_s)) \sum_{k=1}^{m} C'_k$, where $\nu(\boldsymbol{\alpha}) = M - \sum_{j=1}^{M} \alpha_j$ for naive decoding and $\nu(\boldsymbol{\alpha}) = \sum_{j=1}^{M} (1-\alpha_j)^+$ for MMSE-DFE decoding. It is interesting to note that the above upper bound is equivalent to the upper bound derived for the error performance of lattice decoding [23].

Averaging (3.25) over the channels in $\overline{\mathcal{O}}$ set,

$$\begin{aligned} &\Pr(C \geq L, \|\boldsymbol{e}\|^2 \leq R_s^2) \\ &\qquad \dot{\leq} \int_{\overline{\mathcal{O}}} f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \Pr(C \geq L, \|\boldsymbol{e}\|^2 \leq R_s^2 | \boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\ &\qquad \dot{\leq} \rho^{-d_{\text{out}}(r)}, \end{aligned} \tag{3.26}$$

where $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ is the joint probability density function of $\boldsymbol{\alpha}$ which, for all $\boldsymbol{\alpha} \in \overline{\mathcal{O}}$, is asymptotically given by [23]

$$f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \doteq \exp(-\log(\rho) \sum_{i=1}^{M} (2i - 1 + |N - M|)\alpha_i).$$

and $d_{\text{out}}(r)$ is the outage SNR exponent that is given in (3.8) or (3.9) depending on the coding and decoding schemes used in the system.

The behavior of the first term in the RHS of (3.19) at high SNR is also $\rho^{-d_{\text{out}}(r)}$. Therefore, we finally have

$$\Pr(C \geq L) \, \dot{\leq} \, \rho^{-d_{\text{out}}(r)}. \tag{3.27}$$

under the condition that

$$L \geq m + V(\mathcal{S}_m(2R_s)) \sum_{k=1}^{m} \frac{V(\mathcal{S}_k(\sqrt{7}R_s))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}} \boldsymbol{R}_{kk})^{1/2}}.$$

□

The above results reveal that if the number of computations performed by the decoder exceeds

$$L_0 = m + V(\mathcal{S}_m(2R_s)) \sum_{k=1}^{m} \frac{V(\mathcal{S}_k(\sqrt{7}R_s))}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}} \boldsymbol{R}_{kk})^{1/2}}, \tag{3.28}$$

the complexity distribution of the sphere decoder at *any* SNR is upper bounded by the lattice decoding error probability (at high SNR the bound becomes equivalent to the asymptotic outage probability). As a result, one may save on decoding complexity while still achieving near-ML performance by setting a *time-out* limit at the decoder so that when the computational complexity exceeds L_0 the decoder terminates the search. It is clear that this time-out limit does not affect the optimal tradeoff achieved by the modified decoding scheme. To see this, suppose that the sphere decoder imposes a time-out limit so that the search is terminated once the number of computations reaches L_0, and hence the decoder declares an error. Let E_s be the event that the decoder makes an erroneous detection when $L \leq L_0$ (this event occurs when the received signal $\boldsymbol{y} \in \mathcal{V}_{\boldsymbol{x}}(\boldsymbol{MG})$, assuming $\boldsymbol{x}$ was transmitted). In this case, the average error probability is given by

$$P_e(\rho) = \Pr(E_s \cup \{C \geq L_0\}) \leq \Pr(E_s) + \Pr(C \geq L_0) \dot{\leq} \rho^{-d_{\text{out}}(r)}. \tag{3.29}$$

However, since L_0 is random, it would be interesting to calculate the minimum average number of computations required by the decoder to terminate the search.

3.4 Average Sphere Decoding Complexity

It is to be expected that when the channel is ill-conditioned (i.e., in outage) the computational complexity becomes extremely large. Moreover, when the channel is in outage it is highly likely that the decoder performs an erroneous detection. Unfortunately, when the channel is *not* in outage, there is still a non-zero probability that the number of computations will become large (see (3.25)). As such, it is sometimes desirable to terminate the search even when the channel is not in outage, especially when the sphere decoder is used under practically relevant runtime constraints. Therefore, we would like to determine the *minimum* average number of computations that is required in order for the decoder to decide when to terminate the search without affecting the achievability of the optimal tradeoff.

This can be expressed as

$$L_{\text{out}} = \mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\}, \tag{3.30}$$

where $L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})$ denotes the minimum number of computations performed by the decoder to achieve near-ML performance when the channel is not in outage which is given in (3.28).

Since it is very difficult to evaluate L_{out} for any SNR, we would like first to consider the asymptotic (at high SNR) behavior of L_0. As mentioned in Chapter 2, we focus our analysis on nested LAST codes, specifically LAST codes that are generated using construction A.

We consider the Loeliger ensemble of mod-p lattices, where p is a prime. First, we generate the set of all lattices given by

$$\Lambda_p = \kappa(\mathsf{C} + p\mathbb{Z}^{2MT})$$

where $p \to \infty$, $\kappa \to 0$ is a scaling coefficient chosen such that the fundamental volume $V_f = \kappa^{2MT} p^{2MT-1} = 1$, $\mathbb{Z}_p$ denotes the field of mod-p integers, and $\mathsf{C} \subset \mathbb{Z}_p^{2MT}$ is a linear code over $\mathbb{Z}_p$ with generator matrix in systematic form $[\boldsymbol{I}\ \boldsymbol{P}^{\mathsf{T}}]^{\mathsf{T}}$. We use a pair of self-similar lattices for nesting. We take the shaping lattice to be $\Lambda_s = \phi\Lambda_p$, where ϕ is chosen such that the covering radius is $1/2$ in order to satisfy the input power constraint. Finally, the coding lattice is obtained as $\Lambda_c = \rho^{-r/2M}\Lambda_s$ to satisfy the transmission rate constraint $R(\rho) = r\log\rho$. Interestingly, one can construct a generator matrix of Λ_p as (see [6])

$$\boldsymbol{G}_p = \kappa \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{P} & p\boldsymbol{I} \end{pmatrix}, \tag{3.31}$$

which has a lower triangular form. In this case, one can express the generator matrix of Λ_c as $\boldsymbol{G} = \rho^{-r/2M}\boldsymbol{G}'$, where $\boldsymbol{G}' = \phi\boldsymbol{G}_p$. Thanks to the lower triangular format of $\boldsymbol{G}$. If $\boldsymbol{M}$ is an $m \times m$ arbitrary full-rank matrix, and $\boldsymbol{G}$ is an $m \times m$ lower triangular matrix, then one can easily show that

$$\det[(\boldsymbol{MG})_{kk}] = \det(\boldsymbol{M}_{kk})\det(\boldsymbol{G}_{kk}), \tag{3.32}$$

where $(\boldsymbol{MG})_{kk}$, $\boldsymbol{M}_{kk}$, and $\boldsymbol{G}_{kk}$, are the lower $k \times k$ part of $\boldsymbol{MG}$, $\boldsymbol{M}$, and $\boldsymbol{G}$, respectively.

3.4.1 MMSE-DFE Sphere Decoding ($\boldsymbol{M} = \boldsymbol{B}$)

Using the above result, for the case of MMSE-DFE sphere decoding, one can express the determinant that appears in (3.28) as

$$\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) = \det(\boldsymbol{M}_{kk}^{\mathsf{T}}\boldsymbol{M}_{kk})\det(\boldsymbol{G}_{kk}^{\mathsf{T}}\boldsymbol{G}_{kk}) = \rho^{-rk/2M}\det(\boldsymbol{B}_{kk}^{\mathsf{T}}\boldsymbol{B}_{kk})\det({\boldsymbol{G}'}_{kk}^{\mathsf{T}}{\boldsymbol{G}'}_{kk}). \tag{3.33}$$

Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ be the ordered nonzero eigenvalues of $\boldsymbol{B}_{kk}^{\mathsf{T}}\boldsymbol{B}_{kk}$, for $k = 1, \cdots, m$. Then,

$$\det(\boldsymbol{B}_{kk}^{\mathsf{T}}\boldsymbol{B}_{kk}) = \prod_{j=1}^{k}\mu_j.$$

Note that for the special case when $k = m$ we have $\mu_{2(j-1)T+1} = \cdots = \mu_{2jT} = 1 + \rho\lambda_j((\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c)$, for all $j = 1, \cdots, M$.

Denote $\alpha_i' = -\log \mu_i / \log \rho$. Using (3.33), one can asymptotically express L_0 as

$$L_0 = m + (\log \rho)^{m/2} \sum_{k=1}^{m} (\log \rho)^{k/2} \rho^{c_k}, \tag{3.34}$$

where

$$c_k = \frac{1}{2} \sum_{j=1}^{k} \left(\frac{r}{M} - \alpha_j' \right)^+ . \tag{3.35}$$

Now, since c_k is non-decreasing in k, we have at high SNR

$$L_0 = m + (\log \rho)^m \rho^{c_m}, \tag{3.36}$$

where

$$c_m = T \sum_{i=1}^{M} \left(\frac{r}{M} - (1 - \alpha_i)^+ \right)^+ .$$

The average of L_0 at high SNR (averaged over the channel statistics) when the channel is not in outage is given by

$$\begin{aligned} \mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\} &= \int\limits_{\boldsymbol{\alpha} \in \overline{\mathcal{O}}} L_0 f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \\ &= m + (\log \rho)^m \int\limits_{\boldsymbol{\alpha} \in \overline{\mathcal{O}}} \exp\Bigg(\log \rho \Bigg[T \sum_{i=1}^{M} \left(\frac{r}{M} - (1 - \alpha_i)^+ \right)^+ - \\ &\qquad \sum_{i=1}^{M} (2i - 1 + N - M) \alpha_i \Bigg] \Bigg) \, d\boldsymbol{\alpha} \\ &= m + (\log \rho)^m \rho^{l_{\mathrm{MMSE-DFE}}(r)}, \end{aligned}$$

where $\overline{\mathcal{O}} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^M : \sum_{i=1}^{M} (1 - \alpha_i)^+ \geq r \right\}$, and

$$l_{\mathrm{MMSE-DFE}}(r) = \max_{\boldsymbol{\alpha} \in \overline{\mathcal{O}}} \left[T \sum_{i=1}^{M} \left(\frac{r}{M} - (1 - \alpha_i)^+ \right)^+ - \sum_{i=1}^{M} (2i - 1 + N - M) \alpha_i \right] . \tag{3.37}$$

It is not so difficult to see that the optimal channel coefficients that maximize (3.37) are

$$\alpha_i^* = 1, \quad \text{for } i = 1, \cdots, M - k,$$

and

$$\alpha_i^* = 0, \quad \text{for } i = M-k+1, \cdots, M,$$

i.e., the same $\boldsymbol{\alpha}^*$ that achieves the optimal DMT of the channel. Substituting $\boldsymbol{\alpha}^*$ in (3.37), we get

$$l_{\text{MMSE-DFE}}(r) = \frac{Tr(M-r)}{M} - (M-r)(N-r), \tag{3.38}$$

for $r = 0, 1, \cdots, M$. In this case, the asymptotic average computational complexity that is required by the decoder to achieve near-ML performance, when the channel is not in outage, can be expressed as

$$L_{\text{out}}^{\text{MMSE-DFE}} = 2MT + (\log\rho)^{2MT}\rho^{l_{\text{MMSE-DFE}}(r)}. \tag{3.39}$$

3.4.2 Naive Sphere Decoding ($\boldsymbol{M} = \boldsymbol{H}$)

Unfortunately, the equality in (3.32) does not apply for a general $M \times N$ MIMO channel under naive sphere decoding, and applies only to $\boldsymbol{M}$ being a square matrix, i.e., applies only to the case of MMSE-DFE sphere decoding where $\boldsymbol{M} = \boldsymbol{B}$ (the MMSE-DFE feedback matrix). For the case of naive sphere decoding, one may find a lower bound on $\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})$ which yields to an upper bound on the average computational complexity.

The interlacing theorem for bordered matrices (see [65], Theorem 4.3.8) implies that:

$$\lambda_i({\boldsymbol{R}_{kk}}^{\mathsf{T}}\boldsymbol{R}_{kk}) \geq \lambda_i(\boldsymbol{R}^{\mathsf{T}}\boldsymbol{R}), \quad \text{for } i = 1, \cdots, k.$$

Therefore, for the case of naive sphere decoding where $\boldsymbol{M} = \boldsymbol{H}$, we have

$$\begin{aligned}\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) = \prod_{j=1}^{k}\lambda_j(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) &\geq \prod_{j=1}^{k}\lambda_j(\boldsymbol{H}_{kk}^{\mathsf{T}}\boldsymbol{H}_{kk})\lambda_j(\boldsymbol{G}_{kk}^{\mathsf{T}}\boldsymbol{G}_{kk}) \\ &\geq \rho^{-rk/2M}\prod_{j=1}^{k}\lambda_j(\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H})\lambda_j({\boldsymbol{G}'}_{kk}^{\mathsf{T}}{\boldsymbol{G}'}_{kk}).\end{aligned} \tag{3.40}$$

Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ be the ordered nonzero eigenvalues of $\boldsymbol{H}_{kk}^{\mathsf{T}}\boldsymbol{H}_{kk}$, for $k = 1, \cdots, m$. Denote $\alpha_i' = -\log\mu_i/\log\rho$. Using (3.40), one can asymptotically upper bound L_0 as

$$L_0 \leq m + (\log\rho)^{m/2}\sum_{k=1}^{m}(\log\rho)^{k/2}\rho^{c_k}, \tag{3.41}$$

where

$$c_k = \frac{1}{2}\sum_{j=1}^{k}\left(\frac{r}{M} - \alpha'_j\right)^+ . \tag{3.42}$$

Now, since c_k is non-decreasing in k, we have at high SNR

$$L_0 \leq m + (\log\rho)^m \rho^{c_m}, \tag{3.43}$$

where

$$c_m = T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^+ .$$

In this case, the average of L_0 (averaged over channel statistics) when the channel is not in outage, can be upper bounded as

$$\begin{aligned}
\mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\} &= \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} L_0 f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \\
&= m + (\log\rho)^m \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} \exp\Bigg(\log\rho\Bigg[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^+ - \\
&\qquad\qquad \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\Bigg]\Bigg)\, d\boldsymbol{\alpha} \\
&= m + (\log\rho)^m \rho^{l_{\text{naive}}(r)},
\end{aligned}$$

where $\overline{\mathcal{O}} = \left\{\boldsymbol{\alpha} \in \mathbb{R}_+^M : \sum_{i=1}^{M}\alpha_i \leq M - r\right\}$, and

$$l_{\text{naive}}(r) = \max_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}}\left[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^+ - \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\right]. \tag{3.44}$$

Therefore, one can show that when the channel is not in outage we have that the optimal $\boldsymbol{\alpha}$ that maximizes (3.44) is achieved for $\alpha_1 = M - r$, and $\alpha_i = 0$ for all $i > 1$, yielding

$$l_{\text{naive}}(r) = \frac{T(M-1)}{M}(M-r) - (N-M+1)(M-r), \tag{3.45}$$

for $r = 0, 1, \cdots, M$. In this case, the asymptotic average computational complexity that is required by the naive decoder to upper bound the complexity tail distribution by its outage probability can be expressed as

$$L_{\text{out}}^{\text{naive}} \leq 2MT + (\log \rho)^{2MT} \rho^{l_{\text{naive}}(r)}. \tag{3.46}$$

To see the advantage of using the MMSE-DFE prior decoding that results in a huge saving in (average) computational complexity over the naive decoder, consider the case of a MIMO system with $M = N$. Assuming the use of an optimal random nested LAST code of codeword length T and a fixed rate R, i.e., $r = 0$. In this case, one can see that $l_{\text{MMSE-DFE}}(0) < 0$ irrespective to the value of T (i.e., the average complexity is bounded for all T). It is clear that the term $(\log \rho)^{2MT} \rho^{-NM}$ decays quickly to 0 as $\rho \to \infty$. The simulation results (introduced next) agree with the above analysis.

For the case of naive decoding we have $l_{\text{naive}}(0) = T(M-1) - M$ which results into unbounded average complexity except for the case when $T = 1$. However, for the case that corresponds to $T = M = 2$, although it becomes unbounded, the average complexity grows slowly with the SNR as $(\log \rho)^{2MT}$. For $T > 2$, the average complexity grows quickly with SNR as $(\log \rho)^{2MT} \rho^{T(M-1)-M}$ resulting in an unbounded complexity. However, the experimental results (provided in the next section) shows that the average complexity of such a decoder decays (albeit rather slowly) with SNR, for $T \geq 2$. This means that the theoretical bound derived above fails to predict the average complexity behavior of the naive sphere decoder for such values of T. In all cases, the simulation results show that the average complexity becomes extensively high for values of codeword length $T \geq 2$. In general, at any multiplexing gain r, we have that $l_{\text{MMSE-DFE}}(r) < l_{\text{naive}}(r)$. This again proves that employing MMSE-DFE preprocessing at the decoding stage significantly improves the average computational complexity of the decoder at all multiplexing gains.

Moreover, for the case of MMSE-DFE sphere decoding, there exists a *cut-off* multiplexing gain, say r_0, such that the average computational complexity of the decoder remains bounded as long as we operate below such value. This value can be easily found by setting $l_{\text{MMSE-DFE}}(r_0) = 0$. This results in

$$r_0 = \left\lfloor \frac{MN}{M+T} \right\rfloor. \tag{3.47}$$

Interestingly, for the DMT optimal random LAST codes with $T = N + M - 1$, if we let the number of receive antennas $N \to \infty$, then one can achieve a cut-off multiplexing gain $r_0 = M$ which is the maximum multiplexing gain achieved by the channel. This shows that one can dramatically improve the computational complexity of the decoder by increasing the number of antennas at the receiver side.

From the above analysis, one can see that it is impossible for the sphere decoder to maintain very low decoding complexity while achieving the maximal diversity (or the optimal tradeoff) of the channel, especially for the case of nested LAST codes discussed previously. For the case of MMSE-DFE sphere decoding, achieving the maximum diversity MN requires the use of LAST codes with codeword lengths $T \geq N+M-1$. Increasing the number of receive antennas N requires increasing T as well, and hence, the second term in (3.39) does not decay very quickly to zero. It turns out that the sphere decoder may achieve *linear* computational complexity m for high SNR for large enough number of antennas N and *fixed* T, however at the expense of losing the maximum diversity MN (or losing the optimal tradeoff).

3.5 Simulation Results

We consider a MIMO system with $M = N = 2$, $T = 3$ for different rates $R = 4,\ 8$ bits per channel use. The LAST code is selected randomly and is obtained as an (m, p, k) Loeliger construction (refer to Chapter 2 for a detailed description). The computational complexity distribution $\Pr(C > L)$ is plotted when L increases proportionally with SNR as $L = \rho$, for both the naive and the MMSE-DFE sphere decoders at different rates (see Figure. 3.3 and Figure. 3.4). For comparison, the frame error rate of the corresponding decoders, and the outage probability at the same coding rates are also plotted. It is clear from both figures that the curves which correspond to the outage probability, the error performance, and the computational complexity distribution match in slope, i.e., they all exhibit the same behavior at high SNR. In other words, all curves have the same SNR exponent. This basically agrees with the derived theoretical results. Moreover, the average computational complexity are plotted in Figure. 3.5, and Figure. 3.6 and Figure. 3.7 for both MMSE-DFE and naive decoding, respectively. Figure. 3.5 shows how the average number of computations decays very quickly to m at high SNR, even for large values of T. Figure. 3.6 and Figure. 3.7 show how the average computational complexity is affected by the codeword length T, at a fixed rate $(r = 0)$, for the case of naive sphere decoding. In a 2×2 quasi-static MIMO channel under naive sphere decoding, the maximum diversity gain $M = 2$ is achieved when $T \geq 1$. Three random nested LAST codes with codeword lengths $T = 1, 2$, and 3 are used to achieve the same diversity gain. However, as discussed in the previous section, using a codeword length $T \leq 2$ would result in a small average decoding complexity. For $T = 3$ the average computational complexity becomes extensively large. This is clearly depicted in Figure. 3.6 and Figure. 3.7 where, even at high SNR, the average number of computations decays to m at a slower rate compared to the case of MMSE-DFE sphere decoding.

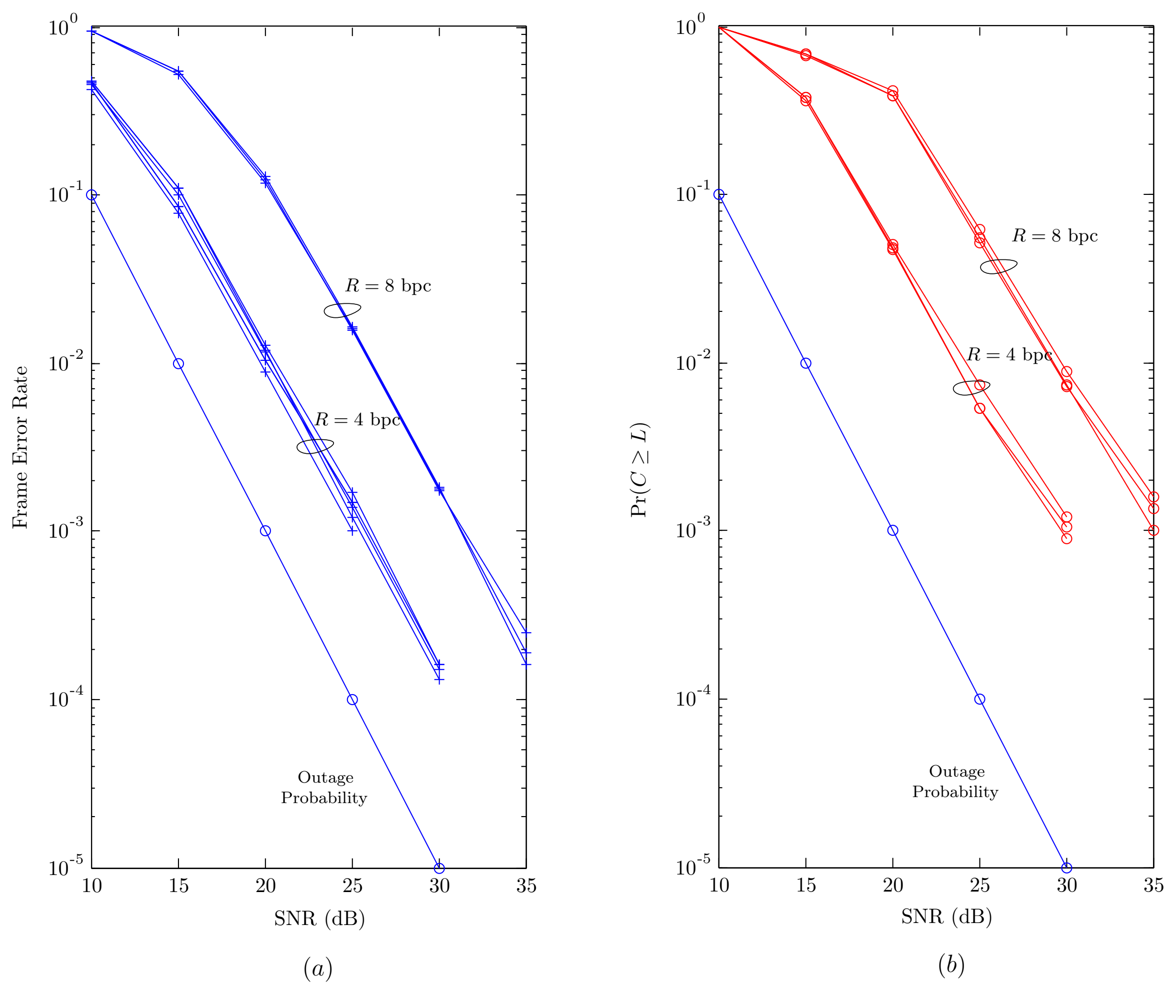


Figure 3.3: (a) Performance and (b) complexity distribution (with $L = \rho$) achieved by the naive sphere decoder for the case of 2×2 LAST coded MIMO channel.

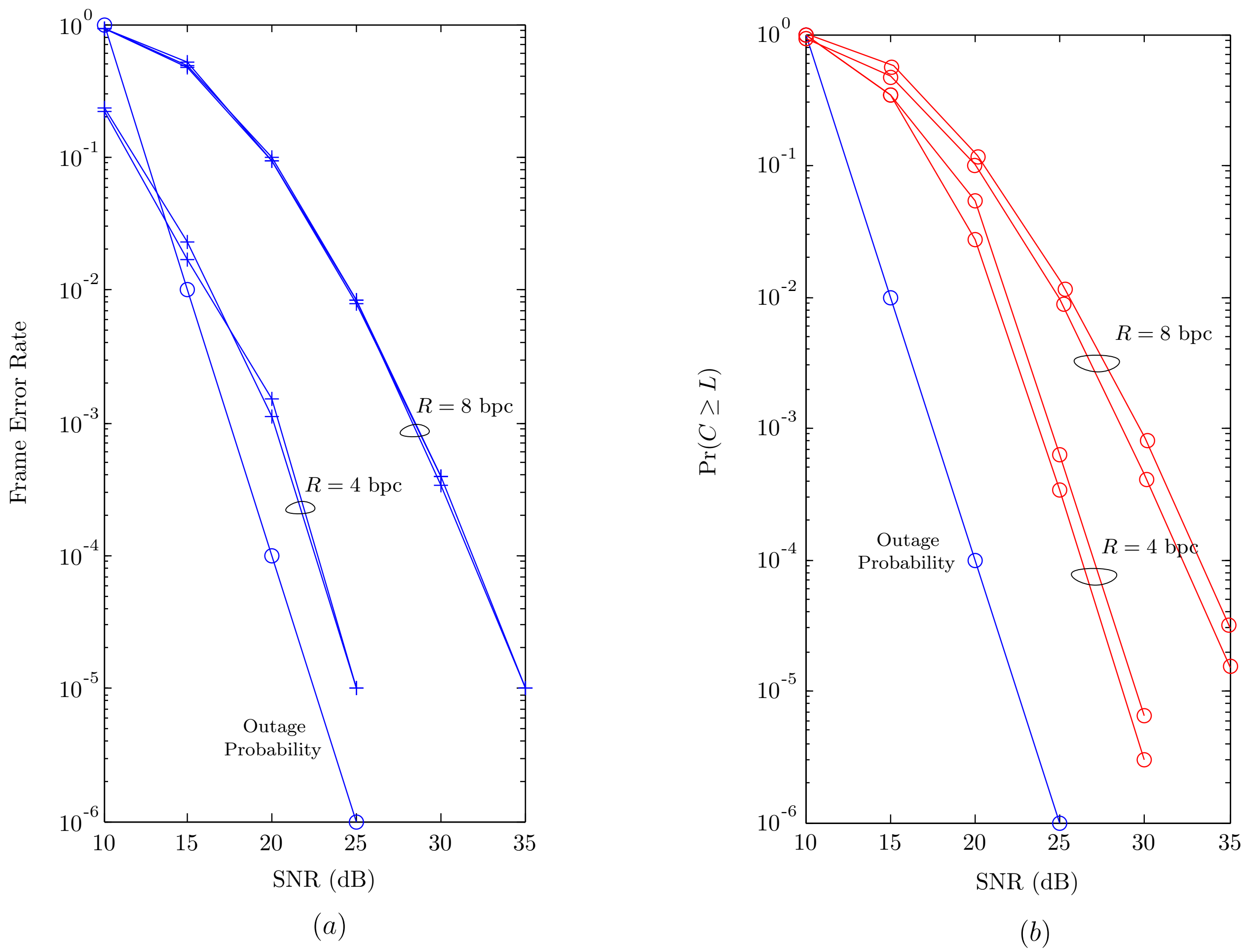


Figure 3.4: (a) Performance and (b) complexity distribution (with $L = \rho$) achieved by the MMSE-DFE sphere decoder for the case of 2×2 LAST coded MIMO channel.

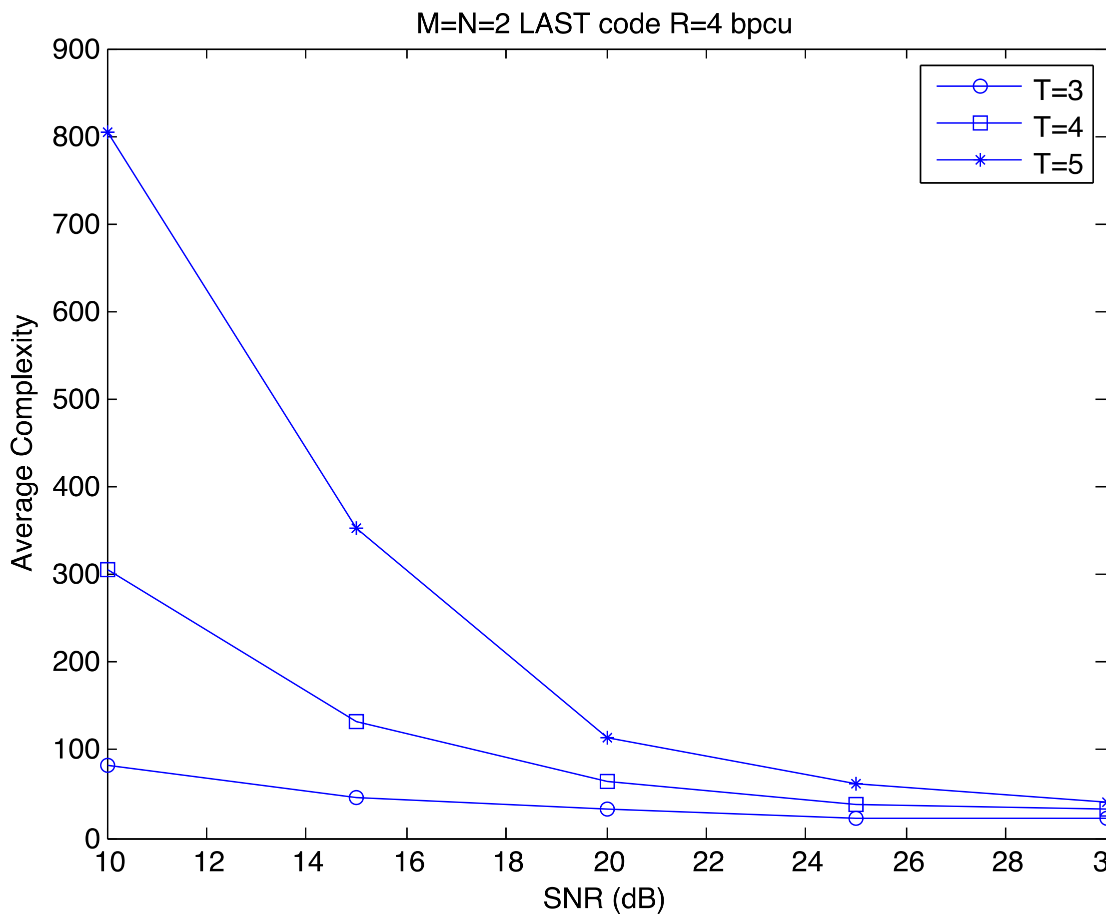


Figure 3.5: The reduction in computational complexity achieved by the MMSE-DFE lattice decoder for all values of T that achieve maximum diversity 4. All curves decays quickly to $m = 2MT = 4T$ at high SNR.

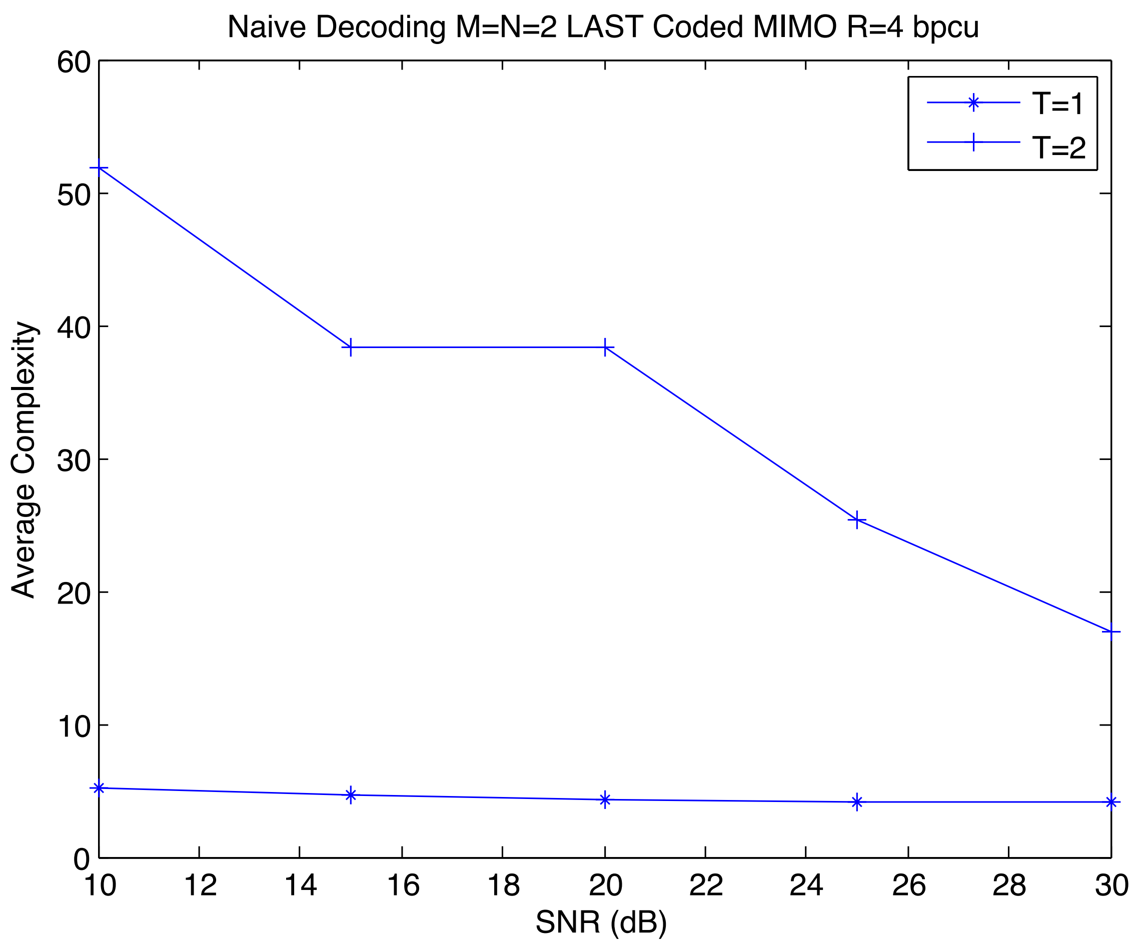


Figure 3.6: The computational complexity achieved by the naive lattice decoder for values of $T = 1$ and $T = 2$, that achieve maximum diversity 2.

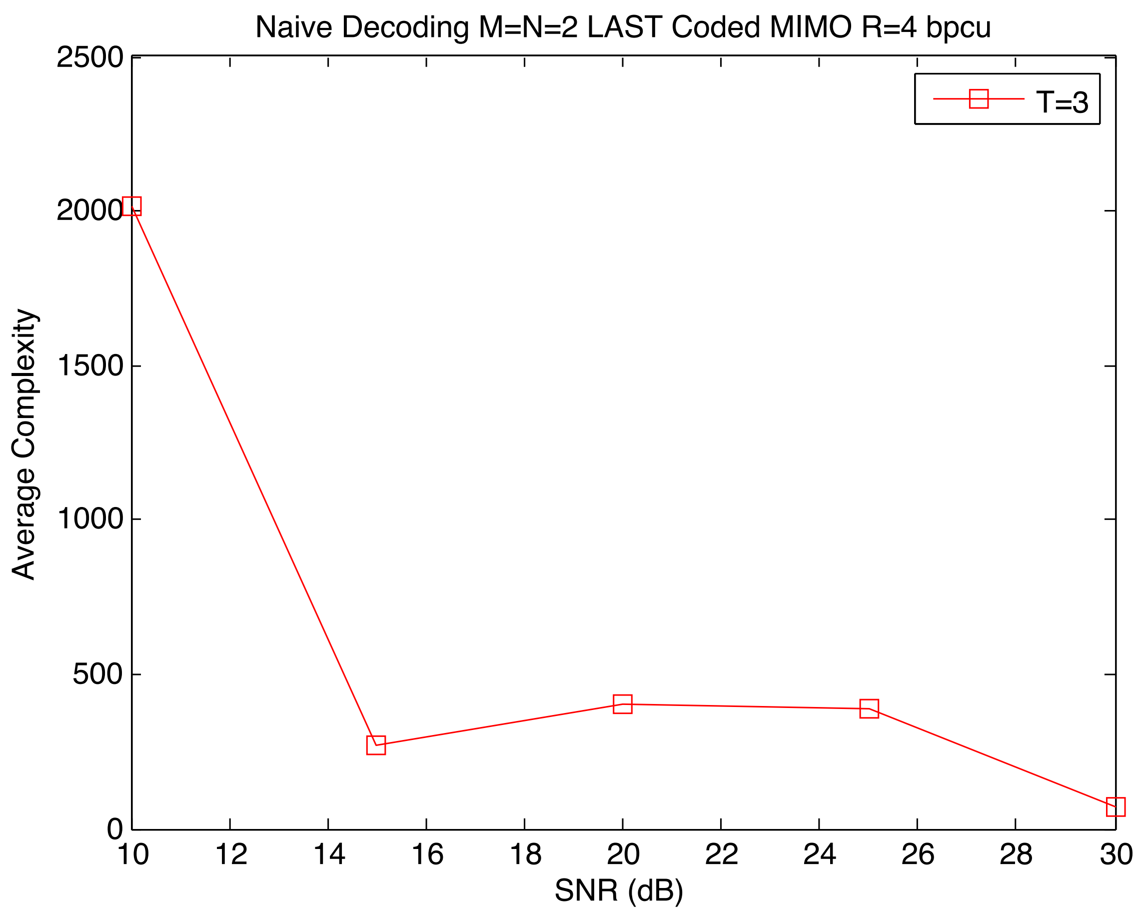


Figure 3.7: The computational complexity achieved by the naive lattice decoder for values of $T = 3$.

3.6 Summary

In this chapter, we have provided a complete analysis for the computational complexity of a fixed radius sphere decoder applied to LAST coded MIMO channel, at the high SNR regime. The sphere radius increases with SNR (as $\log \rho$) but is independent of the lattice and the channel conditions. An upper bound of the asymptotic complexity distribution has been derive. It has been shown that, for both the naive and the MMSE-DFE sphere decoders, if the number of computations performed by the decoder exceeds a certain limit, the complexity's tail distribution becomes upper bounded by the asymptotic outage probability achieved by the LAST coding and sphere decoding schemes. As a result, the tradeoff of the MIMO channel under sphere decoding is naturally extended to include the decoder's complexity. When the channel is well-conditioned, this computations limit can be used as an indication of when the decoder can terminate the search to save on complexity without affecting the achievability of the optimal tradeoff. The average number of computations that is required to terminate the search when the channel is not in outage has been calculated in terms of the system parameters. As expected, MMSE-DFE preprocessing significantly improves the overall computational complexity of the underlying decoding scheme.

It is clear from the previous analysis that the complexity of the sphere decoder depends critically on the system parameters M, N, and T. In order to achieve high order diversity, the number of antennas and the codeword length must be increased simultaneously, causing the complexity of the decoding to increase. The search for low complexity decoders that can achieve near-optimal performance is considered a challenging problem. As will be shown in the next Chapter, we attempt to solve this issue using efficient tree search algorithms to perform lattice decoding that are capable of providing an excellent performance-complexity tradeoff in the outage-limited MIMO channel.

Chapter 4

Lattice Sequential Decoding for The LAST Coded MIMO Channel

WITH the aid of MMSE-DFE at the decoding stage, lattice coding and sphere decoding achieve the *optimal* tradeoff of the channel. However, as we have seen in the previous chapter, sphere decoders are only efficient in the high signal-to-noise ratio (SNR) regime and low signal dimensions, and exhibits exponential (average) complexity for low-to-moderate SNR and large signal dimensions [52], [54]. On the other extreme, linear and non-linear receivers such as zero-forcing, MMSE, and MMSE-DFE decoders, are considered attractive alternatives to lattice decoders in MIMO channels and have been widely used in many practical communication systems [10]–[35]. Unfortunately, the very low decoding complexity advantage that these decoders can provide comes at the expense of poor performance, especially for large signal dimensions. The problem of designing low complexity receivers for the MIMO channel that achieve *near-optimal* performance is considered a challenging problem and has driven much research in the past years. In this work, we analyze the performance of *lattice sequential decoding* that is capable of bridging the gap between sphere decoders and low complexity linear decoders (e.g., MMSE-DFE decoder).

Applying sequential decoders for the detection of signals transmitted via MIMO communication channels introduced an alternative and interesting approach to solve the CLPS problem that is related to the optimum decoding rule in such channels [39], [40]. Morgan et. al. [39] showed that lattice sequential decoders, although sub-optimal, are capable of achieving good, and for some cases near ML, error performance. The analysis was considered *only* for the case of uncoded MIMO channel (i.e., V-BLAST). It was demonstrated that lattice sequential decoders achieve the maximum receive diversity provided by the channel and for low signal dimensions it achieves near-ML performance while significantly reducing decoding complexity compared to lattice decoder. **The performance limits achieved by lattice sequential decoders for (lattice) space-time coded MIMO channel [23], [13], [16] has not yet been studied.**

Conventional sequential decoders (e.g., Fano and Stack algorithms [36],[38]) were originally constructed as an alternative to the ML decoder to decode convolutional codes transmitted via discrete memoryless channel while achieving low (average) decoding complexity. Although sequential decoding algorithms are simple to describe, the analysis of decoding complexity is considered difficult. This is due to the fact that the amount of computations performed by the decoder attempting to decode a message is random. Therefore, sequential decoding complexity is usually analyzed through its computational distribution. For codes transmitted at rate R, the asymptotic computational complexity C of sequential decoding for the above mentioned channel follows a Pareto distribution [37],

$$\Pr(C > L) \approx L^{-e(R)}, \quad L \to \infty, \tag{4.1}$$

where $e(R)$ is the tail distribution exponent that is a function of R. Theoretical analysis showed that $e(R) > 1$ as long as $R < R_0$, where R_0 is the well-known channel *cut-off* rate. In other words, average computational complexity is kept bounded as long as we operate at rates below R_0. For the quasi-static MIMO channel, it is expected that lattice sequential decoders would behave in a similar fashion.

Similar to the discrete memoryless channel, our analysis reveals that there exists a *cut-off* multiplexing gain for which the average computational complexity of the lattice sequential decoder remains bounded as long as we operate below such value. In this work, we show that a tradeoff exists between the (average) computational complexity of the decoder and the multiplexing gain. The tradeoff is characterized by the tail exponent of the computational distribution, which is shown to be equivalent to the diversity-multiplexing tradeoff achieved by such decoding scheme.

In this chapter, the asymptotic performance of the lattice sequential decoder for lattice space-time coded MIMO channel is analyzed. We determine the rates achievable by lattice coding and sequential decoding applied to such a channel. The diversity-multiplexing tradeoff under lattice sequential decoding is derived as a function of its parameter—*the bias term*, which is critical for controlling the amount of computations required at the decoding stage. Achieving low decoding complexity requires increasing the value of the bias term. However, this is done at the expense of losing the optimal tradeoff of the channel. In this work, we derive the tail distribution of the decoder's computational complexity in the high signal-to-noise ratio regime. Our analysis reveals that the tail distribution of such low complexity decoder is dominated by the outage probability of the channel for the underlying coding scheme. Also, the tail exponent of the complexity distribution is shown to be equivalent to the diversity-multiplexing tradeoff achieved by lattice coding and lattice sequential decoding schemes. We show analytically how minimum-mean square-error decision feed-back equalization can significantly improve the tail exponent and as a consequence reduces computational complexity. In particular, we show that there exists a *cut-off* multiplexing gain for which the average computational complexity of the decoder remains bounded.

4.1 System Model

We consider a quasi-static, Rayleigh fading MIMO channel with M-transmit, N-receive antennas, and no CSI at the transmitter and perfect CSI at the receiver. The complex base-band model of the received signal can be mathematically described by

$$\boldsymbol{Y}^c = \sqrt{\rho}\boldsymbol{H}^c\boldsymbol{X}^c + \boldsymbol{W}^c, \tag{4.2}$$

where $\boldsymbol{X}^c \in \mathbb{C}^{M\times T}$ is the transmitted space-time code matrix, T is the number of channel usages, $\boldsymbol{Y}^c \in \mathbb{C}^{N\times T}$ is the received signal matrix, $\boldsymbol{W}^c \in \mathbb{C}^{N\times T}$ is the noise matrix, $\boldsymbol{H}^c \in \mathbb{C}^{N\times M}$ is the channel matrix, and $\rho = \mathsf{SNR}/M$ is the normalized SNR at each receive antenna with respect to M. The elements of both the noise matrix and the channel fading gain matrix are assumed to be i. i. d zero mean circularly symmetric complex Gaussian random variables with variance $\sigma^2 = 1$.

An $M \times T$ space-time coding scheme is a full-dimensional LAttice Space-Time (LAST) code if its vectorized (real) codebook (corresponding to the channel model (2.5)) is a lattice code with dimension $m = 2MT$. As discussed in [23], the design of space-time signals reduces to the construction of a codebook $\mathcal{C} \subseteq \mathbb{R}^{2MT}$ with code rate $R = \frac{1}{T}\log|\mathcal{C}|$, satisfying the input averaging power constraint

$$\frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}\|^2 \leq MT. \tag{4.3}$$

The equivalent real model of (4.2) can be easily shown to be given by (2.5) with

$$\boldsymbol{M} = \sqrt{\rho}\,\boldsymbol{I}_T \otimes \begin{pmatrix} \Re\{\boldsymbol{H}^c\} & -\Im\{\boldsymbol{H}^c\} \\ \Im\{\boldsymbol{H}^c\} & \Re\{\boldsymbol{H}^c\} \end{pmatrix}.$$

where $\otimes$ denotes the Kronecker product.

4.2 Lattice Fano/Stack Sequential Decoder

As we have discussed in Chapter 3, the sphere decoder can be considered as a search in a tree. Generally speaking, a sphere decoding algorithm explores the tree of all possible lattice points and uses the path metric in order to discard paths corresponding to points outside the search sphere. As an alternative to sphere decoding algorithms, sequential decoders comprise a set of efficient and powerful decoding techniques able to perform the tree search. These decoders can achieve near optimal performance without suffering the complexity of the ML or sphere decoder for coding rates not too close to the channel capacity [2], [37].

The sequential search on a tree can be briefly described as follows: the search is attempted one branch at a time. Namely, if the decoder is "located" at a particular node, it will move forward along the most likely branch stemming from it and thus reach a new node, provided that the likelihood of the entire past path up to and including the new node exceeds a certain current threshold. If it does not, then the decoder must return to the preceding node. From there it will try to move forward along an alternate path. It will succeed in this attempt if the value of the likelihood of the new path exceeds a threshold appropriate to it. Thus, the decoder moves forward and backward with the hope that the likely paths are going to be examined so that the average decoding effort will be kept low.

Fano and Stack sequential decoders [36], [38] are efficient tree search algorithms that attempt to find a "best fit" with the received noisy signal. As in conventional sequential decoder, to determine a best fit (path), values are assigned to each node on the tree. This value is called the *metric*. For lattice sequential decoders, this metric [corresponds to (2.5)] is given by (see [39])

$$\mu(\boldsymbol{z}_1^k) = bk - \|\boldsymbol{y'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2, \quad \forall\, 1 \leq k \leq m, \tag{4.4}$$

where $\boldsymbol{z}_1^k = [z_k, \cdots, z_2, z_1]^\mathsf{T}$ denotes the last k components of the integer vector $\boldsymbol{z}$, $\boldsymbol{R}_{kk}$ is the lower $k \times k$ part of the matrix $\boldsymbol{R}$ that corresponds to the QR decomposition of the channel-code matrix $\boldsymbol{MG} = \boldsymbol{QR}$, $\boldsymbol{y'}_1^k$ is the last k components of the vector $\boldsymbol{y'} = \boldsymbol{Q}^\mathsf{T}\boldsymbol{y}$, and $b \geq 0$ is the bias term.

In the Stack algorithm, as the decoder searches the different nodes in the tree, an ordered list of previously examined paths of different lengths is kept in storage. Each stack entry contains a path along with its metric. Each decoding step consists of extending the top (best) path in the stack. The decoding algorithm terminates when the top path in the stack reaches the end of the tree (refer to [38] for more details about the algorithm).

In the Fano algorithm, as the decoder searches nodes, values of the path metric are compared to a certain threshold denoted by $\tau \in \{\cdots, -2\delta, -\delta, 0, \delta, 2\delta, \cdots\}$ where δ is called the step size. The decoder attempts to extend the most probable path by moving "forward" if the path metric stays above the running threshold. Otherwise, it moves "backward" searching for another path that may lead to the most probable transmitted sequence (refer to [36] for more details about the algorithm).

Although the Stack decoder and the Fano algorithm generate essentially the same set of visited nodes (see [39]), the Fano decoder visits some nodes more than once. However, the Fano decoder requires essentially no memory, unlike the Stack algorithm. Also, it must be noted that the way the nodes are generated in both sequential algorithms plays an important role in reducing the computation complexity and for some cases may improve the detection performance. For example, the determination of the best and next best nodes is simplified in the CLPS problem by using the Schnorr-Euchner enumeration [32] which generates nodes with metrics in ascending order given any node $\boldsymbol{z}_1^k$.

4.3 Performance Analysis for Fixed Bias Term: Achieving the Optimal Tradeoff

After the work of [22], the diversity-multiplexing tradeoff — a fundamental tradeoff between rate via *multiplexing* and error probability via *diversity*, has become a standard metric in the characterization of the quasi-static Rayleigh fading MIMO channel. Our goal in this section is to analyze the diversity-multiplexing tradeoff achieved by the lattice sequential decoder when the bias b (defined in (4.4)) is held fixed but not too large. We consider two scenarios: the naive and MMSE-DFE lattice sequential decoders. The latter corresponds to the case when the decoder is preprocessed by MMSE-DFE filtering.

For the case of naive lattice sequential decoding we have the following result:

Theorem 3. *For $N \geq M$ and any block length $T \geq 1$, there exists a sequence of full-dimensional LAST codes that achieves diversity gain*

$$d(r) = \min\{T, N - M + 1\}(M - r), \ \forall r \in [0, M], \tag{4.5}$$

under naive lattice sequential decoding for fixed bias $b \geq 0$.

Proof. See Appendix B. □

It is clear from the above theorem that the naive lattice sequential decoder is not capable of achieving the optimal tradeoff of the channel for any finite $b \geq 0$. This result is expected, since the performance of such a decoder upper bounds the performance of naive lattice decoder (corresponds to $b = 0$), where the latter has been shown in [23] to be sub-optimal, and achieves SNR exponent $d(r)$ as defined in Theorem 1 (Chapter 1).

Similar to the analysis provided in [23], in order to improve the performance of the lattice sequential decoder one could apply MMSE-DFE prior decoding. It has been shown in [23] that, for a fixed, non-random channel matrix $\boldsymbol{H}^c$, the rate

$$R_{\text{mod}}(\boldsymbol{H}^c, \rho) = \log\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right), \tag{4.6}$$

is achievable by *nested* LAST codes (see Chapter 2) and MMSE-DFE lattice decoding. For such coding and decoding schemes, the real channel model can be shown to be expressed by (2.5) with $\boldsymbol{M} = \boldsymbol{B}$ and $n = m$, where $\boldsymbol{B}$ is the feedback matrix of the MMSE-DFE (see [23] for more details) that satisfies

$$\det(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}) = \left[\det\left(\boldsymbol{I}_M + \rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right]^{2T}. \tag{4.7}$$

However, in such scheme, the additive noise becomes non-Gaussian, but for a well-constructed lattice code[1] it is asymptotically (as $T \to \infty$) Gaussian [24], [25]. This creates some difficulty in decoder's performance and complexity analysis in the outage-limited MIMO channel (due to T being finite) which can be overcome as discussed in Chapter 2, Section 2.1.

In our analysis, we apply the same mod-Λ scheme that was described in Chapter 2 with some modification in the decoding stage. Here, we replace the lattice decoder by the lattice sequential decoder. in this case, the decoder first estimates[2] the closest lattice point to $\boldsymbol{y}'$, say $\hat{\boldsymbol{z}}$. Then, the decoded codeword is given by $\hat{\boldsymbol{c}} = [\boldsymbol{G}\hat{\boldsymbol{z}}] \bmod \Lambda_s$. In this case, we have the following result:

Theorem 4. *There exists a sequence of nested LAST codes with block length* $T \geq M + N - 1$ *that achieves the optimal diversity-multiplexing tradeoff curve*

$$d^*(r) = (M - r)(N - r),\ \forall r \in [0, \min\{M, N\}], \tag{4.8}$$

under the mod-Λ *scheme and lattice sequential decoding for fixed bias* $b \geq 0$.

[1] Lattices that satisfy Minkowski-Hlawka theorem (see [7], [8] for more details)

[2] It must be noted that other than the case of $b = 0$, the output of the decoder may not always be *the* closest lattice point to the received signal. Lattice sequential decoding attempts to estimate the decoding rule that is given in (2.8).

Proof. See Appendix C. □

The above theorem indicates that the use of optimal receivers (e.g., ML and lattice decoders) is *not* essential if the main goal is to achieve the optimal tradeoff of the channel. Sub-optimal receivers may do the job. It should be noted, however, that although the optimal diversity-multiplexing tradeoff is achieved by such decoders, the performance gap from ML or lattice decoder increases as b becomes large. To achieve near-ML performance in this case, one has to resort to low values of b.

At this point, one may ask the following question: how large b can be set in order not to loose the optimal tradeoff? For fixed (finite) b, one cannot catch the effect of the bias term on the diversity-multiplexing tradeoff achieved by such decoding scheme. In order to do that, we allow the bias term to vary with SNR and channel coefficients as will be shown in the sequel.

4.4 Achievable Rate & Outage Performance Analysis: Variable Bias Term

In this section, we would like to study the behavior of the outage probability under lattice sequential decoding when the bias term b is allowed to change with SNR. It has been shown in Section 4.3 that the naive lattice decoder cannot achieve the optimal tradeoff of the channel for the any $b \geq 0$. Therefore, in this section we exclude such a decoder from further discussion. In what follows, we consider the use of the MMSE-DFE lattice sequential decoder. As discussed in the previous section, rate up to R_{mod} is achievable by lattice coding and decoding. When the lattice decoder is replaced by the lattice Fano /Stack[3] sequential decoder we get the following result:

Theorem 5. *For a fixed non-random channel matrix* $\boldsymbol{H}^c$, *the rate*

$$R_b(\boldsymbol{H}^c, \rho) \triangleq \max\left\{R_{\text{mod}}(\boldsymbol{H}^c, \rho) - 2M \log\left(\frac{1+\sqrt{1+8\alpha}}{2}\right), 0\right\}, \tag{4.9}$$

[3] For the Fano algorithm, we assume throughout the Chapter that only small values of step size δ is used by the decoder, and hence, its affect on the performance analysis can be neglected (see the proof of Theorem 5). Otherwise, choosing very large values of δ may result in very poor performance. For the Stack algorithm, we have $\delta = 0$.

is achievable by LAST coding and MMSE-DFE lattice Fano/Stack sequential decoding with bias term b, where α is given by

$$\alpha = \left(\frac{r_{\text{eff}}(\boldsymbol{BG_c})}{2r_{\text{pack}}(\boldsymbol{BG_c})}\right)^2 b, \tag{4.10}$$

and $r_{\text{eff}}(\boldsymbol{BG}_c)$ and $r_{\text{pack}}(\boldsymbol{BG}_c)$ are the effective radius and packing radius of the lattice generated using $\boldsymbol{BG}_c$.

Proof. The proof relies on the ambiguity decoder that was introduced in Chapter 2. Consider an $m = 2MT$-dimensional lattice code $\mathcal{C}(\Lambda_c, \boldsymbol{u}_0, \mathcal{R})$ (that corresponds to a generated matrix $\boldsymbol{G}_c$) with fundamental volume V_c that satisfies (4.3), for some fixed translation vector $\boldsymbol{u}_0$, and $\mathcal{R}$ is the $m/2$-dimensional hypersphere with radius $\sqrt{m/2}$ centered at the origin.

The input to the MMSE-DFE lattice sequential decoder is the vector $\boldsymbol{y}' = \boldsymbol{Q}^\mathsf{T}\boldsymbol{y}$, where $\boldsymbol{Q}$ is an orthogonal matrix that corresponds to the QR decomposition of the channel-code matrix $\boldsymbol{MG}_c = \boldsymbol{BG}_c = \boldsymbol{QR}$. The associated path metric in this case is given by (4.4).

Consider the Fano algorithm with bias $b \geq 0$, threshold τ, and step size δ. Let E_f be the event that the Fano decoder makes an erroneous detection, conditioned on $\tau_{\min} > \mu_{\min} - \delta$, where $\tau_{\min}$ is the minimum threshold used by the decoder, $\mu_{\min} = \min\{0, b - \|\boldsymbol{e}'^1_1\|^2, 2b - \|\boldsymbol{e}'^2_1\|^2, \ldots, bm - \|\boldsymbol{e}'^m_1\|^2\}$ is the minimum metric that corresponds to the transmitted path, and $\boldsymbol{e}' = \boldsymbol{Q}^\mathsf{T}\boldsymbol{e}$. Then, $P_e = \mathsf{E}_{\tau_{\min}}\{\Pr(E_f)\}$ is the frame error rate of the lattice Fano sequential decoder. Due to lattice symmetry, we can assume that the all zero codeword, i.e., $\mathbf{0}$, was transmitted. For a given lattice Λ_c,

$$\begin{aligned}
\Pr(E_f|\Lambda_c) &\overset{(a)}{\leq} \Pr\left(\bigcup_{\boldsymbol{z}\in\mathbb{Z}^m\setminus\{\mathbf{0}\}} \{\mu(\boldsymbol{z}) > \mu_{\min} - \delta\}\right) \\
&\overset{(b)}{\leq} \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*} \{\|\boldsymbol{Bx}\|^2 - 2(\boldsymbol{Bx})^\mathsf{T}\boldsymbol{e} < bm + \delta\}\right) \\
&= \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*} \left\{2(\boldsymbol{Bx})^\mathsf{T}\boldsymbol{e} \geq \|\boldsymbol{Bx}\|^2\left(1 - \frac{bm+\delta}{\|\boldsymbol{Bx}\|^2}\right)\right\}\right),
\end{aligned} \tag{4.11}$$

where $\Lambda_c^* = \Lambda_c\setminus\{\mathbf{0}\}$, (a) is due to the fact that in general, $\mu(\boldsymbol{z}) > \mu_{\min} - \delta$ is just a necessary condition for $\boldsymbol{x} = \boldsymbol{G}_c\boldsymbol{z}$ to be decoded by the Fano decoder, and (b) follows by noticing that

$-(\mu_{\min}+\|\boldsymbol{e}'\|^2)\leq 0$. Note the independence of (4.11) on $\tau_{\min}$. It is clear from the above analysis that lattice Fano sequential decoder approaches the performance of lattice decoder as $b,\delta\to 0$. Now, using the fact that

$$\|\boldsymbol{Bx}\|^2\geq\min_{\boldsymbol{x}\in\Lambda_c^*}\|\boldsymbol{Bx}\|^2=(2r_{\text{pack}}(\boldsymbol{BG}_c))^2,$$

where $r_{\text{pack}}(\boldsymbol{BG}_c)$ is the packing radius of the lattice $\Lambda(\boldsymbol{BG}_c)$, we can further upper bound (4.11) as

$$\Pr(E_f|\Lambda_c)\quad\leq\quad\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2(\boldsymbol{B}'\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}\geq\|\boldsymbol{B}'\boldsymbol{x}\|^2\right\}\right),\tag{4.12}$$

where

$$\boldsymbol{B}'=\left(1-\frac{b+\delta/m}{(2r_{\text{pack}}(\boldsymbol{BG}_c))^2/m}\right)\boldsymbol{B}.\tag{4.13}$$

Now, one can show that

$$\begin{aligned}(2r_{\text{pack}}(\boldsymbol{BG}_c))^2/m&=2^{R_{\text{mod}}/M}\frac{\Gamma(m/2+1)^{2/m}V(\mathcal{S}_m(2r_{\text{pack}}(\boldsymbol{BG}_c)))^{2/m}}{m\pi 2^{R_{\text{mod}}/M}}\\&\overset{(a)}{=}2^{R_{\text{mod}}/M}\frac{V(\mathcal{S}_m(2r_{\text{pack}}(\boldsymbol{BG}_c)))^{2/m}}{2V(\mathcal{R})^{2/m}\det(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})^{2/m}}\\&=\frac{2^{R_{\text{mod}}/M}}{2}\frac{V_c^{2/m}}{V(\mathcal{R})^{2/m}}\frac{V(\mathcal{S}_m(2r_{\text{pack}}(\boldsymbol{BG}_c)))^{2/m}}{V_c^{2/m}V(\mathcal{V}(\boldsymbol{B}))^{2/m}}\\&\overset{(b)}{=}\frac{2^{[R_{\text{mod}}-R]/M}}{2}\frac{V(\mathcal{S}_m(2r_{\text{pack}}(\boldsymbol{BG}_c)))^{2/m}}{V(\mathcal{V}(\boldsymbol{BG_c}))^{2/m}}\\&\overset{(c)}{=}\frac{2^{[R_{\text{mod}}-R]/M}}{2}\frac{V(\mathcal{S}_m(2r_{\text{pack}}(\boldsymbol{BG}_c)))^{2/m}}{V(\mathcal{S}_m(r_{\text{eff}}(\boldsymbol{BG}_c)))^{2/m}}\end{aligned}\tag{4.14}$$

where (a) follows from the fact that $V(\mathcal{R})$ is the volume of the m-dimensional hypersphere of radius $\sqrt{m/2}$, (b) follows from the fact that there exists a shifted lattice code $\Lambda_c+\boldsymbol{u}_0^*$ with number of codewords inside the shaping region (see Chapter 2, Lemma 1)

$$|\mathcal{C}(\Lambda_c,\boldsymbol{u}_0^*,\mathcal{R})|=2^{RT}=\frac{V(\mathcal{R})}{V_c},$$

and (c) follows from the definition of the effective radius of the lattice generated using the matrix $\boldsymbol{BG}_c$. Therefore, we can further asymptotically (as $m\to\infty$) upper bound (4.12) as

$$\Pr(E_f|\Lambda_c)\quad\leq\quad\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2(\tilde{\boldsymbol{B}}\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}\geq\|\tilde{\boldsymbol{B}}\boldsymbol{x}\|^2\right\}\right),\tag{4.15}$$

where $\tilde{\boldsymbol{B}}$ is given by

$$\tilde{\boldsymbol{B}} = \left(1 - \frac{2b}{2^{[R_{\text{mod}}-R]/M}(2r_{\text{pack}}(\boldsymbol{BG}_c)/r_{\text{eff}}(\boldsymbol{BG}_c))^2}\right)\boldsymbol{B}, \tag{4.16}$$

where for large m, we have approximated $b + \delta/m \approx b$ for finite δ. The last equation in the upper bound (4.15) corresponds to the probability of decoding error of a received signal $\boldsymbol{y} = \tilde{\boldsymbol{B}}\boldsymbol{x} + \boldsymbol{e}$ decoded using lattice decoding and is valid for all values of $b < 2^{[R_{\text{mod}}-R-1]}(2r_{\text{pack}}(\boldsymbol{BG}_c)/r_{\text{eff}}(\boldsymbol{BG}_c))^2$. Let $b = \frac{1}{2}b'(2r_{\text{pack}}(\boldsymbol{BG}_c)/r_{\text{eff}}(\boldsymbol{BG}_c))^2$, where $b' \geq 0$ is a constant independent of the lattice Λ_c. In this case, we may express $\tilde{\boldsymbol{B}}$ as

$$\tilde{\boldsymbol{B}} = \left(1 - \frac{b'}{2^{[R_{\text{mod}}-R]/M}}\right)\boldsymbol{B}. \tag{4.17}$$

It is clear from (4.17) that $\tilde{\boldsymbol{B}}$ is invertible. In this case, we obtain the equivalent channel output

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{B}}^{-1}\boldsymbol{y}' = \boldsymbol{x} + \tilde{\boldsymbol{e}}.$$

Next, we apply the ambiguity decoder with decision region

$$\mathcal{E}'_{T,\gamma} \triangleq \left\{\boldsymbol{z} \in \mathbb{R}^m : \boldsymbol{z}^{\mathsf{T}}\tilde{\boldsymbol{B}}^{\mathsf{T}}\tilde{\boldsymbol{B}}\boldsymbol{z} \leq MT(1+\gamma)\right\}. \tag{4.18}$$

The probability of making a decoding error using the lattice sequential decoder can then be upper bounded by

$$\Pr(E_f|\Lambda_c) \leq \Pr(\tilde{\boldsymbol{e}} \in \mathcal{E}'_{T,\gamma}) + \Pr(\mathcal{A}(\mathcal{E}'_{T,\gamma})). \tag{4.19}$$

In this case, Lemma 1 can be easily applied to the bound (4.19) with $\boldsymbol{A} = \tilde{\boldsymbol{B}}$, and $r_e^2 = MT$. Noticing that

$$\det\left(\tilde{\boldsymbol{B}}^{\mathsf{T}}\tilde{\boldsymbol{B}}\right) = \left(1 - \frac{b'}{2^{[R_{\text{mod}}-R]/M}}\right)^{2m}\det\left(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}\right),$$

and by solving for R, we achieve the desired result.

The above derivation also applies to the Stack algorithm with minor modifications. In such algorithm, any lattice codeword $\boldsymbol{x} = \boldsymbol{G}_c\boldsymbol{z} \neq \boldsymbol{0}$ can be decoded as the closest lattice point to the received vector only if $\mu(\boldsymbol{z}) \geq \mu_{\min}$. Hence, the average error probability of the stack decoder can be upper bounded by (4.19) (since $\delta = 0$ in such algorithm). □

As discussed earlier, choosing a fixed but not very large values of b may result in achieving the optimal diversity-multiplexing tradeoff of the channel. However, lattice sequential decoders are used as an alternative to ML and lattice decoders to achieve very low decoding complexity and to do so one has to resort to large values of b. As will be shown in the sequel, choosing large values of b may lead to a loss in diversity gain and/or multiplexing gain, and as a result, a loss in the optimal tradeoff.

4.4.1 Outage Performance Analysis

Next, we consider a random channel matrix $\boldsymbol{H}^c$ as defined in (4.2) and obtain an achievable diversity-multiplexing tradeoff for LAST codes under MMSE-DFE lattice sequential decoding when b varies with SNR. Before we do that, we would like to analyze the outage behavior of the lattice sequential decoder and drive its achievable diversity-multiplexing tradeoff. Without loss of generality, we assume that $N \geq M$.

Our goal in this section is to show how the outage performance critically depends on the value of the bias term b. Denote $0 \leq \lambda_1 \leq \cdots \leq \lambda_M$ the eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c$. Consider b as a function of ρ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_M)$, and express it as

$$b(\boldsymbol{\lambda}, \rho) = \frac{1}{2}\frac{\prod_{i=1}^{M}(1+\rho\lambda_i)^{1/M}}{\eta(\boldsymbol{\lambda},\rho)^{1/M}}\left[1 - \left(\frac{\eta(\boldsymbol{\lambda},\rho)}{\prod_{i=1}^{M}(1+\rho\lambda_i)}\right)^{1/2M}\right]\left(\frac{2r_{\text{pack}}(\boldsymbol{BG}_c)}{r_{\text{eff}}(\boldsymbol{BG}_c)}\right)^2. \tag{4.20}$$

In this case, one can easily show that by substituting b in (4.10), we get

$$R_b(\boldsymbol{\lambda}, \rho) = \log \eta(\boldsymbol{\lambda}, \rho). \tag{4.21}$$

Depending on the value of $\eta(\boldsymbol{\lambda}, \rho)$ we obtain different achievable rates and hence different outage performances. For example, setting $\eta(\boldsymbol{\lambda}, \rho) = \prod_{i=1}^{M}(1+\rho\lambda_i)$ we achieve lattice decoder's outage performance, which corresponds to $b = 0$ and $R_b = R_{\text{mod}}$. To analyze the outage performance of lattice sequential decoders, we allow the bias term b to vary with SNR as defined in (4.20). We define the outage event under lattice sequential decoding as $\mathcal{O}_b(\rho) \triangleq \{\boldsymbol{H}^c : R_b(\boldsymbol{H}^c, \rho) < R\}$. Denote $R = r\log\rho$. The probability that the channel is in outage, $P_{\text{out}}(\rho, b) = \Pr(\mathcal{O}_b(\rho))$, can be evaluated as follows:

$$P_{\text{out}}(\rho, b) = \Pr(\log \eta(\boldsymbol{\lambda}, \rho) < R). \tag{4.22}$$

The term $\eta(\boldsymbol{\lambda}, \rho)$ can be chosen freely between 1 and $\prod_{i=1}^{M}(1+\rho\lambda_i)$ (the maximum achievable rate under lattice decoding). However, in our analysis and for the sake of simplicity, we let

$$\eta(\boldsymbol{\lambda}, \rho) = \prod_{i=1}^{M}(1+\rho\lambda_i)^{\zeta_i}, \tag{4.23}$$

where ζ_i, $\forall 1 \leq i \leq M$, are constants that satisfy the following two constraints: $\sum_{i=1}^{M}\zeta_i = M$, and $\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_M \geq 0$.

Now define $\nu_i \triangleq -\log\lambda_i/\log\rho$, then

$$\begin{aligned} P_{\text{out}}(\rho, b) &= \Pr\left(\log\prod_{i=1}^{M}(1+\rho\lambda_i)^{\zeta_i} < r\log\rho\right) \\ &\doteq \Pr\left(\sum_{i=1}^{M}\zeta_i(1-\nu_i)^{+} < r\right), \end{aligned} \tag{4.24}$$

where $(x)^{+} = \max\{0, x\}$. At high SNR, the typical outage event can be written as

$$\mathcal{O}_b^{+}(\zeta_1, \cdots, \zeta_M) \triangleq \left\{\boldsymbol{\nu} \in \mathbb{R}_{+}^{M} : \sum_{i=1}^{M}\zeta_i(1-\nu_i)^{+} < r\right\}.$$

In this case, the outage probability can be evaluated as follows:

$$P_{\text{out}}(\rho, b) = \int_{\mathcal{O}_b^{+}(\zeta_1, \cdots, \zeta_M)} f_{\boldsymbol{\nu}}(\boldsymbol{\nu})\, d\boldsymbol{\nu},$$

where $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ is the joint probability density function of $\boldsymbol{\nu}$ which, for all $\boldsymbol{\nu} \in \mathcal{O}_b^{+}(\zeta_1, \cdots, \zeta_M)$, is asymptotically given by [23]

$$f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \doteq \exp\left(-\log(\rho)\sum_{i=1}^{M}(2i-1+N-M)\nu_i\right). \tag{4.25}$$

Applying Varadhan's lemma as in [22], we obtain

$$P_{\text{out}}(\rho, b) \doteq \rho^{-d_b(r)},$$

where

$$d_b(r) = d(r, \boldsymbol{\zeta}) = \inf_{\boldsymbol{\nu} \in \mathcal{O}_b^{+}(\zeta_1, \cdots, \zeta_M)} \sum_{i=1}^{M}(2i-1+N-M)\nu_i.$$

where $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_M)$. It is clear from the above optimization problem that $d_b(r)$ depends critically on the selected coefficients $\boldsymbol{\zeta}$ (or equivalently b). Since ζ_i are ordered, one can assume without loss of generality of the optimal solution that $1 \geq \nu_1 \geq \cdots \geq \nu_M \geq 0$. The linear optimization problem is therefore equivalent to the following problem

$$\begin{cases} \text{Minimize}: & \sum_{i=1}^{M}(2i-1+N-M)\nu_i \\ \text{Such that}: & 0 \leq \nu_i \leq 1 \quad \forall i \geq 2 \\ & \sum_{i=1}^{M} \zeta_i \nu_i \geq M - r \end{cases}$$

where $\zeta_i \in [0, M]$. We arrive now to the following results:

- *Case 1*: ($0 < \zeta_i < M$, and $\sum_{i=1}^{M} \zeta_i = M$) We have the following:
  - If $r = 0$, the optimal solution is
  $$\nu_1^* = \cdots = \nu_M^* = 1.$$
  - If $r \neq 0$, the optimal solution is
  $$\nu_i^* = \min\left[\frac{1}{\zeta_i}\left(\sum_{j=i}^{M}\zeta_j - r\right)^+, 1\right] \quad \forall i \geq 1, \tag{4.26}$$

  and the diversity-multiplexing tradeoff is given by
  $$\begin{aligned} d_b(0) &= MN, \\ d(r, \boldsymbol{\zeta}) &= \sum_{i=1}^{M}(2i-1+N-M)\nu_i^*. \end{aligned} \tag{4.27}$$

  An interesting remark about this diversity-multiplexing tradeoff is that maximum diversity $d(0, \boldsymbol{\zeta}) = MN$ is independent of $\zeta_i, \forall i \geq 1$. Moreover, other than the uniform assignments[4] of $\boldsymbol{\zeta} = (1, \cdots, 1)$, the optimal diversity-multiplexing tradeoff cannot be achieved.

[4] This corresponds to the case of $b = 0$. However, if we choose $\eta(\boldsymbol{\lambda}, \rho) = \phi \prod_{i=1}^{M}(1+\rho\lambda_i)$ where $0 < \phi < 1$ is a constant independent of ρ, then according to (4.20) we have $b = \frac{1}{2}\phi^{-1/M}[1-\phi^{1/2M}]$, i.e., b is a constant independent of ρ. In this case, at high SNR we have $R \doteq R_{\text{mod}}$ which results in achieving the optimal diversity-multiplexing tradeoff. This agrees with result provided in Theorem 2 for finite bias term.

- *Case 2*: ($\zeta_i = 0$ for some i) For such choices of ζ_i, it is clear that the optimal diversity-multiplexing tradeoff is lost, i.e., $d_b(r) < (M-r)(N-r)$ for all $r = 0, 1, \cdots, M$. The maximum diversity achieved in this scenario can be easily shown to be given by

$$d(0, \boldsymbol{\zeta}) = MN - \sum_{i=1}^{M} (2i - 1 + N - M)\delta(\zeta_i),$$

where $\delta(\zeta_i) = 1$ if $\zeta_i = 0$ and 0 otherwise.

Interestingly, for *Case 1*, one can derive a closed form for the achievable diversity-multiplexing tradeoff as given in the following theorem:

Theorem 6. *The diversity-multiplexing tradeoff, $d_b(r)$, for an M-transmit, N-receive antenna coded MIMO Rayleigh channel under MMSE-DFE lattice Fano/Stack sequential decoder with bias b as given in (4.20) and coefficients $\zeta_i \in (0, M)$, $\forall 1 \leq i \leq M$, is the piecewise-linear function connecting the points (r(k),d(k)), $k = 0, 1, \cdots, M$ where*

$$\begin{aligned} r(0) = 0, \quad r(k) &= \sum_{i=M-k+1}^{M} \zeta_i, \ 1 \leq k \leq M, \\ d(k) &= (M-k)(N-k), \qquad 0 \leq k \leq M. \end{aligned} \tag{4.28}$$

Proof. By solving the above optimization problem, we obtain the following diversity-multiplexing tradeoff:

$$d(r, \boldsymbol{\zeta}) = \begin{cases} \sum_{i=1}^{M-k-1} (2i - 1 + N - M) + \\ \qquad \dfrac{2(M-k) - 1 + N - M}{\zeta_{M-k}} \left(\sum_{j=M-k}^{M} \zeta_j - r \right), & r \in [r_k, r_{k+1}],\ 0 \leq k \leq M-2; \\ \dfrac{N - M + 1}{\zeta_1} \left(\sum_{j=1}^{M} \zeta_j - r \right), & r \in [r_{M-1}, r_M], \end{cases} \tag{4.29}$$

where

$$r_k = \begin{cases} 0, & k = 0; \\ \sum_{i=M-k+1}^{M} \zeta_i, & 1 \leq k \leq M. \end{cases}$$

Substituting r_k in (4.29), we get the diversity-multiplexing tradeoff expression in (4.28). □

Example 1. *Consider a 2×2 MIMO channel. The diversity-multiplexing tradeoff curves achieved with respect to different values of ζ_i that correspond to Case 1 and Case 2 are illustrated in Figure. 4.1. Although the diversity at $r = 0$ is not affected by the coefficients $\zeta_i \neq 0$ ($d(0) = 4$), the more unbalanced the coefficients are, the worse the diversity-multiplexing tradeoff is.*

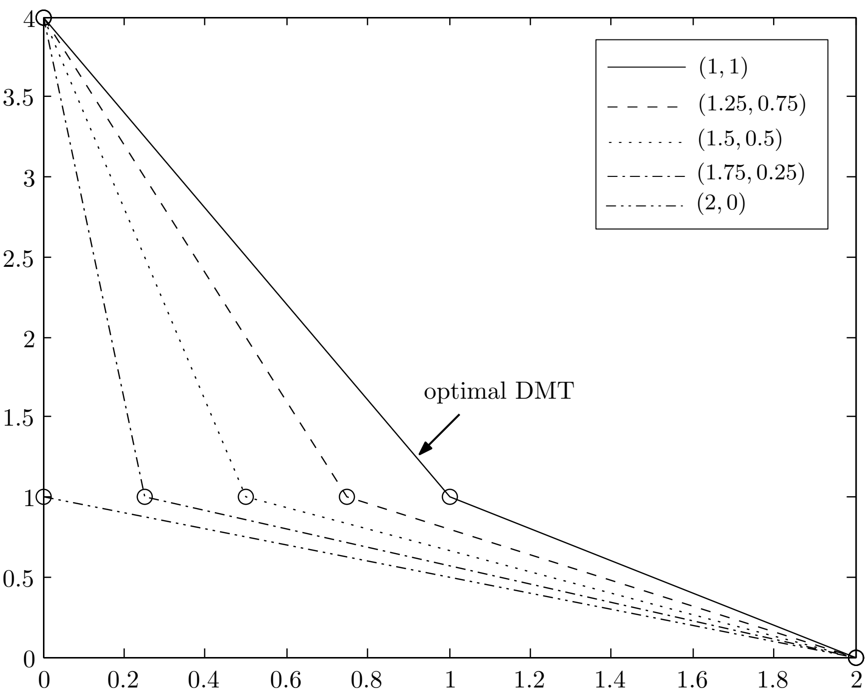
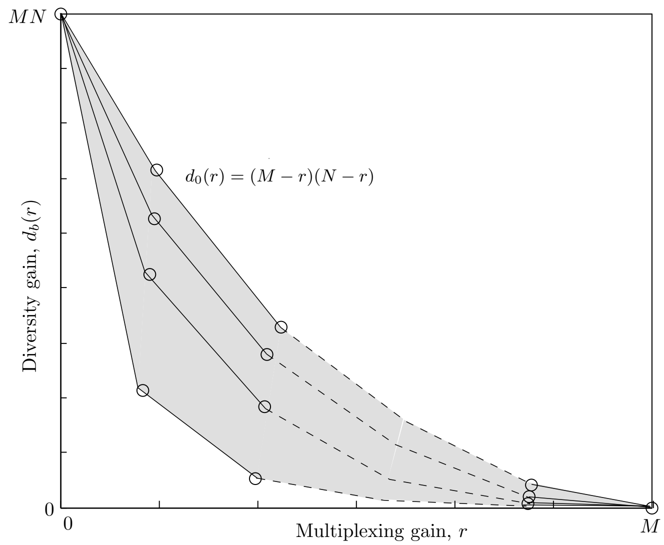


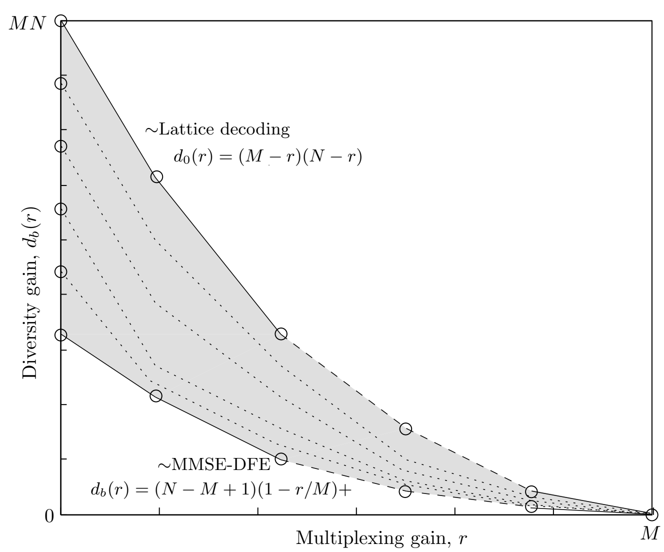
Figure 4.1: diversity-multiplexing tradeoff curves $d_b(r)$ achieved by lattice Fano/Stack sequential decoder for the case of 2×2 MIMO channel for different values of (ζ_1, ζ_2).

It is clear from the above analysis that by varying ζ_i and correspondingly varying b, one can fully control the maximum diversity and multiplexing gains achieved by such decoding scheme. Figure. 4.2 shows the achievable diversity-multiplexing tradeoff curves under lattice sequential decoding for all possible values of ζ_i that satisfy the constraint $\sum_{i=1}^{M} \zeta_i = M$. The figures include both *Case 1* and *Case 2*.

Following the footsteps of [23], we are now ready to prove the following theorem:



(a) Diversity-multiplexing tradeoff curves correspond to *Case* 1 in Theorem 4.



(b) Diversity-multiplexing tradeoff curves correspond to *Case* 2

Figure 4.2: Diversity-multiplexing tradeoff curves $d_b(r)$ achieved by lattice Fano/Stack sequential decoder for different bias b.

Theorem 7. *There exists a sequence of full-dimensional LAST codes with block length $T \geq M + N - 1$ that achieves the diversity-multiplexing tradeoff curve $d_b(r)$ under LAST coding and MMSE-DFE lattice Fano/Stack sequential decoding with variable bias term b that is given in (4.20).*

Proof. See Appendix D. □

4.4.2 Improving the Achievable Rate

It is clear from (4.9) that lattice sequential decoders suffer from very poor performance as b becomes large (achievable rate R_b could reach 0!). The question that may arise here is whether the achievable rate of the decoder can be improved especially for large values of b (for which low decoding complexity is to be expected [39]) and hence improving the error performance.

It turns out that the way the nodes are generated in the algorithm plays an important role in improving both the achievable rate and performance of the decoder without increasing the decoding complexity. For example, Schnorr-Euchner enumeration is considered a good candidate for the use in lattice Fano/Stack sequential decoding algorithms [39]. If the determination of best and next best nodes in the lattice Fano/Stack sequential decoder is based on the Schnorr-Euchner search strategy, then as $b \to \infty$ the decoder reduces to the MMSE-DFE decoder [39], which achieves diversity-multiplexing tradeoff $(N - M + 1)(1 - r/M)^+$ [35].

Corollary 1. *For a fixed non-random channel matrix $\boldsymbol{H}^c$, the rate*

$$R_b(\boldsymbol{H}^c, \rho) \triangleq \max\left\{ R_{\text{mod}}(\boldsymbol{H}^c, \rho) - 2M \log\left(\frac{1 + \sqrt{1 + 8\alpha}}{2}\right), R_{\text{MMSE-DFE}}(\boldsymbol{H}^c, \rho) \right\}, \tag{4.30}$$

is achievable by LAST coding and MMSE-DFE lattice Fano/Stack sequential decoding constructed under the Schnorr-Euchner search strategy, where $R_{\text{MMSE-DFE}}(\boldsymbol{H}^c, \rho)$ is the achievable rate of the MMSE-DFE decoder, and α is as defined in (4.10).

In what follows, we discuss some interesting results about low computational complexity receivers.

4.4.3 MMSE-like Receivers: Large N Analysis

The main role of the bias term b used in the algorithm is to control the amount of computations performed by the decoder. The computational complexity of the lattice sequential decoder is defined as the total number of nodes visited by the decoder during the search. It has been shown in [39] via simulation, that there exists a value of b, say b^*, such that for all $b \geq b^*$, the computational complexity decreases monotonically with b. As $b \to \infty$, the number of visited nodes is always equal to m (computational complexity of MMSE-DFE decoder). In what follows, we discuss a very interesting result.

It is clear from the above analysis that increasing the bias b can affect both diversity and multiplexing gains achieved by such a decoding scheme. However, we would like to show that at $r = 0$ (i.e., at fixed rate R), there exists a lattice sequential decoding algorithm that can simultaneously achieve computational complexity m and maximum diversity $d = MN$.

Consider the bias term given in (4.20) with $\eta(\boldsymbol{\lambda}, \rho) = \prod_{i=1}^{M}(1+\rho\lambda_i)^{\zeta_i}$ where the coefficients $0 < \zeta_i < 1$ are chosen according to *Case 1* such that $\zeta_i = \epsilon$ for all i. In this case, as $\rho \to \infty$, it can be easily verified that $b \doteq \rho^{\frac{(1-\epsilon)}{M}\sum_{i=1}^{M}(1-\alpha_i)^+}$. The probability that b exceeds $\rho^{\kappa/M}$, for $0 < \kappa < M$, can be evaluated as follows:

$$\begin{aligned}\Pr(b \geq \rho^{\kappa/M}) &\doteq \Pr\left((1-\epsilon)\sum_{i=1}^{M}(1-\alpha_i)^+ \geq \kappa\right) = 1 - \Pr\left(\sum_{i=1}^{M}(1-\alpha_i)^+ < \frac{\kappa}{(1-\epsilon)}\right) \\ &\doteq 1 - \rho^{-\left(N-\frac{\kappa}{(1-\epsilon)}\right)^+\left(M-\frac{\kappa}{(1-\epsilon)}\right)^+}.\end{aligned}$$

It is clearly seen that, as N becomes large, with probability close to 1 the bias term $b \to \infty$ as $\rho \to \infty$. Therefore, for such choice of $\eta(\boldsymbol{\lambda}, \rho)$, at high SNR we can achieve *linear* computational complexity but at the expense of losing the optimal tradeoff. However, as argued in the proof of Theorem 4, at $r = 0$ we have $d = MN$. Therefore, as $\rho \to \infty$, linear computational complexity m and maximum diversity gain MN can be achieved simultaneously for large values of N. We can conclude that there exists a lattice sequential decoding algorithm that achieves ML decoder's diversity gain, MN, at $r = 0$ (fixed rate R) when $N \to \infty$.

4.5 Computational Complexity: Tail Distribution in the High SNR Regime

Lattice sequential decoders are constructed as an alternative to sphere decoders (or equivalently lattice decoders) to solve the CLPS problem with much lower computational complexity. Due to the random nature of the channel matrix and the additive noise, the computational complexity of both decoders is considered difficult to analyze in general. As such, most of the work related to such analysis has been performed via first and second order statistics of complexity [52],[53],[54]. However, in their work [55], Seethaler *et. al.* took a different path and analyzed sphere decoder through its complexity tail distribution defined as $\Pr(C \geq L)$, where C is the total number of computations performed by the decoder and L is the distribution parameter. This approach follows naturally from the randomness of the computational complexity of such decoding scheme. It has been shown in [55] that, for large L (i.e., as $L \to \infty$), the complexity distribution of sphere decoder is of a Pareto-type that is given by $L^{-(N-M+1)}$.

As discussed earlier, the bias term b is responsible for the performance-complexity tradeoff achieved by the lattice sequential decoders [39]. For example, setting $b = 0$, we achieve the best performance (performance of sphere decoder) but at the expense of very large decoding complexity. On the other extreme, setting $b = \infty$, lattice sequential decoder that uses Schnorr-Euchner enumeration becomes equivalent to the MMSE-DFE decoder. Although it achieves very low decoding complexity, it suffers from poor performance. In our work, we consider the case of fixed (finite) b. It turns out that for fixed but not large values of b, the complexity distribution's tail exponent $e(r)$ defined by

$$e(r) = \lim_{\rho \to \infty} \frac{-\log \Pr(C \geq L)}{\log \rho},$$

is asymptotically lower bounded by the diversity-multiplexing tradeoff achieved by the LAST coding and sequential decoding schemes, i.e., $e(r) \geq d_{\text{out}}(r)$, and does not depend on the bias term at the high SNR regime. However, increasing the value of b could significantly lower the computational complexity (e.g., as $b \to \infty$, $\Pr(C > L) = 0$ for $L \geq m$) but at the expense of great loss in the achievable diversity-multiplexing tradeoff.

In what follows, we consider only lattice codes that are diversity-multiplexing tradeoff optimal. Also, for the sake of simplicity we consider the Stack algorithm in analyzing the decoder's computational complexity. It must be noted that the following analysis is *only* valid for finite but small values of b.

4.5.1 Naive Lattice Sequential Decoding

In this section, we would like to analyze the computational complexity of the *naive* lattice Stack sequential decoder with bias term $b > 0$, particularly at the high SNR regime. We are interested in bounding the tail distribution of the decoder's computational complexity at high SNR.

Theorem 8. *The asymptotic computational complexity distribution of the naive lattice sequential decoder in an $M \times N$ LAST coded MIMO channel with codeword length $T \geq N + M - 1$, is upper bounded by the asymptotic outage probability, i.e.,*

$$\Pr(C \geq L) \dot{\leq} \rho^{-d(r)}, \tag{4.31}$$

for all L that satisfy

$$L \geq m + \sum_{k=1}^{m} \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MT(1+\log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}, \tag{4.32}$$

where $\boldsymbol{R}_{kk}$ is the lower $k \times k$ part of $\boldsymbol{R} = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{H}\boldsymbol{G}$, and $d(r)$ is given in (4.5).

Proof. The input to the decoder, after QR preprocessing ($\boldsymbol{HG} = \boldsymbol{QR}$) of (2.5), is given by $\boldsymbol{y}' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{Rz} + \boldsymbol{e}'$, where $\boldsymbol{e}' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{e}$. Let $\mu_{\min} = \min\{0, b - \|\boldsymbol{e}'^{1}_{1}\|^2, 2b - \|\boldsymbol{e}'^{2}_{1}\|^2, \ldots, bm - \|\boldsymbol{e}'^{m}_{1}\|^2\}$ be the minimum metric that corresponds to the transmitted path. Without loss of generality, we assume that $N \geq M$. Due to lattice symmetry, we assume that the all zero codeword, i.e., $\mathbf{0}$, was transmitted.

First, let

$$C = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k),$$

be a random variable that denotes the total number of visited nodes during the search, where $\phi(\boldsymbol{z}_1^k)$ is the indicator function defined by

$$\phi(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if node } \boldsymbol{z}_1^k \text{ is extended;} \\ 0, & \text{otherwise.} \end{cases}$$

In this case, the computational complexity tail distribution can be expressed as $\Pr(C \geq L)$, where L is the distribution parameter. Now, a node at level k, i.e., $\boldsymbol{z}_1^k$, may be extended by the Stack decoder if $\mu(\boldsymbol{z}_1^k) > \mu_{\min}$, or equivalently, if $\|\boldsymbol{e}'^{k}_{1} - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk - \mu_{\min}$. The difficulty in analyzing the computational complexity of the lattice Stack sequential decoder stems from the fact that the distribution of the partial matrix $\boldsymbol{R}_{kk}$ is hard to obtain in general. Another factor that may complicate the analysis is $\mu_{\min}$ which is a noise dependent term. However, we can simplify the analysis by considering the following.

First, the complexity tail distribution can be upper bounded as

$$\Pr(C \geq L) \leq \Pr(C \geq L, \|\boldsymbol{e}'\|^2 \leq \alpha^2) + \Pr(\|\boldsymbol{e}'\|^2 > \alpha^2). \tag{4.33}$$

where $\alpha > 0$.

Next, we would like to further upper bound the second term in the RHS of (4.33). We can first write $\phi(\boldsymbol{z}_1^k)$ as

$$\phi(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{e'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk - \mu_{\min}; \\ 0, & \text{otherwise,} \end{cases}$$

Given $\|\boldsymbol{e}'\|^2 \leq \alpha^2$, and by noticing that $-(\mu_{\min} + \|\boldsymbol{e}'\|^2) \leq 0$, we obtain

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k) \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \tag{4.34}$$

where

$$\phi^{'}(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{e'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq bmk + \alpha^2; \\ 0, & \text{otherwise.} \end{cases} \tag{4.35}$$

Now, let

$$\phi_k^{''}(\boldsymbol{z}) = \begin{cases} S_k, & \text{if } \|\boldsymbol{e}' - \boldsymbol{R}\boldsymbol{z}\|^2 \leq bm - \mu_{\min}; \\ 0, & \text{otherwise,} \end{cases}$$

where

$$S_k = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \tag{4.36}$$

then it can be easily shown that

$$C \leq \sum_{k=1}^{m} \sum_{\boldsymbol{z} \in \mathbb{Z}^m} \phi_k^{''}(\boldsymbol{z}) \leq \sum_{k=1}^{m} \sum_{\boldsymbol{x} \in \Lambda_c} \tilde{\phi}_k(\boldsymbol{x}),$$

where

$$\tilde{\phi}_k(\boldsymbol{x}) = \begin{cases} S_k, & \text{if } \|\boldsymbol{H}\boldsymbol{x}\|^2 - 2(\boldsymbol{H}\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e} \leq bm; \\ 0, & \text{otherwise,} \end{cases}.$$

Notice the independence of the above upper bound on $\mu_{\min}$. Consider now the following lemma:

Lemma 4. *In lattice Stack sequential decoder with finite bias $b > 0$, the number of visited nodes at level k, given that $\|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)$, can be upper bounded by*

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k) \leq S_k \leq \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MT(1+\log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}, \tag{4.37}$$

where S_k is as defined in (4.36).

Proof. The proof is similar to the proof of Lemma 3.16 in Chapter 3 with $R_s = \alpha$. □

For a given lattice Λ_c, we have

$$\begin{aligned}\Pr(C \geq L|\Lambda_c, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)) &\leq \Pr(\tilde{C} \geq L-m|\Lambda_c, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)) \\ &\leq \frac{\mathsf{E}_{\boldsymbol{e}'}\{\tilde{C}|\Lambda_c, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)\}}{L-m}, \quad \text{for } L > m,\end{aligned} \tag{4.38}$$

where the last inequality follows from using Markov inequality, and $\tilde{C}$ is defined as

$$\tilde{C} = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k \setminus \{\boldsymbol{0}\}} \phi(\boldsymbol{z}_1^k),$$

since we have assumed that the all-zero lattice point was transmitted.

The conditional average of $\tilde{C}$ with respect to the noise can be further upper bounded as

$$\mathsf{E}_{\boldsymbol{e}'}\{\tilde{C}|\Lambda_c, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)\} \leq \sum_{k=1}^{m} S_k \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{H}\boldsymbol{x}\|^2 - 2(\boldsymbol{H}\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e} < bm) \tag{4.39}$$

Therefore, we have

$$\Pr(C \geq L|\Lambda_c, \|\boldsymbol{e}\|^2 \leq MT(1+\log\rho)) \leq \frac{\sum_{k=1}^{m} S_k}{L-m} \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{H}\boldsymbol{x}\|^2 - 2(\boldsymbol{H}\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e} < bm). \tag{4.40}$$

Following the proof of Theorem 3 (see Appendix B), and by averaging over the ensemble of random lattices we get, for $L > m + \sum_{k=1}^{m} S_k$

$$\Pr(C \geq L) \dot{\leq} \rho^{-T[M - \sum_{j=1}^{M} \nu_j - r]}. \tag{4.41}$$

Define $\mathcal{A} = \{\boldsymbol{\nu} \in \mathbb{R}_+^M : \nu_1 \geq \cdots \geq \nu_M \geq 0,\ \sum_{i=1}^M \nu_i > M - r\}$. Similar to the outage analysis in Section IV, by separating the event $\{\boldsymbol{\nu} \in \mathcal{A}\}$ from its complement, we obtain:

$$\Pr(C \geq L) \leq \Pr(\boldsymbol{\nu} \in \mathcal{A}) + \Pr(\|\boldsymbol{e}'\|^2 > MT(1+\log\rho)) + \Pr(C \geq L, \boldsymbol{\nu} \in \overline{\mathcal{A}}, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)) \tag{4.42}$$

The behavior of the first term in (4.42) at high SNR is $\rho^{-d(r)}$, where $d(r)$ is as defined in Theorem 1. The second term can be shown to be upper bounded by $\rho^{-d(r)}$ (see [23]). Averaging the third term over the channels in $\overline{\mathcal{A}}$ set, we obtain,

$$\Pr(C \geq L) \dot{\leq} \rho^{-d(r)} + \int_{\overline{\mathcal{A}}} f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \Pr(C \geq L|\boldsymbol{\nu})\, d\boldsymbol{\nu} \dot{\leq} \rho^{-d(r)}, \tag{4.43}$$

for all $L \geq m + \sum_{k=1}^m S_k$, where $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ is the joint probability density function of $\boldsymbol{\nu}$ defined in (B.9). □

4.5.2 MMSE-DFE Lattice Sequential Decoding

It is well-known [32] that employing MMSE-DFE preprocessing at the decoding stage significantly reduces the decoder's computational complexity. In this section, we show how MMSE-DFE significantly improves the tail exponent of the computation complexity distribution of lattice sequential decoding compared to the naive decoder. Again, our goal in this section is to analyze the computational complexity of the MMSE-DFE lattice Stack sequential decoder for fixed but small $b > 0$, particularly at the high SNR regime. We are interested in bounding the tail distribution of the decoder's computational complexity at high SNR.

Theorem 9. The asymptotic computational complexity distribution of the MMSE-DFE lattice sequential decoder in an $M \times N$ LAST coded MIMO channel with codeword length $T \geq N + M - 1$, is upper bounded by the asymptotic outage probability, i.e.,

$$\Pr(C \geq L) \dot{\leq} \rho^{-d^*(r)}, \tag{4.44}$$

for all L that satisfy

$$L \geq m + \sum_{k=1}^{m} \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MT(1+\log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2}}, \tag{4.45}$$

where $\boldsymbol{R}_{kk}$ is the lower $k \times k$ part of $\boldsymbol{R} = \boldsymbol{Q}^\mathsf{T}\boldsymbol{B}\boldsymbol{G}$, and $d^*(r)$ is as defined in Theorem 4.

Proof. The input to the decoder, after QR preprocessing ($\boldsymbol{BG} = \boldsymbol{QR}$) of (2.5), is given by $\boldsymbol{y}'' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{y}' = \boldsymbol{Rz} + \boldsymbol{e}''$, where $\boldsymbol{e}'' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{e}'$. Following the same approach used to prove Theorem 6, the tail distribution can be upper bounded as follows

$$\Pr(C \geq L) \leq \Pr(\boldsymbol{\nu} \in \mathcal{B}) + \Pr(\|\boldsymbol{e}'\|^2 > MT(1+\log\rho)) + \Pr(C \geq L, \boldsymbol{\nu} \in \overline{\mathcal{B}}, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)), \tag{4.46}$$

where the set $\mathcal{B} = \{\boldsymbol{\nu} \in \mathbb{R}_+^M : \nu_1 \geq \cdots \geq \nu_M \geq 0,\ \sum_{i=1}^M (1-\nu_i)^+ < r\}$.

Using Lemma 4 and the Markov inequality, one can show that for a given lattice Λ_c

$$\Pr(C \geq L | \Lambda_c, \|\boldsymbol{e}'\|^2 \leq MT(1+\log\rho)) \leq \frac{1}{L-m}\sum_{k=1}^{m} S_k \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{Bx}\|^2 - 2(\boldsymbol{Bx})^{\mathsf{T}}\boldsymbol{e}' < bm). \tag{4.47}$$

Similar to the proof of Theorem 4, one can easily show that

$$\Pr(C \geq L) \dot{\leq} \rho^{-T[\sum_{j=1}^{\min\{M,N\}}(1-\alpha_j)^+ - r]}. \tag{4.48}$$

for any $L > m + \sum_{k=1}^m S_k$. Now, the behavior of the first term in (4.46) at high SNR is $\rho^{-d^*(r)}$, where $d^*(r)$ is as defined in Theorem 2. Following [23], one can show that the second term is upper bounded by $\rho^{-d^*(r)}$. Averaging the third term over the channels in $\overline{\mathcal{B}}$ set, we obtain,

$$\Pr(C \geq L) \dot{\leq} \rho^{-d^*(r)} + \int_{\overline{\mathcal{B}}} f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \Pr(C \geq L|\boldsymbol{\nu})\, d\boldsymbol{\nu} \dot{\leq} \rho^{-d^*(r)}. \tag{4.49}$$

for all $L \geq m + \sum_{k=1}^m S_k$. □

The above results reveal that if the number of computations performed by the decoder exceeds

$$L_0 = m + \sum_{k=1}^{m} \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MT(1+\log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2}}, \tag{4.50}$$

the complexity distribution of the lattice sequential decoder at high SNR is upper bounded by the asymptotic outage probability. However, the MMSE-DFE lattice sequential decoder exhibits larger SNR exponent than the naive one. This implies that the probability of the complexity being atypically large is smaller when MMSE-DFE is applied prior sequential decoding. Now, if a "time-out" limit is imposed at the decoder to terminate the search when the number of computations exceeds this limit, then L_0 represents the minimum value that should be set by the decoder without resulting in a loss in the optimal performance achieved by such decoding scheme. This can be very beneficial in two-ways MIMO communication systems (e.g, MIMO automatic repeat request [41]), where the feedback channel can be used to eliminate the decoding failure probability. In applications where there is a hard-limit on the buffer size, the decoder declares an error when the complexity goes above the limit.

It is clear that this time-out limit does not affect the optimal tradeoff achieved by the modified decoding scheme. To see this, suppose that the lattice (stack) sequential decoder imposes a time-out limit so that the search is terminated once the number of computations reaches L_0, and hence the decoder declares an error. Assuming E_s is the event that the decoder performs an error when $C < L_0$, in this case, the average error probability is given by

$$P_e(\rho) = \Pr(E_s \cup \{C \geq L_0\}) \leq \Pr(E_s) + \Pr(C \geq L_0) \dot{\leq} \rho^{-d_{\text{out}}(r)}. \tag{4.51}$$

It should be noted that the above analysis does not yield the full picture of the decoder's complexity in general. As mentioned previously, the complexity of the decoder depends critically on the bias b chosen in the algorithm. Unfortunately, it is still unclear how the SNR exponent $e(r)$ is affected by the value b in general. However, as $b \to \infty$, the naive or the MMSE-DFE lattice sequential decoder under Schnorr-Euchner enumeration becomes equivalent to the zero-forcing (ZF-DFE) or the MMSE-DFE decoder, respectively. The total number of computations performed by both decoders is always equal to m. This corresponds to an SNR exponent $e(r) = \infty$. Thus, we can conclude that, at high SNR, as b increases the SNR exponent $e(r)$ increases as well.

Another criterion that is used to characterize the computational complexity of such a decoder is through its average complexity. Since L_0 is random, it would be interesting to calculate the minimum average number of computations required by the decoder to terminate the search. This is considered next.

4.6 Average Computational Complexity

It is to be expected that when the channel is ill-conditioned (i.e., in outage) the computational complexity becomes extremely large. Moreover, when the channel is in outage it is highly likely that the decoder performs an erroneous detection. However, when the channel is *not* in outage, there is still a non-zero probability that the number of computations will become large (see (4.43) and (4.49)). As such, it is sometimes desirable to terminate the search even when the channel is not in outage. Therefore, we would like to determine the *minimum* average number of computations that is required in order for the decoder to determine when to terminate the search.

In other words, we would like to find the minimum average number of computations that is required by the decoder to achieve near-optimal performance. This can be expressed as

$$L_{\text{out}} = \mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\}, \tag{4.52}$$

where $L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})$ denotes the total number of computations performed by the decoder to achieve near optimal performance when the channel is not in outage.

Before we do that, we would like first now to study the asymptotic behavior of L_0. Again, in this section we focus our analysis on nested LAST codes, specifically LAST codes that are generated using construction A (see Chapter 2, Section 2.2).

We consider the Loeliger ensemble of mod-p lattices, where p is a prime. First, we generate the set of all lattices given by

$$\Lambda_p = \kappa(\mathsf{C} + p\mathbb{Z}^{2MT})$$

where $p \to \infty$, $\kappa \to 0$ is a scaling coefficient chosen such that the fundamental volume $V_f = \kappa^{2MT} p^{2MT-1} = 1$, $\mathbb{Z}_p$ denotes the field of mod-p integers, and $\mathsf{C} \subset \mathbb{Z}_p^{2MT}$ is a linear code over $\mathbb{Z}_p$ with generator matrix in systematic form $[\boldsymbol{I}\ \boldsymbol{P}^{\mathsf{T}}]^{\mathsf{T}}$. We use a pair of self-similar lattices for nesting. We take the shaping lattice to be $\Lambda_s = \phi\Lambda_p$, where ϕ is chosen such that the covering radius is $1/2$ in order to satisfy the input power constraint. Finally, the coding lattice is obtained as $\Lambda_c = \rho^{-r/2M}\Lambda_s$. Interestingly, one can construct a generator matrix of Λ_p as (see [6])

$$\boldsymbol{G}_p = \kappa \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{P} & p\boldsymbol{I} \end{pmatrix}, \tag{4.53}$$

which has a lower triangular form. In this case, one can express the generator matrix of Λ_c as $\boldsymbol{G} = \rho^{-r/2M}\boldsymbol{G}'$, where $\boldsymbol{G}' = \zeta\boldsymbol{G}_p$. Thanks to the lower triangular format of $\boldsymbol{G}$. If $\boldsymbol{M}$ is an $m \times m$ arbitrary full-rank matrix, and $\boldsymbol{G}$ is an $m \times m$ lower triangular matrix, then one can easily show that

$$\det[(\boldsymbol{M}\boldsymbol{G})_{kk}] = \det(\boldsymbol{M}_{kk})\det(\boldsymbol{G}_{kk}), \tag{4.54}$$

where $(\boldsymbol{M}\boldsymbol{G})_{kk}$, $\boldsymbol{M}_{kk}$, and $\boldsymbol{G}_{kk}$, are the lower $k \times k$ part of $\boldsymbol{M}\boldsymbol{G}$, $\boldsymbol{M}$, and $\boldsymbol{G}$, respectively[5]..

[5]For the sake of simplicity we consider the complexity analysis of naive decoding for a square MIMO system (i.e., for $N = M$). This is due to the fact that for a general $M \times N$ MIMO system under naive decoding, the determinant (4.54) does not exist (except when $N = M$). In this case, one may lower bound $\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) = \prod_{j=1}^{k} \lambda_j(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) \geq \prod_{j=1}^{k} \lambda_j(\boldsymbol{H}_{kk}^{\mathsf{T}}\boldsymbol{H}_{kk})\lambda_j(\boldsymbol{G}_{kk}^{\mathsf{T}}\boldsymbol{G}_{kk})$ to get only an upper bound on L_0. For the case of MMSE-DFE decoding, the combined channel code matrix is always a square matrix.

Using the above result, one can express the determinant that appears in (4.50) as

$$\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk}) = \det(\boldsymbol{M}_{kk}^{\mathsf{T}}\boldsymbol{M}_{kk})\det(\boldsymbol{G}_{kk}^{\mathsf{T}}\boldsymbol{G}_{kk}) = \rho^{-rk/2M}\det(\boldsymbol{M}_{kk}^{\mathsf{T}}\boldsymbol{M}_{kk})\det(\boldsymbol{G'}_{kk}^{\mathsf{T}}\boldsymbol{G'}_{kk}) \quad (4.55)$$

where $\boldsymbol{M}$ is either $\boldsymbol{B}$ or $\boldsymbol{H}$, depending whether the decoder is preprocessed with MMSE-DFE or not. Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ be the ordered nonzero eigenvalues of $\boldsymbol{M}_{kk}^{\mathsf{T}}\boldsymbol{M}_{kk}$, for $k = 1, \cdots, m$. Then,

$$\det(\boldsymbol{M}_{kk}^{\mathsf{T}}\boldsymbol{M}_{kk}) = \prod_{j=1}^{k}\mu_j$$

Note that for the special case when $k = m$ we have $\mu_{2(j-1)T+1} = \cdots = \mu_{2jT} = \rho\lambda_j((\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c)$, for all $j = 1, \cdots, M$ when $\boldsymbol{M} = \boldsymbol{H}$. When $\boldsymbol{M} = \boldsymbol{B}$ we have $\mu_{2(j-1)T+1} = \cdots = \mu_{2jT} = 1 + \rho\lambda_j((\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c)$, for all $j = 1, \cdots, M$.

Denote $\alpha_i' = -\log\mu_i/\log\rho$. Using (4.54), one can asymptotically upper bound L_0 as

$$L_0 = m + \sum_{k=1}^{m}(\log\rho)^{k/2}\rho^{c_k}, \quad (4.56)$$

where

$$c_k = \frac{1}{2}\sum_{j=1}^{k}\left(\frac{r}{M} - \alpha_j'\right)^+. \quad (4.57)$$

Now, since c_k is non-decreasing in k, we have

$$L_0 = m + (\log\rho)^{m/2}\rho^{c_m}, \quad (4.58)$$

where

$$c_m = \begin{cases} T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^+, & \text{for } \boldsymbol{M} = \boldsymbol{H}; \\ T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)^+\right)^+, & \text{for } \boldsymbol{M} = \boldsymbol{B}; \end{cases}$$

Consider the case of MMSE-DFE lattice decoding. The average of L_0 (averaged over channel statistics) when the channel is not in outage is given by

$$\begin{aligned}\mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\} &= \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} L_0 f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \\ &= m + (\log\rho)^{m/2} \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} \exp\Bigg(\log\rho\Bigg[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)^+\right)^+ - \\ &\qquad \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\Bigg]\Bigg)\, d\boldsymbol{\alpha} \\ &= m + (\log\rho)^{m/2}\rho^{l_{\text{mmse}}(r)},\end{aligned}$$

where $\overline{\mathcal{O}} = \left\{\boldsymbol{\alpha}\in\mathbb{R}_+^M : \sum_{i=1}^{M}(1-\alpha_i)^+ \geq r\right\}$, and

$$l_{\text{mmse}}(r) = \max_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}}\left[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)^+\right)^+ - \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\right]. \tag{4.59}$$

It is not so difficult to see that the optimal channel coefficients that maximize (4.59) are

$$\alpha_i^* = 1, \quad \text{for } i = 1,\cdots,M-k,$$

and

$$\alpha_i^* = 0, \quad \text{for } i = M-k+1,\cdots,M,$$

i.e., the same $\boldsymbol{\alpha}^*$ that achieves the optimal diversity-multiplexing tradeoff of the channel. Substituting $\boldsymbol{\alpha}^*$ in (4.59), we get

$$l_{\text{mmse}}(r) = \frac{Tr(M-r)}{M} - (M-r)(N-r), \tag{4.60}$$

for $r = 0,1,\cdots,M$. In this case, the asymptotic minimum average computational complexity, when the channel is not in outage, can be upper bounded as

$$L_{\text{out}}^{\text{mmse}} = 2MT + (\log\rho)^{MT}\rho^{l_{\text{mmse}}(r)}. \tag{4.61}$$

Similarly, the above analysis can be applied to the case of naive lattice decoding, where the average of L_0 (averaged over channel statistics) when the channel is not in outage is given by

$$
\begin{aligned}
\mathsf{E}\{L_0(\boldsymbol{H}^c \in \overline{\mathcal{O}})\} &= \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} L_0 f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \\
&= m + (\log\rho)^{m/2} \int\limits_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}} \exp\Bigg(\log\rho\Bigg[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^{+} - \\
&\qquad \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\Bigg]\Bigg)\, d\boldsymbol{\alpha} \\
&= m + (\log\rho)^{m/2}\rho^{l_{\text{naive}}(r)},
\end{aligned}
$$

where $\overline{\mathcal{O}} = \left\{\boldsymbol{\alpha} \in \mathbb{R}_{+}^{M} : \sum_{i=1}^{M}\alpha_i \leq M - r\right\}$, and

$$
l_{\text{naive}}(r) = \max_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}}\left[T\sum_{i=1}^{M}\left(\frac{r}{M} - (1-\alpha_i)\right)^{+} - \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\right]. \tag{4.62}
$$

In this case, one can show that when the channel is in outage we have that the optimal $\boldsymbol{\alpha}$ that maximizes (4.62) is achieved for $\alpha_1 = M - r$, and $\alpha_i = 0$ for all $i > 1$, yielding

$$
l_{\text{naive}}(r) = \frac{T(M-1)}{M}(M-r) - (N-M+1)(M-r), \tag{4.63}
$$

for $r = 0, 1, \cdots, M$. In this case, the asymptotic minimum average computational complexity can be upper bounded as

$$
L_{\text{out}}^{\text{naive}} \leq 2MT + (\log\rho)^{MT}\rho^{l_{\text{naive}}(r)}. \tag{4.64}
$$

To see the advantage of using the MMSE-DFE prior decoding that results in a huge saving in (average) computational complexity compared to the naive decoder, consider the case of $M = N$. Assuming the use of an optimal random nested LAST code of codeword length T and fixed rate R, i.e., $r = 0$. In this case, one can see that $l_{\text{mmse}}(0) < 0$ irrespective to the value of T, i.e., the average complexity is bounded for all T. It is clear that the term $(\log\rho)^{2MT}\rho^{-NM}$ decays quickly to 0 as $\rho \to \infty$. The simulation results (introduced next) agree with the above analysis.

For the case of naive decoding we have $l_{\text{naive}}(0) = T(M-1) - M$ which makes the RHS of (4.64) unbounded, except for the case when $T = 1$. This can be used as an indication of what values of T may result in large average complexity. However, for the case that corresponds to $T = M = 2$, although it becomes unbounded, the upper bound grows slowly with SNR as $(\log \rho)^4$. For $T > 2$, the upper bound grows quickly with SNR as $(\log \rho)^{MT} \rho^{T(M-1)-M}$. Since, the maximum diversity $d = \min\{T, N-M+1\}M$ is achievable for all values of $T \geq 1$, one should avoid using codeword lengths $T > 2$ to save on complexity. The experimental results (provided in the next section) demonstrate such improvements and agrees with the above theoretical results. In general, at any multiplexing gain r, we have that $l_{\text{mmse}}(r) > l_{\text{naive}}(r)$, for the same codeword length T. This again proves that employing MMSE-DFE preprocessing at the decoding stage significantly improves the average computational complexity of the decoder at all multiplexing gains.

It is interesting to note that, for the case of MMSE-DFE lattice decoding, there exists a *cut-off* multiplexing gain, say r_0, such that the average computational complexity of the decoder remains bounded as long as we operate below such value. This value can be easily found by setting $l_{\text{mmse}}(r_0) = 0$. This results in

$$r_0 = \left\lfloor \frac{MN}{M+T} \right\rfloor.$$

If we let the number of receive antennas $N \to \infty$, then one can achieve a multiplexing gain $r_0 = M$ which is the maximum multiplexing gain achieved by the channel. As discussed in Section 4.4.3, this again shows that one can dramatically improve the computational complexity of the decoder by increasing the number of antennas at the receiver side.

To see the great advantage of using the lattice sequential decoder with constant bias term over the lattice decoder implemented via sphere decoding algorithms, we compare the average computational complexity of both decoders when MMSE-DFE is presented. It has been shown in Chapter 3 (see (**??**)) that the minimum average number of computations required by the MMSE-DFE sphere decoder to achieve the optimal tradeoff, say $L_{\text{sphere}}^{\text{mmse}}$ for a system with $m = 2MT$ signal dimension is given by (assuming fixed rate $r = 0$)

$$L_{\text{sphere}}^{\text{mmse}} = 2MT + \frac{(\log \rho)^{2MT}}{\rho^{MN}}.$$

The ratio of the average complexity of both decoders, say γ, is given by

$$\gamma = \frac{L_{\text{sphere}}}{L_{\text{sequential}}^{\text{mmse}}} \doteq \frac{2MT + (\log \rho)^{2MT}/\rho^{MN}}{2MT + (\log \rho)^{MT}/\rho^{MN}}.$$

This is a huge saving in computational complexity, especially for large signal dimensions and moderate-to-high SNR. For example, consider the case of a 3×3 LAST coded MIMO system with $T = 5$. At $\rho = 10^3$ (30 dB), we have $\gamma \approx 31$, i.e., the sphere decoder's complexity is about 31 times larger than the complexity of the lattice sequential decoder. As will be shown in the sequel, simulation results agree with the above theoretical results. For $\rho < 30$ dB, one would expect the ratio $\gamma \gg 31$. For extremely high SNR values (e.g., $\rho \gg 30$ dB), it seems that $\gamma \rightarrow 1$ as $\rho \rightarrow \infty$.

4.7 Numerical Results

Throughout the simulation study, the fading coefficients are generated as independent identically distributed circularly symmetric complex Gaussian random variables. The LAST code is obtained as an $(m = 2MT, p, k)$ Loeliger construction (see Chapter 2, Section 2.2).

In Figure. 4.3, we compare the performance in terms of the frame error rate of a MIMO system with $M = N = 2$, $T = 3$ and rate $R = 4$ bits per channel use (bpcu) under naive and MMSE-DFE lattice Fano sequential decoding. For both decoders we fix the bias term to $b = 0.6$. It is clear that the MMSE-DFE lattice sequential decoder outperforms the naive one, where the former achieves diversity order of 4 (the maximum diversity gain achieved by the channel) and the latter achieves diversity order of 2. This validates our theoretical claims for fixed rate (i.e. $r = 0$). To validate the achievability of the optimal diversity-multiplexing tradeoff with LAST coding and MMSE-DFE lattice sequential decoding, we consider the performance of a MIMO system with $M = N = 2$, $T = 3$ for different rates $R = 4,\ 8,\ 10.34$ bpcu, which is illustrated in Figure. 4.4. The constant gap between the outage probability and the error performance for different R confirms our theoretical results.

Figure. 4.5 and Figure. 4.6 show the effect of increasing the bias term on diversity order and average computational complexity (number of visited nodes during the search) achieved by lattice sequential decoding. As discussed earlier, increasing the bias term in the decoding algorithm significantly reduces decoding complexity but at the expense of losing diversity. For the 2×2 LAST coded MIMO system with $T = 3$, as $b \rightarrow \infty$ we achieve linear computational complexity $m = 12$ for all SNR, and diversity order 1. For sequential decoding algorithms that implement the Schnorr-Euchner enumeration, this corresponds to the performance and complexity of MMSE-DFE decoder.

In our computational complexity distribution simulation, we consider a MIMO system with $M = N = 2$, $T = 3$ for different rates $R = 4,\ 8$ bits per channel use. First, the frame error rate of the lattice Fano sequential decoder is plotted in Figure. 4.7. (a) and Figure 4.8. (a) when the bias $b = 0.6$ and the step size $\delta = 0.2$ for both cases, the naive and the MMSE-DFE decoding[6]. The computational complexity distribution $\Pr(C > L)$ is plotted for both decoders at different rates, for sufficiently large L (see Figure 4.7. (b) and Figure 4.8. (b)). It is clear from both figures that the curves which correspond to the error probability and the computational complexity distribution match in slope, i.e., they both exhibit the same behavior at high SNR. In other words, both curves have the same SNR exponent. This basically agrees with the derived theoretical results.

The complexity saving advantage that lattice sequential decoders posses over sphere decoders is depicted in Figure. 4.9 and Figure. 4.10, for the same coded MIMO channel with $R = 4$ bits per channel use. One can notice the amount of computations saved by lattice sequential decoders for all values of SNR, especially for large signal dimensions (see Figure 4.10). Even at high SNR, the sphere decoder still exhibits large decoding complexity compared to the lattice sequential decoder. For example, as depicted in Figure. 4.10, at $\rho = 30$ dB, the average complexity of the sphere decoder is about 30 times the complexity of the lattice sequential decoder for an optimal LAST coded MIMO system with dimension $m = 30$. This is achieved at the expense of small loss in performance ($\sim$1 dB). This agrees with the derived theoretical results.

Figure. 4.11 shows how the average computational complexity is affected by the codeword length T, at a fixed rate ($r = 0$), for the case of naive lattice sequential decoding. In a 2×2 quasi-static MIMO channel under naive lattice sequential decoding, the maximum diversity gain $M = 2$ is achieved when $T \geq 1$. Three random nested LAST codes with codeword lengths $T = 1, 2,$ and 3 are used to achieve the same diversity gain. However, as discussed in the previous section, using a codeword length $T \leq 2$ would result in a small average decoding complexity. For $T = 3$ the average computational complexity becomes extensively large. The complexity saving advantage that the MMSE-DFE pre-processing provides over the naive decoder is also shown in Figure 4.12. It is clear that applying MMSE-DFE prior sequential decoding significantly reduces average computational complexity, especially at high SNR. This agrees with the theoretical results derived in this Chapter.

[6] We have noticed that the simulation under lattice Fano sequential decoding with small values decoding parameters (e.g., $b = 0.6$ and $\delta = 0.2$) yields a decoding complexity tail distribution similar to the sphere decoder one (see Chapter 3, Fig. 3.3 and Fig. 3.4). This can be explained from the fact that the Fano decoder visits a node more than once during the search for the closest lattice point and the number of visits increases as b and δ decreases. Since low values of decoding parameters is needed to achieve close to the sphere decoder performance, one should expect computational cost of sequential decoding close to the sphere decoding one. In general, there is a slight improvement in the complexity tail distribution of the lattice sequential decoder compared to the sphere decoder especially for low-to-moderate values of decoding parameters. The improvement is in the coding gain but no the diversity order.

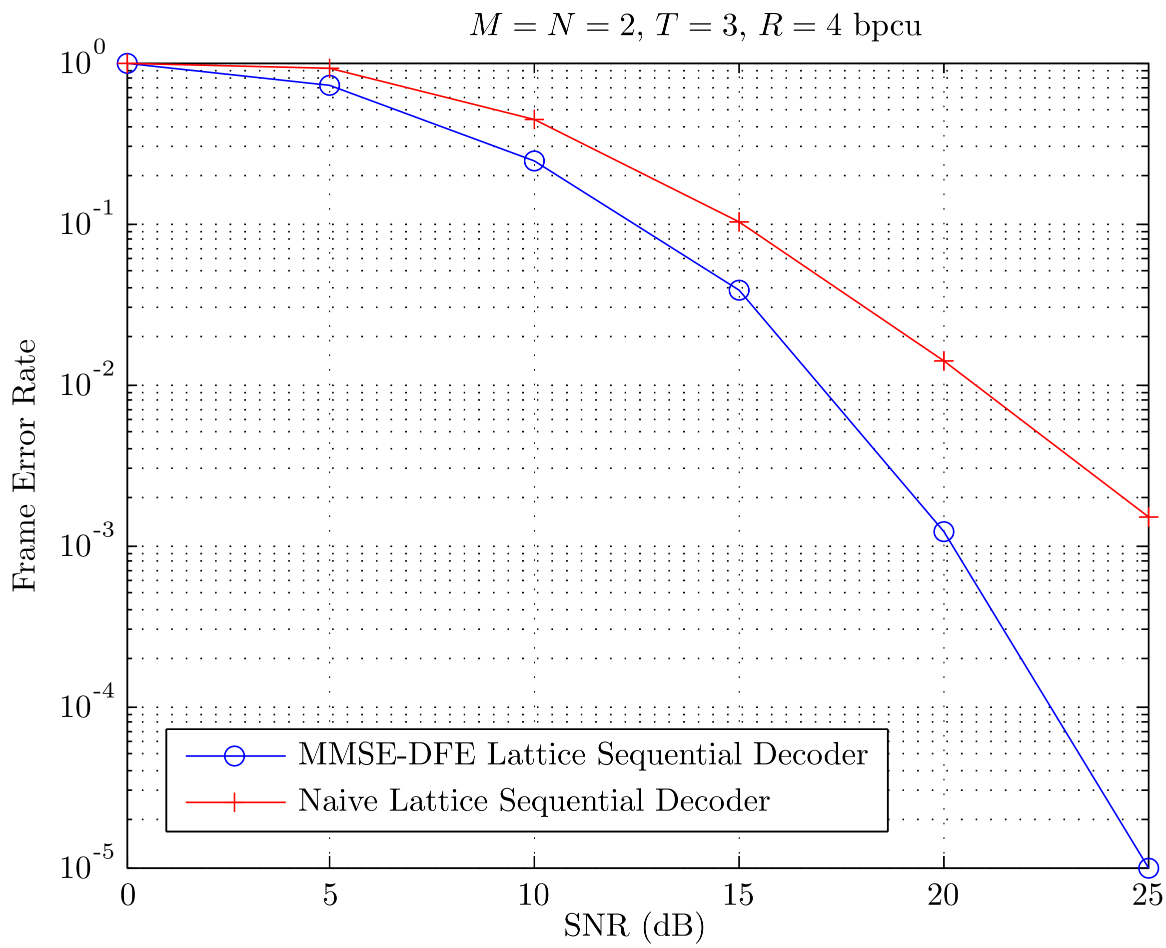


Figure 4.3: Performance comparison between naive and MMSE-DFE lattice sequential decoding with $b = 0.6$ for the case of 2×2 LAST coded MIMO channel with $T = 3$ and $R = 4$ bpcu.

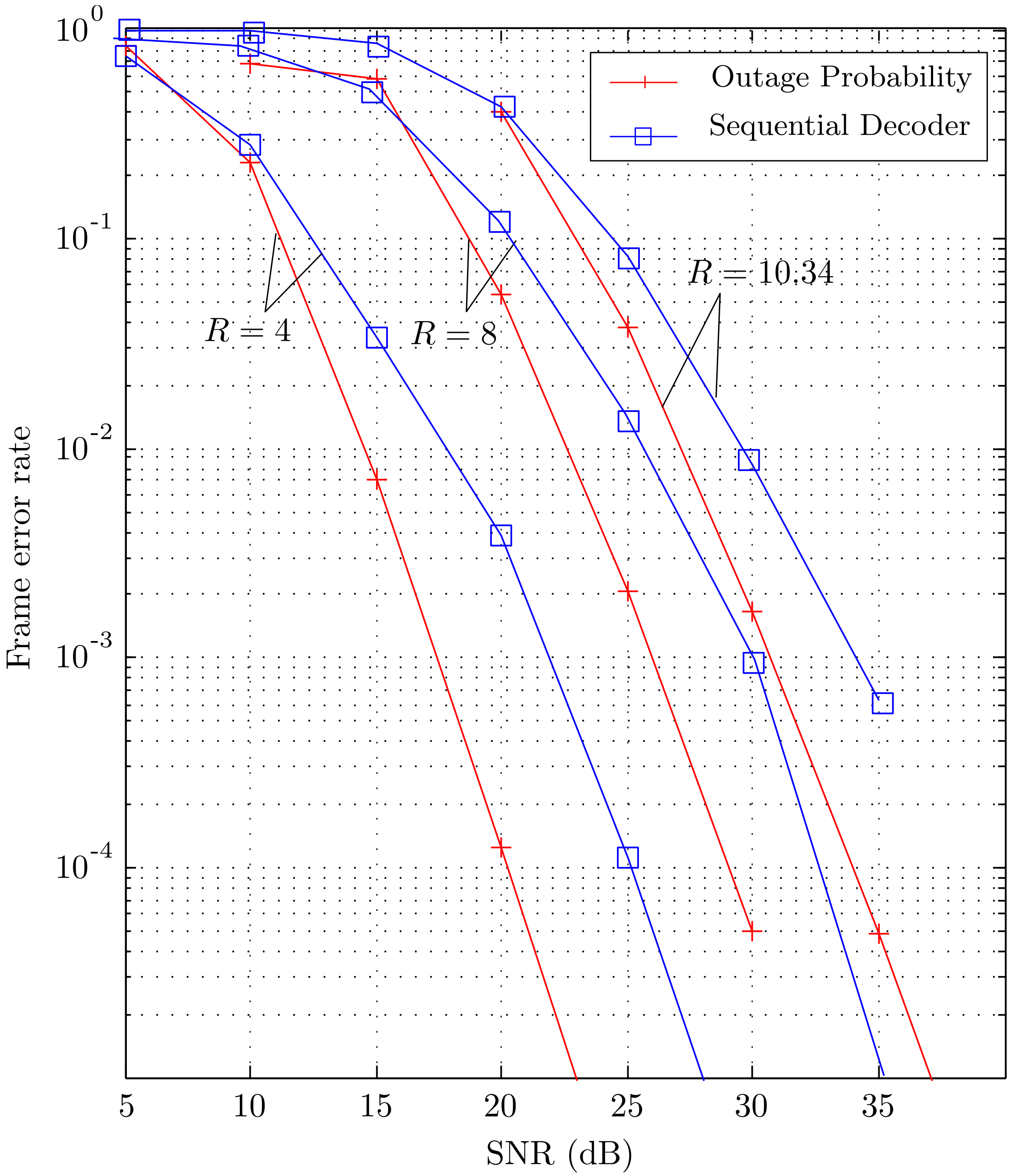


Figure 4.4: Outage probability and error rate performance of lattice sequential decoding with $b = 1$.

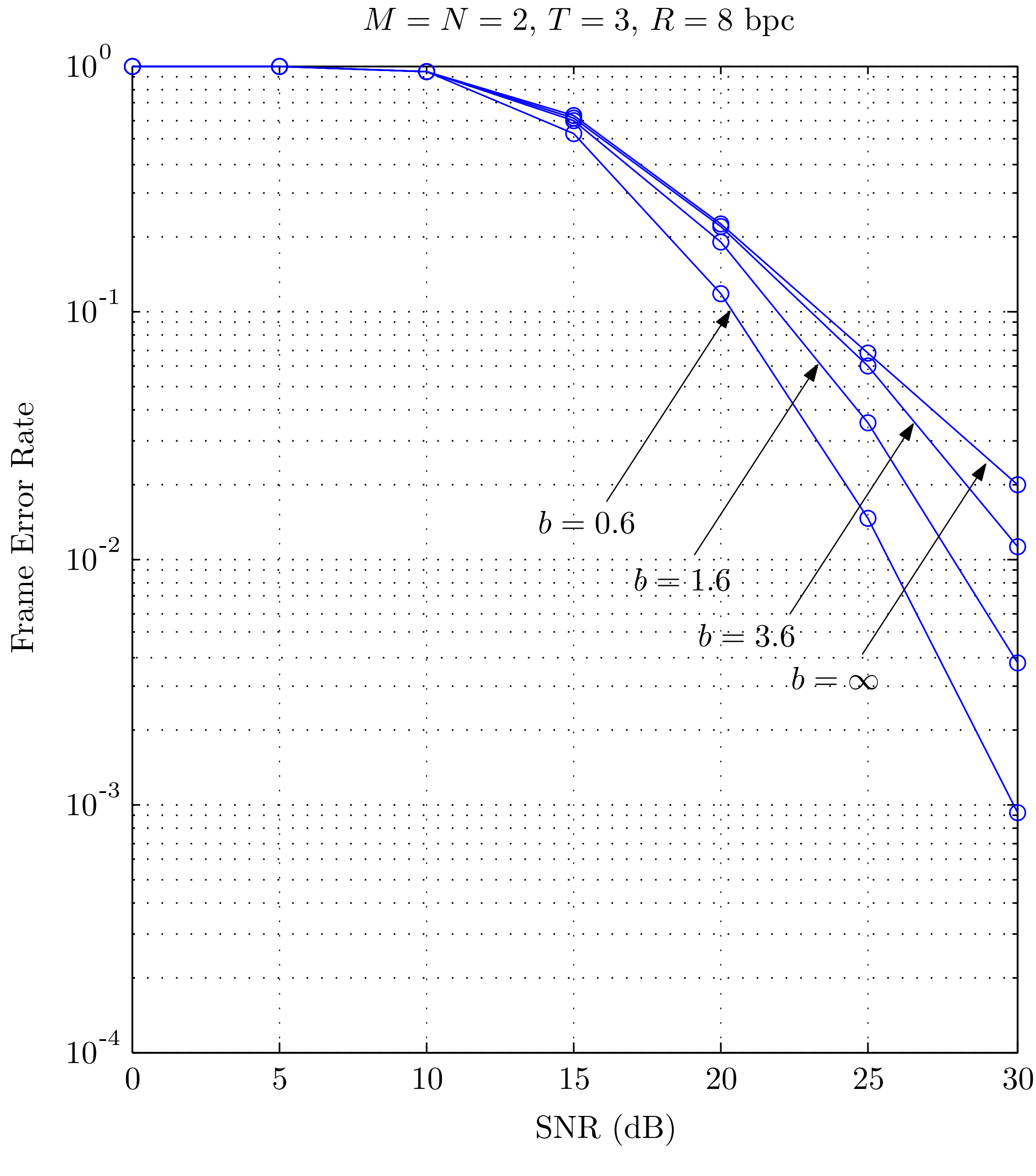


Figure 4.5: Comparison of diversity order achieved by lattice sequential decoding for several values of b.

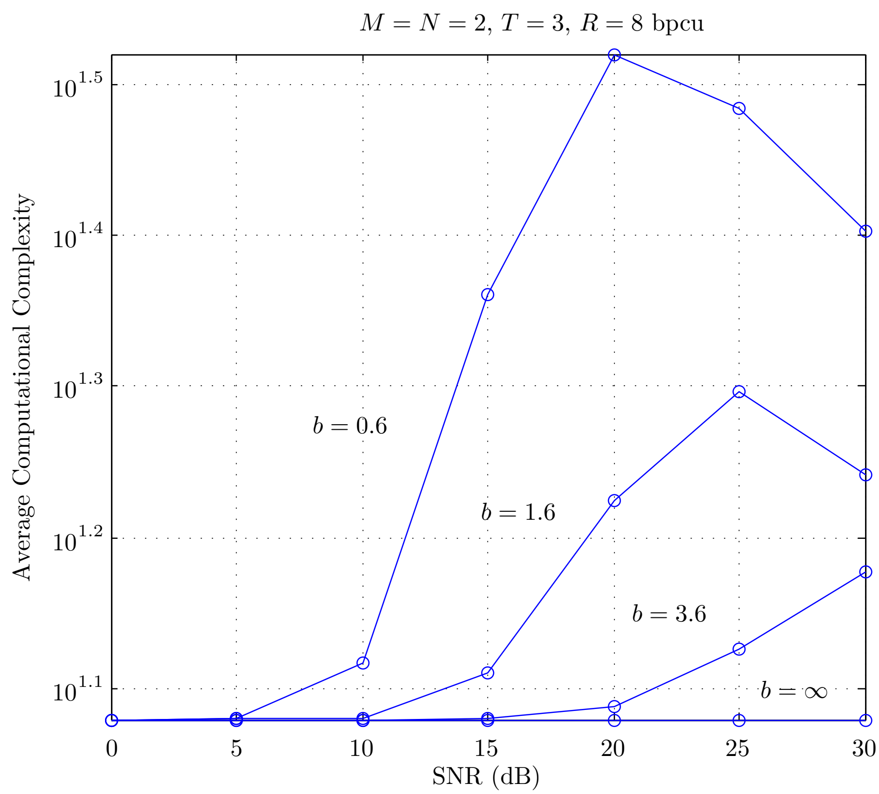


Figure 4.6: Comparison of average computational complexity achieved by lattice sequential decoding for several values of b.

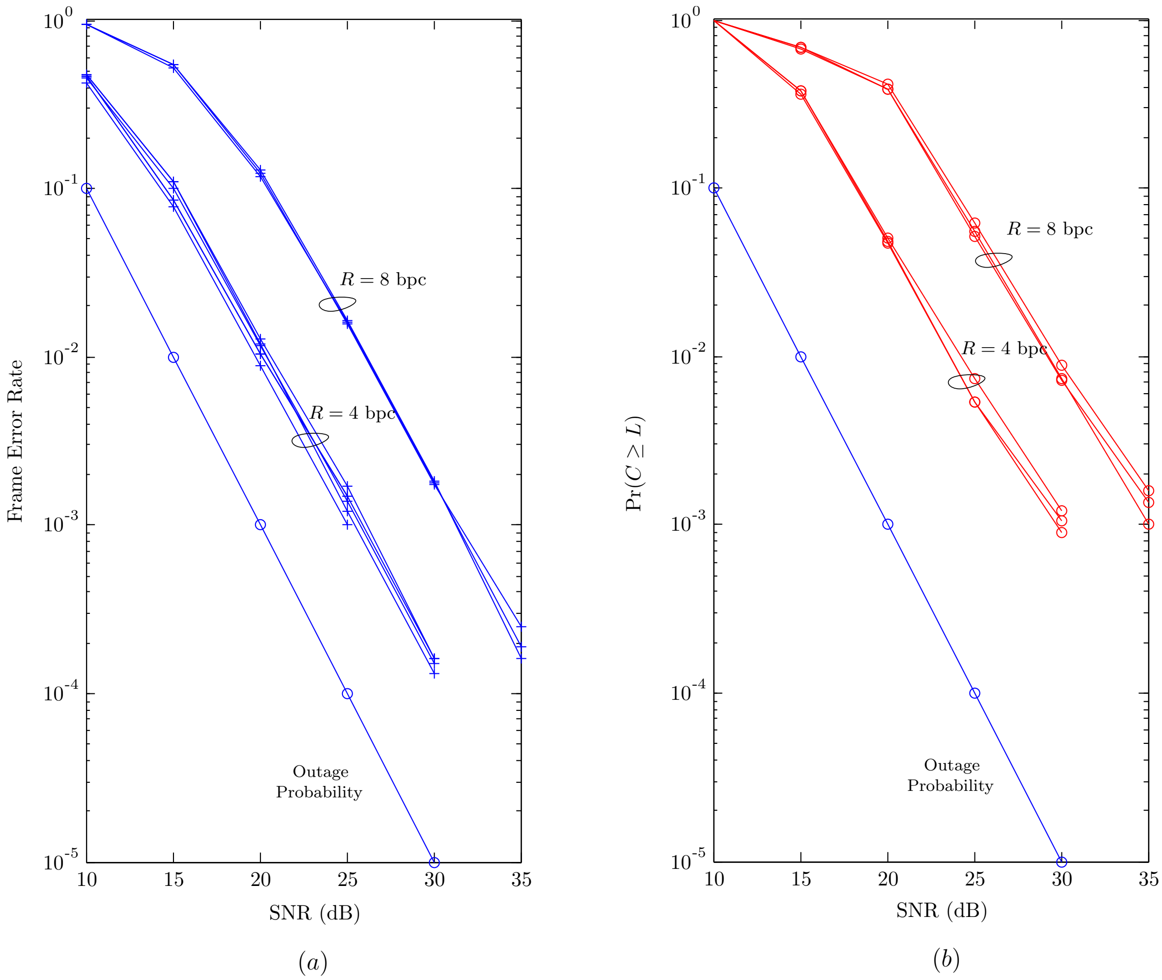
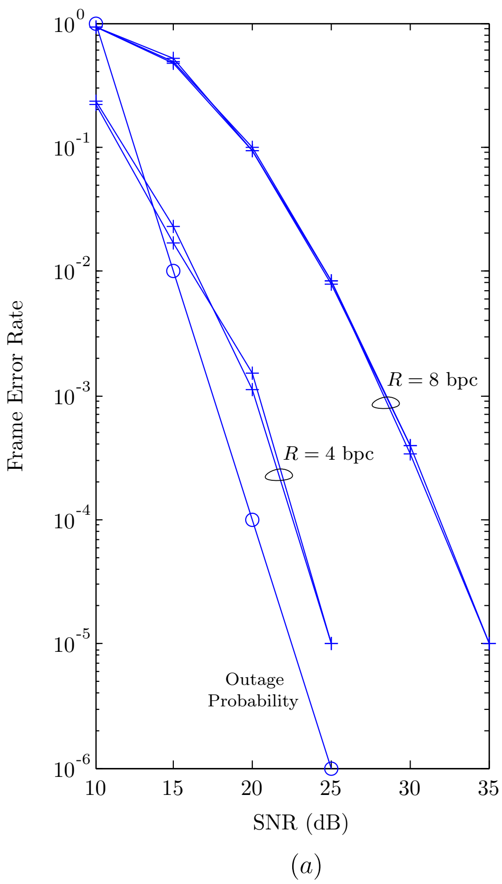


Figure 4.7: (a) Performance and (b) complexity distribution (with $L = \rho$) achieved by the naive lattice Fano sequential decoder ($b = 0.6, \delta = 0.2$) for the case of $2{\times}2$ LAST coded MIMO channel.



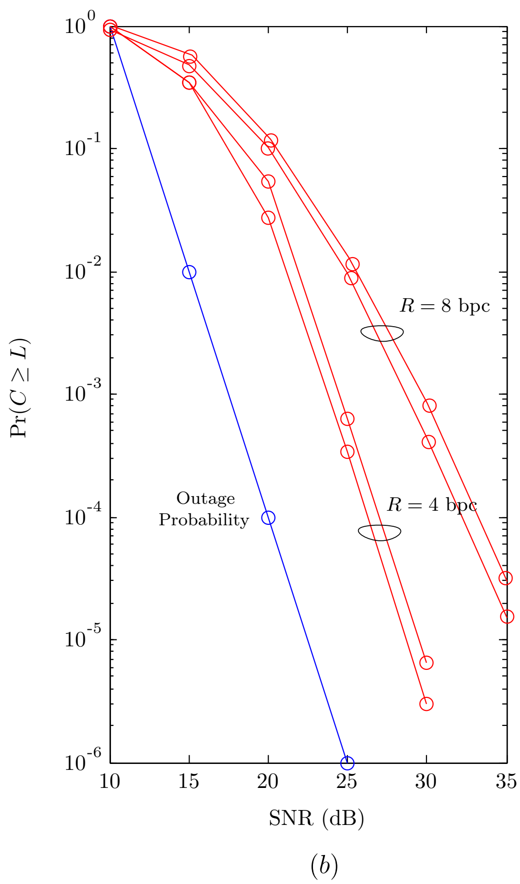


Figure 4.8: (a) Performance and (b) complexity distribution (with $L = \rho$) achieved by the MMSE-DFE lattice sequential decoder ($b = 0.6, \delta = 0.2$) for the case of $2{\times}2$ LAST coded MIMO channel.

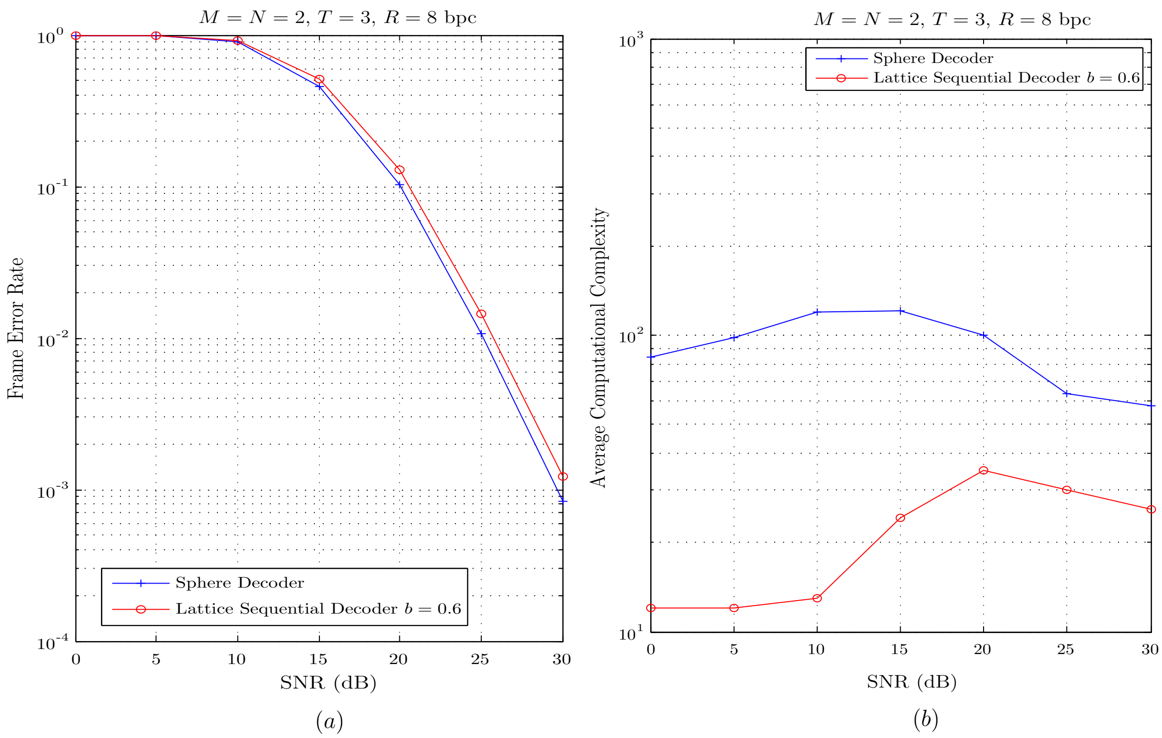


Figure 4.9: (a) Performance and (b) average computational complexity comparison between sphere decoding and lattice sequential decoding for signal with dimension $m = 12$.

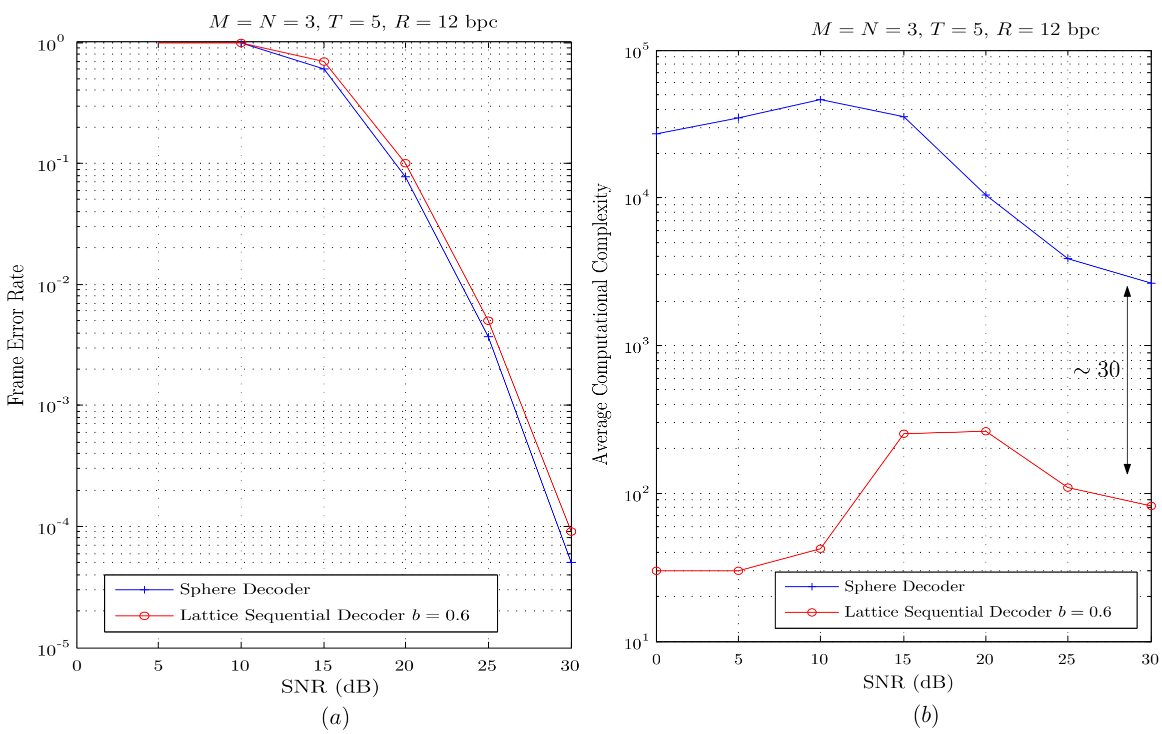


Figure 4.10: (a) Performance and (b) average computational complexity comparison between sphere decoding and lattice sequential decoding for signal with dimension $m = 30$.

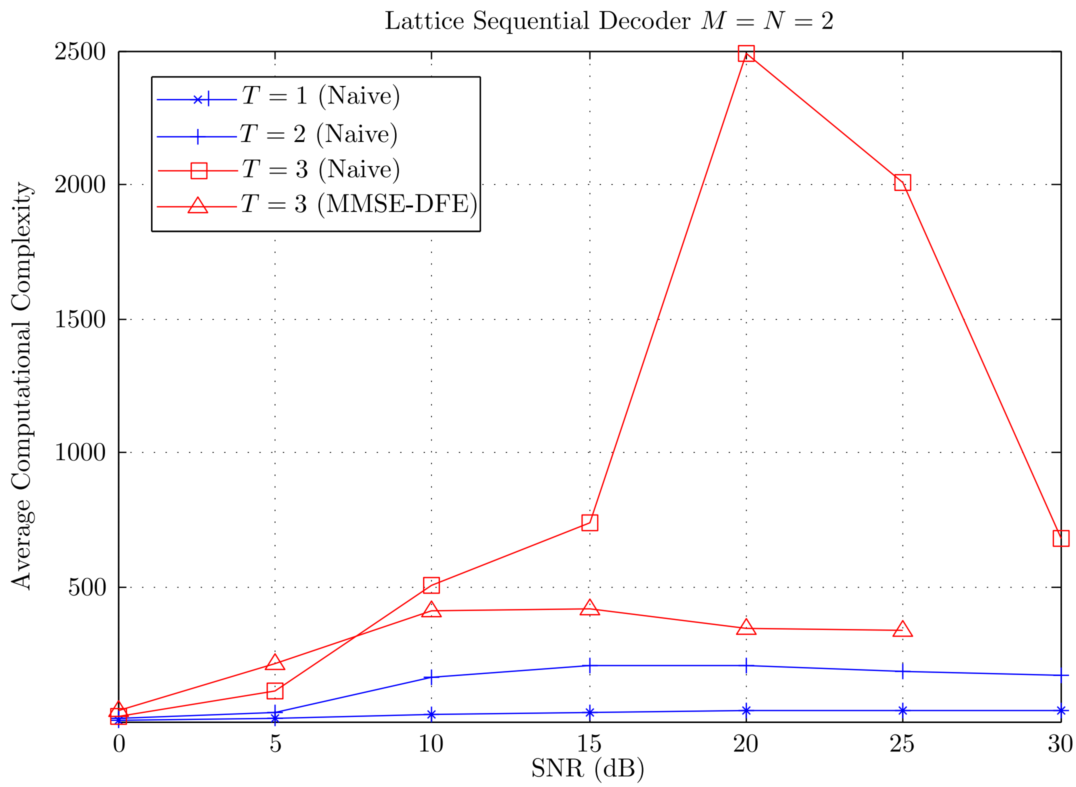


Figure 4.11: The reduction in computational complexity achieved by the MMSE-DFE lattice sequential decoder compared to the naive one for several codeword lengths. The codeword lengths are selected so that the maximum diversity is achieved: $T \geq 1$ for naive decoding, and $T \geq 3$ for MMSE-DFE decoding.

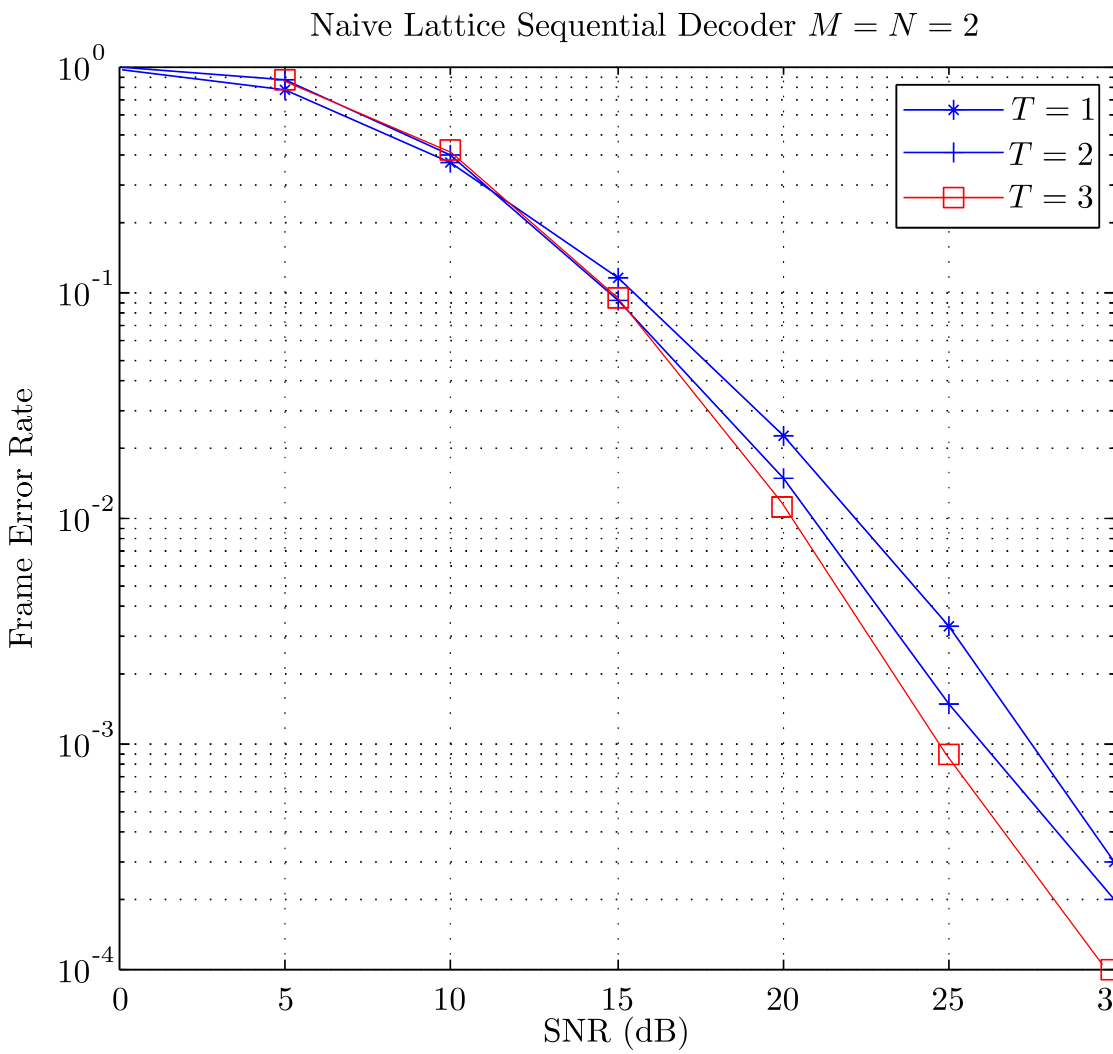


Figure 4.12: Comparison of the performance achieved by the naive lattice sequential decoder with $b = 0.3$ in a 2×2 quasi-static MIMO channel, for different values of codeword length T.

4.8 Summary

In this chapter, we have provided a complete analysis for the performance limits of lattice Fano/Stack sequential decoder applied to LAST coded MIMO system. The achievable rate of such system is derived. It turns out that the achievable rate under lattice sequential decoding depends critically on the decoding parameter, the bias term. The bias term is responsible for the excellent performance-complexity tradeoff achieved by such decoding scheme. For small values of bias, it has been shown that the optimal tradeoff of the channel can be achieved. As the bias grows without bound, lattice sequential decoding achieves linear computational complexity, where the total number of visited nodes during the search is always equal to the signal dimension. As such, lattice sequential decoders bridge the gap between the lattice (sphere) decoder and the low complexity receivers (e.g., the MMSE-DFE decoder). At high SNR, it was argued that there exists a lattice sequential decoding algorithm that can achieve maximum diversity gain at very low multiplexing gain, especially for large number of receive antennas.

We have also provided a complete analysis for the computational complexity of the lattice sequential decoder applied to LAST coded MIMO systems at the high SNR regime. It has been shown that for both the naive and the MMSE-DFE lattice sequential decoders, if the number of computations performed by the decoder exceeds a certain limit, then the complexity's tail distribution becomes dominated by the outage probability with an SNR exponent that is equivalent to the diversity-multiplexing tradeoff achieved by the corresponding coding and decoding schemes. The tradeoff of the channel is naturally extended to include decoding complexity. Moreover, the asymptotic average computational complexity has also been analyzed for both cases. As expected, MMSE-DFE preprocessing significantly improves the overall computational complexity of the underlying decoding scheme. Finally, it has been shown that there exists a cut-off multiplexing gain for which the average complexity remains bounded as long as we operate below such value.

Chapter 5

Time-Out Lattice Sequential Decoder for The MIMO ARQ Channel

AUTOMATIC Repeat reQuest (ARQ) is an efficient communication strategy that uses feedback to achieve high reliability, and is widely used in many wireless networks (e.g., LTE and WiMAX) (refer to [58] for a detailed study about several ARQ schemes). In its early stages, ARQ was used in conjunction with codes with good error detection capabilities. However, such codes increase the number of retransmissions which significantly reduce transmission rate (throughput) and increase delay. This may become undesirable for many communication systems, particularly in wireless fading channels. To overcome such problems, hybrid-ARQ system was introduced which uses forward error correction techniques (e.g., block and convolutional codes) [59]–[63]. This, however, comes at the expense of increasing the complexity of the receiver. The design of low complexity receivers for ARQ systems that achieve *near* optimal performance and high throughput is considered a challenging problem.

The class of sequential decoders is among the most promising decoders that can handle high data rates with low decoding complexity. As we have demonstrated in Chapter 4, there is still a non-zero probability that the decoding complexity (time) becomes excessive, especially when the channel is very noisy. In this situation, the decoder encounters buffer overflow that results in a decoding failure. It is this probability that limits the performance of the sequential decoder. Fortunately, the decoding failure probability can be totally eliminated using systems with *feedback* channel. For that reason, sequential decoders were adopted with ARQ systems [43],[60],[61] due to their ability to detect for retransmission before ending the decoding search which results in huge saving in decoding complexity while maintaining high throughput. All of this makes sequential decoding very promising and attractive for use in systems with repeat request.

Many sequential decoding algorithms (e.g., stack algorithm) were modified for the use of signal detection and decoding in ARQ systems. Among those algorithms that is considered simple but efficient is the so-called *time-out* sequential decoding. In this algorithm, the decoder simply tracks the number of computations performed by the decoder and asks for retransmission if the computations become excessive and exceed a certain predetermined time limit. This results in reducing the decoding complexity by terminating the search during high channel noise. For the case of single-input single-output discrete memoryless channel, it was shown (see [43]) that there exists an optimal time-out limit value that maximizes both performance and throughput while achieving low decoding complexity. In this chapter, we would like to extend the work in [43] to the quasi-static MIMO ARQ channel. In particular, we will study the throughput-performance-complexity tradeoffs in sequential decoding algorithms and the effect of preprocessing and termination strategies such as the time-out algorithm.

5.1 System Model

5.1.1 ARQ MIMO Channel

Consider a MIMO ARQ system with M-transmit and N-receive antennas, a maximum of L rounds, no CSI at the transmitter and perfect CSI at the receiver. For the MIMO ARQ channel model, we follow in the footsteps of El Gamal *et al.* [41] and use the incremental-redundancy ARQ transmission scheme. We restrict ourselves to the one-bit feedback (ACK/NACK) MIMO ARQ model. The ARQ feedback channel is assumed to be zero-delay and error-free. The complex baseband model of the received signal at the ℓ-th round can be mathematically described as

$$\boldsymbol{Y}_\ell^c = \sqrt{\frac{\rho}{M}}\boldsymbol{H}_\ell^c\boldsymbol{X}_\ell^c + \boldsymbol{W}_\ell^c, \tag{5.1}$$

where $\boldsymbol{X}_\ell^c \in \mathbb{C}^{M\times T}$ is the transmitted signal matrix, T is the number of channel uses, $\boldsymbol{Y}_\ell^c \in \mathbb{C}^{N\times T}$ is the received signal matrix, $\boldsymbol{W}_\ell^c \in \mathbb{C}^{N\times T}$ is the noise matrix, $\boldsymbol{H}_\ell^c \in \mathbb{C}^{N\times M}$ is the channel matrix, and ρ is the average signal-to-noise ratio (SNR) per receive antenna. The elements of both the noise matrix and the channel fading gain matrix are assumed to be independent identically distributed zero-mean circularly symmetric complex Gaussian random variables with variance $\sigma^2 = 1$.

In this work, we assume two different scenarios of channel dynamics. The first model being the *long-term static* channel, where the channel coefficients remain constant during all L rounds, i.e., $\boldsymbol{H}_\ell^c = \boldsymbol{H}^c$ for all $1 \leq \ell \leq L$. This scenario applies to very fast ARQ protocols and/or very slow fading environments, such as wireless LANs. The second scenario is the *short-term static* channel, where the channel remains constant during each round and changes independently at each round. This scenario applies to slow ARQ protocols where the time between the consecutive rounds is larger than the channel coherence time, or to frequency-selective fading, where each ARQ transmission takes place at a different frequency according to some frequency hopping scheme. Also, the following short-term average power constraint on the transmitted signal is assumed

$$\mathbb{E}\{\|\boldsymbol{X}_\ell^c\|_F^2\} \leq MT. \tag{5.2}$$

The equivalent real-valued channel model, after ℓ transmission rounds, corresponding to (5.1) can be written as

$$\boldsymbol{y}_\ell = \boldsymbol{H}_\ell \boldsymbol{x} + \boldsymbol{w}_\ell, \tag{5.3}$$

where we define $\boldsymbol{x} = (\boldsymbol{x}_{1,1}^\mathsf{T}, \ldots, \boldsymbol{x}_{L,1}^\mathsf{T}, \ldots, \boldsymbol{x}_{L,T}^\mathsf{T})^\mathsf{T}$, with $\boldsymbol{x}_{\ell,t}^\mathsf{T} = (\Re\{[\boldsymbol{X}_\ell^c]_t\}^\mathsf{T}, \Im\{[\boldsymbol{X}_\ell^c]_t\}^\mathsf{T})^\mathsf{T}$, and $\boldsymbol{w} = (\boldsymbol{w}_{1,1}^\mathsf{T}, \ldots, \boldsymbol{w}_{\ell,1}^\mathsf{T}, \ldots, \boldsymbol{w}_{\ell,T}^\mathsf{T})^\mathsf{T}$, with $\boldsymbol{w}_{\ell,t}^\mathsf{T} = (\Re\{[\boldsymbol{W}_\ell^c]_t\}^\mathsf{T}, \Im\{[\boldsymbol{W}_\ell^c]_t\}^\mathsf{T})^\mathsf{T}$. The vector $\boldsymbol{y}_\ell \in \mathbb{R}^{2NT\ell}$ represents the total signal received over all transmitted blocks from 1 to ℓ. The equivalent real-valued channel matrix, $\boldsymbol{H}_\ell$, has dimension $2NT\ell \times 2MTL$ and is formed by taking the first $2NT\ell$ rows of the matrix $\boldsymbol{H}_L$ which is composed by L diagonal blocks, where each diagonal block takes the form

$$\sqrt{\frac{\rho}{M}} \boldsymbol{I}_T \otimes \begin{pmatrix} \Re\{\boldsymbol{H}_\ell^c\} & -\Im\{\boldsymbol{H}_\ell^c\} \\ \Im\{\boldsymbol{H}_\ell^c\} & \Re\{\boldsymbol{H}_\ell^c\} \end{pmatrix}.$$

5.1.2 IR-LAST Coding Scheme

An m-dimensional lattice code $\mathcal{C}(\Lambda, \boldsymbol{u}_o, \mathcal{R})$ is the finite subset of the lattice translate $\Lambda + \boldsymbol{u}_0$ inside the shaping region $\mathcal{R}$, i.e., $\mathcal{C} = \{\Lambda + \boldsymbol{u}_0\} \cap \mathcal{R}$, where $\mathcal{R}$ is a bounded measurable region of $\mathbb{R}^m$. It is well-known [23] that an $(M \times T) \times L$ space-time coding scheme is a full-dimensional LAST code if its vectorized (real) codebook (corresponding to the channel model (5.3)) is a lattice code with dimension $m = 2MTL$.

We say that a LAST code is nested if the underlying lattice code is nested. Here, the information message is effectively encoded into the cosets Λ_s in Λ_c. As defined in [23], we shall call such codes the mod-Λ scheme. The proposed mod-Λ scheme works as follows. Consider the nested LAST code $\mathcal{C}$ defined by Λ_c (the coding lattice) and by its sublattice Λ_s

(the shaping lattice) in $\mathbb{R}^m$. Assume that Λ_s has a second-order moment $\sigma^2(\Lambda_s) = 1/2$ (so that $\boldsymbol{u}$ uniformly distributed over $\mathcal{V}_s$ satisfies $\mathsf{E}\{\|\boldsymbol{u}\|^2\} = MTL$). The transmitter selects a codeword $\boldsymbol{c} \in \mathcal{C}$, generates a dither signal[1] $\boldsymbol{u}$ with uniform distribution over $\mathcal{V}_s$, and computes $\boldsymbol{x} = [\boldsymbol{c} - \boldsymbol{u}] \bmod \Lambda_s$.

For the MIMO ARQ channel, we use the mod-Λ *incremental redundancy* scheme that was provided in [41]. The signal $\boldsymbol{x}$ is partitioned into L vectors of size $2MT$ each. Those vectors are transmitted, sequentially, in the different ARQ rounds based on the ACK/NACK feedback. Upon completion of the $\ell < L$ transmission, the receiver attempts to decode the message using lattice stack sequential decoder (introduced in Chapter 4) that implements a sort of deadline algorithm. In particular, the received signal, $\boldsymbol{y}_\ell$, is multiplied by the forward filter matrix $\boldsymbol{F}_\ell$ of the minimum mean-square error decision feedback equalization (MMSE-DFE) corresponding to the truncated matrix $\boldsymbol{H}_\ell$, and then the dither signal filtered by the upper triangular feedback filter matrix $\boldsymbol{B}_\ell$ of the MMSE-DFE is added to it (the definitions and some useful properties of the MMSE-DFE matrices $\boldsymbol{F}$, $\boldsymbol{B}$ are given in [23]). In this case, the received signal can be expressed as

$$\boldsymbol{y}'_\ell = \boldsymbol{F}_\ell \boldsymbol{y}_\ell + \boldsymbol{B}_\ell \boldsymbol{u} = \boldsymbol{B}_\ell \boldsymbol{c}' + \boldsymbol{e}', \tag{5.4}$$

where $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{\lambda}$, $\boldsymbol{\lambda} = -Q_{\Lambda_s}(\boldsymbol{c} - \boldsymbol{u})$, and $\boldsymbol{e}' = -[\boldsymbol{B}_\ell - \boldsymbol{F}_\ell \boldsymbol{H}_\ell]\boldsymbol{x} + \boldsymbol{F}_\ell \boldsymbol{e}_\ell$. One can easily verify (see [23]) the relationship between $\boldsymbol{B}_\ell$ and the channel matrix $\boldsymbol{H}_\ell$ through the following equations:

For the case of long-term static channel, we have

$$\det(\boldsymbol{B}_\ell^{\mathsf{T}} \boldsymbol{B}_\ell) = \left(\det \left(\boldsymbol{I} + \frac{\rho}{M} (\boldsymbol{H}^c)^{\mathsf{H}} \boldsymbol{H}^c \right) \right)^{2T\ell}, \tag{5.5}$$

and for the short-term static channel, we have

$$\det(\boldsymbol{B}_\ell^{\mathsf{T}} \boldsymbol{B}_\ell) = \prod_{j=1}^{\ell} \det \left(\boldsymbol{I} + \frac{\rho}{M} (\boldsymbol{H}_j^c)^{\mathsf{H}} \boldsymbol{H}_j^c \right)^{2T}. \tag{5.6}$$

The basic idea in this approach is to use a modified lattice stack sequential decoder for joint error detection and correction. The decoder first check if the channel is in outage. In this case, an error is declared and a NACK is sent back. If not, we use the modified lattice sequential decoder to find a lattice point that satisfies a certain predetermined condition

[1] A dither signal is a random signal that is used to make the MMSE estimation error independent of the transmitted codeword (see [25] for further details).

(e.g., a time-out limit). Now if no point is found, an error is declared, and hence, a NACK bit is fed back. If a point is found to satisfy such condition then we proceed to the next step to find the codeword as $\hat{\boldsymbol{c}} = [\boldsymbol{G}\hat{\boldsymbol{z}}] \bmod \Lambda_s$. The only exception to this rule is at the L-th ARQ round, where the regular lattice stack sequential decoder is used to find the closest lattice point.

5.1.3 Diversity-Multiplexing-Delay Tradeoff

Let η be defined as the average throughput of the ARQ scheme, expressed in transmitted bits per channel use. Following the definition used in [41], η can be expressed as

$$\eta = \frac{R_1}{1 + \sum_{\ell=1}^{L-1} p(\ell)}, \tag{5.7}$$

where R_1 denotes the rate of the first block in bits per channel use, $p(\ell) = \Pr(\overline{\mathcal{A}_1}, \ldots, \overline{\mathcal{A}_\ell})$ with $\mathcal{A}_\ell$ denoting the event that an ACK is fed back at round ℓ.

Let E_ℓ denotes the event that the transmitted message is incorrectly decoded by the ARQ decoder, then the probability of error can be upper bounded as

$$P_e \leq \sum_{\ell=1}^{L-1} \Pr(E_\ell, \mathcal{A}_\ell) + \Pr(E_L), \tag{5.8}$$

where $\Pr(E_\ell, \mathcal{A}_\ell)$ takes the definition of the probability of *undetected error* at round $\ell \leq L-1$.

Define respectively the ARQ multiplexing gain and ARQ diversity gain as

$$r_e = \lim_{\rho\to\infty} \frac{\eta(\rho)}{\log\rho}, \quad d = -\lim_{\rho\to\infty} \frac{\log P_e(\rho)}{\log\rho}.$$

In the MIMO ARQ system, the *delay* introduced by the channel provides a third dimension in the tradeoff region. The tradeoff between the diversity, the multiplexing, and the delay provided by the quasi-static MIMO ARQ channel has been established in the paper by El-Gamal et. al. in [41]. The result is summarized in the following theorem:

Theorem 10. *The optimal tradeoff of the coherent block-fading MIMO ARQ channel with M-transmit, N-receive antennas, L maximum number of ARQ rounds , under the short-term power constraint, is given as follows: In the case of long-term static channel*

$$d^*_{ls}(r_e, L) = f\left(\frac{r_e}{L}\right) \ \forall 0 \leq r_e < \min\{M, N\},$$

which is achieved with code block length $T \geq \lceil (M + N - 1)/L \rceil$, where $f(r) = (M - r)(N - r)$. In the case of short-term static channel

$$d^*_{ss}(r_e, L) = Lf\left(\frac{r_e}{L}\right), \ \forall 0 \leq r_e < \min\{M, N\},$$

which is achieved with code block length $T \geq M + N - 1$.

It has been shown in [41] that LAST codes achieve the optimal *diversity-multiplexing-delay* tradeoff. Achieving the optimal tradeoff in [41] was performed using IR-LAST coding scheme coupled with a *list lattice decoder* for joint error detection and correction. This decoder (corresponds to (5.4)) finds all lattice points that satisfy (see [41])

$$\left\{ \boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{y}' - \boldsymbol{B}_\ell \boldsymbol{x}\|^2 \leq MTL(1 + \gamma \log \rho) \right\},$$

where γ is a constant chosen appropriately to strike the optimal balance between the probability of error (accepting a wrong message) and the probability of declaring a decoding error (i.e., to ensure the achievability of the optimal tradeoff). This decoder can be efficiently implemented using sphere decoding algorithms (see for example [32]). Sphere decoders, however, are computationally very complex especially for low-to-moderate SNR and large signal dimensions where the output of the list sphere decoder can become extensively large that may result in a waste of time trying to decode the message. Hence, it is of great interest to search for a low complexity joint detector and decoder that can achieve the optimal tradeoff. This fact motivates us to replace the list lattice decoder by a more efficient retransmission strategy using lattice sequential decoders. The strategy is based on the stack algorithm and is designed to predict the occurrence of an error in advance by monitoring the number of computations performed by the decoder. Before we do that, we would like to introduce next the sequential decoder for lattice codes and some of its parameters that will be used to construct our new MIMO ARQ joint error detection and decoding scheme.

5.2 Lattice Sequential Decoder: Performance Bounds and Complexity Distribution

5.2.1 Lattice Stack Algorithm

We extend the stack algorithm that was introduced in Chapter 4 to the MIMO ARQ channel. The metric at the ℓ-th round [corresponds to (5.4)] is given by

$$\mu(\boldsymbol{z}_1^k, \ell) = bk - \|\boldsymbol{y}''^{k}_{\ell\,1} - \boldsymbol{R}^{(\ell)}_{kk}\boldsymbol{z}_1^k\|^2, \quad \forall 1 \leq k \leq m, \tag{5.9}$$

where $\boldsymbol{z}_1^k = [z_k, \cdots, z_2, z_1]^T$ denotes the last k components of the integer vector $\boldsymbol{z} \in \mathbb{Z}^m$, $\boldsymbol{R}^{(\ell)}_{kk}$ is the lower $k \times k$ matrix of $\boldsymbol{R}_\ell$ that corresponds to the QR decomposition of the code-channel matrix $\boldsymbol{B}_\ell\boldsymbol{G} = \boldsymbol{Q}_\ell\boldsymbol{R}_\ell$ at the ℓ-th round, $\boldsymbol{y}''^{k}_{\ell\,1}$ is the last k components of the vector $\boldsymbol{y}''_\ell = \boldsymbol{Q}_\ell^{\mathsf{T}}\boldsymbol{y}'_\ell$, and $b \geq 0$ is the bias term. The bias parameter is critical for controlling the amount of computations required at the decoding stage.

The main role of the bias term b used in the algorithm is to control the amount of computations performed by the decoder. In this work, we define the computational complexity of the joint lattice sequential decoder as the *total number of nodes visited by the decoder during the search, accumulated over all ARQ rounds, until a new transmission is started.* Also, the bias term is responsible for the excellent performance-throughput-complexity tradeoff achieved by such decoding scheme. The role that the bias parameter plays in the new efficient decoding algorithm will be discussed in details in the subsequent sections.

5.2.2 Performance Analysis: Lower and Upper Bounds

In this section, we extend the results of Chapter 4 to the MIMO ARQ channel and derive lower and upper bounds on the error performance of the lattice sequential decoder when applied to such a channel. These bounds work as the primary elements for constructing our new efficient decoder for the MIMO ARQ channel.

Consider the detection at the ℓ-th ARQ round. For simplicity, we consider here the long-term static channel (similar arguments can be done for the short-term static channel). Assume the received signal is $\boldsymbol{y}_\ell = \boldsymbol{B}_\ell\boldsymbol{x} + \boldsymbol{e}_\ell$, and denote $E_{ld}(\boldsymbol{B}_\ell)$ and $E_{sd}(\boldsymbol{B}_\ell, b)$ as the events that lattice decoder and lattice sequential decoder make an erroneous detection, respectively, where b is the bias term that was introduced in (5.9). As has been discussed in Chapter 4, lattice decoders outperform lattice sequential decoders for any $b > 0$. In fact, the performance of the lattice decoder serves as a lower bound of the lattice stack sequential decoder. Now, lattice decoding disregard the boundaries of the lattice code and find the point of the underlying (infinite) lattice closest to the received point. As such, due to lattice symmetry, one can assume that the all-zero lattice point is transmitted. For a given lattice Λ_c, we have

$$P(E_{ld}(\boldsymbol{B}_\ell)|\Lambda_c) = \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}_\ell \geq \|\boldsymbol{B}_\ell\boldsymbol{x}\|^2\right\}\right) \leq P(E_{sd}(\boldsymbol{B}_\ell, b)|\Lambda_c), \tag{5.10}$$

where $\Lambda_c^* = \Lambda_c\backslash\{\mathbf{0}\}$.

For the lattice sequential decoder, it seems a bit difficult to obtain an exact expression for its decoding error probability. Instead, we seek to derive an upper bound for the error performance of such a decoder which can be done as follows:

$$\begin{aligned} P(E_{sd}(\boldsymbol{B}_\ell, b)|\Lambda_c) &\overset{(a)}{\leq} \Pr\left(\bigcup_{\boldsymbol{z}\in\mathbb{Z}^m\backslash\{\mathbf{0}\}}\{\mu(\boldsymbol{z},\ell) > \mu_{\min}(\ell)\}\right) \\ &\overset{(b)}{\leq} \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\{\|\boldsymbol{B}_\ell\boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}_\ell < bm\}\right) \\ &= \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}_\ell > \|\boldsymbol{B}_\ell\boldsymbol{x}\|^2\left(1-\frac{bm}{\|\boldsymbol{B}_\ell\boldsymbol{x}\|^2}\right)\right\}\right), \end{aligned} \tag{5.11}$$

where (a) is due to the fact that in general, $\mu(\boldsymbol{z},\ell) > \mu_{\min}(\ell)$ is just a necessary condition for $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{z}$ to be decoded by the stack decoder, $\mu_{\min} = \min\{0, b-\|\boldsymbol{e}'^{1}_{\ell 1}\|^2, 2b-\|\boldsymbol{e}'^{2}_{\ell 1}\|^2, \ldots, bm-\|\boldsymbol{e}'^{m}_{\ell 1}\|^2\}$ is the minimum metric that corresponds to the transmitted path with $\boldsymbol{e}'_\ell = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{e}_\ell$, and (b) follows by noticing that $-(\mu_{\min} + \|\boldsymbol{e}'_\ell\|^2) \leq 0$. Similar to the proof the Theorem 5 in Chapter 4, one can show that (5.11) can be finally upper bounded as

$$P(E_{sd}(\boldsymbol{B}_\ell, b)|\Lambda_c) \leq P(E_{ld}(\tilde{\boldsymbol{B}}_\ell)|\Lambda_c), \tag{5.12}$$

where

$$\tilde{\boldsymbol{B}}_\ell = \left(1 - \frac{b}{2^{[R_{\text{mod}}(\ell)-R]/M}\phi(\ell)}\right)\boldsymbol{B}_\ell, \tag{5.13}$$

where $R_{\text{mod}}(\ell)$ is the rate at round ℓ that can be achieved using MMSE-DFE lattice decoding and according to (5.5) is given by

$$R_{\text{mod}}(\ell) = \log\det(\boldsymbol{B}_\ell^{\mathsf{T}}\boldsymbol{B}_\ell)^{1/2T} = \log\det\left(\boldsymbol{I} + \frac{\rho}{M}(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)^{\ell},$$

R is the transmission rate, and $\phi(\ell) = 0.5(2r_{\text{pack}}(\boldsymbol{B}_\ell\boldsymbol{G})/r_{\text{eff}}(\boldsymbol{B}_\ell\boldsymbol{G}))^2$. Interestingly, one may show that $\phi(\ell)$ is lower bounded by a constant independent of SNR and ℓ and as a result it has no effect on the performance in the SNR scale of interest.

It is clear from the above analysis that the lattice stack sequential decoder approaches the performance of the lattice decoder as $b \to 0$, i.e., $E_{sd}(\boldsymbol{B}_\ell, 0) = E_{ld}(\boldsymbol{B}_\ell)$. Moreover, the upper bound (5.12) corresponds to the probability of decoding error of a received signal $\boldsymbol{y}_\ell = \tilde{\boldsymbol{B}}_\ell\boldsymbol{x} + \boldsymbol{e}_\ell$ decoded using lattice decoding and is valid for all values of $b < 2^{[R_{\text{mod}}(\ell)-R]/M}\phi(\ell)$, i.e., $E_{sd}(\boldsymbol{B}_\ell, b) = E_{ld}(\tilde{\boldsymbol{B}}_\ell)$. Therefore, for a given lattice Λ_c, channel matrix $\boldsymbol{H}_\ell$, and a bias term $b > 0$, one can bound the error performance of the lattice sequential decoder as

$$P(E_{ld}(\boldsymbol{B}_\ell)|\Lambda_c) \leq P(E_{sd}(\boldsymbol{B}_\ell, b)|\Lambda_c) \leq P(E_{ld}(\tilde{\boldsymbol{B}}_\ell)|\Lambda_c). \tag{5.14}$$

By averaging (5.14) over the ensemble of random lattices Λ_c, one can show that for a fixed non-random channel matrix $\boldsymbol{H}_\ell^c$, the rate

$$R_b(\boldsymbol{H}_\ell^c, \rho, \ell) \triangleq \max\left\{R_{\text{mod}}(\boldsymbol{H}_\ell^c, \rho) - 2ML\log\left(\frac{1+\sqrt{1+8\alpha(\ell)}}{2}\right), 0\right\}, \tag{5.15}$$

is achievable by LAST coding and MMSE-DFE lattice Fano/Stack sequential decoding with bias term $b \geq 0$, where α is given by

$$\alpha(\ell) = \left(\frac{r_{\text{eff}}(\boldsymbol{B}_\ell\boldsymbol{G})}{2r_{\text{pack}}(\boldsymbol{B}_\ell\boldsymbol{G})}\right)^2 b. \tag{5.16}$$

This is simply the extension of Theorem 5 to the MIMO ARQ channel.

The equations (5.15) and (5.16) suggest that as long as the channel is well-conditioned, one may use large of values of bias term which is needed to achieve low decoding complexity (as will be discussed later). On the other hand, if the channel is close to outage, very low values of b must be chosen in order to maintain high achievable rates. For example, if $b = 0$, (5.15) reduces simply to $R_b = R_{\text{mod}}$, which corresponds to the rate achievable by the MMSE-DFE lattice decoder.

In general, for the long-term static channel, one can show that if b is allowed to vary with SNR and the channel statistics as

$$b(\boldsymbol{\lambda}, \rho) = \frac{1}{2}\frac{\prod_{i=1}^M(1+\rho\lambda_i)^{\ell/ML}}{\eta(\boldsymbol{\lambda}, \rho)^{\ell/ML}}\left[1 - \left(\frac{\eta(\boldsymbol{\lambda}, \rho)}{\prod_{i=1}^{M}(1+\rho\lambda_i)}\right)^{\ell/2ML}\right]\left(\frac{2r_{\text{pack}}(\boldsymbol{B}_\ell\boldsymbol{G})}{r_{\text{eff}}(\boldsymbol{B}_\ell\boldsymbol{G})}\right)^2, \tag{5.17}$$

the achievable rate can be rewritten as

$$R_b(\boldsymbol{\lambda}, \rho, \ell) = \ell \log \eta(\boldsymbol{\lambda}, \rho). \tag{5.18}$$

The term $\eta(\boldsymbol{\lambda}, \rho)$ can be chosen freely between 1 and $\prod_{i=1}^{M}(1+\rho\lambda_i)$ (the maximum achievable rate under lattice decoding). Depending on the value of $\eta(\boldsymbol{\lambda}, \rho)$ we obtain different achievable rates and hence different outage performances.

We define the outage event under lattice sequential decoding as $\mathcal{O}_b(\rho, \ell) \triangleq \{\boldsymbol{H}_\ell^c : R_b(\boldsymbol{H}_\ell^c, \rho) < R_1\}$. Denote $R_1 = r_1 \log \rho$. The probability that the channel is in outage, $P_{\text{out}}(\ell, b) = \Pr(\mathcal{O}_b(\rho, \ell))$, can be evaluated as follows:

$$P_{\text{out}}(\ell, b) = \Pr(\ell \log \eta(\boldsymbol{\lambda}, \rho) < R_1) \doteq \rho^{-d_b(r_1/\ell)}. \tag{5.19}$$

where $d_b(r)$ is the outage SNR exponent that is achieved by the MIMO channel with no ARQ (i.e., with $L = 1$). For simplicity, we may express

$$\eta(\boldsymbol{\lambda}, \rho) = \phi \prod_{i=1}^{M}(1+\rho\lambda_i)^{\zeta_i}, \tag{5.20}$$

where $0 < \phi < 1$ is a constant independent of ρ, and ζ_i, $\forall 1 \leq i \leq M$, are constants that satisfy the following two constraints: $\sum_{i=1}^{M}\zeta_i \leq M$, and $\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_M \geq 0$.

The above choice of ζ_i leads to a bias term (assuming high SNR)

$$b(\boldsymbol{\lambda}, \rho) \approx \frac{1}{2}\prod_{i=1}^{M}(1+\rho\lambda_i)^{(1-\zeta_i)\ell/ML} \tag{5.21}$$

which may grow exponentially with SNR as ρ^ϵ, $\epsilon > 0$, when the channel is not in outage. This may significantly reduce the decoding complexity. However, as discussed in Chapter 4, although the diversity at $r = 0$ is not affected by the coefficients $\zeta_i \neq 0$ ($d_b(0) = MN$), the more unbalanced the coefficients are, the worse the diversity-multiplexing tradeoff is. By varying the decoder parameter (bias term), one gets different performance-rate-complexity tradeoffs.

It is a simple matter to extend the above result to the ℓ-th ARQ round. One can show that, there exists a sequence of full-dimensional LAST codes with block length $T \geq (M+N-1)/\ell$ that achieves the diversity-multiplexing tradeoff curve $d_b(r_1/\ell)$ which is the piecewise-linear function connecting the points $(r(k), d(k/\ell))$, $k = 0, 1, \cdots, M$, where

$$\begin{aligned} r(0) = 0, \quad r(k) &= \sum_{i=M-k+1}^{M}\zeta_i, \ 1 \leq k \leq M, \\ d(k/\ell) &= \left(M - \frac{k}{\ell}\right)\left(N - \frac{k}{\ell}\right), \qquad 0 \leq k \leq M. \end{aligned} \tag{5.22}$$

In this case, one can show that there exists a lattice code Λ_c such that

$$P(E_{sd}(\boldsymbol{B}_\ell, b)) \doteq \rho^{-d_b(r_1/\ell)}, \quad \forall\, 1 \leq \ell \leq L. \tag{5.23}$$

under the condition $T\ell \geq N + M - 1$.

It must be noted that, except for the case of $\ell = L$ (the final ARQ round), $d_b(r_1/\ell)$ is not the best achievable SNR exponent for the MIMO ARQ channel at any round $\ell < L$. This is because in the above analysis we have ignored the detection capabilities that the MIMO ARQ can have. By carefully designing the lattice sequential decoder to include an efficient error detection mechanism, we will show how the SNR exponent can be raised up to achieve the optimal tradeoff of the channel at any round ℓ.

From a lattice point of view, the event of error under lattice decoding can be expressed as the event that the received signal is located outside the fundamental Voronoi region of the underlying (infinite) lattice. In this case, assuming $\mathbf{0}$ was transmitted, one may express the bounds in (5.14)

$$P(\boldsymbol{e} \notin \mathcal{V}_{\mathbf{0}}(\boldsymbol{B}_\ell \boldsymbol{G})|\Lambda_c) \leq P(E_{sd}(\boldsymbol{B}_\ell, b)|\Lambda_c) \leq P(\boldsymbol{e} \notin \mathcal{V}_{\mathbf{0}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G})|\Lambda_c). \tag{5.24}$$

where $\tilde{\boldsymbol{B}}_\ell$ is as defined in (5.13). Interestingly, the upper bound in (5.24) provides us with the fact that as long as $\boldsymbol{e} \in \mathcal{V}_{\mathbf{0}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G})$, the received signal is correctly decoded using lattice sequential decoding. As will be shown in the sequel, this fact can be used as the basic tool to construct our new efficient joint detection and decoding scheme.

5.2.3 Computational Complexity Distribution

An important parameter of the lattice sequential decoder is the distribution of computation, which characterizes the time needed to decode a message. It is well-known [38] that the number of computations required to decode a message using sequential decoders is highly variable and assume very large values during intervals of high channel noise. Moreover, due to the random nature of the channel matrix and the additive noise, the computational complexity of such decoder is considered difficult to analyze in general. However, in Chapter 4, we have shown that for the MIMO channel with no ARQ under lattice sequential decoding with constant bias term, the tail distribution becomes upper bounded by the asymptotic outage probability with SNR exponent that is equivalent to the optimal diversity-multiplexing tradeoff of the channel. This upper bound is shown to be achieved only of the number of computations performed by the decoder exceeds a certain limit when the channel is not in outage. This result can be easily extended to the MIMO ARQ channel as will be explained in the following.

Consider again decoding the received signal at the ARQ round $\ell = 1, \cdots, L$. Let $\phi(\boldsymbol{z}_1^k, \ell)$ be the indicator function defined by

$$\phi(\boldsymbol{z}_1^k, \ell) = \begin{cases} 1, & \text{if node } \boldsymbol{z}_1^k \text{ is extended;} \\ 0, & \text{otherwise,} \end{cases} \tag{5.25}$$

and let $\mathcal{N}_j(\ell)$ be a random variable that denotes the total number of visited nodes during the search up to dimension j, at round $\ell < L$. In this case, $\mathcal{N}_j(\ell)$ can be expressed as

$$\mathcal{N}_j(\ell) = \sum_{k=1}^{j} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k, \ell). \tag{5.26}$$

For the case of MIMO channel with no ARQ (i.e., $L = 1$), following the footsteps of the proof of Theorem 9 in Chapter 4, one can show that the *asymptotic* computational complexity distribution of the MMSE-DFE lattice sequential decoder (assuming no error detection) is given by (for $T \geq N + M - 1$)

$$\Pr(\mathcal{N}_j \geq \Gamma) \leq \Pr(\mathcal{N}_m \geq \Gamma) \dot{\leq} \rho^{-f(r)}, \tag{5.27}$$

for all Γ that satisfy

$$\Gamma \geq m + \sum_{k=1}^{m} \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MT(1 + \log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{\mathsf{T}} \boldsymbol{R}_{kk})^{1/2}}, \tag{5.28}$$

where $\boldsymbol{R}_{kk}$ is the lower $k \times k$ part of the upper triangular matrix $\boldsymbol{R}^{(\ell)}$ of the QR decomposition of $\boldsymbol{B}_\ell \boldsymbol{G}$, and $f(r) = (M - r)(N - r)$ for $r \in [0, \min\{M, N\})$. It must be noted that (5.28) is only valid for constant (fixed) bias term b, and a lower bound on Γ for the general expression of b (see (5.21) is known yet. However, as will be discuss in the sequel, even for fixed bias, the above lower bound shows a significant improvement in complexity compared to the more complex optimal sphere decoder.

The above important result indicates that there exists a finite probability that the number of computations performed by the decoder may become excessive even at high SNR, irrespective to the channel being ill or well-conditioned! This probability is usually referred to as the probability of a decoding failure. Such probability limits the performance of the lattice sequential decoder, especially for a one-way communication system. For a two-way communication system, such as in our MIMO ARQ system, the feedback channel can be used to eliminate the decoding failure probability. Therefore, our new decoder must be carefully designed to predict in advance the occurrence of decoding failure to avoid wasting the time trying to decode the message. This would result in a huge saving in decoding complexity. As will be shown in the sequel, the above result can be easily extended to the MIMO ARQ channel.

5.3 Time-Out Algorithm

It is well-known that the number of computations required to decode a message using sequential decoders is highly variable and assume very large values during intervals of high channel noise. As such, the decoder is expected to spend longer time attempting to decode the message. For the proposed incremental-redundancy MIMO ARQ system, this condition can be used as an indicator of when the receiver should terminate the search and request the transmitter for additional redundancy bits during any of the $\ell < L$ rounds.

In order to avoid wasting time trying to decode a noisy signal during any of $\ell < L$ ARQ rounds, we implement a *time-out* algorithm in the lattice stack sequential decoder for joint error detection and correction. Such algorithm works as follows: we define a parameter Γ_{out} to be the maximum time (number of computations) allowed to decode a message during any of the $\ell < L$ ARQ rounds. If the decoding time exceeds Γ_{out}, a NACK bit is fed back to the transmitter. The only exception of this rule is when the maximum number of ARQ rounds, L, is reached. In this case, the regular lattice sequential decoder (with no time-out limit) is used, where a NACK bit will be interpreted as an error, and the transmission of the next message is started anyway. Next, we define the *retransmission probability* and the *undetected error probability* from a lattice point of view. Those two quantities are responsible for the performance-throughput tradeoff achieved by the MIMO ARQ system. Throughout the work we assume the use of a small (fixed) bias term during all L ARQ rounds. Before continuing our analysis, we would like to introduce some important definitions related to the MIMO ARQ channel that will be used throughout the Chapter.

Denote $0 \leq \lambda_1^{(j)} \leq \cdots \leq \lambda_{\min\{M,N\}}^{(j)}$ the eigenvalues of $(\boldsymbol{H}_j^c)^{\mathsf{H}}\boldsymbol{H}_j^c$, $\forall 1 \leq j \leq \ell$. Let us first define the outage event for both long-term and short-term static channels under the lattice sequential decoder with bias term b that is given in (5.21) and $\eta(\boldsymbol{\lambda}, \rho)$ as defined in (5.20), with ℓ received blocks as

$$\mathcal{O}_b(\rho, \ell) = \left\{ \boldsymbol{H}_j^c \in \mathcal{C}^{N\times M}\ \forall 1 \leq j \leq \ell\ : R_b(\rho) < R_1 \right\},$$

where

$$R_b(\rho) = \begin{cases} \ell \sum_{i=1}^{M} \zeta_i \log(1+\rho\lambda_i), & \textit{for long-term static channel;} \\ \sum_{j=1}^{\ell}\sum_{i=1}^{M} \zeta_i \log(1+\rho\lambda_i^{(j)}), & \textit{for short-term static channel,} \end{cases} \tag{5.29}$$

Denote, $R_1 = r_1 \log\rho$, and define $\alpha_i \triangleq -\log\lambda_i/\log\rho$, and $(x)^+ = \max\{0, x\}$., then It can be easily verified that at high SNR, the outage event for the long-term static channel model can be expressed as (assuming $N \geq M$)

$$\mathcal{O}_{ls}(\ell) = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^M : \alpha_1 \geq \cdots \geq \alpha_M,\ \sum_{i=1}^{M} \zeta_i (1-\alpha_i)^+ < r_1/\ell \right\},$$

where in such channel we have $\lambda_i^{(j)} = \lambda_i$, $\forall 1 \leq j \leq \ell$. For the short-term static channel we have

$$\mathcal{O}_{ss}(\ell) = \left\{ (\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(\ell)}) \in \mathbb{R}_+^{M\ell} : \alpha_1^{(j)} \geq \cdots \geq \alpha_M^{(j)}, \forall 1 \leq j \leq \ell,\ \sum_{j=1}^{\ell}\sum_{i=1}^{M} \zeta_i \left[1-\alpha_i^{(j)}\right]^+ < r_1 \right\}.$$

Then, the associated asymptotic outage probability is given by (see [41])

$$P_{\text{out}}(\rho, \ell) \doteq \begin{cases} \Pr(\mathcal{O}_{ls}(\ell)) \doteq \rho^{-d_{\text{out}}(\ell)}, & \textit{for long-term static channel;} \\ \Pr(\mathcal{O}_{ss}(\ell)) \doteq \rho^{-\ell d_{\text{out}}(\ell)}, & \textit{for short-term static channel,} \end{cases} \tag{5.30}$$

where $d_{\text{out}}(\ell) = d_b(r_1/\ell)$ where $d_b(r)$ is as defined in (5.19).

5.3.1 Retransmission Request Probability

We make use of the lower and the upper bounds of the lattice sequential decoder's error performance in (5.24) to implement an error control mechanism for our IR-LAST MIMO ARQ system. It is clear from (5.13) and (5.24) that $\mathcal{V}_{\mathbf{0}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G}) \subseteq \mathcal{V}_{\mathbf{0}}(\boldsymbol{B}_\ell\boldsymbol{G})$ for all $b \geq 0$. Therefore, at the decoder side, one can divide each of the Voronoi regions of the channel-code lattice $\Lambda(\boldsymbol{B}_\ell\boldsymbol{G})$, i.e., $\mathcal{V}_{\boldsymbol{u}}(\boldsymbol{B}_\ell\boldsymbol{G})$ (corresponds to a lattice point $\boldsymbol{u} = \boldsymbol{B}_\ell\boldsymbol{G}\boldsymbol{z}$, $\boldsymbol{z} \in \mathbb{Z}^m$) into two disjoint regions — $\mathcal{R}_{\boldsymbol{u}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G})$ and $\mathcal{V}_{\boldsymbol{u}}(\boldsymbol{B}_\ell\boldsymbol{G})\backslash\mathcal{R}_{\boldsymbol{u}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G})$. This is depicted in Figure. 5.1. For convenience, we define $\mathcal{V}_{\boldsymbol{u}}(\ell) = \mathcal{V}_{\boldsymbol{u}}(\boldsymbol{B}_\ell\boldsymbol{G})$ and $\mathcal{R}_{\boldsymbol{u}}(\ell) = \mathcal{R}_{\boldsymbol{u}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G})$.

Now, denote the region $\mathcal{D}(\ell)$ as

$$\mathcal{D}(\ell) = \mathbb{R}^m \backslash \left\{ \bigcup_{\boldsymbol{u}\in\Lambda(\boldsymbol{B}_\ell\boldsymbol{G})} \mathcal{R}_{\boldsymbol{u}}(\ell) \right\}. \tag{5.31}$$

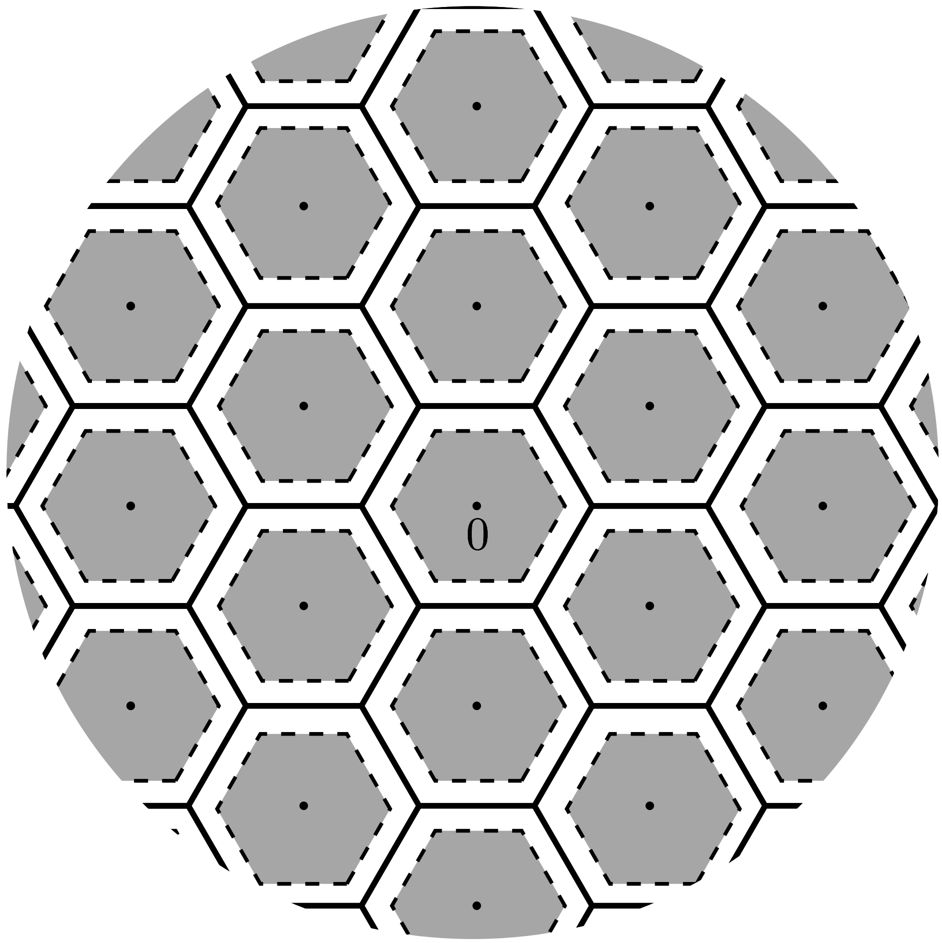


Figure 5.1: The division of the Voronoi cell of the lattice generated by the channel-code matrix $\boldsymbol{B}_\ell\boldsymbol{G}$ into two distinct regions — the shaded region $\mathcal{R}_{\boldsymbol{u}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G})$, and the white region $\mathcal{V}_{\boldsymbol{u}}(\boldsymbol{B}_\ell\boldsymbol{G})\backslash\mathcal{R}_{\boldsymbol{u}}(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G})$. The two dimensional hexagonal lattice is shown for illustration purposes.

We take advantage of the feedback channel by introducing the erasure option at the decoder such that whenever the received signal $\boldsymbol{y}_\ell \in \mathcal{D}(\ell)$, for $\ell < L$, the decoder requests for a repeat transmission or additional bits (as in the case of IR-LAST coding scheme). In this case, at round $\ell < L$, the probability of a retransmission when the channel is *not* in outage is given by

$$\Pr(\overline{\mathcal{A}}_\ell, \text{no outage}) = \Pr(\boldsymbol{y}_\ell \in \mathcal{D}(\ell)). \tag{5.32}$$

Unfortunately, determining whether $\boldsymbol{y}_\ell \in \mathcal{D}(\ell)$ or not is considered by itself a difficult problem. We try to simplify this problem through the use of lattice sequential decoding by tracking the number of computations performed during the search for the closest lattice point.

Following the definition of the number of computations performed by the decoder provided in (5.26), a *retransmission* is requested by the time-out algorithm at round $\ell < L$ if the number of computations exceeds the maximum time allowed before reaching the end of the tree. In other words, a NACK is sent back to the transmitter if, at any $1 \leq j < m$, $\mathcal{N}_j(\ell) > \Gamma_{\text{out}}$. This event could occur during a high channel noise period so that the received signal $\boldsymbol{y}_\ell$ is close to the boundaries of a Voronoi cell of the lattice $\Lambda(\boldsymbol{B}_\ell\boldsymbol{G})$ and as a result, the decoder declares that $\boldsymbol{y}_\ell \in \mathcal{D}(\ell)$. In this case, for the selected value of b, one has to carefully choose the time-out parameter Γ_{out} so that whenever $\mathcal{N}_j(\ell) > \Gamma_{\text{out}}$, the decoder decides that $\boldsymbol{y}_\ell \in \mathcal{D}(\ell)$ (see Figure. 5.2.(a)). Selecting an inappropriate value of Γ_{out} may result in the loss of the optimal tradeoff. Later, we shall make use of the following result for evaluating (5.32), for fixed[2] bias values:

Lemma 5. *For the long-term static ARQ channel, the asymptotic tail distribution of the total computational complexity of the lattice sequential decoder with fixed bias $b > 0$, at round ℓ given the channel is not in outage,* $\Pr(\mathcal{N}_j(\ell) \geq \Gamma_{\text{out}})$, *can be upper bounded by*

$$\Pr(\mathcal{N}_j(\ell) \geq \Gamma_{\text{out}}) \ \dot{\leq}\ \rho^{-d_\ell}, \quad \forall 1 \leq j \leq m, \tag{5.33}$$

under the condition

$$\Gamma_{\text{out}} \geq m + \sum_{k=1}^{m} \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MTL(1+\zeta\log\rho)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{R}_{kk}^{(\ell)})^{1/2}},$$

where ζ is a constant chosen sufficiently large enough so that $MTL\zeta \geq (M - r_1/\ell)(N - r_1/\ell)$, and

[2]The general case of a variable bias term that is given in (5.21) will not be considered here due to the difficulty of obtaining a lower bound on Γ_{out} for such bias values.

$$d_\ell = \inf_{\boldsymbol{\alpha}\in\overline{\mathcal{O}}_{ls}(\ell)} \left\{ \sum_{i=1}^{M}(2i-1+N-M)\alpha_i + T\ell\left[\sum_{i=1}^{M}(1-\alpha_i)^+ - r_1/\ell\right]\right\}, \tag{5.34}$$

and

$$\overline{\mathcal{O}}_{ls}(\ell) = \left\{\boldsymbol{\alpha}\in\mathbb{R}_+^M, \alpha_1 \geq \cdots \geq \alpha_M : \sum_{j=1}^{M}(1-\alpha_i)^+ \geq r_1/\ell\right\}.$$

Proof. see Appendix E. □

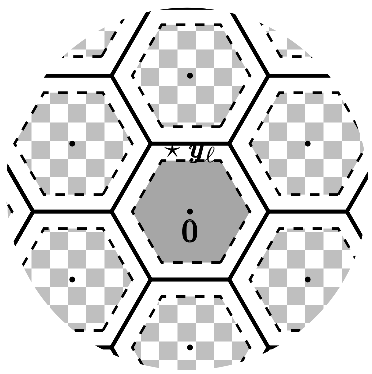
It must be noted that the above result can be easily extended for the short-term static channel. In this case, one can show that $\Pr(\mathcal{N}_j(\ell) \geq \Gamma_{\text{out}}) \dot{\leq} \rho^{-\ell d_\ell}$, $\forall 1 \leq j \leq m$. Next, we derive an upper bound for the undetected error probability.

5.3.2 Undetected Error Probability

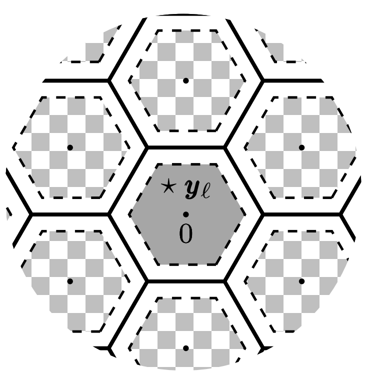
Another important parameter that characterizes the performance of the MIMO ARQ system, is the undetected error probability which was defined in (5.8). In the IR-LAST MIMO ARQ system, an ACK bit is sent back to the transmitter if the decoder correctly decode the received signal subject to the condition that $\mathcal{N}_m(\ell) < \Gamma_{\text{out}}$. This corresponds to the event $\boldsymbol{y}_\ell \in \mathcal{R}_{\mathbf{0}}(\ell)$, assuming $\mathbf{0}$ was transmitted (see Figure. 5.2.(b)). Also, an ACK is sent back to the transmitter, if decoding fails but it is not detected. Such an event occurs when the total number of computations $\mathcal{N}_m(\ell) < \Gamma_{\text{out}}$, but the decoded lattice point is not $\mathbf{0}$. This happens when the received signal $\boldsymbol{y}_\ell \in \mathcal{R}_{\boldsymbol{u}}(\ell)$ for any $\boldsymbol{u} \neq \mathbf{0}$ (see Figure. 5.2.(c)). Using the above argument, the undetected error probability can be expressed as (assuming $\mathbf{0}$ was transmitted)

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) = \Pr\left(\bigcup_{\boldsymbol{u}\in\Lambda^*(\boldsymbol{B}_\ell\boldsymbol{G})} \{\boldsymbol{e}'_\ell \in \mathcal{R}_{\boldsymbol{u}}(\ell)\}\right). \tag{5.35}$$

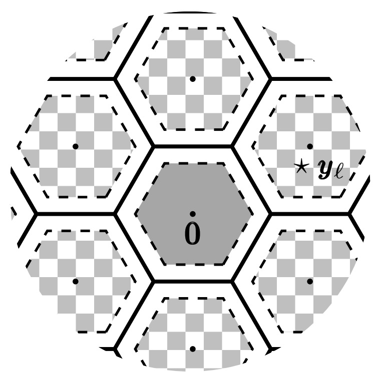
Our goal now is to upper bound (5.35). In this work, we will resort to a geometrical approach to obtain an upper bound on the undetected error probability. Before doing so,



(a) The event of sending a NACK when the channel is not in outage.



(b) The event of sending an ACK with correct decoding.



(c) The event of sending an ACK with decoding failure but not detected.

Figure 5.2: The events of retransmission, correct decoding, and undetected error that occur in the time-out algorithm (assuming $\mathbf{0}$ was transmitted). The correct decoding region is represented in dark color. The chessboard shaded regions represent the undetected error events $(E_\ell, \mathcal{A}_\ell)$. The white region represents the detected error event.

we express the event of sending an ACK, i.e., $\mathcal{A}_\ell$, in terms of the number of computations performed by the lattice sequential decoder as $\mathcal{A}_\ell = \{\mathcal{N}_m(\ell) < \Gamma_{\text{out}}\}$. Consider now the following theorem:

Theorem 11. *For any lattice code Λ_c, the undetected error probability of the quasi-static $M \times N$ MIMO ARQ system with maximum rounds L, and codeword length T, under the time-out MMSE-DFE lattice sequential decoding scheme with parameters b and Γ_{out}, is bounded above as: In the case of long-term static channel*

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) \dot{\leq} \rho^{-f(r_1/L)}, \tag{5.36}$$

which is achieved with code block length $T \geq \lceil (M+N-1)/L \rceil$, where $f(r) = (M-r)(N-r)$, and any b that satisfies (5.21). In the case of short-term static channel

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) \dot{\leq} \rho^{-Lf(r_1/L)}, \tag{5.37}$$

which is achieved with code block length $T \geq M+N-1$.

Proof. It seems a bit difficult to obtain an upper bound directly for the undetected error probability using (5.35). Therefore, we resort to a geometrical approach (see Figure. 5.3) to further upper bound (5.35) by selecting b such that the bound $r_{\text{eff}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G}) \leq r_{\text{pack}}(\boldsymbol{B}_\ell \boldsymbol{G})$ is maintained for all $\ell < L$. The effective radius $r_{\text{eff}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G})$ of the lattice generated by $\tilde{\boldsymbol{B}}_\ell \boldsymbol{G}$ is given by

$$r_{\text{eff}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G}) = \left[\frac{V_c \det(\tilde{\boldsymbol{B}}_\ell^\mathsf{T} \tilde{\boldsymbol{B}}_\ell)^{1/2}}{V(\mathcal{S}_m(1))}\right]^{1/m}. \tag{5.38}$$

In this case, we can upper bound the undetected error probability as

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) = \Pr\left(\bigcup_{\boldsymbol{u} \in \Lambda^*(\boldsymbol{BG})} \{\boldsymbol{e}'_\ell \in \mathcal{R}_{\boldsymbol{u}}(\ell)\}\right) \leq \Pr(\|\boldsymbol{e}'_\ell\|^2 \geq r^2_{\text{eff}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G})). \tag{5.39}$$

It is not yet clear how the RHS of (5.39) can be evaluated. To overcome this problem, we can find a lower bound on $r^2_{\text{eff}}(\tilde{\boldsymbol{B}}_\ell \boldsymbol{G})$ at high SNR as follows:

Asymptotically, one can express bias term that is defined by (5.21) as

$$b \doteq \frac{\rho^{\frac{\ell}{ML}\sum_{i=1}^{M}(1-\alpha_i)^+}}{\eta^{\ell/ML}}\left[1 - \left(\frac{\eta}{\rho^{\sum_{i=1}^{M}(1-\alpha_i)^+}}\right)^{\ell/2ML}\right]. \tag{5.40}$$

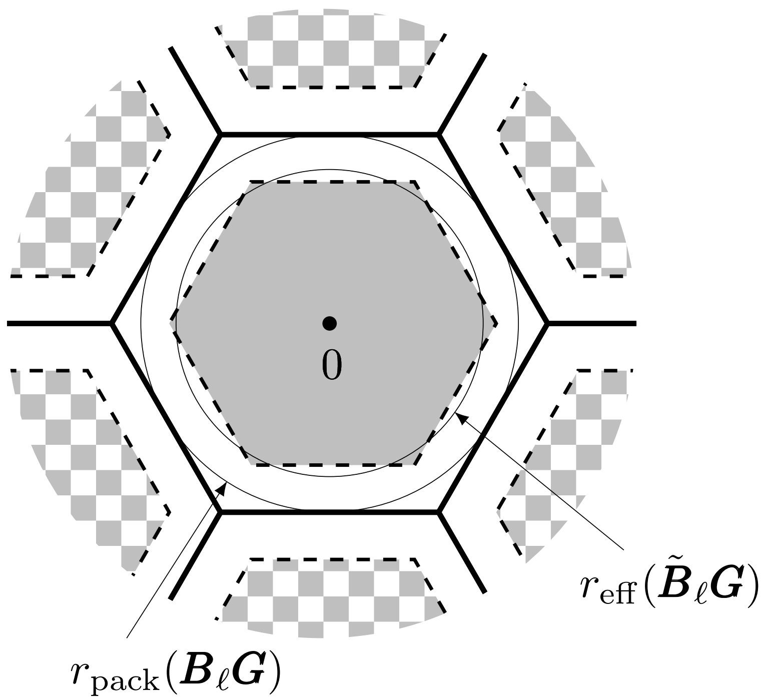


Figure 5.3: A geometric approach used to over bound the undetected error probability under the time-out algorithm.

Substituting (5.40) in (5.13), when the channel is not in outage, one can upper bound $\det(\tilde{\boldsymbol{B}}_\ell^{\mathsf{T}}\tilde{\boldsymbol{B}}_\ell)) \geq \eta^{2T}$, where $\eta \doteq \rho^{\sum_{i=1}^{M}\zeta_i(1-\alpha_i)}$. In this case, we have that

$$\begin{aligned} r_{\text{eff}}^2(\tilde{\boldsymbol{B}}_\ell\boldsymbol{G}) &\geq \left[\frac{V_c}{V(\mathcal{R})}\frac{V(\mathcal{R})}{V(\mathcal{S}_m(1))}\eta^T\right]^{2/m} \\ &\overset{(a)}{\geq} MTL\left[\rho^{-r_1T}\rho^{T\sum_{i=1}^{M}\zeta_i(1-\alpha_i)}\right]^{2/m} \\ &\dot{\geq} MTL\rho^{\nu} \dot{\geq} MTL(1+\gamma\log\rho), \end{aligned} \tag{5.41}$$

where (a) follows from the fact that there exists a shifted lattice code $\Lambda_c + \boldsymbol{u}_0^*$ with number of codewords inside the shaping region,

$$|\mathcal{C}(\Lambda_c, \boldsymbol{u}_0^*, \mathcal{R})| = 2^{R_1T} = \rho^{r_1T} \geq \frac{V(\mathcal{R})}{V_c}. \tag{5.42}$$

Also, $\nu = \frac{\ell}{ML}[\sum_{j=1}^{M}\zeta_i(1-\alpha_j)^+ - r_1/\ell] > 0$ when the channel is not in outage, and the last inequality follows from the fact that $\lim_{\rho\to\infty}(1+\gamma\log\rho)/\rho^\nu = 0$ for any $\nu, \gamma > 0$. Therefore,

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) \dot{\leq} \Pr(\|\boldsymbol{e}_\ell'\|^2 \geq MLT(1+\gamma\log\rho)) \dot{\leq} \rho^{-2MTL\gamma}. \tag{5.43}$$

By choosing a large enough value of γ such that $MTL\gamma \geq f(r_1/L)$, we obtain

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) \dot{\leq} \rho^{-f(r_1/L)}, \tag{5.44}$$

under the condition that $LT \geq M + N - 1$.

The above analysis also applies to the short-term static channel, and one can show that under the condition $T \geq N + M - 1$, we have

$$\Pr(E_{sd}(\boldsymbol{B}_\ell, b), \mathcal{A}_\ell) \dot{\leq} \rho^{-Lf(r_1/L)}. \tag{5.45}$$

□

As will be shown in the sequel, (5.32) and (5.35) play an important role in determining the achievable diversity-multiplexing-delay tradeoff of the MIMO ARQ channel. Reducing the number of retransmissions comes at the expense of increasing the undetected error probability, which is undesirable. It is clear that both probabilities are closely related and hence increasing or decreasing one of them may lead to a loss in the optimal tradeoff of the channel.

5.3.3 Achieving the Optimal Tradeoff: Bias Term vs. Γ_{out}

Our goal here is to prove the optimality of the time-out lattice sequential decoder in terms of the achievable diversity-multiplexing-delay tradeoff. It is well-known that the bias term b controls the amount of the computations performed by the lattice sequential decoder during the search and responsible for the excellent performance-complexity tradeoff achieved by such a decoder. Choosing a very large value of b although greatly reduces decoding complexity, it may lead to a loss in the optimal tradeoff of the channel. Therefore, it is expected that the time-out parameter Γ_{out} will be a function of the bias term b chosen in the algorithm. One may have already noticed that in Lemma 1. It turns out that an optimal value of Γ_{out}, denoted by Γ^*_{out}, exists so that the optimal tradeoff is achieved with a fairly low decoding complexity (i.e., average number of computations) compared to the joint list lattice decoder. The achievability of the optimal tradeoff, for any ***fixed*** bias term, under time-out lattice sequential decoding is summarized in the following theorem:

Theorem 12. *Consider a MIMO ARQ channel under short-power constraint given in (5.2), with M transmit, N receive antennas, a maximum number of ARQ rounds L, an effective multiplexing gain $0 \leq r_e < \min\{M,N\}$. Then, the IR-LAST coding scheme under time-out lattice stack sequential decoding with parameter Γ_{out} and fixed $b > 0$, achieves the optimal tradeoff: In the case of long-term static channel*

$$d^*_{ls}(r_e, L) = \begin{cases} f\left(\dfrac{r_e}{L}\right), & 0 \leq r_e < \min\{M,N\}; \\ 0, & r_e \geq \min\{M,N\}, \end{cases}$$

which is achieved with code block length $T \geq \lceil (M+N-1)/L \rceil$, where $f(r) = (M-r)(N-r)$. In the case of short-term static channel

$$d^*_{ss}(r_e, L) = \begin{cases} Lf\left(\dfrac{r_e}{L}\right), & 0 \leq r_e < \min\{M,N\}; \\ 0, & r_e \geq \min\{M,N\}, \end{cases}$$

which is achieved with code block length $T \geq M+N-1$. The optimal tradeoff is achieved subject to the condition

$$\Gamma_{\text{out}} \geq m + \sum_{k=1}^{m} \frac{(4\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MTL(1+\zeta\log\rho)]^{k/2}}{\det(\boldsymbol{R}^{(\ell)\mathsf{T}}_{kk}\boldsymbol{R}^{(\ell)}_{kk})^{1/2}}. \tag{5.46}$$

Proof. see Appendix F. □

It must be noted that the above theorem is *only* valid for non-zero, but fixed values of b. Although one can fully characterize the achievable tradeoff in terms of the bias term[3] (see Section 5. 2. 2), for non-fixed bias, it is not yet clear how the time-out parameter $\Gamma_{\rm out}$ changes with the bias term for variable b. Therefore, in what follows, and for the purpose of completing the analysis, we will only consider the case of fixed bias term.

Now, since $\Gamma_{\rm out}$ depends on the channel statistics (i.e., it is random), it would be desirable to determine its average value (averaged over channel statistics when it is not in outage). This may shed the light on determining the optimal value of $\Gamma_{\rm out}$ that can be used to achieve the optimal tradeoff of the channel. This can be done as follows:

Consider the long-term static channel. Therefore, the optimal average value of $\Gamma_{\rm out}$, say $\overline{\Gamma^*}_{\rm out}$, may be asymptotically lower bounded by

$$\overline{\Gamma^*}_{\rm out} \doteq m + \sum_{k=1}^{m} (\log\rho)^{k/2} \mathsf{E}_{\boldsymbol{\alpha}\notin\mathcal{O}_{ls}(\ell)} \left\{ \det(\boldsymbol{R}_{kk}^{(\ell)\mathsf{T}} \boldsymbol{R}_{kk}^{(\ell)})^{-1/2} \right\}, \tag{5.47}$$

In this work, we focus our analysis on nested LAST codes, specifically LAST codes that are generated using construction A that is described below.

We consider the Loeliger ensemble of mod-p lattices, where p is a prime. First, we generate the set of all lattices given by

$$\Lambda_p = \kappa(\mathsf{C} + p\mathbb{Z}^{2MTL})$$

where $p \to \infty$, $\kappa \to 0$ is a scaling coefficient chosen such that the fundamental volume $V_f = \kappa^{2MTL} p^{2MTL-1} = 1$, $\mathbb{Z}_p$ denotes the field of mod-p integers, and $\mathsf{C} \subset \mathbb{Z}_p^{2MTL}$ is a linear code over $\mathbb{Z}_p$ with generator matrix in systematic form $[\boldsymbol{I}\ \boldsymbol{P}^{\mathsf{T}}]^{\mathsf{T}}$. We use a pair of self-similar lattices for nesting. We take the shaping lattice to be $\Lambda_s = \zeta\Lambda_p$, where ζ is chosen such that the covering radius is $1/2$ in order to satisfy the input power constraint. Finally, the coding lattice is obtained as $\Lambda_c = \rho^{-r/2ML}\Lambda_s$. Interestingly, one can construct a generator matrix of Λ_p as (see [6])

$$\boldsymbol{G}_p = \kappa \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{P} & p\boldsymbol{I} \end{pmatrix}, \tag{5.48}$$

which has a lower triangular form. In this case, one can express the generator matrix of Λ_c as $\boldsymbol{G} = \rho^{-r/2ML}\boldsymbol{G}'$, where $\boldsymbol{G}' = \zeta\boldsymbol{G}_p$. Thanks to the lower triangular format of $\boldsymbol{G}$. If

[3] For example, if we let b to vary with SNR and channel statistics as given in (5.21) one can achieve an asymptotic SNR exponent $d^*_{ls}(r_e, L) = d_b(r_e/L)$ and $d^*_{ss}(r_e, L) = Ld_b(r_e/L)$ for the long-term and short term static channels, respectively, where $d_b(r)$ is as described in Section 5. 2. 2.

$\boldsymbol{M}$ is an $m \times m$ arbitrary full-rank matrix, where $m = 2MTL$, and $\boldsymbol{G}$ is an $m \times m$ lower triangular matrix, then one can easily show that

$$\det[(\boldsymbol{MG})_{kk}] = \det(\boldsymbol{M}_{kk})\det(\boldsymbol{G}_{kk}), \tag{5.49}$$

where $(\boldsymbol{MG})_{kk}$, $\boldsymbol{M}_{kk}$, and $\boldsymbol{G}_{kk}$, are the lower $k \times k$ part of $\boldsymbol{MG}$, $\boldsymbol{M}$, and $\boldsymbol{G}$, respectively.

Using the above result, one can express the determinant that appears in (5.47) as

$$\det(\boldsymbol{R}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{R}_{kk}^{(\ell)}) = \det(\boldsymbol{B}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{B}_{kk}^{(\ell)})\det(\boldsymbol{G}_{kk}^{\mathsf{T}}\boldsymbol{G}_{kk}) = \rho^{-rk/2ML}\det(\boldsymbol{R}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{R}_{kk}^{(\ell)})\det(\boldsymbol{G'}_{kk}^{\mathsf{T}}\boldsymbol{G'}_{kk}), \tag{5.50}$$

Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ be the ordered nonzero eigenvalues of $\boldsymbol{B}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{B}_{kk}^{(\ell)}$, for $k = 1, \cdots, m$. Then,

$$\det(\boldsymbol{B}_{kk}^{(\ell)\mathsf{T}}\boldsymbol{B}_{kk}^{(\ell)}) = \prod_{j=1}^{k}\mu_j$$

Note that for the special case when $k = m$ we have $\mu_{2(j-1)TL+1} = \cdots = \mu_{2jTL} = 1 + \rho\lambda_j((\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c)$, for all $j = 1, \cdots, M$.

Denote $\alpha'_i = -\log\mu_i/\log\rho$. Using (5.49), one can asymptotically express $\overline{\Gamma^*}_{\text{out}}$ as

$$\overline{\Gamma^*}_{\text{out}} = m + \sum_{k=1}^{m}(\log\rho)^{k/2}\mathsf{E}_{\boldsymbol{\alpha}\notin\mathcal{O}_{ls}(\ell)}\{\rho^{c_k}\}, \tag{5.51}$$

where

$$c_k = \frac{1}{2}\sum_{j=1}^{k}\left(\frac{r_1}{ML} - \alpha'_j\right)^+. \tag{5.52}$$

Now, since c_k is non-decreasing in k, we have

$$\overline{\Gamma^*}_{\text{out}} = m + (\log\rho)^{m/2}\mathsf{E}_{\boldsymbol{\alpha}\notin\mathcal{O}_{sl}(L)}\{\rho^{c_m}\}, \tag{5.53}$$

where

$$c_m = TL\sum_{i=1}^{M}\left(\frac{r_1}{ML} - (1-\alpha_i)^+\right)^+.$$

At multiplexing gain r_1, we have the channel is in outage only when $\sum_{j=1}^{M}(1-\alpha_j)^+ \leq$

r_1/L. In this case, we have

$$\begin{aligned}\mathsf{E}_{\boldsymbol{\alpha}\notin\mathcal{O}_{sl}(L)}\{\rho^{c_m}\} &= \int\limits_{\boldsymbol{\alpha}\notin\mathcal{O}_{sl}(L)} \rho^{c_m} f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \\ &\doteq (\log\rho)^{m/2} \int\limits_{\boldsymbol{\alpha}\notin\mathcal{O}_{sl}(L)} \exp\Bigg(\log\rho\Bigg[TL\sum_{i=1}^{M}\left(\frac{r_1}{ML}-(1-\alpha_i)^+\right)^+ - \\ &\qquad \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\Bigg]\Bigg)\, d\boldsymbol{\alpha} \\ &\doteq (\log\rho)^{m/2}\rho^{l(r_1)},\end{aligned}$$

where $\mathcal{O}_{sl}(L) = \left\{\boldsymbol{\alpha}\in\mathbb{R}_+^M : \sum_{i=1}^M (1-\alpha_i)^+ < r_1/L\right\}$, and

$$l(r_1) = \max_{\boldsymbol{\alpha}\notin\mathcal{O}_{sl}(L)} \left[TL\sum_{i=1}^{M}\left(\frac{r_1}{ML}-(1-\alpha_i)^+\right)^+ - \sum_{i=1}^{M}(2i-1+N-M)\alpha_i\right]. \tag{5.54}$$

It is not so difficult to see that the optimal channel coefficients that maximize (5.54) are

$$\alpha_i^* = 1, \quad \text{for } i = 1,\cdots,M-k,$$

and

$$\alpha_i^* = 0, \quad \text{for } i = M-k+1,\cdots,M.$$

Substituting $\boldsymbol{\alpha}^*$ in (5.54), we get

$$l(r_1) = \frac{Tr_1}{M}\left(M-\frac{r_1}{L}\right) - \left(M-\frac{r_1}{L}\right)\left(N-\frac{r_1}{L}\right), \tag{5.55}$$

for $r_1 = 0,1,\cdots,M$. An since $r_e \doteq r_1$, the asymptotic average computational complexity, when the channel is in outage, can be expressed as[4]

$$\overline{\Gamma^*}_{\text{out}} = 2MTL + (\log\rho)^{MTL}\rho^{l(r_e)}. \tag{5.56}$$

One interesting special case of computing the optimal average time-out parameter $\overline{\Gamma^*}_{\text{out}}$ is when $r_e = 0$, i.e., when using a code with fixed rate R_1. In this case we have

$$\overline{\Gamma^*}_{\text{out}} = 2MTL + \frac{(\log\rho)^{MTL}}{\rho^{MN}}. \tag{5.57}$$

[4]As a reminder, the logarithm that appears in the complexity analysis is to the base 2.

Thus, the above equation describes how the time-out limit is related to the system parameters (M, N, T, L, ρ) at the high SNR regime. As an example, consider a 2×2 MIMO ARQ channel with $L = 2$ rounds. Then, according to Theorem 2, we have $T = 3$ is sufficient to achieve the optimal tradeoff. The signal dimension in this case is $m = 24$. Assume $\rho = 100$ (20 dB). According to (5.56), the optimal time-out limit is given by $\overline{\Gamma^*}_{\text{out}} \geq 98$. As will be shown in the sequel, this theoretical result closely matches the value of $\overline{\Gamma^*}_{\text{out}}$ that is obtained experimentally. Typical values of b that corresponds to $\overline{\Gamma^*}_{\text{out}} \approx 98$ are between 0.6 and 1. For very small values of b, the average number of computations increases and according to that $\overline{\Gamma^*}_{\text{out}} \gg 98$.

To see the great advantage of using the time-out lattice sequential decoder with constant bias term over the incomplete list lattice decoder implemented via sphere decoding algorithms, we compare the average computational complexity of both decoders. However, it should be noted here that the sphere decoder is allowed to time-out according to the results that were derived in Chapter 3, Section 3.4, in order to get a fair complexity comparison between both decoders. It has been shown in Chapter 3 (see Section 3.4) that, for high SNR, the average computations performed by the sphere decoder when the channel is not in outage, say Γ_{sphere} for a system with $m = 2MTL$ signal dimension is given by (assuming fixed rate $r_1 = 0$)

$$\Gamma_{\text{sphere}} = 2MTL + \frac{(\log \rho)^{2MTL}}{\rho^{MN}}. \tag{5.58}$$

The ratio of the average complexity of both decoder, say γ, is given by

$$\gamma = \frac{\Gamma_{\text{sphere}}}{\overline{\Gamma^*}_{\text{out}}} = \frac{2MTL + (\log \rho)^{2MTL}/\rho^{MN}}{2MTL + (\log \rho)^{MTL}/\rho^{MN}}.$$

This is a huge saving in computational complexity, especially for large signal dimensions even at very high SNR. For example, at $\rho = 10^8$ (80 dB), $\gamma \approx 7.4$, i.e., the list sphere decoder's complexity is about 7 times larger than the complexity of the new proposed time-out lattice sequential decoder. For $\rho < 80$ dB, one would expect the ratio $\gamma \gg 7$. For extremely high SNR values (e.g., $\rho \geq 90$ dB), it seems that $\gamma \to 1$ as $\rho \to \infty$.

Moreover, some interesting remarks about the effect of extremely decreasing or increasing the value of b on the performance-throughput-complexity tradeoff are discussed next. For very large SNR values, it is

It is well-known that as $b \to \infty$, lattice sequential decoders based on Schnnor-Euchner enumeration converts to the MMSE-DFE decoder [39]. In terms of the total number of visited nodes, MMSE-DFE decoder achieves *linear* computational complexity in m. In this case, for any $\Gamma_{\rm out} > m$, the message will always be decoded from the first round. Although it achieves high throughput and is computationally efficient, this decoder cannot achieve the optimal tradeoff. Assuming $N \geq M$, the maximum SNR exponent that such a decoder can achieve is $(N-M+1)(1-r_e/ML)^+$ (see [35] for more details about MMSE decoding for the case of MIMO channel with no ARQ).

On the other hand, as $b \to 0$ the decoder achieves the best performance. However, the decoding complexity becomes equivalent to (and for some cases worst than) the complexity of lattice (sphere) decoding which is extensively large. Our main objective of using lattice sequential decoding is to save on decoding complexity. Therefore, one should appropriately select the bias term b so that the optimal tradeoff is achieved while reducing the computational complexity. This may be achieved by the sequential decoder by ensuring that the path metric along the correct path increases on average, while decreases along other paths. In this case, we choose b such that $\mathsf{E}_{\boldsymbol{e}'}\{\mu(\boldsymbol{z}_1^k)\} > 0$ (assuming $\boldsymbol{z}_1^m$ is the correct path). This corresponds to $b > \mathsf{E}\{\|[\boldsymbol{e}']_i\|^2\} = 1/2$. This fact is verified experimentally as will be shown in the next section.

As discussed earlier, the value of the parameter $\Gamma_{\rm out}$ used in the time-out algorithm is critical for achieving the optimal tradeoff of the MIMO ARQ system. In order to achieve the optimal tradeoff, both b and $\Gamma_{\rm out}$ have to be appropriately selected in the time-out algorithm. Remember that the probability of error is upper bounded by

$$P_e \leq \sum_{\ell=1}^{L-1} \underbrace{\Pr(E_\ell, \mathcal{A}_\ell)}_{\text{controlled by } \Gamma_{\rm out}} + \underbrace{\Pr(E_L)}_{\text{controlled by } b} \tag{5.59}$$

where $\Pr(E_\ell, \mathcal{A}_\ell) \leq \rho^{-f(r_e/L)}$ is the probability of undetected error at round ℓ and is mainly controlled by the time-out parameter $\Gamma_{\rm out}$. However, the second term of the RHS of the above upper bound is affected by the value of the bias term b. Therefore, both b and $\Gamma_{\rm out}$ have to be appropriately selected in the time-out algorithm so that a balance is obtained which may lead to achieving the optimal tradeoff of the channel.

Suppose that an optimal value of b is set in the time-out algorithm. Then, choosing a very small or large value of $\Gamma_{\rm out}$ may result in a loss of the optimal tradeoff. This is because $\Gamma_{\rm out}$ can be seen as the parameter responsible for the amount of retransmission when error is detected. Choosing a large value of $\Gamma_{\rm out}$ reduces the probability of retransmission and may result in large performance degradation. In this case, as $\Gamma_{\rm out} \to \infty$, we have $\Pr(\mathcal{A}_\ell) \to 1$. Therefore, the undetected error probability becomes equivalent to (for the long-term static channel)

$$\Pr(E_\ell, \mathcal{A}_\ell) = \Pr(E_\ell) \doteq \rho^{-f(r_1/\ell)}.$$

Moreover, the probability of retransmission given the channel is not in outage approaches 0, which means that $r_e = r_1$ ($p(\ell) \to 0$). And since the average error probability defined in (5.8) is dominated by the term with the smallest SNR exponent, we have

$$P_e(\rho) \dot{\leq} \rho^{-f(r_1/L)} + \sum_{\ell=1}^{L} \rho^{-f(r_1/\ell)} \doteq \rho^{-f(r_e)}. \tag{5.60}$$

This is equivalent to the performance of the MIMO channel with no ARQ. On the other hand, choosing a very small value of Γ_{out} improves the performance[5] at the expense of large throughput loss. In this case, the undetected error probability approaches 0, $p(\ell) \to 1$, and the throughput $\eta \to R_1/L$, i.e., $r_e = r_1/L$. The error probability at high SNR in this case is also given by (5.60).

Therefore, in order to achieve the optimal tradeoff of the MIMO ARQ channel, first b must be chosen to ensure that the achievability of the optimal diversity-multiplexing tradeoff when operating over the whole received signal (i.e., at round L). Then, Γ_{out} is selected accordingly so that the optimal diversity-multiplexing-delay tradeoff is achieved.

5.4 Simulation Results

In our simulation, we consider a long-term static MIMO ARQ link with $M = N = L = 2$, $T = 3$ and $R_1 = 8$ bits per channel use. The incremental redundancy LAST code is obtained as an (m, p, k) Loeliger construction. The frame error rate and computational complexity are plotted in Figure 5.4 and Figure 5.5, respectively, for different values of b used in the time-out algorithm. We measure the computational complexity of the joint lattice sequential decoder as the total number of nodes visited by the decoder during the search, accumulated over all ARQ rounds, until a new transmission is started. For every value of b, the optimal value of Γ_{out}, denoted by Γ^*_{out}, was found via simulation by trial and error. Using those optimal parameters, it is shown that the IR-LAST coding scheme decoded using the time-out lattice sequential decoder can achieve probability of error very close to the one that corresponds to the same IR-LAST coding scheme decoded using the list lattice decoder. This is achieved with significant reduction in complexity compared to the list lattice decoder (see Figure 5.5). It is interesting to see how such IR-LAST coding scheme can achieve probability of error close to the coherent LAST code with half the

[5] Performance improvement is achieved as a coding gain and not in the SNR exponent. In this case, the decoder will accumulate more information about the message before making a decision which is due to the fact that most of the time the decoder asks for retransmission.

rate (4 bpcu). On the other hand, the effective rate, R_e, of the IR-LAST coding scheme decoded under the new proposed decoder is shown to approach $R_1 = 8$ as SNR grows as predicted by the theory. Optimal values of Γ_{out} for some special cases of b are provided in Table 1. As expected, for values of $b < 1/2$, the average computational complexity of the time-out algorithm increases and as a consequence, the value of Γ^*_{out} is proportionally increased. Simulation results demonstrate the excellent performance-complexity tradeoff achieved by the proposed algorithm for all values of b, especially at the moderate-to-high SNR regime (see Figure. 5.5).

Table 5.1: Optimum values of Γ_{out} for some special cases of b used in the time-out algorithm for the case of $M = N = L = 2$ and $T = 3$ MIMO ARQ system using IR-LAST random code

b	Γ^*_{out}
0.6	100
0.4	800
0.1	4×10^4

The error rates are obtained by averaging over at least 10 000 channel realizations at small SNRs and as much channel realization as required to count at least 100 frame errors at high SNRs. It is clear that for values of $b > 1/2$ the decoder exhibits some performance degradation compared to the incomplete list sphere decoder. To improve the performance one need to resort to smaller values of b at the price of increasing computational complexity (see Figure 5.4 and Figure 5.5). In general and for all finite values of b, the time-out lattice stack sequential decoder has much lower computational complexity compared to the incomplete list sphere decoder.

5.5 Summary

In this Chapter, we have demonstrated, analytically and via simulation, the significant improvements achieved by the lattice stack sequential decoder over the incomplete list lattice decoder that is used for joint error detection and correction in the IR-LAST MIMO ARQ channel. A time-out algorithm has been proposed. Theoretical analysis and simulation results show that the optimal tradeoff can be achieved by such algorithm with very low decoding complexity compared to the list lattice decoder, especially for moderate-to-high SNR for which the list output could be extensively large.

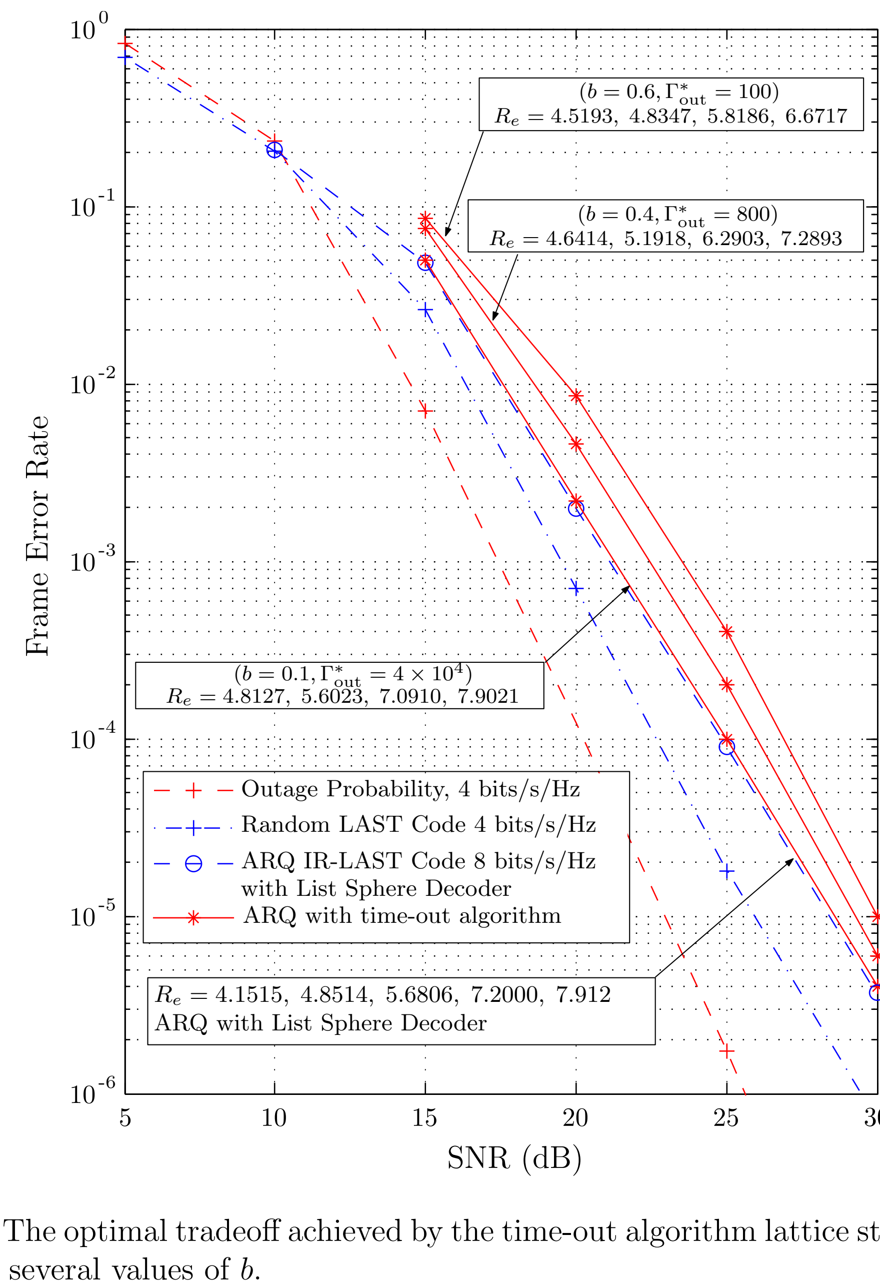


Figure 5.4: The optimal tradeoff achieved by the time-out algorithm lattice stack sequential decoder for several values of b.

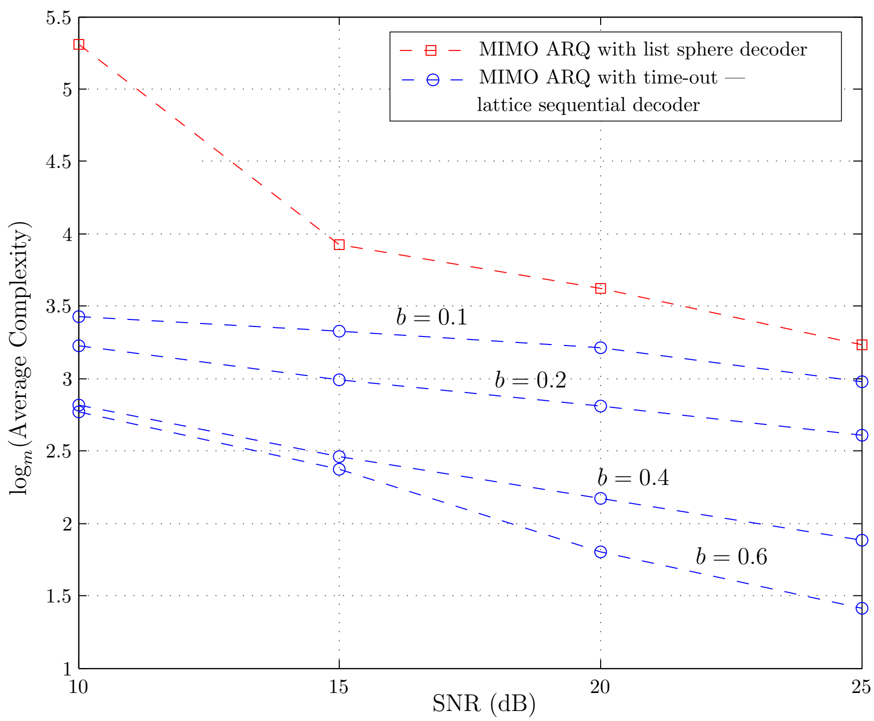


Figure 5.5: Comparison of average computational complexity of the MMSE-DFE list lattice decoder and the time-out lattice stack sequential decoder for several values of b using their corresponding optimal values of Γ_{out} (see Fig. 4).

Chapter 6

Discussion and Conclusion

6.1 Summary

The search for low computational complexity decoders for any communication system is still an active and ongoing research topic up to these days. However, it is considered a very challenging problem.

In this thesis, our main objective is to analyze the computational complexity of some efficient algorithms that perform lattice decoding for lattice codes applied to several wireless communication channels. In general, the sphere decoder is commonly used in communication systems that can be well-described by the linear Gaussian vector channel model. It is well-known that such a decoder achieves ML or near-ML performance. However, the exact complexity analysis of the basic sphere decoder for general space-time codes applied to MIMO wireless channel is known to be difficult. In Chapter 3, we have analyzed the computational complexity of sphere decoding for the quasi-static, LAST coded MIMO channel. Specifically, we have derived an upper bound of the tail distribution of the decoder's computational complexity. We have shown that, when the computational complexity exceeds a certain limit, this upper bound becomes dominated by the outage probability achieved by LAST coding and lattice (sphere) decoding schemes. We have calculated the minimum average computational complexity that is required by the decoder to achieve near optimal performance in terms of the system performance. Moreover, we have shown analytically how MMSE-DFE can significantly improve the tail exponent and as a consequence reduces (average) computational complexity.

However, sphere decoders are only efficient in the high SNR regime and low signal dimensions, and exhibits exponential (average) complexity for low-to-moderate SNR and large signal dimensions. The problem of designing low complexity receivers for the MIMO channel that achieve *near-optimal* performance is considered a challenging problem. In Chapter 4, the asymptotic performance of the lattice sequential decoder for LAST MIMO channel has been analyzed. We have determined the rates achievable by lattice coding and sequential decoding applied to such a channel. The diversity-multiplexing tradeoff under lattice sequential decoding has been derived as a function of its parameter—*the bias term*, which is critical for controlling the amount of computations required at the decoding stage. Achieving low decoding complexity requires increasing the value of the bias term. However, this is done at the expense of losing the optimal tradeoff of the channel. In this work, we have also derived the tail distribution of the decoder's computational complexity in the high signal-to-noise ratio regime. Our analysis reveals that the tail distribution of such low complexity decoder is dominated by the outage probability of the channel for the underlying coding scheme. Also, the tail exponent of the complexity distribution is shown to be equivalent to the diversity-multiplexing tradeoff achieved by lattice coding and sequential decoding schemes. We have shown analytically how MMSE-DFE can significantly improve the tail exponent and as a consequence reduces computational complexity. In particular, we have shown that there exists a *cut-off* multiplexing gain for which the average computational complexity of the decoder remains bounded.

Finally, in Chapter 5, we have considered some applications to lattice sequential decoding for the MIMO ARQ channel. We have proposed an efficient approach for joint detection and decoding based on lattice stack sequential decoding. We have implemented a time-out algorithm at the decoder to predict in advance the occurrence of high channel noise. This results in less wasted time trying to decode a noisy signal and hence improving upon decoding complexity. We have shown that the optimal tradeoff of the channel can be achieved using such a decoder with significant reduction in (average) decoding complexity. We have demonstrated, via analysis and simulation, the significant improvements achieved by the Fano and stack sequential decoders over the incomplete list lattice decoder that is used for joint error detection and correction in the MIMO ARQ channel.

6.2 Suggestion for Further Research

In this work we were able to provide some mathematical analysis regarding the computational complexity of the lattice sequential decoder for the outage-limited MIMO channels. However, our analysis is not complete in the sense that the analysis has been performed for the high-SNR regime. It would be desirable to study the complexity behavior of both sphere decoding and lattice sequential decoding for practical values of SNR (i.e., for low-to-moderate SNR).

Another important issue we would be interesting to investigate for future work is the implementation of a power control algorithm to the MIMO ARQ scheme. Our work analyze the MIMO ARQ system with constant power at the transmitter. In long-term static channel, constant power at the transmitter does not improve diversity at low multiplexing rate. This problem has been solved in [41] by implementing a power control algorithm. Mathematical analysis revealed that diversity order at low multiplexing rate with power control algorithm increases without any additional feedback beyond the standard one-bit ARQ feedback signal. The implementation of power control algorithm for the case of sequential decoding is considered for future work.

Multi-Level Feedback MIMO ARQ Systems: Our MIMO ARQ system model uses only a one-bit feedback for success/failure indication. Recently [64], there has been some interest in MIMO ARQ systems with multi-level feedback. A significant part of such system depends on the accumulative nature of the incremental redundancy ARQ scheme. It has been shown in [64] that performance improvements are possible when additional information is provided through the feedback link. The design of IR-LAST codes for the case of multi-level feedback MIMO ARQ system has not yet been studied. It would interesting to analyze both the performance improvements and computational complexity of the multi-level feedback MIMO ARQ system when sequential decoder is considered. One could examine the advantages of having such low complexity decoders on the performance of the system and try to optimize the search algorithm when multi-level feedback is considered.

Appendix A

Proof of Lemma 1

Without loss of generality, we assume that all-zero lattice point was transmitted. Let

$$\phi'(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{e'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq R_s^2,\ \|\boldsymbol{e'}_1^k\|^2 \leq R_s^2; \\ 0, & \text{otherwise.} \end{cases} \tag{A.1}$$

where $\boldsymbol{e'}_1^k$ is the last k components of $\boldsymbol{e}' = \boldsymbol{Q}^\mathsf{T}\boldsymbol{e}$, and $\boldsymbol{Q}$ is the orthogonal matrix defined in (3.10). Given that $\|\boldsymbol{e}'\|^2 \leq R_s^2$, it must follow that $\|\boldsymbol{e'}_1^k\| \leq R_s^2$, for all $1 \leq k \leq m$. The total number of integer lattice points that satisfy (A.1) is given by

$$C_k = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi'(\boldsymbol{z}_1^k). \tag{A.2}$$

In general one can show that for any random vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, and $R_s > 0$, it holds$\{\|\boldsymbol{u} - \boldsymbol{v}\|^2 \leq R_s^2, \|\boldsymbol{v}\|^2 \leq R_s^2\} \subseteq \{\|\boldsymbol{v}\|^2 \leq 4R_s^2\}$. Therefore, one can easily show that

$$C_k \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \hat{\phi}(\boldsymbol{z}_1^k), \tag{A.3}$$

where

$$\hat{\phi}(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq 4R_s^2; \\ 0, & \text{otherwise.} \end{cases} \tag{A.4}$$

We can further upper bound C_k by introducing an auxiliary random variable that has a uniform distribution in the Voronoi region of the lattice $\Lambda(\boldsymbol{R}_{kk})$. This can be done as follows:

Let

$$\tilde{\phi}(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k) = \begin{cases} 1, & \|\boldsymbol{x}_1^k + \boldsymbol{u}_1^k\|^2 \leq 7R_s^2 \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{u}_1^k$ is a random variable that is uniformly distributed in $\mathcal{V}_{\mathbf{0}}(\boldsymbol{R}_{kk})$ and independent of $\boldsymbol{x}_1^k$. Then, assuming that there exists at least one lattice point $\boldsymbol{x}_1^k \neq \mathbf{0}$ inside the sphere, one can show that

$$C_k \leq \sum_{\boldsymbol{x}_1^k \in \Lambda(\boldsymbol{R}_{kk})} \tilde{\phi}(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k)$$

The indicator function in (A.4) can be rewritten as

$$\begin{aligned} \hat{\phi}(\boldsymbol{x}_1^k) &= \begin{cases} 1, & \|\boldsymbol{x}_1^k\|^2 \leq 4R_s^2,\ \|(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k) - \boldsymbol{u}_1^k\|^2 \leq 4R_s^2 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 1, & \|\boldsymbol{x}_1^k\|^2 \leq 4R_s^2,\ \|(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k)\|^2 \leq 4R_s^2 + 2{\boldsymbol{u}_1^k}^{\mathsf{T}}\boldsymbol{x}_1^k + \|\boldsymbol{u}_1^k\|^2 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where $\boldsymbol{u}_1^k$ is a uniform random variable in the fundamental region of the lattice $\Lambda(\boldsymbol{R}_{kk})$. By noting that $\|\boldsymbol{u}_1^k\|^2 \leq R_s^2$ [since $\boldsymbol{u}_1^k \in \mathcal{V}_{\mathbf{0}}(\Lambda(\boldsymbol{R}_{kk}))$], and ${\boldsymbol{u}_1^k}^{\mathsf{T}}\boldsymbol{x}_1^k \leq \|\boldsymbol{u}_1^k\|\|\boldsymbol{x}_1^k\| \leq R_s^2$ (since $\|\boldsymbol{x}_1^k\| \leq R_s$), we then have

$$\sum_{\boldsymbol{x}_1^k \in \Lambda(\boldsymbol{R}_{kk})} \hat{\phi}(\boldsymbol{x}_1^k) \leq \sum_{\boldsymbol{x}_1^k \in \Lambda(\boldsymbol{R}_{kk})} \tilde{\phi}(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k)$$

Equivalently, we have that

$$C_k \leq \sum_{\boldsymbol{x}_1^k \in \Lambda(\boldsymbol{R}_{kk})} \tilde{\phi}(\boldsymbol{x}_1^k + \boldsymbol{u}_1^k). \tag{A.5}$$

Now, taking the average in both sides of (A.5) over $\boldsymbol{u}_1^k \in \mathcal{V}_{\mathbf{0}}(\boldsymbol{R}_{kk})$ we have (see Lemma 2 in [8])

$$C_k \leq \frac{V(\mathcal{S}_k(\sqrt{7}R_s))}{V_f(\Lambda(\boldsymbol{R}_{kk}))}$$

Appendix B

Proof of Theorem 4

The input to the decoder, after QR preprocessing ($\boldsymbol{HG} = \boldsymbol{QR}$) of (2.5), is given by $\boldsymbol{y}' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{Rz} + \boldsymbol{e}'$, where $\boldsymbol{e}' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{e}$. Let E_s be the event that the lattice Stack sequential decoder makes an erroneous detection, conditioned on $\mu_{\min}$, where $\mu_{\min} = \min\{0, b - \|\boldsymbol{e}'^{1}_{1}\|^2, 2b - \|\boldsymbol{e}'^{2}_{1}\|^2, \ldots, bm - \|\boldsymbol{e}'^{m}_{1}\|^2\}$ is the minimum metric that corresponds to the transmitted path. Then, $P_e = \mathsf{E}_{\mu_{\min}}\{\Pr(E_s)\}$ is the frame error rate of the lattice Stack sequential decoder. Without loss of generality, we assume that $N \geq M$.

Due to lattice symmetry, we assume that the all zero codeword $\mathbf{0}$ was transmitted. Now, any sequence $\boldsymbol{x} = \boldsymbol{Gz} \neq \mathbf{0}$, $\boldsymbol{x} \in \Lambda_c$ can be decoded as the closest lattice point by the decoder only if its metric $\mu(\boldsymbol{z}_1^m)$ is greater than $\mu_{\min}$. Therefore, for a given lattice Λ_c,

$$\begin{aligned}\Pr(E_s|\Lambda_c) &\leq \sum_{\boldsymbol{z}\in\mathbb{Z}^m\setminus\{\mathbf{0}\}} \Pr(\mu(\boldsymbol{z}_1^m) > \mu_{\min}) \\ &= \sum_{\boldsymbol{z}\in\mathbb{Z}^m\setminus\{\mathbf{0}\}} \Pr(\|\boldsymbol{e}' - \boldsymbol{Rz}\|^2 < bm - \mu_{\min}).\end{aligned} \tag{B.1}$$

The upper bound in (B.1) follows from the union bound, and due to the fact that in general, $\mu(\boldsymbol{z}_1^m) > \mu_{\min}$ is just a necessary condition for $\boldsymbol{x}$ to be decoded by the lattice Stack sequential decoder. By noticing that $-(\mu_{\min} + \|\boldsymbol{e}'\|^2) \leq 0$, we get

$$\Pr(E_s|\Lambda_c) \leq \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{Hx}\|^2 - 2(\boldsymbol{Hx})^{\mathsf{T}}\boldsymbol{e} < bm), \tag{B.2}$$

where $\Lambda_c^* = \Lambda_c\setminus\{\mathbf{0}\}$. Note the independence of the upper bound (B.2) of $\mu_{\min}$. We would like now to upper bound the term inside the summation in (B.2). Using Chernoff bound,

$$\Pr(\|\boldsymbol{Hx}\|^2 - 2(\boldsymbol{Hx})^{\mathsf{T}}\boldsymbol{e} < bm) \le \begin{cases} e^{-\|\boldsymbol{Hx}\|^2/8} e^{bm/4}, & \|\boldsymbol{Hx}\|^2 > bm; \\ 1, & \|\boldsymbol{Hx}\|^2 \le bm. \end{cases} \tag{B.3}$$

By taking the expectation over the ensemble of random lattices (see [8], Theorem 4),

$$\begin{aligned} \Pr(E_s) = \mathsf{E}_{\Lambda_c}\{\Pr(E_s|\Lambda_c)\} &\le \frac{1}{V_c}\left\{ \int\limits_{\|\boldsymbol{Hx}\|^2 < bm} d\boldsymbol{x} + e^{bm/4} \int\limits_{\|\boldsymbol{Hx}\|^2 > bm} e^{-\|\boldsymbol{Hx}\|^2/8}\, d\boldsymbol{x} \right\} \\ &\le \frac{1}{V_c}\left\{ \frac{\pi^{m/2}(bm)^{m/2}}{\Gamma(m/2+1)\det(\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H})^{1/2}} + \frac{(8\pi)^{m/2} e^{bm/4}}{\det(\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H})^{1/2}} \right\}. \end{aligned} \tag{B.4}$$

Next, we make use of the fact that there exists a shifted lattice code $\Lambda_c + \boldsymbol{u}_0^*$ with number of codewords inside the shaping region (see [8])

$$|\mathcal{C}(\Lambda_c, \boldsymbol{u}_0^*, \mathcal{R})| = 2^{RT} \ge \frac{V(\mathcal{R})}{V_c}.$$

Also, it is easy to verify that

$$\det(\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H}) = \left(\det\left(\rho(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c\right)\right)^{2T}.$$

Denote $R = r\log\rho$ and $0 \le \lambda_1 \le \cdots \le \lambda_M$ the eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c$, then, the bound (B.4) can be rewritten as (conditioned on channel statistics)

$$\Pr(E_s|\boldsymbol{\nu}) \dot{\le} \mathcal{K}(m,b)\rho^{-T[M-\sum_{j=1}^{M}(1-\nu_j)^+ - r]}, \tag{B.5}$$

where $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_M)$, $\nu_i \triangleq -\log\lambda_i/\log\rho$, $(x)^+ = \max\{0, x\}$, and $\mathcal{K}(m,b)$ is a constant independent of ρ. Now, define the set

$$\mathcal{A} = \left\{ \boldsymbol{\nu} \in \mathbb{R}_+^M : \nu_1 \ge \cdots \ge \nu_M \ge 0,\ \sum_{i=1}^{M} \nu_i > M - r \right\}. \tag{B.6}$$

Using (B.6), the probability of error can be upper bounded as follows:

$$\Pr(E_s) \le \Pr(\boldsymbol{\nu} \in \mathcal{A}) + \Pr(E_s, \boldsymbol{\nu} \in \overline{\mathcal{A}}). \tag{B.7}$$

The behaviour of the first term at high SNR is $\rho^{-d(r)}$. Averaging the second term over the channels in $\overline{\mathcal{A}}$ set, we obtain (see [23]),

$$\Pr(E_s) \quad \dot{\le} \quad \rho^{-d(r)} + \int_{\overline{\mathcal{A}}} f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \Pr(E_s|\boldsymbol{\nu})\, d\boldsymbol{\nu}$$

$$\dot{\leq} \quad \rho^{-d(r)}, \tag{B.8}$$

where $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ is the joint probability density function of $\boldsymbol{\nu}$ which, for all $\boldsymbol{\nu} \in \overline{\mathcal{A}}$, is asymptotically given by (see [23])

$$f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \doteq \exp\left(-\log(\rho) \sum_{i=1}^{\min\{M,N\}} (2i - 1 + |N - M|)\nu_i\right). \tag{B.9}$$

By definition, the error probability of the lattice sequential decoder is lower bounded by the probability of error of the lattice decoder (ld) knowing the channel matrix $\boldsymbol{H}^c$. Hence, it can be easily shown that (see [23])

$$\Pr(E_s) \geq \Pr(E_{\mathrm{ld}}) \doteq \rho^{-d(r)}. \tag{B.10}$$

Appendix C

Proof of Theorem 5

The input to the decoder, after QR preprocessing ($\boldsymbol{BG} = \boldsymbol{QR}$) of (2.5), is given by $\boldsymbol{y}'' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{y}' = \boldsymbol{Rz} + \boldsymbol{e}''$, where $\boldsymbol{e}'' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{e}'$. Let E_s be the event that the lattice Stack sequential decoder makes an erroneous detection, conditioned on $\mu_{\min}$, where $\mu_{\min} = \min\{0, b - \|\boldsymbol{e}'^1_1\|^2, 2b - \|\boldsymbol{e}'^2_1\|^2, \ldots, bm - \|\boldsymbol{e}'^m_1\|^2\}$ is the minimum metric that corresponds to the transmitted path. Then, $P_e = \mathsf{E}_{\mu_{\min}}\{\Pr(E_s)\}$ is the frame error rate of the lattice Stack sequential decoder.

Due to lattice symmetry, we assume that the all zero codeword $\mathbf{0}$ was transmitted. Now, any sequence $\boldsymbol{x} = \boldsymbol{Gz} \neq \mathbf{0}$, $\boldsymbol{x} \in \Lambda_c$ can be decoded as the closest lattice point by the decoder only if its metric $\mu(\boldsymbol{z}_1^m)$ is greater than $\mu_{\min}$. Therefore, for a given lattice Λ_c,

$$\begin{aligned}\Pr(E_s|\Lambda_c) &\leq \sum_{\boldsymbol{z}\in\mathbb{Z}^m\backslash\{\mathbf{0}\}} \Pr(\mu(\boldsymbol{z}_1^m) > \mu_{\min}) \\ &= \sum_{\boldsymbol{z}\in\mathbb{Z}^m\backslash\{\mathbf{0}\}} \Pr(\|\boldsymbol{e}'' - \boldsymbol{Rz}\|^2 < bm - \mu_{\min}).\end{aligned} \tag{C.1}$$

The upper bound in (C.1) follows from the union bound, and due to the fact that in general, $\mu(\boldsymbol{z}_1^m) > \mu_{\min}$ is just a necessary condition for $\boldsymbol{x}$ to be decoded by the lattice Stack sequential decoder. By noticing that $-(\mu_{\min} + \|\boldsymbol{e}''\|^2) \leq 0$, we get

$$\Pr(E_s|\Lambda_c) \leq \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{Bx}\|^2 - 2(\boldsymbol{Bx})^{\mathsf{T}}\boldsymbol{e}' < bm), \tag{C.2}$$

where $\Lambda_c^* = \Lambda_c\backslash\{\mathbf{0}\}$. Note the independence of the upper bound (C.2) of $\mu_{\min}$. We would like now to upper bound the term inside the summation in (C.2). The difficulty here stems from the non-Gaussianity of the random vector $\boldsymbol{e}'$ for any finite T. To overcome this problem, consider the following:

Let

$$\tilde{\boldsymbol{e}} = [\boldsymbol{B} - \boldsymbol{F}\boldsymbol{H}]\boldsymbol{g} + \boldsymbol{F}(\boldsymbol{w} + \boldsymbol{w}_1),$$

where $\boldsymbol{g} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_m)$, $\boldsymbol{w}_1 \sim \mathcal{N}(0, (\sigma^2 - 1/2)\boldsymbol{I}_m)$ and $\sigma^2 \geq 1/2$. Following the footsteps of [23], it can be shown that by appropriately constructing a nested LAST code we have that

$$\Pr(E_s|\Lambda_c) \leq \beta_m \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{B}\boldsymbol{x}\|^2 - 2(\boldsymbol{B}\boldsymbol{x})^{\mathsf{T}} \tilde{\boldsymbol{e}} < bm), \tag{C.3}$$

where $\tilde{\boldsymbol{e}} \sim \mathcal{N}(0, 0.5\boldsymbol{I}_m)$, and β_m is a constant independent of ρ. Using Chernoff bound,

$$\Pr(\|\boldsymbol{B}\boldsymbol{x}\|^2 - 2(\boldsymbol{B}\boldsymbol{x})^{\mathsf{T}} \tilde{\boldsymbol{e}} < bm) \leq \begin{cases} e^{-\|\boldsymbol{B}\boldsymbol{x}\|^2/8} e^{bm/4}, & \|\boldsymbol{B}\boldsymbol{x}\|^2 > bm; \\ 1, & \|\boldsymbol{B}\boldsymbol{x}\|^2 \leq bm. \end{cases} \tag{C.4}$$

By taking the expectation over the ensemble of random lattices (see [8], Theorem 4),

$$\begin{aligned} \Pr(E_s) = \mathsf{E}_{\Lambda_c}\{\Pr(E_s|\Lambda_c)\} &\leq \frac{\beta_m}{V_c} \left\{ \int_{\|\boldsymbol{B}\boldsymbol{x}\|^2 < bm} d\boldsymbol{x} + e^{bm/4} \int_{\|\boldsymbol{B}\boldsymbol{x}\|^2 > bm} e^{-\|\boldsymbol{B}\boldsymbol{x}\|^2/8} \, d\boldsymbol{x} \right\} \\ &\leq \frac{\beta_m}{V_c} \left\{ \frac{\pi^{m/2} (bm)^{m/2}}{\Gamma(m/2+1) \det(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})^{1/2}} + \frac{(8\pi)^{m/2} e^{bm/4}}{\det(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})^{1/2}} \right\}. \end{aligned} \tag{C.5}$$

Next, we make use of the fact that there exists a shifted lattice code $\Lambda_c + \boldsymbol{u}_0^*$ with number of codewords inside the shaping region (see [8])

$$|\mathcal{C}(\Lambda_c, \boldsymbol{u}_0^*, \mathcal{R})| = 2^{RT} \geq \frac{V(\mathcal{R})}{V_c}.$$

Also, it is easy to verify that

$$\det(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}) = \left(\det \left(\boldsymbol{I} + \frac{\rho}{M} (\boldsymbol{H}^c)^{\mathsf{H}} \boldsymbol{H}^c \right) \right)^{2T}.$$

Denote $R = r \log \rho$ and $0 \leq \lambda_1 \leq \cdots \leq \lambda_{\min\{M,N\}}$ the eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}} \boldsymbol{H}^c$, then, the bound (C.5) can be rewritten as (conditioned on channel statistics)

$$\Pr(E_s|\boldsymbol{\nu}) \,\dot{\leq}\, \mathcal{K}(m,b) \rho^{-T[\sum_{j=1}^{\min\{M,N\}} (1-\nu_j)^+ - r]}, \tag{C.6}$$

where $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_{\min\{M,N\}})$, $\nu_i \triangleq -\log \lambda_i / \log \rho$, $(x)^+ = \max\{0, x\}$, and $\mathcal{K}(m,b)$ is a constant independent of ρ. Now, define the set

$$\mathcal{B} = \left\{ \boldsymbol{\nu} \in \mathbb{R}_+^{\min\{M,N\}} : \nu_1 \geq \cdots \geq \nu_{\min\{M,N\}} \geq 0, \sum_{i=1}^{\min\{M,N\}} (1-\nu_i)^+ < r \right\}. \tag{C.7}$$

Using (C.7), the probability of error can be upper bounded as follows:

$$\Pr(E_s) \leq \Pr(\boldsymbol{\nu} \in \mathcal{B}) + \Pr(E_s, \boldsymbol{\nu} \in \overline{\mathcal{B}}). \tag{C.8}$$

The behaviour of the first term at high SNR is $\rho^{-d^*(r)}$. Averaging the second term over the channels in $\overline{\mathcal{B}}$ set,

$$\begin{aligned} \Pr(E_s) &\;\dot{\leq}\; \rho^{-d^*(r)} + \int_{\overline{\mathcal{B}}} f_{\boldsymbol{\nu}}(\boldsymbol{\nu}) \Pr(E_s|\boldsymbol{\nu}) \, d\boldsymbol{\nu} \\ &\;\dot{\leq}\; \rho^{-d^*(r)}, \end{aligned} \tag{C.9}$$

where $f_{\boldsymbol{\nu}}(\boldsymbol{\nu})$ is the joint probability density function of $\boldsymbol{\nu}$ given by (B.9).

By definition, the error probability of the lattice sequential decoder is lower bounded by the probability of error of the lattice decoder (ld) knowing the channel matrix $\boldsymbol{H}^c$. Hence, it can be easily shown that (see [23])

$$\Pr(E_s) \geq \Pr(E_{\text{ld}}) \doteq \rho^{-d^*(r)}. \tag{C.10}$$

Appendix D

Proof of Theorem 7

We consider an ensemble of $2MT$-dimensional random lattices $\{\Lambda_c\}$ with fundamental volume V_c satisfying the Minkowski-Hlawka theorem. The random lattice codebook is $\mathcal{C}(\Lambda, \boldsymbol{u}_0, \mathcal{R})$, for some fixed translation vector $\boldsymbol{u}_0$ and where $\mathcal{R}$ is the $2MT$-dimensional sphere of radius $\sqrt{MT}$ centred at the origin. The average probability of error (average over the channel and lattice ensemble) can be upper bounded as

$$\begin{aligned} \bar{P}_e(\rho) &= \mathsf{E}_\Lambda\{P_e(\rho|\Lambda)\} \\ &\le \mathsf{E}_\Lambda\{\Pr(\text{error}, R_b(\rho) > R(\rho))\} + P_{\text{out}}(\rho, b), \end{aligned} \tag{D.1}$$

where $P_e(\rho|\Lambda)$ is the probability of error for a given choice of Λ. Denote $0 \le \lambda_1 \le \cdots \le \lambda_M$ the eigenvalues of $(\boldsymbol{H}^c)^{\mathsf{H}}\boldsymbol{H}^c$, and let $R = r\log\rho$. As shown in Section 4. 4. 1, by expressing the bias term b as in (4.20), the achievable rate of lattice sequential decoding can be written as $R_b = \log\eta$, where $\eta = \prod_{i=1}^{M}(1+\rho\lambda_i)^{\zeta_i}$. Now, define the outage event $\mathcal{B} = \{\boldsymbol{\beta} \in \mathbb{R}_+^M : \sum_{i=1}^{M}\zeta_i(1-\beta_i)^+ < r\}$, where $\beta_i = -\log\lambda_i/\log\rho$. Then, the second term in the upper bound can be expressed as

$$\begin{aligned} \mathsf{E}_\Lambda\{\Pr(\text{error}, R_b(\rho) > R(\rho))\} &\doteq \int_{\overline{\mathcal{B}}} f_{\boldsymbol{\beta}}(\boldsymbol{\beta})\mathsf{E}_\Lambda\{P_e(\rho|\boldsymbol{\beta},\Lambda)\}\, d\boldsymbol{\beta} \\ &\le \Pr(\|\boldsymbol{e}'\|^2 > MT(1+\gamma)) + \int_{\overline{\mathcal{B}}} f_{\boldsymbol{\beta}}(\boldsymbol{\beta})\Pr(\mathcal{A}|\boldsymbol{\beta})\, d\boldsymbol{\beta}, \end{aligned} \tag{D.2}$$

where $\gamma > 0$, and $f_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ is the joint probability density function of $\boldsymbol{\beta}$ which is asymptotically given by

$$f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \doteq \exp\left(-\log(\rho)\sum_{i=1}^{M}(2i-1+|N-M|)\beta_i\right). \tag{D.3}$$

Consider here the Stack algorithm ($\delta = 0$). In this case, the matrix $\boldsymbol{B}'$ provided in (4.13) can be expressed at high SNR as

$$\tilde{\boldsymbol{B}} = \left(1 - b\rho^{-[\sum_{i=1}^{M}(1-\beta_i)^+ - r]/M}\right)\boldsymbol{B}.$$

Hence, at high SNR we have

$$\det(\tilde{\boldsymbol{B}}^{\mathsf{T}}\tilde{\boldsymbol{B}}) \doteq \left(1 - b\rho^{-[\sum_{i=1}^{M}(1-\beta_i)^+ - r]/M}\right)\rho^{\sum_{i=1}^{M}(1-\beta_i)^+}. \tag{D.4}$$

As $\rho \to \infty$, we can express b [see (4.20)] as

$$b \doteq \frac{\rho^{\sum_{i=1}^{M}(1-\beta_i)^+/M}}{\eta^{1/M}}\left[1 - \left(\frac{\eta}{\rho^{\sum_{i=1}^{M}(1-\beta_i)^+}}\right)^{1/2M}\right]. \tag{D.5}$$

Substituting (D.5) into (D.4), and by realizing that for all $R_b > R$ or equivalently $\eta \dot{>} \rho^r$, we can lower-bound (D.4) as $\det(\boldsymbol{B}'^{\mathsf{T}}\boldsymbol{B}') \geq \eta$. Setting $\boldsymbol{A} = \boldsymbol{B}'$ in Lemma 1, the ambiguity probability can be upper bounded as

$$\Pr(\mathcal{A}|\boldsymbol{\beta}) \dot{\leq} \exp(-T[\log\eta - r\log\rho]). \tag{D.6}$$

It has been shown in [23] that for $T \geq M + N - 1$, the SNR exponent of $\Pr(\|\boldsymbol{e}'\|^2 > MT(1+\gamma))$ with respect to $\log\rho$ is larger than $d_0(r) > d_b(r)$. Substituting (D.6) in (D.2) we get (for $T \geq M + N - 1$)

$$\begin{aligned}
&\mathsf{E}_{\Lambda}\{\Pr(\text{error}, R_b(\rho) > R(\rho))\} \\
&\quad \dot{\leq} \int_{\overline{\mathcal{B}}} \exp\left(-\log(\rho)\sum_{i=1}^{M}(2i-1+|N-M|)\beta_i + T\left[\sum_{i=1}^{M}\zeta_i(1-\beta_i)^+ - r\right]\right) d\boldsymbol{\beta} \\
&\quad \doteq \rho^{-d_b(r)}.
\end{aligned} \tag{D.7}$$

Appendix E

Proof of Lemma 5

Let $\phi'(\boldsymbol{z}_1^k, \ell)$ be the indicator function defined by

$$\phi'(\boldsymbol{z}_1^k, \ell) = \begin{cases} 1, & \text{if } \|\boldsymbol{e}''^{k}_{\ell\,1} - \boldsymbol{R}^{(\ell)}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk - \mu_{\min}(\ell); \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu_{\min}(\ell)$ is the minimum metric along the decoded path. Then, it can be easily verified that

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k, \ell) \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi'(\boldsymbol{z}_1^k, \ell), \tag{E.1}$$

where $\phi(\boldsymbol{z}_1^k, \ell)$ is as defined in (5.25), where a path may be extended by the stack decoder if the partial path metric at that node satisfies $\mu(\boldsymbol{z}_1^k, \ell) \geq \mu_{\min}(\ell)$. Now, given $\|\boldsymbol{e}''_\ell\|^2 \leq MTL(1+\gamma)$, and by noticing that $-(\mu_{\min}(\ell) + \|\boldsymbol{e}''_\ell\|^2) \leq 0$, we obtain

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi'(\boldsymbol{z}_1^k, \ell) \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi''(\boldsymbol{z}_1^k, \ell), \tag{E.2}$$

where

$$\phi''(\boldsymbol{z}_1^k, \ell) = \begin{cases} 1, & \text{if } \|\boldsymbol{e}''^{k}_{\ell\,1} - \boldsymbol{R}^{(\ell)}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk + MTL(1+\gamma); \\ 0, & \text{otherwise.} \end{cases} \tag{E.3}$$

Notice the independence of the upper bound (E.2) on $\mu_{\min}(\ell)$. Now, let

$$\phi_k'''(\boldsymbol{z}, \ell) = \begin{cases} S_k(\ell), & \text{if } \|\boldsymbol{e}''_\ell - \boldsymbol{R}_\ell \boldsymbol{z}\|^2 \leq bm - \mu_{\min}(\ell); \\ 0, & \text{otherwise,} \end{cases}$$

where

$$S_k(\ell) = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{''}(\boldsymbol{z}_1^k, \ell), \tag{E.4}$$

then it can be easily shown that

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k, \ell) \leq \sum_{\boldsymbol{z} \in \mathbb{Z}^m} \phi_k^{'''}(\boldsymbol{z}, \ell) \leq \sum_{\boldsymbol{x} \in \Lambda_c} \tilde{\phi}_k(\boldsymbol{x}, \ell),$$

where

$$\tilde{\phi}_k(\boldsymbol{x}, \ell) = \begin{cases} S_k(\ell), & \text{if } \|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell \boldsymbol{x})^\mathsf{T} \boldsymbol{e}'_\ell \leq bm; \\ 0, & \text{otherwise.} \end{cases}$$

Consider now the following lemma:

Lemma 6. *At the ℓ-th ARQ round, the number of nodes visited by the lattice stack sequential decoder at level $k = 1, \cdots, m$, given that $\|\mathbf{e}'\|^2 \leq MTL(1+\gamma)$, can be upper bounded by (for any finite $b > 0$)*

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k, \ell) \leq S_k(\ell) \leq \frac{(7\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + MTL(1+\gamma)]^{k/2}}{\det(\boldsymbol{R}_{kk}^{(\ell)\mathsf{T}} \boldsymbol{R}_{kk}^{(\ell)})^{1/2}} = S'_k(\ell), \tag{E.5}$$

where $S_k(\ell)$ is as defined in (E.4).

Proof. The proof follows the same footsteps of Lemma 4 in Chapter 4. □

The tail distribution, given the channel is not in outage, can then be upper bounded as follows

$$\begin{aligned}\Pr(\mathcal{N}_m(\ell) \geq \Gamma_{\text{out}} | \overline{\mathcal{O}}_{ls}(\ell)) \leq \Pr(\|\boldsymbol{e}'_\ell\|^2 > & MTL(1+\gamma)) + \\ & \Pr(\mathcal{N}_m(\ell) \geq \Gamma_{\text{out}}, \|\boldsymbol{e}'_\ell\|^2 \leq MTL(1+\gamma) | \overline{\mathcal{O}}_{ls}(\ell)).\end{aligned} \tag{E.6}$$

For a given lattice Λ_c, using Markov inequality, we have

$$\begin{aligned}\Pr(\mathcal{N}_m(\ell) \geq \Gamma_{\text{out}} | \Lambda_c, \overline{\mathcal{O}}_{ls}(\ell), \|\boldsymbol{e}'_\ell\|^2 \leq MTL(1+\gamma)) \\ \leq \Pr(\tilde{\mathcal{N}}_m(\ell) \geq \Gamma_{\text{out}} - m | \Lambda_c, \overline{\mathcal{O}}_{ls}(\ell), \|\boldsymbol{e}'_\ell\|^2 \leq MTL(1+\gamma)) \\ \leq \frac{\mathsf{E}_{\boldsymbol{e}'}\{\tilde{\mathcal{N}}_m(\ell) | \Lambda_c, \overline{\mathcal{O}}_{ls}(\ell), \|\boldsymbol{e}'_\ell\|^2 \leq MTL(1+\gamma)\}}{\Gamma_{\text{out}} - m}, \quad \text{for } \Gamma_{\text{out}} > m,\end{aligned} \tag{E.7}$$

where $\tilde{\mathcal{N}}_m(\ell)$, assuming all-zero codeword was transmitted, is defined as

$$\tilde{\mathcal{N}}_m(\ell) = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k \setminus \{\boldsymbol{0}\}} \phi(\boldsymbol{z}_1^k, \ell).$$

Using Lemma 2, the conditional average of $\tilde{\mathcal{N}}_m(\ell)$ with respect to the noise can be further upper bounded as

$$\mathsf{E}_{\boldsymbol{e}'}\{\tilde{\mathcal{N}}_m(\ell)|\Lambda_c, \overline{\mathcal{O}}_{ls}(\ell), \|\boldsymbol{e}'_\ell\|^2 \le MTL(1+\gamma)\} \le \sum_{k=1}^{m} S'_k(\ell) \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}'_\ell < bm). \tag{E.8}$$

Therefore,

$$\begin{aligned}\Pr(\mathcal{N}_m(\ell) \ge \Gamma_{\text{out}}|\Lambda_c, \overline{\mathcal{O}}_{ls}(\ell), \|\boldsymbol{e}'_\ell\|^2 \le MTL(1+\gamma))& \\ \le \frac{\sum_{k=1}^{m} S'_k(\ell)}{\Gamma_{\text{out}} - m} \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}'_\ell < bm).&\end{aligned} \tag{E.9}$$

We would like now to upper bound the term inside the summation in (E.9). The difficulty here stems from the non-Gaussianity of the random vector $\boldsymbol{e}'_\ell$ for any finite T. To overcome this problem, consider the following:

Let

$$\tilde{\boldsymbol{e}}_\ell = [\boldsymbol{B}_\ell - \boldsymbol{F}_\ell \boldsymbol{H}_\ell]\boldsymbol{g}_\ell + \boldsymbol{F}_\ell(\boldsymbol{e}_\ell + \boldsymbol{w}_\ell),$$

where $\boldsymbol{g}_\ell \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_m)$, $\boldsymbol{w}_\ell \sim \mathcal{N}(0, (\sigma^2 - 1/2)\boldsymbol{I}_m)$ and $\sigma^2 \ge 1/2$. It can be shown that by appropriately constructing a nested LAST code we have that

$$\Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}'_\ell < bm) \le \beta_m \Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\tilde{\boldsymbol{e}}_\ell < bm), \tag{E.10}$$

where $\tilde{\boldsymbol{e}}_\ell \sim \mathcal{N}(0, 0.5\boldsymbol{I}_m)$, and β_m is a constant independent of ρ. Using Chernoff bound,

$$\Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\tilde{\boldsymbol{e}}_\ell < bm) \le \begin{cases} e^{-\|\boldsymbol{B}_\ell\boldsymbol{x}\|^2/8} e^{bm/4}, & \|\boldsymbol{B}_\ell\boldsymbol{x}\|^2 > bm; \\ 1, & \|\boldsymbol{B}_\ell\boldsymbol{x}\|^2 \le bm. \end{cases} \tag{E.11}$$

By taking the expectation over the ensemble of random lattices

$$\mathsf{E}_{\Lambda_c}\left\{\sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(\|\boldsymbol{B}_\ell \boldsymbol{x}\|^2 - 2(\boldsymbol{B}_\ell\boldsymbol{x})^{\mathsf{T}}\boldsymbol{e}'_\ell < bm)\right\} \le \frac{\beta_m}{V_c}\left\{\frac{\pi^{m/2}(bm)^{m/2}}{\Gamma(m/2+1)\det(\boldsymbol{B}_\ell^{\mathsf{T}}\boldsymbol{B}_\ell)^{1/2}} + \frac{(8\pi)^{m/2}e^{bm/4}}{\det(\boldsymbol{B}_\ell^{\mathsf{T}}\boldsymbol{B}_\ell)^{1/2}}\right\}. \tag{E.12}$$

Substituting (5.5) and (5.42) in (E.12), and by selecting $\Gamma_{\text{out}} \geq m + \sum_{k=1}^{m} S'_k(\ell)$, we get

$$\Pr(\mathcal{N}_m \geq \Gamma_{\text{out}} | \overline{\mathcal{O}}_{ls}(\ell)) \,\dot{\leq}\, \rho^{-T\ell[\sum_{i=1}^{M}(1-\alpha_i)^+ - r_1/\ell]}. \tag{E.13}$$

Now, by setting $\gamma = \zeta \log \rho$, the first term in the RHS of (E.6) can be shown to be upper bounded by $\rho^{-f(r_1/\ell)}$, as long as ζ is chosen sufficiently large such that $MTL\zeta \geq f(r_1/\ell)$ (see [41], Appendix IV). Averaging the second term in the RHS of (E.13) over the channels in $\overline{\mathcal{O}}_{ls}(\ell)$ set, we obtain ,

$$\Pr(\mathcal{N}_m(\ell) \geq \Gamma_{\text{out}}) \,\dot{\leq}\, \rho^{-f(r_1/\ell)} + \int_{\overline{\mathcal{O}}_{ls}(\ell)} f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \Pr(\mathcal{N}_m \geq \Gamma_{\text{out}} | \boldsymbol{\alpha}) \, d\boldsymbol{\alpha} \,\dot{\leq}\, \rho^{-f(r_1/\ell)}, \tag{E.14}$$

where $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ is the joint probability density function of $\boldsymbol{\alpha}$ which, for all $\boldsymbol{\alpha} \in \overline{\mathcal{O}}_{ls}(\ell)$, is asymptotically given by (see [23])

$$f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \doteq \exp\left(-\log(\rho) \sum_{i=1}^{M} (2i - 1 + N - M)\alpha_i \right). \tag{E.15}$$

Since $\mathcal{N}_j(\ell) < \mathcal{N}_m(\ell)$, for all $1 \leq j < m$, then

$$\Pr(\mathcal{N}_j(\ell) \geq \Gamma_{\text{out}}) \leq \Pr(\mathcal{N}_m(\ell) \geq \Gamma_{\text{out}}) \,\dot{\leq}\, \rho^{-f(r_1/\ell)}.$$

References

[1] C. E. Shannon, “A Mathematical Theory of Communication”, *Bell Syst. Techn. J.*, vol. 27, pp. 379-423, 623-656, July, October, 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons Inc., New York, 1968.

[3] Forney, G.D., Costello, D.J., “Channel Coding: The Road to Channel Capacity,” *IEEE Trans. Inform. Theory*, vol. 95, pp. 1150-1177, June 2007.

[4] C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” *Bell Syst. Tech. J* , vol. 38, no. 3, pp. 611-656, May 1959.

[5] R. deBuda, “The upper bound of a new near-optimal code”, *IEEE Trans. on Inform. Theory*, vol. IT-21, pp. 441-445, July 1975.

[6] J. H. Conway and N. J. A. Sloane, *SpherePackings, Lattices, and Groups*, 3rd ed. Springer Verlag NewYork, 1999.

[7] G. Polyterv, “On coding without restrictions for the AWGN channel,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 409-417, 1994.

[8] H. Loeliger, “Averaging Bounds for Lattices and Linear Codes”, *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1767-11773, Nov. 1997.

[9] R. Urbanke, and B. Rimoldi, “Lattice Codes Can Achieve Capacity on the AWGN Channel”, *IEEE Trans. on Inform. Theory*, vol. 44, pp. 273-278, Jan. 1998.

[10] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-elements antenna,” *Bell Labs Tech. J. ,* vol. 1, no. 2, pp. 41-59, 1996.

[11] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecommu.* , vol. 10, pp. 585-595, Nov. /Dec. 1999.

[12] L. Ozarow, S. Shamai, and A. Whyner, "Information-theoretic considerations in cellular mobile radio," *IEEE Trans. Veh. Technol.* , vol. 43, pp. 359-378, May 1994.

[13] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-756, Mar. 1998.

[14] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.* , vol. 16, pp. 1451-1458.

[15] V. Tarokh, H. Jafrakhani, and A. R. Calderbak, "Space-time block code from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.

[16] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804-24., Jul. 2002.

[17] F. E. Oggier, G. Rekaya, J. -C. Belfiore and E. Viterbo, "Perfect Space-Time Block Codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3885-3902, Sep. 2006.

[18] J. M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing*, A. Paulraj, V. Roychowdhury, and C. Schaper, Eds. Kluwer, 1997, ch. 4, pp. 79 127.

[19] S. Sandhi A. Paulraj, "Space time block coding: A capacity prespective", *IEEE Comm. Letters*, vol. 4, no. 12, December 2000.

[20] H. El-Gamal, and M. O. Damen, "Universal Space-Time Coding", *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1097-1119, May 2003.

[21] M. O. Damen, H. El-Gamal, N. C. Beaulieu, "Linear threaded algebraic space-time constellations", *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2372-2388, October 2003.

[22] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[23] H. El-Gamal, G. Caire, M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels", *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 968-985, June 2004.

[24] U. Erez and R. Zamir, "Lattice coding can achieve $1/2\log(1+snr)$ on the AWGN channel using nested codes", *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2293-2314, Oct. 2004.

[25] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3401-3416, Oct. 2005.

[26] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. on Inform. Theory*, vol. 42, pp. 1152-1159, 1996.

[27] H. Minkowski, "Zur Geometrie der Zahlen," *Math. Z.*, vol. 49, pp. 285-312, 1944.

[28] C. A. Rogers, *Packing and Covering*, Cambridge, UK: Cambridge Uni. Press, 1964.

[29] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr. , "MMSE decision-feedback equalizers and coding. I. Eequalization results" *IEEE Trans. on Commun.*, vol. 43, no. 10, pp. 25822594, Oct. 1995.

[30] M. O. Damen, H. El Gamal and G. Caire, "MMSE-GDFE Lattice Decoding for Solving Under-determined Linear Systems with Integer Unknowns," *Inter. Symp. on Inform. Theory*, June 2004.

[31] M. O. Damen, A. Chkeif, and J. -C. Belfior, "Lattice codes decoder for space-time codes," *IEEE Commun. Lett.* , vol. 4, no. 5, pp. 161163, May 2000.

[32] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 23892401, Oct. 2003.

[33] B. Hassibi and H. Vikalo, "On sphere decoding algorithm. I. expected complexity," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, pp. 2389-2401, Aug. 2005.

[34] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: A closer study in high SNR regime," *IEEE Trans. on Inform. Theory*,

[35] K. Kumar, G. Caire, and A. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Trans. on Inform. Theory*, vol. 55, no. 10, pp. 4398-4418, Oct. 2009.

[36] R. M. Fano, "A heuristic discussion of probabilistic decoding", *IEEE Trans. Inform. Theory*, vol. IT-9, pp 64-73, Apr. 1963.

[37] I. M. Jacobs, and E. R. Berlekamp, "A lower bound to the distribution of computation for sequential decoding", *IEEE Trans. Inform. Theory*, vol. IT-13, pp 167-174, 1976.

[38] F. Jelinek, "A fast sequential decoding algorithm using a stack", *IBM J. Res. Dev.* , 13:675-685,1969.

[39] A. Murugan, H. El Gamal, M. O. Damen and G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder", *IEEE Trans. Inform. Theory*, vol. 52, no. 3, March 2005.

[40] N. Sommer, M. Feder, and O. Shalvi, "Closest point search in lattices using sequential decoding", *IEEE Int. Symp. Information Theory*, p. 1053-1057, Adelaide, SA, Sept. 2005.

[41] H. El Gamal, G. Caire, M. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff", *IEEE Trans. Inform. Theory*, vol. 52, no. 8, August 2006.

[42] S.A. Pawar, K.R. Kumar, P.V. Kumar, P. Elia, B.A. Sethuraman, "Achieving the DMD tradeoff of the MIMO-ARQ channel", *Inter. Symp. on Inform. Theory*, pp. 901-905, Sept. 2005.

[43] A. Drukarev, and D. J. Costello, "Hybrid ARQ error control using sequential decoding", *IEEE Trans. Inform. Theory*, vol. 5, pp. 521-535, July 1983.

[44] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 22012214, Aug. 2002.

[45] J. M. Wozencraft, and B. Reiffen, *Sequential Decoding*, New York: MIT Press/Wiley, 1961.

[46] K. S. Zigangirov, "Some sequential decoding procedures," *Probl. Pered. Inform.* , pp. 19-35, Nov. 1966.

[47] K. Nguyen, L. Rasmussen, A. Guillen i Fabregas and N. Letzepis, "MIMO ARQ systems with multi-level feedback," (to appear) in *Proc. Int. Symp. Inform. Theory*, 2009.

[48] P. Elia, K.R. Kumar, S.A. Pawar, P.V. Kumar, H. -F. Lu,, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff", *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3869-3884, Sept. 2006.

[49] G. D. Forney, "Coset codes-I: Introduction and geometrical classification", *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1123-1151, 1988.

[50] G. D. Forney, "Coset codes-II: Binary lattices and related codes", *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1152-1187, 1988.

[51] N. Prasad, I. Berenguer, and X. Wang, "Design of Spherical Lattice SpaceTime Codes", *IEEE Trans. Inform. Theory*, vol. 54, no. 11, pp. 4847-4856, Nov. 2008.

[52] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm: Part I, the expected complexity", *IEEE Transactions on Signal Processing*, vol 53, no 8, pages 2806-2818, Aug 2005.

[53] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm: Part II, generalization, second-order statistics, and applications to communication", *IEEE Trans. on Signal Processing*, vol 53, no 8, pages 2819-2834, Aug 2005.

[54] J. Jaldén, B. Ottersten, "On the complexity of sphere decoding in digital communication", *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[55] D. Seethaler, J. Jaldén, C. Studer, and H. Bölcskei, "On the complexity distribution of sphere-decoding", *IEEE Trans. Inform. Theory*, 2011, to appear.

[56] A. H. Banihashemi and A. K. Khandani, "On the complexity of decoding lattices using the Korkin-Zolotarev reduced basis", *IEEE Trans. Info. Theory*, vol. 44, no. 1, pp. 162-171, Jan. 1998.

[57] J. Jaldén, P. Elia, "DMT Optimality of LR-Aided Linear Decoders for a General Class of Channels, Lattice Designs, and System Models", *IEEE Trans. on Inform. Theory*, vol. 56, no. 10, Oct. 2010.

[58] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, 2nd Edition, Englewood Cliffs, NJ: Prentice-Hall,2004.

[59] R. Comroe, D. Costello, "ARQ schemes for data transmission in mobile radio systems". *IEEE Jour. on Selec. Areas in Commun.*, pp. 472481, vol. 2, issue:4, July 1984.

[60] S. Kallel, and D. Haccoun, "Sequential decoding with ARQ and code combining: a robust hybrid FEC/ARQ system", *IEEE Trans. on Commun.*, vol. 36, no. 7, July 1988.

[61] P. Orten, "Sequential decoding of tailbiting convolutional codes for hybrid ARQ on wireless channels", *49th IEEE Vehic. Tech. Conference*, p. 279, July 1999.

[62] J. Hamorsky, U.Wachsmann, J. B. Huber, and A. Cizmar, "Hybrid automatic repeat request scheme with turbo codes", *Int. Symp. on Turbo Codes*, pp. 247-250, Brest, France, Sept. 1997.

[63] D. N. Rowitch and L. B. Milstein, "On the Performance of Hybrid FEC/ARQ Systems Using Rate Compatible Punctured Turbo (RCPT) Codes," *IEEE Trans. on Commun.*, pp. 948-959, vol. 48, no. 6, June 2000.

[64] K. Nguyen, L. Rasmussen, A. Guillen i Fabregas and N. Letzepis, "MIMO ARQ systems with multi-level feedback" *Proc. Int. Symp. Inform. Theory*, pp. 254 - 258, Seoul 2009.

[65] P. M. Gruber, and J. M. Wills, Eds., *Handbook of Convex Geometry*, vol. B, North Holland, Amsterdam: Elsevier, 1993.