ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012ENST005

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Communication et Electronique »

présentée et soutenue publiquement par

Arun Kumar SINGH

le 21 Février 2012

Le compromis Débit-Fiabilité-Complexité dans les systémes MIMO, multi-utilisateurs et coopératifs avec décodeurs ML et Lattice

Directeur de thèse : **Petros ELIA**

Jury

M. Jean-Claude BELFIORE, Professeur, Télécom ParisTech, France	Président
M. Meir FEDER, Professeur, Université de Tel-Aviv, Israël	Rapporteur
M. Erik G. LARSSON, Professeur, Université Linköping, Suède	Rapporteur
M. Aris L. MOUSTAKAS, Maître de Conférences, Université d'Athènes, Grèce	Examinateur
M. Dirk SLOCK, Professeur, EURECOM, France	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

THÈSE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012ENST005

EDITE - ED 130

Doctor of Philosophy ParisTech

D I S S E R T A T I O N

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy from

TELECOM ParisTech

Specialization « Electronics and Communications »

publicly presented and defended by

Arun Kumar SINGH

on 21 February 2012

Rate-Reliability-Complexity limits in ML and lattice based decoding for MIMO, multiuser and cooperative communications

Thesis Advisor : **Petros ELIA**

Jury

Jean-Claude BELFIORE, Professor, Telecom ParisTech, France	President
Meir FEDER, Professor, Tel-Aviv University, Israel	Reviewer
Erik G. LARSSON, Professor, Linköping University, Sweden	Reviewer
Aris L. MOUSTAKAS, Assistant Professor, University of Athens, Greece	Examiner
Dirk SLOCK, Professor, EURECOM, France	Examiner

TELECOM ParisTech
An Institut Telecom school - member of ParisTech

THÈSE

Acknowledgements

It is my pleasure to extend my sincere gratitude and appreciation to all the people who made this dissertation possible.

First and foremost, I would like to thank my advisor Dr. Petros Elia. It has been an honor to be his first Ph.D. student. I appreciate his unrelenting support, continuing interactions and guidance that made my Ph.D. experience productive and stimulating. His encouragement and contagious enthusiasm proved to be instrumental and driving force many a times.

Besides my advisor, I would also like to thank Dr. Joakim Jaldén for insightful technical discussions which provided me with a deeper insight into my research area. My special thanks goes to Kiran T. Gowda for his friendship and stimulating interactions during long hours at work. I would like also to extend my sincere thanks to Prof. David Gesbert for all his support and guidance.

A very special thanks goes to my thesis jury members, Prof. Meir Feder, Prof. Erik G. Larsson, Prof. Jean-Claude Belfiore, Prof. Aris L. Moustakas and Prof. Dirk Slock for their valuable comments and encouragement.

Best regards to all the past and present members of my office ES-201 for providing a very cheerful and lively atmosphere. I wish to express my warm and sincere thanks to Bassem Zayen, Siouar Bensaid, Kaoutar Elkhiyaoui, Paul De Kerret, Fatma Hrizi and Aymen Hafsaoui for helping me out with French translations. I would also like to thank Ashish Jain, Sudesh Jain, Jinhui Chen, Rizwan Gaffar, Erhan Yilmaz, Jinyuan Chen, Lorenzo Maggi, Xiao Lei, Miltiades Filippou, Xinping Yi, Rajendra Jain and Sankar Sivagnanam for being an extremely diverse crowd of people and at the same time being helpful and exciting friends.

Last, but far from least, I wish to express my deepest gratitude to my family for their unflagging love and support throughout my life.

Abstract

In telecommunications, rate-reliability and encoding-decoding computational complexity (floating point operations - flops), are widely considered to be limiting and interrelated bottlenecks. For this reason, any attempt to significantly reduce complexity may be at the expense of a substantial degradation in error-performance. Establishing this intertwined relationship constitutes an important research topic of substantial practical interest. This dissertation deals with the question of establishing fundamental rate, reliability and complexity limits in general outage-limited multiple-input multiple-output (MIMO) communications, and its related point-to-point, multiuser, cooperative, two-directional, and feedback-aided scenarios. Our interest in the outage-limited (non ergodic) setting (reduced or no channel state information at the transmitter, i.e., reduced or no CSIT), stems from the pivotal role of non-ergodic scenarios in modern wireless communications. These include scenarios of delay-limited data transmission in the absence of CSIT, or more importantly scenarios that consider quick and high rate communication of CSIT in the absence of channel reciprocity (over the feedback link), as well as transmission of CSIT in a multiuser setting, by each of the receivers back to their transmitters, in the presence or absence of channel reciprocity. The crucial role of such scenarios brings to the fore the need for maximal reliability, in the presence of the inevitable computational constraints. Finally we explore a large subset of the family of linear lattice encoding methods, and we consider the two main families of decoders ; maximum likelihood (ML) based and lattice-based decoding. Algorithmic analysis focuses on the efficient bounded-search implementations of these decoders, including a large family of sphere decoders.

Specifically, the presented work provides high signal-to-noise (SNR) analysis of the minimum computational reserves (flops or chip size) that allow for a) a certain performance with respect to the diversity-multiplexing gain tradeoff (DMT) and for b) a vanishing gap to the uninterrupted (optimal) ML decoder or a vanishing gap to the exact implementation of (regularized) lattice decoding. The vanishing gap condition guarantees that the decoder's error curve is arbitrarily close, given a sufficiently high SNR, to the optimal error curve of exact solutions, which is a much stronger condition than DMT optimality which only guarantees an error gap that is subpolynomial

in SNR, and can thus be unbounded and generally unacceptable for practical implementations. In some cases, the presented performance-complexity guarantees hold for a very broad setting of channel statistics, MIMO settings and lattice implementations.

Applying the methods of the derived complexity exponent and of the DMT, the work establishes the rate-reliability-complexity tradeoff of both ML-based sphere decoding as well as regularized (i.e., MMSE-preprocessed) lattice decoding. Specifically the work derives the complexity exponent for a given achievable DMT, as well as the achievable DMT for a given constraint on the computational resources. The derived complexity exponent describes the asymptotic rate of exponential increase of complexity, exponential in the number of codeword bits. This complexity exponent, albeit much reduced compared to that of an exhaustive ML decoder, still reveals a very considerable complexity. The same exponent reveals how this complexity can be modulated down to the complexity of linear receivers, at the expense of a specific amount of DMT loss. Interestingly we will see that sacrificing performance does not always pay back in terms of complexity reductions. Another conclusion is that, for a relatively broad setting of practical interest, the families of ML-based and regularized lattice sphere decoding behave identically (in terms of error and complexity exponents). Furthermore the work reinforces the role of halting policies that allow for complexity savings by being selective as to when to decode and when to suspend operations. Interestingly enough, such policies allow for a complexity that is not monotonically increasing in the rate.

The above described exponential complexity tends to render such transceiver algorithms inapplicable to several MIMO scenarios. This served as motivation to explore other decoders, and to provide the first ever lattice decoding solution and halting policy, that jointly achieve a vanishing gap to the exact implementation of regularized lattice decoding with a complexity that is subexponential in the number of codeword bits as well as in the rate. This work was able to, for the first time, rigorously demonstrate and quantify the pivotal role of lattice reduction as a special complexity reducing ingredient in MIMO systems. We here note that while lattice reduction has indeed allowed here for near-optimal behavior at very manageable complexity (with respect to lattice decoding), it is the case that there exist scenarios for which these same lattice reduction methods cannot be readily applied. Such problematic cases include the ubiquitous scenario where outer binary codes are employed. It is for this exact reason that analysis of non LR-aided schemes remains of strong interest.

The work then addresses the fundamental question of establishing the rate-reliability-complexity ramifications of feedback. This setting is very important because it can offer near ergodic behavior (high diversity), even at high multiplexing gains. We focus on two fundamental questions. The first question asks what is the complexity savings that feedback provides for a gi-

ven fixed rate-reliability performance, and the second question asks what is the complexity costs of achieving the full rate-reliability benefits of feedback. The analysis and the constructed feedback schemes tell us how to properly utilize a finite number of feedback bits to alleviate the adverse effects of computational constraints, as those seen in the derived rate-reliability-complexity tradeoffs of the previous chapters. Emphasis is placed on MIMO-ARQ feedback schemes, although we do also consider feedback with antenna selection.

Finally we present preliminary work on extending the rate-reliability-complexity analysis to simple instances of the multiple access, relay, and bidirectional channels, where again we identify the computational reserves that guarantee a DMT optimality, as well as address user/relay selection criteria and communication protocols that provide improved joint reliability-complexity performance in the presence of computational constraints.

The above methodology can be applied to quantify the rate-reliability-complexity performance of novel or existing coding schemes, decoders, as well as cooperative and multiuser protocols. Despite the serious efforts to resolve each problem in their most general form, the current work leaves out ample space for exponential reductions in complexity, and improvements in the joint performance-complexity measure, both on the side of decoders, as well as for encoders, protocols or feedback schemes.

Résumé

Dans les télécommunications, le débit-fiabilité et la complexité de l'encodage et du décodage (opération à virgule flottante-flops) sont largement reconnus comme représentant des facteurs limitant interdépendants. Pour cette raison, tout tentative de réduire la complexité peut venir au prix d'une dégradation substantielle du taux d'erreurs. Cette thèse traite de l'établissement d'un compromis limite fondamental entre la fiabilité et la complexité dans des systèmes de communications outage-limités à entrées et sorties multiples (MIMO), et ses scénarios point-à-point, utilisateurs multiple, bidirectionnels, et aidés de feedback. Nous explorons un large sous-ensemble de la famille des méthodes d'encodage linéaire Lattice, et nous considérons deux familles principales de décodeurs : les décodeurs à maximum de vraisemblance (ML) et les décodeurs Lattice. L'analyse algorithmique est concentrée sur l'implémentation de ces décodeurs ayant comme limitation une recherche bornée, ce qui inclue une large famille de sphère-décodeurs.

En particulier, le travail présenté fournit une analyse à haut rapport Signal-à-Bruit (SNR) de la complexité minimum (flops ou taille de puce électronique) qui permet d'atteindre a) une certaine performance vis-à-vis du compromis diversité-gain de multiplexage et b) une différence tendant vers zéro avec le non-interrompu (optimale) ML décodeur, ou une différence tendant vers zéro comparé à l'implémentation exacte du décodeur (régularisé) Lattice. L'exposant de complexité obtenu décrit la vitesse asymptotique d'accroissement de la complexité, qui est exponentielle en terme du nombre de bits encodés. L'exposant de complexité, quoique beaucoup plus réduit que le décodeur optimal exhaustif comporte toujours une complexité considérable.

La complexité exponentielle décrite ci-dessus tend à rendre tout algorithme d'émission-réception inapplicable au cas de plusieurs sources MIMO. Ceci fournit la motivation pour explorer d'autres décodeurs et pour apporter la première solution de décodage Lattice avec des consignes d'arrêt, qui atteint une différence tendant vers zéro comparée à l'implémentation exacte du décodeur Lattice régularisé avec une complexité qui est sous-exponentiel vis-à-vis du nombre de bits encodés ainsi que vis-à-vis du débit. Ce travail a permis de démontrer rigoureusement pour la première fois et de quantifier le rôle crucial de la réduction Lattice en tant que méthode réduisant la

complexité dans les systèmes MIMO.

Ce travail adresse ensuite la question fondamentale de l'impact du feedback sur la relation débit-fiabilité-complexité. Ce scenario est très important car il offre un comportement presque ergodique (forte diversité), même à haut degré de multiplexage. Nous nous concentrons sur deux questions fondamentales. La première question consiste à se demander quelle est la réduction de complexité que le feedback permet pour une performance donnée en terme de compromis débit-fiabilité, et la seconde question est de se demander quel est le cout en terme de complexité pour atteindre la performance maximale en terme de compromis débit-fiabilité avec du feedback. L'analyse et les méthodes de feedback développées nous montrent comment utiliser efficacement un nombre fini de bits de feedback pour réduire les effets négatifs dus aux contraintes de complexité, comme celles vues dans le compromis débit-fiabilité-complexité dans les chapitres précédents.

Enfin, nous présentons un travail préliminaire sur l'extension de l'analyse du compromis débit-fiabilité-complexité à des cas simples d'accès multiples, de relai, et de canaux bidirectionnels, où nous mettons encore en évidence quelle est la complexité minimum qui garantie un DMT optimal. Nous identifions aussi des critères de sélection des utilisateurs/relais et des protocoles de communication qui permettent une amélioration conjointe des performances fiabilité-complexité en présence de contrainte de complexité.

Table of Contents

List of Figures

Acronyms

Here are the main acronyms used in this dissertation. The meaning of an acronym is usually indicated once, when it first occurs in the text. The English acronyms are also used for the French summary.

ACK	Acknowledgement
ARQ	Automatic Repeat Request
BC	Broadcast Channel
CSI	Channel State Information
CSIR	Channel State Information at Receiver
CSIT	Channel State Information at Transmitter
CDA	Cyclic Division Algebra
DMD	Diversity-Multiplexing-Delay
DSP	Digital Signal Processing
DF	Decode-and-Forward
DMT	Diversity-Multiplexing gain Tradeoff
FLOPS	Floating Point Operations
HBC	Hybrid Broadcast
i.i.d.	Independent and Identically Distributed
ISI	Inter-Symbol Interference
LDPC	Low-Density Parity Check
LLL	Lenstra-Lenstra-Lovász
LR	Lattice Reduction
MAC	Multiple Access Channel
ML	Maximum Likelihood
MIMO	Multi-Input Multi-Output
MIMO-ARQ	MIMO Automatic Repeat Request
MISO	Multi-Input Single-Output
ML-SD	ML-based Sphere Decoder
MMSE	Minimum Mean Square Error
MMSE-DFE	MMSE Decision-Feedback Equalizer
NACK	Negative Acknowledgement
NP	Non-deterministic Polynomial-time
ns-TRC	non-separated Two-way Relay Channel
OAF	Orthogonal Amplify-and-Forward

OFDM	Orthogonal Frequency Division Multiplexing
SD	Sphere Decoder
SIC	Successive Interference Cancelation
SIMO	Single-Input Multi-Output
SNR	Signal-to-Noise Ratio
SISO	Single-Input Single-Output
ST	Space-Time
STBC	Space-Time Block Coding/Code
TDBC	Time Division Broadcast
TRC	Two-Way Relay Channel
ZF	Zero Forcing
V-BLAST	Vertical Bell Laboratories Layered Space-Time

Notations

Throughout the dissertation, we use the bold uppercase letters, e.g., $\mathbf{H}$, to refer to matrices, while bold lowercase letters, e.g., $\mathbf{h}$, to refer to column vectors. We use $\doteq$ to denote *exponential equality*, i.e., we write $f \doteq \rho^B$ to denote $\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = B$, $\dot{\leq}$ and $\dot{\geq}$ are defined similarly. Other notational conventions are summarized as follows :

$\mathbb{R}, \mathbb{C}$	Sets of real and complex numbers, respectively
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{Z}^N$	N dimensional Integer Lattice
$\lvert x \rvert$	The absolute value of a scalar
$\lfloor x \rfloor$	The nearest integer less than or equal to x
$\lceil x \rceil$	The nearest integer greater than or equal to x
$(x)^+$	Maximum of 0 and x, i.e. $\max\{0, x\}$
$\lVert \mathbf{x} \rVert$	The Euclidian (l^2) norm of a vector $\mathbf{x}$
$\lvert \mathcal{X} \rvert$	The cardinality of the set $\mathcal{X}$, i.e. the number of elements in the set $\mathcal{X}$
$\mathcal{CN}(\mu, \sigma^2)$	The circularly symmetric complex Gaussian distribution with mean μ and variance σ^2
$\mathbb{E}_X[.]$	The expectation operator over the RV X
$\mathbf{X}^*$	The conjugate operation on $\mathbf{X}$
$\mathbf{X}^H$	The complex conjugate (Hermitian) operation on $\mathbf{X}$
$\mathbf{X}^T$	The transpose operation on $\mathbf{X}$
$\mathrm{Tr}(\mathbf{X})$	The trace of $\mathbf{X}$
$\det(\mathbf{X})$	The determinant of $\mathbf{X}$
$\mathbf{X}^{-1}$	The inverse of $\mathbf{X}$
$\mathbf{X}^\dagger$	The Moore-Penrose pseudoinverse of $\mathbf{X}$
$\mathrm{vec}(\mathbf{X})$	The vector obtained by stacking the columns of $\mathbf{X}$
$\sigma_i(\mathbf{X})$	The i-th singular value of $\mathbf{X}$
$\mathbf{I}_m$	$m \times m$ identity matrix
$\mathbf{0}$	All zeros matrix of appropriate dimensions

Chapter 1

Introduction

1.1 Multiple Antenna Systems : Error Performance and Complexity

In wireless communications there is an ever increasing demand to support higher data rates, better quality of service (such as error-performance, coverage) and a large number of users, while maintaining the transmit power and bandwidth. This entails considering novel technologies like multiple-input multiple-output (MIMO) communication systems that can improve both, the data rate and the error-performance for the given transmit power and bandwidth, but at the cost of a potentially more challenging decoding task for the receiver. Specifically for the brute force maximum likelihood (ML) decoder [1] the decoding complexity grows exponentially in the number of codeword bits as 2^{RT} for some coding duration T. This number of bits generally increase with the dimensionality.

In MIMO scenarios with large signal dimensionality, ML computational costs can be prohibitively large. This motivates the use of suboptimal transceiver designs, whose sub-optimality can often entirely negate the reliability benefits of MIMO. Most of the present day MIMO receivers employ low computational complexity linear MIMO detection schemes (zero-forcing (ZF), minimum mean square error (MMSE), successive interference cancellation (SIC)) which have been shown to be highly suboptimal (cf. [1]). The recent computational speed advancements have paved the way for the use of state-of-art MIMO decoders employing non-linear MIMO detection schemes

1. The ML decoder is strictly speaking the only optimal decoder, in the sense that it minimizes the probability of codeword error (irrespective of the code), and it is arguably the slowest possible decoder, in the sense that it has to visit all possible codewords.

based on more advanced bounded search algorithms, generally referred to as sphere decoders (SD). Such decoders can provide complexity savings over the brute force ML implementation, at the expense of a small gap to the optimal error performance (cf. [2–7]). This variability of algorithms brings to the fore the natural question as to what is the relationship between complexity and reliability. Answering this question can be challenging, especially since the algorithmic complexity tends to fluctuate randomly with the channel, noise and code word realizations. This randomness also brings to the fore the potential of trading off computational complexity for error performance by selectively choosing when to decode and when not to. Generally though any attempt to significantly reduce complexity may be at the expense of a substantial degradation in error-performance. In the delay-limited outage-limited MIMO settings, establishing this intertwined relationship of the complexity and error performance constitutes an important research topic of substantial practical interest.

1.2 Complexity characterization in outage-limited MIMO

The complexity measure here is simply the amount of computational reserves (floating point operations - flops) that one is allowed to use during the decoding of any given codeword. Given the aforementioned fluctuating nature of the instantaneous algorithmic complexity (fluctuating with, for example, the fading realization), it becomes clear that the presence of computational constraints, can cause error performance degradation, as a result of what can be perceived as a decoding outage caused whenever the run-time constraints are violated (Fig. 1.1).

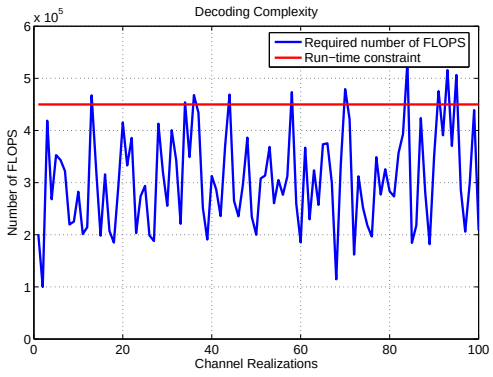


FIGURE 1.1 – Instantaneous algorithmic complexity fluctuations

1.2.1 Complexity Exponent

We denote the complexity constraint by $N_{\max}$, where as stated, $N_{\max}$ specifically describes the computational resources, in flops per T channel uses, that the transceiver is endowed with, in the sense that after $N_{\max}$ flops, the transceiver must simply terminate, potentially prematurely and before completion of its task. In establishing a meaningful representation of the above, we draw from Zheng and Tse's diversity multiplexing tradeoff (DMT), which quantifies the relationship between rate (R) and probability of error P_{err}, in the high SNR measure of multiplexing gain $r = R/\log \text{SNR}$ and the diversity gain $d(r) = -\lim_{\text{SNR}\to\infty} \log P_{\text{err}}/\log \text{SNR}$. In the same spirit of high-SNR tail analysis (SNR will be henceforth denoted as ρ), we here consider the *complexity exponent*, defined here as

$$c(r) := \lim_{\rho\to\infty} \frac{\log N_{\max}}{\log \rho}.$$

We observe that $c(r) > 0$ implies a complexity that is exponential in the rate.

1.2.2 On the suitability and applicability of the complexity measure and the complexity exponent

A reasonable question at this point would pertain as to why the computational resources $N_{\max}$ scale with ρ and are dependent on r, to which we note that the complexity of decoding is generally dependent on the density and cardinality of the codebook, which in turn depends on ρ and r. Furthermore this dependence of the complexity exponent (and by extension of $N_{\max}$) on r, reflects a potential ability to regulate the computational resources depending on the rate.

In further arguing that the above representation is a natural one (rather than a forced one), we also quickly note that $c(r)$ captures the entire complexity range

$$0 \le c(r) \le rT$$

of all reasonable transceivers, with $c(r) = 0$ corresponding to the fastest possible transceiver (requiring a subexponential number of flops per T channel uses), and with $c(r) = rT$ corresponding to the optimal but arguably slowest, full-search uninterrupted ML decoder [2].

2. We here note that strictly speaking, a encoder-decoder may potentially introduce a complexity exponent larger than rT. In such a case though, this encoder-decoder may be substituted by a lookup table implementation of encoder and an unrestricted ML decoder in the presence of a canonical code with multiplexing gain r, i.e., with the cardinality of codebook $|\mathcal{X}_r| = 2^{RT} = \rho^{rT}$. This encoder-decoder will jointly require resources that are a constant multiple of $|\mathcal{X}_r| \doteq \rho^{rT}$ as it has to construct and visit all possible $|\mathcal{X}_r|$ codewords, at a computational cost of a bounded number of flops per codeword visit. It is noted that the number of flops per visited codeword is naturally independent of ρ.

In terms of the suitability of the asymptotic setting, i.e., the suitability of the complexity exponent and the chosen scale of refinement, this is the scale that best captures the entirety of the complexity problem. Similar to the case of the probability of error, which varies from 1 to the values that scale as $K_1 \cdot \rho^{-z}$, bringing to the fore an error exponent $d(r) \in [0, z]$ (here K_1 is a subpolynomial function of ρ), in the case of complexity, we have a variation from 1 flop to a maximum complexity that scales as $K_2 2^{RT} = K_2 \cdot \rho^{rT}$ flops (again K_2 is essentially a constant), which naturally brings to the fore a complexity exponent $c(r) \in [0, rT]$.

One can then argue that having the complexity cost share the same scale as the probability of error, is not a forced choice but rather a natural consequence of the fact that both the probability of error as well as the code size are naturally polynomial functions of ρ, and (in the case of complexity) exponential functions of $R = r \log \rho$. As a result the above error and complexity exponents could be meaningfully combined to form a measure of the overall rate-reliability-complexity capabilities of different transceivers. One such measure could for example take the form of a joint reliability-complexity measure $\Gamma(r) \triangleq d(r) - \gamma c(r)$, for some weighing factor $\gamma \geq 0$. The measure could be applied to describe, for example, the high-SNR error capabilities of a particular encoder-decoder per unit power and per chip area. We note that in nicely traversing the above reliability and complexity exponent tradeoffs, what is important is to find proper encoding-decoding policies can regulate complexity at a limited error performance loss. Such regulation could span the two extremes of brute force ML decoding to the highly inefficient linear receivers. A special ingredient in traversing this tradeoff will be the halting or *regulating policies* that we will employ.

1.2.3 Vanishing Gap

Finally, towards refining our above mentioned exponent analysis (DMT and complexity exponent), we first note that DMT analysis fails to capture potentially infinite (subpolynomial to SNR) gaps to the optimal error performance. Towards this we are also interested in decoders that achieve a *vanishing gap* to the brute force ML performance. This vanishing gap approach is a much stronger condition than DMT optimality which remains insensitive to error gaps that could be unbounded.

In terms of error-performance gaps, one could consider the gap of a given decoder $\mathcal{D}_r$ to ML, i.e, the gap between the error performance P_e of $\mathcal{D}_r$ to the optimal error performance $\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)$ of the brute force ML decoder, where $\mathbf{s}$ denotes the transmitted symbol vector and where $\hat{\mathbf{s}}_{ML}$ denotes brute force ML estimate of the transmitted symbol vector $\mathbf{s}$. Given a certain computational constraint $N_{\max} \doteq \rho^c$ for $\mathcal{D}_r$, this gap is quantified in the high

SNR regime to be

$$g(c) \triangleq \lim_{\rho\to\infty} \frac{P_e}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)}. \tag{1.1}$$

A *vanishing gap* $g(c) = 1$ means that with $N_{\max} \doteq \rho^c$ flops, $\mathcal{D}_r$ can asymptotically have near identical error performance as the optimal ML decoder.

Similarly, when considering any other baseline decoder, such as the regularized (MMSE-preprocessed) lattice sphere decoder, we would be interested in the performance gap to the *exact implementation* of this lattice decoder. As before, in the presence of $N_{\max} \doteq \rho^{c_L}$ flops for the lattice sphere decoder, we would have a vanishing gap when

$$g(c_L) \triangleq \lim_{\rho\to\infty} \frac{P_L}{\mathrm{P}\left(\hat{\mathbf{s}}_L \neq \mathbf{s}\right)} = 1 \tag{1.2}$$

where P_L describes the error probability of the preprocessed lattice sphere decoder, and where $\mathrm{P}\left(\hat{\mathbf{s}}_L \neq \mathbf{s}\right)$ would describe the error probability of the exact solution of MMSE-preprocessed lattice decoder.

One of our interests is to capture the minimum complexity that allows for a vanishing gap to either full ML or full (regularized) lattice decoding. We will study this and specifically expand the work of [7] to a very broad setting.

Remark 1 (Comments on vanishing gap approach : a heuristic interpretation)**.** *The comments in this remark are not rigorous but do not affect at all any of our analysis, but rather present heuristic insights on the utility of the vanishing gap approach.*

Consider, for a fixed code and rate, the error curves (x axis is SNR, y axis is the probability of error) corresponding to two different decoders. For example let the first decoder be the optimal (uninterrupted) ML decoder. Let the second decoder be an approximate implementation of the first decoder. Let $P(\rho)$ *and* $P_{ap}(\rho)$ *denote the* actual *probability of error curves, for increasing* ρ*, respectively for the optimal and suboptimal decoders (*$P(\rho) \leq P_{ap}(\rho)$ *for all* ρ*). Furthermore consider the following two assumptions.*

Assumption 1 *: There exists an SNR value* ρ_1 *(resp.* ρ_2*) after which the error-curve slope* $\frac{d}{d\rho}P$ *(resp.* $\frac{d}{d\rho}P_{ap}$*) is monotonically non-decreasing.*

Assumption 2 *: There exists an SNR value* ρ_0 *at which the error-curve slope* $\frac{d}{d\rho}P$ *of the first (ML) decoder attains its maximum value in the monotonic region, i.e., attains a value* $\frac{d}{d\rho}P|_{\rho_0} = \max_{\rho\geq\rho_1} \frac{d}{d\rho}P$*.*

Given the above two assumptions, one can see that the guarantee on the vanishing gap implies that the two error curves coincide for any $\rho \geq \rho_0$*, i.e.,*

$$\frac{P(\rho)}{P_{ap}(\rho)} = 1, \ \forall \rho \geq \rho_0.$$

We here caution the reader that it is conceivably easy to violate the above two conditions, for example by employing a decoding policy that alternates, for different values of SNR, between two different decoders, or by employing codes with specific distance distributions that may for example result in error floors. Despite this though, when the transceiver structure remains unchanged, the two assumptions accept substantial validation by most empirical observations, but have not been proven to be true.

Furthermore consider a third assumption, which asks that for any SNR value in the monotonic region $\rho \geq \max\{\rho_1, \rho_2\}$*, the error-curve slope of the exact (first) decoder, is not smaller than that of the approximate solution, i.e.,*

$$\frac{d}{d\rho}P \geq \frac{d}{d\rho}P_{ap}, \quad \forall \rho \geq \max\{\rho_1, \rho_2\}.$$

Then the three assumptions, together with the guarantee on the vanishing gap, jointly imply that the error curves of the exact and approximate decoder, match at (and after) the potentially much lower SNR value of $\max\{\rho_1, \rho_2\}$*, i.e., jointly imply that*

$$\frac{P(\rho)}{P_{ap}(\rho)} = 1, \ \forall \rho \geq \max\{\rho_1, \rho_2\}.$$

Such values can indeed be moderate as they do not immediately relate to the faithfulness of asymptotic approximations, but are rather the starting points of monotonic regions. This exposition suggests that under some assumptions, a vanishing performance gap guarantee could carry over to finite SNR regions.

1.3 Dissertation Outline and main Contributions

We proceed with a summary of what is to come, focusing on the results that offer more insight rather than focusing on the results with the broadest scope, which will be presented in the chapters themselves. We hope that our work can give concise insight on pertinent questions such as :

- What is the complexity price to pay for near-optimal implementation of MIMO, multiuser and cooperative communications ?
- How does feedback reduce complexity ?
- What policies can regulate complexity at a limited performance loss ?
- How do complexity-constraints affect reliability in different MIMO settings ?
- How big of a MIMO system (how many transmit antennas or tones or relays or mac users) can your DSP chip sustain ?
- How should multiple users behave in the presence of complexity constraints ?
- What is the role of antenna selection in reducing complexity ?

- What are the cooperative protocols that perform best in the presence of computational constraints ?

Channel Model, Encoders and Decoders - Chapter 2

The next chapter presents the channel model, the lattice designs and basic decoders of interest, as well as provides motivation for the delay-limited outage-limited setting and its relevance to present day communication scenarios.

Complexity for ML decoding - Chapter 3

The purpose of this chapter is two-fold. The first is to extend the scope of the complexity analysis in previous work (cf. [7]), to a very broad setting of lattice designs, decoding policies and fading statistics. This extension is important in allowing us to explore interesting MIMO settings. The second is to provide meaningful rate-reliability-complexity measures and tradeoffs. The corresponding mathematical exposition is for the general MIMO case, whereas the single-letter expressions are derived for the quasi-static MIMO channel. All the results hold for ML-based decoding, and the algorithmic analysis considers the efficient family of ML-based sphere decoding algorithms. The derived complexity exponents of these algorithms describe the sufficient, and in many cases necessary, computational resources required for ML based SD to achieve either a specific DMT performance, or to achieve a vanishing gap to (uninterrupted) ML.

Albeit possibly premature at this point, we note that the results suggest complexity ramifications given a choice of lattice designs, fading statistics and sphere decoding ordering policies. The concepts of lattice designs (and specifically full-rate lattice designs), as well as of ordering policies, are commonly used but they will certainly be discussed in detail later on. We also note that a common theme in this research is that we will often derive universal complexity upper bounds that hold irrespective of the above choices, and then there are tightening results for relatively broad settings.

General MIMO setting Theorem 1 considers the general $m \times n$ ($n \geq m$) MIMO [3] channel setting

$$\mathbf{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w},$$

and considers ML based decoding. In what follows, $I(\boldsymbol{\mu})$ will denote the rate function of $\boldsymbol{\mu} \triangleq (\mu_1, \cdots, \mu_m)$, $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}^H\mathbf{H})}{\log \rho}$, $j = 1, \cdots, m$, corresponding to the ordered singular values σ_j of the channel. We also recall that $rT \log \rho$ is the total number of bits.

3. Slightly less conclusive results hold for the $n \geq m$ case.

Theorem 1 : The complexity exponent of achieving a diversity gain $d(r)$ is upper bounded as

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{m} \min\left(\frac{rT}{m} - \frac{1}{2}(1-\mu_i), \frac{rT}{m}\right)^{+} \tag{1.3a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d(r), \tag{1.3b}$$

$$\mu_1 \geq \cdots \geq \mu_m \geq 0, \tag{1.3c}$$

for all families of fading statistics[4]*, all full rate (or below full rate) lattice designs, and all static or dynamic decoding ordering policies.*

Then Proposition 1 establishes conditional tightness of the above.

Proposition 1 : Irrespective of channel fading statistics and of the full-rate lattice code applied, there exists a fixed decoding order for which the above universal upper bound is tight.

With respect to establishing sufficient resources for achieving a vanishing gap to ML, Corollary 1a simply applies the above theorem after setting $d(r) = d_{\text{ML}}$ to be the optimal DMT of uninterrupted ML decoding of the specific (potentially suboptimal) code.

Corollary 1a : The complexity exponent required by SD for achieving a vanishing gap to optimal ML is upper bounded as

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{m} \min\left(\frac{rT}{m} - \frac{1}{2}(1-\mu_i), \frac{rT}{m}\right)^{+} \tag{1.4a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d_{\text{ML}}(r), \tag{1.4b}$$

$$\mu_1 \geq \cdots \geq \mu_m \geq 0, \tag{1.4c}$$

irrespective of statistics and ordering policies.

Quasi-static MIMO Transitioning to the specific case of the $n_T \times n_R$ ($n_R \geq n_T$) quasi-static point-to-point MIMO channel (with T uses over channel $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$), where now $\boldsymbol{\mu}$ denotes the (asymptotics of) the singular values of $\mathbf{H}_C$, we have Theorem 3 establishing the following universal upper bound.

Theorem 3 The SD complexity exponent of achieving a diversity gain $d(r)$ is

4. Strictly speaking, this holds for the broad family of statistics that accept the large deviation principle [8].

upper bounded as

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} T\sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^+ \tag{1.5a}$$

$$s.t.\quad I(\boldsymbol{\mu}) \leq d(r), \tag{1.5b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \tag{1.5c}$$

for any full rate lattice code.

Furthermore it is worth noting that, directly from Proposition 1, we know that irrespective of the fading statistics and of the full-rate code, there exists a fixed decoding order for which the above universal upper bound is tight.

From Theorem 3 we can now establish a universal upper bound on the complexity to achieve the DMT optimal performance $d^*(r)$ of the $n_T \times n_R$ ($n_T \leq n_R$) MIMO channel.

Theorem 4 The SD complexity exponent of achieving the optimal DMT $d^*(r)$ is upper bounded as

$$c(r) \leq \bar{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right), \tag{1.6}$$

where $\bar{c}(r)$ is a piecewise linear function that, for integer values of r, takes the form

$$\bar{c}(r) = \frac{T}{n_T} r(n_T - r), \quad r = 0, 1, \cdots, n_T.$$

This holds for any set of fading statistics, all DMT optimal full rate code designs, and for any decoding order policy.

Due to the fact that complexity is increasing in $d(r)$, the above can be used as a universal upper bound for all full rate codes. This is described below.

Corollary 4a : The SD complexity exponent is upper bounded as in Theorem 4 for any full rate code, all fading statistics and all decoding order policies.

We note that, as it turns out, this is the tightest upper bound that can hold for all (full-rate) codes and fading statistics (cf. Proposition 1). This upper bound is already useful in itself as it establishes that the SD complexity exponent is much lower than the worst-case SNR exponent rT associated with the brute force ML decoding, irrespective of the codes, channels or ordering policies. For example, a comparison of brute force complexity exponent with the above derived bound is depicted in Fig. 1.2 for $n_T = T = 2$.

Tightness of universal upper bound Remaining in the $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel, and focusing on the case of i.i.d. Rayleigh

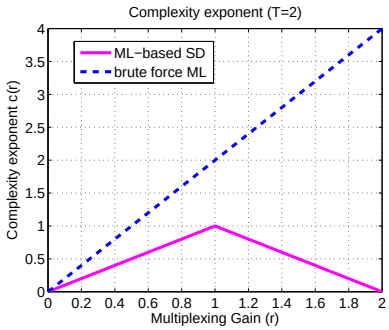


FIGURE 1.2 – Comparison of brute force ML complexity exponent with the universal upper bound for ML-based SD

fading statistics, Theorem 6 establishes that the above universal upper bound is in fact tight for almost all DMT optimal full-rate lattice designs, irrespective of the ordering policies. Specifically it says that if the design is chosen at random (each element of the lattice generator matrix is chosen in an i.i.d. manner from a continuous distribution) but held fixed, then with probability one in the choice of the code, the above bound in Theorem 4 is tight.

Theorem 6 : For the i.i.d. Rayleigh quasi-static MIMO channel, and irrespective of the fixed or dynamically changing decoding ordering policy, the complexity exponent of the ML-based sphere decoder almost surely*, in the choice of the DMT optimal lattice code, matches the universal upper bound in Theorem 4.*

Rate-reliability-complexity tradeoff : achievable DMT in the presence of computational constraints The above derivations suggest that under some conditions, the complexity exponent can decrease with a reduction in the desired diversity gain. Careful study will show that such a tradeoff is not always successful, that reductions in $d(r)$ are not necessarily rewarded with a reduction in the complexity, and that there are ranges of $d(r)$ for which $c(r)$ remains fixed. We seek to understand the achievable diversity gain in the presence of computational constraints. Theorem 7 gives a general expression for this complexity-constrained DMT for the general MIMO setting, and for any set of fading statistics. Focusing on the i.i.d quasi-static Rayleigh fading $n_T \times n_R$ ($n_R \geq n_T$) MIMO channel, and on DMT optimal lattice designs (achieving $d^*(r)$), Corollary 8a gives a lower bound on the complexity-constrained DMT. The bound is tight for the case of DMT optimal layered codes [5].

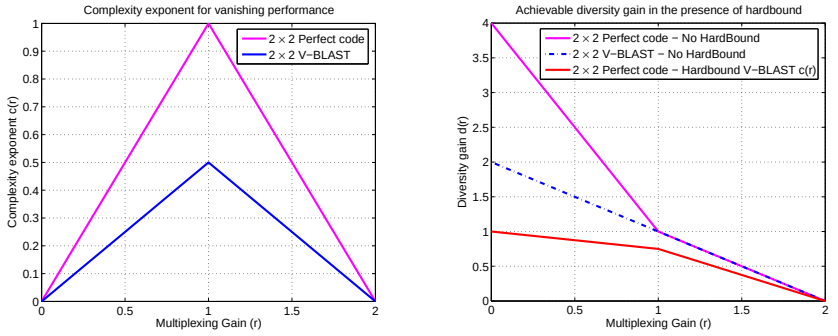
5. We will refer to layered codes in more detail later on. We note that all currently known DMT optimal lattice designs belong in the family of layered codes

Corollary 8a : The complexity-constrained DMT of sphere decoding any DMT optimal lattice design using maximally $N_{\max} \doteq \rho^{c_{\mathcal{D}}(r)}$ *flops, is lower bounded irrespective of the decoding ordering policy as* $d_{\mathcal{D}}(r) = \min\{d^*(r), d_{\mathcal{D}}(r, c_{\mathcal{D}}(r))\}$ *where*

$$d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)) = \sum_{j=1}^{K} (n_R - n_T + 2j - 1) + (n_R - n_T + 2K + 1)\left(\frac{c_{\mathcal{D}}(r)}{T} + 1 - \frac{(K+1)r}{n_T}\right),$$

where $K = \left\lfloor \frac{n_T c_{\mathcal{D}}(r)}{rT} \right\rfloor$. *Furthermore in the presence of DMT optimal layered lattice designs, the above described DMT is the exact complexity constrained DMT given by the natural sphere decoding ordering.*

Example 1. *The complexity-constrained DMT of sphere decoding the* 2×2 *Golden code ([9]) over the i.i.d. Rayleigh channel, is illustrated in Fig. 1.3. The upper line in Fig. 1.3 (a) describes the complexity exponent that would have been needed by the* 2×2 *perfect code to achieve the optimal DMT (upper line in Fig. 1.3(b)). The lower line in Fig. 1.3 (a) describes the complexity exponent that would have been needed by the weaker coding design of V-BLAST to achieve its own optimal DMT (middle line in Fig. 1.3(b)). This same lower line in Fig. 1.3 (a) also describes the complexity limitations that we are assigning to the* 2×2 *perfect code, which now, due to these complexity limitations, is achieving a much reduced DMT (lowest line in Fig. 1.3(b)).*



(a) Variable Complexity Hardbound (b) Achievable Diversity Gain

FIGURE 1.3 – Rate-reliability-complexity tradeoff for 2×2 perfect codes.

Complexity of regularized lattice decoding - Chapter 4

Our interest in lattice decoding stems from the derived proofs of high complexity required in ML based decoding, as well as indications and proofs

that lattice decoding can, in some cases, achieve near-ML performance. This chapter achieves two main results. The first result proves the rate-reliability-complexity equivalence between ML-based SD and regularized lattice SD [6] over a broad and common setting of interest, and the second is to provide the first ever solution (decoder and computational halting policy) that achieves a vanishing gap to the exact implementation of regularized lattice decoding at a complexity that is subexponential in the rate.

Equivalence of ML-based SD and regularized lattice SD The work will quantify the complexity of regularized lattice decoding and its rate-reliability-complexity equivalennce to ML based SD.

Theorem 10 : For the quasi-static $n_T \times n_R$ ($n_R \geq n_T$) i.i.d. Rayleigh fading MIMO channel, the complexity exponent for regularized lattice (sphere) decoding any full-rate threaded lattice design at any regulated DMT, is equal to the complexity exponent of ML-based SD (Theorem 3), for any fixed layer-preserving ordering including the natural decoding ordering.

We clarify that even though both (ML and lattice based) decoders are DMT optimal, the above result incorporates more than just DMT optimal decoding, in the sense that it shows that any timeout policy will tradeoff $d(r)$ with $c(r)$ identically for ML-based and lattice-based sphere decoding. In other words the decoders share the same $d(r)$ and $c(r)$ capabilities, irrespective of the timeout policy.

Furthermore, remaining in the setting of the $n_T \times n_R$ ($n_R \geq n_T$) i.i.d. Rayleigh fading quasi-static MIMO channel, Corollary 10b establishes that the above equivalence in fact holds for almost all DMT optimal full-rate lattice designs, and does so for a very general setting.

Corollary 10b : Irrespective of the fixed or dynamically changing decoding order, the complexity exponent for MMSE preprocessed lattice sphere decoding any (fixed but) randomly and uniformly chosen code (from an ensemble of DMT optimal full-rate linear lattice designs) almost surely, in the choice of DMT optimal lattice code, matches the complexity exponent of ML-based SD, and does so irrespective of the regulated DMT.

Furthermore Corollary 10d reveals the surprising fact that there exists no statistical channel behavior that will allow the removal of the bounding region (lattice decoding) to cause unbounded increases in the complexity of the decoder, i.e., reveals that this complexity bound holds even if the channel statistics are such that the channel realizations cause the decoder to always have to solve the hardest possible lattice search problem. We give this result in its refined form which holds for the quasi-static MIMO channel.

6. A 'regularized' lattice decoder is simply a generalization of the known MMSE-preprocessed lattice decoder.

Corollary 10d The complexity exponent of regularized lattice SD is upper bounded by

$$\overline{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right) \tag{1.7}$$

which for integer r simplifies to

$$\overline{c}(r) = \frac{T}{n_T} r(n_T - r), \tag{1.8}$$

for all fading statistics, all decoding ordering policies, all target DMT values and all full-rate lattice designs.

Vanishing gap to exact regularized lattice decoding at a subexponential complexity With provable evidence of the very high complexity of regularized lattice decoding, we turn to the powerful tool of lattice reduction (LR) and seek to understand its effects on computational complexity. The main result is the following, and it holds for a very broad setting of MIMO scenarios, lattice designs and fading statistics.

Theorem 12 : LR-aided MMSE-preprocessed lattice sphere decoding with a computational constraint activated at $N_{\max}\rho^x$ flops, for any $x > 0$, allows for a vanishing gap to the exact solution of regularized lattice decoding.

The above implies subexponential complexity in the sense that the complexity scales slower than any conceivable exponential function. This work constitutes the first proof that subexponential complexity need not come at the cost of exponential reductions in lattice decoding error performance. This work was also able to, for the first time, rigorously demonstrate and quantify the pivotal role of lattice reduction as a special complexity reducing ingredient in MIMO systems.

We here note that while lattice reduction has indeed allowed here for near-optimal behavior at very manageable complexity (with respect to lattice decoding), it is the case that there exist scenarios for which these same lattice reduction methods cannot be readily applied. Such problematic cases include the ubiquitous scenario where inner binary codes are employed. It is for this exact reason that analysis of non LR-aided schemes remains of strong interest.

Feedback-Aided Complexity of ML and Lattice Decoding - Chapter 5

The work then addresses the fundamental question of establishing the rate-reliability-complexity ramifications of feedback. This setting is very important because it can offer near ergodic behavior (high diversity), even at high multiplexing gains. We focus on two fundamental questions. The first

question asks what is the complexity savings that feedback provides for the optimal DMT ($d^*(r)$) [7], and the second question asks what is the complexity costs of achieving the full rate-reliability benefits of feedback. The analysis and the constructed feedback schemes tell us how to properly utilize a finite number of feedback bits to alleviate the adverse effects of computational constraints, as those seen in the derived rate-reliability-complexity tradeoffs of the previous chapters. Emphasis is placed on MIMO-ARQ feedback schemes, although we do also consider feedback with antenna selection.

All the presented results hold for ML-based decoding as well as MMSE-preprocessed lattice decoding.

Feedback-aided complexity of achieving the optimal DMT For the setting of the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$) i.i.d. regular fading MIMO-ARQ channel with L rounds of ARQ and with T channel uses per round (cf. [10]), the following focuses on the case of $LT = n_{\mathrm{T}}$ and upper bounds the minimum feedback-aided complexity that guarantees, with the assistance of ARQ, the DMT optimal $d^*(r)$. Before proceeding with the result it is worth noting that, as it turns out, the halting policies that decide when to decode and not to decode during the intermediate rounds, play a crucial role in reducing the derived complexity. We proceed with the result.

Theorem 13 : Let $c(r)$ be the minimum complexity exponent required to achieve $d^*(r)$, minimized over all lattice designs, all ARQ schemes with $L \leq n_{\mathrm{T}}$ rounds of ARQ and total delay $LT = n_{\mathrm{T}}$, and all halting and decoding order policies. Then

$$c(r) \leq \overline{c}_{red}(r) = \frac{1}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right],$$

which is a piecewise linear function that, for integer multiplexing gain values, takes the form

$$\overline{c}_{red}(r) = \frac{1}{n_T} r(n_T - r), \ \textit{for } r = 0, 1, \cdots, n_T.$$

To note the complexity reduction, we recall from Theorem 5 and Corollary 5a that in the absence of feedback, the complexity exponent associated to the same $d^*(r)$ took the form

$$c(r) = r(n_T - r), \tag{1.9}$$

(for integer $r = 0, 1, \cdots, n_T$), whereas we have just seen that, for example, in the presence of $L = n_{\mathrm{T}}$ rounds of ARQ, the same DMT is achieved with

7. We stress here that $d^*(r)$ denotes the optimal DMT of the MIMO channel without feedback.

a much reduced feedback-aided complexity of at most

$$c(r) \leq \frac{1}{n_T} r(n_T - r).$$

After that, Propositions 3,4 will present a very simple MIMO ARQ scheme that achieves $d^*(r)$ with $c(r) \leq \bar{c}_{red}(r)$.

Example 2. *Figure 1.4 considers the case of* $n_{\mathrm{T}} = 3 \leq n_{\mathrm{R}}$ *and Rayleigh fading, and compares the above complexity upper bound in the presence of feedback (L-rounds, minimum delay), to the equivalent complexity exponent in* (1.9) *of achieving the same DMT optimal* $d^*(r)$ *without ARQ feedback (Perfect codes and natural, fixed decoding ordering).*

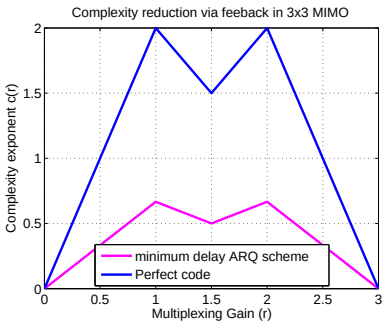


FIGURE 1.4 – Complexity reductions using ARQ feedback.

In regard to the feedback reduction for asymmetric ($n_{\mathrm{R}} < n_{\mathrm{T}}$) channels, Theorem 14 and Proposition 5 will describe the corresponding upper complexity bound and very simple ARQ schemes that achieve this bound for the case of $n_{\mathrm{R}} \leq n_{\mathrm{T}}$, and specifically the case where $n_{\mathrm{R}} | n_{\mathrm{T}}$ (i.e., n_{T} is an integer multiple of n_{R}).

The complexity cost of achieving the optimal feedback-aided DMT $d^*(r/L)$ For establishing the complexity of harvesting the full rate-reliability benefits of ARQ feedback, we again focus on the $n_T \times n_R$ ($n_R \geq n_T$) MIMO-ARQ channel with i.i.d. regular fading statistics. As before all presented results hold for ML-based decoding as well as MMSE-preprocessed lattice decoding. The following holds for the case where $n_{\mathrm{R}} \geq n_{\mathrm{T}}$ and the case where $L | n_{\mathrm{T}}$ (i.e., n_{T} is an integer multiple of L). As has been shown in [10], the optimal feedback-aided DMT performance is given by $d^*(r/L)$.

Theorem 15 : Let $c(r)$ *be the minimum complexity exponent required to achieve the optimal L-round MIMO-ARQ DMT* $d^*(r/L)$*, for any given* $L | n_{\mathrm{T}}$ *(i.e.,* n_{T} *is an integer multiple of L), where the complexity is minimized over*

all lattice designs, all halting policies and all decoding order policies. Then

$$c(r) \leq \overline{c}_{dmd}(r) \triangleq \frac{1}{L}\left[r\left(n_T - \left\lfloor \frac{r}{L} \right\rfloor - 1\right) + \left(Ln_T \left\lfloor \frac{r}{L} \right\rfloor - r(n_T - 1)\right)^+\right],$$

where $\overline{c}_{dmd}(r)$ is a piecewise linear function that, for r being an integer multiple of L, takes the form

$$\overline{c}_{dmd}(r) = \frac{rn_{\mathrm{T}}}{L^2}\left(L - \frac{r}{n_{\mathrm{T}}}\right).$$

Immediately after that, Proposition 6 will present very simple MIMO ARQ scheme that achieves $d^*(r/L)$ with $c(r) \leq \overline{c}_{dmd}(r)$.

Other bounds, ARQ schemes, decoding policies and lattice designs are presented later on in the same chapter. The same chapter also presents several examples for furthering intuition.

Complexity reductions due to antenna selection We here explore another method that feedback might reduce complexity. We will specifically focus on antenna selection schemes. Such schemes employ $\log_2 \binom{n_{\mathrm{T}}}{l_T}$ bits of CSIT to reduce an $n_T \times n_R$ MIMO system to a smaller and more manageable $l_T \times l_R$ system with generally reduced computational requirements. This part of the thesis will analyze the complexity ramifications of antenna selection, focusing on the case where the performance, after antenna selection, remains DMT optimal ($d(r) = d^*_{n_T \times n_R}(r)$). Our work here is preliminary, and it builds only on the greedy selection algorithms in [11]. In this setting, for

$$N_r = \arg \min_{N' \in \{1,\cdots,n_T\}} \left[\left(\arg \min_{p \in \{0,\cdots,N'-1\}} \frac{(n_{\mathrm{T}} - p)(n_{\mathrm{R}} - p)}{N' - p}\right) = \lceil r \rceil\right], \quad (1.10)$$

we have the following that holds for the i.i.d. Rayleigh fading $n_T \times n_R$ ($n_R \geq n_T$) MIMO setting. As before we consider ML based decoding as well as MMSE-preprocessed lattice decoding.

Proposition 8 The minimum, over all antenna selection algorithms, all lattice designs and all halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal DMT $d^*_{n_T \times n_R}(r)$, is upper bounded as

$$c(r) \leq \overline{c}_{as}(r) = \left(r(N_r - \lfloor r \rfloor - 1) + (N_r \lfloor r \rfloor - r(N_r - 1))^+\right),$$

which, for the N_r as above, is a piece-wise linear function that, for integer values of multiplexing gain r, takes the form

$$\overline{c}_{as}(r) = r(N_r - r), \text{ for } r = 0, 1, \cdots, n_{\mathrm{T}}. \quad (1.11)$$

The proper DMT optimal antenna selection scheme that achieves the above bound is also presented in the same chapter.

Complexity analysis for multiuser, cooperative, and bidirectional settings - Chapter 6

The work in this chapter is preliminary and it extends the rate-reliability-complexity analysis to simple instances of the multiple access, relay, and bidirectional channels, where again we identify the computational reserves that guarantee a DMT optimality, as well as address user/relay selection criteria and communication protocols that provide improved joint reliability-complexity performance in the presence of computational constraints.

Multiple access channels This work establishes bounds on the complexity requirements for achieving the optimal MAC-DMT over an i.i.d. Rayleigh fading symmetric multiple access channel (MAC) with K-users, each having the same multiplexing gain r, each having a single transmit antenna, and where the destination has $n_R \leq K$ receive antennas. In some cases the bounds are shown to be tight. The upper bound on complexity is described below, and as before it holds for ML-based decoding and for regularized lattice decoding. In what follows, $K_0 = K$ if K is odd, and $K_0 = K + 1$ if K is even.

Theorem 16 : The minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal DMT of MAC, is upper bounded as

$$c(r) \leq \overline{c}_{mac}(r) = \begin{cases} \overline{c}_v(r) & \text{for } r \leq \frac{n_R}{K+1}, \\ \overline{c}_f(r) & \text{for } \frac{n_R}{K+1} < r \leq \frac{n_R}{K}, \end{cases} \tag{1.12}$$

where

$$\begin{aligned} \overline{c}_v(r) = \max_{\boldsymbol{\mu}} \; & (K-1)r + \sum_{i=1}^{n_R} \left(r - (1-\mu_i)^+\right)^+ \\ s.t. \; & \sum_{i=1}^{n_R} (K - n_R + 2i - 1)\mu_i \leq n_R(1-r), \\ & \mu_1 \geq \cdots \geq \mu_{n_R}, \end{aligned}$$

where

$$\overline{c}_f(r) = (K-1)\, rK_0 + K_0 \left(r(n_R - \lfloor Kr \rfloor - 1) + (\lfloor Kr \rfloor - r(K-1))^+\right),$$

which is a piece-wise linear function that, for $r = 0, \frac{1}{K}, \cdots, \frac{n_R}{K}$, takes the form

$$\overline{c}_f(r) = (K-1)\, rK_0 + K_0 r(n_R - Kr), \text{ for } \; r = 0, \frac{1}{K}, \cdots, \frac{n_R}{K}.$$

Nice intuition can be deducted from the result in Proposition 9 which will describe encoding-decoding policies that achieve the optimal MAC-DMT with $c(r) \leq \overline{c}_{mac}(r)$. For the specific case of $n_R = 1$, the bound in (1.12) takes the simple form

$$\overline{c}_{mac}(r) = \begin{cases} (K-1)r & \text{for } r \leq \frac{1}{K+1}, \\ (K-1)K_0 r & \text{for } \frac{1}{K+1} < r \leq \frac{1}{K}. \end{cases}$$

Cooperative relay channels Again preliminary, this work establishes very early results on the complexity of achieving the optimal DMT of a cooperative network with a source, $n-1$ relays and a destination, each having single transmit/receive antenna and communicating over i.i.d. Rayleigh fading. This is done only for the orthogonal amplify-and-forward (OAF) protocol. The upper bound on complexity is described below.

Proposition 10 : The minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required by ML-based decoding to achieve the optimal DMT , is upper bounded as

$$\begin{aligned} c(r) \leq \overline{c}_{oaf}(r) = & \frac{2n-1}{n} r(n - \lfloor (2n-1)r \rfloor - 1) \\ & + \left(\lfloor (2n-1)r \rfloor - \frac{2n-1}{n} r(n-1) \right)^+, \end{aligned}$$

which is a piecewise linear function that, for $r = 0, \frac{1}{2n-1}, \cdots, \frac{n}{2n-1}$, takes the form

$$\overline{c}_{oaf}(r) = (2n-1)r(1 - \frac{2n-1}{n} r).$$

Proposition 10 also describes encoding-decoding policies that achieve the optimal DMT of OAF with $c(r) \leq \overline{c}_{oaf}(r)$.

For bidirectional channels, Theorem 17 and Theorem 18 describe the optimal DMT of the non-separated two-way relay channel with asymmetric fading and given the decode-and-forward protocol. Using this, Section 6.3.3 presents the complexity analysis and the joint reliability-complexity measure for some of the related two-way protocols.

Finally Chapter 7 presents conclusions and thoughts about future work.

Chapter 2

MIMO Channels, Lattice Code Designs and Bounded Search Decoders

This chapter presents system model describing the underlying MIMO channel model along with the description of encoding and decoding schemes considered for the rate, reliability and complexity tradeoff studies. In this dissertation we consider the general setting of *delay-limited outage-limited MIMO communications*, where communication takes place between several transmitting and receiving nodes, where the transmitting nodes generally do not have substantial information on the forward channel state, whereas receiving nodes have considerable information about the channels (coherent communications). This setting captures several pertinent communication scenarios in modern wireless communications. These include scenarios of delay-

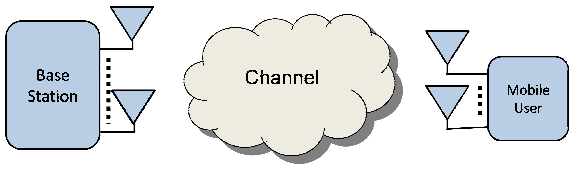


FIGURE 2.1 – Outage-limited channel model - no CSIT

limited data transmission in the absence of channel state information at the transmitter (no CSIT) as shown in Fig. 2.1, or more importantly scenarios that consider quick and high rate communication over the feedback link of CSIT in a multi-user setting in the absence of channel reciprocity as

shown in Fig. 2.2 , as well as the multi-user interference channel shown in Fig. 2.3 where interference management solutions require global CSIT, i.e., each transmitter is required to have channel state information of all forward links and even in the presence of channel reciprocity, CSIT required through feedback scales with the number of users.

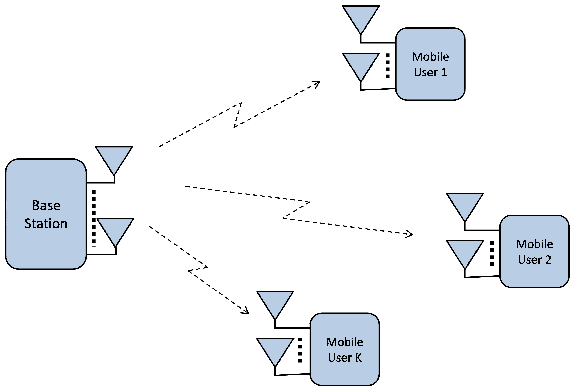


FIGURE 2.2 – Mutliuser MIMO downlink : In the absence of reciprocity each receiver is required to transmit CSIT over feedback link.

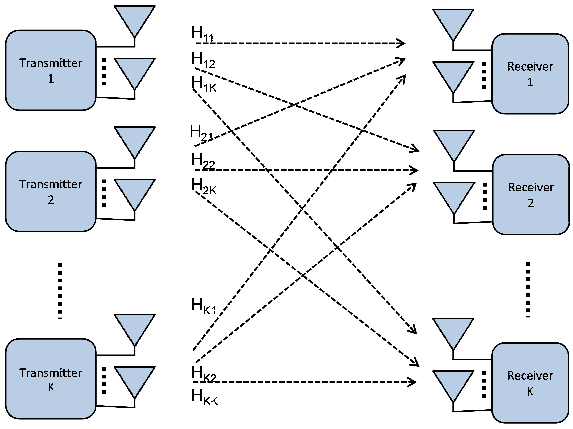


FIGURE 2.3 – Schematic for multi-user interference channel

The CSIT is meant to communicate large amounts of information (that naturally scale with the capacity of the channel) at high rate, firstly because any delays directly reduce the degrees of freedom for the forward channel communication, and secondly so that the CSIT information does not become obsolete. Furthermore, the presence of multiple users make these delay limitations more severe because global CSIT need to be shared among all transmitters and receivers in the network, which scales with the number of users. Also the presence of multiple users, brings to the fore the need for

high reliability because effect of erroneous communication of CSIT can carry through for a substantially long time and may potentially render the outage-limited feedback link as the reliability bottleneck of the entire transmission. Thus, the transmission of CSIT over the feedback link in a multi-user setting asks for reliable communication of large amounts of information at high rate under the risk of outage from the receiver to the transmitter.

The stringent requirements on delay and high performance can be met by employing DMT optimal lattice designs (cf. [12–16]) that offer high reliability with relatively short delays. Also, for these codes computationally efficient ML-based sphere decoding solutions can be used to provide a vanishing gap to brute force ML performance. The very long coding structures, such as for example turbo codes [17], polar codes [18] and low-density parity check (LDPC) codes [19] known to have excellent performance over different channels, often fail to meet these stringent delay requirements. Interesting work on such long codes, as well as their decoding performance and complexity, can be found in [20], see also [21], [22], [23] and a plethora of other works.

2.1 Channel model

We consider the general $m \times n$ MIMO channel representation given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ respectively denote the transmitted codewords, the received signal vectors, and the additive white Gaussian noise with unit variance, where the parameter ρ takes the role of the signal to noise ratio (SNR), and where the fading matrix $\mathbf{H} \in \mathbb{R}^{n\times m}$ is assumed to be random, with elements drawn from arbitrary statistical distributions. We consider that one use of (2.1) corresponds to T uses of some underlying "physical" channel. We further assume the transmitted codewords $\mathbf{x}$ to be uniformly distributed over some codebook $\mathcal{X} \in \mathbb{R}^m$, to be statistically independent of the channel $\mathbf{H}$, and to satisfy the power constraint

$$E\{\|\mathbf{x}\|^2\} \leq T. \tag{2.2}$$

We consider the rate,

$$R = \frac{1}{T}\log|\mathcal{X}|, \tag{2.3}$$

in bits per channel use (bpcu), where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$.

The model in (2.1) is known to encompass many pertinent communication scenarios such as quasi-static MIMO, MIMO orthogonal frequency division multiplexing (MIMO-OFDM), MIMO automatic repeat request (MIMO-ARQ), inter symbol interference (ISI) etc. Each of the above models introduces its own structure on $\mathbf{H}$ and $\mathbf{x}$, its own error performance limits, and generally its own requirements on coding and decoding schemes.

We now proceed to provide few specific MIMO channel models and establish their equivalence to the system model presented in (2.1).

2.1.1 Quasi-static MIMO

The general $n_T \times n_R$ point-to-point quasi-static MIMO channel is given by

$$\mathbf{Y}_C = \sqrt{\rho}\mathbf{H}_C\mathbf{X}_C + \mathbf{W}_C, \tag{2.4}$$

where $\mathbf{X}_C \in \mathbb{C}^{n_T \times T}$, $\mathbf{Y}_C \in \mathbb{C}^{n_R \times T}$ and $\mathbf{W}_C \in \mathbb{C}^{n_R \times T}$ represent the transmitted, received and noise signals over a period of T time slots, where the fading matrix $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$ is assumed to be random, with elements drawn from arbitrary distribution.

After vectorization the real values representation of (2.4) takes the form

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2.5}$$

where

$$\mathbf{H} = \mathbf{I}_T \otimes \mathbf{H}_R \in \mathbb{R}^{2n_RT \times 2n_TT}, \mathbf{H}_R = \begin{bmatrix} Re\{\mathbf{H}_C\} & -Im\{\mathbf{H}_C\} \\ Im\mathbf{H}_C\} & Re\{\mathbf{H}_C\} \end{bmatrix},$$

$$\mathbf{x} = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_TT} \text{ with } \mathbf{x}_t = [Re\{\mathbf{X}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{X}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{w} = (\mathbf{w}_1^{\mathrm{T}}, \cdots, \mathbf{w}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_RT} \text{ with } \mathbf{w}_t = [Re\{\mathbf{W}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{W}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_RT} \text{ with } \mathbf{y}_t = [Re\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

for $t = 1, \cdots, T$, where $\mathbf{X}_{t,C}$, $\mathbf{Y}_{t,C}$ and $\mathbf{W}_{t,C}$ are t-th column of $\mathbf{X}_C$, $\mathbf{W}_C$ and $\mathbf{Y}_C$ respectively.

2.1.2 MIMO-orthogonal frequency division multiplexing

A natural extension of the quasi-static MIMO channel is the $n_T \times n_R$ parallel, or MIMO-OFDM channel. In this setting

$$\mathbf{Y}_C^l = \sqrt{\rho}\mathbf{H}_C^l\mathbf{X}_C^l + \mathbf{W}_C^l, \quad l = 1, \cdots, L, \tag{2.6}$$

where $\mathbf{X}_C^l = [\mathbf{x}_C^{l,1}, \cdots, \mathbf{x}_C^{l,T}] \in \mathbb{C}^{n_T \times T}$ denotes the complex space-time block codeword transmitted over the l-th sub-channel in the T time-slots, and where $\mathbf{H}_C^l \in \mathbb{C}^{n_R \times n_R}$ is the channel matrix for the l-th sub-channel. Similar to the flat fading quasi-static channel in (2.4), it is clear by the linearity of (2.6) that the parallel or MIMO-orthogonal frequency division multiplexing (MIMO-OFDM) channel can be rewritten according to (2.1).

2.1.3 MIMO-automatic repeat request (MIMO-ARQ)

Another extension of the quasi-static MIMO channel is $n_T \times n_R$ MIMO with ARQ signaling. Under ARQ signaling, each message symbol from source is associated with a unique block $[\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^L]$ of matrices, each $\mathbf{X}_C^i \in \mathbb{C}^{n_T \times T}$, $i = 1, \cdots, L$, in such a way that it is possible to uniquely decode the message symbol given $[\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^\ell]$ for any $\ell = 1, \cdots, L$. For the long-term static channel model, which assumes that the channel encountered over the L ARQ rounds are identical received signal $\mathbf{Y}_C^\ell$ after ℓ-th ARQ round is given by

$$\mathbf{Y}_C^\ell = \sqrt{\rho}\mathbf{H}_C\mathbf{X}_C^{ARQ,\ell} + \mathbf{W}_C^\ell, \quad \ell = 1, \cdots, L, \tag{2.7}$$

where $\mathbf{X}_C^{ARQ,\ell} = [\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^\ell] \in \mathbb{C}^{n_T \times \ell T}$ denotes the complex space-time block code transmitted over ℓ rounds, and where $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$ is same as in (2.4). Again, it is clear by the linearity of (2.7) that the MIMO-ARQ channel can be rewritten according to (2.1).

Similarly ISI and MIMO-OFDM channels can be rewritten according to (2.1) as shown in Appendix 2A.

2.2 Encoding

We consider the general class of lattice codes defined by a *generator matrix* $\mathbf{G}$ and a *shaping region* $\mathcal{R}'$. Specifically for a rate R that scales with SNR as a function of the multiplexing gain $r = R/\log\rho \geq 0$, a (sequence of) full-rate linear (lattice) code(s) $\mathcal{X}_r$ is given by

$$\mathcal{X}_r = \Lambda_r \cap \mathcal{R}',$$

where $\mathcal{R}'$ is a compact convex subset of $\mathbb{R}^\kappa$ that is independent of ρ, where

$$\Lambda_r \triangleq \rho^{\frac{-rT}{\kappa}}\Lambda,$$

and where

$$\Lambda \triangleq \{\mathbf{Gs} \mid \mathbf{s} \in \mathbb{Z}^\kappa\},$$

where $\mathbb{Z}^\kappa$ denotes the κ dimensional integer lattice, and where $\mathbf{G} \in \mathbb{R}^{m \times \kappa}$ is full rank and independent of ρ. For full-rate code $\kappa = \min\{m, n\}$. In our analysis we consider only those lattice designs where generator matrix $\mathbf{G}$ remains fixed for all multiplexing gain values. The vectorized codewords take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}}\mathbf{Gs}, \qquad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}}\mathcal{R}, \tag{2.8}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}'$ that preserves the code, and contains the all zero vector $\mathbf{0}$. For simplicity we consider

$\mathcal{R} \triangleq [-1,\ 1]^{\kappa}$ to be a hypercube in $\mathbb{R}^{\kappa}$, although this could be relaxed without affecting the analysis and the results presented in this dissertation, as long as the constellation is the same for all $s_i,\ i = 1, \cdots, \kappa$. The presented general definition of lattice codes covers most of the codes in [9,12–16,24–37]. In this work a specific class of threaded algebraic space times codes [12] is considered to present several closed form expression for complexity exponent over quasi-static MIMO channels. The cyclic division algebra (cf. [13, 25]) based threaded (layered) code designs are the only currently known explicit constructions capable of achieving the DMT for all values of n_{T} ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$) and simultaneously over all $r \in [0,\ n_{\mathrm{T}}]$. All these codes have a common threaded (layered) structure. Specifically an $d \times d$ threaded code is built from d component codes mapped cyclically in threads (or layers) to the codewords $\mathbf{X}$. For example, in the special case of quasi-static MIMO channel with $d = n_{\mathrm{T}} = T = 4$, the thread structure is given by

$$\begin{bmatrix} 1 & 4 & 3 & 2 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

where the numbers $1,\ 2,\ 3,\ 4$ indicate the thread to which a particular entry of $\mathbf{X}$ belongs. In general, symbol j in thread l is mapped to $[\mathbf{X}]_{j,k}$ where $k = \mod(j - l, d) + 1$ and where $\mod(\bullet, d)$ denotes the modulo d operation. In the specific case of perfect codes [15, 28], the code takes the form

$$lay(\mathbf{X}) = \theta \begin{bmatrix} \mathbf{B}_0\mathbf{C} & & \\ & \ddots & \\ & & \mathbf{B}_d\mathbf{C} \end{bmatrix} \begin{bmatrix} s^{(1)} \\ \vdots \\ s^{(d)} \end{bmatrix}$$

where

$$\mathbf{B}_i = \mathrm{Diag}\{\underbrace{1, \cdots, 1}_{d-i \text{ entries}}, \underbrace{\gamma, \cdots, \gamma}_{i \text{ entries}}\},\ \text{for } i = 0, 1, \cdots, d-1$$

are full rank diagonal matrices incorporating a properly chosen thread-separating scalar $\gamma \in \mathbb{C}$, where $\mathbf{C} \in \mathbb{C}^{d \times d}$ is a (unitary) full rank generator matrix for the component code of each thread, $\mathbf{s}^l \in S_r^d$ are the constellation symbols of thread l, and where $lay(\mathbf{X})$ denotes the matrix to vector operation obtained by stacking the elements of $\mathbf{X}$ according to their thread. In this thesis we refer to this stacking as thread-wise grouping or layer-preserving ordering.

2.3 Decoding

Combining (2.1) and (2.8) yields the equivalent system model

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \tag{2.9a}$$

$$\text{where} \quad \mathbf{M} \triangleq \rho^{\frac{1}{2} - \frac{rT}{\kappa}} \mathbf{H}\mathbf{G} \in \mathbb{R}^{n \times \kappa}. \tag{2.9b}$$

Let $\mathbf{QR} = \mathbf{M}$ be the thin QR factorization of the code-channel matrix $\mathbf{M}$ and $\mathbf{r} \triangleq \mathbf{Q}^H\mathbf{y}$, then the equivalent system model in (2.9a) is represented by

$$\mathbf{r} = \mathbf{R}\mathbf{s} + \mathbf{Q}^H\mathbf{w},$$

and the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{S}_r^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2, \tag{2.10}$$

where (2.10) is an optimization problem suitable for the sphere decoder.

2.3.1 Sphere decoder (SD)

The sphere decoding algorithm is a branch-and-bound search on a regular tree. In the settings of coherent delay-limited outage-limited MIMO channels, SD algorithm complexity fluctuates with the fading channel realizations. In the worst case the sphere decoder is in essence forced to perform a complete search over the entire codebook, and its complexity is same as that of the brute force ML decoder. In the presence of constraints on the computational reserves that may be allocated to decoding, the algorithm is faced with the prospect of encountering channel realizations that force it to violate its run-time constraints, thus having to declare decoding outages that inevitably increase the error probability. The decoder can regulate the computational costs and error performance with the proper selection of decoding policies specifying when to decode and when not to. If a small gap to the optimal error performance is acceptable then sphere decoder algorithms [2–7, 38–48] have been known to require reduced computational resources. These sphere decoding solutions can be employed to decode a linear code from real lattices (cf. [3–5]) or complex lattices (cf. [49]) and allow for efficient optimal or near optimal decoding of a large number of high rate space-time codes [4,7,41,47].

The detailed descriptions and most implementation issues of the SD algorithm are found in [2] and the semi-tutorial papers [3–5]. In order to introduce notations that will be required for the subsequent complexity analysis presented in the next chapters, we present important SD algorithm details.

2.3.2 Sphere Decoding Algorithm

As shown in Fig. 2.4 the sphere decoder solves ML problem in (2.10) by recursively enumerating all lattice vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$ within a given sphere of radius $\xi > 0$, i.e., it identifies as candidates the vectors $\hat{\mathbf{s}}$ that satisfy

$$\|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2 \leq \xi^2. \tag{2.11}$$

The algorithm specifically uses the upper-triangular nature of $\mathbf{R}$ to recursively identify partial symbol vectors $\hat{\mathbf{s}}_k$, $k = 1, \cdots, \kappa$, for which

$$\|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2, \tag{2.12}$$

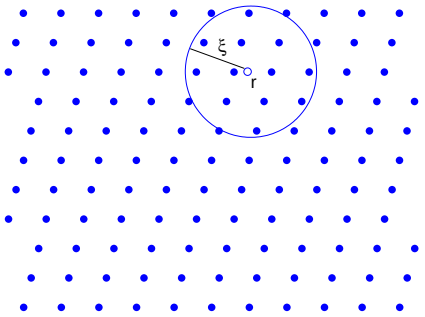


FIGURE 2.4 – Schematic for sphere decoder with search radius ξ.

where $\hat{\mathbf{s}}_k$ and $\mathbf{r}_k$ respectively denote the last k components of $\hat{\mathbf{s}}$ and $\mathbf{r}$, and where $\mathbf{R}_k$ denotes the $k \times k$ lower-right submatrix of $\mathbf{R}$. Clearly any set of vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$, with common last k components that fail to satisfy (2.12), may be excluded from the set of candidate vectors that satisfy (2.11).

The enumeration of partial symbol vectors $\hat{\mathbf{s}}_k$ is equivalent to the traversal of a regular tree with κ layers – one layer per symbol component of the symbol vectors, such that layer k corresponds to the kth component of the transmitted symbol vector [1] $\mathbf{s}$. A sample tree illustrating lattice points search in a $\kappa = 4$ dimensional sphere is shown Fig. 2.5.

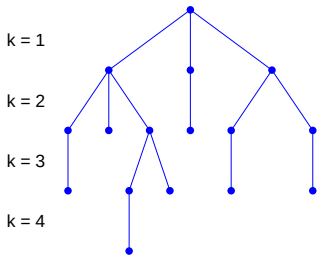


FIGURE 2.5 – Sample tree illustrating lattice points search in a $\kappa = 4$ dimensional sphere.

There is a one-to-one correspondence between the nodes at layer k and the partial vectors $\hat{\mathbf{s}}_k$. We say that a node is visited by the sphere decoder if and only if the corresponding partial vector $\hat{\mathbf{s}}_k$ satisfies (2.12), i.e., there is

1. We will henceforth refer to the symbol vector $\mathbf{s} \in \mathbb{S}_r^\kappa$ corresponding to the transmitted codeword $\mathbf{x} = \rho^{\frac{-rT}{\kappa}} \mathbf{Gs}$ (cf. (3.2)) , simply as the *transmitted symbol vector*.

a bijection between the visited nodes at layer k and the set

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{S}_r^k \mid \|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}. \tag{2.13}$$

Consequently the total number of visited nodes (in all layers of the tree) is given by

$$N_{SD} = \sum_{k=1}^{\kappa} N_k, \tag{2.14}$$

where $N_k \triangleq |\mathcal{N}_k|$ is the number of visited nodes at layer k of the search tree.

Complexity of sphere decoder

The total number of visited nodes (that scales polynomially with ρ) is commonly taken as a measure of the sphere decoder complexity ([3,5,7]). It is easy to show that in the scale of interest the SD complexity exponent $c(r)$ would not change if instead of considering the number of visited nodes, we considered the number of flops spent by the decoder. To see this, we consider that the cost of visiting a node, is independent of ρ. Once at a visited node, this same bounded cost includes the cost of establishing which children-nodes not to visit in the next layer.

The sphere decoder complexity (i.e., the number of visited nodes) is a random variable with a distribution that depends on a number of parameters, e.g., the system dimensions, the SNR, the transmission rate, the code generator matrix, and the search radius. This randomness must be considered when properly analyzing the sphere decoder complexity, unless one resorts to a worst-case analysis, which can be same as that of the brute force ML decoder and is unnecessarily pessimistic. In the next chapter we will describe decoding policies that take into account this randomness to estimates computational reserves required by SD algorithm to achieve vanishing gap to ML performance.

Sphere decoder search radius

The search radius of the sphere decoder affects both its error performance as well as the total number of visited nodes by the sphere decoder. In choosing this radius, we note that for the transmitted symbol vector $\mathbf{s}$, the metric in (2.10) satisfies

$$\|\mathbf{r} - \mathbf{R}\mathbf{s}\|^2 = \|\mathbf{Q}^H\mathbf{w}\|^2,$$

which means that if $\|\mathbf{Q}^H\mathbf{w}\| > \xi$, then the transmitted symbol vector is excluded from the search, resulting in a decoding error. For $\xi = \sqrt{z\log\rho}$ it can be shown that

$$\mathrm{P}\left(\|\mathbf{Q}^H\mathbf{w}\|^2 > \xi^2\right) \doteq \rho^{-z}.$$

We can set $\xi = \sqrt{z \log \rho}$, for some $z > d(r)$ such that

$$\mathrm{P}\left(\|\mathbf{Q}^H \mathbf{w}\|^2 > \xi^2\right) \doteq \rho^{-d(r)},$$

which implies a vanishing probability of excluding the transmitted information vector from the search, and a vanishing degradation of performance. If instead of fixed radius, we considered adaptive radius updates, the resulting complexity exponent would be the same (cf. [7]).

Decoding order policies

A sphere decoder can employ different decoding ordering policies, i.e., fixed decoding ordering and dynamic decoding ordering. The sphere decoder implementation with dynamic decoding ordering offers the possibility to choose specific set of column permutation matrices that minimize the complexity exponent of the sphere decoder by minimizing number of nodes visited by the sphere decoding algorithm for each channel realization. A particular decoding order is obtained by permuting the columns of generator matrix $\mathbf{G}$, i.e., replacing $\mathbf{G}$ with $\mathbf{G\Pi}$ where $\mathbf{\Pi} \in \mathbb{R}^{\kappa \times \kappa}$ is a permutation matrix. Note that choosing a different permutation matrix does not change the underlying code. Instead, the effect such a permutation would have is that it would change the order in which the symbols in $\mathbf{s}$ are enumerated by the sphere decoder. For fixed decoding ordering the column permutation matrix $\mathbf{\Pi}$ remains fixed for the entire duration of communication. The special case of $\mathbf{\Pi} = \mathbf{I}$ is the *natural decoding ordering* for any given code. For the case of layered codes (cf. [50], [15], [14], etc.), another set of decoding orderings will be those that preserve the layered structure of the codes.

For dynamic decoding orderings the column permutation matrix $\mathbf{\Pi}$ is channel dependent and changes dynamically with the channel realizations.

Appendix 2A : Inter-symbol-interference (ISI) channel model

From multi-path to OFDM

Consider the L tap SISO frequency selective channel $\{h_0, \cdots, h_{L-1}\}$ and consider N_c symbols $\mathbf{x} = [x_0, x_1, \cdots, x_{N_c-1}]$. Let

$$\mathbf{z} = \boldsymbol{F}^{-1}\mathbf{x} = [z_0, z_1, \cdots, z_{N_c-1}],$$

where $\boldsymbol{F}^{-1}$ denotes the N_c-point inverse DFT (modulation step). Add cyclic prefix to get

$$\mathbf{u} = \left[z_{N_c-L+1}, z_{N-c-L+2}, \cdots, z_{N_c-1}, z_0, z_1, \cdots, z_{N_c-1}\right],$$

$N_c > L, N_c < \infty$. Transmit $\mathbf{u}$ sequentially in time. The received signal is

$$r_m = \sqrt{\rho}\sum_{l=0}^{L-1} h_l \mathbf{u}_{m-l} + w_m, \quad m = 1 \to N_c + L - 1.$$

After ignoring the first L symbols (guard-band), we have

$$\mathbf{r} = [r_L, r_{L+1}, \cdots, r_{N_c+L-1}]^\dagger_{N_c\times 1},$$
$$\mathbf{w} = [w_L, w_{L+1}, \cdots, w_{N_c+L-1}]^\dagger_{N_c\times 1},$$

and

$$\mathbf{r} = \sqrt{\rho}\mathbf{h} * \mathbf{z} + \mathbf{w},$$

where $*$ denotes circular convolution. Proceed to take the N_c-point DFT to get

$$\begin{aligned}\mathbf{y} =& \boldsymbol{F}\mathbf{r},\\ =& \sqrt{\rho}\boldsymbol{F}(\mathbf{h} * \mathbf{z}) + \widetilde{\mathbf{w}},\end{aligned}$$

where $\widetilde{\mathbf{w}} = \boldsymbol{F}\mathbf{w}$. Let

$$\widetilde{\mathbf{z}} := \mathrm{DFT}(\mathbf{z}) = \boldsymbol{F}\cdot\mathbf{z} = \begin{bmatrix} \widetilde{z}_0 \\ \widetilde{z}_1 \\ \vdots \\ \widetilde{z}_{N_c-1} \end{bmatrix},$$

where F is the N_c-point DFT matrix, and where

$$\widetilde{z}_j := \frac{1}{\sqrt{N_c}}\sum_{m=0}^{N_c-1} z_m exp(-2\imath\pi m/N_c), \quad j = 0, ..., N_c - 1.$$

Hence

$$\mathrm{DFT}[(\mathbf{h}\otimes\mathbf{z})_j] = \sqrt{N_c}\mathrm{DFT}(\mathbf{h})_j \cdot \mathrm{DFT}(\mathbf{z})_j = \widetilde{h}_j\widetilde{z}_j,$$

where $\widetilde{h}_j := \frac{1}{\sqrt{N_c}}\sum_{m=0}^{N_c-1} h_m exp(-2\imath\pi m/N_c)$, $j = 0, ..., N_c - 1$, thus

$$y_j = \sqrt{\rho}\widetilde{h}_j\widetilde{z}_j + \widetilde{w}_j, \quad j = 0, ..., N_c - 1,$$

i.e.,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_c} \end{bmatrix} = \sqrt{\rho}\underbrace{\begin{bmatrix} \widetilde{h}_1 & & & \\ & \widetilde{h}_2 & & \\ & & \ddots & \\ & & & \widetilde{h}_{N_c} \end{bmatrix}}_{\boldsymbol{H}_C}\begin{bmatrix} \widetilde{z}_1 \\ \widetilde{z}_2 \\ \vdots \\ \widetilde{z}_{N_c} \end{bmatrix} + \begin{bmatrix} \widetilde{w}_1 \\ \widetilde{w}_2 \\ \vdots \\ \widetilde{w}_{N_c} \end{bmatrix}.$$

Thus,

$$\mathbf{y} = \sqrt{\rho}\boldsymbol{H}_C\widetilde{\mathbf{z}} + \widetilde{\mathbf{w}} = \sqrt{\rho}\boldsymbol{H}_C\boldsymbol{F}\mathbf{z} + \widetilde{\mathbf{w}} = \sqrt{\rho}\boldsymbol{H}_C\boldsymbol{F}\boldsymbol{F}^{-1}\mathbf{x} + \widetilde{\mathbf{w}} = \sqrt{\rho}\boldsymbol{H}_C\mathbf{x} + \widetilde{\mathbf{w}},$$

which can be rewritten according to (2.1).

From MIMO-ISI to MIMO-OFDM

Consider the $n_T \times n_T$ L-tap MIMO channel $\{\boldsymbol{G}_0, \cdots, \boldsymbol{G}_{L-1}\}$, and $N_c \geq L$ tones. Let $n := n_T N_c$. Let $\mathbf{x}(k)$ be the $n \times 1$ vector corresponding to the kth time slot, and let $\mathbf{z}'(k) := (\boldsymbol{F}^{-1} \otimes \boldsymbol{I}_{n_T})\mathbf{x}(k) \in \mathbb{R}^n$ come from tone modulation. After adding cyclic prefix, we get $\mathbf{u}'(k) = \boldsymbol{A}_1 \mathbf{z}'(k) \in \mathbb{R}^{(L+N_c)n_t}$, where $\boldsymbol{A}_1 := \left(\underbrace{\begin{bmatrix} \mathbf{0} & \boldsymbol{I}_L \\ & \boldsymbol{I}_{N_c} \end{bmatrix}}_{\in \mathbb{R}^{(L+N_c)\times N_c}} \otimes \boldsymbol{I}_{n_T} \right) \in \mathbb{R}^{(L+N_c)n_T \times N_c n_T}$. Sequentially transmit $(L+N_c)$ times, the n_T-tuples (in space) of $\mathbf{u}'(k)$. The received signal is $\mathbf{r}'(k) = \sqrt{\rho}\boldsymbol{G}'\mathbf{u}'(k) + \mathbf{v}'(k) + \boldsymbol{G}_{\text{isi}}\mathbf{u}'(k-1)$, where $\boldsymbol{G}' := \begin{bmatrix} \boldsymbol{G}_0 & & & \\ & \boldsymbol{G}_0 & & \\ & & \ddots & \\ \boldsymbol{G}_{L-1} & & & \\ & \ddots & & \\ & & \boldsymbol{G}_{L-1} & \cdots \quad \boldsymbol{G}_0 \end{bmatrix}$ is a Toeplitz matrix. Let $\mathbf{r}(k) = \boldsymbol{A}_2 \mathbf{r}'(k)$, denote the action of nulling/dropping the first L n_T-tuples of $\mathbf{r}'(k)$, where $\boldsymbol{A}_2 := \left(\begin{bmatrix} \mathbf{0}_{N_c \times L} & \boldsymbol{I}_{N_c} \end{bmatrix} \otimes \boldsymbol{I}_{n_R} \right)$. Proceed to take the N_c-point DFT to get

$$\begin{aligned}
\mathbf{y}(k) &= (\boldsymbol{F} \otimes \boldsymbol{I}_{n_r})\mathbf{r}(k), \\
&= (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \mathbf{r}'(k), \\
&= \sqrt{\rho}(\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \boldsymbol{G}' \mathbf{u}'(k) + (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \mathbf{v}'(k), \\
&= \sqrt{\rho}(\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \boldsymbol{G}' \boldsymbol{A}_1 \mathbf{z}'(k) + (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \mathbf{v}'(k), \\
&= \sqrt{\rho}(\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \boldsymbol{G}' \boldsymbol{A}_1 (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\mathbf{x}(k) + (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \mathbf{v}'(k).
\end{aligned}$$

Given that $\boldsymbol{A}_2 \boldsymbol{G}' \boldsymbol{A}_1$ is circulant, and given that left-right DFT, IDFT diagonalizes circulant matrices, we get

$$\mathbf{y}(k) = \underbrace{\begin{bmatrix} \boldsymbol{H}(0)_{n_R \times n_T} & & \\ & \ddots & \\ & & \boldsymbol{H}(N_c - 1)_{n_R \times n_T} \end{bmatrix}}_{\boldsymbol{H}_C} \mathbf{x}(k) + \mathbf{v}(k),$$

where $\boldsymbol{H}_C = (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \boldsymbol{G}' \boldsymbol{A}_1 (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})$ and where $\mathbf{v}(k) = (\boldsymbol{F} \otimes \boldsymbol{I}_{n_R})\boldsymbol{A}_2 \mathbf{v}'(k)$. For $k = 1 \to T$, it takes the form

$$\begin{bmatrix} \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(T) \end{bmatrix} = \sqrt{\rho}(\boldsymbol{H}_C \otimes \boldsymbol{I}_T) \begin{bmatrix} \mathbf{x}(1) \\ \vdots \\ \mathbf{x}(T) \end{bmatrix} + \begin{bmatrix} \mathbf{v}(1) \\ \vdots \\ \mathbf{v}(T) \end{bmatrix},$$

which can be easily rewritten according to (2.1).

Chapter 3

Complexity of Maximum Likelihood Decoding

3.1 Introduction

The purpose of this chapter is two-fold. The first task is to extend the scope of the complexity analysis in previous work (cf. [7]), to a very broad setting of lattice designs, decoding policies and fading statistics. This extension is important in further exposition. The second task is to provide meaningful rate-reliability-complexity measures and tradeoffs. The mathematical exposition is for the general MIMO case, and the single-letter expressions are derived for the quasi-static MIMO channel. All the results hold for ML-based decoding, and the algorithmic analysis considers the efficient family of ML-based sphere decoding algorithms. The derived complexity exponents of these algorithms describe the sufficient, and in many cases necessary, computational resources required for ML based SD to achieve either a specific DMT performance, or to achieve a vanishing gap to (full) ML.

Before proceeding with the analysis, we note that the choice of lattice designs and fading statistics may conceivably play a crucial role in defining the complexity behavior of different decoders. Another aspect that could have substantial complexity ramification is that of the sphere decoding ordering policies. We here provide universal upper bounds on complexity that hold irrespective of such choices on lattice designs, fading statistics and decoding ordering policies. We then proceed to establish the tightness of these bounds in a very broad setting (of designs, statistics, and ordering policies). Even though this tightness appears in a very broad setting, it still remains to be shown whether any of these choices can in fact substantially reduce the

complexity behavior.

We consider the general $m \times n$ point-to-point MIMO channel, which can be represented as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{3.1}$$

where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ respectively denote the transmitted codewords, the received signal vectors, and the additive white Gaussian noise with unit variance, where ρ denotes the signal to noise ratio (SNR), and where the fading matrix $\mathbf{H} \in \mathbb{R}^{n\times m}$ is assumed to be random, with elements drawn from arbitrary statistical distributions.

For the lattice codes described in the previous chapter, the codewords take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}}\mathbf{G}\mathbf{s}, \qquad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}}\mathcal{R}, \tag{3.2}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}'$ that preserves the code, and contains the all zero vector $\mathbf{0}$. For simplicity we consider $\mathcal{R} \triangleq [-1,\ 1]^\kappa$ to be a hypercube in $\mathbb{R}^\kappa$, although this could be relaxed. Combining (3.1) and (3.2) yields the equivalent system model

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \tag{3.3a}$$

$$\text{where} \quad \mathbf{M} \triangleq \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{H}\mathbf{G} \in \mathbb{R}^{n\times\kappa}. \tag{3.3b}$$

Let $\mathbf{QR} = \mathbf{M}$ be the thin QR factorization of the code-channel matrix $\mathbf{M}$ and $\mathbf{r} \triangleq \mathbf{Q}^H\mathbf{y}$, then the equivalent system model in (3.3a) is represented by

$$\mathbf{r} = \mathbf{R}\mathbf{s} + \mathbf{Q}^H\mathbf{w},$$

and the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{S}_r^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2, \tag{3.4}$$

which is then solved by the sphere decoder which recursively enumerates all candidate vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$ within a given search sphere of radius $\xi > 0$. We can set a fixed search radius $\xi = \sqrt{z\log\rho}$ for some $z > d(r)$ such that

$$\mathrm{P}\left(\|\mathbf{Q}^H\mathbf{w}\|^2 > \xi^2\right) \doteq \rho^{-d(r)}, \tag{3.5}$$

which implies a vanishing probability of excluding the transmitted information vector from the search.

Rate-reliability-complexity tradeoff

In the high SNR regime, a given encoder $\mathcal{X}_r$ and decoder $\mathcal{D}_r$ are said to achieve a *multiplexing gain* r and *diversity gain* $d_\mathcal{D}(r)$ if (cf. [51])

$$\lim_{\rho\to\infty}\frac{R(\rho)}{\log\rho} = r, \qquad \text{and} \qquad -\lim_{\rho\to\infty}\frac{\log P_e}{\log\rho} = d_\mathcal{D}(r) \tag{3.6}$$

where P_e denotes the probability of codeword error with a ML-based sphere decoder $\mathcal{D}_r$ employing time-out policies.

As stated before for $N_{\max}$ denoting the amount of computational reserves, in floating point operations (flops) per T channel uses, that the transceiver is endowed with, the complexity exponent then takes the form

$$c(r) := \lim_{\rho\to\infty} \frac{\log N_{\max}}{\log \rho}. \tag{3.7}$$

In order to establish the complexity exponent ($c(r)$) that describes the computational resources required to achieve a vanishing gap to brute force ML performance we quantify, in the high SNR regime, the ML-based SD the error performance gap as

$$g(c) \triangleq \lim_{\rho\to\infty} \frac{P_e}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)} = 1, \tag{3.8}$$

where $\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right) \doteq \rho^{-d(r)}$ describes the error probability of the brute force ML decoder.

Now we proceed to identify the amount of computational reserves required to achieve a vanishing gap to the brute force ML performance. While this issue was first addressed and partially answered in [7] for the specific settings of i.i.d. Rayleigh fading quasi-static channels with specific channel dimensions, specific codes, and specific permutation orderings, we here provide answers for the most general MIMO settings, i.e., all fading distributions from rational numbers, all channel dimensions, all MIMO scenarios and all full-rate and below-full-rate lattice codes. After that Section 3.3 will present the joint reliability-complexity measure that can be used to compare worth of encoding-decoding policies, and Section 3.4 presents rate-reliability-complexity tradeoff for ML-based decoders.

3.2 Complexity analysis ML-based sphere decoding

As discussed in the previous chapter the total number of visited nodes is commonly taken as a measure of the sphere decoder complexity. We recall that the total number of visited nodes (in all layers of the tree) is given by

$$N_{SD} = \sum_{k=1}^{\kappa} N_k, \tag{3.9}$$

where N_k denotes the number of visited nodes at layer k that corresponds to the kth component of the transmitted symbol vector $\mathbf{s}$ and is given by $N_k \triangleq |\mathcal{N}_k|$ where

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{S}_r^{\kappa} \mid \|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}.$$

Before proceeding further we want to clarify that the analysis presented here is specific to sphere decoding, and that it does not account for any other ML based solutions that could, under some (arguably rare) circumstances, be more efficient. A classical example of such rare circumstances would be a MIMO scenario, or equivalently a set of fade statistics, that always generate diagonal channel matrices. Another example would be having codes drawn from orthogonal designs which introduce very small decoding complexity, but which are provably shown to be highly suboptimal except for very few unique cases like the 2×1 quasi-static MIMO case [52]. [1]

3.2.1 Complexity for vanishing gap to ML performance

We are interested in the ML-based SD complexity required to achieve a vanishing performance gap to brute force ML. We recall that a ML-based SD with run-time constraints, in addition to making the ML errors ($\hat{\mathbf{s}}_{ML} \neq \mathbf{s}$), also makes errors when the run-time limit of ρ^x flops for $x > c(r)$ becomes active, as well as when the fixed search radius ξ causes $\mathcal{N}_\kappa = \emptyset$. Consequently the corresponding performance gap to the brute force ML decoder, takes the form (cf. (3.8))

$$g(x) = \lim_{\rho \to \infty} \frac{\mathrm{P}\left(\{\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\} \cup \{N_{SD} \geq \rho^x\} \cup \{\mathcal{N}_\kappa = \emptyset\}\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)}.$$

To bound the above gap, we apply the union bound along with the fact that (see Section 2.3.2)

$$\mathrm{P}\left(\mathcal{N}_\kappa = \emptyset\right) \leq \mathrm{P}\left(\|\mathbf{Q}^H \mathbf{w}\| > \xi\right),$$

to get that

$$g(x) \quad \leq \quad \lim_{\rho \to \infty} \frac{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right) + \mathrm{P}\left(N_{SD} \geq \rho^x\right) + \mathrm{P}\left(\|\mathbf{Q}^H \mathbf{w}\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)}. \quad (3.10)$$

We recall that we set the search radius $\xi = \sqrt{z \log \rho}$, for some $z > d(r)$ such that $\mathrm{P}\left(\|\mathbf{Q}^H \mathbf{w}\| > \xi\right) \doteq \rho^{-d(r)}$, it then follows that

$$\lim_{\rho \to \infty} \frac{\mathrm{P}\left(\|\mathbf{Q}^H \mathbf{w}\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)} = 0. \quad (3.11)$$

1. *Conjecture :* We here conjecture that for a $n_T \times n_R$ quasi-static MIMO, in the limit of a large number n_T of transmit antennas, and given encoding drawn from orthogonal designs, the complexity of ML decoding one degree of freedom scales exponentially with at least $n_\mathrm{T}/2$. This conjecture is supported by the fact that, as n_T increases, orthogonal designs maximally allow for $1/2$ degrees of freedom, at a delay of $T = O(2^{n_\mathrm{T}/2})$ (cf. [52]), and by the fact that square orthogonal designs with $T \approx n_\mathrm{T}$, allow for approximately $\frac{\log n_\mathrm{T}}{n_\mathrm{T}}$ degrees of freedom, at a decoding complexity that scales as $O(n_\mathrm{T})$. We remind the reader that the number of degrees of freedom is, in our setting, generally equal to $\kappa/2T$, or equivalently equal to the maximum achievable multiplexing gain that allows for positive diversity gain. To put the above conjecture in perspective, we note that for brute force ML decoding, in the presence of uncoded (V-BLAST, $T = 1$) BPSK symbols (n_T of them), the complexity required per one degree of freedom scales as $O(2^{n_\mathrm{T}})$.

Thus a vanishing gap to the brute force ML decoding ($g(x) = 1$) requires that

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right)} = 0.$$

Now going back to (3.7), and having in mind appropriate timeout policies that guarantee a *vanishing* gap, the complexity exponent $c(r)$ can be bounded as $\underline{c}(r) \leq c(r) \leq \overline{c}(r)$, where

$$\overline{c}(r) \triangleq \inf\{x \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\log \rho} > d(r)\}, \text{ and} \tag{3.12a}$$

$$\underline{c}(r) \triangleq \sup\{x \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\log \rho} < d(r)\}. \tag{3.12b}$$

We note that $\overline{c}(r)$ in (3.12a) denotes a sufficient condition that guarantee a vanishing performance gap to optimal performance, while $\underline{c}(r)$ in (3.12b) denotes a necessary condition for a vanishing gap to optimal performance.

3.2.2 Complexity for fixed decoding order

In this section we consider ML-based SD employing fixed decoding order. We first establish an upper bound that holds irrespective of the any fixed or dynamically changing decoding order and then provide a lower bound that matches this upper bound for any fixed decoding order. The complexity results presented here are valid for all channel dimensions and for all-rate lattice codes. For simplicity of analysis without loss of any generality [2] we consider the case of $m \leq n$ and $\kappa = m$. We define $\mu_i \triangleq -\frac{\log \sigma_i(\mathbf{H}^H\mathbf{H})}{\log \rho}$, $i = 1, \cdots, m$, where $\sigma_i(\mathbf{H}^H\mathbf{H})$ denotes i-th singular value of the channel matrix $\mathbf{H}^H\mathbf{H}$. It follows that

$$\begin{aligned}
\sigma_i(\mathbf{R}) &= \sigma_i(\mathbf{M}), \\
&\geq \rho^{\frac{1}{2}-\frac{rT}{\kappa}} \sigma_{\min}(\mathbf{G})\sigma_{(i)}(\mathbf{H})), \\
&\doteq \rho^{\frac{-rT}{\kappa}+\frac{1}{2}(1-\mu_i)}, \quad i = 1, \cdots, \kappa,
\end{aligned} \tag{3.13}$$

2. For the considered case $n \geq m$ and full-rate codes ($\kappa = m$), generator matrix $\mathbf{G} \in \mathbb{R}^{m\times m}$ is full-rank and square matrix, but for $n < m$ and full-rate codes ($\kappa = n$), generator matrix $\mathbf{G}$ is full-rank but not a square matrix. In order to have uniformity in analysis, for the case of $n < m$, we can substitute $\mathbf{G}$ with a new generator matrix $\tilde{\mathbf{G}} = \tilde{\mathbf{U}}^{\mathbf{T}}\mathbf{G} \in \mathbb{R}^{n\times n}$ which is full-rank and square matrix, and where $\tilde{\mathbf{U}} \in \mathbb{R}^{m\times n}$ has orthogonal columns, this substitution results in a new channel matrix $\tilde{\mathbf{H}} = \mathbf{H}\tilde{\mathbf{U}} \in \mathbb{R}^{n\times n}$ such that the code-channel matrix remains unaltered, i.e., $\mathbf{Ms} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{HGs} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{H}\tilde{\mathbf{U}}\tilde{\mathbf{G}}\mathbf{s} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\tilde{\mathbf{H}}\tilde{\mathbf{G}}\mathbf{s}$. As no explicit assumption is made regarding the fading distribution of $\mathbf{H}$, the complexity analysis and results will directly apply for $n < m$ with new matrices $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$. The analysis for the below-full-rate code ($\kappa < \min\{m, n\}$) can be handled by making similar substitutions.

where the asymptotic equality is due to the fact that $\sigma_{\min}(\mathbf{G}) \doteq \rho^0$. We recall from (3.9) that the total number of the visited nodes for any given channel realization $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_m)$ takes the form

$$N_{SD}(\boldsymbol{\mu}) = \sum_{k=1}^{\kappa} N_k(\boldsymbol{\mu}),$$

where from [7, Lemma 1]

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + \min\{ \frac{2\xi}{\sigma_i(\mathbf{R}_k)}, 2\sqrt{k}\rho^{\frac{rT}{\kappa}} \} \right].$$

From the interlacing property of singular values of sub-matrices [53] we have that $\sigma_i(\mathbf{R}_k) \geq \sigma_i(\mathbf{R})$. It follows that

$$N_k(\boldsymbol{\mu}) \dot{\leq} \rho^{\sum_{i=1}^{k} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^+}.$$

Consequently, we have that

$$N_{SD}(\boldsymbol{\mu}) \dot{\leq} \rho^{\sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^+}. \tag{3.14}$$

Now let

$$\mathcal{T}(x) \triangleq \left\{ \boldsymbol{\mu} | \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa} \right)^+ \geq x \right\}, \tag{3.15}$$

and note that for any $y < x$, then (3.14) and $\boldsymbol{\mu} \notin \mathcal{T}(y)$ jointly imply that $N_{SD} < \rho^x$, which in turn implies that $\mathrm{P}\left(\boldsymbol{\mu} \notin \mathcal{T}(y)\right) \leq \mathrm{P}\left(N_{SD} < \rho^x\right)$ and consequently that $\mathrm{P}\left(\boldsymbol{\mu} \in \mathcal{T}(x)\right) \geq \mathrm{P}\left(N_{SD} \geq \rho^x\right)$. Thus for any $y < x$ it follows that

$$-\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\log \rho} \geq -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(\boldsymbol{\mu} \in \mathcal{T}(y)\right)}{\log \rho}. \tag{3.16}$$

In evaluating the right hand side of (3.16) we note that $\mathcal{T}(y)$ is a closed set and thus, applying the large deviation principle (cf. [8]), we have that

$$-\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(\boldsymbol{\mu} \in \mathcal{T}(y)\right)}{\log \rho} \geq \inf_{\boldsymbol{\mu}\in\mathcal{T}(y)} I(\boldsymbol{\mu}), \tag{3.17}$$

where $\boldsymbol{\mu} \triangleq (\mu_1, \cdots, \mu_\kappa)$ satisfies the large deviation principle with rate function $I(\boldsymbol{\mu})$. Consequently from (3.16) and (3.17), it follows that

$$-\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\log \rho} \geq \inf_{\boldsymbol{\mu}\in\mathcal{T}(y)} I(\boldsymbol{\mu}). \tag{3.18}$$

This lower bound specified in (3.18) holds for any $y < x$. Consequently to get the tightest possible bound, we need to find $\sup_{y<x} \inf_{\boldsymbol{\mu}\in\mathcal{T}(y)} I(\boldsymbol{\mu})$. As $\inf_{\boldsymbol{\mu}\in\mathcal{T}(y)} I(\boldsymbol{\mu})$ is non-decreasing and left-continuous in y, it follows that

$$\sup_{y<x} \inf_{\boldsymbol{\mu}\in\mathcal{T}(y)} I(\boldsymbol{\mu}) = \inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu}).$$

Consequently, it follows that

$$-\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N_{SD} \geq \rho^x\right)}{\log \rho} \geq \inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu}), \tag{3.19}$$

which in conjunction with (3.12a) gives that

$$\begin{aligned}\overline{c}(r) \leq \tilde{c}(r) &\triangleq \inf\{x| \inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu}) > d(r)\},\\ &= \sup\{x| \inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu}) \leq d(r)\},\\ &= \max\{x| \inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu}) \leq d(r)\},\end{aligned} \tag{3.20}$$

where the above follows from the aforementioned fact that $-\lim_{\rho\to\infty} \frac{\log \mathrm{P}(N_{SD}\geq\rho^x)}{\log\rho}$ (and by extension also $\inf_{\boldsymbol{\mu}\in\mathcal{T}(x)} I(\boldsymbol{\mu})$) is continuous and nondecreasing in x, and from the fact that $\mathcal{T}(x)$ is a closed set. Consequently $\tilde{c}(r)$ takes the form

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} x \tag{3.21a}$$

$$\text{s.t.} \quad \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^+ \geq x, \tag{3.21b}$$

$$I(\boldsymbol{\mu}) \leq d(r), \tag{3.21c}$$

$$\mu_1 \geq \cdots \geq \mu_\kappa \geq 0. \tag{3.21d}$$

Furthermore since $\mathcal{T}(x)$ is a closed set, the maximum x in (3.21) must be such that (3.21b) is satisfied with equality, in which case the upper bound $\tilde{c}(r)$ can be obtained as the solution to a constrained maximization problem according to

$$\begin{aligned}\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \; & \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^+\\ \text{s.t.} \; & I(\boldsymbol{\mu}) \leq d(r),\\ & \mu_1 \geq \cdots \geq \mu_\kappa \geq 0.\end{aligned}$$

This results in following theorem

Theorem 1. *The complexity exponent of sphere decoding any full-rate lattice code at $d(r)$ is upper bounded, irrespective of the fading statistics and fixed or dynamically changing decoding orderings, as $c(r) \leq \tilde{c}(r)$ where*

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^+ \tag{3.22a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d(r), \tag{3.22b}$$

$$\mu_1 \geq \cdots \geq \mu_\kappa \geq 0, \tag{3.22c}$$

Let $d(r) = d_{\text{ML}}$ be the optimal DMT of uninterrupted ML decoding of the specific (potentially suboptimal) code. Then we have the following

Corollary 1a. *The complexity exponent of achieving a vanishing gap to optimal ML is upper bounded as*

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{m} \min\left(\frac{rT}{m} - \frac{1}{2}(1-\mu_i), \frac{rT}{m}\right)^+ \tag{3.23a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d_{\text{ML}}(r), \tag{3.23b}$$

$$\mu_1 \geq \cdots \geq \mu_m \geq 0, \tag{3.23c}$$

irrespective of the full-rate code, the channel statistics and the ordering policies.

The above holds directly from Theorem 1.

Lower Bound on Complexity

In this section we establish that $\underline{c}(r) = \tilde{c}(r)$, i.e., the sphere decoder visits a total number of nodes that is close to $\rho^{\tilde{c}(r)}$ with a probability that is large compared to the probability of decoding error $\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right) \doteq \rho^{-d(r)}$.

For $\boldsymbol{\mu}^* = (\mu_1^*, \cdots, \mu_\kappa^*)$ being one of the maximizing vectors such that $I(\boldsymbol{\mu}^*) = d(r)$, the upper bound takes the form

$$\tilde{c}(r) = \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i^*), \frac{rT}{\kappa}\right)^+.$$

Furthermore given the monotonicity of the rate function $I(\boldsymbol{\mu})$, and the fact that the objective function in (3.22) does not increase in μ_i beyond $\mu_i = 1$, we may also assume without loss of generality that $\mu_i^* \leq 1$ for $i = 1, \cdots, \kappa$. It then follows that

$$\tilde{c}(r) = \sum_{i=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i^*)\right)^+. \tag{3.24}$$

We let $q \in [1, \kappa]$ be the largest integer for which $\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_q^*) > 0$, in which case (3.24) takes the form

$$\tilde{c}(r) = \sum_{i=1}^{q} \left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i^*) \right). \tag{3.25}$$

We quickly note that without loss of generality we can assume that $q \geq 1$ as otherwise $\bar{c}(r) = c(r) = 0$. Consequently it is the case that $\mu_i^* > 0$ for $i = 1, \cdots, q$.

We proceed to define three events Ω_1, Ω_2 and Ω_3 which we will prove to be jointly sufficient so that the total number of nodes visited by the sphere decoder, employing a channel dependent fixed decoding order, is close to $\rho^{\tilde{c}(r)}$. These events are given by

$$\begin{aligned} \Omega_1 \triangleq \{ & \mu_i^* - 2\delta < \mu_i < \mu_i^* - \delta, j = 1, \cdots, q \\ & 0 < \mu_i < \delta, i = q+1, \cdots, \kappa \}, \end{aligned} \tag{3.26}$$

for a given small $\delta > 0$,

$$\Omega_2 \triangleq \left\{ \|\mathbf{w}\|^2 < \xi^2 \right\}, \tag{3.27}$$

$$\Omega_3 \triangleq \left\{ \|\mathbf{s}\| < \frac{1}{2} \rho^{\frac{rT}{\kappa}} \right\}. \tag{3.28}$$

Note also that by choosing δ sufficiently small, and using the fact that $\mu_i^* > 0$ for $i = 1, \cdots, q$, we may without loss of generality assume that Ω_1 implies that $\mu_i > 0$ for all $i = 1, \cdots, \kappa$.

We begin the proof by first showing that in the presence of events Ω_1, Ω_2 and Ω_3 we can remove the boundary constraints of ML-based SD in (3.4), which allows us to lower bound on the number of nodes visited at layer k as (cf. [7, Lemma 1])

$$N_k \geq \prod_{i=1}^{k} \left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k} \right]^+. \tag{3.29}$$

For boundary removal we let $\hat{\mathbf{x}}_k \in \mathbb{S}_\infty^k$ be an arbitrary point in the k-dimensional infinite constellation that satisfies the sphere constraint at layer k, i.e.

$$\|\mathbf{r}_k - \mathbf{R}_k \hat{\mathbf{s}}_k\| \leq \xi. \tag{3.30}$$

Note that $\mathbf{r}_k = \mathbf{R}_k \mathbf{s}_k + \mathbf{w}_k$, where $\mathbf{s}_k$ denotes the last k components of the transmitted sysmbol vector $\mathbf{s} \in \mathbb{S}_r^\kappa$ and where $\mathbf{w}_k$ denotes the last k components of $\mathbf{Q}^H \mathbf{w}$. It follows that

$$\begin{aligned} \|\mathbf{r}_k - \mathbf{R}_k \hat{\mathbf{s}}_k\| = & \|\mathbf{R}_k(\mathbf{s}_k - \hat{\mathbf{s}}_k) + \mathbf{w}_k\|, \\ \geq & \sigma_1(\mathbf{R}_k)\|(\hat{\mathbf{s}}_k - \mathbf{s}_k)\| - \|\mathbf{w}_k\|, \end{aligned}$$

substituting this back in (3.30) implies that

$$\|\hat{\mathbf{s}}_k\| \leq \frac{1}{\sigma_1(\mathbf{R}_k)}(\xi + \|\mathbf{w}_k\|) + \|\mathbf{s}_k\|. \tag{3.31}$$

From (3.13), (3.26) and fact that $\mu_1^* \leq 1$ it further follows that

$$\sigma_1(\mathbf{R}_k) \geq \sigma_1(\mathbf{R}) \dot{\geq} \rho^{-\frac{rT}{\kappa}+\frac{\delta}{2}}.$$

As $\xi \doteq \rho^0$ and $\|\mathbf{w}_k\| \leq \|\mathbf{Q}^H\mathbf{w}\| \leq \xi$ (cf. (3.27)) it follows that

$$\frac{1}{\sigma_1(\mathbf{R}_k)}(\xi + \|\mathbf{w}_k\|) \dot{\leq} \rho^{\frac{rT}{\kappa}-\frac{\delta}{2}}.$$

Also from (3.28), $\|\mathbf{s}_k\| \leq \|\mathbf{s}\| \leq \frac{1}{2}\rho^{\frac{rT}{\kappa}}$, substituting this in (3.31) gives that

$$\|\hat{\mathbf{s}}_k\| \leq \rho^{\frac{rT}{\kappa}}.$$

This implies that $\hat{\mathbf{x}}_k \in \mathbb{S}_r^k$. Thus, any integer point that satisfies the sphere constraint must also belong to the constellation and it allows for the removal of the boundary constraints of ML-based SD in (3.4).

In the following, and up until (3.41), we will work toward upper bounding $\sigma_i(\mathbf{R}_k)$ for the case of $q \in [1, \kappa-1]$, the case of $q = \kappa$ is treated separately later on. Towards this we first consider a Greedy QR decomposition (cf. [11]) of $\mathbf{M}$ resulting in a column permutation matrix $\mathbf{\Pi}$ such that $\mathbf{M\Pi} = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$ where $\mathbf{\Pi} = \mathbf{\Pi}_1 \cdots \mathbf{\Pi}_p$ and unitary matrix $\tilde{\mathbf{Q}} \triangleq \tilde{\mathbf{Q}}_1 \cdots \tilde{\mathbf{Q}}_p$ is obtained by applying p ($p \triangleq \kappa - q$) recursive steps of Greedy QR decomposition. The diagonal elements of $\tilde{\mathbf{R}}$ satisfy $\tilde{r}_{11} \geq \cdots \geq \tilde{r}_{pp}$. Let $\mathbf{M}_{|p} \in \mathbb{R}^{n\times p}$ contains the first p columns of $\mathbf{M\Pi}$. It then follows that

$$\mathbf{M}_{|p} \triangleq \mathbf{M\Pi}_p = \tilde{\mathbf{Q}}\tilde{\mathbf{R}}_p, \tag{3.32}$$

where, $\mathbf{\Pi}_p$ and $\tilde{\mathbf{R}}_p$ denote the sub matrices consisting of the first p columns of $\mathbf{\Pi}$ and $\tilde{\mathbf{R}}$ respectively. Now let $\mathbf{R}_p$ be the $p \times p$ upper triangular matrix consisting of the first p rows of $\tilde{\mathbf{R}}_p$, then we get that

$$\sigma_i(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) = \sigma_i(\mathbf{R}_p^H\mathbf{R}_p), \quad i = 1, \cdots, p.$$

For $\mathbf{R}_p^H\mathbf{R}_p$, having diagonal entries $\tilde{r}_{11}^2 \geq \cdots \geq \tilde{r}_{pp}^2$ and singular values $\sigma_1(\mathbf{R}_p^H\mathbf{R}_p) \leq \cdots \leq \sigma_p(\mathbf{R}_p^H\mathbf{R}_p)$, from [54, Theorem 2.3] we have that

$$\prod_{i=1}^{k} \tilde{r}_{ii}^2 \leq \prod_{i=1}^{k} \sigma_{p-i+1}(\mathbf{R}_p^H\mathbf{R}_p), \quad k = 1, \cdots, p.$$

Consequently, it follows that

$$\prod_{i=1}^{k} \tilde{r}_{ii}^2 \leq \prod_{i=1}^{k} \sigma_{p-i+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}), \quad k = 1, \cdots, p. \tag{3.33}$$

From [11, Lemma 4.3] regarding the Greedy QR decomposition, we have that

$$\tilde{r}_{kk}^2 \geq \frac{\sigma_{\kappa-k+1}(\mathbf{M}^H\mathbf{M})}{\kappa-k+1}, \quad k=1,\cdots,p,$$

and it follows that

$$\prod_{i=1}^{k}\tilde{r}_{ii}^2 \geq \prod_{i=1}^{k}\frac{\sigma_{\kappa-i+1}(\mathbf{M}^H\mathbf{M})}{\kappa-i+1}, \quad k=1,\cdots,p. \tag{3.34}$$

From (3.33) and (3.34) we have that

$$\prod_{i=1}^{k}\sigma_{p-i+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \geq \prod_{i=1}^{k}\frac{\sigma_{\kappa-i+1}(\mathbf{M}^H\mathbf{M})}{\kappa-i+1}, \quad k=1,\cdots,p.$$

Consequently, using [53, Theorem 4.3.15] which gives that $\sigma_{p-k+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \leq \sigma_{\kappa-k+1}(\mathbf{M}^H\mathbf{M})$, it can be shown recursively that

$$\sigma_{p-k+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \geq \sigma_{\kappa-k+1}(\mathbf{M}^H\mathbf{M})\prod_{i=1}^{k}\frac{1}{\kappa-i+1}, \quad k=1,\cdots,p. \tag{3.35}$$

Now using (3.35) along with the fact that $\sigma_{p-k+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \leq \sigma_{\kappa-k+1}(\mathbf{M}^H\mathbf{M})$, $k=1,\cdots,p$ (cf. [53, Theorem 4.3.15]), we have that

$$\sigma_{p-k+1}(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \doteq \sigma_{\kappa-k+1}(\mathbf{M}^H\mathbf{M}), \quad k=1,\cdots,p.$$

Consequently, for $k=p$ it follows that

$$\sigma_1(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \doteq \sigma_{q+1}(\mathbf{M}^H\mathbf{M}), \tag{3.36}$$

where we have used the fact that $q=\kappa-p$. Recalling that $\sigma_1(\mathbf{M}^H\mathbf{M}) \leq \cdots\sigma_m(\mathbf{M}^H\mathbf{M})$, we have that

$$\sigma_1(\mathbf{M}_{|p}^H\mathbf{M}_{|p}) \geq \sigma_i(\mathbf{M}^H\mathbf{M}) \quad \text{for} \quad i=1,\cdots,q. \tag{3.37}$$

The above inequality allows us to apply Lemma 3 from [7], which in turn gives that

$$\sigma_i(\mathbf{R}_k) \leq \left[\frac{\sigma_\kappa(\mathbf{M})}{\sigma_1(\mathbf{M}_{|p})}+1\right]\sigma_i(\mathbf{M}) \doteq \left[\frac{\sigma_\kappa(\mathbf{M})}{\sigma_{q+1}(\mathbf{M})}+1\right]\sigma_i(\mathbf{M}), \tag{3.38}$$

for $i=1,\cdots,q$, where exponential equality follows from (3.36). From (3.13) we have that

$$\sigma_i(\mathbf{M}) \doteq \rho^{\frac{-rT}{\kappa}+\frac{1}{2}(1-\mu_i)}, \quad i=1,\cdots,\kappa. \tag{3.39}$$

Furthermore (3.26) gives that

$$\sigma_i(\mathbf{M}) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\delta+\frac{1}{2}(1-\mu_i^*)} \quad \text{for} \quad i=1,\cdots,q, \tag{3.40a}$$

$$\sigma_\kappa(\mathbf{M}) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\mu_\kappa)} \leq \rho^{\frac{1}{2}-\frac{rT}{\kappa}}, \tag{3.40b}$$

$$\sigma_{q+1}(\mathbf{M}) \doteq \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\mu_{q+1})} \geq \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\delta)}. \tag{3.40c}$$

Substituting (3.40) in (3.38) gives that

$$\sigma_i(\mathbf{R}_k) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\frac{3}{2}\delta+\frac{1}{2}(1-\mu_i^*)}, \quad i=1,\cdots,q. \tag{3.41}$$

Consequently, going back to (3.29), we have that

$$\left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)}-\sqrt{k}\right]^+ \dot{\geq} \rho^{\left(\frac{rT}{\kappa}-\frac{3}{2}\delta-\frac{1}{2}(1-\mu_i^*)\right)}. \tag{3.42}$$

As a result, for $k=q$ with $q\in[1,\kappa-1]$ we have that

$$N_q \dot{\geq} \rho^{\left(\sum_{i=1}^q\left(\frac{rT}{\kappa}-\frac{1}{2}(1-\mu_i^*)\right)-\frac{3}{2}q\delta\right)} = \rho^{(\tilde{c}(r)-\frac{3}{2}q\delta)}, \tag{3.43}$$

where the last equality follows from (3.25). For the case of $q=\kappa$, from (3.29) and (3.40a) we have that

$$N_q \dot{\geq} \rho^{\sum_{i=1}^\kappa\left(\frac{rT}{\kappa}-\delta-\frac{1}{2}(1-\mu_i^*)\right)} = \rho^{(\tilde{c}(r)-\kappa\delta)}. \tag{3.44}$$

Consequently for $q\in[1,\kappa]$ we have that $N_{SD} \dot{\geq} \rho^{\tilde{c}(r)-K\delta}$ for small $\delta>0$, where $K\in\{\frac{3}{2}q,\kappa\}$.

We note that (3.26)-(3.28) jointly imply that $N_{SD} \dot{\geq} \rho^{\tilde{c}(r)-K\delta}$. For some $\delta' \triangleq K\delta+\delta_1$, where $\delta>\delta_1>0$, it follows that

$$\mathrm{P}\left(N_{SD}\geq\rho^{\tilde{c}(r)-\delta'}\right) \geq \mathrm{P}\left(\Omega_1\cap\Omega_2\cap\Omega_3\right) \doteq \mathrm{P}\left(\Omega_1\right), \tag{3.45}$$

where exponential equality follows from the independence of the events Ω_1, Ω_2 and Ω_3 and from the fact that $\mathrm{P}(\Omega_2)\doteq\rho^0$ (cf.(3.5)) and $\mathrm{P}(\Omega_3)\doteq\rho^0$. For any $r>0$ in which case the the subset of the constellation defined by Ω_3 contains an asymptotically deterministic and strictly positive fraction of the full constellation. With Ω_1 being an open set, we have that

$$-\lim_{\rho\to\infty}\frac{\mathrm{P}(\Omega_1)}{\log\rho} \leq \inf_{\boldsymbol{\mu}\in\Omega_1} I(\boldsymbol{\mu}) = I(\tilde{\boldsymbol{\mu}}) < I(\boldsymbol{\mu}^*) = d(r), \tag{3.46}$$

where $\tilde{\boldsymbol{\mu}}=\{\mu_1^*-2\delta\cdots,\mu_q^*-2\delta,0,\cdots,0\}$, where the last inequality follows from the monotonicity of the rate function $I(\boldsymbol{\mu})$ and where the last equality follows from the fact that, by definition, $I(\boldsymbol{\mu}^*)=d(r)$.

Consequently (3.45) and (3.46) along with the definition of the lower bound in (3.12b) imply that $\underline{c}(r)=\tilde{c}(r)$, for arbitrarily small $\delta>0$. The following proposition directly holds corresponding to a vanishing performance gap.

Proposition 1. *Irrespective of channel fading statistics and of the full-rate code applied, for every realization of channel there exists a channel dependent column permutation matrix* $\mathbf{\Pi}$ *such that the ML-based sphere decoder with decoding order* $\mathbf{\Pi}$ *has the complexity exponent* $c(r) = \tilde{c}(r)$ *where*

$$\tilde{c}(r) = \max_{\boldsymbol{\mu}} \sum_{i=1}^{\kappa} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_i), \frac{rT}{\kappa}\right)^{+} \tag{3.47a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d(r), \tag{3.47b}$$

$$\mu_1 \geq \cdots \geq \mu_\kappa \geq 0. \tag{3.47c}$$

This proposition also establishes that $\tilde{c}(r)$ is the tightest upper bound that can hold for any full-rate code and ML-based sphere decoder with fixed decoding order. To show the dependence of $\mathbf{\Pi}$ on $\boldsymbol{M}$, we henceforth use $\mathbf{\Pi}_M$ instead of $\mathbf{\Pi}$. Under the assumption that each column permutation matrix 'appears' with non-zero probability, then for every column permutation matrix $\mathbf{\Pi}_\kappa \in \mathbb{R}^{\kappa\times\kappa}$ we have that $\mathrm{P}\left(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa\right) \doteq \rho^0$, where probability is taken over random $\boldsymbol{M}$. Then the following theorem is a consequence of Proposition 1.

Theorem 2. *For any full-rate code and fading distribution such that* $\mathrm{P}\left(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa\right) > \epsilon \quad \forall \quad \mathbf{\Pi}_\kappa$ *for some* $\epsilon > 0$*, the SD complexity exponent is* $c(r) = \tilde{c}(r)$ *for any fixed decoding order.*

After establishing complexity exponent for the general MIMO case, in the following we consider the specific case of quasi-static MIMO channels and derive the single-letter expressions for complexity exponent for decoding over the quasi-static MIMO channels.

3.2.3 Quasi-static MIMO with fixed decoding order

Complexity upper bound

Recalling the equivalence of quasi-static MIMO channel and general MIMO from Section 2.1.1 and transitioning to the specific case of the $n_T \times n_R$ ($n_R \geq n_T$) quasi-static point-to-point MIMO channel (with T uses over channel $\mathbf{H}_C \in \mathbb{C}^{n_R\times n_T}$), where now $\boldsymbol{\mu}$ denotes the (asymptotics of) the singular values of $\mathbf{H}_C$, i.e., $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho}$, $j = 1, \cdots, n_T$, where $\mu_1 \geq \cdots \geq \mu_{n_T}$ and where $\boldsymbol{\mu} \triangleq (\mu_1, \cdots, \mu_{n_T})$ satisfy the large deviation principle with rate function $I(\boldsymbol{\mu})$, the following holds directly from Theorem 1.

Theorem 3. *The SD complexity exponent of achieving a diversity gain* $d(r)$

is upper bounded as

$$\tilde{c}_{quasi}(r) \triangleq \max_{\boldsymbol{\mu}} T\sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^+ \tag{3.48a}$$

$$s.t.\ \ I(\boldsymbol{\mu}) \leq d(r), \tag{3.48b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \tag{3.48c}$$

irrespective of the full rate lattice code.

Furthermore it is worth noting that, directly from Proposition 1, we know that irrespective of the fading statistics and of the full-rate code, there exists a fixed decoding order for which the above universal upper bound is tight.

From Theorem 3 we can now establish a universal upper bound on the complexity to achieve the DMT optimal performance $d^*(r)$ of the $n_T \times n_R$ ($n_T \leq n_R$) MIMO channel.

Theorem 4. *The SD complexity exponent of achieving the optimal DMT $d^*(r)$ is upper bounded as*

$$c(r) \leq \bar{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right), \tag{3.49}$$

where $\bar{c}(r)$ is a piecewise linear function that, for integer values of r, takes the form

$$\bar{c}(r) = \frac{T}{n_T} r(n_T - r), \quad r = 0, 1, \cdots, n_T.$$

This holds for any set of fading statistics, all DMT optimal full rate code designs, and for any decoding order policy.

The proof for this theorem is given in Appendix 3A.

Universal upper bound

Furthermore we can see from (3.48) that, regardless of the fading statistics and the corresponding $I(\boldsymbol{\mu})$, the upper bound on complexity exponent is non-decreasing in $d(r)$ and is hence maximized when $d(r)$ is itself maximized, i.e., it is maximized in the presence of DMT optimal encoding and decoding. Combining it with the fact that the corresponding maximization problem in (3.48) does not depend on the fading distribution, other than the natural fact that its tail must vanish exponentially fast. The following corollary then holds for all full-rate codes.

Corollary 4a. *For $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel, the SD complexity exponent is upper bounded as in Theorem 4 for any full rate code, all fading statistics and all decoding order policies. This takes the form*

$$c(r) \leq \bar{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right). \tag{3.50}$$

which, for integer r, simplifies to

$$\frac{T}{n_T} r(n_T - r), \quad r = 0, 1, \cdots, n_T.$$

This is the tightest upper bound that can hold for all (full-rate) codes and fading statistics (cf. Proposition 1).

SISO and SIMO systems

Corollary 4b. *For quasi-static SISO and SIMO systems, irrespective of the fading statistics and given any full-rate code of arbitrary DMT performance, the ML-based sphere decoder has zero complexity exponent.*

This corollary follows directly from the universal upper bound as a special case for $n_T = 1$.

Tightness of complexity upper bound

For quasi-static MIMO channels following theorem holds directly from Theorem 2 which describes complexity exponent for decoding over general MIMO channels.

Theorem 5. *For $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel with any full-rate code and fading distribution such that $\mathrm{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) > \epsilon \quad \forall \quad \mathbf{\Pi}_\kappa$, for some $\epsilon > 0$, the complexity exponent of the ML-based sphere decoder with any fixed decoding order is given by*

$$c_{quasi}(r) = \max_{\boldsymbol{\mu}} T \sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1 - \mu_j), \frac{r}{n_T}\right)^+ \tag{3.51a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d(r), \tag{3.51b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \tag{3.51c}$$

Complexity for DMT optimal codes

Corollary 5a. *For $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel with DMT optimal codes, the complexity exponent for ML-based SD with fixed decoding order takes the form*

$$c_{n_T \times n_R}(r) = \frac{T}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right], \tag{3.52}$$

which is a piece-wise linear function that, for integer values of multiplexing gain r, takes the form

$$c_{n_T \times n_R}(r) = \frac{T}{n_T} r(n_T - r), \text{ for } r = 0, 1, \cdots, n_T.$$

This corollary follows directly from universal upper bound and Theorem 5 and holds under the same settings as described in Theorem 5.

Example 3. *For the specific case of decoding threaded minimum delay ($T = n_T$) DMT optimal codes over $n_T \times n_R$ ($n_R \geq n_T$) MIMO, the complexity exponent takes the form*

$$c_{n_T \times n_R}(r) = r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+,$$

which again is a piece-wise linear function that, for integer values of multiplexing gain r, takes the form

$$c_{n_T \times n_R}(r) = r(n_T - r), \text{ for } r = 0, 1, \cdots, n_T.$$

The Fig 3.1 describes the complexity exponents that are needed by the 2×2 ($n_T = T = 2$) and 3×3 ($n_T = T = 3$) threaded minimum delay codes to achieve the optimal DMT.

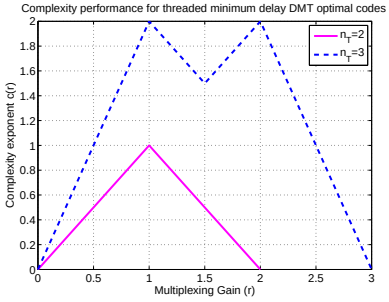


FIGURE 3.1 – The complexity exponent $c(r)$ for decoding threaded minimum delay DMT optimal codes over $n_T \times n_R$ ($n_R \geq n_T$) MIMO

V-BLAST

Corollary 5b. *For $n \times n$ i.i.d. Rayleigh MIMO channel with V-BLAST, the complexity exponent for ML-based SD with fixed decoding order takes the form*

$$c_v(r) = \frac{r \lfloor \sqrt{n-r} \rfloor}{n} + \left(\frac{r}{n} - 1 + \frac{n - r - (\lfloor \sqrt{n-r} \rfloor)^2}{2 \lfloor \sqrt{n-r} \rfloor + 1} \right)^+ . \tag{3.53}$$

The proof for this corollary is given in Appendix 3B. For example, the complexity exponent that is needed to achieve the optimal DMT for V-BLAST over $n \times n$ i.i.d. Rayleigh channel for $n = 2, 3, 4$ is shown in Fig 3.2.

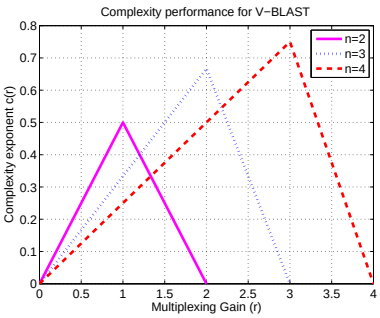


FIGURE 3.2 – The complexity exponent $c(r)$ for decoding V-BLAST over $n \times n$ i.i.d. Rayleigh channel

MISO systems

Proposition 2. *The minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required by ML-based decoding to achieve the optimal DMT of $n_T \times 1$ MISO channel, is upper bounded as*

$$c(r) \leq \overline{c}_{miso}(r) = \left[r(n_T - \lfloor n_T r \rfloor - 1) + (\lfloor n_T r \rfloor - r(n_T - 1))^+\right],$$

which is a piece-wise linear function that, for $r = 0, \frac{1}{n_T}, \cdots, 1$, takes the form

$$\overline{c}_{miso}(r) = n_T r(1 - r).$$

The proof for this proposition is given in Appendix 3C. The comparison of the MISO complexity exponent upper bound with the complexity exponent for $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel with DMT optimal codes in (3.52) reveals that

$$\overline{c}_{miso}(r) = \frac{c_{n_T \times n_R}(n_T r)}{n_T}.$$

For example, the upper bound on complexity exponent for decoding over $n_T \times 1$ MISO for $n_T = 2, 3$ is shown in Fig 3.3.

3.2.4 Complexity for dynamically changing decoding order

In the previous section we established complexity requirements for the fixed decoding order implementation of SD. It turns out that for the fixed decoding order implementation, the complexity exponent meets the universal upper bound. A natural improvement can be a sphere decoder employing dynamically changing decoding orders. In this section we wish to quantify the advantages, if any, of using dynamically changing decoding order implementations.

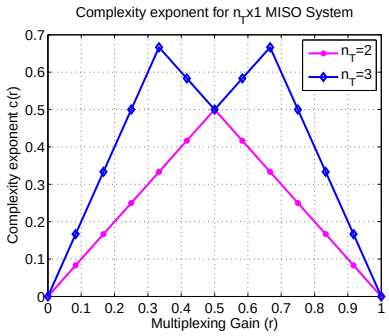


FIGURE 3.3 – The complexity exponent $c(r)$ for decoding diagonal DMT optimal codes over $n_T \times 1$ MISO

The results of analysis presented in this section establish that for any randomly picked lattice code (randomly and uniformly drawn from an ensemble of lattice designs), irrespective of the fixed or dynamically changing decoding order, the complexity exponent of the ML-based sphere decoder almost surely, in the choice of the lattice code, matches the upper bound derived in the previous section. For the analysis presented in this section we consider slightly restricted quasi-static MIMO model with i.i.d. Rayleigh fading statistics.

Theorem 6. *For the i.i.d. Rayleigh quasi-static MIMO channel, and irrespective of the fixed or dynamically changing decoding ordering policy, the complexity exponent of the ML-based sphere decoder* almost surely, *in the choice of the DMT optimal full-rate lattice code, matches the universal upper bound in Theorem 4.*

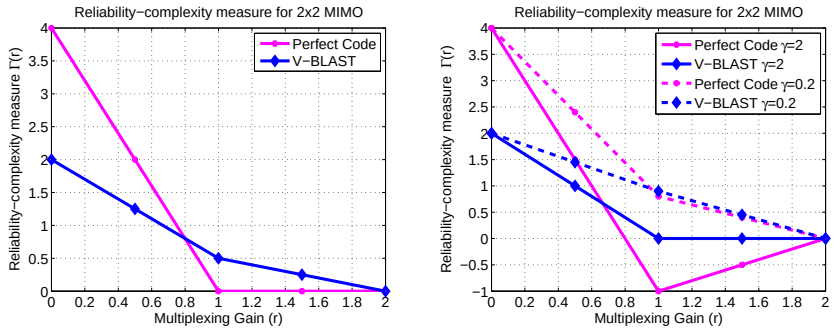
The proof for this theorem is given in Appendix 3D. At this point we want to clarify that Theorem 6 does not imply that dynamic decoding ordering policies can not reduce complexity exponent for any DMT optimal code. For any given specific DMT optimal code finding dynamically changing decoding orders that can guarantee reduction in the complexity exponent as compared to the universal upper bound in Theorem 4, remains a challenging open problem.

3.3 Reliability-Complexity Measure

Aforementioned complexity characterization provides basis to quantify the performance of specific encoders/decoders. A joint reliability-complexity measure, described as $\Gamma(r) \triangleq d(r) - \gamma c(r)$ for $\gamma \geq 0$, can be used as a meaningful unified performance metric to compare the worth of encoders and decoders. The weight factor γ compares the price of flop vs error and can be derived from the system design requirements (what is more critical - error

performance or computational resources), e.g., high γ signifies scarcity of computational resources.

Example 4. *Figure 3.4 depicting the joint reliability-complexity measure for* 2×2 *perfect codes and V-BLAST over* 2×2 *MIMO system clearly illustrates that a policy that employs perfect codes up to multiplexing gain of* $r = \frac{4}{4+\gamma}$ *and then switches to V-BLAST scheme results in a preferable reliability-complexity measure. In this dissertation all the joint reliability-complexity measure results, except for this section, unless stated clearly consider the metric with* $\gamma = 1$, *i.e.,* $\Gamma(r) \triangleq d(r) - c(r)$. *In the Fig. 3.4 (b) the diversity-complexity tradeoff of* 2×2 *perfect codes and V-BLAST is shown as dotted lines for* $\gamma = 0.2$ *and as solid lines for* $\gamma = 2$. *The figure suggests that for very low values of* γ *using perfect codes will result in uniformly better joint reliability-complexity measure.*



(a) For $\Gamma(r) = d(r) - c(r)$ (b) For $\Gamma(r) = d(r) - \gamma c(r)$

FIGURE 3.4 – Reliability-Complexity Measure

3.4 Rate-Reliability-Complexity Tradeoff

In this section we present rate-reliability-complexity tradeoff for ML-based SD. The motivation for using ML-based SD is evident from the fact that it can achieve vanishing gap to ML error performance with significantly smaller computational reserves as established in the previous section. The scenarios where ML-based SD is used in the absence of sufficient computational reserves required to achieve vanishing gap, there exists a rate-reliability-complexity tradeoff describing the optimal diversity performance that can be achieved in such scenarios. This tradeoff analysis is of particular interest for the scenarios involving transceiver designs (like software defined radios) where computational reserves might vary with time and render it insufficient to achieve a vanishing performance gap.

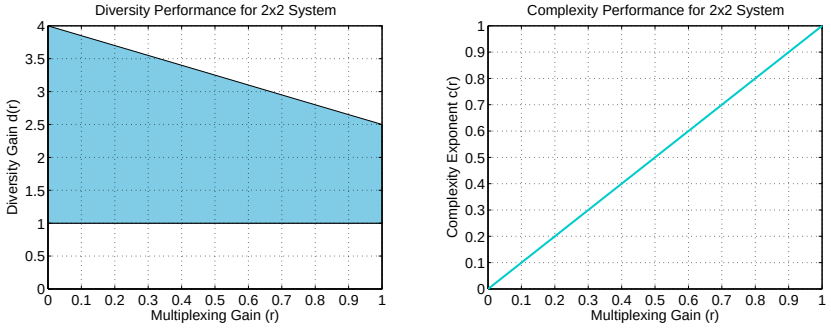
From the concise closed form expressions for complexity exponent provided in Theorem 2 and Theorem 6, it is evident that one can potentially

tradeoff diversity gain with complexity exponent, i.e., complexity exponent can decrease with reduction in diversity gain. Careful study though will show that such a tradeoff is not always successful, that reductions in $d(r)$ are not necessarily rewarded with a reduction in the required complexity, and that there are ranges of $d(r)$ for which $c(r)$ remains fixed. Such ranges can be described relatively concisely.

Example 5. *For a $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO system a diversity gain,*

$$d(r) \in \left[\sum_{j=1}^{n_T - \lfloor r \rfloor - 1} a_j, \quad \sum_{j=1}^{n_T - \lfloor r \rfloor - 1} a_j + a_{n_T - \lfloor r \rfloor} (1 - \frac{r}{n_T}) \right]$$

with a_j, $j = 0, \cdots, n_T$ denoting the coefficients of rate function $I(\boldsymbol{\mu})$, requires computational reserves denoted by $c(r) = \frac{rT}{n_T}(n_T - \lfloor r \rfloor - 1)$, e.g., for a 2×2 MIMO system for $0 \leq r \leq 1$, all diversity gains $d(r) \in [1, 4 - \frac{3r}{2}]$ shown with shaded area in Fig. 3.5 (a) require the same complexity exponent r as shown as a line in Fig. 3.5 (b). Although premature at this point, we hasten to note that any attempt to reduce complexity exponent below specified $c(r) = \frac{rT}{n_T}(n_T - \lfloor r \rfloor - 1)$ results in a reduction in achievable diversity gain, causing it to sink below $\sum_{j=1}^{n_T - \lfloor r \rfloor - 1} a_j$.



(a) Diversity Gain (b) Complexity Exponent

FIGURE 3.5 – The figure depicts the diversity gain region for which the required complexity exponent is given by $c(r)$.

The practical implication of this result is that any small reduction in the the complexity exponent might result in potentially high error performance loss. The complexity and diversity gain relationship can be succinctly described by a rate-reliability-complexity tradeoff which identifies optimal diversity gain achievable in the presence of any run-time constraint imposed due to the unavailability of enough computational resources required to achieve a vanishing gap.

Theorem 7. *For any full-rate code with ML-based diversity gain $d(r)$ and any fading distribution such that $\mathrm{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) > \epsilon \ \ \forall \ \ \mathbf{\Pi}_\kappa$, for some $\epsilon > 0$, the achievable diversity performance $d_{\mathcal{D}}(r)$ for ML-based SD with any fixed decoding order and a run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops is uniquely described by*

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{\mathcal{D}}(r, x_{\mathcal{D}})\} \ \ \forall \ \ c_{\mathcal{D}}(r) \geq 0, \tag{3.54}$$

where $d_{\mathcal{D}}(r, x_{\mathcal{D}}) \triangleq \lim_{\epsilon \to 0^+} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon)$, and where

$$\begin{aligned} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon) \triangleq & \inf I(\boldsymbol{\mu}) \\ & s.t. \ \ \sum_{i=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i) \right)^+ \geq c_{\mathcal{D}}(r) + \epsilon, \\ & 1 \geq \mu_1 \geq \cdots \geq \mu_\kappa \geq 0. \end{aligned}$$

The proof for this theorem is given in Appendix 3E.

3.4.1 Quasi-static MIMO channels

For an $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO, the following holds directly holds from Theorem 7

Theorem 8. *For any full-rate code and any fading distribution such that $\mathrm{P}(\mathbf{\Pi}_M = \mathbf{\Pi}_\kappa) > \epsilon \ \ \forall \ \ \mathbf{\Pi}_\kappa$, for some $\epsilon > 0$, the achievable diversity performance $d_{\mathcal{D}}(r)$ for ML-based SD with any fixed decoding order and a run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops is uniquely described by*

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{\mathcal{D}}(r, c_{\mathcal{D}}(r))\} \ \ \forall \ \ c_{\mathcal{D}}(r) \geq 0,$$

where $d(r)$ is the optimal diversity gain (of uninterrupted brute force ML, for the given code), where $d_{\mathcal{D}}(r, x) \triangleq \lim_{\epsilon \to 0^+} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon)$, and where

$$\begin{aligned} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon) \triangleq & \inf I(\boldsymbol{\mu}) \\ & s.t. \ \ T\sum_{j=1}^{n_T} \left(\frac{r}{n_T} - (1 - \mu_j) \right)^+ \geq c_{\mathcal{D}}(r) + \epsilon, \\ & 1 \geq \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \end{aligned}$$

Focusing on the quasi-static i.i.d Rayleigh fading $n_T \times n_R$ ($n_R \geq n_T$) MIMO channel, and on DMT optimal lattice designs (achieving the optimal DMT $d^*(r)$), the following Corollary gives a lower bound on the complexity-constrained DMT. The bound is tight for the case of DMT optimal layered lattice designs, which we will refer to in more detail later on, and which are currently the only known DMT optimal lattice designs. The corollary follows, after basic algebraic manipulation, directly from Theorem 8 and [7, Theorem 5].

Corollary 8a. *The best possible complexity-constrained DMT of ML-based decoding any DMT optimal lattice design using maximally* $N_{\max} \doteq \rho^{c_{\mathcal{D}}(r)}$ *flops, is lower bounded as* $d_{\mathcal{D}}(r) = \min\{d^*(r), d_{\mathcal{D}}(r, c_{\mathcal{D}}(r))\}$ *where*

$$d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)) = \sum_{j=1}^{K} (n_R - n_T + 2j - 1) + (n_R - n_T + 2K + 1)(\frac{c_{\mathcal{D}}(r)}{T} + 1 - \frac{(K+1)r}{n_T}),$$

where $K = \left\lfloor \frac{n_T c_{\mathcal{D}}(r)}{rT} \right\rfloor$, *irrespective of the decoding ordering policy. Furthermore in the presence of DMT optimal layered lattice designs, the above described DMT is the exact complexity constrained DMT given by the natural sphere decoding ordering.*

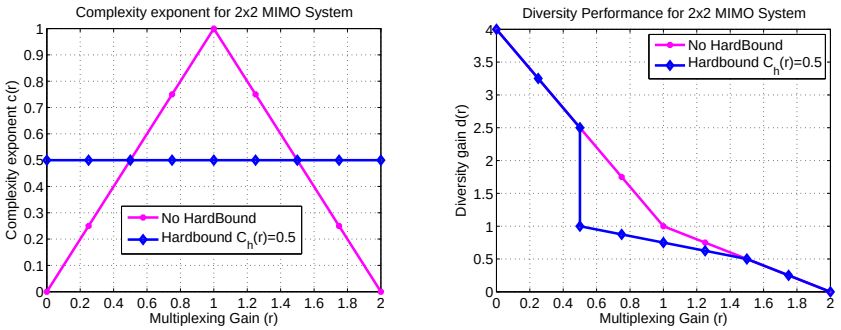
The following directly holds directly from the Corollary 8a.

Corollary 8b. *For a* $n_T \times n_R$ $(n_R \geq n_T)$ *i.i.d. Rayleigh fading channel with a DMT optimal threaded code and ML-based SD employing natural decoding order with a run-time constraint of* ρ^0 *flops, the achievable diversity gain is uniquely described by* $d_{\mathcal{D}}(r) = \min\{d^*(r), d_{\mathcal{D}}(r, c_{\mathcal{D}}(r))\}$, *where* $d^*(r)$ *is the is the optimal DMT of channel and where* $d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)) = (n_R - n_T + 1)(1 - \frac{r}{n_T})$.

In the following we present several clarifying examples corresponding to Corollary 8a under the natural decoding order.

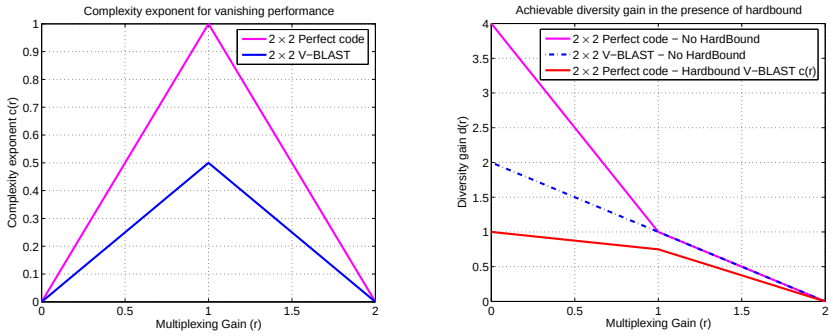
Example 6. *Figure 3.6 presents achievable diversity gain for* 2×2 *Perfect code (cf. [15]) over i.i.d. Rayleigh channel in the presence of a fixed run-time constraint. The upper solid line in Fig. 3.6 (a) describes the complexity exponent that would have been needed by the* 2×2 *perfect code to achieve the optimal DMT (upper solid line in Fig. 3.6(b)). The lower straight line in Fig. 3.6 (a) describes the complexity limitations that we are assigning to the* 2×2 *perfect code, which now, due to these complexity limitations, is achieving a much reduced DMT (blue line in Fig. 3.6(b)) wherever the complexity limit becomes active.*

Example 7. *Figure 3.7 presents achievable diversity gain for* 2×2 *Perfect code (cf. [15]) over i.i.d. Rayleigh channel in the presence of a run-time constraint that allows for computational resources required by V-BLAST to achieve its own optimal DMT. The upper line in Fig. 3.7 (a) describes the complexity exponent that would have been needed by the* 2×2 *perfect code to achieve the optimal DMT (upper line in Fig. 3.7(b)). The lower line in Fig. 3.7 (a) describes the complexity exponent that would have been needed by V-BLAST to achieve its own optimal DMT (middle line in Fig. 3.7(b)). This same lower line in Fig. 3.7 (a) also describes the complexity limitations that we are assigning to the* 2×2 *perfect code, which now, due to these complexity limitations, is achieving a much reduced DMT (lowest line in Fig. 3.7(b)).*



(a) Complexity Hardbound (b) Achievable Diversity Gain

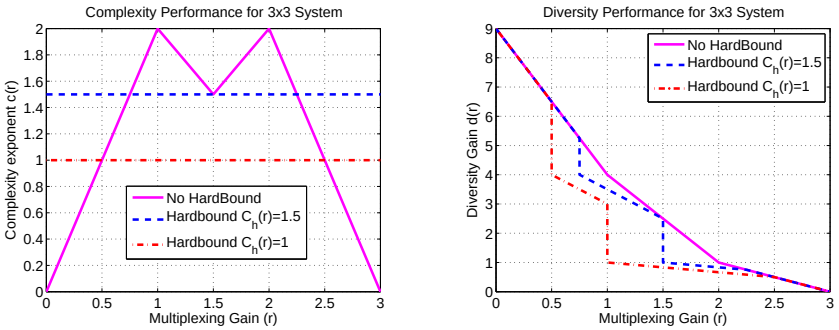
FIGURE 3.6 – Achievable diversity gain for 2×2 Perfect code in the presence of a complexity hardbound of $\rho^{\frac{1}{2}}$ flops.



(a) Variable Complexity Hardbound (b) Achievable Diversity Gain

FIGURE 3.7 – Achievable diversity gain for 2×2 Perfect code in the presence of a complexity hardbound that only allows for the computational resources required by 2×2 V-BLAST.

Example 8. *Figure 3.8 presents achievable diversity gain for* 3×3 *Perfect code (cf. [15]) over i.i.d. Rayleigh channel again in the presence of fixed run-time constraints. The upper solid line in Fig. 3.8 (a) describes the complexity exponent that would have been needed by the* 3×3 *perfect code to achieve the optimal DMT (upper solid line in Fig. 3.8(b)). The middle blue line and lowest red line in Fig. 3.8 (a) describe the complexity limitations that we are assigning to the* 3×3 *perfect code, which now, due to these complexity limitations, is achieving a much reduced DMT (middle blue line and lowest red line respectively in Fig. 3.8(b)) wherever the complexity limits become active.*



(a) Complexity Hardbound (b) Achievable Diversity Gain

FIGURE 3.8 – Achievable diversity gain for 3×3 Perfect code in the presence of a complexity hardbound of $\rho^{\frac{3}{2}}$ flops and ρ flops.

Discussion on rate-reliability-complexity tradeoff results

As we see in Fig. 3.6, Fig. 3.7 and Fig. 3.8 that the effect of run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ on the performance of DMT optimal codes is more severe at lower multiplexing gains, this phenomenon can be explained by understanding the behavior of the decoding policies in different multiplexing gain regions.

In high multiplexing gain region, say close to maximum multiplexing gain (for $n_T - 1 \leq r \leq n_T$), the optimizing vector $\boldsymbol{\mu}^*$ for optimal DMT is of the form

$$\mu_1^* = n_T - r, \ \mu_2^* = \cdots = \mu_{n_T}^* = 0.$$

and results in the complexity exponent $c(r)$ required to achieve optimal DMT. Any small reduction in complexity exponent $c(r)$ by some positive constant $\epsilon > 0$ will result in a new optimal solution $\tilde{\mu}_1^* = n_T - r - \epsilon, \ \tilde{\mu}_2^* = \cdots = \tilde{\mu}_{n_T}^* = 0$ and the achievable diversity gain is reduced by $(n_R - n_T + 1)\epsilon$. However, in the low multiplexing gain region, for say multiplexing gain close to zero (for $0 \leq r \leq 1$), the optimizing vector for optimal DMT is of the form

$$\mu_1^* = \cdots = \mu_{n_T-1}^* = 1, \ \mu_{n_T}^* = 1 - r.$$

Again the identical small reduction in complexity exponent by $\epsilon > 0$ will result in a new optimal solution $\tilde{\mu}_1^* = \cdots = \tilde{\mu}_{n_T-2}^* = 1, \ \tilde{\mu}_{n_T-1}^* = 1 - \epsilon, \ \mu_{n_T}^* = 0$, that corresponds to a reduction in achievable diversity gain by $(n_R + n_T + 1)(1 - r) + (n_R + n_T - 1)\epsilon$, which is much larger than the reduction at higher multiplexing gains.

The result of Theorem 6 allows for a meaningful rate-reliability-complexity tradeoff that holds, almost surely in the random choice of the DMT optimal full-rate lattice design, and which holds irrespective of the decoding ordering policy. This tradeoff is defined in the following

Theorem 9. *With probability one in the random choice of a full-rate lattice design, the achievable diversity performance $d_{\mathcal{D}}(r)$ for ML-based SD with a run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops, is uniquely described by*

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{\mathcal{D}}(r,x)\} \quad \forall \quad c_{\mathcal{D}}(r) \geq 0, \tag{3.56}$$

where $d(r)$ is the optimal diversity gain (of uninterrupted brute force ML, for the given code), where $d_{\mathcal{D}}(r,x) \triangleq \lim_{\epsilon \to 0^+} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)+\epsilon)$, and where

$$\begin{aligned} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)+\epsilon) \triangleq & \inf I(\boldsymbol{\mu}) \\ & s.t. \;\; T\sum_{j=1}^{n_T}\left(\frac{r}{n_T} - (1-\mu_j)\right)^+ \geq c_{\mathcal{D}}(r)+\epsilon, \\ & 1 \geq \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \end{aligned}$$

and where the above holds irrespective of the fixed or dynamically changing decoding order.

The proof follows directly from the footsteps of the proof of Theorem 7 and the lower bound proof for Theorem 6.

Appendix 3A : Proof of Theorem 4

For DMT optimal performance we have that

$$\begin{aligned} \tilde{c}_{quasi}(r) = \max_{\boldsymbol{\mu}} & \; T\sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^+ \\ \text{s.t.} & \;\; \sum_{j=1}^{n_T}(1-\mu_j)^+ \geq r, \\ & \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \end{aligned}$$

For any $r \in [k, \;\; k+1)$, solution to $\sum_{j=1}^{n_T}\mu_j \leq n_T - r$ results in a $\boldsymbol{\mu}^*$ given by

$$\mu_1^* = \cdots = \mu_{n_T-k-1}^* = 1,$$

$$\mu_{n_T-k}^* = k+1-r,$$

$$\mu_{n_T-k+1}^* = \cdots = \mu_m^* = 0.$$

A sample optimization problem solution for $0 \leq r \leq 1$ and $n_T = 2$ is shown in Fig. 3.9. Substituting the above solution back into the objective function of the maximization problem of (3.48) proves Theorem 4. $\square$

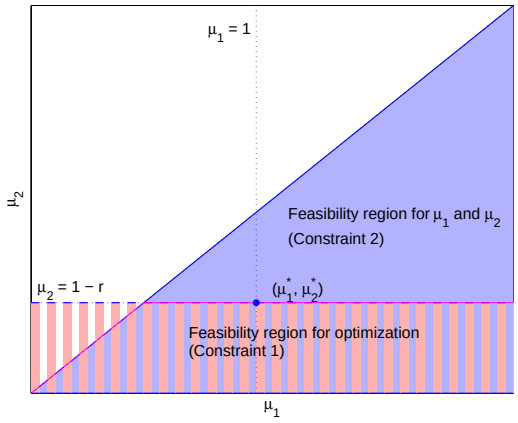


FIGURE 3.9 – Sample complexity optimization problem solution for $n_T = 2$ MIMO system.

Appendix 3B : Proof of Corollary 5b

For $n \times n$ i.i.d. Rayleigh MIMO channel with V-BLAST, the complexity exponent for ML-based SD with fixed decoding order takes the form

$$\begin{aligned} c_v(r) = \max_{\boldsymbol{\mu}} T \sum_{j=1}^{n} \min\left(\frac{r}{n} - (1-\mu_j), \frac{r}{n}\right)^+ \\ \text{s.t.} \quad \sum_{j=1}^{n} (2j-1)\mu_j \leq n - r, \\ \mu_1 \geq \cdots \geq \mu_n \geq 0. \end{aligned}$$

For any r, solution to $\sum_{j=1}^{n}(2j-1)\mu_j \leq n - r$ results in a $\boldsymbol{\mu}^*$ given by

$$\mu_1^* = \cdots = \mu_p^* = 1,$$

$$\mu_{p+1}^* = \frac{n - r - p^2}{2p+1},$$

$$\mu_{p+2}^* = \cdots = \mu_n^* = 0,$$

where $p \triangleq \left\lfloor \sqrt{n-r} \right\rfloor$. Substituting the above solution back into the objective function of the maximization problem of (3.48) proves Corollary 5b. □

Appendix 3C : Proof of Proposition 2

Let us consider a $n_T \times 1$ quasi-static MISO channel $\mathbf{h}^T = [h_1 \ h_2 \ \cdots \ h_{n_T}]$ with a DMT optimal code $\theta \mathbf{X}_{miso}$, where $\theta^2 = \rho^{1-r}$ and where

$$\mathbf{X}_{miso} = \begin{bmatrix} \ell_0 & 0 & \cdots & 0 \\ 0 & \sigma(\ell_0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^{n_T-1}(\ell_0) \end{bmatrix}, \tag{3.58}$$

where $\ell_0 \in \mathcal{A}_{QAM}(\beta_1, \cdots, \beta_{n_T})$ (see below (3.59)) and where σ is the generator of the cyclic Galois group $Gal(\mathbb{L}/\mathbb{F})$ with $\mathbb{L}$ being a degree-n_T cyclic Galois extension field of $\mathbb{F} = \mathbb{Q}(i)$, where $\mathbb{Q}(i) = \{a + ib | a, b, \in \mathbb{Q}\}$. $\mathbb{Q}$ is the set of all rational numbers. Let $\mathcal{O}_{\mathbb{F}}$ and $\mathcal{O}_{\mathbb{L}}$ denote the ring of algebraic integers in $\mathbb{F}$ and $\mathbb{L}$, respectively. Let $\{\beta_1, \cdots, \beta_{n_T}\}$ be an integral basis for $\mathcal{O}_{\mathbb{L}}/\mathcal{O}_{\mathbb{F}}$ and for M even, let $\mathcal{A}_{QAM}$ denote the M^2-QAM constellation given by $\mathcal{A}_{QAM} = \{a + ib | |a|, |b| \leq M-1, a, b, \text{odd}\}$ and

$$\mathcal{A}_{QAM}(\beta_1, \cdots, \beta_{n_T}) = \left\{ \sum_i a_i \beta_i | a_i \in \mathcal{A}_{QAM} \right\}. \tag{3.59}$$

The received signal vector $\mathbf{y}^T = [y_1 \ y_2 \ \cdots \ y_{n_T}]$ for $n_T \times 1$ quasi-static MISO channel with the space-time code $\theta \mathbf{X}_{miso}$ is given by

$$\mathbf{y}^T = \theta \mathbf{h}^T \mathbf{X}_{miso} + \mathbf{w}^T,$$

where $\mathbf{w}$ is the noise vector. The equivalent system model for this MISO system takes the form

$$\begin{aligned} \mathbf{y} =& \theta \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n_T} \end{bmatrix} \begin{bmatrix} \ell_0 \\ \sigma(\ell_0) \\ \cdots \\ \sigma^{n_T-1}(\ell_0) \end{bmatrix} + \mathbf{w}, \\ =& \theta \mathbf{H}_{miso} \mathbf{G}_{miso} \mathbf{s} + \mathbf{w}, \end{aligned}$$

where the code generator matrix $\mathbf{G}_{miso} \in \mathbb{C}^{n_T \times n_T}$ is such that $\mathbf{G}_{miso}\mathbf{s} = [\ell_0 \sigma(\ell_0) \cdots \sigma^{n_T-1}(\ell_0)]^T$ and where

$$\mathbf{H}_{miso} = \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n_T} \end{bmatrix}.$$

For i.i.d. Rayleigh fading channel we define $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}_{miso}^H \mathbf{H}_{miso})}{\log \rho}$, $j = 1, \cdots, n_T$. Following the footsteps of the complexity analysis presented in

Section 3.2.2 the upper bound complexity exponent takes the form

$$\overline{c}_{miso}(r) = \max_{\boldsymbol{\mu}} \sum_{j=1}^{n_T} \min\left(r-(1-\mu_j), r\right)^+ \tag{3.60a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n_T} \mu_j \leq n_T(1-r), \tag{3.60b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \tag{3.60c}$$

The solution to this optimization problem is given by

$$\overline{c}_{miso}(r) = \left[r(n_T - \lfloor n_T r \rfloor - 1) + \left(\lfloor n_T r \rfloor - r(n_T-1)\right)^+\right], \tag{3.61}$$

which is a piece-wise linear function that, for $r = 0, \frac{1}{n_T}, \cdots, 1$, takes the form

$$\overline{c}_{miso}(r) = n_T r(1-r).$$

This proves Proposition 2. □

Appendix 3D : Proof of Theorem 6

The general $n_T \times n_R$ point-to-point quasi-static MIMO channel is given by

$$\mathbf{Y}_C = \sqrt{\rho}\mathbf{H}_C\mathbf{X}_C + \mathbf{W}_C, \tag{3.62}$$

where $\mathbf{X}_C \in \mathbb{C}^{n_T \times T}$, $\mathbf{Y}_C \in \mathbb{C}^{n_R \times T}$ and $\mathbf{W}_C \in \mathbb{C}^{n_R \times T}$ represent the transmitted, received and noise signals over a period of T time slots, where the fading matrix $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$ is assumed to be random, with i.i.d. Rayleigh fading statistics. After vectorization the real values representation of (3.62) takes the form

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{3.63}$$

where

$$\mathbf{H} = \mathbf{I}_T \otimes \mathbf{H}_R \in \mathbb{R}^{2n_R T \times 2n_T T}, \mathbf{H}_R = \begin{bmatrix} Re\{\mathbf{H}_C\} & -Im\{\mathbf{H}_C\} \\ Im\mathbf{H}_C\} & Re\{\mathbf{H}_C\} \end{bmatrix},$$

$$\mathbf{x} = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_T T} \text{ with } \mathbf{x}_t = [Re\{\mathbf{X}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{X}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{w} = (\mathbf{w}_1^{\mathrm{T}}, \cdots, \mathbf{w}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R T} \text{ with } \mathbf{w}_t = [Re\{\mathbf{W}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{W}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R T} \text{ with } \mathbf{y}_t = [Re\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

for $t = 1, \cdots, T$, where $\mathbf{X}_{t,C}$, $\mathbf{Y}_{t,C}$ and $\mathbf{W}_{t,C}$ are t-th column of $\mathbf{X}_C$, $\mathbf{W}_C$ and $\mathbf{Y}_C$ respectively.

For the lattice codes described in the previous chapter, the codewords take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}} \mathbf{G}\mathbf{s}, \quad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}} \mathcal{R}, \tag{3.64}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}^{'}$ that preserves the code, and contains the all zero vector $\mathbf{0}$ and for full-rate code $\kappa = 2n_T T$. For simplicity we consider $\mathcal{R} \triangleq [-1,\ 1]^\kappa$ to be a hypercube in $\mathbb{R}^\kappa$, although this could be relaxed. Combining (3.63) and (3.64) yields the equivalent system model

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \tag{3.65a}$$

$$\text{where} \quad \mathbf{M} \triangleq \rho^{\frac{1}{2} - \frac{r}{2n_T}} \mathbf{H}\mathbf{G} \in \mathbb{R}^{n \times \kappa}. \tag{3.65b}$$

Let $\mathbf{Q}\mathbf{R} = \mathbf{M}$ be the thin QR factorization of the modified code-channel matrix $\mathbf{M}$ and $\mathbf{r} \triangleq \mathbf{Q}^H \mathbf{y}$, then the equivalent system model in (3.65a) is represented by

$$\mathbf{r} = \mathbf{R}\mathbf{s} + \mathbf{Q}^H \mathbf{w},$$

and the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa} \left\| \mathbf{r} - \mathbf{R}\hat{\mathbf{s}} \right\|^2, \tag{3.66}$$

which is then solved by the sphere decoder which recursively enumerates all candidate vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$ within a given sphere of radius $\xi > 0$.

We define $\mu_j \triangleq -\frac{\log \sigma_i(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho},\ j = 1, \cdots, n_T$, where $\mu_1 \geq \cdots \geq \mu_{n_T}$. For full-rate code it follows that

$$\begin{aligned} \sigma_i(\mathbf{M}) &\leq \rho^{\frac{1}{2} - \frac{r}{2n_T}} \sigma_{\max}(\mathbf{G}) \sigma_i(\mathbf{H}), \\ &\doteq \rho^{-\frac{r}{2n_T} + \frac{1}{2}(1 - \mu_{l_{2T}(i)})}, \quad i = 1, \cdots, 2n_T T, \end{aligned} \tag{3.67}$$

where the asymptotic equality is due to the fact that $\sigma_{\max}(\mathbf{G}) \doteq \rho^0$ and where $l_{2T}(i) \triangleq \left\lceil \frac{i}{2T} \right\rceil$.

The upper bound $\tilde{c}(r)$ holds irrespective of the fixed or dynamically changing decoding order. For quasi-static MIMO the upper bound in (3.24) takes the form

$$\tilde{c}(r) = T \sum_{j=1}^{n_T} \min \left(\frac{r}{n_T} - (1 - \mu_j^*) \right)^+, \tag{3.68}$$

where $\boldsymbol{\mu}^* = (\mu_1^*, \cdots, \mu_{n_T}^*)$ is one of the maximizing vectors such that $I(\boldsymbol{\mu}^*) = d(r)$.

In order to establish a lower bound that matches the upper bound in (3.68) irrespective of the decoding ordering policy, we define the following lemma.

Lemma 1. *Let $\mathbf{G}_p$ be the matrix consisting of any $2pT$ columns of the (fixed but) randomly chosen generator matrix $\mathbf{G}$ where entries of $\mathbf{G}$ are independently chosen from a continuous distribution over the real numbers. Let $\mathbf{V}_p$ be the matrix consisting of the $2p$ columns of the unitary matrix $\mathbf{V}$ corresponding to the $2p$ largest singular values of $\mathbf{H}_R$, where $\mathbf{V}$ is such that $\mathbf{H}_R = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where*

$$\mathbf{\Sigma} \triangleq \mathrm{diag}\{\sigma_1(\mathbf{H}_R), \cdots, \sigma_{2n_T}(\mathbf{H}_R)\}$$

and $\mathbf{V}\mathbf{V}^H = \mathbf{I}$. Then almost surely, in the choice of $\mathbf{G}$ we have that

$$rank((I_T \otimes \mathbf{V}_p^H)\mathbf{G}_p) = 2pT. \tag{3.69}$$

It then follows that for any fixed or dynamically changing column permutation matrix $\mathbf{\Pi}$, and for $\mathbf{G}_{|p}$ denoting the first $2pT$ columns of the matrix $\mathbf{G}\mathbf{\Pi}$, it holds that

$$\mathrm{P}\left(\{\sigma_1\left((\mathbf{I}_T \otimes \mathbf{V}_p^H)\mathbf{G}_{|p}\right) \geq u\}\right) \doteq \rho^0, \quad u > 0. \tag{3.70}$$

Proof. For $\mathbf{v}_i^H,\ i = 1, \cdots, 2pT$ denoting the $2pT$ linearly independent rows of the matrix $I_T \otimes \mathbf{V}_p^H$ with rank $2pT$ and for $\mathbf{g}_i,\ i = 1, \cdots, \kappa$ denoting the κ linearly independent columns of the full rank matrix $\mathbf{G}$, then

$$(I_T \otimes \mathbf{V}_p^H)\mathbf{G} = \begin{bmatrix} \mathbf{v}_1^H\mathbf{g}_1 & \cdots & \mathbf{v}_1^H\mathbf{g}_\kappa \\ \vdots & \ddots & \vdots \\ \mathbf{v}_{2pT}^H\mathbf{g}_1 & \cdots & \mathbf{v}_{2pT}^H\mathbf{g}_\kappa \end{bmatrix}.$$

Since $\mathbf{v}_i^H,\ i = 1, \cdots, 2pT$ are fixed and linearly independent, any $2pT$ columns of $(I_T \otimes \mathbf{V}_p^H)\mathbf{G}$ are linearly independent ($rank((I_T \otimes \mathbf{V}_p^H)\mathbf{G}_p) = 2pT$), with probability one. This in turn implies that, given such linear lattice codes that are drawn with probability one, there exists a unitary matrix $\mathbf{V}_p$ such that irrespective of the fixed or dynamically changing column permutation matrix $\mathbf{\Pi}$, it is the case that

$$\mathrm{rank}\big((\mathbf{I}_T \otimes (\mathbf{V}_p)^H)\mathbf{G}_{|p}\big) = 2pT.$$

Consequently, it follows that

$$\sigma_1\left((\mathbf{I}_T \otimes (\mathbf{V}_p)^H)\mathbf{G}_{|p}\right) > 0.$$

Now, by the continuity of singular values [53], it follows for sufficiently small $u > 0$ that

$$\mathrm{P}\left(\{\sigma_1\left((\mathbf{I}_T \otimes \mathbf{V}_p^H)\mathbf{G}_{|p}\right) \geq u\}\right) > 0,$$

which implies[3] that

$$\mathrm{P}\left(\{\sigma_1\left((\mathbf{I}_T \otimes \mathbf{V}_p^H)\mathbf{G}_{|p}\right) \geq u\}\right) \doteq \rho^0$$

as $(\mathbf{I}_T \otimes (\mathbf{V}_p)^H)\mathbf{G}_{|p}$ is independent of ρ. This proves Lemma 1. □

To tighten the lower bound and show that $\underline{c}(r) = \tilde{c}(r)$, we will use Lemma 1 to show that, irrespective of the ordering policy, the sphere decoder visits a total number of nodes that approaches $\rho^{\tilde{c}(r)}$, and does so with probability that is large compared to the ML error probability. We consider the codes that satisfy (3.69) and which appear almost surely as shown in Lemma 1.

Towards this we let $q \in [1, \kappa]$ be the largest integer for which $\frac{r}{n_T} - (1 - \mu_q^*) > 0$, in which case (3.68) takes the form

$$\tilde{c}(r) = T \sum_{j=1}^{q} \left(\frac{r}{n_T} - (1 - \mu_j^*) \right). \tag{3.71}$$

We quickly note that without loss of generality we can assume that $q \geq 1$ as otherwise $\bar{c}(r) = c(r) = 0$. Consequently it is the case that $\mu_j^* > 0$ for $j = 1, \cdots, q$.

We proceed to define four events Ω_4, Ω_5, Ω_6 and Ω_7 which we will prove to be jointly sufficient so that that at layer $k = 2qT$, for some $q \in [1, n_T]$ the total number of nodes visited by sphere decoder is close to $\rho^{\tilde{c}(r)}$. These events are given by

$$\begin{aligned} \Omega_4 \triangleq \{&\mu_j^* - 2\delta < \mu_j < \mu_j^* - \delta, j = 1, \cdots, q \\ &0 < \mu_j < \delta, j = q+1, \cdots, n_T\}, \end{aligned} \tag{3.72}$$

for a given small $\delta > 0$,

$$\Omega_5 \triangleq \{\sigma_1\left((\mathbf{I}_T \otimes \mathbf{V}_p^H)\mathbf{G}_{|p}\right) \geq u\}, \tag{3.73}$$

for some given $u > 0$ independent of ρ, where for $p \triangleq n_T - q$,

$$\Omega_6 \triangleq \left\{\|\mathbf{w}\|^2 < \xi^2\right\}, \tag{3.74}$$

$$\Omega_7 \triangleq \left\{\|\mathbf{s}\| < \frac{1}{2}\rho^{\frac{rT}{\kappa}}\right\}. \tag{3.75}$$

Note also that by choosing δ sufficiently small, and using the fact that $\mu_j^* > 0$ for $j = 1, \cdots, q$, we may without loss of generality assume that Ω_4 implies that $\mu_j > 0$ for all $j = 1, \cdots, n_T$.

3. In light of the fact that event $\mathbf{V}_p$ has zero measure, what the continuity of eigenvalues guarantees is that we can construct a neighborhood of matrices around $\mathbf{V}_p$ which are full rank, and which have a non zero measure. We also note that the matrices $\mathbf{V}_p$ can be created recursively, starting from a single matrix $\mathbf{V}_{2n_T}$.

Similar to Section 3.2.2 it can be shown that in the presence of events Ω_4, Ω_6 and Ω_7 we can remove the boundary constraints of ML-based SD in (3.4), which allows us to lower bound on the number of nodes visited at layer k as (cf. [7, Lemma 1])

$$N_k \geq \prod_{i=1}^{k} \left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k} \right]^+. \tag{3.76}$$

In the following, and up until (3.83), we will work towards upper bounding $\sigma_i(\mathbf{R}_k)$ so that we can then lower bound N_k. Towards this let

$$\mathbf{M}_{|p} \triangleq \rho^{\frac{1}{2}-\frac{r}{2n_T}} \mathbf{H}\mathbf{G}_{|p} \in \mathbb{R}^{2n_R T \times 2pT}$$

contain the first $2pT$ columns of $\mathbf{M}$, and note that

$$(\mathbf{M}_{|p})^H \mathbf{M}_{|p} = \rho^{1-\frac{r}{n_T}} \mathbf{G}_{|p}^H \mathbf{H}^H \mathbf{H} \mathbf{G}_{|p},$$

and by substituting for $\mathbf{H}$ we get

$$(\mathbf{M}_{|p})^H \mathbf{M}_{|p} = \rho^{1-\frac{r}{n_T}} \mathbf{G}_{|p}^H (\mathbf{I}_T \otimes \mathbf{H}_R^H \mathbf{H}_R) \mathbf{G}_{|p}. \tag{3.77}$$

Since

$$\begin{aligned}
\mathbf{H}_R^H \mathbf{H}_R &= \mathbf{V}(\mathrm{diag}\{\sigma_1(\mathbf{H}_R^H \mathbf{H}_R), \cdots, \sigma_{2n_T}(\mathbf{H}_R^H \mathbf{H}_R)\})\mathbf{V}^H, \\
&= \mathbf{V}(\mathrm{diag}\{\sigma_1(\mathbf{H}_R^H \mathbf{H}_R), \cdots, \sigma_{2n_T}(\mathbf{H}_R^H \mathbf{H}_R)\} \\
&\quad - \sigma_{(2q+1)}(\mathbf{H}_R^H \mathbf{H}_R)\mathbf{V}(\mathrm{diag}\{\underbrace{0, \cdots, 0}_{2q}, \underbrace{1, \cdots, 1}_{2p}\})\mathbf{V}^H \\
&\quad + \sigma_{(2q+1)}(\mathbf{H}_R^H \mathbf{H}_R)\mathbf{V}(\mathrm{diag}\{\underbrace{0, \cdots, 0}_{2q}, \underbrace{1, \cdots, 1}_{2p}\})\mathbf{V}^H,
\end{aligned}$$

hence we have that

$$\begin{aligned}
\mathbf{H}_R^H \mathbf{H}_R &\succeq \sigma_{(2q+1)}(\mathbf{H}_R^H \mathbf{H}_R)\mathbf{V}(\mathrm{diag}\{\underbrace{0, \cdots, 0}_{2q}, \underbrace{1, \cdots, 1}_{2p}\})\mathbf{V}^H, \\
&= \sigma_{(2q+1)}(\mathbf{H}_R^H \mathbf{H}_R)\mathbf{V}_p \mathbf{V}_p^H,
\end{aligned} \tag{3.78}$$

where $\mathbf{A} \succeq \mathbf{B}$ denotes that $\mathbf{A} - \mathbf{B}$ is positive-semidefinite. Since $\sigma_i(\mathbf{H}^H\mathbf{H}) \in \mathbb{R}$ and since the Kronecker product induces singular value multiplicity, from (3.78) it follows that

$$(\mathbf{M}_{|p})^H \mathbf{M}_{|p} \succeq \rho^{1-\frac{r}{n_T}} \sigma_{(2q+1)}(\mathbf{H}_R \mathbf{H}_R^H) \mathbf{G}_{|p}^H (\mathbf{I}_T \otimes \mathbf{V}_p \mathbf{V}_p^H) \mathbf{G}_{|p}.$$

With respect to the smallest singular value of $(\mathbf{M}_{|p})^H \mathbf{M}_{|p}$ we have

$$\sigma_1((\mathbf{M}_{|p})^H \mathbf{M}_{|p}) \geq \rho^{1-\frac{r}{n_T}} \sigma_{(2q+1)}(\mathbf{H}_R \mathbf{H}_R^H) \sigma_1 \left(\mathbf{G}_{|p}^H (\mathbf{I}_T \otimes \mathbf{V}_p \mathbf{V}_p^H) \mathbf{G}_{|p} \right),$$

and consequently, given that $\mathbf{H}_R \in \Omega_5$, we have that

$$\begin{aligned}\sigma_1(\mathbf{M}_{|p}) &\geq u\rho^{\frac{1}{2}-\frac{r}{2n_T}}\sigma_{l_2(2q+1)}(\mathbf{H}_C),\\ &\doteq \rho^{\frac{1}{2}-\frac{r}{2n_T}-\frac{\mu_{q+1}}{2}},\\ &\geq \rho^{\frac{1}{2}-\frac{r}{2n_T}-\frac{\delta}{2}}.\end{aligned} \tag{3.79}$$

Furthermore (3.72) and (3.67) give that for $i = 1, \cdots, 2qT$ then

$$\sigma_i(\mathbf{M}) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})}. \tag{3.80}$$

Given that $\mu^*_j > 0$ for $j = 1, \cdots, q$, then for sufficiently small δ and for $i = 1, \cdots, 2qT$, we have that

$$-\frac{rT}{\kappa}+\frac{1}{2}(1-\delta)^+ \geq -\frac{rT}{\kappa}+\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+,$$

which means that for sufficiently small δ, a comparison of (3.79) and (3.80) yields

$$\sigma_i(\mathbf{M}) < \sigma_1(\mathbf{M}_{|p}), \text{ for } i = 1, \cdots, 2qT.$$

The above inequality allows us to apply Lemma 3 in [7], which in turn gives that

$$\sigma_i(\mathbf{R}_k) \leq \left[\frac{\sigma_\kappa(\mathbf{M})}{\sigma_1(\mathbf{M}_{|p})}+1\right]\sigma_i(\mathbf{M}), \text{ for } i = 1, \cdots, 2qT. \tag{3.81}$$

Furthermore (3.26), (3.81) and (3.67) give that

$$\sigma_i(\mathbf{M}) \dot{\leq} \rho^{-\frac{r}{2n_T}+\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})} \text{ for } i = 1, \cdots, 2qT, \tag{3.82a}$$

$$\sigma_\kappa(\mathbf{M}) \dot{\leq} \rho^{-\frac{r}{2n_T}+\frac{1}{2}(1-\mu^*_{l_{2T}(2n_TT)})} \leq \rho^{\frac{1}{2}-\frac{rT}{\kappa}}, \tag{3.82b}$$

$$\sigma_1(\mathbf{M}_{|p}) \dot{\geq} \rho^{-\frac{r}{2n_T}+\frac{1}{2}(1-\mu^*_{l_{2T}(2qT+1)})} \geq \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\delta)}. \tag{3.82c}$$

Substituting (3.82) in (3.81) gives that

$$\sigma_i(\mathbf{R}_k) \dot{\leq} \rho^{-\frac{r}{2n_T}+\frac{3}{2}\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})}, \quad i = 1, \cdots, q. \tag{3.83}$$

Consequently, going back to (3.76), we have that

$$\left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)}-\sqrt{k}\right]^+ \dot{\geq} \rho^{\left(\frac{r}{2n_T}-\frac{3}{2}\delta-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})\right)}. \tag{3.84}$$

As a result, for $k = 2qT$ with $q \in [1, n_T - 1]$ we have that

$$N_{2qT} \dot{\geq} \rho^{\left(\sum_{i=1}^{2qT}\left(\frac{r}{2n_T}-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})\right)-3qT\delta\right)} = \rho^{(\tilde{c}(r)-3qT\delta)},$$

where the last equality follows from (3.71).

For the case of $q = n_T$, it can be shown that

$$N_{2qT} \dot{\geq} \rho^{T\sum_{j=1}^{n_T}\left(\frac{r}{n_T}-2\delta-(1-\mu_j^*)\right)} = \rho^{(\tilde{c}(r)-2n_TT\delta)}.$$

Consequently for $q \in [1, n_T]$ we have that $N_{SD} \dot{\geq} \rho^{\tilde{c}(r)-K\delta}$ for small $\delta > 0$, where $K \in \{3qT, 2n_TT\}$. We note that (3.72)-(3.75) jointly imply that $N_{SD} \dot{\geq} \rho^{\tilde{c}(r)-K\delta}$. For some $\delta' \triangleq K\delta + \delta_1$, where $\delta > \delta_1 > 0$, it follows that

$$\mathrm{P}\left(N_{SD} \geq \rho^{\tilde{c}(r)-\delta'}\right) \geq \mathrm{P}\left(\Omega_4 \cap \Omega_5 \cap \Omega_6 \cap \Omega_7\right) \doteq \mathrm{P}\left(\Omega_4\right), \tag{3.85}$$

where exponential equality follows from the independence of the events Ω_4, Ω_5, Ω_6 and Ω_7 for i.i.d. Rayleigh fading statistics (cf. [55]) and from the fact that $\mathrm{P}\left(\Omega_5\right) \doteq \rho^0$ (cf. (3.70)), $\mathrm{P}\left(\Omega_6\right) \doteq \rho^0$ (cf.(3.5)) and $\mathrm{P}\left(\Omega_7\right) \doteq \rho^0$.With Ω_4 being an open set, we have that

$$-\lim_{\rho\to\infty}\frac{\mathrm{P}\left(\Omega_4\right)}{\log\rho} \leq \inf_{\boldsymbol{\mu}\in\Omega_4} I(\boldsymbol{\mu}) = I(\tilde{\boldsymbol{\mu}}) < I(\boldsymbol{\mu}^*) = d(r), \tag{3.86}$$

where $\tilde{\boldsymbol{\mu}} = \{\mu_1^* - 2\delta \cdots, \mu_q^* - 2\delta, 0, \cdots, 0\}$, where the last inequality follows from the monotonicity of the rate function $I(\boldsymbol{\mu})$ and where last equality follows from the fact that, by definition, $I(\boldsymbol{\mu}^*) = d(r)$. Consequently (3.85) and (3.86) along with the definition of the lower bound in (3.12b) imply that $\underline{c}(r) = \tilde{c}(r)$, for arbitrarily small $\delta > 0$. This proves Theorem 6. □

Appendix 3E : Proof of Theorem 7

For a ML-based SD with run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops, (3.10) and (3.11) tell us that for some $x_{\mathcal{D}} > c_{\mathcal{D}}(r)$, it is then the case that

$$P_e \dot{\leq} \max\{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right), \mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)\}.$$

Combining this with the trivial lower bound

$$\max\{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right), \mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)\} \leq P_e,$$

gives that

$$P_e \doteq \max\{\mathrm{P}\left(\hat{\mathbf{s}}_{ML} \neq \mathbf{s}\right), \mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)\}. \tag{3.87}$$

If we can derive the precise optimum SNR exponent for $\mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)$ then we have the DMT performance of SD implemented with run-time constraint $\rho^{c_{\mathcal{D}}(r)}$ flops. Let

$$d_{hb}(r) \triangleq -\lim_{\rho\to\infty}\frac{\log\mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)}{\log\rho}$$

denotes the precise optimum SNR exponent for $\mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}}\right)$, then in the presence of complexity hardbound at $\rho^{x_{\mathcal{D}}}$ flops, the achievable diversity gain $d_{\mathcal{D}}(r)$ is given by

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{hb}(r)\}.$$

We recall that from (3.15) and (3.19)

$$d_{hb}(r) \geq \inf_{\boldsymbol{\mu} \in \mathcal{T}(x_{\mathcal{D}})} I(\boldsymbol{\mu}).$$

It then follows that

$$d_{hb}(r) \geq d_{\mathcal{D}}(r, x_{\mathcal{D}}), \tag{3.88}$$

where $d_{\mathcal{D}}(r, x_{\mathcal{D}}) \triangleq \lim_{\epsilon \to 0^+} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon)$ and where

$$\begin{aligned} d_{\mathcal{D}}(r, c_{\mathcal{D}}(r) + \epsilon) \triangleq & \inf I(\boldsymbol{\mu}) && \text{(3.89a)} \\ \text{s.t.} \quad & \sum_{j=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1 - \mu_i) \right)^+ \geq c_{\mathcal{D}}(r) + \epsilon, \\ & 1 \geq \mu_1 \geq \cdots \geq \mu_\kappa \geq 0. \end{aligned}$$

Let $\tilde{\boldsymbol{\mu}}^*$ be one of the maximizing vectors of the optimization problem in (3.89). It then follows that

$$d_{\mathcal{D}}(r, x_{\mathcal{D}}) = I(\tilde{\boldsymbol{\mu}}^*). \tag{3.90}$$

In the following we derive a upper bound on $d_{hb}(r)$ and show that it matches the lower bound $d_{\mathcal{D}}(r, x_{\mathcal{D}})$. Going back to (3.90) with $0 \leq \tilde{\mu}_i^* \leq 1$, we let q be the largest integer for which $\frac{rT}{\kappa} - (1 - \tilde{\mu}_q^*) > 0$, and similar to Section 3.2.2 we define the set Ω_{1-hb} as

$$\begin{aligned} \Omega_{1-hb} \triangleq \{ & \tilde{\mu}_j^* - 2\delta < \mu_j < \tilde{\mu}_j^* - \delta, j = 1, \cdots, q \\ & 0 < \mu_j < \delta, j = q+1, \cdots, n_T \}, \end{aligned} \tag{3.91}$$

for a given small $\delta' > 0$. Then from (3.45) it follows that

$$\mathrm{P}\left(N_{SD} \geq \rho^{c_{\mathcal{D}}(r) + \epsilon - \delta'} \right) \dot{\geq} \mathrm{P}\left(\Omega_{1-hb} \right). \tag{3.92}$$

Choosing $\epsilon > \delta' > 0$, for $x_{\mathcal{D}} > c_{\mathcal{D}}(r)$ gives that

$$\mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}} \right) \geq \mathrm{P}\left(N_{SD} \geq \rho^{c_{\mathcal{D}}(r) + \epsilon - \delta'} \right),$$

and it follows that

$$\mathrm{P}\left(N_{SD} \geq \rho^{x_{\mathcal{D}}} \right) \dot{\geq} \mathrm{P}\left(\Omega_{1-hb} \right), \tag{3.93}$$

We now compute $\mathrm{P}(\Omega_{1-hb})$ similar to $\mathrm{P}(\Omega_1)$ in (3.46), to get that

$$\mathrm{P}(\Omega_{1-hb}) \dot{\geq} \rho^{-d_{\mathcal{D}}(r,x_{\mathcal{D}})}.$$

Substituting this in (3.93) gives that

$$\mathrm{P}(N_{SD} \geq \rho^{x_{\mathcal{D}}}) \dot{\geq} \rho^{-d_{\mathcal{D}}(r,x_{\mathcal{D}})}.$$

Consequently, we have that $d_{hb}(r) \leq d_{\mathcal{D}}(r, x_{\mathcal{D}})$. This along with (3.88) gives that $d_{hb}(r) = d_{\mathcal{D}}(r, x_{\mathcal{D}})$. Thus, it follows that

$$d_{\mathcal{D}}(r) = \min\{d(r), d_{\mathcal{D}}(r, x_{\mathcal{D}})\}.$$

This proves Theorem 7. □

Chapter 4

Complexity of Lattice Decoding

4.1 Introduction

The complexity analysis of the ML-based sphere decoding solutions presented in this work has revealed that, to achieve a vanishing gap to ML solutions, these decoders generally require computational resources that, albeit significantly smaller than those required by an exhaustive ML decoder, again grow exponentially in the rate and the dimensionality, and remain prohibitive for several MIMO scenarios. As an indicative example of this increased complexity we recall that such SD algorithms, when applied for decoding a large family of high-performing codes including all known full-rate DMT optimal codes, over the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$) quasi-static MIMO channel, introduce a complexity exponent [1] of the form

$$c(r) = \frac{T}{n_{\mathrm{T}}}\big(r(n_{\mathrm{T}} - \lfloor r \rfloor - 1) + (n_{\mathrm{T}} \lfloor r \rfloor - r(n_{\mathrm{T}} - 1))^{+}\big). \tag{4.1}$$

The exponent, which simplifies to $c(r) = \frac{T}{n_{\mathrm{T}}} r(n_{\mathrm{T}} - r)$ for integer values of r, reaches at $r = n_{\mathrm{T}}/2$ (for even values of n_{T}) an overall maximum value of $n_{\mathrm{T}}T/4$ which, for the aforementioned codes is equal to $\kappa/8$, corresponding to complexity in the order of $2^{\frac{1}{8}\kappa \log \rho} = \rho^{\kappa/8} = \sqrt{|\mathcal{X}|}$. At any fixed multiplexing gain, these required computational resources can be seen to be in the order of

1. Although premature at this point, we hasten to note this complexity indeed holds irrespective of the radius updating policy, irrespective of the decoding ordering, and as we will see later on, holds even in the presence of MMSE preprocessing.

$2^{RT(\frac{n_{\mathrm{T}}-r}{n_{\mathrm{T}}})}$ flops which reveals a complexity that is exponential in the number of codeword bits, and a corresponding exponential slope of $\frac{n_{\mathrm{T}}-r}{n_{\mathrm{T}}}$.

4.1.1 Transition to lattice decoding for reducing complexity

This high complexity required by ML-based decoding solutions, serves as further motivation for exploring other families of decoding methods. A natural alternative is lattice decoding obtained by simply removing the constellation boundaries of the ML-based search, an action that loosely speaking exploits a certain symmetry which in turn may yield faster implementations and takes the general form

$$\hat{\mathbf{x}}_L = \arg \min_{\hat{\mathbf{x}} \in \Lambda_r} \|\mathbf{y} - \sqrt{\rho}\mathbf{H}\hat{\mathbf{x}}\|^2 . \tag{4.2}$$

Naturally when $\hat{\mathbf{x}}_L \notin \mathcal{X}_r$, the decoder declares an error.

The use of lattice decoding, and specifically of preprocessed lattice decoding in MIMO communications has received substantial attention from works like [3], [5] and [56], where the latter proved that lattice decoding in the presence of MMSE preprocessing achieves the optimal DMT for specific MIMO channels and statistics, and for DMT-optimal random codes. The use of lattice decoding as an alternative to computationally expensive ML based solutions, was recently further validated on the one hand by the aforementioned work in [7] which revealed the large computational disadvantages of ML based solutions, and on the other hand by the work in [57] which further confirmed the performance advantages of lattice decoding by showing that regularized (MMSE-preprocessed) [2] lattice decoding achieves the optimal DMT performance, for almost all MIMO scenarios and fading statistics, and all non-random lattice codes, irrespective of the codes' ML performance

It is the case though that even with lattice decoding, the computational complexity can be prohibitive : finding the exact solution to the lattice decoding problem is generally an non-deterministic polynomial-time (NP) hard problem (cf. [38]). At the same time though, the other extreme of very early terminations of lattice decoding, such as linear solutions, have been known to achieve computational efficiency at the expense though of a very sizable, and often unbounded, gap to the exact solution of the lattice decoding problem. In this work we explore lattice decoding solutions that, in conjunction with terminating policies, strike the proper balance between this exponential complexity and exponential gap.

The aforementioned extreme complexity of exact lattice decoding solutions, in conjunction with the potentially unbounded error-performance

2. We will interchangeably use *MMSE-preprocessed decoder* and *regularized decoder*, with the first term being more commonly used, and with the second implying a more general family of decoders (cf. [57]). Even though in the asymptotic setting of interest, the two accept the same results throughout the work, some extra error-performance gains can be achieved by proper optimization of the regularized decoder (cf. [58]).

degradation (gap) of very early terminations (as opposed to exact implementations) of lattice decoding, bring to the fore the need for approximate lattice decoding solutions that better balance the very sizable complexity and gap. Specifically for any simplified variant $\mathcal{D}_r$ of the baseline (exact) MMSE-preprocessed lattice decoder, this gap can, in the high SNR regime, be quantified as

$$g_L(c) \triangleq \lim_{\rho\to\infty} \frac{P_e}{\mathrm{P}\left(\hat{\mathbf{x}}_L \neq \mathbf{x}\right)} \tag{4.3}$$

where $\mathrm{P}\left(\hat{\mathbf{x}}_L \neq \mathbf{x}\right)$ describes the probability of error of the exact MMSE-preprocessed lattice decoder, where P_e denotes the probability of error of $\mathcal{D}_r$, and where c (i.e., $c(r)$) is the complexity exponent that describes the (asymptotic rate of increase of the) computational resources required to achieve this performance gap. The clear task has remained for some time to construct decoders that optimally traverse this tradeoff between g and c, i.e., that reduce the performance gap to the exact lattice decoding solution, with reasonable computational complexity. Equivalently for $N_{\max}(g)$ denoting the computational resources in flops required to achieve a certain gap g to the baseline exact MMSE-preprocessed lattice decoder, the above task can be described, in the high SNR regime, as trying to minimize

$$\lim_{\rho\to\infty} \frac{\log N_{\max}(g)}{\log \rho}.$$

This will be achieved later on.

In Section 4.2 we first show that the computational complexity required by the MMSE-preprocessed (unconstrained) lattice sphere decoder, asymptotically matches the complexity of the (constrained) ML-based (MMSE-preprocessed or not) sphere decoders, and is commonly exponential in the dimensionality and the number of codeword bits. This is established for a large class of codes of arbitrary error-performance, a large class of fading statistics, and specifically for the quasi-static MIMO channel – for example the complexity required for DMT optimal lattice sphere decoding, in the presence of a large family of DMT optimal codes, takes the previously seen simple piecewise linear form in (4.1). We also provide a universal upper bound on the complexity of regularized lattice sphere decoding, which holds irrespective of the lattice code applied and irrespective of the fading statistics. This upper bound again takes the form in (4.1), matching that in the case of constrained ML-based sphere decoding, thus revealing the surprising fact that there exists no statistical channel behavior that will allow the removal of the bounding region to cause unbounded increases in the complexity of the decoder [3].

3. In other words, this complexity bound holds even if the channel statistics are such that the channel realizations cause the decoder to always have to solve the hardest possible lattice search problem.

With provable evidence of the very high complexity of regularized lattice decoding, in Section 4.3 we turn to the powerful tool of lattice reduction (LR) and seek to understand its effects on computational complexity. While there has existed a general agreement in the community that lattice reduction does reduce complexity, cf. [4], this has not yet been supported analytically in any relevant communication settings. In fact, and quite opposite to common wisdom, it was recently shown that for a fixed radius sphere decoding implementation of the naive lattice decoder [59], LR does not improve the sphere decoder complexity tail exponent.

What this work shows is that lattice reduction reduces an ML-like exponentially increasing complexity, to very manageable subexponential values. We specifically proceed to prove that the LR-aided regularized lattice decoder, implemented by a fixed-radius sphere decoder and timeout policies that occasionally abort decoding and declare an error, achieves

$$g_L(\epsilon) = 1, \quad \lim_{\rho\to\infty} \frac{\log N_{\max}(g)}{\log \rho} = 0 \qquad \forall \epsilon > 0, g \geq 1,$$

i.e., achieves a vanishing gap to the exact implementation of regularized lattice decoding and does so with a complexity exponent that vanishes to zero, which in turn implies subexponential complexity in the sense that the complexity scales slower than any conceivable exponential function. It is noted that this vanishing gap approach serves the practical purpose of an analytical refinement over basic diversity analysis which generally fails to address potentially massive gaps between theory and practice.

4.2 Regularized Lattice Sphere Decoding

We proceed to describe the preprocessed lattice decoder, its sphere decoding implementation, and for a practical setting of interest that includes the quasi-static MIMO channel and common codes, to establish the decoder's computational complexity.

4.2.1 Lattice sphere decoding

We again start with the general $m \times n$ point-to-point multiple-input multiple-output model given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{4.4}$$

For the class of lattice codes considered here, the codewords take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}} \mathbf{G}\mathbf{s}, \qquad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{rT}{\kappa}} \mathcal{R}, \tag{4.5}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}'$ that preserves the code, and where $\mathcal{R}$ contains the all zero vector $\mathbf{0}$. Combining (4.4) and (4.5) yields the equivalent model

$$\mathbf{y} = \mathbf{M}_r \mathbf{s} + \mathbf{w} \tag{4.6}$$

where

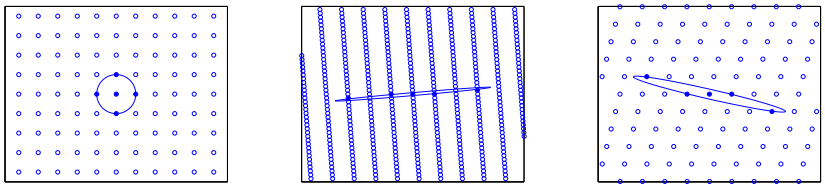
$$\mathbf{M}_r = \rho^{\frac{1}{2} - \frac{rT}{\kappa}} \mathbf{HG} \in \mathbb{R}^{n \times \kappa} \tag{4.7}$$

is a function of the multiplexing gain [4] r.

Consequently the corresponding naive lattice decoder in (4.2) takes the form (see for example [57], also [4])

$$\hat{\mathbf{s}}_L = \arg \min_{\hat{\mathbf{s}} \in \mathbb{Z}^\kappa} \|\mathbf{y} - \mathbf{M}\hat{\mathbf{s}}\|^2 . \tag{4.8}$$

As a result though of neglecting the boundary region, the above decoder declares additional errors if $\hat{\mathbf{s}}_L \notin \mathbb{S}_r^\kappa$, resulting in possible performance costs. These costs motivated the use of MMSE preprocessing which essentially regularizes the decision metric to penalize vectors outside the boundary constraint $\mathbb{S}_r^\kappa$ (cf. [57]). A pictorial representation of the MMSE-preprocessed lattice is shown in Fig. 4.1.



(a) Transmitted lattice (b) Received lattice (c) Regularized lattice

FIGURE 4.1 – 2-dimensional lattice for regularized lattice decoder

Specifically the MMSE-preprocessed lattice decoder is obtained by implementing an unconstrained search over the MMSE-preprocessed lattice, and takes the form

$$\hat{\mathbf{s}}_{r-ld} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{Z}^\kappa} \|\mathbf{F}\mathbf{y} - \mathbf{R}\hat{\mathbf{s}}\|^2 , \tag{4.9}$$

where $\mathbf{F}$ and $\mathbf{R}$ are respectively the MMSE forward and feedback filters such that $\mathbf{F} = \mathbf{R}^{-H}\mathbf{M}^H$, where

$$\mathbf{R}^H\mathbf{R} = \mathbf{M}^H\mathbf{M} + \alpha_r^2\mathbf{I}, \tag{4.10}$$

4. For simplicity of notation we will, in most cases, denote $\mathbf{M}_r$ with $\mathbf{M}$.

where $\alpha_r = \rho^{\frac{-rT}{\kappa}}$ and where $\mathbf{R}$ is an upper-triangular matrix. For $\mathbf{r} \triangleq \mathbf{F}\mathbf{y}$, the model transitions from (4.6) to

$$\begin{aligned} \mathbf{r} &= \mathbf{R}^{-H}\mathbf{M}^H\mathbf{M}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}^{-H}(\mathbf{R}^H\mathbf{R} - \alpha_r^2\mathbf{I})\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}\mathbf{s} - {\alpha_r}^2\mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= \mathbf{R}\mathbf{s} + \mathbf{w}' \end{aligned} \tag{4.11}$$

where

$$\mathbf{w}^{'} = -\alpha_r^2\mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \tag{4.12}$$

is the equivalent noise that includes self-interference (first summand) and colored Gaussian noise. Consequently the corresponding regularized lattice decoder takes the form

$$\hat{\mathbf{s}}_{r-ld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2, \tag{4.13}$$

which is then solved by the sphere decoder which recursively enumerates all lattice vectors $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ within a given sphere of radius $\xi > 0$.

We recall that there is a one-to-one correspondence between the nodes at layer k and the partial vectors $\hat{\mathbf{s}}_k$. We say that a node is visited by the sphere decoder if and only if the corresponding partial vector $\hat{\mathbf{s}}_k$ satisfies (2.12), i.e., there is a bijection between the visited nodes at layer k and the set

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{Z}^k \mid \|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}. \tag{4.14}$$

Consequently the total number of visited nodes (in all layers of the tree) is given by

$$N_{SD} = \sum_{k=1}^{\kappa} N_k, \tag{4.15}$$

where $N_k \triangleq |\mathcal{N}_k|$ is the number of visited nodes at layer k of the search tree. Again, the total number of visited nodes is taken as a measure of the sphere decoder complexity.

In choosing this radius, we note that for the transmitted symbol vector $\mathbf{s}$, the metric in (4.21) satisfies

$$\|\mathbf{r} - \mathbf{R}\mathbf{s}\|^2 = \|\mathbf{w}^{'}\|^2,$$

which means that if $\|\mathbf{w}^{'}\| > \xi$, then the transmitted symbol vector is excluded from the search, resulting in a decoding error. As Lemma 3 will later argue taking into consideration the self-interference and non-Gaussianity of $\mathbf{w}^{'}$, we can set $\xi = \sqrt{z \log \rho}$, for some $z > d_L(r)$ such that

$$\mathrm{P}\left(\|\mathbf{w}^{'}\|^2 > \xi^2\right) \doteq \rho^{-d_L(r)},$$

which implies a vanishing probability of excluding the transmitted information vector from the search, and a vanishing degradation of performance.

4.2.2 Complexity of regularized lattice decoding

Regarding complexity, we note that the MMSE-preprocessed lattice sphere decoder differs from its ML-based equivalent in two aspects : the presence of MMSE preprocessing and the absence of a bounding region to constrain the search. These two aspects are generally perceived to have an opposite effect on the complexity. On the one hand, MMSE preprocessing, which we recall from (4.14) to introduce unpruned sets

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{Z}^k \mid \|\mathbf{r}_k - \mathbf{R}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}, \quad k = 1, \cdots, \kappa,$$

is associated to reduced complexity in lattice-based SD solutions (cf. [59]) due to the resulting penalization of faraway lattice points (cf. [57]). On the other hand, the absence of boundary constraints can be associated to increased complexity as it introduces an unbounded number of candidate vectors. We proceed to show that in terms of the complexity exponent, under common MIMO scenarios and codes, these two aspects exactly cancel each other out, and that consequently MMSE-preprocessed lattice sphere decoding introduces a complexity exponent that matches that of ML-based sphere decoding, which it self is shown here to also match the complexity exponent of ML-based SD in the presence of MMSE preprocessing [5].

In light of this, in this section only, we mainly focus on the widely considered $n_T \times n_R$ ($n_R \geq n_T$) i.i.d. and quasi-static MIMO setting and on the large but specific family of full-rate ($\kappa = 2\min\{n_\mathrm{T}, n_\mathrm{R}\}T = 2n_\mathrm{T}T$) threaded codes (cf. [12–15]), which includes all known DMT optimal codes as well as uncoded transmission (V-BLAST).

We proceed with the main result of the section, which applies to the case of the $n_T \times n_R$ ($n_R \geq n_T$) quasi-static MIMO channel with i.i.d. Rayleigh fading statistics, and which applies under the natural detection ordering (cf. [5, 7]).

Theorem 10. *Under natural decoding ordering, the complexity exponent of regularized lattice sphere decoding any full-rate threaded code, is equal to the complexity exponent of ML-based SD with or without MMSE preprocessing.*

The proof for this theorem is given in Appendix 4A. We clarify that even though all three decoders are DMT optimal, the above result incorporates more than just DMT optimal decoding, in the sense that any timeout policy will tradeoff $d(r)$ with $c(r)$ identically for ML-based and lattice-based sphere decoding. In other words the three decoders share the same $d(r)$ and $c(r)$ capabilities, irrespective of the timeout policy. Furthermore, considering different SD detection orderings (cf. [5]), the following extends the range of

5. We clarify that ML-based SD in the presence of MMSE preprocessing, corresponds to unpruned sets $\mathcal{N}_k \cap \mathbb{S}_r^k$ where $\mathbb{S}_r^k$ is the k-dimensional set resulting from the natural reduction of $\mathbb{S}_r^\kappa$ from (4.5).

codes for which the ML-based and lattice-based SD share a similar complexity.

Corollary 10a. *Given any full-rate code of arbitrary DMT performance, there is always at least one non-random fixed permutation of the columns of $\boldsymbol{G}$, for which the complexity exponent of the MMSE-preprocessed lattice sphere decoder matches that of the ML based sphere decoder.*

The proof follows from the proof of Theorem 10 and from Theorem 4 in [7].

Corollary 10b. *Irrespective of the fixed or dynamically changing decoding order, the complexity exponent for MMSE preprocessed lattice sphere decoding any (fixed but) randomly and uniformly chosen code (from an ensemble of DMT optimal full-rate lattice designs) over the quasi-static MIMO channel with i.i.d. Rayleigh fading statistics almost surely, in the choice of DMT optimal lattice design, matches the complexity exponent of ML-based SD with or without MMSE preprocessing.*

The proof follows from the proof of Theorem 10 and from Lemma 1. This corollary establishes the fact that ML and (MMSE-preprocessed) lattice decoding share the same complexity exponent for a very broad setting, which includes almost any code (randomly and uniformly drawn from an ensemble of lattice designs) and all decoding order policies. The following focuses on a specific example of practical interest.

Corollary 10c. *The complexity exponent for DMT optimal MMSE preprocessed lattice sphere decoding of minimum delay ($T = n_T$) DMT optimal threaded codes over the quasi-static MIMO channel with i.i.d. Rayleigh fading statistics, takes the following form*

$$c_{r-ld}(r) = r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+. \tag{4.16}$$

The proof follows directly from Theorem 10 and Corollary 5a.

Further evidence that connects the complexity behavior of MMSE preprocessed (unconstrained) lattice-based SD, with that of its ML-based (constrained) counterpart, now comes in the form of a non-trivial universal bound that is shared by the two methods. This is particularly relevant because unconstrained lattice decoding could conceivably require unbounded computational resources given the unbounded number of candidate lattice points. Specifically the following universal upper bound on the complexity, presented in Corollary 4a for the ML case, is shown here to apply also to the lattice based sphere decoders. The bound holds irrespective of the lattice code applied and irrespective of the fading statistics. The generality with respect to the fading statistics is important because it guarantees that no set of fading statistics, even those that always generate infinitely dense lattices, can cause

an unbounded increase in the complexity due to removal of the boundary constraints.

Corollary 10d. *For any decoding ordering policy, and set of fading statistics and any full-rate lattice design, the complexity exponents of regularized lattice SD is upper bounded as*

$$c_{r-ld}(r) \leq \overline{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right) \quad (4.17)$$

which simplifies to

$$\overline{c}(r) = \frac{T}{n_T} r(n_T - r) \quad (4.18)$$

for integer r.

The proof for this corollary is given in Appendix 4B.

Coming back to the main focus of the work, we recall that Theorem 10 revealed that for common practical scenarios of interest, MMSE-preprocessed lattice decoding introduces very high complexity that often matches the exponential complexity of ML based solutions. In the following, and after reverting to the most general setting of MIMO scenarios, statistics and full-rate lattice codes, we show how proper utilization of lattice sphere decoding and LR techniques can indeed reduce the complexity exponent to zero, at an error-performance cost that vanishes in the high SNR limit.

4.3 LR-aided Regularized Lattice Sphere Decoding

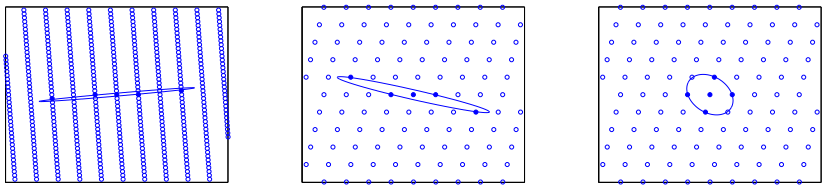
Lattice reduction techniques have been typically used in the MIMO setting to improve the error performance of suboptimal decoders (cf. [60], [61], see also [62], [63]). In the current setting the LR algorithm, which is employed at the receiver after the action of MMSE preprocessing, modifies the search of the MMSE-preprocessed lattice decoder, from

$$\hat{\mathbf{s}}_{rld} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2$$

(cf. (4.13)), to the new

$$\tilde{\mathbf{s}}_{lr-rld} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{Z}^\kappa} \|\mathbf{r} - \mathbf{RT}\hat{\mathbf{s}}\|^2, \quad (4.19)$$

by accepting as input the MMSE-preprocessed lattice generator matrix $\mathbf{R}$, and producing as output the matrix $\mathbf{T} \in \mathbb{Z}^{\kappa \times \kappa}$ which is unimodular meaning that it has integer coefficients and unit-norm determinant, and which is designed so that $\mathbf{RT}$ is (loosely speaking) more orthogonal than $\mathbf{R}$. As a result of this unimodularity, we have that $\mathbf{T}^{-1}\mathbb{Z}^\kappa = \mathbb{Z}^\kappa$, and consequently



(a) Received lattice (b) Regularized lattice (c) LR-Regularized lattice

FIGURE 4.2 – 2-dimensional lattice for LR-aided regularized lattice decoder

the new search in (4.19) corresponds to yet another *lattice* decoder, referred to as the LR-aided MMSE-preprocessed lattice decoder, which operates over a generally better conditioned channel matrix $\mathbf{RT}$. A pictorial representation of the LR-aided MMSE-preprocessed channel matrix is shown in Fig. 4.2.

Finally with sphere decoding in mind, the LR algorithm is followed by the QR decomposition [6] of the new lattice-reduced MMSE-preprocessed matrix $\mathbf{RT}$, resulting in a new upper-triangular model

$$\tilde{\mathbf{r}} \quad = \quad \tilde{\mathbf{R}}\tilde{\mathbf{s}} + \mathbf{w}'' \tag{4.20}$$

and the new LR-aided MMSE-preprocessed lattice search, which accepts the application of the sphere decoder, and which takes the form

$$\tilde{\mathbf{s}}_{lr-rld} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{Z}^{\kappa}} \left\| \tilde{\mathbf{r}} - \tilde{\mathbf{R}}\hat{\mathbf{s}} \right\|^2, \tag{4.21}$$

where $\tilde{\mathbf{Q}}\tilde{\mathbf{R}} = \mathbf{RT}$ corresponds to the QR-decomposition of $\mathbf{RT}$, where $\tilde{\mathbf{R}}$ is upper-triangular, where $\tilde{\mathbf{r}} \triangleq \tilde{\mathbf{Q}}^H \mathbf{r}$, $\tilde{\mathbf{s}} = \mathbf{T}^{-1}\mathbf{s}$, and where $\mathbf{w}'' = \tilde{\mathbf{Q}}^H \mathbf{w}'$.

At the very end,

$$\hat{\mathbf{s}}_{lr-rld} = \mathbf{T}\tilde{\mathbf{s}}_{lr-rld}, \tag{4.22}$$

allows for calculation of the estimate of the transmitted symbol vector $\mathbf{s}$ in (4.6).

We note here that this (exact) solution of the LR-aided MMSE preprocessed lattice decoder defined by (4.21), (4.22), is identical to the exact solution

6. A more proper statement would be that the QR decomposition is performed by the LR algorithm it self.

of the MMSE-preprocessed lattice decoder given by (4.13), because

$$\begin{aligned}
\min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \|\mathbf{r}-\mathbf{R}\hat{\mathbf{s}}\|^2 &= \min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \left\|\mathbf{r}-\mathbf{R}\mathbf{T}\mathbf{T}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(a)}{=} \min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \left\|\mathbf{r}-\tilde{\mathbf{Q}}\tilde{\mathbf{R}}\mathbf{T}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(b)}{=} \min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \left\|\tilde{\mathbf{r}}-\tilde{\mathbf{R}}\mathbf{T}^{-1}\hat{\mathbf{s}}\right\|^2 \\
&= \min_{\hat{\mathbf{s}}\in\mathbf{T}^{-1}\mathbb{Z}^\kappa} \left\|\tilde{\mathbf{r}}-\tilde{\mathbf{R}}\hat{\mathbf{s}}\right\|^2 \\
&\overset{(c)}{=} \min_{\hat{\mathbf{s}}\in\mathbb{Z}^\kappa} \left\|\tilde{\mathbf{r}}-\tilde{\mathbf{R}}\hat{\mathbf{s}}\right\|^2, \qquad (4.23)
\end{aligned}$$

where (a) follows from the fact that $\tilde{\mathbf{Q}}\tilde{\mathbf{R}} = \mathbf{R}\mathbf{T}$, (b) follows from the rotational invariance of the Euclidean norm, and (c) follows from the fact that $\mathbf{T}^{-1}\mathbb{Z}^\kappa = \mathbb{Z}^\kappa$.

While though the two lattice decoding solutions (with and without LR) provide identical error performance in the setting of exact implementations, we proceed to rigorously prove that, in terms of complexity, lattice reduction techniques, and specifically a proper utilization of the LLL algorithm [64], can provide dramatic improvements.

4.3.1 Complexity of LR-aided regularized lattice decoding

We are here interested in establishing the complexity of the LR-aided regularized lattice sphere decoder. Given that the costs of implementing MMSE preprocessing and of implementing the linear transformation in (4.22) are negligible in the scale of interest [7], we limit our focus on establishing the cost of lattice reduction, and then the cost of the SD implementation of the search in (4.21). Starting with the SD complexity, as in (4.14), we identify the corresponding unpruned set at layer k to be

$$\mathcal{N}_k \triangleq \{\hat{\mathbf{s}}_k \in \mathbb{Z}^k \mid \|\tilde{\mathbf{r}}_k - \tilde{\mathbf{R}}_k\hat{\mathbf{s}}_k\|^2 \leq \xi^2\}, \qquad (4.24)$$

and in bounding the size of the above, we first focus on understanding the statistical behavior of the $k \times k$ lower-right submatrices $\tilde{\mathbf{R}}_k$ of matrix $\tilde{\mathbf{R}}$ $(k = 1, \cdots, \kappa)$, where we recall that $\tilde{\mathbf{R}}$ is the upper triangular code-channel matrix, after MMSE preprocessing and LLL lattice reduction. Towards this, and for $d_L(r-\epsilon)$ denoting the diversity gain of the exact implementation of the regularized lattice decoder at multiplexing gain $r-\epsilon$, we have the following lemma on the smallest singular value of $\tilde{\mathbf{R}}_k$.

7. Even though the work here focuses on decoding, we can also quickly state the obvious fact that the cost of constructing the codewords is also negligible in the scale of interest because it again only involves a finite-dimensional linear transformation (cf. (4.5)).

Lemma 2. *The smallest singular value* $\sigma_{min}(\tilde{\mathbf{R}}_k)$ *of submatrix* $\tilde{\mathbf{R}}_k$, $k = 1, \cdots, \kappa$, *satisfies*

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_k) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}, \text{ for all } r \geq \epsilon > 0. \tag{4.25}$$

The proof for this lemma is given in Appendix 4C. To bound the cardinality N_k of $\mathcal{N}_k$ (cf. (4.24)), and eventually the total number $N_{SD} = \sum_{k=1}^{\kappa} N_k$ of lattice points visited by the SD, we proceed along the lines of the work in [7], making the proper modifications to account for MMSE preprocessing, for the removal of the bounding region, and for lattice reduction.

Towards this we see that, after removing the boundary constraint, Lemma 1 in [7] tells us that

$$N_k \triangleq |\mathcal{N}_k| \leq \prod_{i=1}^{k} \left[\sqrt{k} + \frac{2\xi}{\sigma_i(\tilde{\mathbf{R}}_k)}\right],$$

where

$$\sigma_{min}(\tilde{\mathbf{R}}_k) = \sigma_1(\tilde{\mathbf{R}}_k) \leq \cdots \leq \sigma_k(\tilde{\mathbf{R}}_k)$$

are the singular values of $\tilde{\mathbf{R}}_k$. Consequently we have that

$$N_k \quad \leq \quad \left[\sqrt{k} + \frac{2\xi}{\sigma_{min}(\tilde{\mathbf{R}}_k)}\right]^k. \tag{4.26}$$

As a result, for any $\tilde{\mathbf{R}}_k$ such that

$$\sigma_{min}(\tilde{\mathbf{R}}_k) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}, \tag{4.27}$$

and given that $\xi = \sqrt{z \log \rho}$ for some finite z, then

$$N_k \dot{\leq} \left(\sqrt{k} + \frac{2\sqrt{z \log \rho}}{\rho^{\frac{-\epsilon T}{\kappa}}}\right)^k \doteq \rho^{\frac{\epsilon T k}{\kappa}}, \tag{4.28}$$

which guarantees that the total number of visited lattice points is upper bounded as

$$N_{SD} = \sum_{k=1}^{\kappa} N_k \dot{\leq} \sum_{k=1}^{\kappa} \rho^{\frac{\epsilon T k}{\kappa}} \doteq \rho^{\epsilon T}. \tag{4.29}$$

Consequently, directly from Lemma 2, we have that

$$\mathrm{P}\left(N_{SD} \dot{\geq} \rho^{\epsilon T}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}. \tag{4.30}$$

A similar approach deals with the complexity of the LLL algorithm, which is known (cf. [65]) to be generally unbounded. The LLL-based lattice reduction

algorithm is applied here in the presence of specific time-out policies which will limit the otherwise unbounded complexity of the LLL algorithm by timing out and declaring an error whenever the number of LLL flops exceeds a certain threshold.

Specifically drawing from [57, Lemma 2], under the natural assumption of power-limited channels [8] (cf. [57]), under the natural assumption that $d_L(r-\epsilon) > d_L(r)$ for all $\epsilon > 0$, and for N_{LR} denoting the number of flops spent by the LLL algorithm, one can readily conclude that

$$\mathrm{P}\left(N_{LR} \geq \gamma \log \rho\right) \stackrel{\cdot}{\leq} \rho^{-d_L(r-\epsilon)}, \tag{4.31}$$

for any $\gamma > \frac{1}{2}(d_L(r-\epsilon))$. Consequently the overall complexity

$$N \doteq N_{SD} + N_{LR},$$

in flops, for the LR-aided MMSE preprocessed lattice sphere decoder, satisfies the following

$$\begin{aligned} \mathrm{P}\left(N \dot{\geq} \rho^{\epsilon T}\right) &\doteq \mathrm{P}\left(\{N_{SD} \dot{\geq} \rho^{\epsilon T}\} \cup \{N_{LR} \dot{\geq} \rho^{\epsilon T}\}\right) \\ &\dot{\leq} \rho^{-d_L(r-\epsilon)}. \end{aligned} \tag{4.32}$$

Now going back to (3.12a), and having in mind appropriate timeout policies that specifically guarantee a *vanishing* gap to the exact solution of regularized lattice decoding, we can see that the complexity exponent $c_{lr-rld}(r)$ is upper bounded as

$$\overline{c}_{lr-rld}(r) = \inf\{x \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N \geq \rho^x\right)}{\log \rho} > d_L(r)\}. \tag{4.33}$$

Applying (4.32) we see that for any positive $\epsilon_1 < \epsilon$, it is the case that

$$\overline{c}_{lr-rld}(r) = \inf\{\epsilon \mid -\lim_{\rho\to\infty} \frac{\log \mathrm{P}\left(N \geq \rho^{\epsilon T+\epsilon_1}\right)}{\log \rho} > d_L(r)\} \tag{4.34}$$

which vanishes arbitrarily close to zero, resulting in a zero complexity exponent, and in the aforementioned subexponential complexity. The following theorem then holds.

Theorem 11. *LR-aided MMSE-preprocessed lattice sphere decoding introduces a zero complexity exponent.*

8. This is a moderate assumption that asks that $\mathrm{E}\left\{\|\boldsymbol{H}\|^2\mathrm{F}\right\} \dot{\leq} \rho$. We note that this holds true for any telecommunications setting.

4.3.2 Gap to the exact solution of regularized lattice decoding

We here prove that the LR-aided regularized lattice sphere decoder and the associated time-out policies that guarantee a vanishing complexity exponent, also guarantee a vanishing gap to the error performance of the exact lattice decoding implementation. This result is motivated by potentially exponential gaps in the performance of other DMT optimal decoders (cf. [57]), where these gaps may grow exponentially up to $2^{\frac{\kappa}{2}}$ (cf. [66]) or may potentially be unbounded [67].

Towards establishing this gap, we recall that the exact MMSE preprocessed lattice decoder in (4.9) makes errors when $\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}$. On the other hand the LLL-reduced MMSE-preprocessed lattice sphere decoder with run-time constraints, in addition to making the same errors ($\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}$), also makes errors when the run-time limit of ρ^x flops becomes active, i.e., when $N \geq \rho^x$, as well as when a small search radius causes $\mathcal{N}_\kappa = \emptyset$. Consequently the corresponding performance gap to the exact regularized decoder, takes the form

$$g_L(x) = \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\{\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\} \cup \{N \geq \rho^x\} \cup \{\mathcal{N}_\kappa = \emptyset\}\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)}.$$

To bound the above gap, we apply the union bound and the fact that

$$\mathrm{P}\left(\mathcal{N}_\kappa = \emptyset\right) \leq \mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)$$

to get that

$$\begin{aligned} g_L(x) \quad \leq \quad & \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)} + \lim_{\rho\to\infty} \frac{\mathrm{P}\left(N \geq \rho^x\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)} \\ & + \lim_{\rho\to\infty} \frac{\mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)}. \end{aligned} \tag{4.35}$$

Furthermore from (4.23) we observe that

$$\mathrm{P}\left(\hat{\mathbf{s}}_{lr-rld} \neq \mathbf{s}\right) = \mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right), \tag{4.36}$$

and from (4.32) we recall that

$$\mathrm{P}\left(N \dot{\geq} \rho^{\epsilon T}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}$$

which implies that for any $x > 0$ it holds that

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(N \geq \rho^x\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)} = 0. \tag{4.37}$$

Finally the last term in (4.35) relates to the search radius ξ, and to the behavior of the noise $\mathbf{w}''$ which was shown in (4.12), (4.20) to take the form

$$\mathbf{w}'' = \tilde{\mathbf{Q}}^H \left(-\alpha_r^2 \mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w}\right). \tag{4.38}$$

The following lemma accounts for the fact that $\mathbf{w}''$ includes self interference and colored noise, to bound the last term in (4.35).

Lemma 3. *There exist a finite $z > d_L(r)$ for which a search radius $\xi = \sqrt{z \log \rho}$ guarantees that*

$$\lim_{\rho \to \infty} \frac{\mathrm{P}\left(\|\mathbf{w}''\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)} = 0. \tag{4.39}$$

The proof for this lemma is given in Appendix 4D.

Consequently combining (4.36), (4.37) and (4.39) gives that $g_L(x) = 1, \ \forall x > 0$. The following directly holds.

Theorem 12. *LR-aided MMSE-preprocessed lattice sphere decoding with a computational constraint activated at ρ^x flops, allows for a vanishing gap to the exact solution of MMSE-preprocessed lattice decoding, for any $x > 0$. Equivalently the same LR-aided decoder guarantees that*

$$g_L(\epsilon) = 1 \quad \textit{and} \quad \lim_{\rho \to \infty} \frac{\log N_{\max}(g)}{\log \rho} = 0 \qquad \forall \epsilon > 0, g \geq 1,$$

for all fading statistics, all MIMO scenarios, and all full-rate lattice codes.

Summary

The work identified the first lattice decoding solution that achieves, in the most general outage-limited MIMO setting and the high rate and high SNR limit, both a vanishing gap to the error-performance of the (DMT optimal) exact solution of preprocessed lattice decoding, as well as a computational complexity that is subexponential in the number of codeword bits. The proposed solution employs lattice reduction (LR)-aided regularized lattice sphere decoding and proper timeout policies. As it turns out, lattice reduction is a special ingredient that allows for complexity reductions; a role that was rigorously demonstrated here for the first time, by proving that without lattice reduction, for most common codes, the complexity cost for asymptotically optimal regularized lattice sphere decoding is exponential in the number of codeword bits, and in many cases it in fact matches the complexity cost of ML sphere decoding.

In light of the fact that, prior to this work, a vanishing error performance gap was generally attributed only to near-full lattice searches that have exponential complexity, in conjunction with the fact that subexponential complexity was generally attributed to early-terminated (linear) solutions which have though a performance gap that can be up to exponential in dimension and/or rate, the work constitutes the first proof that subexponential complexity need not come at the cost of exponential reductions in lattice decoding error performance.

Appendix 4A : Proof of Theorem 10

We begin by providing an upper bound on the complexity exponent of MMSE-preprocessed (unconstrained) lattice sphere decoding, where this bound holds for the general quasi-static MIMO channel, for all fading statistics and for any full-rate lattice code. We then proceed to provide a lower bound on the complexity exponent of the same decoder, where this bound, under the extra assumptions of i.i.d. Rayleigh fading statistics and of layered codes, will in fact match the above mentioned upper bound to prove the theorem and the associated corollaries. We recall from Section 2.1.1 that any $n_T \times n_R$ ($n_R \geq n_T$) quasi-static point-to-point MIMO channel corresponds to the general MIMO channel model. Specifically, from (3.63) for quasi-static channel the real-valued model takes the form

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{w} \tag{4.40}$$

as in (4.4) with $m = 2n_T T$, $n = 2n_R T$, and where

$$\mathbf{H} = \mathbf{I}_T \otimes \mathbf{H}_R. \tag{4.41}$$

As before the vectorized codewords $\mathbf{x}$, associated to the full-rate code, take the form

$$\mathbf{x} = \rho^{\frac{-rT}{\kappa}}\mathbf{G}\mathbf{s},\ \mathbf{s} \in \mathbb{Z}^{\kappa}, \tag{4.42}$$

where $\kappa = 2n_T T = m$, which allows us to rewrite the model as

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \tag{4.43}$$

for

$$\mathbf{M} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{H}\mathbf{G} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}(\mathbf{I}_T \otimes \mathbf{H}_R)\mathbf{G}. \tag{4.44}$$

Finally the corresponding coherent MMSE-preprocessed lattice decoder for $\mathbf{s}$ can be expressed, as in (4.13), to be

$$\hat{\mathbf{s}}_{r-ld} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{Z}^{\kappa}} \|\mathbf{r} - \mathbf{R}\hat{\mathbf{s}}\|^2, \tag{4.45}$$

where $\mathbf{r} = \mathbf{Q}_1^H\mathbf{y}$ and where $\mathbf{R} \in \mathbb{R}^{m\times m}$ and $\mathbf{Q}_1 \in \mathbb{R}^{n\times m}$ are obtained by considering the thin QR decomposition of an equivalent representation of the MMSE-preprocessed lattice decoder (cf. [68])

$$\mathbf{M}^{reg} \triangleq \begin{bmatrix} \mathbf{M} \\ \alpha_r\mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix}\mathbf{R} = \mathbf{Q}\mathbf{R} \in \mathbb{R}^{(n+m)\times m} \tag{4.46}$$

where $\mathbf{Q}_1 = \mathbf{R}^{-1}\mathbf{M} \in \mathbb{R}^{n\times m}$, $\mathbf{Q}_2 = \alpha_r\mathbf{R}^{-1} \in \mathbb{R}^{m\times m}$, $\alpha_r = \rho^{\frac{-rT}{\kappa}}$ and where $\mathbf{R}^H\mathbf{R} = \mathbf{M}^H\mathbf{M} + \alpha_r^2\mathbf{I}$.

Upper bound on complexity of regularized lattice SD

In establishing the upper bound, we consider Lemma 1 in [7], which we properly modify to account for MMSE preprocessing and for the removal of the constellation boundaries, and get that the number N_k of nodes visited at layer k by the MMSE-preprocessed lattice sphere decoder, is upper bounded as

$$N_k = |\mathcal{N}_k| \leq \prod_{i=1}^{k} \left\lfloor \sqrt{2k} + \frac{2\xi}{\sigma_i(\mathbf{R}_k)} \right\rfloor, \tag{4.47}$$

where $\sigma_i(\mathbf{R}_k)$, $i = 1, \cdots, k$ denote the singular values of $\mathbf{R}_k$ in increasing order.

Towards lower bounding $\sigma_i(\mathbf{R}_k)$, we note that

$$\sigma_i(\mathbf{R}_k) \geq \sigma_i(\mathbf{R}) = \sigma_i(\mathbf{M}^{reg}) = \sqrt{\alpha_r^2 + \sigma_i(\mathbf{M}^H\mathbf{M})}, \tag{4.48}$$

where the first inequality makes use of the interlacing property of singular values of sub-matrices [53]. Furthermore for

$$\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}_C^H\mathbf{H}_C)}{\log \rho},\ j = 1, \cdots, n_T \tag{4.49}$$

and $\mu_1 \geq \cdots \geq \mu_{n_T}$, we see that $\sigma_j(\mathbf{H}_C) = \rho^{-\frac{1}{2}\mu_j}$, and from (4.44) that

$$\begin{aligned} \sigma_i(\mathbf{M}) &\geq \rho^{\frac{1}{2}-\frac{rT}{\kappa}} \sigma_{\min}(\mathbf{G})\sigma_{(i)}(\mathbf{I}_T \otimes \mathbf{H}_R)) \\ &\doteq \rho^{\frac{1}{2}-\frac{rT}{\kappa}} \sigma_{l_{2T}(i)}(\mathbf{H}_C) \\ &= \rho^{\frac{-rT}{\kappa}+\frac{1}{2}(1-\mu_{l_{2T}(i)})}, \end{aligned} \tag{4.50}$$

where $l_{2T}(i) \triangleq \left\lceil \frac{i}{2T} \right\rceil$, and where the asymptotic equality is due to the fact that $\sigma_{\min}(\mathbf{G}) \doteq \rho^0$. Substituting from (4.50) in (4.48) we now have that

$$\sigma_i(\mathbf{R}_k) \dot{\geq} \rho^{\frac{-rT}{\kappa}+\frac{1}{2}(1-\mu_{l_{2T}(i)})^+}, \quad i = 1, \cdots, \kappa. \tag{4.51}$$

Corresponding to (4.47) we see that

$$\left\lfloor \sqrt{2k} + \frac{2\xi}{\sigma_i(\mathbf{R}_k)} \right\rfloor \dot{\leq} \rho^{\left(\frac{rT}{\kappa}-\frac{1}{2}(1-\mu_{l_{2T}(i)})^+\right)^+},$$

for any $i = 1, \cdots, 2n_T T$, and from (4.47) we have that

$$N_k(\boldsymbol{\mu}) \dot{\leq} \rho^{\sum_{i=1}^{k}\left(\frac{rT}{\kappa}-\frac{1}{2}(1-\mu_{l_{2T}(i)})^+\right)^+}, \tag{4.52}$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_{n_T})$. It follows that

$$\begin{aligned} N_{SD}(\boldsymbol{\mu}) = \sum_{k=1}^{\kappa} N_k(\boldsymbol{\mu}) &\dot{\leq} \sum_{k=1}^{\kappa} \rho^{\sum_{i=1}^{k} \left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_{l_{2T^{(i)}}})^+ \right)^+} \\ &\doteq \rho^{\sum_{i=1}^{\kappa} \left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_{l_{2T^{(i)}}})^+ \right)^+} \\ &\doteq \rho^{T \sum_{j=1}^{n_T} \left(\frac{r}{n_T} - (1-\mu_j)^+ \right)^+}, \end{aligned} \tag{4.53}$$

where the last asymptotic equality is due to the multiplicity of the singular values.

Following the footsteps of the complexity analysis in Section 3.2.2 an upper bound on the complexity exponent $c_{r-ld}(r)$ can be obtained as the solution to a constrained maximization problem according to

$$c_{r-ld}(r) \leq \overline{c}_{r-ld}(r) \triangleq \max_{\boldsymbol{\mu}} T \sum_{j=1}^{n_T} \left(\frac{r}{n_T} - (1-\mu_j)^+ \right)^+ \tag{4.54a}$$

$$\text{s.t.} \quad I(\boldsymbol{\mu}) \leq d_L(r), \tag{4.54b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \tag{4.54c}$$

Equivalently for $\boldsymbol{\mu}^* = (\mu_1^*, \cdots, \mu_{n_T}^*)$ being one of the maximizing vectors [9], i.e., such that $\boldsymbol{\mu}^* \in \Im(x)$ and $I(\boldsymbol{\mu}^*) = d_L(r)$, then $\overline{c}_{r-ld}(r)$ takes the form

$$\overline{c}_{r-ld}(r) = T \sum_{j=1}^{n_T} \left(\frac{r}{n_T} - (1-\mu_j^*)^+ \right)^+. \tag{4.55}$$

As we will now show, the above bound is also shared by the ML-based sphere decoder, with or without MMSE preprocessing, irrespective of the full-rate code and the fading statistics. Directly from Corollary 5, and taking into consideration that MMSE-preprocessed lattice decoding is DMT optimal for any code [57], we recall that the equivalent upper bound for the ML-based sphere decoder, without MMSE preprocessing, takes the form

$$\overline{c}_{quasi}(r) \triangleq \max_{\boldsymbol{\mu}} T \sum_{j=1}^{n_T} \min \left(\frac{r}{n_T} - 1 + \mu_j, \frac{r}{n_T} \right)^+ \tag{4.56a}$$

$$\text{s.t.} \quad I(\boldsymbol{\mu}) \leq d_L(r), \tag{4.56b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \tag{4.56c}$$

Comparing (3.22) and (4.56) we are able to conclude that both the objective functions (3.22a) and (4.56a) as well as both pairs of constraints are identical.

9. In general, (3.22) does not have a unique optimal point because $(a)^+$ is constant in a for $a \leq 0$.

To see this, we first note that for $0 \leq \mu_j \leq 1$, then

$$\min\left(\frac{r}{n_T} - 1 + \mu_j, \frac{r}{n_T}\right)^+ = \left(\frac{r}{n_T} - 1 + \mu_j\right)^+,$$
$$\left(\frac{r}{n_T} - (1-\mu_j)^+\right)^+ = \left(\frac{r}{n_T} - 1 + \mu_j\right)^+,$$

and furthermore we note that for $\mu_j > 1$, then

$$\min\left(\frac{r}{n_T} - 1 + \mu_j, \frac{r}{n_T}\right)^+ = \left(\frac{r}{n_T} - (1-\mu_j)^+\right)^+ = \frac{r}{n_T},$$

which proves that $\overline{c}_{quasi}(r)$ and $\overline{c}_{r-ld}(r)$ are identical.

In considering the case of MMSE-preprocessed ML SD, it is easy to see that the summands in the objective function in (4.56a) will be modified to take the form $\min\left(\frac{r}{n_T} - (1-\mu_j)^+, \frac{r}{n_T}\right)^+$ which can be seen to match (3.22a) for all $\mu_j \geq 0$, which in turn concludes the proof that the upper bound $\overline{c}_{r-ld}(r)$ for MMSE-preprocessed lattice SD is also shared by the ML-based sphere decoder, with or without MMSE preprocessing, irrespective of the full-rate code, and for all fade statistics represented by monotonic rate functions.

Lower bound on complexity of regularized lattice SD

We will here, under the extra assumptions of i.i.d. Rayleigh fading statistics and of layered codes with natural decoding order, provide a lower bound that matches the upper bound in (4.55). The same bound and tightness will also apply to any full-rate code, under the assumption of a fixed, worst case decoding ordering.

The goal here is to show that at layer $k = 2qT$, for some $q \in [1, n_{\mathrm{T}}]$, the sphere decoder visits close to $\rho^{\overline{c}_{r-ld}(r)}$ nodes with a probability that is large compared to the probability of decoding error $\mathrm{P}(\mathbf{s}_L \neq \mathbf{s}) \doteq \rho^{-d_L(r)}$, which from the expression of the complexity exponent (3.12b), will prove that $c_{r-ld}(r) = \overline{c}_{r-ld}(r)$.

Going back to (4.55), we let q be the largest integer for which

$$\frac{r}{n_T} - (1-\mu_q^*)^+ > 0, \tag{4.57}$$

in which case (4.55) takes the form

$$\overline{c}_{r-ld}(r) = T\sum_{j=1}^{q} \frac{r}{n_T} - (1-\mu_j^*)^+. \tag{4.58}$$

We recall from (4.49) that $\mu_j = -\frac{\log \sigma_j(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho}$, $j = 1, \cdots, n_T$, and that $\boldsymbol{\mu}^* \in \mathcal{T}(x)$ satisfies $I(\boldsymbol{\mu}^*) = d_L(r)$ and maximizes (4.54). We also note that

without loss of generality we can assume that $q \geq 1$ as otherwise $\overline{c}_{r-ld}(r) = 0$ (cf. (4.55)). Consequently it is the case that $\mu_j^* > 0$ for $j = 1, \cdots, q$. Furthermore given the monotonicity of the rate function $I(\boldsymbol{\mu})$, and the fact that the objective function in (4.54a) does not increase in μ_j beyond $\mu_j = 1$, we may also assume without loss of generality that $\mu_j^* \leq 1$ for $j = 1, \cdots, n_T$.

We proceed to define two events Ω_1 and Ω_2 which we will prove to be jointly sufficient so that, at layer $k = 2qT$, the sphere decoder visits close to $\rho^{\overline{c}_{r-ld}(r)}$ nodes. These are given by

$$\begin{aligned} \Omega_1 \triangleq \{&\mu_j^* - 2\delta < \mu_j < \mu_j^* - \delta, j = 1, \cdots, q \\ &0 < \mu_j < \delta, j = q+1, \cdots, n_T\}, \end{aligned} \tag{4.59}$$

for a given small $\delta > 0$, and

$$\Omega_2 \triangleq \{\sigma_1\left((\mathbf{I}_T \otimes \mathbf{V}_p^H)\mathbf{G}_{|p}\right) \geq u\}, \tag{4.60}$$

for some given $u > 0$, where for $p \triangleq n_T - q$ then $\mathbf{G}_{|p}$ denotes the first $2pT$ columns of $\mathbf{G}$, and where $\mathbf{V}_p$ denotes the last $2p$ columns of $\mathbf{V}$ obtained by applying the singular value decomposition on $\mathbf{H}_R$, i.e., $\mathbf{H}_R = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where

$$\boldsymbol{\Sigma} \triangleq \mathrm{diag}\{\sigma_1(\mathbf{H}_R), \cdots, \sigma_{2n_T}(\mathbf{H}_R)\}$$

with $\sigma_1(\mathbf{H}_R) \leq \cdots \leq \sigma_{2n_T}(\mathbf{H}_R)$ and $\mathbf{V}\mathbf{V}^H = \mathbf{I}$. Hence, $\mathbf{V}_p^H$ corresponds to the $2p$ largest singular values of $\mathbf{H}_R$.

Note also that by choosing δ sufficiently small, and using the fact that $\mu_i^* > 0$ for $i = 1, \cdots, q$, we may without loss of generality assume that Ω_1 implies that $\mu_j > 0$ for all $j = 1, \cdots, n_T$.

Modifying the approach in [7, Theorem 1] to account for MMSE pre-processing and unconstrained decoding, the lower bound on the number of nodes visited at layer k by the sphere decoder, is given by

$$N_k \geq \prod_{i=1}^{k} \left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k} \right]^+ . \tag{4.61}$$

In the following, and up until (4.67), we will work towards upper bounding $\sigma_i(\mathbf{R}_k)$ so that we can then lower bound N_k.

Towards this let

$$\mathbf{M}_{|p}^{reg} \triangleq \begin{bmatrix} \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{H}\mathbf{G}_{|p} \\ \alpha_r \mathbf{I}_{|p} \end{bmatrix} \in \mathbb{R}^{2(n_R+n_T)T \times 2pT}$$

contain the first $2pT$ columns of $\mathbf{M}^{reg}$ from (4.46), and note that

$$(\mathbf{M}_{|p}^{reg})^H \mathbf{M}_{|p}^{reg} = \rho^{1-\frac{2rT}{\kappa}} \mathbf{G}_{|p}^H \mathbf{H}^H \mathbf{H} \mathbf{G}_{|p} + \alpha_r^2 \mathbf{I} \ ,$$

and that from (4.41) we get

$$(\mathbf{M}_{|p}^{reg})^H\mathbf{M}_{|p}^{reg} = \rho^{1-\frac{2rT}{\kappa}}\mathbf{G}_{|p}^H(\mathbf{I}_T \otimes \mathbf{H}_R^H\mathbf{H}_R)\mathbf{G}_{|p} + \alpha_r^2\mathbf{I}.$$

Since

$$\begin{aligned}\mathbf{H}_R^H\mathbf{H}_R &= \mathbf{V}(\mathrm{diag}\{\sigma_1(\mathbf{H}_R^H\mathbf{H}_R),\cdots,\sigma_{2n_T}(\mathbf{H}_R^H\mathbf{H}_R)\})\mathbf{V}^H\\ &= \mathbf{V}(\mathrm{diag}\{\sigma_1(\mathbf{H}_R^H\mathbf{H}_R),\cdots,\sigma_{2n_T}(\mathbf{H}_R^H\mathbf{H}_R)\}\\ &\quad - \sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\mathbf{V}\mathrm{diag}\{\underbrace{0,\cdots,0}_{2q},\underbrace{1,\cdots,1}_{2p}\})\mathbf{V}^H\\ &\quad + \sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\mathbf{V}(\mathrm{diag}\{\underbrace{0,\cdots,0}_{2q},\underbrace{1,\cdots,1}_{2p}\})\mathbf{V}^H,\end{aligned}$$

we have that

$$\begin{aligned}\mathbf{H}_R^H\mathbf{H}_R &\succeq \sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\mathbf{V}(\mathrm{diag}\{\underbrace{0,\cdots,0}_{2q},\underbrace{1,\cdots,1}_{2p}\})\mathbf{V}^H\\ &= \sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\mathbf{V}_p\mathbf{V}_p^H\end{aligned}$$

where the last equality follows from the fact that $\mathbf{V}_p$ contains the last $2p$ columns of $\mathbf{V}$ and where $\mathbf{A} \succeq \mathbf{B}$ denotes that $\mathbf{A}-\mathbf{B}$ is positive-semidefinite. Since $\sigma_i(\mathbf{H}^H\mathbf{H}) \in \mathbb{R}$ and since the Kronecker product induces singular value multiplicity, it follows that

$$(\mathbf{M}_{|p}^{reg})^H\mathbf{M}_{|p}^{reg} \succeq \rho^{1-\frac{2rT}{\kappa}}\sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\mathbf{G}_{|p}^H(\mathbf{I}_T \otimes \mathbf{V}_p\mathbf{V}_p^H)\mathbf{G}_{|p} + \alpha_r^2\mathbf{I}.$$

With respect to the smallest singular value of $(\mathbf{M}_{|p}^{reg})^H\mathbf{M}_{|p}^{reg}$ we have

$$\sigma_1((\mathbf{M}_{|p}^{reg})^H\mathbf{M}_{|p}^{reg}) \geq \rho^{1-\frac{2rT}{\kappa}}\sigma_{(2q+1)}(\mathbf{H}_R^H\mathbf{H}_R)\sigma_1\left(\mathbf{G}_{|p}^H(\mathbf{I}_T \otimes \mathbf{V}_p\mathbf{V}_p^H)\mathbf{G}_{|p}\right) + \alpha_r^2$$

and consequently, given that $\mathbf{H}_R \in \Omega_2$, we have that

$$\begin{aligned}\sigma_1(\mathbf{M}_{|p}^{reg}) &\geq \rho^{-\frac{rT}{\kappa}}\sqrt{u^2\rho\sigma_{l_2(2q+1)}(\mathbf{H}_C^H\mathbf{H}_C)+1}\\ &\doteq \rho^{-\frac{rT}{\kappa}}\rho^{\frac{1}{2}(1-\mu_{q+1})^+}\\ &\geq \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\delta)^+},\end{aligned} \tag{4.62}$$

where the first inequality follows from (4.60), the exponential equality follows from (4.49) and from the fact that $u > 0$ is fixed and independent of ρ, and the last inequality follows from (4.59).

From (4.44) we have that

$$\begin{aligned}\sigma_i(\mathbf{M}^{reg}) &\leq \rho^{\frac{-rT}{\kappa}}\sqrt{(1+\rho(\sigma_\kappa(\mathbf{G})\sigma_{l_{2T}(i)}(\mathbf{H}_C))^2)}\\ &\doteq \rho^{\frac{-rT}{\kappa}+\frac{1}{2}(1-\mu_{l_{2T}(i)})^+}, \quad i=1,\cdots,2n_TT,\end{aligned} \tag{4.63}$$

where the asymptotic equality follows from the fact that $\sigma_\kappa(\mathbf{G})$ is fixed and independent of ρ. Furthermore (4.59) gives that for $i = 1, \cdots, 2qT$ then

$$\sigma_i(\mathbf{M}^{reg}) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+}, \tag{4.64}$$

where we have made use of the fact that $\mu_j^* \leq 1$ for $j = 1, \cdots, n_T$.

Given that $\mu_j^* > 0$ for $j = 1, \cdots, q$, then for sufficiently small δ and for $i = 1, \cdots, 2qT$, we have that

$$-\frac{rT}{\kappa} + \frac{1}{2}(1-\delta)^+ \geq -\frac{rT}{\kappa} + \delta + \frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+,$$

which means that for sufficiently small δ, a comparison of (4.62) and (4.64) yields

$$\sigma_i(\mathbf{M}^{reg}) < \sigma_1(\mathbf{M}^{reg}_{|p}), \quad i = 1, \cdots, 2qT.$$

The above inequality allows us to apply Lemma 3 in [7], which in turn gives that

$$\sigma_i(\mathbf{R}_k) \leq \left[\frac{\sigma_\kappa(\mathbf{M}^{reg})}{\sigma_1(\mathbf{M}^{reg}_{|p})} + 1\right] \sigma_i(\mathbf{M}^{reg}), \quad i = 1, \cdots, 2qT. \tag{4.65}$$

Setting $i = \kappa$ in (4.63) upper bounds the maximum singular value of $\mathbf{M}^{reg}$ as

$$\sigma_\kappa(\mathbf{M}^{reg}) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\frac{1}{2}(1-\mu_{n_T})^+} \leq \rho^{\frac{1}{2}-\frac{rT}{\kappa}}, \tag{4.66}$$

where the last inequality is due to the fact that $\mu_j \geq 0$. Consequently combining (4.66) and (4.62) gives that

$$\left[\frac{\sigma_\kappa(\mathbf{M}^{reg})}{\sigma_1(\mathbf{M}^{reg}_{|p})} + 1\right] \dot{\leq} \rho^{\frac{1}{2}\delta},$$

which together with (4.64) and (4.65) gives that

$$\sigma_i(\mathbf{R}_k) \dot{\leq} \rho^{-\frac{rT}{\kappa}+\frac{3}{2}\delta+\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+}, \quad i = 1, \cdots, 2qT. \tag{4.67}$$

Consequently, going back to (4.61), we have that

$$\left[\frac{2\xi}{\sqrt{k}\sigma_i(\mathbf{R}_k)} - \sqrt{k}\right]^+ \dot{\geq} \rho^{\left(\frac{rT}{\kappa}-\frac{3}{2}\delta-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+\right)} > 0 \tag{4.68}$$

and furthermore for $i = 1, \cdots, 2qT$, we have that $\frac{rT}{\kappa} - \frac{3}{2}\delta - \frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+ > 0$ directly from definition of q and for sufficiently small δ. As a result, for $k \leq 2qT$ we have that

$$N_k \dot{\geq} \prod_{i=1}^{k} \rho^{\left(\frac{rT}{\kappa}-\frac{3}{2}\delta-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})\right)} \tag{4.69}$$

$$= \rho^{\sum_{i=1}^{k}\left(\frac{rT}{\kappa}-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+\right)-\frac{3}{2}k\delta}, \tag{4.70}$$

and setting $k = 2qT$ we have that

$$N_{2qT} = \rho^{\left(\sum_{i=1}^{2qT}\left(\frac{rT}{\kappa}-\frac{1}{2}(1-\mu^*_{l_{2T}(i)})^+\right)-3qT\delta\right)} \tag{4.71}$$

$$= \rho^{\left(T\sum_{j=1}^{q}\left(\frac{rT}{\kappa}-(1-\mu^*_j)^+\right)-3qT\delta\right)} \tag{4.72}$$

$$= \rho^{(\bar{c}_{r-ld}(r)-3qT\delta)}, \tag{4.73}$$

where the last equality follows from (4.58). Consequently

$$N_{SD} \geq N_{2qT} \doteq \rho^{\bar{c}_{r-ld}(r)-3qT\delta},$$

for small $\delta > 0$. Given that δ can be chosen arbitrarily small, and given that events Ω_1 and Ω_2 occur, then the number of nodes visited by the SD at layer $2qT$ is arbitrarily close to the upper bound of $\rho^{\bar{c}_{r-ld}(r)}$.

Now we just have to prove that $-\lim_{\rho\to\infty}\dfrac{\mathrm{P}\left(N \doteq \rho^{\bar{c}_{r-ld}(r)-3qT\delta}\right)}{\log\rho} < d_L(r)$. Toward this we note that as (4.59) and (4.60) imply that $N \doteq \rho^{\bar{c}_{r-ld}(r)-3qT\delta}$, it follows that

$$\mathrm{P}\left(N \doteq \rho^{\bar{c}_{r-ld}(r)-3qT\delta}\right) \geq \mathrm{P}\left(\Omega_1 \cap \Omega_2\right) = \mathrm{P}\left(\Omega_1\right)\mathrm{P}\left(\Omega_2\right)$$

where the equality follows from the i.i.d. Rayleigh assumption on the entries in $\mathbf{H}_C$, which makes the singular values of $\mathbf{H}_C^H\mathbf{H}_C$ independent of the singular vectors of $\mathbf{H}_C^H\mathbf{H}_C$ [55], and which in turn also implies independence of the singular values of $\mathbf{H}_C^H\mathbf{H}_C$ (event Ω_1) from the singular vectors of $\mathbf{H}_R^H\mathbf{H}_R$ (event Ω_2).

We now turn to [7, Lemma 2] and recall that for the layered codes assumed here, as well as for any full-rate design and some non-random fixed decoding ordering (corresponding to a permutation of the columns of $\boldsymbol{G}$), there exists a unitary matrix $\mathbf{V}_p^{'}$ such that $\mathrm{rank}\Big((\mathbf{I}_T \otimes (\mathbf{V}_p^{'})^H)\mathbf{G}_{|p}\Big) = 2pT$ i.e., that

$$\sigma_1\left((\mathbf{I}_T \otimes (\mathbf{V}_p^{'})^H)\mathbf{G}_{|p}\right) > 0.$$

However, by continuity of singular values [53] it follows for sufficiently small $u > 0$ (cf.(4.60)) that $\mathrm{P}\left(\Omega_2\right) > 0$, which implies [10] that $\mathrm{P}\left(\Omega_2\right) \doteq \rho^0$ as Ω_2 is independent of ρ. This in turn implies that

$$\mathrm{P}\left(N \doteq \rho^{\bar{c}_{r-ld}(r)-3qT\delta}\right) \doteq \mathrm{P}\left(\Omega_1\right). \tag{4.74}$$

10. In light of the fact that event $\mathbf{V}_p^{'}$ has zero measure, what the continuity of eigenvalues guarantees is that we can construct a neighborhood of matrices around $\mathbf{V}_p^{'}$ which are full rank, and which have a non zero measure. We also note that the matrices $\mathbf{V}_p^{'}$ can be created recursively, starting from a single matrix $\mathbf{V}_{n_T}^{'}$.

With Ω_1 being an open set, we have that

$$\begin{aligned} -\lim_{\rho\to\infty}\frac{\mathrm{P}(\Omega_1)}{\log\rho} &\leq \inf_{\boldsymbol{\mu}\in\Omega_1} I(\boldsymbol{\mu}), \\ &= \sum_{j=1}^{q}(n_R - n_T| + 2j - 1)(\mu_j^* - 2\delta), \\ &= d_L(r) - 2(|n_T - n_R| + q)q\delta, \\ &< d_L(r), \end{aligned} \tag{4.75}$$

where the above follows from the monotonicity of the rate function

$$I(\boldsymbol{\mu}) = \sum_{j=1}^{n_T}(n_R - n_T + 2j - 1)\mu_i,$$

evaluated at

$$\{\mu_1^* - 2\delta\cdots,\mu_q^* - 2\delta, 0,\cdots,0\} = \arg\inf_{\boldsymbol{\mu}\in\Omega_1} I(\boldsymbol{\mu}),$$

and also follows from the fact that, by definition, $I(\boldsymbol{\mu}^*) = d_L(r)$.

Consequently from (4.74) we have that

$$-\lim_{\rho\to\infty}\frac{\mathrm{P}\left(N \dot{\geq} \rho^{\overline{c}_{r-ld}(r)-3qT\delta}\right)}{\log\rho} < d_L(r), \tag{4.76}$$

and directly from the definition of the complexity exponent, we have that $c_{r-ld}(r) \geq \overline{c}_{r-ld}(r) - 3qT\delta$. As the bound holds for arbitrarily small $\delta > 0$, it follows that $c_{r-ld}(r) = \overline{c}_{r-ld}(r)$. Directly from Corollary 5 which analyzes the ML-based complexity exponent $c_{quasi}(r)$, together with the fact that the ML-based sphere decoder, with or without MMSE preprocessing, shares the same upper bound $\overline{c}_{r-ld}(r)$ as the MMSE-preprocessed lattice decoder, gives that $c_{quasi}(r) = \overline{c}_{r-ld}(r)$, which in turns implies that

$$c_{r-ld}(r) = c_{quasi}(r).$$

This establishes Theorem 10. □

Appendix 4B : Proof of Corollary 10d

We can see from (4.54) that, regardless of the fading statistics and the corresponding $I(\boldsymbol{\mu})$, the exponent $\overline{c}_{r-ld}(r)$ is non-decreasing in $d_L(r)$ and is hence maximized when $d_L(r)$ is itself maximized, i.e., it is maximized in the presence of DMT optimal encoding and decoding. Combined with the fact that the corresponding maximization problem does not depend on the fading distribution, other than the natural fact that its tail must vanish

exponentially fast, results in the fact that, for any full-rate code and statistical characterization of the channel, the complexity of MMSE-preprocessed lattice SD is universally upper bounded as (cf. Corollary 4a)

$$\frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right). \tag{4.77}$$

This proves Corollary 10d. □

Appendix 4C : Proof of Lemma 2

For $\mathbf{R}_r^H \mathbf{R}_r = \mathbf{M}_r^H \mathbf{M}_r + \alpha_r^2 \mathbf{I}$ (cf. (4.10)) [11], it follows by the bounded orthogonality defect of LLL reduced bases that there is a constant $K_\kappa > 0$ independent of $\mathbf{R}_r$ and ρ, for which (cf. [64] and the proof in [69])

$$\sigma_{max}(\tilde{\mathbf{R}}_r^{-1}) \leq \frac{K_\kappa}{\lambda(\mathbf{R}_r)} \tag{4.78}$$

where

$$\lambda(\mathbf{R}_r) \triangleq \min_{\mathbf{c} \in \mathbb{Z}^\kappa \setminus \mathbf{0}} \|\mathbf{R}_r \mathbf{c}\| \tag{4.79}$$

denotes the shortest vector in the lattice generated by $\mathbf{R}_r$. As a result we have that

$$\sigma_{min}(\tilde{\mathbf{R}}_r) \geq \frac{\lambda(\mathbf{R}_r)}{K_\kappa}. \tag{4.80}$$

Looking to lower bound $\sigma_{min}(\tilde{\mathbf{R}}_r)$, we seek a bound on $\lambda(\mathbf{R}_r)$. Towards this let $r' = r - \gamma$ for some $r \geq \gamma > 0$, in which case for $\mathbf{s}$ being the transmitted symbol vector, and for any $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ such that $\hat{\mathbf{s}} \neq \mathbf{s}$, it follows that

$$\begin{aligned} \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| &= \|(\mathbf{r} - \mathbf{R}_{r'}\mathbf{s}) + \mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| \\ &\leq \|(\mathbf{r} - \mathbf{R}_{r'}\mathbf{s})\| + \|\mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| \end{aligned} \tag{4.81}$$

and

$$\begin{aligned} \|\mathbf{R}_{r'}(\mathbf{s} - \hat{\mathbf{s}})\| &\geq \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| - \|(\mathbf{r} - \mathbf{R}_{r'}\mathbf{s})\| \\ &= \|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\| - \|\mathbf{w}\|. \end{aligned} \tag{4.82}$$

From (4.82) it is clear that to find a lower bound on $\lambda(\mathbf{R}_{r'})$, we need to lower bound $\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|$ for all $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ and upper bound $\|\mathbf{w}\|$. Let us, for now, assume that $\|\mathbf{w}\|^2 \leq \rho^b$. To lower bound $\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\|$, we draw from the

11. Note the transition to the notation reflecting the dependence of $\mathbf{R}$ on r.

equivalence of MMSE preprocessing and the regularized metric (cf. equation (45) in [57]), and rewrite

$$\|\mathbf{r}-\mathbf{R}_{r'}\hat{\mathbf{s}}\|^2 = \|\mathbf{y}-\mathbf{M}_{r'}\hat{\mathbf{s}}\|^2 + \alpha_{r'}^2\|\hat{\mathbf{s}}\|^2 - c, \tag{4.83}$$

where $c \triangleq \mathbf{y}^H[\mathbf{I}-\mathbf{M}_{r'}^H(\mathbf{M}_{r'}^H\mathbf{M}_{r'}+\alpha_{r'}^2\mathbf{I})^{-1}\mathbf{M}_{r'}]\mathbf{y} \geq 0$. We now note that for $\hat{\mathbf{s}}=\mathbf{s}$ then $\|\mathbf{y}-\mathbf{M}_{r'}\mathbf{s}\|^2+\alpha_{r'}^2\|\mathbf{s}\|^2 = \|\mathbf{w}\|^2 \leq \rho^b$, and since the left hand side of (4.83) cannot be negative, and furthermore given that c is independent of $\hat{\mathbf{s}}$, we conclude that $c \leq \rho^b$.

We will now proceed to lower bound $\|\mathbf{y}-\mathbf{M}_{r'}\hat{\mathbf{s}}\|^2+\alpha_{r'}^2\|\hat{\mathbf{s}}\|^2$ and then use (4.83) to lower bound $\|\mathbf{r}-\mathbf{R}_{r'}\hat{\mathbf{s}}\|$. Towards lower bounding $\|\mathbf{y}-\mathbf{M}_{r'}\hat{\mathbf{s}}\|^2+\alpha_{r'}^2\|\hat{\mathbf{s}}\|^2$ we draw from Theorem 1 in [57] and we let $\mathcal{B}$ be the spherical region given by

$$\mathcal{B} \triangleq \{d \in \mathbb{R}^\kappa | \, \|\mathbf{d}\|^2 \leq \Gamma^2\}$$

where the radius $\Gamma > 0$ is independent of ρ and is chosen so that $\mathbf{d}_1+\mathbf{d}_2 \in \mathcal{R}$ for any $\mathbf{d}_1, \mathbf{d}_2 \in \mathcal{B}$. The existence of the set $\mathcal{B}$ follows by the assumption that $\mathbf{0}$ is contained in the interior of $\mathcal{R}$. Now let

$$\nu_{r'} \triangleq \min_{\mathbf{d} \in \rho^{\frac{r'T}{\kappa}}\mathcal{B}\cap\mathbb{Z}^\kappa : \mathbf{d}\neq\mathbf{0}} \frac{1}{4}\|\mathbf{M}_{r'}\mathbf{d}\|^2,$$

and for given $\gamma > \zeta > 0$ choose $b > 0$ such that

$$\frac{2\zeta T}{\kappa} > b > 0.$$

This may clearly be done for arbitrary $\zeta > 0$. We will in the following temporarily assume that $\nu_{r'+\zeta} \geq 1$ and prove that, together with $\|\mathbf{w}\|^2 \leq \rho^b$, the two conditions are sufficient for $\lambda(\tilde{\mathbf{R}}_{r'}) \dot{\geq} \rho^{\frac{\zeta T}{\kappa}}$ to hold.

In order to bound the metric for $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ where $\hat{\mathbf{s}} \neq \mathbf{s}$, we note that $\nu_{r'+\zeta} \geq 1$ implies that $\forall \mathbf{d} \in \rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B}\cap\mathbb{Z}^\kappa, \mathbf{d} \neq \mathbf{0}$ it is the case that

$$\begin{aligned}
\frac{1}{4}\left\|\mathbf{M}_{r'+\zeta}\mathbf{d}\right\|^2 &\geq 1 \\
\frac{1}{4}\left\|\rho^{\frac{1}{2}-\frac{(r'+\zeta)T}{\kappa}}\mathbf{HGd}\right\|^2 &\overset{(a)}{\geq} 1 \\
\frac{1}{4}\left\|\rho^{\frac{1}{2}-\frac{r'T}{\kappa}}\mathbf{HGd}\right\|^2 &\geq \rho^{\frac{2\zeta T}{\kappa}}
\end{aligned}$$

where (a) follows from the fact that $\mathbf{M}_r = \rho^{\frac{1}{2}-\frac{rT}{\kappa}}\mathbf{HG}$. Consequently

$$\frac{1}{4}\|\mathbf{M}_{r'}\mathbf{d}\|^2 \geq \rho^{\frac{2\zeta T}{\kappa}}, \ \forall \mathbf{d} \in \rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B}\cap\mathbb{Z}^\kappa, \mathbf{d} \neq \mathbf{0}. \tag{4.84}$$

As $\mathcal{R}$ is bounded, and as $\zeta > 0$, it holds that $\mathcal{R} \subset \frac{1}{2}\rho^{\frac{\zeta T}{\kappa}}\mathcal{B}$ for all $\rho \geq \rho_1$, for a sufficiently large ρ_1. This implies that $\mathbf{s} \in \frac{1}{2}\rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B}$ for $\rho \geq \rho_1$ since $\mathbf{s} \in \rho^{\frac{r'T}{\kappa}}\mathcal{R}$.

For $\mathbf{s}, \mathbf{d} \in \frac{1}{2}\rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B} \cap \mathbb{Z}^\kappa$, there exists an $\hat{\mathbf{s}} \in \rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B} \cap \mathbb{Z}^\kappa$, $\hat{\mathbf{s}} \neq \mathbf{s}$, such that $\hat{\mathbf{s}} = \mathbf{d} + \mathbf{s}$. Hence for any $\hat{\mathbf{s}} \in \rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B} \cap \mathbb{Z}^\kappa$, we have from (4.84) that

$$\frac{1}{4}\left\|\mathbf{M}_{r'}(\hat{\mathbf{s}} - \mathbf{s})\right\|^2 = \frac{1}{4}\left\|\mathbf{M}_{r'}\mathbf{d}\right\|^2 \geq \rho^{\frac{2\zeta T}{\kappa}}. \tag{4.85}$$

As $\|\mathbf{w}\|^2 \leq \rho^b$, it follows that $\frac{1}{4}\left\|\mathbf{M}_{r'}\mathbf{d}\right\|^2 \geq \|\mathbf{w}\|^2$ for large ρ, and that

$$\left\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\right\|^2 = \left\|\mathbf{M}_{r'}(\mathbf{s} - \hat{\mathbf{s}}) + \mathbf{w}\right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \tag{4.86}$$

Consequently

$$\left\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\right\|^2 + \alpha_{r'}^2\left\|\hat{\mathbf{s}}\right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \tag{4.87}$$

On the other hand if $\hat{\mathbf{s}} \notin \rho^{\frac{(r'+\zeta)T}{\kappa}}\mathcal{B}$, then by definition of $\mathcal{B}$ we have that $\alpha_{r'}^2\left\|\hat{\mathbf{s}}\right\|^2 \geq \frac{1}{4}\Gamma^2\rho^{\frac{2\zeta T}{\kappa}}$, and consequently that

$$\left\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\right\|^2 + \alpha_{r'}^2\left\|\hat{\mathbf{s}}\right\|^2 \geq \frac{1}{4}\Gamma^2\rho^{\frac{2\zeta T}{\kappa}}. \tag{4.88}$$

From (4.87) and (4.88) we then conclude that

$$\left\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\right\|^2 + \alpha_{r'}^2\left\|\hat{\mathbf{s}}\right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \tag{4.89}$$

Given (4.87) and (4.89), for any $\hat{\mathbf{s}} \in \mathbb{Z}^\kappa$ such that $\hat{\mathbf{s}} \neq \mathbf{s}$, it is the case that $\left\|\mathbf{y} - \mathbf{M}_{r'}\hat{\mathbf{s}}\right\|^2 + \alpha_{r'}^2\left\|\hat{\mathbf{s}}\right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}$, which combined with $c \leq \rho^b$ allows for (4.83) to give that

$$\left\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\right\|^2 \dot{\geq} \rho^{\frac{2\zeta T}{\kappa}}. \tag{4.90}$$

Applying (4.79) and (4.82), we have

$$\begin{aligned}\lambda(\mathbf{R}_{r'}) &\geq \left\|\mathbf{r} - \mathbf{R}_{r'}\hat{\mathbf{s}}\right\| - \|\mathbf{w}\| \\ &\dot{\geq} \rho^{\frac{\zeta T}{\kappa}} - \rho^{\frac{b}{2}} \\ &\doteq \rho^{\frac{\zeta T}{\kappa}}\end{aligned} \tag{4.91}$$

where the exponential inequality follows from (4.90). Furthermore we know that

$$\lambda(\mathbf{R}_r) = \rho^{\frac{-\gamma T}{\kappa}}\lambda(\mathbf{R}_{r'}) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}} \tag{4.92}$$

where $\epsilon = \gamma - \zeta$, $r \geq \epsilon > 0$, and from (4.80) and (4.92) it follows that $\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}$.

We now note that the above implies that for $\nu_{r'+\zeta} \geq 1$ and $\|\mathbf{w}\|^2 \leq \rho^b$ then $\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{\geq} \rho^{\frac{-\epsilon T}{\kappa}}$, and thus applying the union bound yields

$$\begin{aligned} \mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) &= \mathrm{P}\left((\nu_{r'+\zeta} < 1) \cup (\|\mathbf{w}\|^2 > \rho^b)\right) \\ &\leq \mathrm{P}\left(\nu_{r'+\zeta} < 1\right) + \mathrm{P}\left(\|\mathbf{w}\|^2 > \rho^b\right). \end{aligned}$$

We know from the exponential tail of the Gaussian distribution that $\mathrm{P}\left(\|\mathbf{w}\|^2 > \rho^b\right) \doteq \rho^{-\infty}$ and from [57, Lemma 1] that $\mathrm{P}\left(\nu_{r'+\zeta} < 1\right) \dot{\leq} \rho^{-d_{ML}(r'+\zeta)}$. Hence

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_r) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_{ML}(r-\epsilon)}$$

for all $r \geq \epsilon > 0$.

The association with the singular values

$$\sigma_1(\tilde{\mathbf{R}}_{r,k}) \leq \cdots \leq \sigma_k(\tilde{\mathbf{R}}_{r,k})$$

is made using the interlacing property of singular values of sub-matrices, which gives that

$$\sigma_i(\tilde{\mathbf{R}}_{r,k}) \geq \sigma_i(\tilde{\mathbf{R}}_r),\ i \leq k = 1, \cdots, \kappa, \tag{4.93}$$

and for $k = 1, \cdots, \kappa$, that

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_{r,k}) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_{ML}(r-\epsilon)}.$$

Finally from the DMT optimality of the exact implementation of the regularized lattice decoder [56], [57], we have that

$$\mathrm{P}\left(\sigma_{min}(\tilde{\mathbf{R}}_{r,k}) \dot{<} \rho^{\frac{-\epsilon T}{\kappa}}\right) \dot{\leq} \rho^{-d_L(r-\epsilon)}.$$

This proves Lemma 2.□

Appendix 4D : Proof of Lemma 3

For a search radius that grows as $\xi = \sqrt{z \log \rho} \doteq \rho^0$, we first prove that for $z > d_L(r)$

$$\mathrm{P}\left(\|\mathbf{w}'\|^2 > \xi^2\right) \dot{<} \rho^{-d_L(r)}.$$

From (4.12) and (4.46) the equivalent term for MMSE-preprocessed lattice decoding is given by

$$\begin{aligned} \mathbf{w}' &= -\alpha_r^2 \mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w} \\ &= -\alpha_r \mathbf{Q}_2^H\mathbf{s} + \mathbf{Q}_1^H\mathbf{w}. \end{aligned} \tag{4.94}$$

Consequently we calculate

$$\begin{aligned}
\mathrm{P}\left(\|\mathbf{w}^{'}\| > \xi\right) &\leq \mathrm{P}\left(\{\| - \alpha_r \mathbf{Q}_2^H \mathbf{s}\| + \|\mathbf{Q}_1^H \mathbf{w}\|\} > \xi\right) \\
&\overset{(a)}{=} \mathrm{P}\left(\left\{\| - \alpha_r \mathbf{Q}^H \begin{bmatrix} \mathbf{0} \\ \mathbf{s} \end{bmatrix} \| + \|\mathbf{Q}^H \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} \|\right\} > \xi\right) \\
&\leq \mathrm{P}\left(\left\{\sup_{\mathbf{s}\in\mathbb{S}_r^{\kappa}} \| - \alpha_r \mathbf{s}\| + \|\mathbf{w}\|\right\} > \xi\right) \\
&\overset{(b)}{\leq} \mathrm{P}\left(\{K + \|\mathbf{w}\|\} > \xi\right) \\
&= \mathrm{P}\left(\|\mathbf{w}\| > \{\xi - K\}\right) \\
&\overset{(c)}{\leq} \mathrm{P}\left(\|\mathbf{w}\|^2 > z_1 \log \rho\right) \\
&\dot{<} \quad \rho^{-d_L(r)}
\end{aligned} \tag{4.95}$$

where (a) follows from the MMSE-preprocessed equivalent channel representation in (4.46), where the inequality in (b) follows for some fixed $K > 0$ independent of ρ such that $K \geq \sup_{\mathbf{s}\in\mathbb{S}_r^{\kappa}} \| - \alpha_r \mathbf{s}\|$, where the inequality in (c) follows for some arbitrary $z > z_1 > 0$ independent of ρ such that $(\xi - K)^2 \geq z_1 \log \rho$ and where the last asymptotic inequality follows for some $z_1 > d_L(r) > 0$. Consequently

$$\mathrm{P}\left(\|\mathbf{w}^{''}\| > \xi\right) = \mathrm{P}\left(\|\tilde{\mathbf{Q}}^H \mathbf{w}^{'}\| > \xi\right) \dot{<} \rho^{-d_L(r)}$$

and as a result

$$\lim_{\rho\to\infty} \frac{\mathrm{P}\left(\|\mathbf{w}^{''}\| > \xi\right)}{\mathrm{P}\left(\hat{\mathbf{s}}_{r-ld} \neq \mathbf{s}\right)} = 0.$$

This proves Lemma 3. □

Chapter 5

Feedback-Aided Complexity of ML and Lattice Decoding

5.1 Introduction and MIMO-ARQ signaling

The work in the previous chapters has focused on deriving complexity exponents for the case where there is no feedback information sent to the transmitter (no CSIT). Such feedback is known to improve the rate-reliability performance in MIMO systems. What we show here is that feedback also changes the required computational complexity. We focus on two fundamental questions. The first question asks what is the complexity savings that feedback provides for a given fixed rate-reliability performance, and the second question asks what is the complexity costs of achieving the full rate-reliability benefits of feedback. The analysis and the constructed feedback schemes tell us how to properly utilize a finite number of feedback bits to alleviate the adverse effects of computational constraints, as those seen in the derived rate-reliability-complexity tradeoffs of the previous chapters. Emphasis is placed on MIMO-ARQ feedback schemes, although we do also consider feedback with antenna selection.

In the MIMO-ARQ setting (Fig. 5.1), communication of a certain message is divided into rounds, each of duration T. At the end of each ARQ round, the receiver feeds back to the transmitter a single bit (known as an ACK or a NACK response). An ACK implies that the receiver has successfully decoded, in which case the transmitter moves on to the next information message. On the other hand, upon receiving a NACK, the transmitter retransmits a (possibly different) encoded version of the same message. There can be up to L such ARQ rounds.

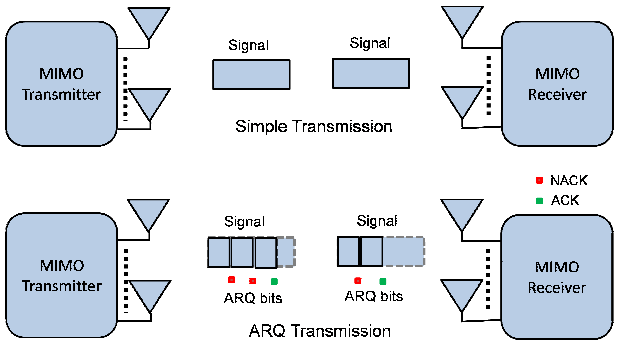


FIGURE 5.1 – Schematic for MIMO-ARQ system

5.1.1 Reliability and complexity implications of feedback : from DMT to the improved DMD

The work in [10] has quantified the benefits of such a MIMO-ARQ scheme, and has shown that it can increase the rate-reliability performance, elevating it from the optimal DMT $d^*(r)$ of the MIMO channel without feedback, up to a much improved [1] $d^*(\frac{r}{L})$. This optimal performance $d^*(r/L)$ is generally referred to as the optimal DMD tradeoff (diversity multiplexing delay tradeoff, cf. [10]). Figure 5.2 illustrates these performance gains for the case of $L = 4$-round ARQ over the Rayleigh fading 4×4 MIMO channel. A practical implication of such feedback schemes is that in the presence of a moderate number of transmit and receive antennas, and a moderate number of $(L-1)$ bits of feedback, one can enjoy near ergodic behavior (i.e., very substantial diversity), even for rates close to the ergodic capacity (high r). Such gains may entail though increased complexity. While the complexity

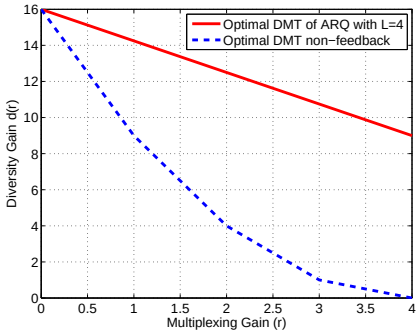


FIGURE 5.2 – Optimal DMT and DMD for 4×4 MIMO.

implications will be clarified later on, loosely speaking, to achieve the optimal DMD $d^*(r/L)$ for specific settings, the last round decoding must be

1. We focus on the quasi-static case and on the case of no power adaptation.

sufficiently reliable, something that can happen only if the associated lattice designs encountered by the decoder in the last round are themselves sufficiently strong, which in turn implies even more heavily packed and complex coding schemes for the intermediate rounds (cf. [36]).

The other side of the coin is that, if one were willing to sacrifice some of the feedback reliability gains, this aforementioned complexity may be reduced even below the original complexity corresponding to the case of no feedback, i.e., below the complexity corresponding quantified by the complexity exponents derived in Theorem 5 and Corollary 5a.

With the above in mind, we seek to understand the computational costs that allow for a) the optimal DMT $d^*(r)$ and b) the optimal DMD $d^*(r/L)$, both with the assistance of L-rounds of ARQ feedback. Specifically we will derive bounds on the minimum complexity exponent required by ML and lattice based SD to achieve the optimal DMT $d^*(r)$ and the optimal DMD $(d^*(r/L))$, where this complexity is minimized over all lattice designs (which must vary accordingly depending on the setting), all ARQ schemes ($L \leq n_{\mathrm{T}}$), all policies of computational halting, as well as all policies on decoding ordering. The results are presented for ML-based (sphere) decoding as well as (MMSE-preprocessed) lattice decoding. The derivations focus on ML-based decoding, but extend automatically to the lattice decoding case, as we have learned from Chapter 4. The validity of the presented results depends on the existence of actual schemes that meet them. These schemes will be here provided, together with the associated lattice designs, decoders, as well as halting and ordering polices.

5.1.2 MIMO-ARQ signaling

We here present the general $n_T \times n_R$ MIMO-ARQ signaling setting, and focus on the details which are necessary for our exposition. For further understanding of the MIMO-ARQ channel, the reader is referred to [10], see also [36].

Under ARQ signaling, each message is associated to a unique block $[\mathbf{X}_C^1\, \mathbf{X}_C^2 \cdots \mathbf{X}_C^L]$ of signaling matrices, where each $\mathbf{X}_C^i \in \mathbb{C}^{n_T \times T}$, $i = 1, \cdots, L$, corresponds to the $n_{\mathrm{T}} \times T$ matrix of signals sent during the ith round. The accumulated code matrix at the end of round ℓ, $\ell = 1, \cdots, L$, takes the form $\mathbf{X}_C^{ARQ,\ell} = [\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^\ell] \in \mathbb{C}^{n_T \times \ell T}$. We note that the signals $\mathbf{X}_C^{ARQ,L}$ are drawn from a lattice design that ensures unique decodability at every round [2].

In the quasi-static case of interest, the received signal accumulated at the end of the ℓ-th round takes the form

$$\mathbf{Y}_C^\ell = \sqrt{\rho}\mathbf{H}_C\mathbf{X}_C^{ARQ,\ell} + \mathbf{W}_C^\ell, \quad \ell = 1, \cdots, L, \tag{5.1}$$

2. Loosely speaking, unique decodability means that, for any $\ell = 1, \cdots, L$, the corresponding $\mathbf{X}_C^{ARQ,\ell}$ carries all bits of information.

where $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$.

We proceed with quantifying the complexity reductions due to ARQ feedback. After that Section 5.3 will consider the complexity that allows for harvesting the full rate-reliability benefits of ARQ feedback, and Section 5.4 presents similar results for the antenna selection case. The rest of the sections are dedicated to proofs of the main results.

5.2 Complexity reduction using ARQ feedback

We here seek to analyze the complexity reductions due to MIMO ARQ feedback. Specifically for $d^*(r)$ denoting the optimal DMT of the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ MIMO channel in the absence of feedback, we here seek to describe the complexity exponent required to meet the same $d^*(r)$ with the assistance now of an L-round ARQ scheme. This exponent is to be compared with the exponent in Theorem 5 and Corollary 5a which describes the complexity that guarantees the same rate-reliability performance ($d^*(r)$) but without feedback.

The following holds for the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$) i.i.d. regular fading [3] MIMO channel.

All the presented results hold for ML-based decoding as well as MMSE-preprocessed lattice decoding.

Theorem 13. *Let $c(r)$ be the minimum complexity exponent required to achieve $d^*(r)$, minimized over all lattice designs, all ARQ schemes with $L \leq n_{\mathrm{T}}$ rounds of ARQ, all halting policies and all decoding order policies. Then*

$$c(r) \leq \overline{c}_{red}(r) \triangleq \frac{1}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right],$$

which is a piecewise linear function that, for integer r, takes the form

$$\overline{c}_{red}(r) = \frac{1}{n_T} r(n_T - r), \text{ for } r = 0, 1, \cdots, n_T.$$

The proof of the above theorem will be presented later on, together with the proofs for the upcoming Propositions 3 and 4, and it will include the derivation of the upper bound, and the constructive achievement of this bound which is presented in Propositions 3,4. The constructive part of the proof is based on designing ARQ schemes and implementations (lattice designs and halting policies) that meet the bound. We proceed with these propositions

3. The i.i.d. regular fading statistics satisfy the general set of conditions as described in [70], where a) the near-zero behavior of the fading coefficients h is bounded in probability as $c_1|h|^t \leq p(h) \leq c_2|h|^t$ for some positive and finite c_1, c_2 and t, where b) the tail behavior of h is bounded in probability as $p(h) \leq c_2 e^{-b|h|^\beta}$ for some positive and finite c_2, b and β, and where c) $p(h)$ is upper bounded by a constant K.

where we identify cases for which the above complexity bound suffices to achieve $d^*(r)$ with the help of feedback.

An important aspect in ARQ schemes is knowing when to decode and when not to decode across the different rounds. Towards this we have the following definition.

Definition 1 (Aggressive intermediate halting policies)**.** *We define* aggressive intermediate halting policies *to be the family of policies that halt decoding in the first round whenever the minimum singular value of the channel scales as $\rho^{-\epsilon}$ for some $\epsilon > 0$, which do not decode in the second to the L-1 round, and which decode at the last round iff a) they have not decoded in the first round and b) the channel is not in outage with respect to the effective rate of ARQ scheme.*

Furthermore we will henceforth use the term *minimum delay ARQ scheme* to refer to any ARQ scheme with $T = 1$. We will also use the term *ARQ-compatible, minimum delay, NVD, rate-1 lattice designs* to refer to the family of $n_{\mathrm{T}} \times n_{\mathrm{T}}$ lattice designs $\mathbf{X}_C^{ARQ,L}$ with $\kappa = 2n_{\mathrm{T}}$, with non-vanishing determinant (NVD) [4] for $r \leq 1$, and with all the information appearing in all rounds.

Proposition 3. *A minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$, irrespective of the ARQ-compatible, minimum delay, NVD, rate-1 lattice design, for any aggressive intermediate halting policy, and any sphere decoding order policy.*

The following describes a very simple MIMO ARQ coding implementation that achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$. The proof of this proposition will appear later on, and is crucial in the achievability part of the proof of Theorem 13.

Proposition 4. *The minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds, implemented with any aggressive intermediate halting policy, any sphere decoding order policy, and a rate-1 lattice design $\mathbf{X}_C^{ARQ,L}$ drawn from the center of perfect codes (cf. [13, 25])* [5]*, achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$.*

Theorem 13 has quantified the computational reserves that are sufficient to achieve DMT optimality. These computational reserves can be seen to be

4. A code has a non-vanishing determinant if, without power normalization, there is a lower bound on the minimum determinant that does not depend on the constellation size. The determinant of any non-normalized difference matrix is lower bounded by a constant independent of ρ (see [25]).

5. For general lattice designs derived from cyclic division algebra (CDA) (cf. [13, 25]), $\mathbb{F}$ and $\mathbb{L}$ are number fields, with $\mathbb{L}$ a finite, cyclic Galois extension of $\mathbb{F}$ of degree n. Let σ denote a generator of the Galois group $\mathrm{Gal}(\mathbb{L}/\mathbb{F})$. Let z be an indeterminate satisfying $lz = z\sigma(l)$, $\forall\ l \in \mathbb{L}$ and $z^n = \gamma$ for some non-norm element $\gamma \in \mathbb{F}^*$. Then the set of all elements of the form $\sum_{i=0}^{n-1} z^i l_i$ forms a CDA $D(\mathbb{L}/\mathbb{F}, \sigma, \gamma)$ with center $\mathbb{F}$ and maximal subfield $\mathbb{L}$. The mentioned codes are limited in the center of the division algebra and take

smaller than those required to achieve the same optimal DMT $d^*(r)$ without feedback, and in the presence of any known minimum delay DMT optimal designs. Specifically, in the absence of ARQ, the latter designs have been shown in Theorem 5 and Corollary 5a to require a complexity exponent of

$$c(r) = r(n_T - r), \tag{5.3}$$

(for integer $r = 0, 1, \cdots, n_T$), whereas we have just seen that, for example, in the presence of n_{T} rounds of ARQ, the same DMT is achieved with a much reduced

$$c(r) \leq \frac{1}{n_T} r(n_T - r).$$

We proceed with a few examples.

Example 9 (Corresponding to Theorem 13 and Proposition4)**.** *For the general $n_{\mathrm{T}} \times n_{\mathrm{R}}$ setting with $n_{\mathrm{R}} \geq n_{\mathrm{T}}$, and for $r = n_{\mathrm{T}}/2$, the computational resources required to achieve the optimal $d^*(r)$ with existing DMT optimal (minimum delay) non-feedback schemes (cf. Corollary 5a), scales as*

$$N_{\max} \doteq \rho^{n_{\mathrm{T}}^2/4} \doteq 2^{Rn_{\mathrm{T}}/2},$$

whereas the feedback aided complexity required by the feedback scheme in Proposition4 scales as

$$N_{\max} \doteq \rho^{n_{\mathrm{T}}/4} \doteq 2^{R/2}.$$

Generally, given a rate that scales linearly with $\min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$), in the absence of feedback the complexity exponent of achieving $d^*(r)$ scales with n_{T}^2, whereas the feedback aided complexity exponent scales with n_{T}.

Example 10. *Figure 5.3 considers the case of $n_{\mathrm{T}} = 4 \leq n_{\mathrm{R}}$ and Rayleigh fading, and compares the above complexity upper bound in the presence of feedback (L-rounds, minimum delay), to the equivalent complexity exponent in* (5.3) *of achieving the same DMT optimal $d^*(r)$ without ARQ feedback (Perfect codes and natural, fixed decoding ordering).*

Example 11. *Figure 5.4 considers, for the same $4 \times n_{\mathrm{R}}$ channel $n_{\mathrm{R}} \geq n_{\mathrm{T}}$, the joint performance-complexity measures*

$$\Gamma(r) = d(r) - c(r)$$

for the above feedback aided and non-aided schemes of the previous example. Naturally the feedback aided scheme exhibits a uniformly increased measure.

---

the simple form

$$\mathbf{X}_C^{ARQ,L} = \begin{bmatrix} f_0 & \gamma f_{n_T-1} & \cdots & \gamma f_1 \\ f_1 & f_0 & \cdots & \gamma f_2 \\ f_2 & f_1 & \cdots & \gamma f_3 \\ \vdots & \vdots & \ddots & \vdots \\ f_{n_T-1} & f_{n_T-2} & \cdots & f_0 \end{bmatrix} \in \mathbb{C}^{n_T \times n_T}, \tag{5.2}$$

where f_i belong to the QAM constellation.

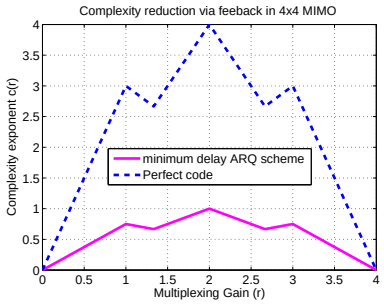


FIGURE 5.3 – Complexity reduction with minimum delay ARQ schemes.

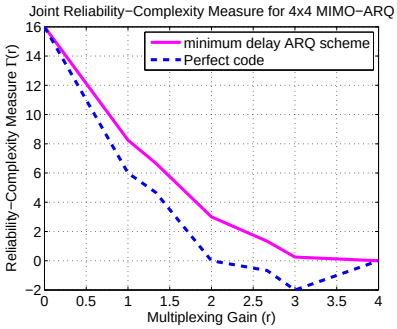


FIGURE 5.4 – Joint performance-complexity measures with minimum delay ARQ schemes.

5.2.1 Feedback reduction for asymmetric channels : $n_{\mathrm{R}} \leq n_{\mathrm{T}}$

We now consider the case of $n_{\mathrm{R}} \leq n_{\mathrm{T}}$, and specifically the case where $n_{\mathrm{R}} | n_{\mathrm{T}}$ (i.e., n_{T} is an integer multiple of n_{R}), to observe again how simple implementations offer substantial reductions in complexity. In terms of statistics, the results hold for any i.i.d. regular fading distribution. As before the results apply to ML-based decoding as well as to MMSE-preprocessed lattice decoding.

Theorem 14. *In the MIMO ARQ channel with $n_{\mathrm{R}} | n_{\mathrm{T}}$, the minimum complexity exponent $c(r)$ required to achieve $d^*(r)$, minimized over all lattice designs, all halting policies, and all minimum delay ARQ schemes with $L \leq n_{\mathrm{T}}$ rounds of ARQ, is bounded as*

$$c(r) \leq \overline{c}_{red}(r) \triangleq = \frac{1}{n_R} \left[r(n_R - \lfloor r \rfloor - 1) + (n_R \lfloor r \rfloor - r(n_R - 1))^+ \right],$$

which is a piecewise linear function that, for integer r, takes the form

$$\overline{c}_{red}(r) = \frac{1}{n_{\mathrm{R}}} r(n_{\mathrm{R}} - r), \ \textit{for } r = 0, 1, \cdots, n_{\mathrm{R}}.$$

Applying as the constructive part of the proof of the above theorem, the following describes a very simple MIMO ARQ block-diagonal repetition coding implementation that achieves $d^*(r)$ with a much reduced $c(r) \leq \overline{c}_{red}(r)$.

Proposition 5. *A minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds, implemented with any aggressive intermediate halting policy, any sphere decoding order policy, and a rate-$\frac{n_{\mathrm{R}}}{n_{\mathrm{T}}}$ block-diagonal repetition lattice design $\mathbf{X}_C^{ARQ,L}$ where the (rate-1) block component code is drawn from the center of $n_{\mathrm{R}} \times n_{\mathrm{R}}$ perfect codes, achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$ from Theorem 14.*

Of interest is the special MISO-ARQ case of $n_{\mathrm{R}} = 1$, where the above described scheme will allow for a zero complexity exponent, and for a complexity that scales as a subpolynomial function of ρ and as a subexponential function of the number of codeword bits and of the rate.

Corollary 14a. *Over the $n_{\mathrm{T}} \times 1$ MISO channel, the minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds, implemented with a rate-$\frac{1}{n_{\mathrm{T}}}$ repetition QAM design $\mathbf{X}_C^{ARQ,L}$, achieves $d^*(r)$ with $c(r) = 0$.*

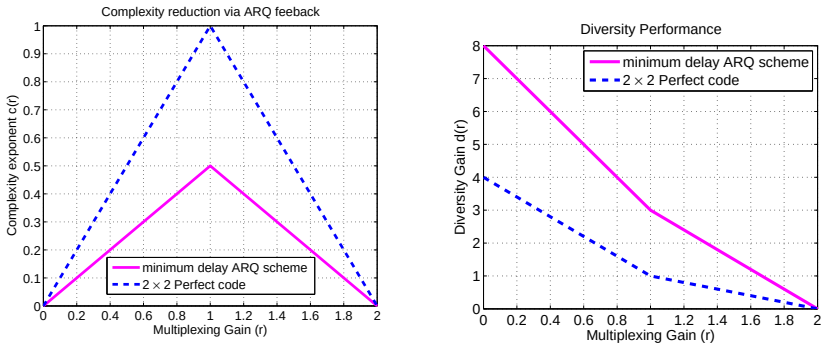
This corollary follows directly from Theorem 14.
We proceed with a few examples.

Example 12 (Corresponding to Theorem 14 and Proposition 5)**.** *For the 4×2 MIMO channel with an $L = 2$-round ARQ, applying a lattice design of the form*

$$\mathbf{X}_C^{ARQ,L} = \begin{bmatrix} f_0 & \gamma f_1 & 0 & 0 \\ f_1 & f_0 & 0 & 0 \\ 0 & 0 & f_0 & \gamma f_1 \\ 0 & 0 & f_1 & f_0 \end{bmatrix} \in \mathbb{C}^{4\times 4},$$

where $f_0, f_1 \sim QAM$, together with an aggressive intermediate halting policy for the first round decoder, and with any sphere decoding ordering policy, can achieve the optimal $d^*(r)$ of the 4×2 channel, and can do so with computational resources of $N_{\max} \doteq \rho^{\overline{c}_{red}(r)}$ flops, which for integer r translates to $N_{\max} \doteq \rho^{\frac{1}{n_{\mathrm{R}}} r(n_{\mathrm{R}}-r)} = \rho^{\frac{r}{2}(2-r)}$.

Example 13 (Corresponding to Theorem 14 and Proposition 5)**.** *Figure 5.5 compares two schemes : the 2×2 MIMO channel (minimum delay, DMT optimal lattice design), and the 4×2 minimum delay MIMO-ARQ channel with $L = n_{\mathrm{T}} = 4$, 3 bits of feedback, and the implementation of Proposition 5. We see a considerably reduced complexity of the feedback aided scheme (Fig. 5.5(a), lower line) which, at the same time, achieves a much higher DMT performance (Fig. 5.5(b), upper line) than its non-feedback counterpart.*



(a) Complexity exponent (b) DMT

FIGURE 5.5 – Complexity reduction $n_R|n_T$ i.i.d. Rayleigh channel with ARQ feedback

Example 14 (Corresponding to Corollary 14a)**.** *For the $n_{\mathrm{T}} \times 1$ MISO case, the corresponding feedback-aided complexity is shown in Corollary 14a to be subexponential in the rate and the number of codeword bits. This dramatic reduction is depicted in Fig. 5.6 which compares it with the corresponding complexity (required to achieve the same $d^*(r)$) in the absence of feedback. This is done for the case of $n_{\mathrm{T}} = 3$.*

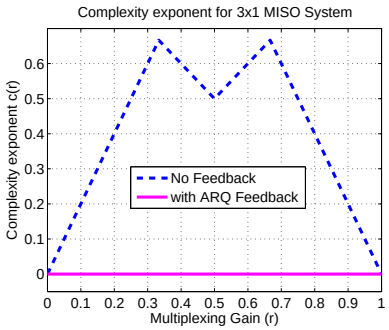


FIGURE 5.6 – Complexity reduction for MISO channel with one bit feedback.

We proceed with the second question of interest.

5.3 Complexity for harvesting the full rate-reliability benefits of feedback

As shown in [10], the presence of L-rounds of ARQ can improve the DMT performance, from the optimal $d^*(r)$ achieved without feedback, to the DMD optimal $d^*(r/L)$ which offers closer to ergodic behavior even at high

multiplexing gains. We here study the complexity implications of achieving this DMD optimal performance $d^*(r/L)$.

The following holds for i.i.d. regular fading statistics. For now we will focus on the case where $n_{\mathrm{R}} \geq n_{\mathrm{T}}$ and the case where $L|n_{\mathrm{T}}$ (i.e., n_{T} is an integer multiple of L). As before all presented results hold for ML-based decoding as well as MMSE-preprocessed lattice decoding.

Theorem 15. *Let $c(r)$ be the minimum complexity exponent required to achieve the optimal L-round MIMO-ARQ DMD $d^*(r/L)$, for any given $L|n_{\mathrm{T}}$, where the complexity is minimized over all lattice designs, all halting policies and all decoding order policies. Then*

$$c(r) \leq \overline{c}_{dmd}(r) \triangleq \frac{1}{L}\left[r\left(n_T - \left\lfloor \frac{r}{L} \right\rfloor - 1\right) + \left(Ln_T \left\lfloor \frac{r}{L} \right\rfloor - r(n_T - 1)\right)^+\right],$$

where $\overline{c}_{dmd}(r)$ is a piecewise linear function that, for r being an integer multiple of L, takes the form

$$\overline{c}_{dmd}(r) = \frac{rn_{\mathrm{T}}}{L^2}\left(L - \frac{r}{n_{\mathrm{T}}}\right).$$

For a clarifying example see Example 15 and Fig. 5.7.

An important note is in order here with respect to the above theorem. We specifically note that one can clearly construct arbitrarily inefficient schemes that meet the optimal $d^*(r/L)$ with a complexity that far exceeds the derived bound on the minimized complexity. For that case, we will be able to say that such a scheme is provably suboptimal in terms of complexity. At the same time, there might exist schemes that do better than the above upper bound. In such a case we will have a tightening of the bound.

A second important note is that the bound is proven valid because there do exist ARQ schemes and implementations (lattice designs and policies) that in fact meet this bound. These designs and policies will be shown in Proposition 6 to provide for DMD optimal ARQ schemes.

These above mentioned constructed schemes happen to incur minimum possible overall delay. This brings us to a third important note. As stated, it is entirely conceivable that the bound in the theorem can be tightened. This is because the bound is met, as we will see in the next proposition, by ARQ schemes that incur an overall delay that remains fixed at a minimum of $TL \leq n_{\mathrm{T}}$. This limitation partially explains why the complexity described by the bound decreases with L, despite the substantial increases in the DMD. In brief, given a fixed LT, an increasing L results in a decreasing T for the first round, which nicely translates to complexity savings.

As a result, in the presence of a choice for any $L|n_{\mathrm{T}}, L \geq 2$, the choice of $L = n_{\mathrm{T}}$ surprisingly minimizes the above bound. For this special case, we have the following.

Corollary 15a. *The minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal DMD $d^*(r/n_{\mathrm{T}})$ of the $(L = n_{\mathrm{T}})$-round ARQ, is upper bounded as*

$$c(r) \leq \overline{c}_{DMD}(r) = \left(1 - \frac{1}{n_T}\right) r. \tag{5.4}$$

From Corollary 15a we can see that $N_{\max} \doteq \rho^{\frac{n_{\mathrm{T}}-1}{n_{\mathrm{T}}}r}$ suffices to achieve the DMD optimal $d^*(r/n_{\mathrm{T}})$. This translates to $N_{\max} \doteq 2^{\frac{n_{\mathrm{T}}-1}{n_{\mathrm{T}}}R}$ for any r, and approaches $N_{\max} \doteq 2^R$ for increased values of n_{T}. From the same corollary it is also easy to see that the presence (and proper utilization) of feedback, while guaranteeing a much increased DMT performance ($d^*(r/L) > d^*(r)$), also results in a provably reduced complexity for any $r \leq \frac{n_T^2}{n_T+1}$, compared to the $c(r)$ required to achieve the reduced $d^*(r)$ without feedback (cf. Corollary 5a). On the other hand, in the small region of $\frac{n_T^2}{n_T+1} \leq r \leq n_{\mathrm{T}}$, this feedback-aided complexity $N_{\max} \doteq 2^R$ does not vanish and exceeds the complexity of the no feedback case which, as r increases, can vanish by employing halting policies that capitalize on the fact that $d^*(r)$ also vanishes. For a clarifying example see Example 16 and Fig. 5.8.

With respect to the achievability of Theorem 15 and Corollary 15a, and under the same conditions as above, we identify cases for which the derived complexity bounds suffice to achieve $d^*(r/L)$. The presented schemes also happen to incur the minimum possible overall delay, minimized over all delays that could allow for DMD optimality. The implementation considers ARQ-compatible, minimum-delay, full-rate NVD lattice designs (cf. [36]) which exist for all cases of interest here.

Proposition 6. *An L-round $(L|n_{\mathrm{T}})$ MIMO ARQ scheme achieves $d^*(r/L)$ with $c(r) \leq \overline{c}_{dmd}(r)$, for any ARQ-compatible square $(LT = n_{\mathrm{T}})$ NVD lattice design, for a search radius $\xi > \sqrt{d^*(\frac{r}{n_T}) \log \rho}$, for any aggressive intermediate halting policy, and any decoding order policy.*

We proceed with a few clarifying examples.

Example 15 (Corresponding to Theorem 15)**.** *Figure 5.7 shows the complexity bounds from Theorem 15, describing the derived sufficient complexity resources for achieving the optimal DMD $d^*(r/L)$ for the cases of $6 \times n_{\mathrm{R}}$, $L = 2, 3, 6$.*

Example 16 (Corresponding to Corollary 15a)**.** *Figure 5.8 plots the complexity bounds from Corollary 15a describing the derived sufficient complexity resources for achieving the optimal DMD $d^*(r/L)$ for the cases of $2 \times n_{\mathrm{R}}$, $L = n_{\mathrm{T}} = 2$ and of $3 \times n_{\mathrm{R}}$, $L = n_{\mathrm{T}} = 3$. These are compared with the corresponding complexity exponents that guarantee $d^*(r)$ in the absence of feedback.*

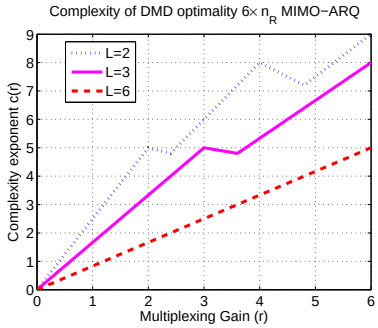
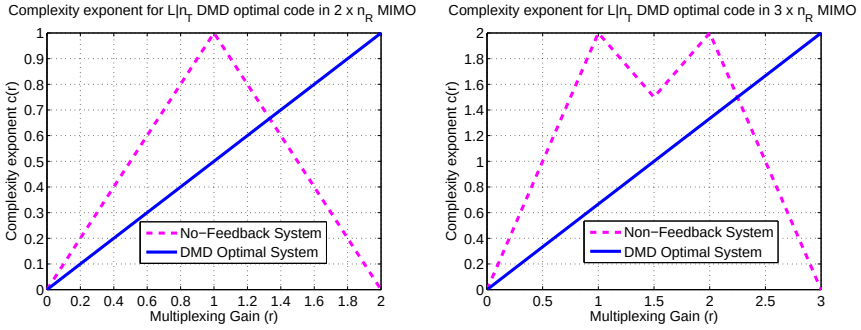


FIGURE 5.7 – Complexity exponent for $6 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 6$) DMD-optimal MIMO system. $L = 2$ (top line), $L = 3$ and $L = 6$ (lowest line).



(a) For $2 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 2$) MIMO (b) For $3 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 3$) MIMO

FIGURE 5.8 – Upper bound on complexity exponent for DMD optimal performance

Example 17 (Comparing Corollary 15a to Theorem 13)**.** *Figure 5.9 plots the complexity bound from Corollary 15a (upper dotted line) describing the derived sufficient complexity resources for achieving the optimal DMD $d^*(r/L)$, and compares this with the (much reduced) complexity bound from Theorem 13 describing the derived sufficient complexity resources for achieving the optimal DMT $d^*(r)$ for the case of $4 \times n_{\mathrm{R}}$, $L = n_{\mathrm{T}} = 4$.*

The question arises as to whether it is better to use the scheme that provides for $d^*(r)$ with reduced complexity (upper line in Fig. 5.9, corresponding to Theorem 13), or the scheme that provides for the much improved $d^*(r/L) = d^*(r/n_{\mathrm{T}})$ but does so with much increased complexity (upper line in Fig. 5.9, corresponding to Corollary 15a). Towards deciding this, one must first decide what is the price of flops, compared to the price of errors. A meaningful joint rate-reliability-complexity measure could for example be chosen to take the form $\Gamma(r) = d(r) - \gamma(r)$ where γ reflects the relative cost of chip size vs reliability; the higher the γ the more importance is placed on computational efficiency rather than reliability. The cases of $\gamma = 1$ and $\gamma = 10$ are

respectively shown in Figures 5.10(a) and 5.10(b). We observe that if the price of flops, compared to the price of errors, justifies a $\gamma > 10$, then the computationally efficient DMT (not DMD) achieving scheme with $L = n_{\mathrm{T}}$ rounds, seems to be a better choice.

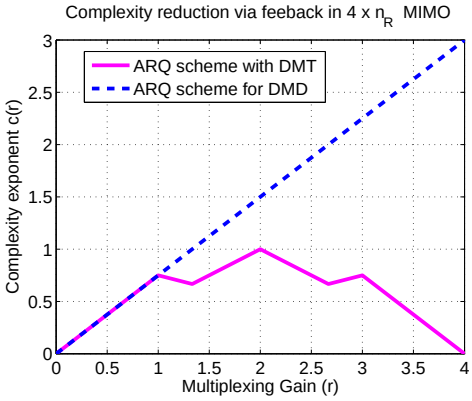
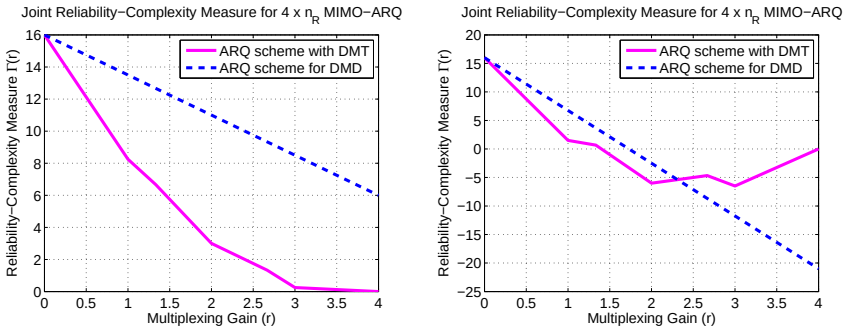


FIGURE 5.9 – Complexity exponent for $4 \times n_{\mathrm{R}}$ MIMO system

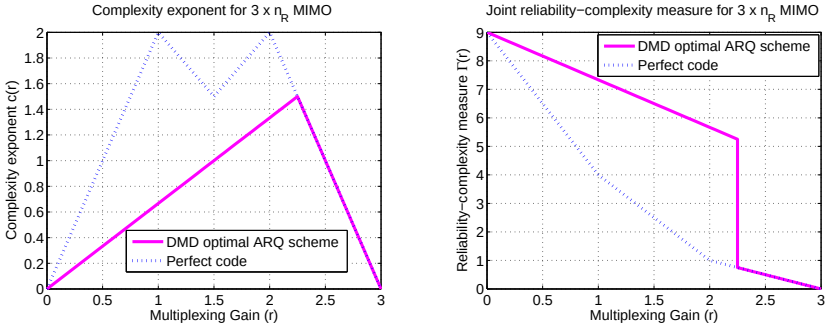


(a) $\gamma = 1$ (b) $\gamma = 10$

FIGURE 5.10 – Joint reliability-complexity measure for DMT and DMD optimal ARQ schemes

Example 18 (Drawing from Corollary 15a)**.** *In an $n_{\mathrm{T}} \times n_{\mathrm{R}}$ MIMO setting ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$), consider a legacy system that achieves $d^*(r)$, that does not have an ARQ mechanism, and which is limited to computational resources that suffice only to achieve this DMT optimality (Theorem 5, Corollary 5a, dotted line in Fig. 5.11(a)). What if, for the same exact system, and the same resources, we are now given the ARQ option with $L = n_{\mathrm{T}}$ bits of feedback. One way to utilize this would be to limit ARQ to the region of $r \leq \frac{n_{\mathrm{T}}^2}{n_{\mathrm{T}}+1}$ (achieving $d^*(r/n_{\mathrm{T}})$) and switch it off otherwise (going down to $d^*(r)$). In terms of implementation, for $r \leq \frac{n_{\mathrm{T}}^2}{n_{\mathrm{T}}+1}$, one could simply decode after the*

first time slot, employing an aggressive intermediate halting policy in the first round, and then wait $n_{\mathrm{T}}-1$ *slots to decode if and only if out of outage. For* $r > \frac{n_{\mathrm{T}}^2}{n_{\mathrm{T}}+1}$*, to switch off ARQ, the only thing that would change would be the intermediate halting policy that would now never decode. Figure 5.11 illustrates the above for the case of* $n_{\mathrm{T}} = 3$*.*



(a) Complexity exponent

(b) Joint reliability-complexity measure ($\gamma = 1$)

FIGURE 5.11 – Achieving partial DMD with reduced computational resources. $3 \times n_{\mathrm{R}}$ MIMO system. The left figure shows the complexity exponent of the proposed scheme, interrupted for high r to not exceed the complexity exponent of the feedback scheme.

5.3.1 Complexity costs for DMD optimality : the case of $L > n_{\mathrm{T}}$

For completeness we present bounds on the minimum required complexity to achieve the DMD optimal $d^*(r/L)$ provided by L bits of feedback. The following holds for the $n_T \times n_R$ ($n_R \geq n_T$) MIMO-ARQ channel with i.i.d. regular fading statistics.

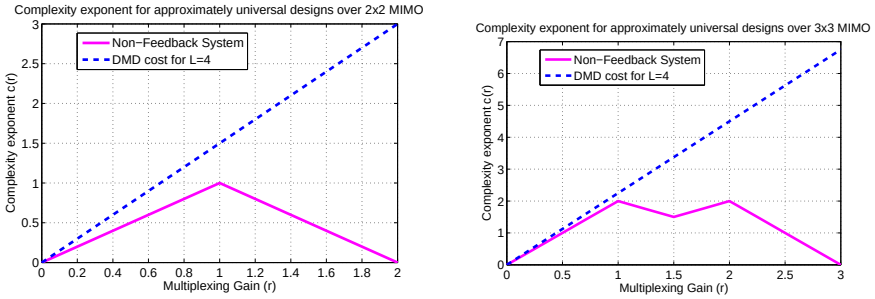
Proposition 7. *The minimum, over all lattice designs and halting and decoding order policies, complexity exponent* $c(r)$ *required to achieve the optimal DMD* $d^*(r/L)$ *of the* $(L > n_{\mathrm{T}})$*-round ARQ, is upper bounded as*

$$c(r) \leq \bar{c}_{DMD}(r) = \left(1 - \frac{1}{L}\right) n_{\mathrm{T}} r. \tag{5.5}$$

The above described bound is sufficient for DMD optimality given any ARQ-compatible, $T = n_{\mathrm{T}}$*, full-rate NVD lattice design, any aggressive intermediate halting policy, any decoding order policy, and given a search radius* $\xi > \sqrt{d^*(\frac{r}{L}) \log \rho}$*.*

We note that the above designs exist for all n_{T} and all $L \geq n_{\mathrm{T}}$ (cf. [36]). We also note that for all known existing lattice design implementations (cf. [36]), the bound is tight for at least one fixed decoding order.

Example 19 (Corresponding to Proposition 7)**.** *We consider the complexity cost of achieving DMD optimality ($d^*(r/L)$) for the example cases of $n_T = 2$ and $n_T = 3$, and for a fixed number $L = 4$ of bits of feedback, and we compare this complexity (upper line in Fig. 5.12(a)-(b) for 2×2 and 3×3 respectively, see Proposition 7) with the complexity cost of achieving the optimal DMT $d^*(r)$ without feedback (lower line in Fig. 5.12(a)-(b) for 2×2 and 3×3 respectively, see Theorem 5, Corollary 5a).*



(a) For 2×2 MIMO (b) For 3×3 MIMO

FIGURE 5.12 – Plot corresponding to Proposition 7

We proceed with a different utilization of feedback, and briefly study the complexity ramifications of antenna selection.

5.4 Complexity reduction using antenna selection

Transmit/receive antenna selection (Fig. 5.13) is a means of utilizing CSIT feedback to reduce the system size while maintaining high DMT performance. Such selection schemes employ $\log_2 \binom{n_T}{l_T}$ bits of CSIT to reduce an $n_T \times n_R$ MIMO system to a smaller and more manageable $l_T \times l_R$ system with generally reduced computational requirements. The extend of system reduction is naturally limited by the rate-reliability requirements. We will here explore the complexity ramifications of antenna selection, and focus on the case where the performance, after antenna selection, remains DMT optimal ($d(r) = d^*_{n_T \times n_R}(r)$). Our work here is preliminary, and it builds on the greedy selection algorithms in [11]. This scheme, in the presence of a requirement for achieving $d^*_{n_T \times n_R}(r)$, places a reduction constraint on $\min\{l_T, l_R\} =: N_r$ where this constraint takes the form

$$N_r = \arg \min_{N' \in \{1,\cdots,n_T\}} \left[\left(\arg \min_{p \in \{0,\cdots,N'-1\}} \frac{(n_{\mathrm{T}} - p)(n_{\mathrm{R}} - p)}{N' - p} \right) = \lceil r \rceil \right]. \quad (5.6)$$

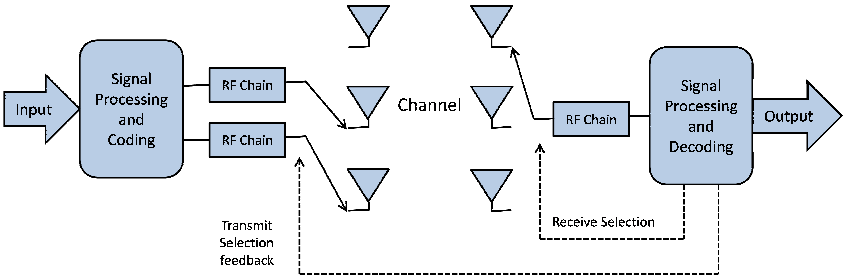


FIGURE 5.13 – Antenna selection

Naturally N_r depends on r; it increases with r up to a certain point after which there is no antenna pruning ($N_r = n_{\mathrm{T}}$).

The following holds for the i.i.d. Rayleigh fading $n_T \times n_R$ ($n_R \geq n_T$) MIMO setting. As before we consider ML based decoding as well as MMSE-preprocessed lattice decoding. Furthermore the results also describe the complexity that guarantees, given any specific antenna selection policy, a vanishing gap to the optimal ML performance, and a vanishing gap to the exact implementation of regularized lattice decoding.

Proposition 8. *The minimum, over all antenna selection algorithms, all lattice designs and all halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal DMT $d^*_{n_T \times n_R}(r)$, is upper bounded as*

$$c(r) \leq \overline{c}_{as}(r) = \left(r(N_r - \lfloor r \rfloor - 1) + (N_r \lfloor r \rfloor - r(N_r - 1))^+\right),$$

which, for the N_r in (5.6), is a piece-wise linear function that, for integer values of multiplexing gain r, takes the form

$$\overline{c}_{as}(r) = r(N_r - r), \text{ for } r = 0, 1, \cdots, n_{\mathrm{T}}. \tag{5.7}$$

The above described bound $\overline{c}_{as}(r)$ is sufficient to achieve the optimal DMT $d^*_{n_T \times n_R}(r)$, given the square ($l_T = l_R = N_r$) greedy antenna selection algorithm[6] in [11], given any $N_r \times N_r$ full-rate NVD (CDA) lattice design, given a sphere decoder with a search radius $\xi > \sqrt{d^*_{n_T \times n_R}(r) \log \rho}$, a decoding halting policy that decodes iff out of outage of the original (unpruned system), and any decoding order policy. Furthermore, given the above implementation and natural decoding ordering, the above described bound is also necessary.

Example 20. *For the 4×4 MIMO channel, and given a target of $d^*_{4\times 4}(r)$, we compare the previously derived ARQ-aided complexity with the current*

6. It is easy to show that the complexity of implementing the greedy algorithm does not affect the complexity exponent.

complexity described above. We first note that no meaningful antenna selection can take place with less than three bits of feedback. For fairness we also consider an $L = 4$-round ARQ system which also employs $L - 1 = 3$ bits of feedback[7]*. The above result reveals (see Fig. 5.14) that the ARQ-aided complexity is uniformly less than that of the specific implementation of antenna selection, which in fact yields no complexity reduction for any $r > 2$ ($N_r = 3$ for $r \leq 2$ and $N_r = 4$ for $r > 2$) with respect to the complexity exponent that would have been needed (perfect code implementation) to achieve the optimal DMT $d^*_{4\times 4}(r)$ without feedback (dotted blue line, Theorem 5, Corollary 5a). The lowest line corresponds to ARQ (Theorem 13).*

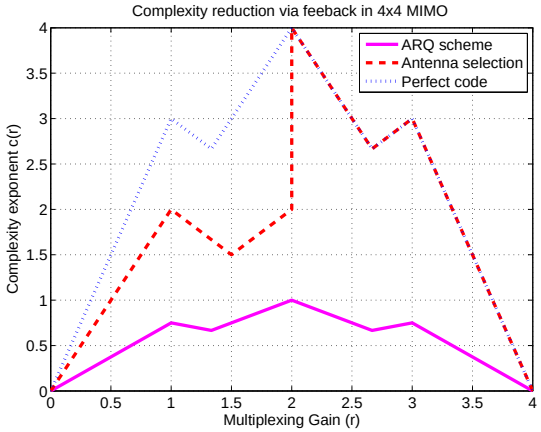


FIGURE 5.14 – Complexity savings : antenna selection vs ARQ

Remark 2. *It is interesting to observe that the expression in* (5.7) *surprisingly corresponds to the complexity of achieving the DMT optimal performance $d^*_{l_T\times l_R}(r)$* of the $l_T \times l_R$ pruned system*, rather than the actual achieved (target) DMT performance $d^*_{n_T\times n_R}(r) > d^*_{l_T\times l_R}(r)$. This is surprising because, whereas complexity is generally increasing with $d(r)$ (for a fixed system dimensionality), in our case there seems to be no complexity penalty on this DMT increase from $d^*_{l_T\times l_R}(r)$ to $d^*_{n_T\times n_R}(r)$. The explanation for this lies in the fact that, despite having less freedom to halt (as compared to the case of a target $d^*_{l_T\times l_R}(r)$), the decoder operates over improved channels (complements of the antenna selection algorithm), which are generally associated with reduced lattice search complexity.*

We proceed with the proofs for the whole chapter.

7. The rationale here is that the case of $L = 1$ entails no feedback.

5.5 Proof of Theorem 13

Now we proceed to prove Theorem 13. We first establish necessary conditions for ARQ signaling to achieve optimal DMT of non-feedback system. For an $n_T \times n_R$ ($n_R \geq n_T$) MIMO system any DMT optimal signaling must satisfy two conditions

Condition 1 : To achieve maximum diversity gain of $n_T.n_R$, the total number of channel uses need to be greater than or equal to n_T, i.e., $LT \geq n_T$, where T is the number of channel uses in each ARQ round. For minimum delay ($T = 1$) L-round ARQ schemes with $L \leq n_{\mathrm{T}}$, it then follows that $L = n_T$.

Condition 2 : To achieve maximum multiplexing gain $r_{max} = n_T$ (as $n_R \geq n_T$), the total number of integers transmitted needs to be greater than or equal to $2n_T T$, which means that $\kappa \geq 2n_T$.

It can be seen from (3.9) that the complexity of sphere decoder

$$N_{SD} = \sum_{k=1}^{\kappa} N_k$$

increases with κ. Thus, for minimum delay L-round ARQ schemes with $L \leq n_{\mathrm{T}}$ tightest upper bound on the complexity exponent is established by considering $T = 1$, $L = n_T$ and $\kappa = 2n_T$. Setting $T = 1$ and $L = n_T$ implies use of at least rate-1 lattice designs for L-round ARQ scheme for achieving DMT performance of $d^*(r)$ for $0 \leq r \leq n_T$. We will later show that encoding-decoding policy described in Proposition 3 and Proposition 4 achieve DMT performance of $d^*(r)$ for $T = 1$, $L = n_T$ and $\kappa = 2n_T$, which in turn implies that $\lim_{\rho\to\infty} r = r_1$, where r_1 is the multiplexing gain for the first round of ARQ.

Having established the necessary parameters κ and L we proceed to prove the upper bound in the Theorem. Following the footsteps of the proof of Theorem 1 we can show that for decoding full-rate code at the end of first round the total number of visited nodes is given by

$$N_{SD}(\boldsymbol{\mu}) \dot{\leq} \rho^{\sum_{j=1}^{n_T} \min\left(\frac{r}{n_T}-(1-\mu_j), \frac{r}{n_T}\right)^+},$$

where $\mu_j = -\frac{\log \sigma_j(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho}$, $j = 1, \cdots, n_T$ with $\mu_1 \geq \cdots \geq \mu_{n_T}$ and where we have made use of the fact that $r = r_1$. Recalling that first round decoding is performed iff $\sigma_1(\mathbf{H}_C^H \mathbf{H}_C) > \rho^{-\epsilon}$ for some $\epsilon > 0$, the upper bound on the complexity exponent for first round decoder can be obtained as the solution to a constrained maximization problem according to

$$\overline{c}_1(r) \triangleq \max_{\{\mu_1 < \epsilon,\ \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0\}} \sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^+,$$

which simplifies (in the limit $\epsilon \to 0$) to take the form

$$\bar{c}_1(r) = 0.$$

To establish the L-th round complexity cost, we proceed with the system model given by

$$\mathbf{Y}_C^L = \theta \mathbf{H}_C \mathbf{X}_C^{ARQ,L} + \mathbf{W}_C^L,$$

where for rate-1 lattice designs we have $\theta^2 = \rho^{1-r_L}$ where $r_L = \frac{r}{L}$ denotes multiplexing gain for L-th round of ARQ. The vectorized real valued representation of L-th round system model takes the form

$$\mathbf{y}^L = \theta \mathbf{H}^L \mathbf{x}^L + \mathbf{w}^L, \tag{5.8}$$

where

$$\mathbf{H}^L = \mathbf{I}_L \otimes \mathbf{H}_R \in \mathbb{R}^{2n_R L \times 2n_T L}, \mathbf{H}_R = \begin{bmatrix} Re\{\mathbf{H}_C\} & -Im\{\mathbf{H}_C\} \\ Im\mathbf{H}_C\} & Re\{\mathbf{H}_C\} \end{bmatrix},$$

$$\mathbf{x}^L = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_T L} \text{ with } \mathbf{x}_t = [Re\{\mathbf{X}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{X}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$
$$\mathbf{w}^L = (\mathbf{w}_1^{\mathrm{T}}, \cdots, \mathbf{w}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R L} \text{ with } \mathbf{w}_t = [Re\{\mathbf{W}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{W}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$
$$\mathbf{y}^L = (\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R L} \text{ with } \mathbf{y}_t = [Re\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

for $t = 1, \cdots, L$, where $\mathbf{X}_{t,C}$, $\mathbf{Y}_{t,C}$ and $\mathbf{W}_{t,C}$ are t-th column of $\mathbf{X}_C^{ARQ,L}$, $\mathbf{W}_C^L$ and $\mathbf{Y}_C^L$ respectively. The vectorized codeword $\mathbf{x}_L$ takes the form (cf. (3.2))

$$\mathbf{x}^L = \mathbf{G}\mathbf{s}, \quad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{r_L}{2}} \mathcal{R}, \tag{5.9}$$

where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the shaping region $\mathcal{R}'$ that preserves the code, and contains the all zero vector $\mathbf{0}$. For simplicity we consider $\mathcal{R} \triangleq [-1,\ 1]^\kappa$ to be a hypercube in $\mathbb{R}^\kappa$, although this could be relaxed. Combining (5.8) and (5.9) yields the equivalent system model

$$\mathbf{y}^L = \mathbf{M}^L \mathbf{s} + \mathbf{w}^L, \tag{5.10a}$$
$$\text{where} \quad \mathbf{M}^L \triangleq \rho^{\frac{1}{2}-\frac{r_L}{2}} \mathbf{H}^L \mathbf{G} \in \mathbb{R}^{2n_R L \times \kappa}. \tag{5.10b}$$

Let $\mathbf{G} = [\mathbf{\Gamma}_1^T \ \ \mathbf{\Gamma}_2^T \cdots \mathbf{\Gamma}_L^T]^T$, where $\mathbf{\Gamma}_i \in \mathbb{C}^{2n_T \times 2n_T}$, for $i = 1, \cdots, L$.
Then the equivalent code-channel matrix ($\boldsymbol{M}^L$) takes the form

$$\mathbf{M}^L = \rho^{\frac{1}{2}-\frac{r_L}{2}} \begin{bmatrix} \mathbf{H}_R & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_R & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_R \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \\ \vdots \\ \mathbf{\Gamma}_L \end{bmatrix}, \tag{5.11}$$
$$= \rho^{\frac{1}{2}-\frac{r_L}{2}} [\mathbf{\Gamma}_1^T \mathbf{H}_R^T \ \ \mathbf{\Gamma}_2^T \mathbf{H}_R^T \cdots \mathbf{\Gamma}_{L1}^T \mathbf{H}_R^T]^T. \tag{5.12}$$

In order to compute the singular values of $\boldsymbol{M}^L$ we note that

$$\begin{aligned}(\mathbf{M}^L)^H(\mathbf{M}^L) =&\rho^{1-r_L}(\boldsymbol{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2^H\mathbf{H}_R^H\mathbf{H}_R\boldsymbol{\Gamma}_2+\cdots+\boldsymbol{\Gamma}_L^H\mathbf{H}_C^H\mathbf{H}_C\boldsymbol{\Gamma}_L),\\ \succeq&\rho^{1-r_L}\boldsymbol{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\boldsymbol{\Gamma}_1,\end{aligned} \tag{5.13}$$

where $\mathbf{A}\succeq\mathbf{B}$ denotes that $\mathbf{A}-\mathbf{B}$ is positive-semidefinite. Without loss of generality we can assume that $\boldsymbol{\Gamma}_1$ is full-rank. It then follows that the singular values of $\mathbf{M}^L$ can be lower bounded as

$$\begin{aligned}\sigma_i((\mathbf{M}^L)^H(\mathbf{M}^L)) \geq&\rho^{1-r_L}\sigma_i(\boldsymbol{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\boldsymbol{\Gamma}_1),\\ \dot{\geq}&\rho^{1-r_L-\mu_{\iota_2(i)}}.\end{aligned} \tag{5.14}$$

where we have made use of the fact that $\sigma_{min}(\boldsymbol{\Gamma}_1)\doteq\rho^0$ and where $\iota_2(i)\triangleq\lceil\frac{i}{2}\rceil$.

Let $\mathbf{QR}=\mathbf{M}^L$ be the thin QR factorization of the code-channel matrix $\mathbf{M}^L$ and $\mathbf{r}\triangleq\mathbf{Q}^H\mathbf{y}^L$, then the equivalent system model in (5.10a) is represented by

$$\mathbf{r}=\mathbf{R}\mathbf{s}+\mathbf{Q}^H\mathbf{w}^L,$$

and the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML}=\arg\min_{\hat{\mathbf{s}}\in\mathbb{S}_r^\kappa}\|\mathbf{r}-\mathbf{R}\hat{\mathbf{s}}\|^2, \tag{5.15}$$

which is then solved by the sphere decoder which recursively enumerates all candidate vectors $\hat{\mathbf{s}}\in\mathbb{S}_r^\kappa$ within a given sphere of search radius $\xi=\sqrt{z\log\rho}$, for some $z>d^*(\frac{r}{L})$.

To establish an upper bound on the complexity exponent for L-th round decoding, we note that $\sigma_i(\mathbf{R})=\sigma_i(\mathbf{M}^L)$ and recall from [7, Lemma 1] that the number of nodes visited at layer-k of the SD is upper bounded as

$$\begin{aligned}N_k(\boldsymbol{\mu})\leq&\prod_{i=1}^{k}\left[\sqrt{k}+\min\{\frac{2\xi}{\sigma_i(\mathbf{R})},2\sqrt{k}\rho^{\frac{r_L}{2}}\}\right],\\ =&\prod_{i=1}^{k}\left[\sqrt{k}+\min\{\frac{2\xi}{\sigma_i(\mathbf{M})},2\sqrt{k}\rho^{\frac{r_L}{2}}\}\right],\\ \dot{\leq}&\rho^{\sum_{i=1}^{k}\min\left(\frac{r_L}{2}-\frac{1}{2}(1-\mu_{\iota_2(i)}),\frac{r_L}{2}\right)^+}.\end{aligned}$$

Consequently, the total number of nodes visited by SD is upper bounded as

$$\begin{aligned}N_{SD}(\boldsymbol{\mu})=&\sum_{k=1}^{\kappa}N_k(\boldsymbol{\mu}),\\ \dot{\leq}&\rho^{\sum_{i=1}^{\kappa}\left(\frac{r_L}{2}-\frac{1}{2}(1-\mu_{\iota_2(i)}),\frac{r_L}{2}\right)^+}.\end{aligned} \tag{5.16}$$

Following the footsteps of the proof of Theorem 1 in Section 3.2.2, the upper bound on the complexity exponent for the L-th round decoding of minimum

delay L-round ARQ schemes achieving $d^*(r)$ can be obtained as the solution to a constrained maximization problem according to

$$\overline{c}_{red}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_i), \frac{r}{n_T}\right)^+ \tag{5.17a}$$

$$\text{s.t. } I(\boldsymbol{\mu}) \leq d^*(r), \tag{5.17b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \tag{5.17c}$$

where we have made use of the fact that $L = n_T$ and $r_L = \frac{r}{L}$. The solution to this optimization problem takes the form

$$\overline{c}_{red}(r) = \frac{1}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right],$$

which for integer multiplexing gain values simplifies to

$$\overline{c}_{red}(r) = \frac{1}{n_T} r(n_T - r), \text{ for } r = 0, 1, \cdots, n_T.$$

For the proof to be complete we must now prove that the family of ARQ schemes, halting policies and lattice designs can indeed achieve the desired DMT $d^*(r)$. To achieve this we make use of following lemma

Lemma 4. *For an i.i.d. regular fading channel, a minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds achieves $d^*(r)$ for all ARQ-compatible, minimum delay, NVD, rate-1 lattice designs, all intermediate aggressive halting policies and a sphere decoder with search radius $\xi > \sqrt{d^*(\frac{r}{L})\log\rho}$.*

The proof for lemma is given in Appendix 5A. In the presence of Lemma 4 and upper bound $\overline{c}_{red}(r)$, it is direct to see that a minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$, irrespective of the ARQ-compatible, minimum delay, NVD, rate-1 lattice design, for any aggressive intermediate halting policy, and any decoding order policy. This proves Theorem 13, Proposition 3 and Proposition 4. □

5.6 Proof of Theorem 14

To prove this theorem we first prove Proposition 5. We proceed to consider minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds, implemented with any aggressive intermediate halting policy, any decoding order policy, and a rate-$\frac{n_{\mathrm{R}}}{n_{\mathrm{T}}}$ block-repetition lattice design $\mathbf{X}_C^{ARQ,L}$ where the (rate-1) block component code is drawn from the center of $n_{\mathrm{R}} \times n_{\mathrm{R}}$ perfect codes. Let lattice design $\theta\mathbf{X}_C^{ARQ,L}$ is given by

$$\mathbf{X}_C^{ARQ,L} = \begin{bmatrix} \mathbf{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X} \end{bmatrix} \in \mathbb{C}^{n_T \times n_T}, \tag{5.18}$$

where $\theta^2 = \rho^{1-\frac{r_1}{n_R}}$, and where block component code $\mathbf{X} \in \mathbb{C}^{n_R \times n_R}$. The L-th round received signal is given by

$$\mathbf{Y}_C^L = \theta \mathbf{H}_C \mathbf{X}_C^{ARQ,L} + \mathbf{W}_C^L,$$

where $\theta^2 = \rho^{1-\frac{n_T r_L}{n_R}}$, where we have made use of the fact that $r_L = \frac{r_1}{n_T}$. Let

$$\mathbf{H}_C = \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_{\frac{n_T}{n_R}} \end{bmatrix} \in \mathbb{C}^{n_R \times n_T},$$

where $\mathbf{H}_i \in \mathbb{C}^{n_R \times n_R}$, for $i = 1, \cdots, \frac{n_T}{n_R}$. After substituting for $\mathbf{H}_C$ and $\mathbf{X}_C^{ARQ,L}$ we get that

$$\begin{aligned} \mathbf{Y}_L =& \theta \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_{\frac{n_T}{n_R}} \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X} \end{bmatrix} + \mathbf{W}_C^L, \\ =& \theta \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_{\frac{n_T}{n_R}} \end{bmatrix} \mathbf{X} + \mathbf{W}_C^L, \\ =& \theta \tilde{\mathbf{H}}_C \mathbf{X} + \mathbf{W}_C^L, \end{aligned}$$

where $\tilde{\mathbf{H}}_C = [\mathbf{H}_1^T \ \cdots \ \mathbf{H}_{\frac{n_T}{n_R}}^T]^T \in \mathbb{C}^{n_T \times n_R}$.

We observe that the lattice design in (5.18) converts the system into an equivalent channel $\tilde{\mathbf{H}}_C$ with inverted channel dimensions. It follows that for this equivalent channel $\tilde{\mathbf{H}}_C$ the system parameters are given by $n_T' = n_R$, $n_R' = n_T$, $L' = n_T'$, $T' = 1$, $L' = n_T'$ and $r_L' = \frac{n_T}{n_R} r_L$. Since for the new system we have $n_R' \geq n_T'$, the upper bound and its achievability can be established by following the footsteps of the proof of Theorem 13. After making proper substitutions in (5.17) the upper bound on complexity exponent can be obtained as the solution to a constrained maximization problem according to

$$\begin{aligned} \bar{c}_{red}(r) = \max_{\boldsymbol{\mu}} & \sum_{j=1}^{n_R} \left(\frac{n_T}{n_R} r_L - (1 - \mu_j) \right)^+ \\ \text{s.t.} \ & I(\boldsymbol{\mu}) \leq d^*(r), \\ & 1 \geq \mu_1 \geq \cdots \geq \mu_{n_R} \geq 0, \end{aligned}$$

where we have made use of the fact that $d^*(r_L'.L') = d^*(r_L.L) = d^*(r)$. The solution to this optimization problem takes the form

$$\bar{c}_{red}(r) = \frac{1}{n_R} \left[r(n_R - \lfloor r \rfloor - 1) + (n_R \lfloor r \rfloor - r(n_R - 1))^+ \right],$$

which is a piecewise linear function that, for integer multiplexing gain values, takes the form

$$\bar{c}_{red}(r) = \frac{1}{n_R} r(n_R - r), \text{ for } r = 0, 1, \cdots, n_R.$$

This proves Proposition 5 and Theorem 14. □

5.7 Proof of Theorem 15 and Corollary 15a

We first establish necessary conditions for ARQ signaling to achieve DMD optimality over $n_T \times n_R$ ($n_R \geq n_T$) MIMO-ARQ channel. Any L round ARQ signaling scheme that achieves optimal DMD performance must satisfy two conditions

Condition 1 : To achieve maximum diversity gain of $n_T.n_R$, the total number of channel uses needs to be greater than or equal to n_T, i.e., $LT \geq n_T$, where T is the number of channel uses in one ARQ round.

Condition 2 : For a full-rate lattice design with $n_R \geq n_T$, we have that $\kappa \geq 2n_T LT \geq 2n_T^2$.

It can be seen from (3.9) that the complexity of sphere decoder

$$N_{SD} = \sum_{k=1}^{\kappa} N_k$$

increases with κ. Also, for L-th round decoder which operates over a full-rate DMT optimal code $\mathbf{X}_C^{ARQ,L}$ the complexity exponent in (3.52) indicates that complexity increases with the number of channel uses LT. Thus, tightest upper bound on the complexity exponent is established by considering $LT = n_T$ and $\kappa = 2n_T^2$.

Having established the necessary parameters κ and L we proceed to prove the upper bound in the Theorem. In order to establish decoding complexity for the first round decoding we proceed with the vectorized real valued representation of ℓ-th round signal model in (5.1), which takes the form

$$\mathbf{y}^\ell = \sqrt{\rho}\mathbf{H}^\ell \mathbf{x}^\ell + \mathbf{w}^\ell, \tag{5.19}$$

where

$$\mathbf{H}^\ell = \mathbf{I}_{\ell T} \otimes \mathbf{H}_R \in \mathbb{R}^{2n_R \ell T \times 2n_T \ell T}, \mathbf{H}_R = \begin{bmatrix} Re\{\mathbf{H}_C\} & -Im\{\mathbf{H}_C\} \\ Im\mathbf{H}_C\} & Re\{\mathbf{H}_C\} \end{bmatrix},$$

$$\mathbf{x}^\ell = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_T \ell T} \text{ with } \mathbf{x}_t = [Re\{\mathbf{X}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{X}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{w}^\ell = (\mathbf{w}_1^{\mathrm{T}}, \cdots, \mathbf{w}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R \ell T} \text{ with } \mathbf{w}_t = [Re\{\mathbf{W}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{W}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\mathbf{y}^\ell = (\mathbf{y}_1^{\mathrm{T}}, \cdots, \mathbf{y}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_R \ell T} \text{ with } \mathbf{y}_t = [Re\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}, Im\{\mathbf{Y}_{t,C}\}^{\mathrm{T}}]^{\mathrm{T}},$$

for $t = 1, \cdots, \ell T$, where $\mathbf{X}_{t,C}$, $\mathbf{Y}_{t,C}$ and $\mathbf{W}_{t,C}$ are t-th column of $\mathbf{X}_C^{ARQ,\ell}$, $\mathbf{W}_C^{\ell}$ and $\mathbf{Y}_C^{\ell}$ respectively.

For the lattice codes described in Chapter 2 and ℓ-th round multiplexing gain r_ℓ, the vectorized codewords take the form

$$\mathbf{x}^{\ell} = \rho^{\frac{-r_\ell T_\ell}{\kappa}} \mathbf{G}\mathbf{s}, \quad \mathbf{s} \in \mathbb{S}_r^{\kappa} \triangleq \mathbb{Z}^{\kappa} \cap \rho^{\frac{r_\ell T_\ell}{\kappa}} \mathcal{R}, \tag{5.20}$$

where $T_\ell = \ell T$, and where $\mathcal{R} \subset \mathbb{R}^{\kappa}$ is a natural bijection of the shaping region $\mathcal{R}'$ that preserves the code, and contains the all zero vector $\mathbf{0}$. For simplicity we consider $\mathcal{R} \triangleq [-1,\ 1]^{\kappa}$ to be a hypercube in $\mathbb{R}^{\kappa}$, although this could be relaxed. We note that for ℓ-th ARQ round $r_\ell T_\ell = rT$ for $\ell = 1, \cdots, L$, where r is the effective multiplexing gain of ARQ signaling and we will later argue that for the DMD optimal scheme $\lim_{\rho\to\infty} r = r_1$. Combining (5.19) and (5.20) yields the equivalent system model

$$\mathbf{y}^{\ell} = \mathbf{M}^{\ell}\mathbf{s} + \mathbf{w}^{\ell}, \tag{5.21a}$$

$$\text{where} \quad \mathbf{M}^{\ell} \triangleq \rho^{\frac{1}{2}-\frac{rT}{\kappa}} \mathbf{H}^{\ell}\mathbf{G} \in \mathbb{R}^{2n_R\ell T \times \kappa}. \tag{5.21b}$$

For the complexity analysis we need a square and full-rank generator matrix (see Section 3.2.2), we proceed to consider a new generator matrix $\mathbf{G'}$ obtained by adding $\kappa - 2n_T\ell T$ linearly independent rows to $\mathbf{G}$, such that $\mathbf{G'} \in \mathbb{R}^{\kappa\times\kappa}$ is full-rank and square matrix. This substitution calls for a new channel matrix $\mathbf{H'} \in \mathbb{R}^{n\times\kappa}$ obtained by appending $\kappa - 2n_T\ell T$ columns containing zeros to $\mathbf{H}^{\ell}$, such that the code-channel matrix $\mathbf{M}$ remains unaltered. The combined code-channel matrix then takes the form

$$\mathbf{M}^{\ell} = \rho^{\frac{1}{2}-\frac{rT}{\kappa}} \mathbf{H'}\mathbf{G'}. \tag{5.22}$$

As discussed before, to implement ML-based SD for first $L-1$ rounds decoder needs to perform MMSE-preprocessing. Let $\mathbf{Q}^{\ell}\mathbf{R}^{\ell} = \tilde{\mathbf{M}}^{\ell}$ be the thin QR factorization of the MMSE-preprocessed code-channel matrix $\tilde{\mathbf{M}}^{\ell} \triangleq \begin{bmatrix} \mathbf{M}^{\ell} \\ \alpha\mathbf{I} \end{bmatrix}$, where $\alpha = \rho^{\frac{-rT}{\kappa}}$, then (5.21a) yields

$$\mathbf{r}^{\ell} = \mathbf{R}^{\ell}\mathbf{s} + \tilde{\mathbf{w}}^{\ell},$$

where $\mathbf{r}^{\ell} = (\mathbf{R}^{\ell})^{-H}(\mathbf{M}^{\ell})^{H}\mathbf{y}^{\ell}$, where $(\mathbf{R}^{\ell})^{H}(\mathbf{R}^{\ell}) = (\mathbf{M}^{\ell})^{H}(\mathbf{M}^{\ell}) + \alpha^2\mathbf{I}$, and where the equivalent noise term

$$\tilde{\mathbf{w}}^{\ell} = -\alpha^2(\mathbf{R}^{\ell})^{-H}\mathbf{s} + (\mathbf{R}^{\ell})^{-H}(\mathbf{M}^{\ell})^{H}\mathbf{w}^{\ell}.$$

Consequently, the ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg\min_{\hat{\mathbf{s}}\in\mathbb{S}_r^{\kappa}} \left\| \mathbf{r}^{\ell} - \mathbf{R}^{\ell}\hat{\mathbf{s}} \right\|^2, \tag{5.23}$$

which is then solved by the sphere decoder which recursively enumerates all candidate vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^\kappa$ within a given sphere of radius $\xi = \sqrt{z \log \rho}$, for some $z > d^*(\frac{r}{L})$.

Following the similar approach as in Section 3.2.2, for $\mu_i \triangleq -\frac{\log \sigma_i(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho}$, for $i = 1, \cdots, n_T$, it follows that

$$\sigma_i(\mathbf{R}^\ell) = \sigma_i(\tilde{\mathbf{M}}^\ell) = \sqrt{\alpha^2 + \sigma_i((\mathbf{M}^\ell)^H(\mathbf{M}^\ell))},$$

which results in

$$\sigma_i(\mathbf{R}^\ell) \dot{\geq} \begin{cases} \rho^{\frac{-rT}{\kappa}} & \text{for } i = 1, \cdots, \kappa - 2n_T\ell T, \\ \rho^{\frac{-rT}{\kappa} + \frac{1}{2}(1-\mu_{\iota_{2\ell T}(i-(\kappa-2n_T\ell T))})^+} & \text{for } i = \kappa - 2n_T\ell T + 1, \cdots, \kappa. \end{cases} \tag{5.24}$$

where we have made use of the $2\ell T$ multiplicity of the singular values of $\mathbf{H}^\ell$ and the fact that $\sigma_{\min}(\mathbf{G}') \doteq \rho^0$ and where $\iota_{2\ell T}(i) \triangleq \lceil \frac{i}{2\ell T} \rceil$. We recall from (2.14) that the total number of the visited nodes for any given channel realization $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_m)$ takes the form

$$N_{SD}(\boldsymbol{\mu}) = \sum_{k=1}^{\kappa} N_k(\boldsymbol{\mu}),$$

where from [7, Lemma 1]

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + \min\{ \frac{2\xi}{\sigma_i(\mathbf{R}_k^\ell)}, 2\sqrt{k}\rho^{\frac{rT}{\kappa}} \} \right].$$

From the interlacing property of singular values of sub-matrices [53] we have that $\sigma_i(\mathbf{R}_k^\ell) \geq \sigma_i(\mathbf{R}^\ell)$. It follows that

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + \min\left\{ \frac{2\xi}{\sigma_i(\mathbf{R}^\ell)}, 2\sqrt{k}\rho^{\frac{rT}{\kappa}} \right\} \right].$$

Consequently, we have that

$$\begin{aligned} N_{SD}(\boldsymbol{\mu}) \leq & \sum_{k=1}^{\kappa} \left\{ \prod_{i=1}^{k} \left[\sqrt{k} + \min\left\{ \frac{2\xi}{\sigma_i(\mathbf{R}^\ell)}, 2\sqrt{k}\rho^{\frac{rT}{\kappa}} \right\} \right] \right\}, \\ \dot{\leq} & \rho^{(1-\frac{2n_T\ell T}{\kappa})rT + \sum_{i=1}^{2n_T\ell T} \min\left(\frac{rT}{\kappa} - \frac{1}{2}(1-\mu_{\iota_{2\ell T}(i)})^+, \frac{rT}{\kappa} \right)^+}, \\ = & \rho^{(1-\frac{2n_T\ell T}{\kappa})rT + \ell T \sum_{j=1}^{n_T} \left(\frac{2rT}{\kappa} - (1-\mu_j)^+ \right)^+}, \end{aligned} \tag{5.25}$$

where we have made use of the fact that

$$\min\left(\frac{2rT}{\kappa} - (1-\mu_j)^+, \frac{2rT}{\kappa} \right)^+ = \left(\frac{2rT}{\kappa} - (1-\mu_j)^+ \right)^+.$$

After substituting $\kappa = 2n_T^2$ and $LT = n_T$ in (5.25) we have that

$$N_{SD}(\boldsymbol{\mu}) \doteq \rho^{(L-\ell)\frac{rn_T}{L^2}+\frac{\ell n_T}{L}\sum_{j=1}^{n_T}\left(\frac{r}{Ln_T}-(1-\mu_j)^+\right)^+}.$$

Recalling that first round decoding is performed iff $\sigma_1(\mathbf{H}_C^H\mathbf{H}_C) > \rho^{-\epsilon}$ for some $\epsilon > 0$, for a given L the upper bound on the complexity exponent for first round ($\overline{c}_1(r)$) can be obtained as the solution to a constrained maximization problem according to

$$\overline{c}_1(r) := \max_{\{\mu_1<\epsilon,\ \mu_1\geq\cdots\geq\mu_{n_T}\geq 0\}} (L-1)\frac{rn_T}{L^2} + \frac{n_T}{L}\sum_{j=1}^{n_T}\left(\frac{r}{Ln_T}-(1-\mu_j)^+\right)^+,$$

which simplifies to take the form

$$\overline{c}_1(r) = (L-1)\frac{rn_T}{L^2}.$$

For a given L, the upper bound on the complexity exponent of the L-th round decoder ($\overline{c}_L(r)$) which operates over a full-rate DMT optimal code $\mathbf{X}_C^{ARQ,L}$ follows from (3.52). After making proper substitutions for total number of channel uses LT and effective multiplexing gain $r_L = \frac{r}{L}$, the upper bound takes the form

$$\overline{c}_L(r) := \frac{1}{L}\left[r\left(n_T - \left\lfloor\frac{r}{L}\right\rfloor - 1\right) + \left(Ln_T\left\lfloor\frac{r}{L}\right\rfloor - r(n_T-1)\right)^+\right],$$

which is a piece-wise linear function that, for multiplexing gain values equal to the integer multiples of L, takes the form

$$\overline{c}_L(r) = \frac{rn_T}{L^2}\left(L - \frac{r}{n_T}\right).$$

The upper bound on the overall complexity exponent of ARQ signaling then takes the form

$$\overline{c}_{DMD}(r) = \max\left(\overline{c}_1(r), \overline{c}_L(r)\right).$$

It turns out that for all feasible values of L such that $L|n_T$, $\overline{c}_L(r) \geq \overline{c}_1(r)$. It then follows that

$$\overline{c}_{DMD}(r) = \overline{c}_L(r).$$

For the proof to be complete we must now prove that the family of ARQ schemes, halting policies and lattice designs can indeed achieve the desired DMT $d^*(r/L)$. To achieve this we make use of following lemma

Lemma 5. *For an i.i.d. regular fading channel, an L rounds ARQ scheme over achieves $d^*(r/L)$ for all ARQ-compatible, minimum delay, NVD, full-rate lattice designs, all intermediate aggressive halting policies and a sphere decoder with search radius $\xi > \sqrt{d^*(\frac{r}{L})\log\rho}$.*

The proof for lemma is given in Appendix 5B. In the presence of Lemma 5 and upper bound it is direct to see that an L-round $(L|n_{\mathrm{T}})$ MIMO ARQ scheme achieves $d^*(r/L)$ with $c(r) \leq \overline{c}_{dmd}(r)$, for any ARQ-compatible square $(LT = n_{\mathrm{T}})$ NVD lattice design, for a search radius $\xi > \sqrt{d^*(\frac{r}{n_T})\log\rho}$, for any aggressive intermediate halting policy, and any decoding order policy. This proves Proposition 6 and Theorem 15. □

For $L = n_{\mathrm{T}}$, we have that

$$\overline{c}_1(r) = \overline{c}_L(r) = \left(1 - \frac{1}{n_T}\right) r.$$

It then follows that

$$\overline{c}_{DMD}(r) = \left(1 - \frac{1}{n_T}\right) r.$$

This proves Corollary 15a. □

Note on decoding policy : We here note that the statement of Theorem 15 also holds for a decoding policy where decoding is performed after each ARQ round instead of decoding only after first and last round. For this decoding policy, the decoder of intermediate rounds $\ell = 1, \cdots, L-1$ should apply aggressive halting strategy that allows decoding with ACK iff the minimum singular value of channel $\sigma_1(\mathbf{H}_C^H \mathbf{H}_C) > \rho^{-\epsilon}$ for some $\epsilon > 0$ and NACK otherwise. For L-th round the decoding is performed as before, i.e., whenever channel is not is outage. For intermediate rounds the complexity upper bound follows from (5.25), for a given L this upper bound on the complexity exponent for ℓ-th round can be obtained as the solution to a constrained maximization problem according to

$$\overline{c}_\ell(r) := \max_{\{\mu_1 < \epsilon,\ \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0\}} (L-\ell)\frac{r n_T}{L^2} + \frac{\ell n_T}{L}\sum_{j=1}^{n_T}\left(\frac{r}{L n_T} - (1-\mu_j)^+\right)^+,$$

which simplifies to take the form

$$\overline{c}_\ell(r) = (L-\ell)\frac{r n_T}{L^2}.$$

We can easily see that $\overline{c}_1(r) > \overline{c}_2(r) > \cdots > \overline{c}_{L-1}(r)$. Consequently, the overall complexity exponent of ARQ signaling remains unchanged. The DMD achievability again holds directly from Lemma 5.

5.8 Proof of Proposition 7

The DMD optimality of any $L > n_T$ ARQ-compatible, $T = n_{\mathrm{T}}$, full-rate NVD lattice design, any aggressive intermediate halting policy, any decoding order policy, and given a search radius $\xi > \sqrt{d^*(\frac{r}{L})\log\rho}$ is direct from [36,

Proof of Theorem 9] and Lemma 5. In order to prove the claim of proposition we now look to establish the upper bound on the complexity exponent. The full-rate DMD optimal lattice designs with L ARQ rounds transmit $\kappa = 2Ln_T^2$ integers in each round of $T = n_T$ channel uses.

In order to establish complexity cost of the first round decoding we follow the footsteps of the proof of Theorem 15. Following from (5.25) with $\kappa = 2Ln_T^2$ and $T = n_T$, the first round decoding the total number of visited nodes is given by

$$N_{SD}(\boldsymbol{\mu}) \doteq \rho^{(1-\frac{1}{L})rn_T + n_T \sum_{j=1}^{n_T} \left(\frac{r}{Ln_T} - (1-\mu_j)^+\right)^+}, \tag{5.26}$$

Recalling that first round decoding is performed iff $\sigma_1(\mathbf{H}_C^H \mathbf{H}_C) > \rho^{-\epsilon}$ for some $\epsilon > 0$, for a given L the upper bound on the complexity exponent for first round can be obtained as the solution to a constrained maximization problem according to

$$\overline{c}_1(r) \triangleq \max_{\{\mu_1 < \epsilon,\ \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0\}} (L-1)\frac{rn_T}{L} + n_T \sum_{j=1}^{n_T} \left(\frac{r}{Ln_T} - (1-\mu_j)^+\right)^+,$$

which simplifies to take the form

$$\overline{c}_1(r) = \left(1 - \frac{1}{L}\right) rn_T.$$

For a given L, the upper bound on the complexity exponent of the L-th round decoder which operates over a full-rate DMT optimal code $\mathbf{X}_C^{ARQ,L}$ can be obtained as the solution to a constrained optimization problem

$$\begin{aligned} \overline{c}_L(r) = \max_{\boldsymbol{\mu}}\ & Ln_T \sum_{j=1}^{n_T} \min\left(\frac{r}{Ln_T} - (1-\mu_j), \frac{r}{Ln_T}\right)^+ \\ \text{s.t.}\ & \sum_{j=1}^{n_T} (1-\mu_j)^+ \geq \frac{r}{L}, \\ & \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \end{aligned}$$

The solution to this optimization problem takes the form

$$\overline{c}_L(r) = r\left(n_T - \left\lfloor \frac{r}{L} \right\rfloor - 1\right) + \left(Ln_T \left\lfloor \frac{r}{L} \right\rfloor - r(n_T - 1)\right)^+, \tag{5.27}$$

which is a piece-wise linear function that, for multiplexing gain values equal to the integer multiples of L, takes the form

$$\overline{c}_L(r) = r\left(n_T - \frac{r}{L}\right).$$

For $L \geq n_T$, the complexity exponent $c_L(r)$ simplifies to take the form

$$\overline{c}_L(r) = \left(1 - \frac{1}{n_T}\right) r n_T.$$

The upper bound on overall complexity exponent is given by

$$\overline{c}_{DMD}(r) = \max\left(\overline{c}_1(r), \overline{c}_L(r)\right),$$

which for the case of $L \geq n_T$, takes the form

$$\overline{c}_{DMD}(r) = \left(1 - \frac{1}{L}\right) r n_T.$$

Furthermore, following the footsteps of the proof of Proposition 1 it can be shown that irrespective of ARQ-compatible, $T = n_{\mathrm{T}}$, full-rate NVD lattice design, any aggressive intermediate halting policy, and given a search radius $\xi > \sqrt{d^*(\frac{r}{L})\log\rho}$ there exists one fixed decoding order for which upper bound $\overline{c}_{DMD}(r)$ is met. This proves Proposition 7.□

5.9 Proof of Proposition 8

The antenna selection algorithm in [11] guarantees that N_r singular values of the $N_r \times N_r$ pruned system are exponentially equal to the N_r largest singular values of the full system, i.e.,

$$\tilde{\sigma}_k \doteq \sigma_{k+(n_T - N_r)}, \ k = 1, \cdots, N, \tag{5.28}$$

where σ_k denotes singular values for the full system and $\tilde{\sigma}_k$ denotes singular values for the pruned system. The DMT performance of pruned system matches with the optimal DMT of full system up to multiplexing gain threshold P, where

$$\begin{aligned} P = \arg\min_{p} \ & \frac{(n_R - p)(n_T - p)}{N - p}, \\ \text{subject to} \ \ & 0 \leq P \leq N_r - 1, p \in \mathbb{Z}. \end{aligned}$$

For any given operating multiplexing gain r, the complexity exponent is minimized by selecting the smallest number N_r such that $P = \lceil r \rceil$, i.e.,

$$N_r = \arg \min_{N' \in \{1, \cdots, n_T\}} \left[\left(\arg \min_{p \in \{0, \cdots, N'-1\}} \frac{(n_{\mathrm{T}} - p)(n_{\mathrm{R}} - p)}{N' - p} \right) = \lceil r \rceil \right].$$

It is shown in [11, Theorem 4.1] that $N_r \times N_r$ achieve optimal DMT $d^*(r)$ with lattice designs of [14] with $T = N_r$. In these settings of the complexity

exponent of ML-based SD with fixed decoding order is given by (cf. Theorem 5)

$$\begin{aligned}\overline{c}_{as}(r) = \max_{\boldsymbol{\mu}} N_r \sum_{j=1}^{N_r} \min\left(\frac{r}{N_r} - (1-\mu_{(j+n_T-N_r)}), \frac{r}{N_r}\right)^+ \\ \text{s.t.} \quad \sum_{j=1}^{n_T} \mu_j \leq n_T - r, \\ \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0. \end{aligned} \tag{5.29}$$

Solution to this optimization problem takes the form

$$\overline{c}_{as}(r) = \left(r(N_r - \lfloor r \rfloor - 1) + (N_r \lfloor r \rfloor - r(N_r - 1))^+\right). \tag{5.30}$$

Furthermore, following the footsteps of the proof of Proposition 1 it can be shown that there exists at least one fixed decoding order for which upper bound $\overline{c}_{as}(r)$ is met. Also, from [7, Theorem 5] it follows that ML-based decoding with the natural decoding order meets the upper bound $\overline{c}_{as}(r)$. This proves Proposition 8. □

Appendix 5A : Proof of Lemma 4

To prove DMT optimality for ARQ-compatible, minimum delay, NVD, rate-1 lattice designs with intermediate aggressive halting policies and a sphere decoder with search radius $\xi^2 = z \log \rho$, for some $z > d^*(\frac{r}{L})$, it is sufficient to show

- that with high probability there will be just a single ARQ round, i.e., $\mathrm{P}\left(\overline{\mathcal{A}}_1\right) \doteq \rho^{-\mathcal{T}}$, where $\mathcal{T} > 0 \; \forall \; 0 \leq r \leq n_T$,
- the error probability of the first decoder applied to the task of ML sphere decoding the ST code $\mathbf{X}_C^{ARQ,1}$ is no larger than that incurred by the ML sphere decoder applied to the task of decoding the ST code $\mathbf{X}_C^{ARQ,L}$, i.e., $\mathrm{P}\left(r\right)_{ARQ,1} \dot{\leq} \mathrm{P}\left(r_L\right)_{ARQ,L}$, and
- the full-length ST code $\mathbf{X}_C^{ARQ,L}$ in L-th round achieves diversity gain $d_{ARQ,L}(r_L) = d^*(r_L.L) = d^*(r)$ with $r_L = \frac{r}{L}$.

Towards proving the first condition above, we recall that in the presence of an aggressive intermediate halting policy, the decoder of first round sends NACK iff $\sigma_1(\mathbf{H}_C^H \mathbf{H}_C) \leq \rho^{-\epsilon}$ for some $\epsilon > 0$ and ACK otherwise. Thus, the probability of a NACK being received at the end of the first round is given by

$$\mathrm{P}\left(\overline{\mathcal{A}}_1\right) = \mathrm{P}\left(\sigma_1(\mathbf{H}_C^H \mathbf{H}_C) \leq \rho^{-\epsilon}\right).$$

For i.i.d. regular fading channel $\mathbf{H}_C$, from [70] it follows that

$$\mathrm{P}\left(\sigma_1(\mathbf{H}_C^H\mathbf{H}_C) \leq \rho^{-\epsilon}\right) \dot{\leq} \int_{\mathcal{A}} \rho^{-I(\boldsymbol{\mu})} \doteq \rho^{-I(\boldsymbol{\mu}^*)},$$

where asymptotic equality comes from Varadhan's lemma [71], where [8]

$$I(\boldsymbol{\mu}) = \sum_{j=1}^{n_T}(n_R - n_T + 2j - 1)\mu_i + \frac{n_R n_T t}{2}\mu_{n_T},$$

$$\mathcal{A} = \{\boldsymbol{\mu} \mid \mu_1 \geq \epsilon, \mu_i \geq 0, \text{ for } i = 2, \cdots, n_T\},$$

and where

$$\boldsymbol{\mu}^* = \arg\inf_{\mathcal{A}} I(\boldsymbol{\mu}).$$

It follows that $I(\boldsymbol{\mu}^*) = (n_R - n_T + 1)\epsilon$ and consequently, we have that

$$\mathrm{P}\left(\overline{\mathcal{A}}_1\right) \dot{\leq} \rho^{-(n_R - n_T + 1)\epsilon}.$$

It proves that $\mathrm{P}\left(\overline{\mathcal{A}}_1\right) \dot{\leq} \rho^{-\mathcal{T}}$ with $\mathcal{T} > 0 \; \forall \; 0 \leq r_1 \leq n_T$. It also implies that $\lim_{\rho\to\infty} r = r_1$. It is in the proof of this condition that we make use of the fact that communication takes place over i.i.d. regular fading statistics, rest of the proof holds irrespective of the fading statistics.

For second condition we need to show that $\mathrm{P}\left(r\right)_{ARQ,1} \dot{\leq} \mathrm{P}\left(r_L\right)_{ARQ,L}$. For first decoder with aggressive halting policy it follows that

$$\begin{aligned} d_e^2(\mathbf{H}_C, \Delta\mathbf{X}_C^{ARQ,1}) =& \|\theta\mathbf{H}_C\Delta\mathbf{X}_C^{ARQ,1}\|^2, \\ \geq& \sigma_1(\mathbf{H}_C^H\mathbf{H}_C)\|\theta\Delta\mathbf{X}_C^{ARQ,1}\|^2, \\ \dot{\geq}& \rho^{1-\frac{r}{n_T}-\epsilon}, \end{aligned}$$

where we have made use of the fact that $\|\theta\Delta\mathbf{X}_C^{ARQ,1}\|^2 \dot{\geq} \rho^{1-\frac{r}{n_T}}$ (cf. [36, Proof of Theorem 5]). Note that $\delta := 1 - \frac{r}{n_T} - \epsilon > 0$ for $0 \leq r \leq n_T - \epsilon$. In the limit $\epsilon \to 0$ whenever first decoder sends ACK the minimum euclidean distance $d_e^2(\mathbf{H}_C, \Delta\mathbf{X}_C^{ARQ,1}) > \rho^{\delta}, \; \delta > 0$ for $0 \leq r \leq n_T$. This implies that whenever ACK is sent only one received codeword can be inside the search sphere of radius $\xi = \sqrt{z\log\rho}$. Thus, from the definition of first decoder it is clear that error probability of the first decoder is the probability of the event that

- $\theta\boldsymbol{H}_C\mathbf{X}_C^{ARQ,1}$ is not included in the search sphere of radius ξ centred around $\boldsymbol{Y}_C^1$, or
- $\theta\boldsymbol{H}_C\hat{\mathbf{X}}_C^{ARQ,1}$ is the unique matrix included in the sphere for some erroneous codeword $\theta\hat{\mathbf{X}}_C^{ARQ,1}$.

8. Recall that parameter t was introduced as a parameter that regulates the near zero behavior of the random variable.

Now $\mathrm{P}\left(r\right)_{ARQ,1}$ can be upper-bounded by the probability that $\theta\mathbf{X}_C^{ARQ,1}$ was transmitted and the additive noise was such that $\theta\boldsymbol{H}_C\mathbf{X}_C^{ARQ,1}$ is not in the search sphere of the corresponding received matrix $\boldsymbol{Y}_C^1$. We thus have

$$\mathrm{P}\left(r\right)_{ARQ,1} \leq \mathrm{P}\left(\|\mathbf{W}_C^1\|_F^2 > \xi^2\right).$$

We note that search radius $\xi^2 = z\log\rho$, for some $z > d^*(\frac{r}{L})$, it then follows that

$$\mathrm{P}\left(\|\mathbf{W}_C^1\|_F^2 > \xi^2\right) \doteq \rho^{-d^*(\frac{r}{L})}$$

which implies that probability of excluding the transmitted codeword from the search is exponentially smaller than probability of error for the L-th round, i.e.,

$$\mathrm{P}\left(r\right)_{ARQ,1} \doteq \mathrm{P}\left(r_L\right)_{ARQ,L}.$$

To satisfy third condition we prove that $d_{ARQ}(r_L) \geq d^*(r)$ by proving that $\mathrm{P}\left(r_L\right)_{ARQ,L} \doteq \rho^{-d^*(r_L.L)}$. For this purpose we proceed to consider the pairwise error probability (PEP) of the decoder when the channel is not in outage for rate r_1. Let $\nu_1 \geq \cdots \geq \nu_{n_T}$ be the ordered eigenvalues of $(\Delta\mathbf{X}_C^{ARQ,L})(\Delta\mathbf{X}_C^{ARQ,L})^H$. then using the mismatched eigenvalue bound [14] we have that

$$\begin{aligned}
d_e^2(\mathbf{H}_C, \Delta\mathbf{X}_C^{ARQ,L}) =& \theta^2 \mathrm{Tr}(\mathbf{H}_C \left(\Delta\mathbf{X}_C^{ARQ,L}\right)\left(\Delta\mathbf{X}_C^{ARQ,L}\right)^H \mathbf{H}_C^H), \\
\doteq& \theta^2 \sum_{i=1}^{n_T} \rho^{-\mu_i}\nu_i, \\
\geq& \theta^2 \sum_{i=n_T-J}^{n_T} \rho^{-\mu_i}\nu_i, \quad 1 \leq J < n_T, \\
\overset{(a)}{\doteq}& \theta^2 \left(\prod_{i=n_T-J}^{n_T} \rho^{-\mu_i}\right)^{\frac{1}{J+1}} \left(\prod_{i=n_T-J}^{n_T} \nu_i\right)^{\frac{1}{J+1}}, \\
\overset{(b)}{\doteq}& \theta^2 \left(\prod_{i=n_T-J}^{n_T} \rho^{-\mu_i}\right)^{\frac{1}{J+1}} \left(\frac{1}{\rho^{(n_T-J-1)r_L}}\right)^{\frac{1}{J+1}}, \\
=& \rho^{\frac{J+1-n_T r_L - \sum_{i=n_T-J}^{n_T}\mu_i}{J+1}},
\end{aligned} \tag{5.31}$$

where (a) follows from arithmetic mean-geometric mean inequality and (b) follows from the facts that the product of the eigenvalues is equal to the determinant, the non-vanishing determinant property which implies that $\det\{(\Delta\mathbf{X}_C^{ARQ,L})(\Delta\mathbf{X}_C^{ARQ,L})^H\} \doteq \rho^0$ and the fact that the eigenvalues of a matrix are upper bounded by the trace. We will now show that if the block fading channel $(\mathbf{H}_C)$ is not in outage for rate r_1, that for some J, $1 \leq J < n_T$,

$J+1-n_Tr_L-\sum_{i=n_T-J}^{n_T}\mu_i>0$. If the block fading channel is not in outage for rate $r_1+\epsilon,\quad \epsilon>0$, we must have

$$\begin{aligned}\log\det\{\mathbf{I}_{n_T}+\rho\mathbf{H}_C^H\mathbf{H}_C\}\geq&(Lr_L+\epsilon)\log\rho,\\ \Rightarrow\sum_{i=1}^{n_T}(1-\mu_i)^+\geq&Lr_L+\epsilon.\end{aligned}$$

Let n_T-J be the smallest index i for which $\mu_i<1$, then for $L=n_T$ we have that

$$\begin{aligned}\sum_{i=n_T-J}^{n_T}(1-\mu_i)\geq&n_Tr_L+\epsilon,\\ \Rightarrow J+1-n_Tr_L-\sum_{i=n_T-J}^{n_T}\mu_i\geq&\epsilon.\end{aligned}$$

By taking the limit as $\epsilon\to 0$ we see that the probability of error is negligible in no outage region of $\mathbf{H}_C$. Consequently, from (5.13) probability of error is negligible in no outage region of $\mathbf{M}^L$. Again from (5.13) the outage probability for the channel $\mathbf{M}^L$ can be upper bounded as

$$\begin{aligned}\mathrm{P}\left(\mathcal{O}\right)\dot{\leq}&\mathrm{P}\left(\log\det\{\mathbf{I}_{n_T}+\rho\mathbf{H}_C^H\mathbf{H}_C\}\leq Lr_L\log\rho\right),\\ \doteq&\rho^{-d^*(Lr_L)}.\end{aligned}$$

It follows that $\mathrm{P}\left(r_L\right)_{ARQ,L}\dot{\leq}\rho^{-d^*(r_L.L)}$. Thus the lattice design $\mathbf{X}_C^{ARQ,L}$ achieves diversity gain

$$d_{ARQ,L}(r_L)\geq d^*(r_L.L)=d^*(r).$$

This proves Lemma 4. □

Appendix 5B : Proof of Lemma 5

To prove DMD optimality for ARQ-compatible, minimum delay, NVD, full-rate lattice designs with intermediate aggressive halting policies and a sphere decoder with search radius $\xi^2=z\log\rho$, for some $z>d^*(\frac{r}{L})$, it is sufficient to show

- that with high probability there will be just a single ARQ round, i.e., $\mathrm{P}\left(\overline{\mathcal{A}}_1\right)\doteq\rho^{-\mathcal{T}}$, where $\mathcal{T}>0\ \forall\ 0\leq r\leq n_T$,
- the error probability of the first decoder applied to the task of ML sphere decoding the ST code $\mathbf{X}_C^{ARQ,1}$ is no larger than that incurred by the ML sphere decoder applied to the task of decoding the ST code $\mathbf{X}_C^{ARQ,L}$, i.e., $\mathrm{P}\left(r\right)_{ARQ,1}\dot{\leq}\mathrm{P}\left(r_L\right)_{ARQ,L}$, and

- the full-length ST code $\mathbf{X}_C^{ARQ,L}$ achieves diversity gain $d_{ARQ,L}(r_L) = d^*(\frac{r}{L})$.

Towards proving the first condition above, we recall that in the presence of an aggressive intermediate halting policy, the decoder of first round sends NACK iff $\sigma_1(\mathbf{H}_C^H\mathbf{H}_C) \leq \rho^{-\epsilon}$ for some $\epsilon > 0$ and ACK otherwise. Thus, the probability of a NACK being received at the end of the first round is given by

$$\mathrm{P}\left(\overline{\mathcal{A}}_1\right) = \mathrm{P}\left(\sigma_1(\mathbf{H}_C^H\mathbf{H}_C) \leq \rho^{-\epsilon}\right).$$

For i.i.d. regular fading channel $\mathbf{H}_C$, from [70] it follows that

$$\mathrm{P}\left(\sigma_1(\mathbf{H}_C^H\mathbf{H}_C) \leq \rho^{-\epsilon}\right) \dot{\leq} \int_{\mathcal{A}} \rho^{-I(\boldsymbol{\mu})} \doteq \rho^{-I(\boldsymbol{\mu}^*)},$$

where asymptotic equality comes from Varadhan's lemma [71], where

$$I(\boldsymbol{\mu}) = \sum_{j=1}^{n_T}(n_R - n_T + 2j - 1)\mu_i + \frac{n_R n_T t}{2}\mu_{n_T},$$

$$\mathcal{A} = \{\boldsymbol{\mu} \mid \mu_1 \geq \epsilon, \mu_i \geq 0, \text{ for } i = 2, \cdots, n_T\},$$

and where

$$\boldsymbol{\mu}^* = \arg\inf_{\mathcal{A}} I(\boldsymbol{\mu}).$$

It follows that $I(\boldsymbol{\mu}^*) = (n_R - n_T + 1)\epsilon$ and consequently, we have that

$$\mathrm{P}\left(\overline{\mathcal{A}}_1\right) \dot{\leq} \rho^{-(n_R - n_T + 1)\epsilon}.$$

It proves that $\mathrm{P}\left(\overline{\mathcal{A}}_1\right) \dot{\leq} \rho^{-\mathcal{T}}$ with $\mathcal{T} > 0 \; \forall \; 0 \leq r_1 \leq n_T$. It also implies that $\lim_{\rho\to\infty} r = r_1$. It is in the proof of this condition that we make use of the fact that communication takes place over i.i.d. regular fading statistics, rest of the proof holds irrespective of the fading statistics.

For second condition we need to show that $\mathrm{P}\left(r\right)_{ARQ,1} \dot{\leq} \mathrm{P}\left(r_L\right)_{ARQ,L}$. For first decoder with aggressive halting policy it follows that

$$\begin{aligned} d_e^2(\mathbf{H}_C, \Delta\mathbf{X}_C^{ARQ,1}) =& \|\theta\mathbf{H}_C\Delta\mathbf{X}_C^{ARQ,1}\|^2, \\ \geq& \sigma_1(\mathbf{H}_C^H\mathbf{H}_C)\|\theta\Delta\mathbf{X}_C^{ARQ,1}\|^2, \\ \dot{\geq}& \rho^{1-\frac{r}{n_T}-\epsilon}, \end{aligned}$$

where we have made use of the fact that $\|\theta\Delta\mathbf{X}_C^{ARQ,1}\|^2 \dot{\geq} \rho^{1-\frac{r}{n_T}}$ (cf. [36, Proof of Theorem 5]). Note that $\delta := 1 - \frac{r}{n_T} - \epsilon > 0$ for $0 \leq r \leq n_T - \epsilon$. In the limit $\epsilon \to 0$ whenever first decoder sends ACK the minimum euclidean distance $d_e^2(\mathbf{H}_C, \Delta\mathbf{X}_C^{ARQ,1}) > \rho^{\delta}, \; \delta > 0$ for $0 \leq r \leq n_T$. This implies that whenever ACK is sent only one received codeword can be inside the search sphere of radius $\xi = \sqrt{z\log\rho}$. Thus, from the definition of first decoder it is clear that error probability of the first decoder is the probability of the event that

- $\theta \boldsymbol{H}_C \mathbf{X}_C^{ARQ,1}$ is not included in the search sphere of radius ξ centered around $\boldsymbol{Y}_C^1$, or
- $\theta \boldsymbol{H}_C \hat{\mathbf{X}}_C^{ARQ,1}$ is the unique matrix included in the sphere for some erroneous codeword $\theta \hat{\mathbf{X}}_C^{ARQ,1}$.

Now $\mathrm{P}\left(r\right)_{ARQ,1}$ can be upper-bounded by the probability that $\theta \mathbf{X}_C^{ARQ,1}$ was transmitted and the additive noise was such that $\theta \boldsymbol{H}_C \mathbf{X}_C^{ARQ,1}$ is not in the search sphere of the corresponding received matrix $\boldsymbol{Y}_C^1$. We thus have

$$\mathrm{P}\left(r\right)_{ARQ,1} \leq \mathrm{P}\left(\|\tilde{\mathbf{w}}^1\|^2 > \xi^2\right).$$

where $\tilde{\mathbf{w}}^1 = -\alpha^2 (\mathbf{R}^1)^{-H} \mathbf{s} + (\mathbf{R}^1)^{-H} (\mathbf{M}^1)^H \mathbf{w}^1$, where $\mathbf{w}^1$ is the real valued representation of noise vector $\mathbf{W}_C^1$. Let

$$\begin{bmatrix} \mathbf{Q}_1^1 \\ \mathbf{Q}_2^1 \end{bmatrix} \mathbf{R}^1 = \mathbf{Q}^1 \mathbf{R}^1 \in \mathbb{R}^{(2n_R T + \kappa) \times \kappa}, \tag{5.32}$$

where $\mathbf{Q}_1^1 = (\mathbf{R}^1)^{-1} (\mathbf{M}^1) \in \mathbb{R}^{n \times m}$ and $\mathbf{Q}_2^1 = \alpha_r (\mathbf{R}^1)^{-1} \in \mathbb{R}^{m \times m}$, then it follows that

$$\begin{aligned} \tilde{\mathbf{w}}^1 &= -\alpha^2 (\mathbf{R}^1)^{-H} \mathbf{s} + (\mathbf{R}^1)^{-H} (\mathbf{M}^1)^H \mathbf{w}^1 \\ &= -\alpha_r (\mathbf{Q}_2^1)^H \mathbf{s} + (\mathbf{Q}_1^1)^H \mathbf{w}^1. \end{aligned}$$

Consequently we calculate

$$\begin{aligned} \mathrm{P}\left(\|\tilde{\mathbf{w}}^1\| > \xi\right) &\leq \mathrm{P}\left(\{\| - \alpha_r (\mathbf{Q}_2^1)^H \mathbf{s}\| + \|(\mathbf{Q}_1^1)^H \mathbf{w}^1\|\} > \xi\right), \\ &\overset{(a)}{=} \mathrm{P}\left(\left\{\| - \alpha_r (\mathbf{Q}^1)^H \begin{bmatrix} \mathbf{0} \\ \mathbf{s} \end{bmatrix} \| + \|(\mathbf{Q}^1)^H \begin{bmatrix} \mathbf{w}^1 \\ \mathbf{0} \end{bmatrix} \|\right\} > \xi\right), \\ &\leq \mathrm{P}\left(\left\{\sup_{\mathbf{s} \in \mathbb{S}_r^\kappa} \| - \alpha_r \mathbf{s}\| + \|\mathbf{w}^1\|\right\} > \xi\right), \\ &\overset{(b)}{\leq} \mathrm{P}\left(\{K + \|\mathbf{w}^1\|\} > \xi\right), \\ &= \mathrm{P}\left(\|\mathbf{w}^1\| > \{\xi - K\}\right), \\ &\overset{(c)}{\leq} \mathrm{P}\left(\|\mathbf{w}^1\|^2 > z_1 \log \rho\right), \\ &\dot{<} \quad \rho^{-d^*(\frac{r}{L})}. \end{aligned}$$

where (a) follows from (5.32), where the inequality in (b) follows for some fixed $K > 0$ independent of ρ such that $K \geq \sup_{\mathbf{s} \in \mathbb{S}_r^\kappa} \| - \alpha_r \mathbf{s}\|$, where the inequality in (c) follows for some arbitrary $z > z_1 > 0$ independent of ρ such that $(\xi - K)^2 \geq z_1 \log \rho$ and where the last asymptotic inequality follows for some $z_1 > d^*(\frac{r}{L}) > 0$.

Thus, the probability of excluding the transmitted codeword from the search is exponentially smaller than probability of error for the L-th round, i.e.,

$$\mathrm{P}\left(r\right)_{ARQ,1} \dot{<} \mathrm{P}\left(r_L\right)_{ARQ,L}.$$

Consequently, it follows that the degradations caused by MMSE-preprocessing in first round are not visible in the overall performance.

For third condition we note that the L-th round code is DMT optimal and is operating with effective multiplexing gain r_L. It then follows that $d_{ARQ,L}(r_L) = d^*(r_L)$ which implies that $d_{ARQ,L}(r_L) = d^*(\frac{r}{L})$. This proves Lemma 5. $\square$

Chapter 6

Complexity analysis for multiuser, cooperative and bidirectional networks

The work in this chapter is preliminary. We here present preliminary work on extending the rate-reliability-complexity analysis to simple instances of the multiple access, relay, and bidirectional channels, where again we identify the computational reserves that guarantee a DMT optimality, as well as address user/relay selection criteria and communication protocols that provide improved joint reliability-complexity performance in the presence of computational constraints.

We begin with the multiple access channel (MAC).

6.1 Multiple Access Channel

The work considers the symmetric MAC with K single-antenna transmitters and a receiver having n_R antennas shown in Fig. 6.1. We consider that each user operates at a multiplexing gain r, and that the receiver employs a joint ML (or joint lattice) decoder. The emphasis on joint decoders stems from their DMT optimality for symmetric MAC (cf. [72,73]). These joint ML or lattice decoders are implemented as bounded search sphere decoder (SD) algorithms with a search radius $\xi := \sqrt{z \log \rho}$, for a properly chosen $z > 0$, and where ρ denotes the signal to noise ratio.

In order to establish complexity requirements for the symmetric MAC, we quickly recall that the optimal DMT performance of the MAC system

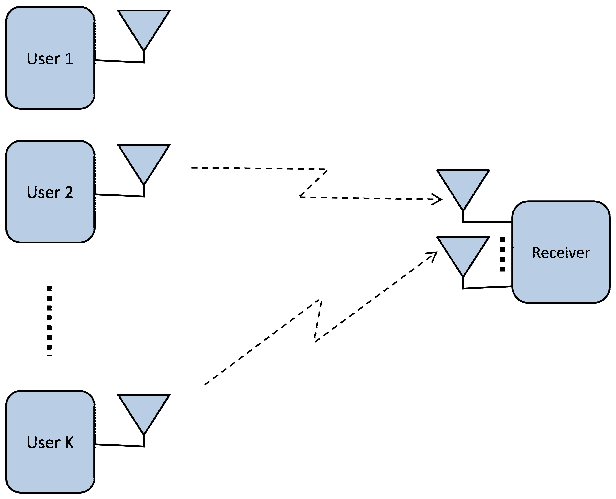


FIGURE 6.1 – Symmetric multiple access channel with K-users

under consideration is given by (cf. [72])

$$d^*_{mac}(r) = \begin{cases} n_R(1-r) & \text{for } r \leq \frac{n_R}{K+1}, \\ d^*(Kr) & \text{for } \frac{n_R}{K+1} < r \leq \frac{n_R}{K}, \end{cases} \tag{6.1}$$

where $d^*(r)$ denotes the optimal DMT of $K \times n_R$ channel. We note that the DMT region $d^*_{mac}(r)$ can be divided into two regimes, which are referred to as the lightly-loaded regime ($r \leq \frac{n_R}{K+1}$) where DMT is independent of multi-user interference, and the heavily-loaded regime ($\frac{n_R}{K+1} < r \leq \frac{n_R}{K}$) where specific lattice designs are required to achieve the DMT optimality (cf. [72]). For $K < n_R$, the MAC system always stays in the lightly loaded regime, and the DMT performance of the MAC is the same as if only a single user is active ; hence our interest is in the case of $K \geq n_R$, which we henceforth consider.

We focus on establishing upper bounds on the complexity exponent that guarantees DMT optimal ML-based (or lattice-based) decoding. This will be achieved by considering specific codes, decoders and halting policies, as these are presented in upcoming Proposition 9. It is interesting to note that for some ordering policies, these bounds are tight. The following holds for i.i.d. Rayleigh fading statistics.

Theorem 16. *For the $K \times n_R$ ($K \geq n_R$) MAC, the minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal DMT $d^*_{mac}(r)$, is upper bounded as*

$$c(r) \leq \overline{c}_{mac}(r) = \begin{cases} \overline{c}_v(r) & \textit{for } r \leq \frac{n_R}{K+1}, \\ \overline{c}_f(r) & \textit{for } \frac{n_R}{K+1} < r \leq \frac{n_R}{K}, \end{cases} \tag{6.2}$$

where

$$\begin{aligned}\overline{c}_v(r) = \max_{\boldsymbol{\mu}} \; & (K-n_R)r + \sum_{i=1}^{n_R} \left(r-(1-\mu_i)^+\right)^+ \\ s.t. \; & \sum_{i=1}^{n_R}(K-n_R+2i-1)\mu_i \le n_R(1-r), \\ & \mu_1 \ge \cdots \ge \mu_{n_R},\end{aligned}$$

where

$$\begin{aligned}\overline{c}_f(r) = & \left(K-n_R\right) rK_0 \\ & + K_0\left(r(n_R-\lfloor Kr\rfloor -1)+(\lfloor Kr\rfloor -r(K-1))^+\right),\end{aligned}$$

which is a piece-wise linear function that, for $r = 0, \frac{1}{K}, \cdots, \frac{n_R}{K}$*, takes the form*

$$\overline{c}_f(r) = K_0 Kr(1-r),$$

and where $K_0 = K$ *for* K *odd and* $K_0 = K+1$ *for* K *even.*

Proposition 9. *The above described bound* $\overline{c}_{mac}(r)$ *is sufficient to achieve the optimal DMT* $d^*_{mac}(r)$*, with uncoded QAM symbols for* $r \le \frac{n_R}{K+1}$*, lattice designs given in [73] for* $\frac{n_R}{K+1} < r \le \frac{n_R}{K}$*, given a sphere decoder with a search radius* $\xi > \sqrt{d^*_{mac}(r)\log\rho}$*, a decoding halting policy that halts decoding if* $N_{max} \ge \rho^x$ *for* $x > \overline{c}_{mac}(r)$*, and any decoding order policy. Furthermore there exists a fixed decoding order policy for which the above described bound is also necessary.*

The proofs for Theorem 16 and Proposition 9 are presented in Appendix 6A. To provide more meaningful insights regarding the upper bound presented in Theorem 16, we present examples for two specific cases with $n_R = 1$ and $n_R = K$.

Example 21. *For the specific case of* $n_R = 1$*, the bound in* (6.2) *takes the simple form*

$$\overline{c}_{mac}(r) = \begin{cases}(K-1)r & \text{for } r \le \frac{1}{K+1}, \\ (K-1)K_0 r & \text{for } \frac{1}{K+1} < r \le \frac{1}{K}.\end{cases}$$

Figure 6.2 describes the upper bounds for the $K = 4$ *and* $K = 5$ *user cases with single antenna receiver (*$n_{\mathrm{R}} = 1$*).*

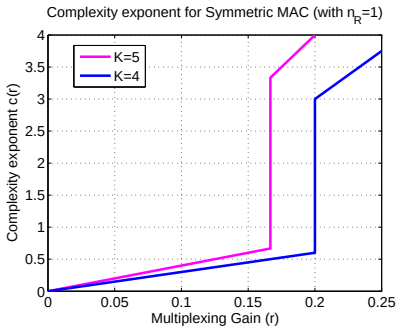


FIGURE 6.2 – Complexity bounds for K-user MAC with $n_{\mathrm{R}} = 1$

Example 22. *For the specific case of $n_R = K$, the optimal DMT of* (6.1) *can be achieved by V-BLAST (cf. [72]). The complexity exponent for this case takes the form (cf.* (6.18)*)*

$$c_{mac}(r) = r\left\lfloor\sqrt{K(1-r)}\right\rfloor + \left(r - 1 + \frac{K(1-r) - (\left\lfloor\sqrt{K(1-r)}\right\rfloor)^2}{2\left\lfloor\sqrt{K(1-r)}\right\rfloor + 1}\right)^+ .$$

Figure 6.3 describes the upper bounds for the $K = 3$ and $K = 4$ user cases with $n_R = K$.

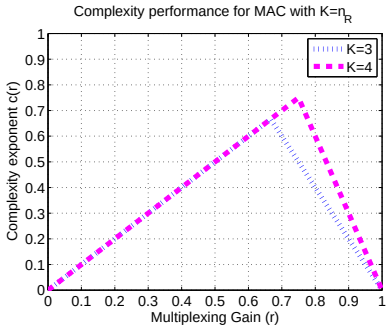


FIGURE 6.3 – Complexity exponent for K-user MAC with $n_{\mathrm{R}} = K$.

6.2 Relay channels with orthogonal amplify-and-forward protocol (OAF)

We consider a cooperative network with a source, $n-1$ relays and a destination, each having single transmit/receive antenna as shown in the Fig. 6.4. In this work we establish very early and preliminary results on the

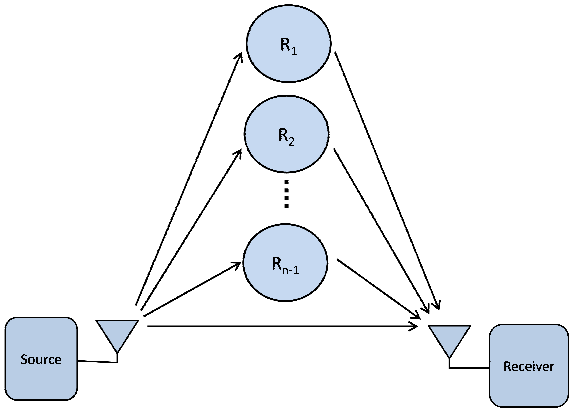


FIGURE 6.4 – System model for relay channel

complexity of achieving the optimal DMT of a cooperative network with a source, $n-1$ relays and a destination, each having single transmit/receive antenna and communicating over i.i.d. Rayleigh fading. This is done only for the orthogonal amplify-and-forward (OAF) protocol. For DMT optimal OAF protocol the source transmits a signal to the relays and to the destination for n channel uses. Over the next $n-1$ channel uses, each relay applies a linear transformation to the received signal and simultaneously transmits it to the destination, while source remains silent. The orthogonal amply-and-forward protocol is shown in Fig. 6.5. The optimal DMT of OAF protocol when the

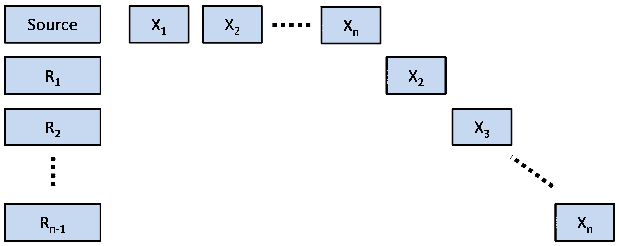


FIGURE 6.5 – System model for relay channel

number of relay is $n-1$ is given by

$$d_{oaf}(r) = \begin{cases} n\left(1 - \frac{(2n-1)r}{n}\right), & \text{for} \quad 0 \leq r \leq \frac{1}{2}, \\ 1-r, & \text{for} \quad \frac{1}{2} < r \leq 1. \end{cases}$$

The upper bound on complexity is described below, and as before it holds for ML-based decoding and for regularized lattice decoding. The following holds for i.i.d. Rayleigh fading statistics.

Proposition 10. *The minimum, over all lattice designs and halting and decoding order policies, complexity exponent $c(r)$ required to achieve the optimal*

DMT $d_{oaf}(r)$, is upper bounded as

$$c(r) \leq \overline{c}_{oaf}(r) = \frac{2n-1}{n} r(n - \lfloor (2n-1)r \rfloor - 1) + \left(\lfloor (2n-1)r \rfloor - \frac{2n-1}{n} r(n-1) \right)^{+},$$

which is a piecewise linear function that, for $r = 0, \frac{1}{2n-1}, \cdots, \frac{n}{2n-1}$, takes the form

$$\overline{c}_{oaf}(r) = (2n-1)r(1 - \frac{2n-1}{n} r).$$

The above described bound $\overline{c}_{oaf}(r)$ is sufficient to achieve the optimal DMT $d_{oaf}(r)$, with lattice designs and OAF protocol of [34], given a sphere decoder with a search radius $\xi > \sqrt{d_{oaf}(r) \log \rho}$, a decoding halting policy that halts decoding if $N_{max} \geq \rho^x$ for $x > \overline{c}_{oaf}(r)$, and any decoding order policy. Furthermore there exists a fixed decoding order policy for which the above described bound is also necessary.

The proof for Proposition 10 is presented in Appendix 6B. The complexity exponent for single and two relay case in shown in Fig. 6.6.

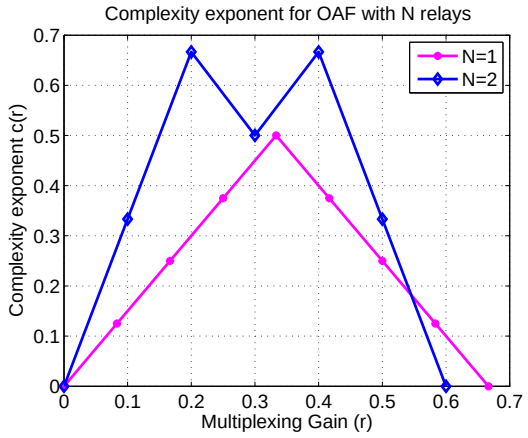


FIGURE 6.6 – Complexity performance of OAF

It is clear that decoding complexity scales exponentially with the number of relay used in communication. In the presence of complexity constraints a subset of relays can be selected (say set of n^* relays) to keep decoding complexity within the allocated computational resources while achieving the optimal DMT of smaller relay network. A schematic depicting the relay selection scheme is shown in in Fig. 6.7.

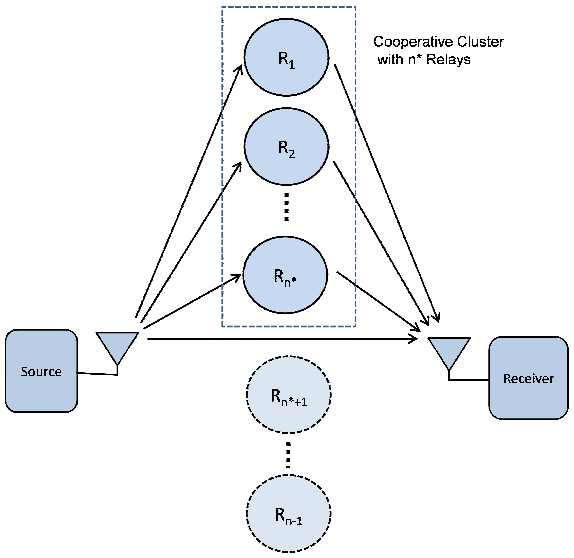


FIGURE 6.7 – Relay cluster selection in the presence of run-time constraints

6.3 Two-way relay channel (TRC)

In this section we demonstrate how reliability-complexity measure could be utilized as a unified performance metric to compare worth of different communication protocols over a non-separated two-way relay channel (TRC). In a TRC communication nodes exchange messages with cooperation of an intermediate relay node employing intelligent two-way relaying strategies [74], [75]. The fundamental advantage of TRC over classical one-way relay channels is that the duplexing loss due to half-duplex constraint (a node cannot transmit and receive simultaneously) can be avoided.

We consider a TRC with an asymmetric fading channel model, which corresponds to a very pertinent communication scenario where source-relay links are statistically different from the direct link between sources. In our channel model, the source relay links are i.i.d. Rayleigh distributed but the direct link between sources is left more general as i.i.d. Nakagami-m (c.f. [76]) distributed (Nakagami-m includes Rayleigh for $m = 1$). We refer to this channel model as a *non-separated two-way relay channel* (ns-TRC) because of existence of direct link between sources. This model is more general than many TRC scenarios considered in previous works (c.f. [77], [78], [79], [80], [81]). We assume that all the channels are frequency flat, quasi-static, and they are all independent of each other. We assume perfect channel state information (CSI) at the receiver (CSIR) of each link, but no CSI at the transmitters (CSIT).

The system model consists of two source nodes A and B who want to exchange information in the presence of an assisting relay node R as shown

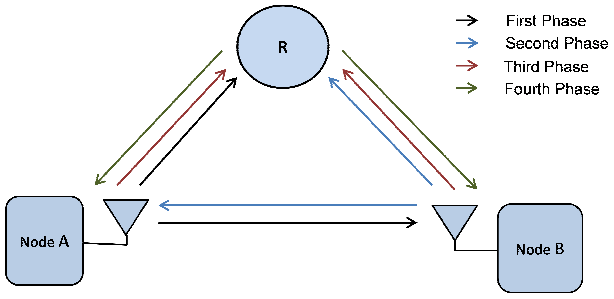


FIGURE 6.8 – non-separated Two-way Relay Channel model

in Fig. 6.8. The relay employs decode-and-forward (DF) strategy and does not have any information of its own to transmit. Each node uses a single antenna and operates under half-duplex constraint. Note that this system model is statistically symmetrical with respect to the two source nodes.

Communication between nodes A and B takes place via a four-phase hybrid broadcast (HBC) protocol as shown in Fig. 6.8 with the direction of arrow indicating the transmit/receive mode of each node. For instance, in the first phase node A transmits while all other nodes listen. We denote the fraction of time-slot allocated for the i^{th} phase with Δ_i, and hence $\sum \Delta_i = 1$. The third and the fourth phase are called *multiple access channel (MAC) phase* and *broadcast channel (BC) phase* respectively. Note that this is the most general scheduling possible, with the half-duplex constraint imposed.

The relative phase durations affect both the error performance as well as the achievable transmission rates. We want to investigate for any possible gains in the achievable DMT and reliability-complexity measure using four-phase HBC protocol, as compared to three-phase time division broadcast (TDBC) protocol. TDBC protocol is obtained by removing third MAC phase from the HBC protocol.

6.3.1 Diversity-Multiplexing Tradeoff

The following holds for four-phase protocol for ns-TRC DF channel.

Lemma 6. *For the asymmetric fading model, the DMT for a static four-phase protocol with fixed phase-durations* $\Delta_1, \Delta_3, \Delta_4$ *satisfying* $\Delta_1 = \Delta_2$ *and* $\Delta_1 \geq m\Delta_4$*, is*

$$d_{4-ph}(\Delta_i, r) = \min\{(d_1(r) + d_2(r)), d_3(r)\}. \tag{6.3}$$

where

$$d_1(r) = m(1 - \frac{r}{\Delta_1}) \quad for \quad 0 \le r \le \Delta_1, \tag{6.4}$$

$$d_2(r) = \begin{cases} 1 - \frac{r}{\Delta_1+\Delta_3} & for \quad 0 \le r \le \Delta_1 \\ 1 + \frac{\Delta_1 - 2r}{\Delta_1+\Delta_3} & for \quad \Delta_1 < r \le \Delta_1 + \frac{\Delta_3}{2}, \end{cases} \tag{6.5}$$

$$d_3(r) = \begin{cases} 1 + m - \frac{r}{\Delta_4} & for\ r \le \Delta_4 \\ m(1 + \frac{\Delta_4 - r}{\Delta_1}) & for\ \Delta_4 < r \le \Delta_1 + \Delta_4 \end{cases} \tag{6.6}$$

The proof for this lemma is presented in Appendix 6C.

Remark 3.
- *For a static protocol with $\Delta_1 < m\Delta_4$, the DMT expression is similar as in Theorem 6, but with*

$$d_3(r) = \begin{cases} 1 + m - \frac{mr}{\Delta_1} & for\ r \le \Delta_1 \\ 1 + \frac{\Delta_1}{\Delta_4} - \frac{r}{\Delta_4} & for\ \Delta_1 < r \le \Delta_1 + \Delta_4 \end{cases}$$

- *For $m = 1$, we get the DMT for symmetric Rayleigh fading model.*

In order to obtain optimal DMT of ns-TRC DF system, we need to solve the following problem for any r :

$$d_{OPT}(r) = \max_{\Delta_1, \Delta_3, \Delta_4} d_{4-ph}(\Delta_i, r)$$
$$\text{subject to} \qquad 2\Delta_1 + \Delta_3 + \Delta_4 = 1, \qquad \Delta_i \ge 0. \tag{6.7}$$

Optimal DMT for $m \ge 1$

The optimal DMT achievable for $m \ge 1$, which is a situation that can be interpreted as the direct link having similar or a more stable fading statistic compared to source-relay links, is given by :

Theorem 17. *For the given system settings with $m \ge 1$, optimal DMT is*

$$d_{OPT}(r) = \begin{cases} 1 + m - (2m+3)r & for\ r \le \frac{1}{(2m+3)} \\ \frac{m(1+m)(1-2r)}{r+m} & for\ \frac{1}{(2m+3)} < r \le \frac{1}{2} \end{cases}, \tag{6.8}$$

and it can be achieved using TDBC protocol with

$$\Delta_1 = \begin{cases} \frac{(m+1)}{(2m+3)} & for\ r \le \frac{1}{(2m+3)} \\ \frac{r+m}{1+2m} & for\ \frac{1}{(2m+3)} < r \le \frac{1}{2} \end{cases} \tag{6.9}$$

$\Delta_2 = \Delta_1$, $\Delta_3 = 0$ *and* $\Delta_4 = 1 - 2\Delta_1$.

Corollary 17a. *For the Rayleigh fading case ($m = 1$), the optimal DMT $d_{OPT}(r)$ is attained by an orthogonal three-phase TDBC protocol.*

The proofs for Theorem 17 and Corollary 17a are presented in Appendix 6D.

Optimal DMT for $\frac{1}{2} \leq m < 1$

Now we consider the case $\frac{1}{2} \leq m < 1$, which corresponds to the case where direct link between sources is less stable compared to source-relay links.

Theorem 18. *For the given system settings with $\frac{1}{2} \leq m < 1$, optimal DMT is*

$$d_{OPT}(r) = 1 + m - \frac{r}{\kappa} \quad for \quad r \leq \kappa, \tag{6.10}$$

where,

$$\kappa = \frac{(2+m) - (1-m)\Delta_3 - X}{3+2m},$$
$$X = \sqrt{(m^2+4)\Delta_3^2 + (2m^2+2m-4)\Delta_3 + (m+1)^2}$$

and it can be achieved using HBC protocol with

$$\Delta_3 = \frac{(1-m)(2+m-2\sqrt{m})}{4+m^2}, \quad \Delta_4 = \kappa,$$

and

$$\Delta_2 = \Delta_1 = \frac{1-\Delta_3-\Delta_4}{2}.$$

The proof for Theorem 18 is presented in Appendix 6E. For $\kappa < r \leq \frac{1}{2}$, we do not have a closed form expression for the optimal DMT, however, this optimization problem can be solved using linear programming. A discussion on this optimization problem is presented in Appendix 6E. Numerical results are shown in Fig. 6.9 for $m = \frac{1}{2}$. Theorem 18 establishes DMT optimality of HBC protocol for the case where direct link between sources is less stable compared to source-relay links.

6.3.2 Complexity Analysis

The overall complexity of the communication protocol is the sum total of the number of flops spent in each phase of the communication, i.e., the total complexity of the communication protocol takes the form

$$N_{protocol} = \sum_{i=1}^{p} N_i,$$

where N_i is the number of flops spent during phase i, and where p is the total number of phases for the protocol. Let $N_i \doteq \rho^{c_i(r)}$, where $c_i(r)$ is the complexity exponent for i-th phase, then the overall complexity exponent for the protocol is given by

$$c_{protocol}(r) = \max\{c_1(r), \cdots, c_p(r).\} \tag{6.11}$$

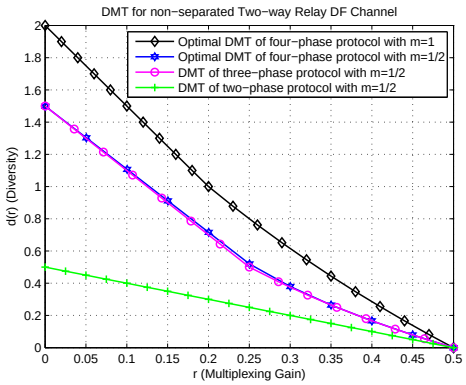


FIGURE 6.9 – DMT Comparison

Complexity of TDBC protocol

In order to perform complexity analysis for the TDBC and HBC protocols, we first define the decoding operations that are required for these protocols. For TDBC protocol in first phase SISO decoding is performed at node B and relay R ($c_{1,tdbc}(r) = 0$), in the second phase again SISO decoding is performed at node A and relay R ($c_{2,tdbc}(r) = 0$). In the third phase each of the source nodes A and B perform decoding after removing self-interference term, which is again equivalent to SISO decoding ($c_{3,tdbc}(r) = 0$). Hence, the overall complexity for TDBC protocol is the is given by (cf. (6.11))

$$c_{TDBC}(r) = 0.$$

Complexity of HBC protocol

For HBC protocol, the first two rounds and the last round are similar to TDBC protocol, i.e., $c_{1,hbc}(r) = c_{2,hbc}(r) = c_{4,hbc}(r) = 0$, but there is an additional third round where decoding is performed at relay R over a multiple access channel with sources nodes A and B. The complexity exponent for this third round is given by (cf. Theorem 16)

$$c_{3,hbc}(r) = \begin{cases} \Delta_3 r & \text{for } r \leq \frac{1}{3}, \\ 3\Delta_3 r & \text{for } \frac{1}{3} < r \leq \frac{1}{2}. \end{cases}$$

It follows that the overall complexity exponent for the HBC takes the form (cf. (6.11))

$$c_{HBC}(r) = \begin{cases} \Delta_3 r & \text{for } r \leq \frac{1}{3}, \\ 3\Delta_3 r & \text{for } \frac{1}{3} < r \leq \frac{1}{2}. \end{cases}$$

6.3.3 Reliability-Complexity Measure for HBC and TDBC Protocols

The question arises as to whether it is best to use the TDBC protocol that provides smaller DMT (line marked with circles in Fig. 6.9) for $\frac{1}{2} \leq m < 1$ with zero complexity exponent, or the HBC protocol that provides for higher DMT (line marked with stars in Fig. 6.9) but does so with much increased complexity. Towards deciding this, one must first decide what is the price of flops, compared to the price of errors, and design a joint rate-reliability-complexity measure. Such a measure could for example be chosen to take the form $\Gamma(r) = d(r) - \gamma c(r)$ where again γ compares the price of flop vs error. As shown in Fig. 6.10 the reliability-complexity measure (with $\gamma = 1$) reveals that even for the statistically weaker direct link ($\frac{1}{2} \leq m < 1$) the TDBC protocol provide higher reliability-complexity performance as compared to HBC protocol. In fact for all $\gamma \geq 1$ TDBC protocol results in higher reliability-complexity measure.

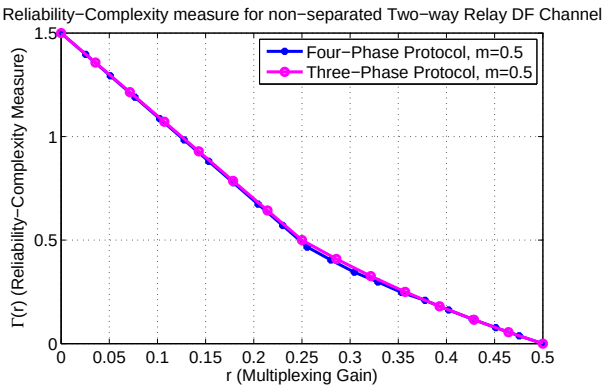


FIGURE 6.10 – Reliability-Complexity Measure

Summary

The analysis indicates that for a DMT optimal protocol, duration of MAC phase decreases with improvement in direct link fading stability relative to source-relay links. For $m \geq 1$, which is a situation that can be interpreted as the direct link having similar or a more stable fading statistic compared to source-relay links, the optimal DMT is shown to be achieved by the orthogonal three-phase TDBC protocol. The operational meaning of this result is that the MAC phase of HBC protocol is not necessary to achieve optimal performance, this results in a simplification of the communication protocol. For less stable direct link ($\frac{1}{2} \leq m < 1$), the MAC phase of HBC protocol is necessary to achieve optimal DMT but a simpler TDBC protocol achieves preferable joint reliability-complexity performance.

Appendix 6A : Proof of Theorem 16

In the following we establish an upper bound on the minimum complexity exponent $c(r)$ required by ML-based decoding to achieve optimal DMT of MAC. The joint ML decoder sees a $K \times n_{\mathrm{R}}$ equivalent MIMO-MAC channel with sum multiplexing gain $r_t = Kr$, the optimal DMT in terms of sum rate r_t can be described as

$$d_{joint}(r_t) = \begin{cases} n_R(1 - \frac{r_t}{K}) & \text{for } r_t \leq \frac{n_R K}{K+1}, \\ d^*(r_t) & \text{for } \frac{n_R K}{K+1} < r_t \leq n_R. \end{cases} \tag{6.12}$$

For complexity analysis we proceed with representation equivalent MIMO-MAC channel $\mathbf{H}_C \in \mathbb{R}^{n_R \times K}$ with T channel uses (cf. (3.63))

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{w}, \tag{6.13a}$$

$$\text{where} \quad \mathbf{M} \triangleq \rho^{\frac{1}{2} - \frac{r_t T}{\kappa}} \mathbf{H}\mathbf{G} \in \mathbb{R}^{2n_R T \times \kappa}, \tag{6.13b}$$

where $\mathbf{H} = I_T \otimes \mathbf{H}_R \in \mathbb{R}^{2n_R T \times 2KT}$, $\mathbf{H}_R$ is real valued representation of $\mathbf{H}_C$ and where $\mathbf{G} \in \mathbb{R}^{2KT \times \kappa}$. As $\kappa = 2KT > 2n_R T$, we need MMSE-preprocessing to have SD implementation of ML decoding. Let $\mathbf{QR} = \tilde{\mathbf{M}}$ be the thin QR factorization of the MMSE-preprocessed code-channel matrix $\tilde{\mathbf{M}} \triangleq \begin{bmatrix} \mathbf{M} \\ \alpha \mathbf{I} \end{bmatrix}$ where $\alpha = \rho^{\frac{-r_t}{2K}}$, then (6.13a) yields

$$\mathbf{r} = \mathbf{R}\mathbf{s} + \mathbf{w}',$$

where $\mathbf{r} = \mathbf{R}^{-H}\mathbf{M}^H\mathbf{y}$, and where $\mathbf{w}' = -\alpha^2 \mathbf{R}^{-H}\mathbf{s} + \mathbf{R}^{-H}\mathbf{M}^H\mathbf{w}$. The ML decoder for this system takes the form

$$\hat{\mathbf{s}}_{ML} = \arg \min_{\hat{\mathbf{s}} \in \mathbb{S}_r^{\kappa}} \left\| \mathbf{r} - \mathbf{R}\hat{\mathbf{s}} \right\|^2, \tag{6.14}$$

that is implemented by the sphere decoder which recursively enumerates all candidate vectors $\hat{\mathbf{s}} \in \mathbb{S}_r^{\kappa}$ within a given search sphere of radius $\xi = \sqrt{z \log \rho}$, for some $z > d^*_{mac}(r)$. To compute an upper bound on the complexity exponent, we follow the approach similar to Chapter 3 and define $\mu_i \triangleq -\frac{\log \sigma_i(\mathbf{H}_C^H \mathbf{H}_C)}{\log \rho}$ with $\mu_1 \geq \cdots \geq \mu_{n_R}$. It then follows that

$$\sigma_i(\mathbf{R}) = \sigma_i(\tilde{\mathbf{M}}) = \sqrt{\alpha^2 + \sigma_i(\mathbf{M}^H\mathbf{M})},$$

which results in

$$\sigma_i(\mathbf{R}) \dot{\geq} \begin{cases} \rho^{\frac{-r_t}{2K}} \text{ for } i = 1, \cdots, 2T(K - n_R), \\ \rho^{\frac{-r_t}{2K} + \frac{1}{2}(1 - \mu_{\iota_{2T}(i - 2T(K-n_R))})^+} \\ \quad \text{for } i = 2T(K - n_R) + 1, \cdots, 2KT, \end{cases} \tag{6.15}$$

where we have made use of the fact that $\sigma_{\min}(\mathbf{G}) \doteq \rho^0$ and where $\iota_{2T}(i) \triangleq \left\lceil \frac{i}{2T} \right\rceil$.

SD Complexity

The total number of visited nodes is commonly taken as a measure of the SD complexity (cf. [7]). For any given channel realization $\boldsymbol{\mu} \triangleq (\mu_1, \cdots, \mu_m)$, this total number of visited nodes is given by

$$N_{SD}(\boldsymbol{\mu}) \triangleq \sum_{k=1}^{\kappa} N_k(\boldsymbol{\mu}).$$

where from [7, Lemma 1] we know that

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + \min\{ \frac{2\xi}{\sigma_i(\mathbf{R}_k)}, 2\sqrt{k}\rho^{\frac{r_t T}{\kappa}} \} \right].$$

From the interlacing property of singular values of sub-matrices [53] we have that $\sigma_i(\mathbf{R}_k) \geq \sigma_i(\mathbf{R})$. It follows that

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + \min\left\{ \frac{2\xi}{\sigma_i(\mathbf{R})}, 2\sqrt{k}\rho^{\frac{r_t T}{\kappa}} \right\} \right].$$

Consequently, we have that

$$\begin{aligned} N_{SD}(\boldsymbol{\mu}) \leq & \sum_{k=1}^{\kappa} \left\{ \prod_{i=1}^{k} \left[\sqrt{k} + \min\left\{ \frac{2\xi}{\sigma_i(\mathbf{R})}, 2\sqrt{k}\rho^{\frac{r_t T}{\kappa}} \right\} \right] \right\}, \\ \dot{\leq} & \rho^{(1-\frac{n_R}{K}) r_t T + \sum_{i=1}^{2n_R T} \min\left(\frac{r_t}{2K} - \frac{1}{2}(1-\mu_{\iota_{2T}(i)})^+, \frac{r_t}{2K} \right)^+}, \\ = & \rho^{(1-\frac{n_R}{K}) r_t T + T \sum_{i=1}^{n_R} \left(\frac{r_t}{K} - (1-\mu_i)^+ \right)^+}, \end{aligned} \tag{6.16}$$

where for the last equality we made use of the $2T$ multiplicity of the singular values of $\mathbf{H}$, and of the fact that

$$\min\left(\frac{r_t}{K} - (1-\mu_1)^+, \frac{r_t}{K} \right)^+ = \left(\frac{r_t}{K} - (1-\mu_1)^+ \right)^+.$$

Following the footsteps of the complexity analysis in Chapter 3 for the MMSE-preprocessed channel matrix $\mathbf{R}$, the upper bound on the complexity exponent can be obtained as the solution to a constrained maximization problem according to $c(r) \leq \overline{c}_{mac}(r_t)$, where

$$\overline{c}_{mac}(r_t) \triangleq \max_{(\mu)} \left(1 - \frac{n_R}{K}\right) r_t T + T \sum_{i=1}^{n_R} \left(\frac{r_t}{K} - (1-\mu_i)^+ \right)^+ \tag{6.17a}$$

$$\text{s.t.} \quad I((\mu)) \leq d_{joint}(r_t), \tag{6.17b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_R}, \tag{6.17c}$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_\kappa)$ satisfies the large deviation principle with rate function $I(\boldsymbol{\mu})$.

Multiplexing gain region $r_t \leq \frac{n_R K}{K+1}$

Towards establishing tight complexity bounds, we vary the dimensionality of the codes, and by noting that uncoded QAM is optimal for $r_t \leq \frac{n_R K}{K+1}$, we get for the same multiplexing gain region that

$$\begin{aligned}\overline{c}_{mac}(r_t) = \max_{\boldsymbol{\mu}} \ & \left(1-\frac{n_R}{K}\right) r_t + \sum_{i=1}^{n_R} \left(\frac{r_t}{K} - (1-\mu_i)^+\right)^+ \\ \text{s.t.} \ & \sum_{i=1}^{n_R} (K - n_R + 2i - 1)\mu_i \leq n_R(1-\frac{r_t}{K}), \\ & \mu_1 \geq \cdots \geq \mu_{n_R}.\end{aligned}$$

For the case of $n_R = 1$, the above simplifies to

$$\overline{c}_{mac}(r_t) = \left(1-\frac{1}{K}\right) r_t,$$

and for the case of $n_R = K$ it simplifies to

$$\begin{aligned}\overline{c}_{mac}(r_t) =& \frac{r_t \left\lfloor \sqrt{n_R - r_t} \right\rfloor}{n_R} \\ &+ \left(\frac{r_t}{n_R} - 1 + \frac{n_R - r_t - (\left\lfloor \sqrt{n_R - r_t} \right\rfloor)^2}{2\left\lfloor \sqrt{n_R - r_t} \right\rfloor + 1}\right)^+ . \end{aligned} \tag{6.18}$$

Multiplexing gain region $\frac{n_R K}{K+1} < r_t \leq n_R$

For $\frac{n_R K}{K+1} < r_t \leq n_R$, we consider the DMT optimal lattice designs proposed in [73]. These designs incur $T = K_0$ and $\kappa = 2KK_0$, where $K_0 = K$ for K odd and $K_0 = K+1$ for K even. Consequently, the upper bound for $\frac{n_R K}{K+1} < r_t \leq n_R$ is given by

$$\begin{aligned}\overline{c}_{mac}(r_t) = \max_{\mu_1} \ & (1-\frac{n_R}{K}) r_t T + T \sum_{i=1}^{n_R} \left(\frac{r_t}{K} - (1-\mu_i)^+\right)^+ \\ \text{s.t.} \ & I((\mu)) \leq d^*(r_t), \\ & \mu_1 \geq \cdots \geq \mu_{n_R}.\end{aligned}$$

It can be shown that the solution to this optimization problem takes the form

$$\begin{aligned}\overline{c}_{mac}(r) =& \left(1-\frac{n_R}{K}\right) r_t K_0 \\ &+ \frac{K_0}{K}\left(r_t(n_R - \lfloor r_t \rfloor - 1) + (K\lfloor r_t \rfloor - r_t(K-1))^+\right),\end{aligned}$$

of a piece-wise linear function that, for integer r_t, is of the form

$$\overline{c}_{mac}(r) = \frac{K_0}{K} r_t(K - r_t).$$

Recalling that $r_t = Kr$, results in the bound specified in the theorem.

Furthermore, following the footsteps of the proof of Proposition 1 it can be shown that with QAM symbols for $r \leq \frac{n_R}{K+1}$, lattice designs given in [73] for $\frac{n_R}{K+1} < r \leq \frac{n_R}{K}$, given a sphere decoder with a search radius $\xi > \sqrt{d_{mac}(r) \log \rho}$, a decoding halting policy that halts decoding if $N_{max} \geq \rho^x$ for $x > \overline{c}_{mac}(r)$, there exists one fixed decoding order for which the above described bound is also necessary. This proves Theorem 16 and Proposition 9. □

Appendix 6B : Proof of Proposition 10

Let h_{SD}, h_{SR_j} and h_{R_jD}, for $j = 2, \cdots, n$, denote the channel fading coefficients for source-destination channel, source-to-j^{th} relay channel and j^{th} relay-to-destination channel respectively. The induced channel model for this OAF protocol takes the following form (cf. [34]) :

$$\mathbf{H}_{oaf} = \begin{bmatrix} h_{SD} & 0 & \cdots & 0 \\ 0 & h_{SR_2}h_{R_2D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{SR_n}h_{R_nD} \end{bmatrix}. \tag{6.19}$$

The overall system model takes the form

$$\mathbf{y} = \theta \begin{bmatrix} h_{SD} & 0 & \cdots & 0 \\ 0 & h_{SR_2}h_{R_2D} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{SR_n}h_{R_nD} \end{bmatrix} \mathbf{x} + \mathbf{w}, \tag{6.20}$$

where $\theta^2 = \rho^{1-\frac{2n-1}{n}r}$, for the DMT optimal performance the code vector $\mathbf{x}$ is given by (cf. [34])

$$\mathbf{x} = [\ell_0 \sigma(\ell_0) \cdots \sigma^{n_T-1}(\ell_0)]^T,$$

where $\ell_0 \in \mathcal{A}_{QAM}(\beta_1, \cdots, \beta_{n_T})$ (see below (6.21)) and where σ is the generator of the cyclic Galois group $Gal(\mathbb{L}/\mathbb{F})$ with $\mathbb{L}$ being a degree-n_T cyclic Galois extension field of $\mathbb{F} = \mathbb{Q}(i)$, where $\mathbb{Q}(i) = \{a + ib | a, b, \in \mathbb{Q}\}$. $\mathbb{Q}$ is the set of all rational numbers. Let $\mathcal{O}_{\mathbb{F}}$ and $\mathcal{O}_{\mathbb{L}}$ denote the ring of algebraic integers in $\mathbb{F}$ and $\mathbb{L}$, respectively. Let $\{\beta_1, \cdots, \beta_{n_T}\}$ be an integral basis for

$\mathcal{O}_{\mathbb{L}}/\mathcal{O}_{\mathbb{F}}$ and for M even, let $\mathcal{A}_{QAM}$ denote the M^2-QAM constellation given by $\mathcal{A}_{QAM} = \{a+ib||a|,|b| \leq M-1, a, b, \text{odd}\}$ and

$$\mathcal{A}_{QAM}(\beta_1, \cdots, \beta_{n_T}) = \left\{ \sum_i a_i \beta_i | a_i \in \mathcal{A}_{QAM} \right\}. \tag{6.21}$$

Following the footsteps of the complexity analysis presented in Section 3.2.2 and defining $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}_{oaf}^H \mathbf{H}_{oaf})}{\log \rho}$, $j = 1, \cdots, n$, the complexity exponent required to achieve a vanishing performance gap is upper bounded as

$$c(r) \leq \overline{c}_{oaf}(r) = \max_{\boldsymbol{\mu}} \sum_{j=1}^{n} \min\left(\frac{2n-1}{n} r - (1-\mu_j), r\right)^+ \tag{6.22a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \mu_j \leq n(1 - \frac{2n-1}{n} r), \tag{6.22b}$$

$$\mu_1 \geq \cdots \geq \mu_n \geq 0. \tag{6.22c}$$

The solution to this optimization problem takes the form

$$\overline{c}_{oaf}(r) = \frac{2n-1}{n} r(n - \lfloor (2n-1)r \rfloor - 1) + (\lfloor (2n-1)r \rfloor - \frac{2n-1}{n} r(n-1))^+, \tag{6.23}$$

which for $r = 0, \frac{1}{2n-1}, \cdots, \frac{n}{2n-1}$ simplifies to

$$\overline{c}_{oaf}(r) = (2n-1)r(1 - \frac{2n-1}{n} r).$$

The comparison of the OAF complexity exponent in (6.23) with the complexity exponent for $n \times 1$ quasi-static MISO channel Proposition 2 reveals that

$$c_{oaf}(r) = \overline{c}_{miso}(\frac{2n-1}{n} r).$$

Furthermore, following the footsteps of the proof of Proposition 1 it can be shown that with lattice designs and OAF protocol of [34], given a sphere decoder with a search radius $\xi > \sqrt{d_{oaf}(r) \log \rho}$, a decoding halting policy that halts decoding if $N_{max} \geq \rho^x$ for $x > \overline{c}_{oaf}(r)$, there exists one fixed decoding order for which the above described bound is also necessary. This proves Proposition 10. □

Appendix 6C : Proof of Lemma 6

Proof. (sketch) We here derive the DMT $d_{4-ph}(\Delta_i, r)$ for a *static* four-phase protocol with fixed phase durations. We are interested in the case where both source nodes transmit with the same rate R, and also, both demand the same

DMT performance. Since the system model is symmetric with respect to the two source nodes, there is no loss of generality in making the two phase durations Δ_1 and Δ_2 equal.

To analyze the system performance for fixed phase durations, we define the average error probability $P_e = P[\{\mathcal{E}_{\rm A} \cup \mathcal{E}_{\rm B}\}]$, where $\mathcal{E}_{\rm A}$ and $\mathcal{E}_{\rm B}$, defined later, denote the error events at the receive node A and B respectively.

In order to analyze the information flow from source node A to B, we define the following error event $\mathcal{E}_{\rm B} \triangleq \{\overline{\overline{\mathcal{E}}_{\rm B,1} \cup \{\overline{\mathcal{E}}_{\rm R,3} \cap \overline{\mathcal{E}}_{\rm B,4}\}}\}$, where denotes complement of , and where

- $\overline{\mathcal{E}}_{\rm B,1}$ - occurs if B is able to decode message from A at the end of phase 1.
- $\overline{\mathcal{E}}_{\rm R,3}$ - occurs if R is able to decode message from A and B at the end of phase 3.
- $\overline{\mathcal{E}}_{\rm B,4}$ - occurs if B is able to decode message from A at the end of phase 4, conditioned on the occurrence of event $\overline{\mathcal{E}}_{\rm R,3}$.

Note that $\overline{\mathcal{E}}_{\rm B,1} \subset \overline{\mathcal{E}}_{\rm B,4}$. Error event corresponding to $\mathcal{E}_{\rm A}$ can be similarly defined by interchanging A and B in the above expressions. With these definitions, we have the upper bound

$$\begin{aligned} P[\mathcal{E}_{\rm B}] &= P[\overline{\overline{\mathcal{E}}_{\rm B,1} \cup \{\overline{\mathcal{E}}_{R,3} \cap \overline{\mathcal{E}}_{\rm B,4}\}}] \\ &= P[\mathcal{E}_{\rm B,1} \cap \{\mathcal{E}_{R,3} \cup \mathcal{E}_{\rm B,4}\}] \\ &= P[\{\mathcal{E}_{\rm B,1} \cap \ \mathcal{E}_{R,3}\} \cup \{\mathcal{E}_{\rm B,1} \cap \mathcal{E}_{\rm B,4}\}] \\ &= P[\{\mathcal{E}_{\rm B,1} \cap \ \mathcal{E}_{R,3}\} \cup \mathcal{E}_{\rm B,4}] \\ &\overset{(f)}{\leq} P[\{\mathcal{E}_{\rm B,1} \cap \ \mathcal{E}_{R,3}\}] + P[\mathcal{E}_{\rm B,4}] \\ &= P[\mathcal{E}_{\rm B,1}]P[\mathcal{E}_{R,3}] + P[\mathcal{E}_{\rm B,4}], \end{aligned} \tag{6.24}$$

where (f) follows from union bound, . We also have a lower bound,

$$\max\{P[\mathcal{E}_{\rm B,1}]P[\mathcal{E}_{R,3}], P[\mathcal{E}_{\rm B,4}]\} \leq P[\mathcal{E}_{\rm B}]. \tag{6.25}$$

Bounds corresponding to $\mathcal{E}_{\rm A}$ can be obtained similarly by replacing B with A in the expressions (6.24) and (6.25).

Now, since

$$\max\{P[\mathcal{E}_{\rm A}], P[\mathcal{E}_{\rm B}]\} \leq P_e \leq P[\mathcal{E}_{\rm A}] + P[\mathcal{E}_{\rm B}], \tag{6.26}$$

if we can derive the precise optimum SNR exponents for each of the $P[\mathcal{E}_{B,1}]$, $P[\mathcal{E}_{R,3}]$ and $P[\mathcal{E}_{B,4}]$ defined above and also show that $P[\mathcal{E}_{\rm A}] \doteq P[\mathcal{E}_{\rm B}]$, then we have the DMT expression for the static four phase protocol.

Toward this end, we define the outage events for information flow from A to B. Let $X_J^{(i)}$ denote the transmit signal, $Y_J^{(k)}$ denote the receive signal for any node $J \in \{\rm A, B, R\}$, in the i-th and the k-th phase, $i, k \in \{1, 2, 3, 4\}$. The outage event $\mathcal{O}_{J,k}$ for the respective error event $\mathcal{E}_{J,k}$ is defined as :

$$\mathcal{O}_{\mathrm{B},1} \triangleq \{h_{AB} : \Delta_1 I(X_{\mathrm{A}}^{(1)}; Y_{\mathrm{B}}^{(1)} | h_{AB}) < R\}, \tag{6.27}$$

$$\mathcal{O}_{\mathrm{R},3} = \{\mathcal{O}_{\mathrm{R},3,1} \cup \mathcal{O}_{\mathrm{R},3,2}\}$$

$$\begin{aligned}\mathcal{O}_{\mathrm{R},3,1} \triangleq \Big\{ & (h_{A\mathrm{R}}, h_{B\mathrm{R}}) : \Delta_1 I(X_{\mathrm{A}}^{(1)}; Y_{\mathrm{R}}^{(1)} | h_{A\mathrm{R}}) \\ & + \Delta_3 I(X_{\mathrm{A}}^{(3)}; Y_{\mathrm{R}}^{(3)} | X_{\mathrm{B}}^{(3)}, h_{A\mathrm{R}}) < R \Big\},\end{aligned} \tag{6.28}$$

$$\begin{aligned}\mathcal{O}_{\mathrm{R},3,2} \triangleq \Big\{ & (h_{A\mathrm{R}}, h_{B\mathrm{R}}) : \Delta_1 I(X_{\mathrm{A}}^{(1)}; Y_{\mathrm{R}}^{(1)} | h_{A\mathrm{R}}) \\ & + \Delta_2 I(X_{\mathrm{B}}^{(2)}; Y_{\mathrm{R}}^{(2)} | h_{B\mathrm{R}}) \\ & + \Delta_3 I(X_{\mathrm{A}}^{(3)}, X_{\mathrm{B}}^{(3)}; Y_{\mathrm{R}}^{(3)} | h_{A\mathrm{R}}, h_{B\mathrm{R}}) < 2R \Big\},\end{aligned} \tag{6.29}$$

$$\begin{aligned}\mathcal{O}_{\mathrm{B},4} \triangleq \Big\{ & (h_{AB}, h_{\mathrm{R}B}) : \Delta_1 I(X_{\mathrm{A}}^{(1)}; Y_{\mathrm{B}}^{(1)} | h_{AB}) \\ & + \Delta_4 I(X_{\mathrm{R}}^{(4)}; Y_{\mathrm{B}}^{(4)} | h_{\mathrm{R}B}) < R \Big\}.\end{aligned} \tag{6.30}$$

Again, $\mathcal{O}_{\mathrm{B},4} \subset \mathcal{O}_{\mathrm{B},1}$, where h_{JK} denotes channel from node J to node K ($J, K \in \{\mathrm{A}, \mathrm{B}, \mathrm{R}\}$). The corresponding events for information flow from B to A are similarly defined by interchanging A and B in all the mutual information expressions. Now we claim that

$$\begin{aligned} P[\mathcal{E}_{\mathrm{B},1}] &\doteq P[\mathcal{O}_{\mathrm{B},1}] \doteq \rho^{-d_1(r)}, \\ P[\mathcal{E}_{\mathrm{R},3}] &\doteq P[\mathcal{O}_{\mathrm{R},3}] \doteq \rho^{-d_2(r)}, \\ P[\mathcal{E}_{\mathrm{B},4}] &\doteq P[\mathcal{O}_{\mathrm{B},4}] \doteq \rho^{-d_3(r)}. \end{aligned} \tag{6.31}$$

To prove (6.31), one can follow similar analysis as in [51,72,81]. For each of the events $\mathcal{O}_{J,k}$ and $\mathcal{E}_{J,k}$ defined above, we have

$$P(\mathcal{O}_{J,k}) \leq P(\mathcal{E}_{J,k}) \leq P(\mathcal{O}_{J,k}) + P(\mathcal{E}_{J,k} | \overline{\mathcal{O}}_{J,k}). \tag{6.32}$$

Then, for sufficiently long, independent random Gaussian encoding at every transmitter, joint maximum-likelihood decode-and-forward relaying at the relay node, we show that

$$P[\mathcal{E}_{\mathrm{B},1} | \overline{\mathcal{O}}_{\mathrm{B},1}] \doteq P[\mathcal{O}_{\mathrm{B},1}] \doteq \rho^{-d_1(r)}, \tag{6.33}$$

$$P[\mathcal{E}_{\mathrm{R},3} | \overline{\mathcal{O}}_{\mathrm{R},3}] \doteq P[\mathcal{O}_{\mathrm{R},3}] \doteq \rho^{-d_2(r)}, \tag{6.34}$$

$$P[\mathcal{E}_{\mathrm{B},4} | \overline{\mathcal{O}}_{\mathrm{B},4}] \doteq P[\mathcal{O}_{\mathrm{B},4}] \doteq \rho^{-d_3(r)}. \tag{6.35}$$

In particular, (6.33) follows from DMT for single antenna point-to-point transmission [51], (6.34) follows from the DMT for MAC [72] and (6.35) follows from the DMT for parallel channels [81]. These along with (6.32) gives

$$P[\mathcal{E}_{\mathrm{B}}] \doteq \rho^{-\min\{(d_1(r)+d_2(r)), d_3(r)\}}, \tag{6.36}$$

where, (6.36) follows from (6.24), (6.25) and Varadhan's lemma [71]. Due to symmetry we have $P[\mathcal{E}_{\mathrm{A}}] \doteq P[\mathcal{E}_{\mathrm{B}}]$ and applying Varadhan's lemma to (6.26), we can write

$$P_e \doteq \rho^{-\min\{(d_1(r)+d_2(r)),d_3(r)\}}. \tag{6.37}$$

This completes the proof of Lemma 6. □

Appendix 6D : Proof of Theorem 17

From Lemma 6 we have :

$$\begin{aligned} &d_1(r)+d_2(r) \\ &= \begin{cases} 1+m-\frac{mr}{\Delta_1}-\frac{r}{\Delta_1+\Delta_3} & \text{for } r \le \Delta_1 \\ 1+\frac{\Delta_1}{\Delta_1+\Delta_3}-\frac{2r}{\Delta_1+\Delta_3} & \text{for } \Delta_1 < r \le \Delta_1+\frac{\Delta_3}{2} \end{cases} \end{aligned} \tag{6.38}$$

We are looking at $d_{OPT} = \max_{\Delta_i} \min\{d_1(r)+d_2(r), d_3(r)\}$ under the sum-constraint and the positivity constraints in (6.7). First we prove that any optimal protocol must have $\Delta_3 = 0$, for $m \ge 1$.

This can be proved by contradiction - suppose (Δ_i^*) maximizes the DMT $(d_{OPT}(r))$ with $\Delta_3^* > 0$. Now consider the four phase protocol with new phase durations $\Delta_1 = \Delta_1^* + \frac{\Delta_3^*}{2}$, $\Delta_4 = \Delta_4^*$ and $\Delta_3 = 0$. These new phase durations are in the feasible set (6.7), and substituting these in (6.38) and (6.6) we get

$$\begin{aligned} d_1(\Delta_i, r) + d_2(\Delta_i, r) \ge d_1(\Delta_i^*, r) + d_2(\Delta_i^*, r) \quad \text{and} \\ d_3(\Delta_i, r) \ge d_3(\Delta_i^*, r) \end{aligned}$$

for all values of r. This contradicts the optimality of (Δ_i^*) and show that for $m \ge 1$ any optimal protocol must have $\Delta_3 = 0$.

The optimization problem in (6.7) can now be written as

$$\begin{aligned} d_{OPT}(r) = \max_{\Delta_1,\Delta_4} d_{4-ph}(\Delta_1, \Delta_4, r) \\ \text{subject to} \qquad 2\Delta_1 + \Delta_4 = 1, \qquad \Delta_1, \Delta_4 \ge 0, \end{aligned}$$

where we now have

$$d_1(r) + d_2(r) = (m+1)\left(1 - \frac{r}{\Delta_1}\right) \quad \text{for} \quad 0 \le r \le \Delta_1.$$

Since we are looking at a max-min problem, for any given multiplexing gain r, the optimal DMT is achieved when the two DMT curves $d_1(r)+d_2(r)$ and $d_3(r)$ meet at r. When $\Delta_1 < m\Delta_4$ it is easy to see that the curves meet only at $r = 0$. So we only need to consider the case $\Delta_1 \ge m\Delta_4$. Now we have

$$\mathrm{d}_3(r) = \begin{cases} 1+m-\frac{r}{\Delta_4} & \text{for } r \le \Delta_4 \\ m(1+\frac{\Delta_4 - r}{\Delta_1}) & \text{for } \Delta_4 < r \le \Delta_1+\Delta_4 \end{cases}$$

The expression for $d_3(r)$ suggests that the optimization can be divided in to two exclusive cases : (i) $r \leq \Delta_4$ (ii) $\Delta_4 < r \leq \Delta_1$.

Case $r \leq \Delta_4$

With $d_1(r) + d_2(r) = d_3(r)$, and all other constraints, it is straightforward to check that the optimum solution must satisfy $\Delta_1 = (m+1)\Delta_4$. Substituting this in the sum-constraint, we get

$$\Delta_4 = \frac{1}{(2m+3)}, \quad \Delta_1 = \frac{(m+1)}{(2m+3)},$$

and optimum DMT as $d_{OPT}(r) = 1 + m - (2m+3)r$ for any $r \leq \frac{1}{(2m+3)}$.

Case $\frac{1}{(2m+3)} < r \leq \frac{1}{2}$

Again with $d_1(r) + d_2(r) = d_3(r)$, we get

$$\Delta_1 = \frac{r+m}{1+2m}, \quad \Delta_3 = 0, \qquad \text{and} \quad \Delta_4 = 1 - 2\Delta_1.$$

and optimum DMT as

$$d_{OPT}(r) = \frac{m(1+m)(1-2r)}{r+m} \text{ for } \frac{1}{(2m+3)} < r \leq \frac{1}{2}.$$

This proves Theorem 17. Also, for all values of r and $m = 1$, we have proved that $\Delta_3 = 0$ is optimal, which proves Corollary 17a. □

Appendix 6E : Proof of Theorem 18

Again we are looking at a max-min problem, for any given multiplexing gain r, the optimal DMT is achieved when the two DMT curves $d_1(r)+d_2(r)$ and $d_3(r)$ meet at r. Solving for $d_1(r) + d_2(r) = d_3(r)$ from (6.38) and (6.6) and sum constraint $2\Delta_1 + \Delta_3 + \Delta_4 = 1$, we get

$$\Delta_3 = \frac{(1-m)(2+m-2\sqrt{m})}{4+m^2}, \quad \Delta_4 = \kappa,$$

and

$$\Delta_1 = \frac{1 - \Delta_3 - \Delta_4}{2}$$

and optimum DMT as $d_{OPT}(r) = 1 + m - \frac{r}{\kappa}$ for $r \leq \kappa$. This proves Theorem 18. □

Optimization problem for $\kappa < r \leq \frac{1}{2}$

For $\kappa < r \leq \frac{1}{2}$, again, optimal DMT is computed by solving for $d_1(r) + d_2(r) = d_3(r)$ from (6.38) and (6.6) and sum constraint $2\Delta_1 + \Delta_3 + \Delta_4 = 1$. Now we have,

$$\Delta_1 = \frac{m + r - (3m+1)\Delta_3 + Y}{2(2m+1)}, \tag{6.39}$$

where,

$$Y = \sqrt{(m + r - (3m+1)\Delta_3)^2 + 4m(2m+1)\Delta_3(1-\Delta_3)}.$$

Substituting for Δ_1 and $\Delta_4 = 1 - 2\Delta_1 - \Delta_3$, the optimization problem in (6.7) can now be written as,

$$\min_{0 \leq \Delta_3 \leq 1-r} f(\Delta_3, r),$$

where

$$f(\Delta_3, r) = \frac{m + r - (3m+1)\Delta_3 + Y}{1 - r - \Delta_3}.$$

This optimization problem can be solved using linear programming.

Chapter 7

Conclusions and Future Work

This dissertation deals with the question of establishing fundamental rate-reliability-complexity limits in general MIMO, cooperative and multiple access communications. The work succinctly described the high-SNR fundamental tradeoff among rate, reliability and computational complexity in the form of the achievable DMT for a given *complexity exponent*, which is a fundamental (not heuristic) measure originating from this work. The work then proceeded to constructively meet these high-SNR limits by identifying fast, reliable and efficient MIMO and cooperative encoders, decoders, and protocols that optimally tradeoff the DMT performance with the complexity exponent. This approach constitutes, to the best of our knowledge, the first time that the complexity-performance tradeoff has been succinctly quantified. Specifically we have been able to identify the general complexity-performance tradeoff for a broad family of ML-based and lattice based decoders, as well as provide the first ever lattice decoding solution, and the corresponding activation policy that jointly achieve a vanishing gap to the exact implementation of (regularized) lattice decoding with complexity that is subexponential in the rate. The derived reliability and complexity guarantees hold for most multiple-input multiple-output scenarios, all reasonable fading statistics, all channel dimensions and all-rate lattice codes.

In a nutshell, we provided simple answers and insights to the following pertinent complexity issues :

- Computational reserves required to achieve a vanishing gap to ML or (regularized) lattice decoding performance.
- Impact of computational constraints on the system reliability for ML-based decoding implementations.
- Encoding-decoding policies that can regulate complexity at a limited

performance loss.

- Complexity costs of ripping full rate-reliability benefits of ARQ feedback and also the use of ARQ feedback to reduce decoding complexity.
- Joint reliability-complexity measure that can be used to compare worth of encoding-decoding policies.
- The role of antenna selection in reducing decoding complexity.
- Multiple/cooperative users selection/scheduling in the presence of complexity constraints.

Future Work

The work presented methodology that can be applied to quantify the rate-reliability-complexity performance of novel or existing coding schemes, decoders, as well as cooperative and multiuser protocols. One interesting scenario where this rate-reliability-complexity tardeoff can be applied is capacity scaling analysis for the large networks [82–84]. In large networks decoding complexity will also scale exponentially with cluster size and in the presence of computational constraints, optimal capacity scaling might not be achievable. Another interesting application could be use of rate-reliability-complexity tardeoff for enhancing the physical layer security [85, 86].

Despite the serious efforts to resolve each problem in their most general form, the current work leaves out ample space for exponential reductions in complexity, and improvements in the joint performance-complexity measure, both on the side of decoders, as well as for encoders, protocols or feedback schemes. One such interesting open problem is finding dynamically changing decoding orders that provide guaranteed reductions in the decoding complexity for DMT optimal threaded CDA based codes.

Chapter 8

French Summary

Dans les télécommunications, le débit-fiabilité et la complexité de l'encodage et du décodage (opération à virgule flottante-flops) sont largement reconnus comme représentant des facteurs limitant interdépendants. Pour cette raison, tout tentative de réduire la complexité peut venir au prix d'une dégradation substantielle du taux d'erreurs. Cette thèse traite de l'établissement d'un compromis limite fondamental entre la fiabilité et la complexité dans des systèmes de communications outage-limités à entrées et sorties multiples (MIMO), et ses scénarios point-à-point, utilisateurs multiple, bidirectionnels, et aidés de feedback. Nous explorons un large sous-ensemble de la famille des méthodes d'encodage linéaire Lattice, et nous considérons deux familles principales de décodeurs : les décodeurs à maximum de vraisemblance (ML) et les décodeurs Lattice.

8.1 Systèmes à antennes multiples : performance en terme d'erreur de décodage et de complexité

Avoir une transmission des données à haut débit, une meilleure qualité de service en terme de reduction d'erreurs de décodage et de covarage, et avec un grand nombre d'utilisateurs ; tout en gardant la même puissance de transmission et la même bande passante est une demande qui ne cesse d'augmenter en communications sans fil. Cela exige de considérer des nouveaux technologies comme les systèmes de communications à entrées-sorties multiples (MIMO) qui peuvent améliorer les deux demandes, le débit des données et la reduction d'erreur de décodage pour une puissance de transmission et une bande passante données, mais cela au coût d'une tâche de décodage plus laborieuse à la reception. Spécifiquement, pour le decodeurde maximum de

vraisemblance force brute [1] la complexité de décodage augmente exponentiellement en mot de code bit comme 2^{RT} pour une durée de codage T. En général, ce nombre de bits augmente avec la dimension.

Dans les scénarios MIMO avec des signaux de dimensions élevées, le calcul de maximum de vraisemblance peut être très coûteux. Cela donne toujours la motivation pour utiliser des designs émetteur-récepteur sous-optimaux dont la sous-optimalité peut souvent annuler la fiabilité du système MIMO. La plupart des récepteurs MIMO de nos jours utilisent des régimes de détection MIMO linéaires de complexité réduite (zero-forcing (ZF), minimum mean square error (MMSE), successive interference cancellation (SIC)) dont la sous-optimalité a été prouvée trés élevée (cf. [1]).

Les avancées récentes de vitesse de calcul ont permis l'utilisation de décodeurs MIMO qui emploient des méthodes de détection non-linéaires qui à leur tour se basent sur les algorithmes de recherche bornée, connus sous le nom de décodeurs sphérique. Ces décodeurs sont plus performants que ceux qui utilisent ML mais en introduisant plus d'erreur. (cf. [2–7]).

La diversité de ces algorithmes nous mènent à poser la question concernant la relation entre complexité et fiabilité. La réponse à cette question n'est pas triviale, vu que la complexité algorithmique fluctue aléatoirement avec le canal, le bruit et les réalisations de codeword. Ce comportement aléatoire met en évidence la possibilité d'introduire plus d'erreur pour améliorer la complexité. En général, toute tentative visant l'optimisation de la complexité algorithmique est au détriment de fiabilité de transmission. Dans le cadre de MIMO à délai limité ou de MIMO à panne limitée, la définition et la compréhension de cette relation étroite entre complexité et fiabilité constituent un sujet de recherche très important dont l'apport pratique est très intéressant.

8.2 Caractérisation de complexité en MIMO avec une outage-limitée

Ici la complexité est estimée par le nombre d'opérations calculatoires (en termes d'opérations arithmétiques et de flops) qu'on peut utiliser pendant le décodage d'un codeword. Quand on prend en compte la nature fluctuante de la complexité algorithmique instantanée, ca devient évident que la présence de limitations calculatoires peut entrainer une dégradation de fiabilité. Ceci peut être interprété comme une panne de décodage liée à la violation du délai de temps d'exécution. (Fig. 8.1).

1. le décodeur ML est strictement le seul décodeur optimal en terme de minimization de probabilité d'erreur par mot de code, et il est le plus lent dans le sens qu'il doit visiter tous les mots de code possibles.

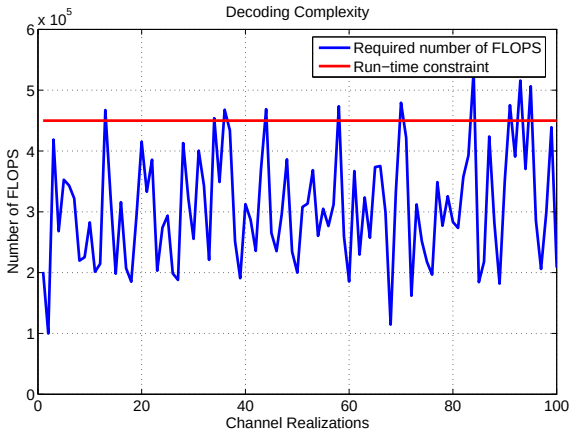


FIGURE 8.1 – Fluctuations de complexité algorithmique instantanées

8.2.1 Exponentiel de complexité

Soit $N_{\max}$ la contrainte de complexité, qui décrit les capacités calculatoires en flops d'un émetteur-récepteur qui a le droit à utiliser le canal T fois. Après $N_{\max}$ flops, l'émetteur-récepteur doit tout simplement s'arrêter, éventuellement avant l'achèvement de sa tache. Pour représenter ce qu'on a discuté ci-dessus, on utilise le compromis de la diversité de multiplexage proposé par Zheng and Tse qui quantifie la relation entre le débit (R) et la probabilité d'erreur P_{err} quand on mesure le gain de multiplexage $r = R/\log \text{SNR}$ et le gain de diversité $d(r) = \lim_{\text{SNR}\to\infty} \log P_{\text{err}}/\log \text{SNR}$ dans un environnement a SNR élevé. Dans le même cadre du tail-analyse de SNR-élevé (le SNR sera noté désormais ρ), on définit l'exposant de complexité comme suit :

$$c(r) := \lim_{\rho\to\infty} \frac{N_{\max}}{\log \rho}.$$

On observe qu'un $c(r) > 0$ implique que la complexité est une fonction exponentielle du débit.

8.2.2 Pertinence et applicabilité de la mesure de complexité et l'exposant de la complexité

En ce qui concerne l'adéquation de la configuration asymptotique, à savoir l'adéquation de l'exposant de complexité et de l'échelle choisie de raffinement, c'est l'échelle qui capture mieux l'intégralité du probléme de complexité. Comme dans le cas de la probabilité d'erreur, qui varie de 1 jusqu'au $K_1 \cdot \rho^{-z}$, mettant en évidence un exposant d'erreur $d(r) \in [0, z]$ (ici K_1 est une fonction sous-polynomiale de ρ), dans le cas de la complexité, on a une variation de 1 flop pour une complexité maximale qui évolue comme

$K_2 2^{RT} = K_2 \cdot \rho^{rT}$ flops (K_2 est essentiellement une constante), ce qui donne un exposant de complexité $c(r) \in [0, rT]$.

Ayant le coût de complexité et la probabilité d'erreur de la même échelle, n'est pas un choix forcé, mais ça correspond à une conséquence triviale du fait que la probabilité d'erreur ainsi que la taille du code sont des fonctions naturellement polynomiales de ρ, et (dans le cas de la complexité) exponentielles de $R = r \log \rho$. En conséquence, l'erreur, mentionnée ci-dessus, et les exposants de la complexité pourraient être combinées pour former une mesure des capabilités taux-fiabilité-complexité des différents émetteurs-récepteurs. Une telle mesure pourrait par exemple prendre la forme d'une mesure commune sur la fiabilité-complexité $\Gamma(r) \triangleq d(r) - \gamma c(r)$, pour un facteur pesant $\gamma \geq 0$. La mesure pourrait être appliquée pour décrire, par exemple, les capacités d'erreur high-SNR d'un codeur-décodeur particulier par unité de puissance et surface sur la puce.

On note que, en observant les compromis entre la fiabilité et la complexité mentionnés ci-dessus, le plus important est de trouver les procedures appropriées de codage-décodage qui peuvent régulariser la complexité en conservant une perte limitée de performance. Une telle réglementation pourrait étendre les deux extrêmes de décodage brute force ML aux récepteurs linéaires inefficaces. En observant ce compromis, un élément necessaire serait le halting ou bien les *procedures de régulation* que nous allons utiliser.

8.2.3 Vanishing écart

Pour raffiner notre analyse de l'exposant, mentionnée ci-dessus (DMT et la complexité de l'exposant), nous notons d'abord que l'analyse de DMT ne parvient pas à capturer potentiellement les écarts infinis (sous-polynomiale à SNR) à la performance optimale. Nous sommes également intéressés aux décodeurs qui permettent d'atteindre un *vanishing écart* à la performance ML. Cette approche de vanishing écart est une condition plus forte que l'optimalité de DMT qui reste insensible aux écarts d'erreur qui pourraient être illimités.

En termes d'écarts d'erreur-performance, on pourrait envisager l'écart d'un décodeur donné $\mathcal{D}_r$ à ML, c-à-d, l'écart entre la performance P_e de $\mathcal{D}_r$ et la performance optimale $\mathrm{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})$ du décodeur brute force ML. Étant donnée une certaine contrainte de calcul $N_{\max} \doteq \rho^c$ pour $\mathcal{D}_r$, cet écart est quantifié dans le régime de high-SNR d'être :

$$g(c) \triangleq \lim_{\rho \to \infty} \frac{P_e}{\mathrm{P}(\hat{\mathbf{s}}_{ML} \neq \mathbf{s})}.$$

Un *vanishing écart* $g(c) = 1$ signifie que, avec $N_{\max} \doteq \rho^c$ flops, $\mathcal{D}_r$ peut asymptotiquement avoir une performance identique à celle du décodeur brute force ML optimal.

De même, lorsqu'on considère tout autre décodeur de base, comme par exemple le décodeur de lattice sphere regulariseé (MMSE-prétraité), nous serions intéressés à l'écart de performance qu'à l'*implémentation exacte* de ce décodeur. Comme précédemment, en présence de $N_{\max} \doteq \rho^{c_L}$ flops pour le décodeur de lattice sphere, nous avons un vanishing gap lorsque :

$$g(c_L) \triangleq \lim_{\rho\to\infty} \frac{P_L}{\mathrm{P}\left(\hat{\mathbf{s}}_L \neq \mathbf{s}\right)} = 1$$

où P_L décrit la probabilité d'erreur du décodeur de lattice sphere prétraité, et $\mathrm{P}\left(\hat{\mathbf{s}}_L \neq \mathbf{s}\right)$ décrit la probabilité d'erreur de la solution exacte du décodeur de lattice sphere MMSE-prétraité.

Un de nos intérêts est de capturer la complexité minimale qui permet un vanishing écart pour plein ML ou plein décodage de lattice (régularisé). Nous allons étudier cela et en particulier développer le travail de [7] dans un cadre plus large.

Remark 4 (Commentaires sur l'approche de vanishing écart : une interpretation heuristique)**.** *Les commentaires dans cette remarque ne sont pas rigoureux, mais n'affectent pas notre analyse, et présentent plutôt des heuristiques sur l'utilité de l'approche de vanishing gap.*

Si on considère, pour un code et un taux fixes, les courbes d'erreur (l'axe x represente le SNR, l'axe y est la probabilité d'erreur) correspondants à deux décodeurs différents. Par exemple, supposant que le premier décodeur soit un décodeur ML optimal (sans interruption). Supposant que le second décodeur soit une implémentation approximative du premier décodeur.

Soit $P(\rho)$ et $P_{ap}(\rho)$ désignant la probabilité réelle *des courbes d'erreur, pour accroître ρ, respectivement, pour les décodeurs optimales et sous-optimales ($P(\rho) \leq P_{ap}(\rho)$ pour tout ρ). En outre, supposant que les deux hypothèses suivantes soient justes.*

Hypothèse 1 *: Il existe une valeur SNR ρ_1 (resp. ρ_2), après laquelle la pente de la courbe d'erreur $\frac{d}{d\rho}P$ (resp. $\frac{d}{d\rho}P_{ap}$) est monotoniquement non décroissante.*

Hypothèse 2 *: Il existe une valeur SNR ρ_0 pour laquelle la pente de la courbe d'erreur $\frac{d}{d\rho}P$ de la première (ML) décodeur atteint sa valeur maximale dans la région monotone, c-à-d, atteint une valeur de $\frac{d}{d\rho}P|_{\rho_0} = \max_{\rho\geq\rho_1} \frac{d}{d\rho}P$.*

Compte tenu des deux hypothèses mentionnées ci-dessus, on observe que la garantie sur vanishing écart implique que les deux courbes d'erreurs coïncident pour tout $\rho \geq \rho_0$, c-à-d,

$$\frac{P(\rho)}{P_{ap}(\rho)} = 1, \ \forall \rho \geq \rho_0.$$

Nous ici nous avisons le lecteur qu'il est concevable facile de violer les deux conditions ci-dessus, par exemple en employant une politique de déco-

dage qui alterne, pour différentes valeurs de SNR, entre deux décodeurs différents, ou en utilisant des codes à distance avec des distributions spécifiques qui peuvent être par exemple entraîner étages d'erreur. Malgré cela, cependant, lorsque la structure émetteur-récepteur demeure inchangée, les deux hypothèses accepter la validation importante par la plupart des observations empiriques, mais n'ont pas été prouvées pour être vrai.

En outre envisager une troisième hypothèse, qui demande que pour toute valeur SNR dans la région monotone $\rho \geq \max\{\rho_1, \rho_2\}$*, la pente d'erreur exacte courbe de la (première) décodeur, n'est pas plus petite que celle de la solution approximative, c-à-d,*

$$\frac{d}{d\rho}P \geq \frac{d}{d\rho}P_{ap}, \quad \forall \rho \geq \max\{\rho_1, \rho_2\}.$$

Ensuite, les trois hypothèses, conjointement avec la garantie sur vanishing écart, conjointement impliquent que les courbes d'erreur du décodeur exacte et approximative, correspondent au (et après) la valeur potentiellement beaucoup plus faible SNR de $\max\{\rho_1, \rho_2\}$*, c-à-d, conjointement impliquent que*

$$\frac{P(\rho)}{P_{ap}(\rho)} = 1, \ \forall \rho \geq \max\{\rho_1, \rho_2\}.$$

Ces valeurs peuvent en effet être modérée, car ils ne se rapportent pas immédiatement à la fidélité des approximations asymptotiques, mais sont plutôt des points de départ des régions monotones. Cette exposition suggère que, dans certaines hypothèses, une garantie sur vanishing écart pourrait se répercuter à finies régions SNR.

8.3 Contribution majeure et plan de thèse

Nous procédons dans le reste de ce résumé à une synthèse de la problématique, en se concentrant sur les résultats qui offrent une meilleure perspicacité plutôt que de se concentrer sur les résultats avec plus large champ d'application, qui sera présenté dans les sections qui suivent. Nous espérons que notre travail peut donner un aperçu sur des questions pertinentes telles que :

- Quel est le prix à payer pour une complexité quasi-optimale lors de l'implémentation des communications coopératives MIMO et multi-utilisateurs ?
- Comment le retour d'information pourra réduire la complexité ?
- Quelles politiques peuvent réduire la complexité pour une perte des performances limitées ?
- Comment les contraintes de complexité peuvent affecter la fiabilité dans des milieux MIMO différents ?
- Quelle est la taille du système MIMO (nombre d'antennes d'émission, nombre de relais ou nombre d'utilisateurs MAC) pour que la puce DSP peut être configuré ?

- Comment les utilisateurs multiples doivent se comporter en présence de contraintes de complexité ?
- Quel est le rôle de la sélection d'antenne dans la réduction de la complexité ?
- Quels sont les protocoles de coopération qui obtiennent les meilleurs résultats en présence des contraintes de calcul ?

Modèle de canal, codeurs et décodeurs - Chapitre 2

Ce chapitre présente le modèle du système décrivant le modèle de canal MIMO sous-jacent ainsi que la description de l'encodage et de décodage régimes considérés comme des études de compromis du taux, de fiabilité et de la complexité. Dans cette thèse nous considérons le cadre général de la *retard-limités outage-limités systèmes de communications MIMO*. Ce réglage capture plusieurs scénarios de communication modernes pertinentes dans les communications sans fil.

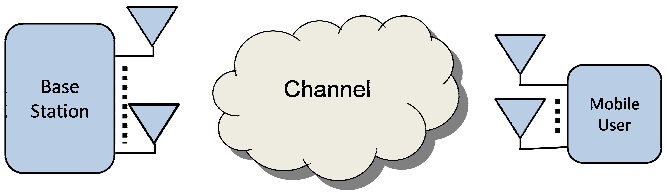


FIGURE 8.2 – Outage-limités systèmes de communications - no CSIT

Il s'agit notamment des scénarios de retard-limités de transmission de données en l'absence d'informations d'état de canal à l'émetteur (pas CSIT) comme le montre la Fig. 8.2, ou, plus important scénarios qui tiennent compte de la communication rapide de la vitesse et à haute au cours de la lien rétroaction de la CSIT dans un cadre multi-utilisateur en l'absence de réciprocité de canal comme indiqué sur la Fig. 8.3, ainsi que le canal d'interférence multi-utilisateur représenté sur la Fig. 8.4 où les solutions de gestion des interférences exigent mondiale CSIT, c-à-d, chaque émetteur est nécessaire de disposer d'informations d'état de canal de tous les liens vers l'avant et même en présence de réciprocité du canal, la CSIT nécessaire à travers les échelles de rétroaction avec le nombre d'utilisateurs.

Nous explorons un large sous-ensemble de la famille des méthodes d'encodage linéaire Lattice, et nous considérons deux familles principales de décodeurs : les décodeurs à maximum de vraisemblance (ML) et les décodeurs Lattice. L'analyse algorithmique est concentrée sur l'implémentation de ces décodeurs ayant comme limitation une recherche bornée, ce qui inclue une large famille de sphère-décodeurs.

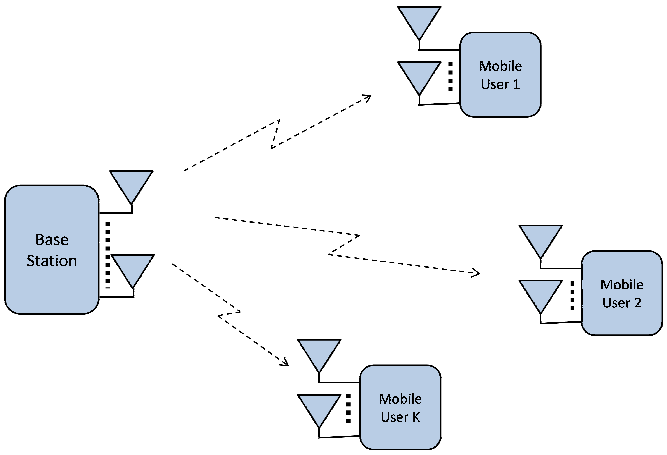


FIGURE 8.3 – Mutliuser MIMO downlink - en l'absence de réciprocité de canal

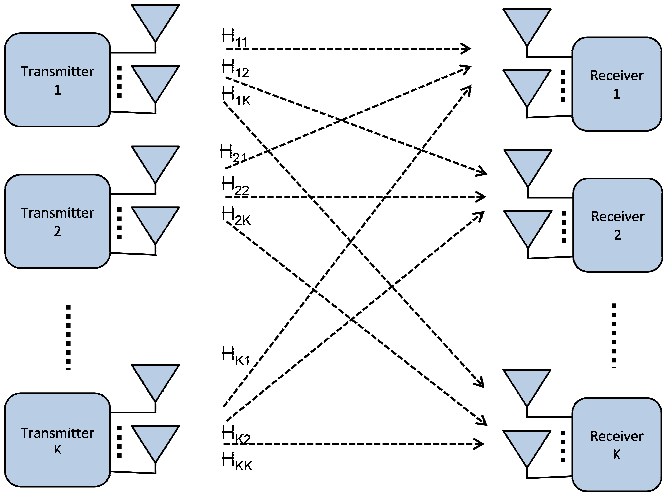


FIGURE 8.4 – Schématique pour canal d'interférence multi-utilisateur

Complexité pour un décodage ML - Chapitre 3

Ce chapitre a un double intérêt. Le premier intérêt consiste à étendre l'analyse de la complexité des précédents travaux ([7]), dans un cadre plus large de modèles en lattice, des politiques de décodage et de statistiques à la décoloration. Cette extension est important dans le sens qu'elle nous permettra d'explorer les paramètres intéressants du modèle MIMO. Le second intérêt est de fournir un taux de fiabilité de complexité, de mesures et des compromis significatives. Le développement mathématique correspondant est pour le cas MIMO général, tandis que les expressions d'une seule-lettre sont dérivées pour le canal quasi-statique MIMO. Tous les résultats

sont valables pour la base ML de décodage, et l'analyse algorithmique considère que la famille ML efficace des algorithmes basés sur décodage par sphère. Les exposants issus de la complexité de ces algorithmes décris le suffisante, et dans de nombreux cas, les ressources de calcul nécessaires requises pour le ML basée sur SD pour atteindre soit une performance DMT spécifique, ou à obtenir un écart de fuite pour (sans interruption) le ML.

Peut-être prématuré à ce stade, nous notons que les résultats suggèrent des ramifications de la complexité pour un choix de design en lattice, les statistiques de décoloration et politiques en sphère de décodage de commande. Les concepts de designs en lattice (et en particulier designs en lattice à taux plein), ainsi que les politiques de commande, sont couramment utilisés, mais ils vont certainement être discuté en détail plus tard.

Nous notons également que le thème commun dans cette recherche est que nous allons souvent obtenir des bornes supérieures universelles de la complexité qui détiennent indépendamment des choix ci-dessus, et par la suite un contraction de résultats pour les paramètres relativement larges.

Réglage MIMO général *Théorème 1 Considerant le cas général pour $m \times n$ ($n \geq m$) MIMO*[2] *de la spécification du canal :*

$$\mathbf{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w},$$

et considerant un ML basée sur le décodage. Dans ce qui suit, $I(\boldsymbol{\mu})$ représentera la fonction du débit de $\boldsymbol{\mu} \triangleq (\mu_1, \cdots, \mu_m)$, $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}^H\mathbf{H})}{\log \rho}$, $j = 1, \cdots, m$, correspondant à des valeurs ordonnées singuliers σ_j du canal. Nous rappelons aussi que $rT \log \rho$ est le nombre total de bits.

Théorème 1 : L'exposant de la complexité de la réalisation d'un gain de diversité $d(r)$ est supérieure délimitée comme pour toutes les familles de statistiques de déperdition[3]*, tous les taux plein (ou en dessous du taux plein) conçoit en lattice, et toutes les politiques statiques ou dynamiques de décodage de commande.*

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{m} \min\left(\frac{rT}{m} - \frac{1}{2}(1-\mu_i), \frac{rT}{m}\right)^+ \tag{8.1a}$$

$$\text{s.t.} \quad I(\boldsymbol{\mu}) \leq d(r), \tag{8.1b}$$

$$\mu_1 \geq \cdots \geq \mu_m \geq 0, \tag{8.1c}$$

2. des résultats moins concluants considère le cas $n \geq m$.

3. Strictement parlant, ceci est valable pour la grande famille des statistiques qui acceptent le principe de grandes déviations [8].

Puis Proposition 1. Établit l'étanchéité conditionnelle de ce qui précède

Proposition 1 : Indépendamment des statistiques canal à évanouissement et du code en lattice, plein taux appliqué, il existe un ordre fixe de décodage pour lesquels la partie supérieure au-dessus universelle liée est serré.

En ce qui concerne l'établissement des ressources suffisantes pour atteindre un écart de fuite à ML, Corollaire 1a applique simplement le théorème ci-dessus après la mise en $d(r) = d_{\text{ML}}$ pour être le DMT optimale ininterrompue de ML décodage du spécifique (potentiellement sous-optimale) du code.

Corollaire 1a : L'exposant la complexité requise par SD pour atteindre un écart de fuite à ML optimale est supérieure délimitée comme

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} \sum_{i=1}^{m} \min\left(\frac{rT}{m} - \frac{1}{2}(1-\mu_i), \frac{rT}{m}\right)^{+} \tag{8.2a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d_{\text{ML}}(r), \tag{8.2b}$$

$$\mu_1 \geq \cdots \geq \mu_m \geq 0, \tag{8.2c}$$

indépendamment de statistiques et les politiques de commande.

Quasi-statique MIMO La transition vers le cas spécifique de la $n_T \times n_R$ ($n_R \geq n_T$) quasi-statique point-à-point de canal MIMO (avec T utilise sur le canal $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$), où maintenant $\boldsymbol{\mu}$ désigne le (asymptotique de) les valeurs singulières de $\mathbf{H}_C$, nous avons Théorème 3 instituant la suivante universelle borne supérieure.

Théorème 3 : L'exposant la complexité SD de parvenir à un gain de diversité $d(r)$ est supérieure délimitée comme

$$\tilde{c}(r) \triangleq \max_{\boldsymbol{\mu}} T\sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^{+} \tag{8.3a}$$

$$s.t. \quad I(\boldsymbol{\mu}) \leq d(r), \tag{8.3b}$$

$$\mu_1 \geq \cdots \geq \mu_{n_T} \geq 0, \tag{8.3c}$$

pour n'importe quel code en lattice à taux plein.

En outre, il convient de noter que, directement à partir de Proposition 1, nous savons que, indépendamment de la déperdition des statistiques et du code à taux plein, il existe un ordre fixe de décodage pour lesquels la partie supérieure au-dessus universelle liée est serré.

D'après Théorème 3 , nous pouvons maintenant établir une limite supérieure universelle liée à la complexité, pour atteindre la performance optimale de DMT $d^*(r)$ de $n_T \times n_R$ ($n_T \leq n_R$) pout le canal MIMO.

Théorème 4 : L'exposant de la complexité SD, pour d'atteindre la DMT optimale $d^*(r)$ est délimitée comme

$$c(r) \leq \bar{c}(r) = \frac{T}{n_T}\left(r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right),$$

où $\bar{c}(r)$ est une fonction linéaire par morceaux qui, pour les valeurs entières de r, prend la forme

$$\bar{c}(r) = \frac{T}{n_T} r(n_T - r), \quad r = 0, 1, \cdots, n_T.$$

Cela vaut pour n'importe quel ensemble de statistiques d'évanouissement, tous les DMT optimales, et de toute politique de l'ordre de décodage.

En raison du fait que la complexité est de plus en plus $d(r)$, ce qui précède peut être utilisé comme une borne supérieure universelle à destination de tous les codes à taux plein. Cette procédure est décrite ci-dessous.

Corollary 4a :L'exposant la complexité SD est majorée comme dans Théorème 4 pour n'importe quel code à taux plein, toutes les statistiques d'évanouissement et toutes les politiques ordre de décodage.

Nous notons que, comme il s'avère, ce n'est le plus serré limite supérieure qui peut contenir tous les codes (à taux plein) et les statistiques fading (cf. Proposition 1). Cette limite supérieure est déjà utile en soi, car il établit que l'exposant de la complexité SD est beaucoup plus faible que le pire des cas SNR exposant rT associé à la brute de décodage ML vigueur, quelles que soient les codes, les canaux ou les politiques de commande. Par exemple, une comparaison de l'exposant vigueur de la complexité brute avec la borne ci-dessus dérivée est représentée sur Fig. 8.5 pour $n_T = T = 2$.

Etanchéité de la borne supérieure universelle En restant dans le $n_T \times n_R$ ($n_R \geq n_T$) quasi-statique canal MIMO, et en se concentrant sur le cas de i.i.d. Statistiques Rayleigh, Théorème 6 établit que la limite supérieure ci-dessus universelle est en effet serré pour presque tous les DMT optimaux en lattice à taux plein, quelles que soient les politiques de commande. Plus précisément, il dit que si la conception est choisi au hasard (chaque élément de la matrice génératrice réseau est choisi de manière iid d'une distribution continue), mais maintenu fixe, puis avec une probabilité dans le choix du code, ce qui précède liée au Théorème 4. est serré

Théorème 6 : Pour un canal Rayleigh MIMO quasi-statique i.i.d., et quelle que soit l'ordre de décodage, fixe ou dynamique, l'exponentiel de complexité de la sphère du décodeur basé sur ML, dans le choix du code DMT Lattice optimal , correspond presque sûrement *à la bande supérieure universelle dans Théorème 4.*

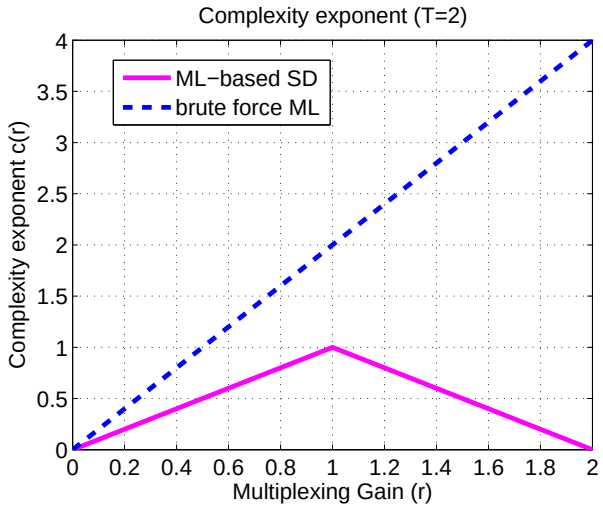


FIGURE 8.5 – Comparaison de l'exposant vigueur de la complexité brute ML avec la partie supérieure universelle à destination de ML à base de SD

Débit-Fiabilité-complexité compromis : DMT réalisable en présence des contraintes de calculs Les dérivations précédentes suggèrent que,sous certaines conditions,l'exponentiel en complexité peut décroître avec une réduction dans le gain désiré de diversité. Une étude prudente montrera plus tard que un tel compromis n'est pas toujours réussi, que des réductions en $d(r)$ ne donnent pas nécessairement une réduction en complexité, et qu'il y a des gammes de $d(r)$ pour lesquelles $c(r)$ reste fixe. On cherche à comprendre le gain de diversité réalisable en présence des contraintes de calcul. Théorème 7 donne une expression générale de ce DMT avec contrainte de complexité pour un cadre MIMO général, et pour tout ensemble de statistiques évanouissantes. En se concentrant sur le canal MIMO Rayleigh évanouissant, i.i.d. et quasi-statique $n_T \times n_R$ ($n_R \geq n_T$), et sur les designs DMT Lattice optimaux (réalisant $d^*(r)$), Corollaire 8a donne une borne infèrieure pour le DMT avec contrainte de complexité. La borne est serrée pour le cas des codes DMT à couches optimales [4].

Corollaire 8a : La contrainte de la complexité DMT de décodage par sphère de n'importe quelle conception DMT utilisant au maximum $N_{\max} \doteq \rho^{c_{\mathcal{D}}(r)}$ *flops, est minoré indépendamment de la politique de décodage de com-*

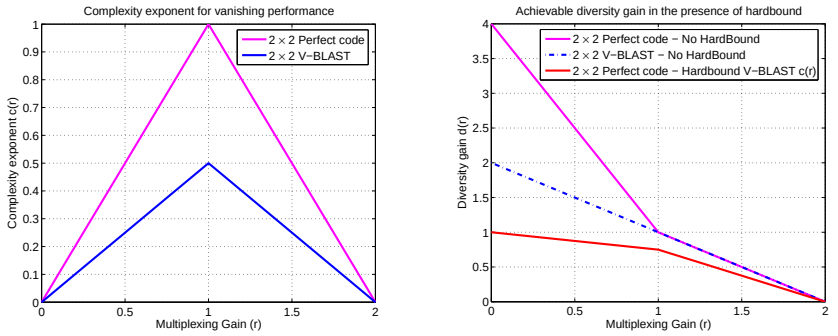
4. Nous reviendrons sur les codes à couches plus en détails dans la suite. On remarque que tous les designs DMT Lattice optimaux connus jusqu'à présent appartiennent à la famille des codes à couches

mande en tant que $d_{\mathcal{D}}(r) = \min\{d^*(r), d_{\mathcal{D}}(r, c_{\mathcal{D}}(r))\}$ *avec*

$$d_{\mathcal{D}}(r, c_{\mathcal{D}}(r)) = \sum_{j=1}^{K} (n_R - n_T + 2j - 1) + (n_R - n_T + 2K + 1)(\frac{c_{\mathcal{D}}(r)}{T} + 1 - \frac{(K+1)r}{n_T}),$$

avec $K = \left\lfloor \frac{n_T c_{\mathcal{D}}(r)}{rT} \right\rfloor$. *De plus, en présence de conceptions optimales du lattice en couches, la DMT décrite en dessus correspond à la complexité de DMT à contraintes donnée par le décodage-ordonnancement naturel en sphère.*

Example 23. *La complexité de DMT à contraintes de décodage en sphère de* 2×2 *du Golden code ([9]) sur canal du Rayleigh i.i.d. est illustrée par la Fig. 8.6. La ligne en haut sur Fig. 8.6 décrit l'exposant de complexité qu'un code idéal de* 2×2 *pourrait exiger pour atteindre le DMT optimal. La ligne en milieu sur Fig. 8.6 (b) décrit l'exposant de complexité qu'un codage moins optimal du VBLAST pourrait exiger pour atteindre son DMT optimal.Cette même ligne sur Fig. 8.6 (a) décrit aussi les limitations de complexité qu'on a attribué un code idéal* 2×2*, qui maintenant à cause de ces limitation ne peut réaliser quŠun DMT réduit (ligne en bas Fig. 8.6(b)).*



(a) Complexité variable Hardbound (b) Gain de diversité possible d'atteindre

FIGURE 8.6 – Compromis de taux de fiabilité-complexité pour des codes 2×2 parfaits.

Complexité du décodage en lattice régularisé- Chapitre 4

Notre intérêt en décodage par lattice découle du fait que le décodage à base de ML exige une grande complexité et que le décodage par lattice atteint parfois des performances comparables a celles du décodage ML. Ce chapitre permet d'aboutir à deux principaux résultats. Le premier résultat prouve

l'équivalence taux-fiabilité-complexité entre SD basé sur ML et SD lattice régularisé [5], et le second est de fournir la première solution (décodeur et procedure de calcul) qui permet d'obtenir un vanishing gap à l'implementation exacte de décodage de lattice regularisé à une complexité sous-exponentielle en termes de taux.

Equivalence de ML-based SD et lattice SD régularisé Ce travail permettra de quantifier la complexité de décodage de lattice regularisé et de son équivalence taux-fiabilité-complexité à SD basé sur ML.

Théorème 10 : Pour la quasi-statique $n_T \times n_R$ ($n_R \geq n_T$) i.i.d. canal MIMO Rayleigh avec fading, l'exposant de complexité du lattice régularisé (sphère) décodant toute conception d'un "full-rate threaded lattice" pour tout DMT régulée, est égal à l'exposant de complexité de SD basé sur ML (Théorème 3), pour tout ordre fixé "layer-preserving" comprenant l'ordre naturel de décodage.

On précise ici que, même si les deux décodeurs (ML et le décodeur basé sur lattice) sont DMT optimales, le résultat ci-dessus intègre plus qu'un simple décodage optimal DMT, dans le sens où il montre que toute procedure de délai sera un compromis entre $d(r)$ et $c(r)$ identiquemet pour les décodages basé sur ML ou basé sur lattice sphere. En d'autres termes, les décodeurs partagent les mêmes capacités $d(r)$ et $c(r)$, indépendamment de la procedure de temporisation.

Par ailleurs, pour le réglage de $n_T \times n_R$ ($n_R \geq n_T$) i.i.d. canal MIMO quasi-statique Rayleigh avec fading, Corollaire 10b démontre que l'équivalence ci-dessus est valable pour pratiquement tous les DMT optimaux à conception full-rate lattice, et ce, pour un réglage très général.

Corollaire 10b : Indépendamment de l'ordre de ayant fixe ou dynamique, l'exposant de complexité de MMSE preprocessed lattice sphere, décodant tout (fixe, mais) code aléatoirement et uniformément choisi (à partir d'un ensemble de DMT optimal à conception full-rate linear lattice) presque sûrement, dans le choix DMT optimal lattice code, correspond à l'exposant de complexité de SD basé sur ML, et indépendamment de la DMT réglementé.

De plus, Corollaire 10d révèle le fait surprenant qu'il n'existe pas de comportement de canal statistique qui permettra la suppression de la région de limites (décodage de lattice) pour provoquer des augmentations illimitées dans la complexité du décodeur, c-à-d, cette limite de complexité existe, même si les statistiques du canal sont tels que les réalisations de canaux qui provoquent le décodeur d'avoir toujours á résoudre le problème le plus dur de la recherche de lattice. Nous donnons ce résultat sous sa forme la plus

5. Un décodeur de lattice "régularisé" est une généralisation du décodeur de lattice MMSE-prétraité.

raffinée qui correspond ai canal MIMO quasi-statique.

Corollaire 10d : L'exposant de la complexité de lattice régularisé SD est majorée par :

$$\overline{c}(r) = \frac{T}{N_t}\left(r(N_t - \lfloor r \rfloor - 1) + (N_t \lfloor r \rfloor - r(N_t - 1))^+\right) \tag{8.4}$$

qui, pour l'entier r simplifie à

$$\overline{c}(r) = \frac{T}{N_t} r(N_t - r), \tag{8.5}$$

pour toutes les statistiques avec fading, toute procedure d'ordre de décodage, toutes les valeurs cibles de DMT et tous les conceptions de full-rate lattice.

Ecart d'erreur pour un décodage régularisé en lattice à une complexité sous-exponentielle Avec la preuve démontrable de la complexité très élevé de la régularisation de décodage en lattice, nous nous tournons vers l'outil de réduction de réseau (LR) et chercher à comprendre ses effets sur la complexité algorithmique. Les résultats principaux seront présentés, et ils sont valables pour un cadre très large de scénarios MIMO, des designs en lattice et des statistiques à la décoloration.

Théorème 12 : Le décodage en sphère lattice du LR-assistée par MMSE-prétraité avec une contrainte de calcul activé à $N_{\max} = \rho^x$, pour $x > 0$, permet un écart de fuite à la solution exacte de régularisation de décodage en lattice.

La complexité sous-exponentielle ci-dessus implique en ce sens que les échelles de complexité plus lent que toute les fonctions exponentielle imaginables. Ce travail constitue la première preuve que la complexité sous-exponentielle n'intervient pas au coût de réduction de la performance de décodage. Ce travail a également été en mesure de, pour la première fois, rigoureusement démontrer et de quantifier le rôle central de la réduction de réseau comme une complexité particulière et de réduire l'ingrédient dans les systèmes MIMO.

Notant aussi que, si la réduction de réseau a en effet permis un comportement quasi-optimale à la complexité très maniable (par rapport au décodage en lattice), c'est le cas qu'il existe des scénarios pour lesquels ces mêmes méthodes ne peuvent pas être facilement appliqués. Ces cas peuvent inclure le scénario omniprésente, où les codes binaires internes sont employés. C'est pour cette raison précise que l'analyse des régimes LR non-assistés reste d'un grand intérêt.

L'impact du feedback sur complexité de ML et les décodeurs lattice - Chapitre 5

Le travail aborde ensuite la question fondamentale de l'établissement des taux de fiabilité de la complexité de la feedback. Ce paramètre est très important car il peut offrir près de comportement ergodique (grande diversité), même à d'importants quels gains de multiplexage. Nous nous concentrons sur deux questions fondamentales.

La première question demande ce qui est des économies de la complexité que la feedback prévoit de la DMT optimale $((d^*(r))$ [6], et la seconde question demande ce qui est des coûts de la complexité de la réalisation des pleins avantages et de la fiabilité du taux de retour. L'analyse et les systèmes de feedback construits nous dira comment utiliser correctement un nombre fini de bits de feedback pour atténuer les effets néfastes des contraintes de calcul, comme ceux observés dans les dérivés de taux de fiabilité de complexité des arbitrages dans les chapitres précédents. L'accent est mis sur les régimes de feedback MIMO-ARQ, bien que nous ne considérons pas également la feedback avec la sélection d'antenne.

Tous les résultats présentés sont valables pour ML-base de décodage ainsi que MMSE-prétraité lattice de décodage.

Commentaires assistée par la complexité de la réalisation de l'DMT optimale Pour le réglage de la $n_{\mathrm{T}} \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$) régulière d'évanouissement MIMO-ARQ canal avec des rondes L de ARQ et avec le canal T utilise par tour (cf. [10]), ce qui suit se concentre sur le cas de $LT = n_{\mathrm{T}}$ et supérieure délimite le minimum feedback assistée par la complexité qui garantit, avec l'aide de l'ARQ, la DMT optimale $d^*(r)$. Avant de procéder avec le résultat, il est intéressant de noter que, comme il s'avère, les politiques qui décident quand à décoder et à ne pas décoder au cours des tours intermédiaires, jouent un rôle crucial dans la réduction de la complexité dérivée. Nous procédons avec le résultat.

Théorème 13 : Soit $c(r)$ est l'exposant de la complexité minimum requis pour atteindre $d^*(r)$, minimisé sur tous les modèles de lattice, tous les régimes ARQ avec $L \leq n_{\mathrm{T}}$ rounds de ARQ et le retard total de $LT = n_{\mathrm{T}}$, et tout stopper et politiques ordre de décodage. puis

$$c(r) \leq \overline{c}_{red}(r) = \frac{1}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right],$$

qui est une fonction linéaire par morceaux que, pour des valeurs de gain de multiplexage entiers, prend la forme

$$\overline{c}_{red}(r) = \frac{1}{n_T} r(n_T - r), \ \textit{for } r = 0, 1, \cdots, n_T.$$

6. Nous signalons ici que $d^*(r)$ désigne la DMT optimale du canal MIMO sans feedback.

A noter la réduction de la complexité, nous rappelons Théorème 5 et Corollaire 5a que dans l'absence de feedback, l'exposant la complexité associée à la même $d^*(r)$ a pris la forme

$$c(r) = r(n_T - r), \tag{8.6}$$

(pour l'entier $r = 0, 1, \cdots, n_T$), alors que nous venons de voir que, par exemple, en présence de $L = n_{\mathrm{T}}$ rounds de ARQ, le DMT même est réalisée avec un très réduite feedback assistée par la complexité d'au plus

$$c(r) \leq \frac{1}{n_T} r(n_T - r).$$

Apres cela, Proposition 3,4 va présenter un schéma très simple de MIMO ARQ qui atteint $d^*(r)$ avec $c(r) \leq \overline{c}_{red}(r)$.

Example 24. *Figure 8.7 traite le cas $n_{\mathrm{T}} = 3 \leq n_{\mathrm{R}}$ dans le cas d'un canal de Rayleigh, et compare la borne supérieur pour complexité qui est décrite ci-dessus en présence de feedback (L-étapes, délai minimum), avec l'exposant de complexité équivalent dans* (8.6) *pour atteindre le même DMT optimal $d^*(r)$ sans ARQ feedback (codes parfaits et naturel, ordre de décodage fixe).*

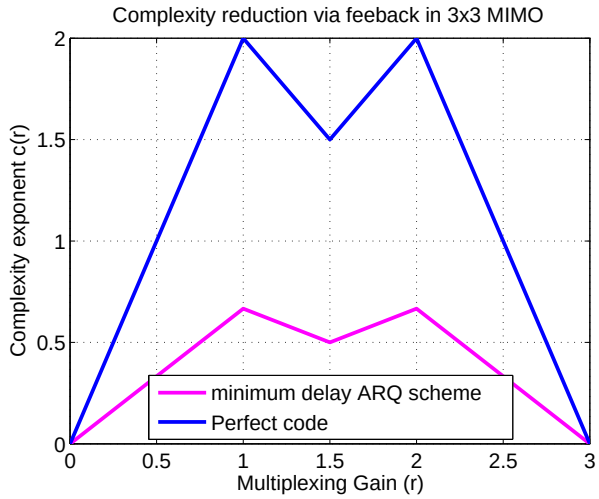
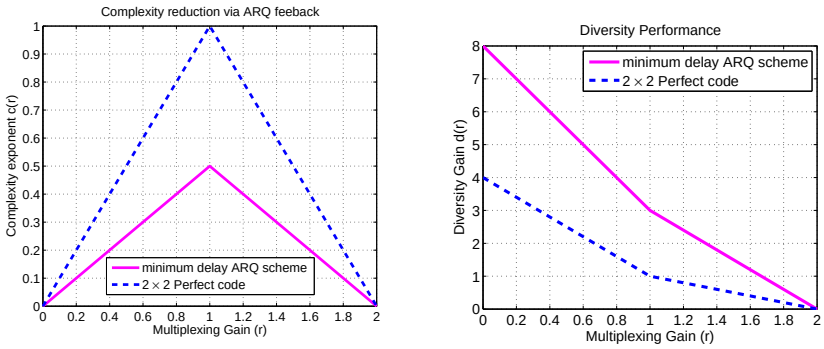


FIGURE 8.7 – Réductions de complexité à l'aide du retour d'informations ARQ.

Concernant la réduction de feedback pour les canaux asymétrique ($n_{\mathrm{R}} < n_{\mathrm{T}}$), Théorème 14 et Proposition 5 décrivent la borne supérieur correspondante pour la complexité et des schémas très simple pour l'ARQ qui atteignent cette borne pour le cas $n_{\mathrm{R}} \leq n_{\mathrm{T}}$, et particulièrement pour le cas $n_{\mathrm{R}} | n_{\mathrm{T}}$ (i.e., n_{T} est un entier multiple de n_{R}).

Example 25 (Correspondant à Theorem 14 et Proposition 5)**.** *Figure 8.8 compare deux régimes : le* 2×2 *canal MIMO (délai minimum, DMT conception réseau optimal), et le montant de* 4×2 *délai minimum MIMO-ARQ canal avec* $L = n_{\mathrm{T}} = 4$*, 3 bits de feedback, et la mise en oeuvre de Proposition 5. Nous voyons une complexité considérablement réduite de la voie de retour aidé (Fig. 8.8(a), ligne inférieure) qui, dans le même temps, réalise une performance DMT beaucoup plus élevé (Fig. 8.8(b), la ligne supérieure) que son homologue.*



(a) Complexity exponent (b) DMT

FIGURE 8.8 – Réduction de complexité pour $n_R|n_T$ canal de Rayleigh i.i.d. avec ARQ feedback

Example 26 (Correspondant à Corollary 14a)**.** *Pour le* $n_{\mathrm{T}} \times 1$*, correspondant feedback assistée par la complexité est montré dans Corollary 14a être sous-exponentielle dans le tarif et le nombre de bits de mots de code. Cette réduction spectaculaire est représentée dans Fig. 8.9 qui la compare à la complexité correspondante (nécessaire pour atteindre le même* $d^*(r)$*) en l'absence de feedback. Ceci est fait pour le cas de* $n_{\mathrm{T}} = 3$*.*

Complexité nécessaire pour atteindre le DMT optimal en présence de feedback $d^*(r/L)$ Pour établir la complexité liée a l'obtention du bénéfice maximum débit-fiabilité de l'ARQ feedback, nous nous concentrons une fois de plus sur le canal ARQ MIMO $n_T \times n_R$ ($n_R \geq n_T$) avec des statistiques i.e. régulières. Comme précédemment, tous les résultats présentés sont valides pour le cas du décodage ML ainsi que pour le cas du décodage à base de lattice MMSE-prétraité. Le résultat suivant est valide pour le case où $n_{\mathrm{R}} \geq n_{\mathrm{T}}$ et le cas ou $L|n_{\mathrm{T}}$ (i.e., n_{T} est un entier multiple de L). Comme il a été montré dans [10], les performances optimales du DMT avec feedback sont données par $d^*(r/L)$.

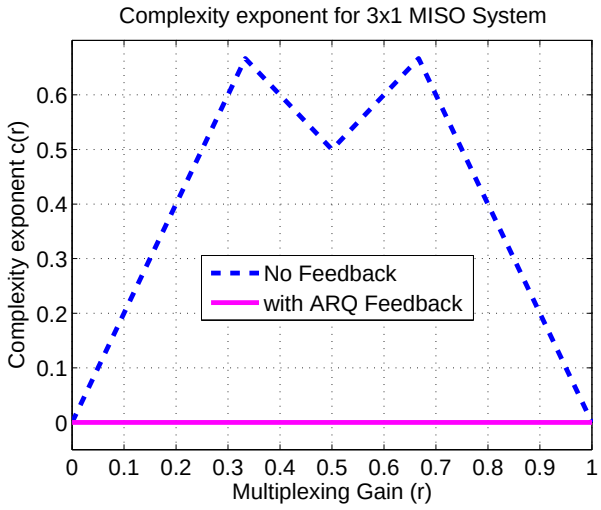


FIGURE 8.9 – Réduction de complexité pour canal de MISO avec un bit feedback.

Théorème 15 : Soit $c(r)$ l'exposant de complexité minimum nécessaire pour atteindre le DMT optimal pour un canal L-étapes MIMO-ARQ $d^*(r/L)$, pour tout $L|n_{\mathrm{T}}$ (i.e., n_{T} est un entier multiple de L), quand la complexité est minimisée sur toutes les conceptions possibles de lattice, tous les critères dŠarrêt, et tous les ordres de décodage. Alors,

$$c(r) \leq \overline{c}_{dmd}(r) \triangleq \frac{1}{L}\left[r\left(n_T - \left\lfloor \frac{r}{L} \right\rfloor - 1\right) + \left(Ln_T \left\lfloor \frac{r}{L} \right\rfloor - r(n_T - 1)\right)^+\right],$$

où $\overline{c}_{dmd}(r)$ est une fonction affine par morceau, qui, pour le cas de r étant un entier multiple de L, prend la forme

$$\overline{c}_{dmd}(r) = \frac{rn_{\mathrm{T}}}{L^2}\left(L - \frac{r}{n_{\mathrm{T}}}\right).$$

Corollary 15a : Le minimum, sur toutes les conceptions de lattice, tous les critères dŠarrêt, et tous les ordres de décodage, qui est nécessaire pour atteindre le optimal DMD $d^*(r/n_{\mathrm{T}})$ de $(L = n_{\mathrm{T}})$-round ARQ, est borné supérieurement par

$$c(r) \leq \overline{c}_{dmd}(r) = \left(1 - \frac{1}{n_T}\right) r.$$

Suivant directement ce résultat, Proposition 6 va présenter des schémas très simple de MIMO ARQ qui atteignent $d^*(r/L)$ avec $c(r) \leq \overline{c}_{dmd}(r)$. DŠautres bornes, schémas ARQ, types de décodages, et conceptions de lattice seront présentés plus tard dans le même chapitre. Ce même chapitre présente aussi plusieurs exemples pour approfondir l'intuition du lecteur.

Example 27 (Correspondant à Theorem 15)**.** *Figure 8.10 montre les limites de la complexité de Theorem 15, décrivant les ressources suffisantes de la complexité dérivées pour atteindre l'optimal DMD* $d^*(r/L)$ *pour les cas de* $6 \times n_{\mathrm{R}}$, $L = 2, 3, 6$.

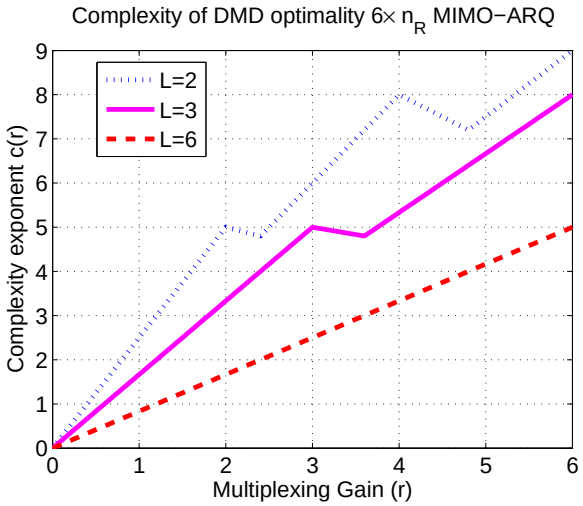
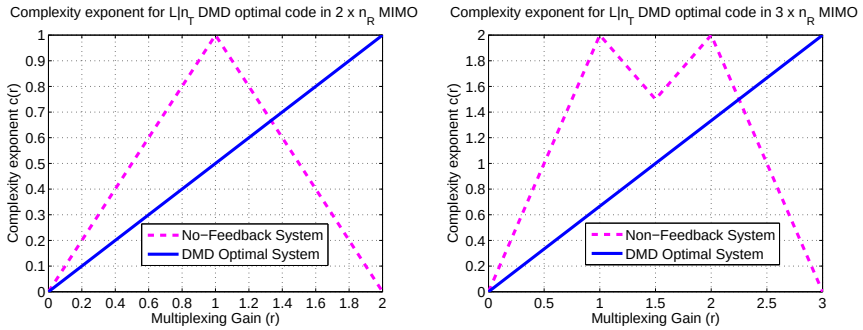


FIGURE 8.10 – Complexité exposant pour $6 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 6$) pour atteindre l'optimal DMD. $L = 2$ (top line), $L = 3$ and $L = 6$ (lowest line).

Example 28 (Correspondant à Corollary 15a)**.** *Figure 8.11 parcelles les bornes de complexité de Corollary 15a décrivant les ressources suffisantes de la complexité dérivées pour atteindre l'optimal DMD* $d^*(r/L)$ *pour les cas des* $2 \times n_{\mathrm{R}}$, $L = n_{\mathrm{T}} = 2$ *et* $3 \times n_{\mathrm{R}}$, $L = n_{\mathrm{T}} = 3$. *Ces résultats sont comparés avec les exposants de la complexité correspondants qui garantissent* $d^*(r)$ *en l'absence de feedback.*

Réductions de complexité due à la sélection d'antennes Nous explorons ici une autre méthode avec laquelle le feedback peut réduire la complexité. Nous nous concentrons particulièrement sur des schémas de sélection d'antennes. Ces schémas utilisent $\log_2 \binom{n_{\mathrm{T}}}{l_T}$ bits de CSIT pour réduire le système MIMO $n_T \times n_R$ en un système de dimensions plus réduites $l_T \times l_R$, et ainsi plus gérable avec en général des besoins en terme de complexité plus réduits. Cette partie de la thèse va analyser la complexité de la sélection d'antennes en se concentrant sur le cas ou les performances, après sélection d'antennes, restent optimal en terme de DMT ($d(r) = d^*_{n_T \times n_R}(r)$). Ce travail est préliminaire, et est basé seulement sur les algorithmes de sélection incrémentielle [11]. Dans cette configuration, pour



(a) For $2 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 2$) MIMO (b) For $3 \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq 3$) MIMO

FIGURE 8.11 – Borne supérieure sur les exposants de la complexité pour atteindre l'optimal DMD.

$$N_r = \arg \min_{N' \in \{1,\cdots,n_T\}} \left[\left(\arg \min_{p \in \{0,\cdots,N'-1\}} \frac{(n_{\mathrm{T}}-p)(n_{\mathrm{R}}-p)}{N'-p} \right) = \lceil r \rceil \right], \quad (8.7)$$

Nous avons le résultat suivant qui est valide pour les canaux MIMO de Rayleigh i.i.d. $n_T \times n_R$ ($n_R \geq n_T$) . Comme précédemment, nous considérons le décodage ML ainsi que le décodage a base de lattice MMSE-prétraités.

Proposition 8 L'exposant de complexité $c(r)$ minimum sur tous les algorithmes de sélection d'antennes, toutes les conceptions de lattice, tous les critères d'arrêt, et tous les ordres de décodage, qui est nécessaire pour atteindre le DMT optimal $d^*_{n_T \times n_R}(r)$ est borné supérieurement par

$$c(r) \leq \overline{c}_{as}(r) = \left(r(N_r - \lfloor r \rfloor - 1) + (N_r \lfloor r \rfloor - r(N_r - 1))^+ \right),$$

qui est, pour le N_r ci-dessus, une fonction affine par morceau, qui, pour valeurs entières de gain de multiplexage r, prend la forme

$$\overline{c}_{as}(r) = r(N_r - r), \ \textit{for } r = 0, 1, \cdots, n_{\mathrm{T}}. \quad (8.8)$$

Le schéma de sélection d'antennes optimal en terme de DMT qui atteint la borne supérieur ci-dessus est aussi présenté dans le même chapitre.

Analyse de complexité pour multi-utilisateurs, cooperative et bidirectionnel-Chapitre 6

Le travail dans ce chapitre est préliminaire et étend l'analyse débit-fiabilité-complexité aux simples examples des canaux à accés multiples, relais

et bidirectionnels, où encore nous identifions les précautions de calculs qui garantissent un DMT optimal , nous adressons également le critère de sélection utilisateur/relais et les protocoles de communications qui améliorent la performance conjointe fiabilité-complexité en présence de contraintes de calculs.

Canaux à accès multiples Dans ce travail, on établit les bornes exigées en terme de complexité pour réaliser le MAC-DMT optimal sur un canal Rayleigh évanouissant i.i.d. à accès multiple (MAC) avec K-utilisateurs, ayant chacun le même gain de muliplexage r, ayant chacun un seul antenne de transmission, et où la destination a $n_R \leq K$ antennes de réception. dans certains cas, les bornes sont montrées être étroites. La borne supérieure de complexité est décrite en dessous, et comme avant, elle est valable pour le décodage basé sur ML et pour le décodage Lattice régularisé. Dans ce qui suit, $K_0 = K$ si K est impair, and $K_0 = K + 1$ si K est pair.

Théorème 16 : Le minimum, sur tous les modèles de Lattice et de halting et l'ordre de décodage, l'exponentiel de complexité $c(r)$ exigée pour atteindre le DMT optimal de MAC, est borné supérieurement comme le suivant

$$c(r) \leq \overline{c}_{mac}(r) = \begin{cases} \overline{c}_v(r) & \text{for } r \leq \frac{n_R}{K+1}, \\ \overline{c}_f(r) & \text{for } \frac{n_R}{K+1} < r \leq \frac{n_R}{K}, \end{cases} \tag{8.9}$$

où

$$\begin{aligned} \overline{c}_v(r) = \max_{\boldsymbol{\mu}} \; & (K-1)r + \sum_{i=1}^{n_R} \left(r - (1-\mu_i)^+\right)^+ \\ s.t. \; & \sum_{i=1}^{n_R} (K - n_R + 2i - 1)\mu_i \leq n_R(1-r), \\ & \mu_1 \geq \cdots \geq \mu_{n_R}, \end{aligned}$$

où

$$\overline{c}_f(r) = (K-1)\, rK_0 + K_0 \left(r(n_R - \lfloor Kr \rfloor - 1) + (\lfloor Kr \rfloor - r(K-1))^+\right),$$

est une fonction linéaire par morceaux telle que, pour $r = 0, \frac{1}{K}, \cdots, \frac{n_R}{K}$, elle prend la forme

$$\overline{c}_f(r) = (K-1)\, rK_0 + K_0 r(n_R - Kr), \; \text{for} \;\; r = 0, \frac{1}{K}, \cdots, \frac{n_R}{K}.$$

Une intuition subtile est déduite du résultat de Proposition 9 qui va décrire la politique de codage/décodage qui atteind le MAC-DMT optimal avec $c(r) \leq \overline{c}_{mac}(r)$.

Example 29. *Pour le cas spécifique où* $n_R = 1$, *la borne dans* (8.9) *prend la forme simple*

$$\overline{c}_{mac}(r) = \begin{cases} (K-1)r & \text{for } r \leq \frac{1}{K+1}, \\ (K-1)K_0 r & \text{for } \frac{1}{K+1} < r \leq \frac{1}{K}. \end{cases}$$

Figure 8.12 décrit les limites supérieures de cas de $K = 4$ *et* $K = 5$ *utilisateurs avec récepteur à antenne unique* ($n_{\mathrm{R}} = 1$).

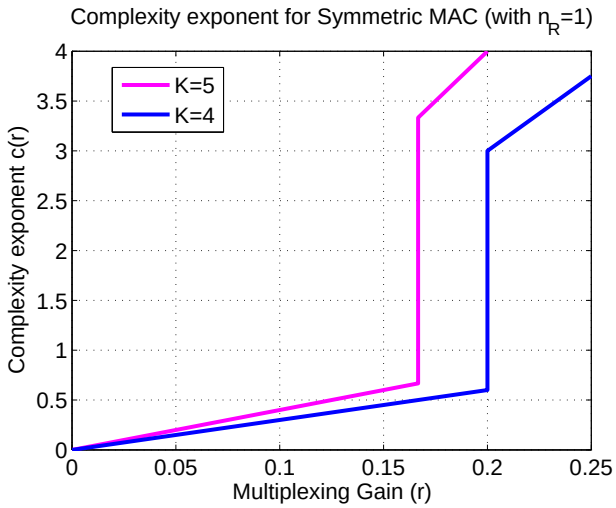


FIGURE 8.12 – Limites supérieures de complexité pour K-utilisateurs MAC avec $n_{\mathrm{R}} = 1$

Canaux à relais coopératives Encore préliminaire, ce travail établit des premiers résultats dans la complexité d'atteindre le DMT optimal d'un réseau coopérative avec une source, $n-1$ relais et une destination, ayant chacun une seule antenne transmetteur/récepteur et communicant sur un canal Rayleigh évanouissant i.i.d.. Cela est fait seulement pour le protocole amplify-and-forward (OAF) orthogonal. la borne supérieure de complexité est décrite en dessous.

Proposition 10 : Le minimum, sur tous les modèles Lattice et halting et politiques d'ordre de décodage, l'exponentiel de complexité $c(r)$ *exigée par le décodage basé sur ML pour atteindre le DMT optimal , est borné supérieurement comme le suivant*

$$\begin{aligned} c(r) \leq \overline{c}_{oaf}(r) = & \frac{2n-1}{n} r(n - \lfloor (2n-1)r \rfloor - 1) \\ & + \left(\lfloor (2n-1)r \rfloor - \frac{2n-1}{n} r(n-1) \right)^+, \end{aligned}$$

qui est une fonction linéaire par morceaux telle que, pour $r = 0, \frac{1}{2n-1}, \cdots, \frac{n}{2n-1}$, *elle prend la forme*

$$\overline{c}_{oaf}(r) = (2n-1)r(1 - \frac{2n-1}{n}r).$$

Proposition 10 décrit aussi les politiques de codage-décodage qui réalisent le DMT optimal de OAF avec $c(r) \leq \overline{c}_{oaf}(r)$. L'exposant la complexité des cas de seul et deux relais en montre dans Fig. 8.13.

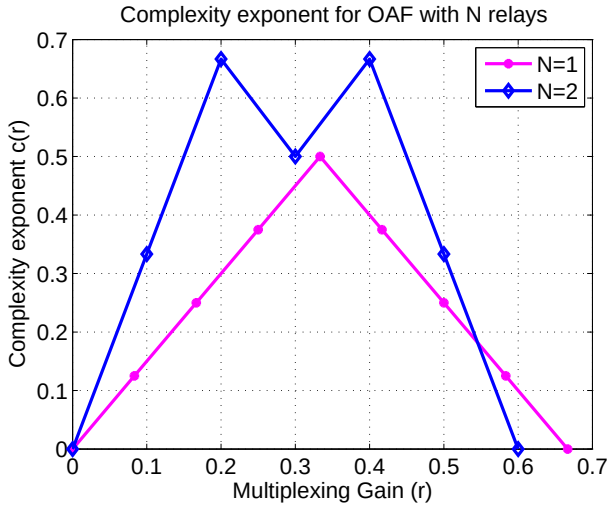


FIGURE 8.13 – Complexité des OAF

Pour les canaux bidirectionnels, Théorème 17 and Théorème 18 décrivent le DMT optimal du relais canal à deux-chemins non-séparables avec asymétrique évanouissant et protocole décodage-et-forward donné. Utilisant cela, Section 6.3.3 présente l'analyse de complexité et la mesure de fiabilité-complexité conjoint pour certains protocoles à deux-chemins.

Conclusions et perspectives - Chapitre 7

Cette thèse traite de la question de la création fondamentaux taux de fiabilité de complexité limites en général MIMO, les communications d'accès de coopération et multiples. Le travail décrit succinctement le compromis de haut SNR fondamentale entre le taux, la fiabilité et la complexité de calcul sous la forme de la DMT réalisable pour une donnée emph exposant la complexité, qui est un droit fondamental (non heuristique) mesure en provenance de ce travail. Le travail a ensuite procédé à répondre à ces limites de manière constructive à haute-SNR en identifiant MIMO rapide, fiable et efficace et des encodeurs de coopération, de décodeurs, et les protocoles de façon optimale compromis la performance DMT avec l'exposant

la complexité. Cette approche constitue, au mieux de notre connaissance, la première fois que le compromis complexité-performances a été quantifiée de façon succincte. Plus précisément, nous avons pu identifier le général de la complexité-performances compromis pour une grande famille de ML à base de lattice et de décodeurs à base, ainsi que de fournir la première solution de lattice de décodage, et la politique d'activation correspondant qui, ensemble, obtenir un écart de fuite à la la mise en oeuvre exacte de (régularisé) lattice de décodage de la complexité qui est sous-exponentielle dans le tarif. La fiabilité et les garanties provenant de la complexité valent pour la plupart Multiple-Input Multiple-output scénarios, toutes les statistiques raisonnables décoloration, toutes les dimensions des canaux et des codes en lattice tous les taux.

Perspectives Le travail présenté méthodologie qui peut être appliquée pour quantifier la performance du taux de fiabilité de complexité du roman ou de systèmes de codage existants, décodeurs, ainsi que des protocoles de coopération et multi-utilisateurs. Un scénario intéressant où cette tardeoff taux de fiabilité de complexité peut être appliquée est mise à l'échelle d'analyse des capacités des grands réseaux [82–84]. Dans les grands réseaux de la complexité de décodage sera également l'échelle de façon exponentielle avec la taille de cluster et en présence de contraintes de calcul, mise à l'échelle capacité optimale peut ne pas être réalisable. Une autre application intéressante pourrait être l'utilisation du taux de fiabilité de complexité tardeoff pour améliorer la sécurité de la couche physique [85, 86].

Malgré les efforts sérieux pour résoudre chaque problème dans leur forme la plus générale, les travaux en cours laisse de côté suffisamment d'espace pour des réductions exponentielles dans la complexité, et des améliorations dans la commune la performance de la complexité mesure, à la fois sur le côté de décodeurs, ainsi que pour les codeurs, protocoles ou les schémas de feedback. Une telle problématique intéressante ouvert est de trouver changeant dynamiquement des commandes de décodage qui permettent des réductions garantis dans la complexité de décodage pour les codes de l'ADC DMT optimales filetés base.

Bibliography

[1] K. Kumar, G. Caire, and A. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4398–4418, Oct. 2009.

[2] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 1639–1642, Jul. 1999.

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

[4] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2401, Oct. 2003.

[5] A. D. Murugan, H. E. Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding : rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933– 953, Mar. 2006.

[6] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[7] J. Jaldén and P. Elia, "Sphere decoding complexity exponent for decoding full rate codes over the quasi-static mimo channel," Feb. 2011, submitted to *IEEE Trans. Inform. Theory*, available on arXiv :1102.1265 [cs.IT].

[8] A. Dembo and O. Zeitouni, "Large deviations techniques and applications," *Springer*, 1998.

[9] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code : A 2×2 full-rate space-time code with non-vanishing determinants," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, Apr. 2005.

[10] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel : Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.

[11] Y. Jiang and M. Varanasi, "Diversity-Multiplexing Tradeoff of MIMO Systems with Antenna Selection," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2007, pp. 2836 – 2840.

[12] H. El Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[13] B. A. Sethuraman, B. Sundar Rajan, and V. Shashidhar, "Full-diversity, high-rate, space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[14] P. Elia, K. R. Kumar, S. A. Pawar, P. Vijay Kumar, and H.-F. Lu, "Explicit space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3869–3884, Sep. 2006.

[15] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3885–3902, Sep. 2006.

[16] Y. Yona and M. Feder, "The Fundamental Limits of Infinite Constellations in MIMO Fading Channels," 2011, available on arXiv :1102.2935 [cs.IT].

[17] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding : Turbo-codes," in *Proc. IEEE Int. Conf. Communications (ICC)*, Paris, France, May 1993.

[18] E. Arikan, "Channel polarization : a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051 – 3073, Jul. 2009.

[19] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, MIT : Cambridge, MA, 1963.

[20] T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge University Press, 2008.

[21] I. Sason and R. L. Urbanke, "Complexity versus performance of capacity-achieving irregular repeat-accumulate codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1247 – 1256, Jun. 2004.

[22] H. D. Pfister, I. Sason, and R. L. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2352 – 2379, Jul. 2005.

[23] I. Sason and G. Wiechman, "Bounds on the number of iterations for turbo-like ensembles over the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2602 – 2617, Jun. 2009.

[24] M. O. Damen, A. Tewfik, and J.-C. Belfiore, "A construction of a space-time code based on number theory," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 753–760, Mar. 2002.

[25] J.-C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003.

[26] T. Kiran and B. Sundar Rajan, "STBC-schemes with non-vanishing determinant for certain number of transmit antennas," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2984–2992, Aug. 2005.

[27] S. Yang, J.-C. Belfiore, and G. Rekaya - Ben Othman, "Perfect space-time block codes for parallel MIMO channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seattle, Washington, USA, Jul. 2006.

[28] P. Elia, B. A. Sethuraman, and P. Vijay Kumar, "Perfect space-time codes for any number of transmit antennas," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3853–3868, Nov. 2007.

[29] F. Oggier, J.-C. Belfiore, and E. Viterbo, "Cyclic division algebras : A tool for space-time coding," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 1, pp. 1–95, 2007.

[30] S. Yang and J.-C. Belfiore, "Optimal space-time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 647 –663, Feb. 2007.

[31] J. Paredes, A. B. Gershman, and M. Gharavi-Alkhansari, "A 2×2 space-time code with non-vanishing determinant and fast maximum likelihood decoding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, USA, Apr. 2007.

[32] H.-F. Lu, "Constructions of multiblock space-time coding schemes that achieve the diversity multiplexing tradeoff," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3790–3796, Aug. 2008.

[33] P. Elia and P. Vijay Kumar, "Space-time codes that are approximately universal for the parallel, multi-block and cooperative DDF channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seoul, Korea, 2009.

[34] P. Elia, K. Vinodh, M. Anand, and P. Vijay Kumar, "D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, Jul. 2009.

[35] E. Biglieri, Y. Hong, and E. Viterbo, "On fast-decodable space-time block codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 524–530, Feb. 2009.

[36] S. A. Pawar, K. Raj Kumar, P. Elia, P. Vijay Kumar, and B. A. Sethuraman, "Space-time codes achieving the DMD tradeoff of the MIMO-ARQ channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, Jul. 2009.

[37] S. Karmakar and B. Sundar Rajan, "Multigroup-decodable STBCs from Clifford algebra," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 223–231, Jan. 2009.

[38] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212 –1215, Mar. 2001.

[39] W. Zhao and G. B. Giannakis, "Sphere decoding algorithm with improved radius search," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1104–1109, Jul. 2005.

[40] ——, "Reduced complexity closest point decoding algorithms for random lattices," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 101–111, Jan. 2006.

[41] A. Y. C. Peng, I. M. Kim, and S. Yousefi, "Low-Complexity Sphere Decoding Algorithm for Quasi-Orthogonal SpaceTime Block Codes," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 1104–1109, Mar. 2006.

[42] R. Gowaikar and B. Hassibi, "Statistical pruning for near-maximum likelihood decoding," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2661–2675, Jun. 2007.

[43] B. Shim and I. Kang, "Sphere decoding with a probabilistic tree pruning," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4867–4878, Oct. 2008.

[44] C. C. Studer and H. Bölcskei, "Soft-input soft-output sphere decoding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2008, pp. 2007 –2011.

[45] A. Ghasemmehdi and E. Agrell, "Optimal projection method in sphere decoding," 2009, available on arXiv :0906.0249 [cs.IT].

[46] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.

[47] L. P. Natarajan, K. P. Srinath, and B. S. Rajan, "On the Sphere Decoding Complexity of STBCs for Asymmetric MIMO Systems," 2011, available on arXiv :1104.0640 [cs.IT].

[48] A. Ghasemmehdi and E. Agrell, "Faster Recursions in Sphere Decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3530 –3536, Jun. 2011.

[49] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806– 2818, Aug. 2005.

[50] H. El Gamal and A. R. Hammons Jr., "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2321–2334, Jun. 2001.

[51] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing : A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[52] J. Radon, *Lineare Scharen Orthogonaler Matrizen.* Abh. Math. Sem. Hamburg I, 1923.

[53] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[54] Y. Jiang, W. W. Hager, and J. Li, "The Generalized Triangular Decomposition," *Mathematics of Computation*, vol. 77, no. 262, pp. 1037 –1056, Apr. 2008.

[55] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, June 2004.

[56] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, Jun. 2004.

[57] J. Jaldén and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4765–4780, Oct. 2010.

[58] J. Pan and W.-K. Ma, "A lagrangian dual relaxation approach to ML MIMO detection : reinterpreting regularized lattice decoding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.

[59] D. Seethaler, J. Jalden, C. Studer, and H. Bölcskei, "Tail behavior of sphere-decoding complexity in random lattices," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2009, pp. 729 – 733.

[60] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Conf. Communications (GLOBECOM)*, Taipei, Taiwan, Nov. 2002.

[61] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003.

[62] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction : A survey with with applications to wireless communication," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 70 – 91, Jun. 2011.

[63] C. Ling, "On the proximity factors of lattice reduction-aided decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2795 – 2808, 2011.

[64] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Matematische Annalen*, vol. 261, no. 4, pp. 1432–1807, Dec. 1982.

[65] J. Jaldén, D. Seethaler, and G. Matz, "Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems," in *Proc.*

IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, Nevada, USA, Apr. 2008.

[66] L. Babai, "On lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.

[67] S. Loyka and G. Levin, "Finite-snr diversity-multiplexing tradeoff via asymptotic analysis of large mimo systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4781 – 4792, Oct. 2010.

[68] D. Wübben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice reduction," in *Proc. IEEE Int. Conf. Communications (ICC)*, Paris, France, Jun. 2004.

[69] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec. 2007.

[70] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1547–1557, Apr. 2007.

[71] S. R. S. Varadhan, "Asymptotic probabilities and differential equations," *Communications on Pure and Applied Mathematics*, vol. 19, no. 3, pp. 261–286, 1966.

[72] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.

[73] H. feng (Francis) Lu, C. J. Hollanti, R. I. Vehkalahti, and J. Lahtonen, "DMT optimal codes constructions for multiple-access MIMO channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3594 –3617, Jun. 2011.

[74] P. Popovski and T. Koike-Akino, *New Directions in Wireless Communications Research*. Springer US, 2009.

[75] S. J. Kim, P. Mitran, and V. Tarokh, "Performance bounds for bidirectional coded cooperation protocols," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5235–5241, Nov. 2008.

[76] M. Nakagami, "The m-distribution- A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Oxford : Pergamon Press, 1960, pp. 3–36.

[77] T. Kim and H. Poor, "On the dmt of bidirectional relaying with limited feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2009, pp. 334–338.

[78] D. Gündüz, A. J. Goldsmith, and H. Vincent Poor, "MIMO Two-way Relay Channel : Diversity-Multiplexing Tradeoff Analysis," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Oct. 2008, pp. 1474 – 1478.

[79] T. T. Kim and H. V. Poor, "Diversity Multiplexing Tradeoff in Adaptive Two-way Relaying," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4235 –4254, july 2011.

[80] R. Vaze and R. W. Heath Jr, "On the Capacity and Diversity-Multiplexing Tradeoff of the Two-Way Relay Channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4219 –4234, Jul. 2011.

[81] P. Mitran, "The diversity-multiplexing tradeoff for independent parallel mimo channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2008, pp. 2366–2370.

[82] J. Ghaderi, L.-L. Xie, and X. Shen, "Hierarchical Cooperation in Ad Hoc Networks : Optimal Clustering and Achievable Throughput," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3425 –3436, Aug. 2009.

[83] A. Ozgur, R. Johari, D. Tse, and O. Leveque, "Information-Theoretic Operating Regimes of Large Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 427 –437, Jan. 2010.

[84] H. Bölcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433 –1444, Jun. 2006.

[85] Y. Shiu, S. Chang, H. Wu, S. Huang, and H. Chen, "Physical layer security in wireless networks : a tutorial," *IEEE Wireless Communications Magazine*, vol. 18, no. 2, pp. 66 –74, Apr. 2011.

[86] L. Dong, Z. Han, A. P. Petropulu, and H. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1875 –1888, Mar. 2010.