On the Analysis and Design of Wireless Communication Systems Using Tools from Statistical Physics

Rodrigo Vicente de Miguel de Juan

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

PHILOSOPHIAE DOCTOR

Department of Electronics and Telecommunications
Norwegian University of Science and Technology
2008
Abstract

The present dissertation is concerned with the analysis and design of wireless communication systems using mathematical tools imported from the field of statistical mechanics. In particular, methods traditionally used to study macroscopic properties of disordered physical systems are used to analyze certain wireless channels as the number of transmitting and receiving elements gets large. The thesis consists of three parts: an introduction, a collection of three papers, and three independent appendices.

The introduction starts by presenting the basic theory of statistical mechanics as it pertains to the analysis of disordered systems, where randomness appears not only in the microscopic state of the system, but also in the interactions between its constituent parts; the concept of self-averaging is presented and emphasized. Then a connection is drawn between the randomness in the wireless multiple-input/multiple-output (MIMO) channel and that exhibited by the interactions in systems with quenched disorder. Afterwards, specific mathematical methods dealing with quenched disorder are presented and justified, in particular the so called replica method and also the annealed approximation, which constitute the basic machinery in the analysis presented in the remainder of the thesis. The introduction concludes with a summary of the papers presented in the second part, some suggestions for future research, and a brief overview of the appendices in the third part.

The second part of the thesis consists of a collection of three papers. In these papers different schemes are explored which allow for low complexity detection at the receiver end of multiuser vector channels. Specifically we propose and explore different non-linear vector precoding schemes for both single-user and multiuser MIMO channels in high-dimensions; convex spaces are suggested to relax original constellations and the replica method is used to characterize the energy penalty in the transmission process. We shall also explore a single-user vector channel where the input is preselected to conform with a strict complexity constraint at the receiver; in this case the annealed approximation is used to obtain the information
capacity of the channel.

The three independent appendices at the end of the thesis show some results that, although not part of any published work, are of theoretical interest and potentially useful in future research within vector precoding.
Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU). I have carried out this work under the supervision of Prof. Ralf R. Müller between April of 2005 and August of 2008.

Part of this work was done in cooperation with Dr. Ori Shental and Prof. Ido Kanter at the Physics Department in Bar Ilan University (Israel), whom I visited as a scholar in July of 2006.

Funding and support was provided by NTNU’s Department of Electronics and Telecommunications, and by the Research Council of Norway through the project Statistical Physics of Advanced Multiuser Communications (SPAM). Some traveling to scientific events was funded by the EU FP6 Network of Excellence in Wireless Communications (NEWCOM) and by TOTAL E&P Norway.
Acknowledgments

First of all, I am greatly indebted to my supervisor, Prof. Ralf Müller, whose wise guidance and incredible insight have made this thesis possible.

I have also had the pleasure of collaborating with Prof. Ido Kanter (Bar Ilan University) and Dr. Ori Shental (University of Tel Aviv). Working with them has been a personal and professional privilege. In particular, I have great memories of time spent together with my good friend Ori, with whom I enjoyed many interesting and illuminating discussions, both technical and mundane.

I would also like to thank Dr. Benjamin M. Zaidel (NTNU) and Prof. Aris L. Moustakas (National and Kapodistrian University of Athens) with whom I have cooperated in the latest stages of my research. I would especially like to thank my friend Benny for coping with me and, with infinite patience, helping me understand information theoretic concepts.

I would like to thank Kirsten Ekseth, our friend and guardian angel in the Signal Processing group, for being kind and helpful, and for making practicalities more practical. The Signal Processing group has been a most stimulating social and scientific environment; every one of my colleagues in Signal Processing has made it possible. Great memories of friendship and comradery will stay with me forever. Thank you.

Throughout these years many friends and colleagues had a positive impact on my life and made my days in Trondheim more pleasant. For our great scientific and social interaction, I would like to thank Anders Gjendemsjø, Benny Zaidel, Changmiang Wang, Geir Øien, José Marçal, Linus Boman, Marco Siniscalchi, Miguel Rubí, Mikko Vehkaperä, Pierluigi Salvo Rossi, Prasanna Chaporkar, Pål Anders Floor, Ralf Müller and Sebastién de la Kethulle. I especially thank Rikke Soligard, who stod by me during difficult times.

Finally, I must say that this thesis wouldn’t have been remotely possible had it not been for my parents. My mother, Mariví, and my father, Sátur, my best friend. My whole family made it all possible for me through their support and inspiration. I could not possibly express enough gratitude to them.

Trondheim, October 2008
Rodrigo Vicente de Miguel de Juan
Contents

Abstract iii
Preface v
Acknowledgments vi
Contents vii

I Introduction 1
1 Statistical Mechanics 3
   Microstates and Macrostates 4
   Temperature 5
   Boltzmann Statistics 7
   Quenched Disorder 9
2 Quenched Disorder and Vector Precoding in MIMO Channels 10
3 The Replica Method 12
   Degeneracy and Ergodicity Breaking 13
   Replica Symmetry? 15
4 The Annealed Approximation 17
5 Overview of Included Papers 19
   Paper A 19
   Paper B 20
   Paper C 20
6 Papers Not Included in this Thesis 22
7 Suggestions for Further Research 22
8 Overview of Appendices 24

References 25
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A Radially Invariant Replica Space</td>
<td>139</td>
</tr>
<tr>
<td>4</td>
<td>Replica Symmetry</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>Radially Invariant One-Dimensional Binary Alphabet</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Approaching the minimum eigenvalue of a Marchenko-Pastur matrix</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Approaching the minimum eigenvalue of a complex inverse Marchenko-Pastur matrix</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>Approaching the minimum eigenvalue of the inverse of the real part of a Marchenko-Pastur matrix</td>
<td>144</td>
</tr>
<tr>
<td>6</td>
<td>Symmetric, Radially Invariant, $m$-PSK Alphabets ($m \geq 2$)</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>Approaching the minimum eigenvalue of a complex inverse Marchenko-Pastur matrix</td>
<td>146</td>
</tr>
<tr>
<td>7</td>
<td>Some Impressions</td>
<td>147</td>
</tr>
</tbody>
</table>

**References** 149
Part I

Introduction
Introduction

The advent of statistical mechanics occurred in the context studying the properties of large physical systems exhibiting many degrees of freedom. Although the use of statistical methods to study physical systems was introduced by Maxwell in 1859 (in the context of kinetic theory of gases [1, 2]) and Boltzmann in 1872 (his famous H-theorem [3]), it wasn’t until 1884 that the term statistical mechanics was coined by Gibbs, who published the first treatise on the subject in 1902 [4]. Statistical mechanics has since evolved and its powerful methods are used not only in virtually every field of the physical sciences, but they have transcended the physics community and are nowadays applied in every discipline where the issue of complexity appears. To name a few examples I’ll mention biology (see e.g. [5, 6]), ecology (see e.g. [7, 8]), sociology (see e.g. [9, 10]), economics (see e.g. [11, 12]), wireless communications (see e.g. [13, 14]) and coding theory (see e.g. [15, 16]). In the following sections I shall try to provide the reader with a modest introduction to the tools of statistical mechanics used in Part II of this thesis, where mathematical methods are imported from the study of disordered magnetic systems and used in the analysis and design of wireless communications systems.

‘The land of statistical physics is broad, with many dales, hills, valleys and peaks to explore that are of relevance to the real world and to our ways of thinking about it.’


1 Statistical Mechanics

There is a multitude of texts which offer quite complete, yet friendly treatments of modern statistical mechanics (see e.g. [17], [18]). Unfortunately they assume the reader is familiar with physics and they go far beyond what is needed to understand this thesis. In this section I attempt to present
a clear and concise introduction to the basic concepts of statistical mechanics upon which this thesis is built.

Consider an impressionist painting viewed from across the room. The individual brush strokes cannot be distinguished at that distance; however, the image is well defined. As the eyes average the contributions of the individual brush strokes, it is even conceivable that different arrangements of such strokes may yield the same image.

Just like with impressionist paintings, when analyzing macroscopic (i.e. large) systems we can deal with the state of the system at two levels: the microscopic level and the macroscopic level. The microscopic state is the detailed configuration of each internal degree of freedom. However, if we observe the system at the macroscopic level, we can, at best, distinguish some of its gross characteristics.

**Microstates and Macrostates**

Suppose that you flip three distinguishable coins. Each coin might land head up (H) or tail up (T). There are eight possible outcomes: \{HHH\}, \{HHT\}, \{HTH\}, \{THH\}, \{HTT\}, \{THT\}, \{TTH\}, and \{TTT\}. Each of the eight outcomes is called a microstate. To specify a microstate we need to specify the state of each particle in the system, in this case the state of each coin. If we only specify the state of the system more generally by saying how many heads or tails there are, we are then specifying the macrostate of the system. Therefore knowledge of the microstate implies knowledge of the macrostate, but the converse is not true. The number of microstates corresponding to a given macrostate is called the multiplicity of that macrostate. For example, the multiplicity of the macrostate which has two tails and one head is 3.

Let us denote the multiplicity by \(\Omega\). Then for the example at hand \(\Omega(0\text{ heads}) = 1\), \(\Omega(1\text{ heads}) = 3\), \(\Omega(2\text{ heads}) = 3\), \(\Omega(3\text{ heads}) = 1\). The sum of all the multiplicities equals the total number of microstates, that is \(\Omega(\text{all}) = 8\). Then, the probability of \(n\) heads is directly proportional to its multiplicity as follows

\[
p(n) = \frac{\Omega(n)}{\Omega(\text{all})}.
\]

In the example above we have observed how, although all microstates are equally probable, some macrostates are more probable than others. Suppose now that we have not just three coins, but \(N\) coins. Then the multiplicity of the \(n\)-head macrostate is given by the combinatorial formula

\[
\Omega(N,n) = \frac{N!}{n!(N-n)!}.
\]
1. **Statistical Mechanics**

**FIGURE 0.1:** Multiplicity graphs for coin tossing experiments with 16 coins (left) and 256 coins (right). As the size of the system increases the peak becomes very narrow compared to the horizontal scale.

and it’s Boltzmann entropy is defined as\(^1\)

\[
\sigma(N, n) \equiv \ln \Omega(N, n). \tag{3}
\]

Fig. 0.1 shows how, as \(N\) increases, certain macrostates become much more probable than all others. When \(N\) grows towards infinity, the binomial distribution becomes a delta function centered at \(n = 0.5N\) and only the macrostate represented by the greatest number of microstates (the one with the most entropy) will be observed. This large system behavior is known as **self-averaging**. Although infinite systems do not exist, systems that are large enough can be approximated by self-averaging. These systems are known as **thermodynamic systems**\(^2\).

**Temperature**

Now that we have some notion of what microstates and macrostates are, we can start talking about statistical mechanics. Suppose there are two large systems somewhat similar to the type of system I described in the previous section. These two systems, of possibly different size, are initially isolated from one another. But now we change our language slightly: instead of characterizing these two systems, called \(S_a\) and \(S_b\), by the number of heads and tails, we will say instead that they are characterized by another intrinsic property. We call this characteristic property of the systems, the **energy**,\(^1\)

\(^1\)Note that Shannon’s entropy reduces to the Boltzmann entropy when equiprobability is assumed.

\(^2\)This name comes from physics, where thermodynamics is nothing but a self-averaging approximation of quantum behavior which turns out very accurate for sufficiently large systems.
and we say that the systems have energies $E_a$ and $E_b$ with multiplicities $\Omega_a(E_a)$ and $\Omega_b(E_b)$, respectively. Don’t be intimidated by the term: you can think of the energy as being any constraint you want, for example, the number of heads up after a coin toss. So far so good; nothing has changed.

Suppose we allow these two systems to exchange some of this thing called energy with each other. The two subsystems together now form a larger isolated system $S = S_a + S_b$ where the total energy $E = E_a + E_b$ can now be redistributed. The question is: how will the energy be distributed? We can start answering this question by noting that the overall multiplicity function $\Omega(E)$ is the product of both multiplicity functions:

\[
\Omega(E) = \Omega_a(E_a) \times \Omega_b(E_b),
\]

and as a result the total entropy (3) of the system is

\[
\sigma(E) = \ln \Omega(E) = \ln \Omega_a(E_a) + \ln \Omega_b(E_b) = \sigma(E_a) + \sigma(E_b).
\]

Since the total system is completely isolated, its total energy $E$ is fixed. When the energy of an isolated system is fixed we can assume that all the microstates of the system are equiprobable as there is no a priori reason to think otherwise. However, as we can see from (4), the total number of accessible states $\Omega$ will depend on the division of the total energy $E$ between $E_a$ and $E_b$. For a certain value of $E_a$ (and therefore of $E_b = E - E_a$), $\Omega$ will be maximum, i.e. the entropy (5) will be maximum. In other words: we will be more likely, on average, to find the system in states with that value of $E_a$ than with any other value of $E_a$.

In other words, the flow of energy between systems $S_a$ and $S_b$ won’t be governed by equalization of either energy or entropy themselves. It will be governed by the equalization of the rate of change of entropy with energy, $\sigma_a / \sigma_b$. To see why, suppose that the rate of change is not equalized and at a given time it obeys

\[
\frac{\partial \sigma_b}{\partial E_b} > \frac{\partial \sigma_a}{\partial E_a} > 0.
\]

Then removing a small amount of energy from $S_a$ will decrease its entropy, but not as much as the entropy in $S_b$ will be increased by the arrival of that energy. Thus the net entropy will be increased with the transfer of energy from $S_a$ to $S_b$. We can only be sure that the net entropy is maximum when both rates of change are equal:

\[
\frac{\partial \sigma_a}{\partial E_a} = \frac{\partial \sigma_b}{\partial E_b} \equiv \frac{1}{T}.
\]

\[\text{3This is commonly referred to as the fundamental postulate of statistical mechanics.}\]
The constant $T$ is called temperature, and we might roughly define it as that quantity which is the same in two systems when they are in equilibrium.

Now we have relation between energy and entropy: in any given system at any given temperature

$$dE - Td\sigma = 0. \tag{8}$$

This means that there is another quantity,

$$F \equiv E - T\sigma, \tag{9}$$

which, just like the total entropy, reaches an extremum at equilibrium. And since the entropy $\sigma$ reaches a maximum at equilibrium, then this quantity must reach a minimum. We shall call this quantity $F$ the free energy.

**Boltzmann Statistics**

A basic assumption in equilibrium statistical mechanics is that if we impose some constraints on a system (we force it to be in some particular macrostate) and wait long enough, the system will eventually flow through the whole ensemble of microstates compatible with these constraints. This leads directly to the ergodic hypothesis: it is believed that after an infinitely long observation time the system visits its different microstates many times, and as a result averaging the value of any macroscopic observable over time is equivalent to taking the so called ensemble average over the different microstates. But in order to take the ensemble average we must first know what the general form of the probability distribution is. In this section I shall present a simple formula for the probability of finding a system with temperature $T$ in a particular microstate.

Let us go back to the two systems, $S_a$ and $S_b$, from the previous section. Let us assume that both systems are in equilibrium and concern ourselves with the probability distribution for the microstates in $S_a$. The probability of finding $S_a$ in a particular microstate depends on how many other microstates there are. Then, the problem can be simplified by looking only at the ratio of probabilities for two particular microstates of interest. We might label these two microstates of $S_a$ as $a_1$ and $a_2$, their energies $E_a(a_1)$ and $E_a(a_2)$ and their probabilities $P_a(a_1)$ and $P_a(a_2)$.

We know that if $S_a$ is in microstate $a_1$ with energy $E_a(a_1)$, then $S_b$ will be in a macrostate dictated by how much energy is left. This macrostate of $S_b$ will have $\Omega_b(a_1)$ accessible equiprobable microstates. Now, as the microstates of the total system are all equiprobable, we can infer that the probability of finding $S_a$ in a particular microstate is directly proportional
to the number of microstates which are available to $S_b$:

$$\frac{P_a(a_1)}{P_a(a_2)} = \frac{\Omega_b(a_1)}{\Omega_b(a_2)}. \quad (10)$$

We might now rewrite each multiplicity $\Omega$ in terms of the Boltzmann entropy $S$:

$$\frac{P_a(a_2)}{P_a(a_1)} = e^{\sigma_b(a_2) - \sigma_b(a_1)}. \quad (11)$$

The exponent in the right hand side contains the change in entropy of $S_b$ when $S_a$ undergoes a transition from state $a_1$ to state $a_2$. Because any energy lost by one of the subsystems is gained by the other, the equilibrium condition (7) implies

$$\sigma_b(a_2) - \sigma_b(a_1) = \frac{1}{T} [E_b(a_2) - E_b(a_1)] = -\frac{1}{T} [E_a(a_2) - E_a(a_1)]. \quad (12)$$

Equation (12) might be inserted into (11) yielding

$$\frac{e^{-E_a(a_2)/T}}{P_a(a_2)} = \frac{e^{-E_a(a_1)/T}}{P_a(a_1)}. \quad (13)$$

The exponential factor $e^{E_a(a_i)/T}$ is known as the Boltzmann factor. As the left(right) side of (13) is independent of $a_1(a_2)$, so must be the right(left) side. Then both sides must be equal to some constant, $Z_a$. This constant, called the partition function, converts the Boltzmann factor into a probability. Then, the probability for any microstate $a_i$ in system $S_a$ at temperature $T$ is

$$P_a(a_i) = \frac{1}{Z_a} e^{-E_a(a_i)/T}. \quad (14)$$

As the total probability must equal 1, the partition function $Z$ (I now drop indices for simplicity) is given by the sum of the Boltzmann factors for all possible microstates:

$$Z = \sum x e^{-E(x)/T}. \quad (15)$$

The partition function is a temperature dependent "constant" which weighs the states that are thermally available to the system. At very low temperature $Z$ is small as higher energy states are very unavailable and have very low Boltzmann factors. But as the temperature increases higher energy states become more available and $Z$ becomes much larger.

The ensemble average of any variable of interest can now be computed in a straight forward manner. If the variable is called $Q$ and it has a value
Q(x) when the system with temperature T is in microstate x, then its ensemble average is given by

$$\langle Q \rangle = \sum_x Q(x)P(x) = \frac{1}{Z} \sum_x Q(x)e^{-\beta E(x)},$$  \hspace{1cm} (16)

where $\beta \equiv 1/T$, and the brackets $\langle \cdot \rangle$ represent the average with respect to the system’s microstates. As discussed earlier, in the thermodynamic limit (very large system) self-averaging occurs, and the only observable value is the average value. Then, it is straightforward to prove that the thermodynamic energy is given by

$$E \rightarrow \langle E \rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\frac{\partial}{\partial \beta} \ln Z,$$  \hspace{1cm} (17)

and the thermodynamic free energy (9) is given by

$$F \rightarrow \langle F \rangle = -\beta^{-1} \ln Z.$$  \hspace{1cm} (18)

**Quenched Disorder**

So far all I have said about energy is that it is some characteristic quantity which is equal in all microstates corresponding to the same macrostate. And I’m not going to say much more. However, because the system is made of many degrees of freedom (coins), the energy must be some collective function of these degrees of freedom.

There are cases where the degrees of freedom interact with one another. For example, suppose that instead of three coins I had talked, say, about three magnets on the vertices of an equilateral triangle. To retain the binary nature of coin tossing, imagine that the magnets could only have an orientation perpendicular to the plane of the triangle (up or down). Then, just like with the coins, the intrinsic characteristic of the system would be given by how these three magnets are oriented. But since magnets do interact with one another, the so-called energy of this system would be dictated by some $3 \times 3$ interaction matrix $E$ which couples each of the magnet orientations.

If instead of three magnets there were $K$ magnets occupying the vertices of a regularly spaced lattice, then the interaction matrix $E$ which dictates the energy would be a well determined $K \times K$ matrix. Then, just like before, for large $K$ (thermodynamic limit) and a given interaction matrix $E$ the partition function (15) is given by

$$Z_E = \sum_x e^{-\beta E(x)},$$  \hspace{1cm} (19)
and the free energy (18) is

\[ F_E = -\beta^{-1} \ln Z_E. \quad (20) \]

As we saw before, the energy of a system is the sum of the energies of its parts. This is also true for the entropy, as we saw in (5). Then the same must hold for the free energy (9). This means that the values of these quantities are due to the sum of the many interactions contained in the large matrix \( E \). If the magnets were occupying only some random sites on the lattice (only a few fixed yet randomly chosen vertices), then the randomness of the positions would appear in the interaction matrix \( E \), which, although random, is in a quenched state (i.e. it is what it is and it is not changing any time soon). For a well behaved statistical distribution, the sum of many random quantities can be represented by their average value multiplied by their number. Therefore, the free energy of such a disordered thermodynamic \((K \to \infty)\) system must be self-averaging over the realizations of the random interactions:

\[ F = \bar{F}_E = -\bar{\beta}^{-1} \ln \bar{Z}_E, \quad (21) \]

where the overbar denotes averaging over the distribution of random interactions \( E \).

2 Quenched Disorder and Vector Precoding in MIMO Channels

In wireless multiple-input/multiple-output (MIMO) channels electromagnetic pulses are used to simultaneously convey information from a group of collocated transmitting antennas to a group of receiving antennas. The electromagnetic pulses are not orthogonal, yet they all travel over the same physical medium and bandwidth. The wireless MIMO channel might be represented by the following vector equality:

\[ \mathbf{r} = \mathbf{Ht} + \mathbf{n}, \quad (22) \]

where \( \mathbf{t} \) is a vector containing the electromagnetic pulses transmitted by the transmitting antennas, \( \mathbf{r} \) is a vector containing the pulses measured by the receiving antennas, \( \mathbf{n} \) is a random vector containing additive noise components, and \( \mathbf{H} \) is a transfer matrix dictated by the characteristics of the physical medium making up the wireless channel.

The wireless channel is shaped by scatterers such as buildings, mountains, trees, etc. Due to the random positioning of the scatterers (and possibly the antennas) there is some randomness in the entries of the transfer
matrix $H$. However, I assume these scatterers are in a fixed position during a transmission interval. I also assume that all the antennas are at fixed positions during this interval. In other words, the channel matrix $H$, although random, is in a quenched state.

Many times, a technique called vector precoding is used in order to ease the information processing by the receiving antennas [19–22]: the transmission protocol is set up in such a way that the transmitted vector $t$ is in reality a linear transformation of another relevant vector $x$, i.e.

$$t = Tx,$$

where the matrix $T$ is some function of the channel matrix $H$.

If the intended communication process could be fulfilled with more than one choice of $x$, then the collocated transmitting antennas would obviously like to choose $x$ (subject to some constraints) such that the transmitted energy per transmitted symbol in $x$ is minimum:

$$\min_{x} \frac{x^\dagger Ex}{K} \equiv E,$$

where $K \equiv \dim x$, and $E \equiv T^\dagger T$ is a function of the channel matrix $H$. Now notice that this expression can be rewritten as follows:

$$E = - \lim_{\beta \to \infty} \beta^{-1} K^{-1} \ln \sum_{x} e^{-\beta E(x)},$$

where

$$E(x, x) \equiv x^\dagger Ex.$$

Expression (25) just picks the realization of $x$ which yields the lowest interaction energy (this $x$ is called the ground state), and then it evaluates that energy (the ground state energy). The problem would be trivial if it weren’t for the fact that there are some constraints both in the magnitudes and signs of the components in $x$. Note that, if $K$ were very large (thermodynamic limit), then the argument of the $\beta$ limit would be a thermodynamic free energy (per microscopic degree of freedom in $x$), as clearly shown by eqs. (19) and (20).

We just made a connection between the concepts of statistical mechanics presented in the previous section and a measurable characteristic of the wireless MIMO channel. The choices of $x$ which can be made to fulfill the intended communication process are equivalent to what I referred to as the microstates of the system. The quenched state of the channel matrix $H$ gives rise to a quenched metric $E$ which is equivalent to the interaction matrix between the elements in a disordered system. A legitimate question is:
how many antennas does the MIMO channel need in order for the thermodynamic approximation to be valid? This question is addressed in Part II of this thesis.

The good news is that, should we want to find out the value of $\mathcal{E}$, we know that appropriate mathematical tools might be found in the statistical mechanics literature (after all, it is just a free energy). Such tools are presented in the next section, but we must not forget that the machinery of statistical mechanics relies on the thermodynamic approximation. This means that the number of antennas in the MIMO channel has to be large enough such that we can approximate the random contributions from $\mathbf{E}$ as their average, like we did in eq. (21). So, in order to use tools from statistical mechanics and treat the argument of the $\beta$ limit in eq. (25) as a thermodynamic free energy, we should write:

\[
\mathcal{E} = - \lim_{\beta \to \infty} \lim_{K \to \infty} \beta^{-1} K^{-1} \ln \sum_x e^{-\beta \mathbf{E}(\mathbf{E}, x)}, \tag{27}
\]

As discussed at the end of Section 1, the free energy is, in the thermodynamic limit, self-averaging over the realizations of the random interactions. Hence, we might rewrite (27) as follows:

\[
\mathcal{E} = \overline{\mathcal{E}} = - \lim_{\beta \to \infty} \lim_{K \to \infty} \beta^{-1} K^{-1} \ln \sum_x e^{-\beta \mathbf{E}(\mathbf{E}, x)}, \tag{28}
\]

where the overbar denotes averaging over the distribution of the random interactions $\mathbf{E}$.

3 The Replica Method

In eqs. (21) and (28) we are presented with the challenge of taking the average of the logarithm of a partition function over the random interactions in $\mathbf{E}$, which is far from trivial. In order to tackle the problem we can exploit the somewhat mystifying equality [23, sec. 6.8]

\[
\frac{c \ln U}{x} \bigg|_{x=0} = \lim_{x \to 0} \frac{c U^x - c}{x}, \tag{29}
\]

which for convenience I write as

\[
\overline{c \ln U} = \frac{\partial}{\partial n} c \ln \overline{U^n} \bigg|_{n=0}, \tag{30}
\]

where $c$ is just some constant upon which the probability distribution has no effect.
We might use (30) to rewrite the average in (28) as

$$\lim_{K \to \infty} \beta^{-1} K^{-1} \ln Z = \lim_{n \to 0} \frac{\partial}{\partial n} \lim_{K \to \infty} \beta^{-1} K^{-1} \ln Z^n,$$

(31)

where

$$Z \equiv \sum_x e^{-\beta E(x)},$$

(32)

The task is now reformulated in terms of finding the average of a power of the partition function $Z$. This is also far from trivial, unless the power $n$ is an integer, in which case we are just dealing with moments of $Z$. The variable $n$ is, however, a positive real number. In the following we make the key assumption that, although we shall take $n$ to be a positive integer, we can nevertheless continually extend the result of the average to $n \to 0$. This critical assumption together with expression (30), is known as the replica trick, and it is the cornerstone of the so-called replica method, which was developed in statistical physics to analyze disordered systems. The name replica refers to the fact that we are now taking the expectation of the product of $n$ identical replicas $(1, 2, \cdots, n)$ of $Z$:

$$Z^n = \prod_{a=1}^n Z_a.$$

(33)

This ingenious trick was first proposed by Kac in 1968 while analyzing lattice vibrations at the Trondheim Theoretical Physics Seminar [24]. However, its proving ground was the spin glass (a disordered magnetic system in which the local couplings vary randomly in sign and magnitude) where Edwards and Anderson boiled the problem down to its physical essentials in 1975 [25]. The method initially considered all copies of the system $(a = 1, \cdots, n)$ as identical, and was successful in finding known correct results for some simple models (see e.g. [26]). However, the very important infinite-ranged model by Sherrington and Kirkpatrick [27, 28] proved extremely difficult to treat and unphysical solutions were being produced. In 1978 Almeida and Thouless suggested the possibility of assuming that, somehow, perhaps different copies of the system shouldn’t be regarded as completely identical [29]. But it wasn’t until 1980 that Parisi formally proposed the replica symmetry breaking approach [30], which took many years to interpret and has only recently been proven to be the correct solution [31]. The symmetry amongst the replicas is addressed in the next subsections.

### Degeneracy and Ergodicity Breaking

Consider the possibility that there was more than one allowed ground state $x$ (states with the same energy are referred to as degenerate). This means
that there could be different identical systems which have all relaxed to the same ground state energy, yet they are in different microscopic states. A quite intuitive example: think of a few magnets with no external magnetic field (the magnets only interact with one another through some interaction $E$); then reversing the orientation of all the magnets would result in the same energy. Of course, this example was so trivial that it wasn’t really an example; both states were related through a very simple transformation and were, in essence, the same state looked at from two different sides. But as the variety in the states $x$ the system can adopt gets richer, and the interaction matrix $E$ exhibits more randomness, we might come across some cases which are not so trivial, where the degenerate states have no apparent relation to one another and the degeneracy factor (the number of such states) is very high.

But as we learned in Section 1, the ergodic hypothesis tells us that, after a long enough observation time, any given of these identical systems should flow through each and all of the equienergetic ground states with equal frequency. However, at sufficiently low temperatures ($\beta \rightarrow \infty$ certainly is) we encounter a new phenomenon: the minima in the energy landscape are now separated by energy barriers whose height is proportional to the size of the system, which, in our case is $K \rightarrow \infty$.

What do I mean by an energy landscape? Just think of a two dimensional plot whose vertical axis represents energy, and whose horizontal axis represents some space of adjacent microstates. You might unfold the horizontal axis onto several dimensions (as many as degrees of freedom there are in the microstates $x$). Then we will encounter a number of energy valleys (microstates with lower energies) separated by energy barriers (microstates with higher energies). So long as these barriers are of finite height, the Boltzmann probability distribution (14) allows systems to flow from one valley to another (from one low-energy microstate to another). This causes the ergodicity we discussed earlier: all microstates with similar energy are visited with equal frequency, therefore a time average is the same as an ensemble average using (14). However, if the energy barriers become infinitely high, the system will be trapped in a valley, and the only accessible microstates will be those corresponding to that energy valley. Hence, when the system is at zero temperature, it will only find the bottom of that specific valley as an accessible state, whether it is a global or a local minimum. This is referred to as ergodicity breaking, and it causes different identical systems to have completely different sets of microstates available to them.

But, why should we worry about all this? By looking at eq. (27), it is clear that degeneracy shouldn’t be a problem even if it were of the order $\exp K$. Neither should ergodicity breaking. This expression will always
return the ground state energy. So, what was this all about? In the next subsection I discuss how ergodicity breaking can be a problem when the replica trick is employed to tackle expression (28).

**Replica Symmetry?**

We might use (26), (31), (32) and (33) to rewrite eq. (28) as

$$\mathcal{E} = - \lim_{\beta \to \infty} \lim_{n \to 0^+} \frac{\partial}{\partial n} \beta^{-1} K^{-1} \ln \prod_{a=1}^{n} \sum_{x_a} e^{-\beta x_a^\dagger x_a}.$$  \hfill (34)

As we just learned, at low enough temperatures, the system could never jump over the infinite energy barrier separating two spaces of microstates. As a result the ergodic hypothesis which made it possible to apply the machinery of statistical mechanics simply won’t work. That is, it won’t work unless we restrict the partition function (32) to account only for the microstates \( x \) within the energy valley where the system is trapped. This means that, if there were in fact different energy valleys, each of the sums in (34) could potentially run over different subsets within the set of microstates. That is, different identical replicas of the system \((a = 1, \ldots, n)\) could be trapped in different valleys.

As we consider the zero temperature limit \((\beta \to \infty)\) only the microstates corresponding to the bottom of each valley become relevant in eq. (34). As we shall see in Part II of this thesis (see Appendix B in Paper B), the manner in which these microstates relate to one another turns out to be a critical factor when evaluating the average in (34). In particular, we shall find that this average depends strongly on the eigenstructure of a \(n \times n\) matrix \(Q\) whose elements are \(Q_{ab} \equiv x_a^\dagger x_b / K\), where \(x_a\) and \(x_b\) are the ground states of copies \(a\) and \(b\) of the system, respectively.

As a first approximation we could naively assume that, at zero temperature, all replicas are in the very same microstate. Or, should they be in different microstates, all of these ground states are trivially related to one another (they form a family) such that their overlap \(Q_{aa} = Q_{ab} = q_1\) for any pair \((a, b)\). This is known as the replica symmetry ansatz (RS) and, as we shall see and discuss in Part II of this thesis (Papers A and B), it can turn out to be a good assumption in certain systems.

The replica symmetry ansatz is, however, not the correct solution in general. While for some specific systems it might turn out completely correct (for example, if (24) were a convex optimization problem), for other systems it is just an approximation which might turn out more or less crude. This basic approximation might be refined by assuming that the RS
family of ground states is divided in two subfamilies such that $Q_{ab} = q_1$ if replicas $a$ and $b$ have their ground states in the same subfamily, or $Q_{ab} = q_2$ if replicas $a$ and $b$ have their ground states in different subfamilies. This is known as first order replica symmetry breaking (1RSB).

The process of refinement could go on to second order replica symmetry breaking (2RSB), where each of the two 1RSB families is divided into two subfamilies, and if replicas $a$ and $b$ have microstates in the same subfamily then their overlap is $Q_{ab} = q_1$. If they are in different subfamilies which originated from the same 1RSB family, then $Q_{ab} = q_2$. And if they are in different subfamilies, each from a different 1RSB family, then $Q_{ab} = q_3$. This nested process, shown in Fig. 0.2 could go on and on\(^4\) to add more and more degrees of freedom in the matrix $Q$.

The successive refinements in the replica structure bring us closer to the true value of $\langle 2 \rangle$, but, some times, it might be that RS or 1RSB yield the exact solution and adding more degrees of freedom is just redundant.

\(^4\)This is the replica symmetry breaking scheme proposed by Parisi in 1980 [30].
Unfortunately, there is no easy way to know. Replica symmetry breaking is, no doubt, the most puzzling and challenging aspect of this introduction. I have tried to spare the reader from tedious mathematics and keep it down to its very essentials.\footnote{Those looking for a thrill might find a more thorough (yet still quite pedagogical) treatment of replica theory in [32].} In Part II of this thesis (Papers A and B) I show how to use the replica method in the context of vector precoding in MIMO channels; fortunately, the optimization space in (24) is convex and RS can safely be assumed as, due to convexity, we expect that there is one and only one ground state.

4 The Annealed Approximation

In this section I shall present and justify a method which, in some cases, might save us from having to apply the tedious replica method. Let us first summarize what we have learned about disordered systems so far. We had a system whose internal degrees of freedom allow it to exist in different microstates \( x \) whose probabilities obey the Boltzmann distribution (14). These degrees of freedom could interact with one another with energy \( E(\mathbf{E}, x) \) according to a random interaction matrix \( \mathbf{E} \). The partition function for the microstates of the system is given by (19)

\[
Z_E = \sum_x e^{-\beta E(\mathbf{E},x)}, \quad (35)
\]

and the free energy is (20)

\[
F_E = -\beta^{-1} \ln Z_E. \quad (36)
\]

Then we argued that if the interaction matrix was in a quenched state, then the thermodynamic free energy \( F_E \) would be self-averaging with respect to the randomness in \( \mathbf{E} \) (21).

Now consider the case that the random interactions in \( \mathbf{E} \) were not quenched. Assume that just like the degrees of freedom in \( x \), the components in the random interaction matrix \( \mathbf{E} \) can change their values. Then we have to account not just for one, but for many possible states \( \mathbf{E} \). This means we have to create a new partition function; just like we had a partition function (35) for the states of \( x \), we now need one for the states of \( \mathbf{E} \), whose characteristic energy is \( F_E \). Given the non-discrete nature of \( \mathbf{E} \), the partition function won’t be a sum over matrix states \( \mathbf{E} \), but rather an average over the random realizations of \( \mathbf{E} \):

\[
Z \equiv \exp \left( -\tilde{\beta} F_E \right), \quad (37)
\]
where $\tilde{\beta}$ is the inverse temperature associated with the degrees of freedom of the interactions. We might insert (36) into (37) and obtain

$$Z = \exp \left( \tilde{\beta} \beta^{-1} \ln Z_E \right) = Z_E^{\beta/\tilde{\beta}},$$  \hspace{1cm} (38)

Just like before, we can now obtain the free energy of the system as

$$F = -\tilde{\beta}^{-1} \ln Z = -\tilde{\beta}^{-1} \ln Z_E^{\beta/\tilde{\beta}}.$$ \hspace{1cm} (39)

Expression (39) is very interesting. If we let $\tilde{\beta} = n\beta$, then we arrive at a replica formalism (33) again. Let us make this simple substitution:

$$F = -\tilde{\beta}^{-1} \ln Z = -\frac{1}{n} \beta^{-1} \ln Z_E^n.$$ \hspace{1cm} (40)

If we assume that the temperature of the interactions’ degrees of freedom is much larger than the temperature of the microstates’ degrees of freedom, i.e. $n \to 0$, then we recover (31)! This makes sense if you think about it: when the temperature of the matrix $E$ is much larger than the temperature of the microstates $x$, then the state of the matrix won’t be at all affected by the changes in the microstates $x$. Therefore we recover the quenched $E$ formalism.

On the other hand, if both temperatures are similar such that $\tilde{\beta} \approx \beta$ and as a result $n \approx 1$, then we can get away with taking just the average of the partition function, instead of the average of its logarithm. This means that, if we suspected that, somehow, as $K \to \infty$ each system configuration $x$ induces its own coupling matrix $E$ then there is no need for replica analysis. Think, if you will, of a data traffic network where the administrator has the power to change the topology depending on the data pressure across each pair of nodes, as opposed to a road traffic network where the interaction topology is quenched for long periods of time.

In Part II of this thesis (Paper C) I shall use the annealed approximation to analyze the information capacity of a complexity constrained vector channel.
5 Overview of Included Papers

This thesis consists of three papers, labeled by the capital letters A to C. Below I shall provide a summary of these papers.

Paper A


In this paper we consider a single-user MIMO broadcast channel where the transmitter wants to ease the task of detection at the multi-antenna receiver by removing all interference before transmission. However, if the number of transmit antennas is smaller than the number of receive antennas, direct channel inversion is not possible. Nevertheless, this problem may be overcome by modifying the channel inversion technique such that the receiver can apply a spatial matched filter.

Regardless of whether a spatial matched filter or distributed antenna detection is used by the receiver, channel inversion techniques at the transmitter come at an increased transmission energy cost. In order to contain the transmission energy, the quaternary signaling space is relaxed onto a convex superset of the original constellation. We show that, even when transmission occurs over singular channels, such convex relaxation scheme can be used to remove all interference. This implies that, at least under some conditions, the use of spatial matched filtering or distributed antenna detection comes down to a choice.

We analyze the performance (in terms of transmitted energy vs. uncoded transmission rate) of both spatial matched filtering and distributed antenna detection. This problem is similar to the one described in Section 2 but, as the optimization space is convex, we need not worry about degeneracy or ergodicity breaking. Therefore symmetry amongst the replicas of the system is safely assumed.

We find that, which of both detection schemes is most advantageous depends on the ratio of the number of antennas at the transmitter to that...
at the receiver. While for some ratios the choice is clear, for other ratios tradeoff situations occur.

**Paper B**

Rodrigo de Miguel and Ralf R. Müller, "On Convex Vector Precoding for Multiuser MIMO Broadcast Channels," under revision for possible publication in the *IEEE Transactions on Signal Processing*.


In this paper we consider a multiuser MIMO broadcast channel where the transmitter wants to remove interference before transmission in order to ease the task of detection by the single-antenna users. As opposed to the scenario in Paper A, spatial matched filtering is not an option as the receiving antennas are no longer collocated. But just as in Paper A channel inversion comes at the price of higher transmission energy, and in order to contain this energy the original signaling space is relaxed onto a convex superset. In this paper we explore and compare different convex precoding schemes to transmit one, two, or three uncoded bits. We examine all possible ratios of users to transmitting antennas, including those resulting in singular channels, and compute the probability that our methods fail to produce a signal which can be transmitted free of interference.

As in Paper A, replica symmetry is used to analyze and compare the large system performance (in terms of transmitted energy vs. uncoded transmission rate) of the different precoding schemes proposed. Quite intuitively, an extended optimization space allows for lower transmission energies than the original constellation points. And as the space is extended further, the transmission energies can be even lower. This suggests the use of both the real and the complex dimension to relax the original alphabets. Yet, when it comes to binary signals, one would think that a single dimension should be enough; in this paper we also propose an alternative channel inversion technique which allows the precoding of binary signals in real space to result in the same performance as in complex space.

**Paper C**

Rodrigo de Miguel, Ori Shental, Ralf R. Müller, and Ido Kanter, "Information and Multiaccess Interference in a Complexity Constrained Vector


In this paper we consider a simple single-user vector channel with a strict complexity constraint at the receiver. According to this constraint, detected bits must be obtained by performing hard decisions directly on the channel’s matched filter output. Using tools from statistical mechanics we show that, under a bounded noise assumption, such a channel exhibits a non-trivial Shannon-theoretic capacity.

Specifically we consider the situation where an input \( i = (i_1, i_2, \cdots, i_T) \) is transformed into an output \( \phi = (\phi_1, \phi_2, \cdots, \phi_N) \) as follows:

\[
\phi = \frac{1}{N} S^\dagger S A i + n,
\]

where \( S \) is a random channel matrix, \( A \) is a diagonal matrix of amplitudes, and \( n \) is a random additive noise vector whose components are all bounded in magnitude. In order to assess the information capacity of this channel we compute how many binary vectors \( i \in \{\pm 1\}^K \) there are, for each matrix \( S \), which fulfill the following preselection rule:

\[
i = \text{sgn}(\phi).
\]

For this purpose we employ a large system analysis and apply the annealed approximation discussed in Section 4 as, for a given set of constraints in \( A \) and \( n \), each binary state \( i \) has its very own relevant interaction matrix \( S^\dagger S \) in the large system limit. The validity of the annealed approximation is confirmed with finite size simulations for \( T = 20 \).

We find that, contrary to intuition, this channel yields a non-trivial capacity. We also analyze the effect that the amplitude distribution \( A \) has on the capacity. Although equal power interference is found to be the worst case interference scenario for linear multiuser receivers \[33\] we find that, in this channel, optimal cooperative interference occurs when all transmitters use equal amplitudes. Indeed we find that cooperation through interference is a key feature of this channel, which is in contrast to systems without input preselection, where higher channel capacity is obtained by interference mitigation at the receiver \[34, 33, 35\]. It is important to note that the input preselection rule employed in this paper is conceptually dif-
ferent from the vector precoding technique used in Papers A and B, where there is no restriction in the input signaling.

6 Papers Not Included in this Thesis

In addition to Papers A-C, I have participated in the elaboration of two other papers whose contents should be highlighted in the context of this thesis. These are:


In these papers we consider vector precoding for transmitting quaternary symbols over a MIMO broadcast channel. However, unlike in Papers A and B, we also consider a non-convex optimization space whereby the original quaternary phase shift keying constellation is relaxed onto a super-lattice. Then, the optimization problem is no longer convex and the replica symmetry ansatz simply does not hold. This forces us to consider first order replica symmetry breaking. As far as we know it is the first time replica symmetry breaking is invoked in the context of wireless communications.

7 Suggestions for Further Research

In the following, I briefly put forth some ideas which might be of interest for future research:

- An important factor to consider when it comes to implementing the precoding schemes proposed in Papers A and B is what the achievable information rate is for coded communications. Although in \[36\] we investigated the achievable rate for the quaternary convex precoding scheme in a MIMO broadcast channel, the capacity of the oc-
7. Suggestions for Further Research

...tonary schemes in Paper B remains to be investigated.

- Dirty paper coding \cite{37} techniques can be used to approach the capacity of the Gaussian MIMO broadcast channel. This is done using optimum lattice quantizers whose dimension approaches infinity \cite{38, 39}, suggesting the appealing use of lattice based precoding schemes.

In \cite{36} we compared the quaternary convex precoding scheme proposed in Papers A and B with a particular lattice precoding scheme whose analysis required replica symmetry breaking. When optimizing the transmitted energy, we found that the latter yields a significantly lower energy penalty per transmitted symbol than the former. However, when comparing the achievable rates of both precoding schemes, the results are mixed. In particular, for low signal to noise ratios the convex scheme yields higher rates than the lattice scheme; furthermore, in this regime, the lattice based scheme is detrimental in terms of achievable rate with respect to no-precoding.

Given the mixed results, one might wonder if there is a way to choose an optimal vector precoding space in terms of achievable information rate before proceeding to minimize the energy. Could we devise a target function (some sort of free energy) to optimize?

- In Section 3 we explored the reasons for replica symmetry breaking. However, in terms of vector precoding, our only conclusion was that the replica symmetry ansatz will hold for convex schemes. But as just discussed, we might want to consider lattice based precoding. If so, will there be replica symmetry breaking? Is convexity a necessary condition for replica symmetry? Or are there other, more general characteristics we should consider when trying to design a replica symmetric (much easier to analyze asymptotically) precoding space?

- Even if we were able to devise a lattice based precoding scheme achieving higher rates than convex schemes, and even if it allowed for a simple asymptotic analysis, we should also consider the complexity at the implementation stage. While optimizing over a lattice is a non-deterministic polynomial-time hard problem, optimizing the energy function over a convex set allows for efficient polynomial time algorithms.
8 Overview of Appendices

The three independent appendices in Part III of this thesis show some results that, although not part of any published work, are of theoretical interest and potentially useful in future work within vector precoding.

Appendix A: A Connected Extension of the Tomlinson-Harashima Lattice

In Papers A and B we considered minimizing quadratic forms using alphabets that were both continuous and convex. In Appendix A I shall consider an alphabet set that, although continuous, is non-convex. This alphabet is a connected complex extension of a binary Tomlinson-Harashima lattice [40][41]. In this Appendix I derive expressions for the replica symmetric parameters corresponding to this complex alphabet.

Appendix B: The R-transform of $\gamma \left( \frac{HH^\dagger}{N} + \gamma I \right)^{-1}$

When the goal of vector precoding is minimizing the mean square error (as opposed to minimizing the transmit power, as we did in Papers A and B), then the relevant metric whose eigendistribution must be known is of the form $\gamma \left( \frac{HH^\dagger}{N} + \gamma I \right)^{-1}$, where $H$ is the channel matrix [42]. In this Appendix the R-transform of the eigenvalue distribution corresponding to this matrix is derived and presented.

Appendix C: A Radially Invariant Replica Analysis

In this Appendix I consider the minimization of a quadratic form involving normalized vectors, i.e. a weighted average of the eigenvalues of a matrix. For a quenched matrix, such a quadratic form is uniquely defined by the phase of the vector. Thus, this problem naturally lends itself to a replica analysis where both the replica matrix and the subshells are radially invariant as well. I use this approach to analyze how close typical states drawn from radially invariant alphabets can get to the minimum eigenmode of certain random matrices with known eigenspectra.
References


Part II

Included papers
Paper A

On Convex Vector Precoding for Single-User MIMO Channels

Rodrigo de Miguel, Vesna Gardasevic and Ralf R. Müller

Under revision for possible publication in IEEE Transactions on Wireless Communications
Abstract

Convex vector precoding is proposed for interference-free reception in single user MIMO channels using either spatial matched filtering or distributed antenna detection at the receiver. An expression is presented for the probability that interference-free communication is possible over singular channels when a symmetric convex space is used to relax a QPSK alphabet. We use the replica method of statistical physics to compare the large system performance (in terms of transmitted energy vs. uncoded transmission rate) of both the matched filter and distributed detection schemes. We conclude that, with the proposed convex relaxation, distributed detection is most advantageous only when the ratio of antennas at the receiver to that at the transmitter is either smaller than 0.5 or between 1 and 1.22. Matched filter detection is most advantageous for ratios greater than 2. In the two remaining regions the choice precoding for matched filter or distributed detection results in a tradeoff between energy and transmission rate.

This material is based upon work partly supported by the Research Council of Norway under Grant No. 171133/V30.

This paper was presented in part at the 2008 IEEE International Zürich Seminar on Communications, Zürich, Switzerland, the 2008 IEEE International Symposium on Applied Sciences in Biomedical and Communication Technologies, Aalborg, Denmark, and the 2008 IEEE International Symposium on Information Theory and its Applications, Auckland, New Zealand.

R. de Miguel, V. Gardasevic and R. R. Müller are with the Department of Electronics and Telecommunications, The Norwegian University of Science and Technology, Trondheim, Norway, (e-mail: {demiguel,gardasevic,ralf}@iet.ntnu.no).
1 Introduction

In multiple-input/multiple-output (MIMO) channels information is conveyed simultaneously from a group of transmitting antennas to a group of receiving antennas. As these transmissions are not orthogonal, yet they occur over the same physical medium and bandwidth, crosstalk becomes unavoidable. As a result signal processing needs to be done at the receiver and/or transmitter side of the channel if significant data rates are to be achieved. In MIMO broadcast channels the transmitting antennas are collocated and they can jointly generate and pre-process the data streams to be transmitted. In contrast to multiuser channels, in a single-user MIMO channel, depicted in Fig. A.1, the receiving antennas are also collocated and they do have the possibility of jointly processing data.

However, in the context of low cost receivers with limited processing power, it might be advantageous to shift most of the signal processing to the transmitter side. One technique which might be employed by the transmitter in order to keep the receiver from doing any signal processing is channel inversion before transmission (provided that the transmitter has complete channel state information). If the number of transmit antennas is smaller than the number of receive antennas, direct channel inversion is not possible. However, this problem can be overcome modifying the channel...
inversion process and applying a spatial matched filter at the receiver.

Regardless of whether a spatial matched filter or distributed antenna detection is applied at the receiver, channel inversion at the transmitter comes at an increased transmission energy cost. One technique which may be used to contain the transmit power while inverting the channel is non-linear vector precoding (henceforth vector precoding) [1–4]. The vector precoding technique, outlined in Section 2, consists of extending the input alphabets representing different information states; this permits the search for symbols which draw less energy when transmitted with channel inversion. In [5, 6] we used the replica method of statistical physics to analyze different vector precoding techniques in asymptotic multiuser MIMO broadcast channels.

In this contribution we analyze vector precoding for matched filter and distributed antenna detection by a single multiantenna user and consider all transmitter to receiver antenna ratios, including those resulting in singular channels. Using a convex extension for quaternary phase shift keying (QPSK) alphabets proposed in [5], we compute the probability that our precoding method fails to remove all interference. We show that this probability goes to zero exponentially with the size of the MIMO system when matched filter detection is used by a receiver that has more than half as many antennas as the transmitter. This asymptotic convergence also occurs for systems with distributed detection if the number of transmitting antennas is more than half the number of data streams.

This paper is organized as follows. The vector precoding technique is presented in Section 2. In Section 3 we find the probability of interference-free transmission over singular MIMO channels. In Section 4 methods from statistical physics are used to derive the transmitted energy in the many antenna limit. The results are presented in Section 5. Some technical details regarding asymptotics of random matrices are relegated to the Appendix.

2 Vector precoding

The single-user MIMO channel may be represented by the following vector equality:

\[ r = Ht + n, \]  

(A.1)

where \( t \) is the \( N \)-dimensional input to the channel, \( r \) is a vector containing the \( K \) received signals, \( n \) is a random vector containing additive noise components, and the complex channel matrix \( H \) has independent and identically distributed Gaussian entries with zero mean and unit variance.
The transmitted vector $t$ is a linear transformation of the information symbols (contained in $x$) intended for the $K$ antenna elements at the receiver, thus we might write

$$ t \equiv Tx. \quad \text{(A.2)} $$

In order to guarantee simple detection by the multiantenna receiver, the transmitter, who has complete channel state information, might construct $T$ such that the information symbols in $x$ can be received interference free (up to additive noise) by a simple linear operation $\hat{\Omega}$ on $r$:

$$ \hat{\Omega}r = x + \hat{\Omega}n. \quad \text{(A.3)} $$

Using this transmission scheme, the transmitted energy per symbol is given by

$$ K^{-1}t^tt = K^{-1}x^TEx, \quad \text{(A.4)} $$

where the energy metric $E$ is given by

$$ E = T^TT. \quad \text{(A.5)} $$

**Matched filter detection**

When the receiver has more antennas than the transmitter ($K \geq N$) and the state of the channel is known at both ends, then a complex data vector $x$ containing $N$ entries might be detected interference free with a matched filter if the matrix $T$ and the operator $\hat{\Omega}$ are constructed as follows:

$$ T \rightarrow T_M \equiv \left( \frac{H^TH}{K} \right)^{-1}, \quad \text{(A.6)} $$

$$ \hat{\Omega} \rightarrow \hat{\Omega}_M \equiv \frac{1}{K}H^T. \quad \text{(A.7)} $$

Under this transmission scheme, the transmitted energy per symbol is given by

$$ K^{-1}t^tt = K^{-1}x^T E_Mx, \quad \text{(A.8)} $$

where

$$ E_M \equiv T_M^TT_M = \left( \frac{H^TH}{K} \right)^{-2}. \quad \text{(A.9)} $$
Distributed antenna detection

On the other hand, if the transmitter has more antennas than the receiver (i.e. \( K \leq N \)) and only the transmitter has channel knowledge, then distributed antenna detection at the receiver may be carried out if the matrix \( T \) and the operator \( \hat{\Omega} \) are constructed as follows \([5, 6]\):

\[ T \rightarrow T_D \equiv \frac{H^*}{\sqrt{N}} \left( \frac{HH^*}{N} \right)^{-1}, \quad (A.10) \]

\[ \hat{\Omega} \rightarrow \hat{\Omega}_D \equiv \frac{1}{\sqrt{N}} I. \quad (A.11) \]

Then the transmitted energy per symbol is given by

\[ K^{-1} t^t t = K^{-1} x^t E_D x, \quad (A.12) \]

where

\[ E_D \equiv T_D^t T_D = \left( \frac{HH^*}{N} \right)^{-1}. \quad (A.13) \]

Minimizing the transmitted energy

Although channel inversion by the transmitter eases the task of detection at the receiver, it might come at the cost of a high transmission energy. The goal of the vector precoding technique is minimizing the cost of the channel inversion process, i.e. minimizing eq. \((A.4)\). For this purpose, it is agreed between the transmitter and the receiver that, although there must be a minimum distance between any two symbols representing different information states, each state might be represented by more than one symbol. This gives the transmitter greater freedom to construct the information vector \( x \) with symbols which faithfully represent the intended information, yet they are chosen so as to minimize eq. \((A.4)\).

The \( i \)th component of the vector \( x \) represents the information state \( s_i \). The symbols which might represent the state \( s_i \) are those contained in the set \( A_{s_i} \). Then the information vector \( x \) is constructed such that \( x \in A \), where \( A = \prod A_{s_i} \). The transmitter chooses the symbol representation in \( A \) which minimizes the transmitted energy:

\[ x \equiv \arg \min_{x \in A} \tilde{x}^t E \tilde{x}. \quad (A.14) \]

If the symbol alphabets are discrete, then \( x \) can only be found performing an exhaustive search, which becomes prohibitively expensive when the
3. Vector Precoding in Singular Channels

Alphabet relaxation for quaternary phase shift keying

We consider now a source of information consisting of four equiprobable states: $\uparrow\uparrow$, $\uparrow\downarrow$, $\downarrow\downarrow$ and $\downarrow\uparrow$. When no vector precoding is employed the entries in $\mathbf{x}$ are usually selected from the unit QPSK alphabets $A_{\uparrow\uparrow} = \{1 + j\}$, $A_{\downarrow\downarrow} = \{-1 - j\}$, $A_{\uparrow\downarrow} = \{-1 + j\}$ and $A_{\downarrow\uparrow} = \{1 - j\}$. We consider the convex alphabet relaxation shown in Fig. A.2 which was proposed in [5]. This extended quaternary alphabet is given by

$$A_{\uparrow\uparrow} = -A_{\downarrow\downarrow} = \{\xi \text{ s.t. } \Re \xi \geq 1 \& \Im \xi \geq 1\},$$
$$A_{\uparrow\downarrow} = -A_{\downarrow\uparrow} = \{\xi \text{ s.t. } \Re \xi \geq 1 \& \Im \xi \leq -1\}. \quad (A.15)$$

3 Vector Precoding in Singular Channels

In the previous section we considered only the inversion of full rank matrices and as a result the choice of precoding scheme was conditioned on whether $K \leq N$ or $N \leq K$. However, the transmitter could, in principle, invert the relevant eigendirections of singular channel-based matrices: when
A. ON CONVEX VECTOR PRECODING FOR SINGLE-USER MIMO CHANNELS

A matrix $M$ is hermitian, as are both $HH^\dagger$ and $H^\dagger H$, one might write

$$M = U\Lambda U^\dagger,$$

where $U$ is unitary and $\Lambda \equiv \text{diag} (\lambda_1, \lambda_2, \cdots , \lambda_K)$ is a diagonal matrix containing the $K$ eigenvalues of $M$. We might then define the pseudoinverse of $M$ as

$$M^{-1} = U\tilde{\Lambda}^{-1}U^\dagger,$$

where

$$\tilde{\Lambda}^{-1} \equiv \lim_{\epsilon \to 0} \text{diag} \left( \frac{1-\delta_{\lambda_1,0}}{\lambda_1 + \epsilon}, \frac{1-\delta_{\lambda_2,0}}{\lambda_2 + \epsilon}, \cdots , \frac{1-\delta_{\lambda_K,0}}{\lambda_K + \epsilon} \right)$$

and $\delta_{i,j}$ is the Kronecker delta.

If the precoded vector $x$ in (A.14) could be constructed in the $\min\{N, K\}$-dimensional space spanned by $H^\dagger H$ (denoted as $S_M$), then this generalized channel inversion technique could be used to construct $T_M$ and matched filter precoding would be possible irrespective of the ratio between the number of antennas at transmitter and receiver. Analogously, if the precoded vector $x$ could be constructed in the span of $HH^\dagger$ (denoted as $S_D$), then this channel inversion technique would make precoding for distributed antenna detection possible irrespective of the ratio of antennas. This brings us to the question: can the transmitter construct such a vector $x$? Because the regions in the space $A$ in Fig. A.2 have an infinite area where the vector components can be tuned, it is conceivable that the components of a typical QPSK vector $x$ can be freely relaxed in the two dimensional regions until the vector lies in the lower dimensional span of the relevant matrix. This possibility for overloading is a distinct advantage that two dimensional convex relaxation schemes offer compared to lattice based schemes such as those proposed in [1, 3, 9].

Whether or not overloaded transmission is possible boils down to the probability that the space $A$ and the span of the matrix intersect. Due to the unitary invariance of $HH^\dagger$ and $H^\dagger H$ and the symmetry in the alphabet $A$ depicted in Fig. A.2, we might take $s_i = \uparrow \uparrow \forall i$ without loss of generality. Furthermore, the span is invariant to the norm of its basis vectors; we might, therefore, for the purpose of analysis, take $\uparrow \uparrow$ in Fig. A.2 to occupy the entire first quadrant. Then the question reduces to finding the probability that a complex vector all of whose components lie in the first quadrant lies also in the span of the matrix. For a $K \times N$ matrix $H$ with independent and identically distributed complex Gaussian entries the probability that the span contains such a vector is given by [10]

$$P_{A \cap S_M} (K, N) = f(K, N),$$

(A.19)
Figure A.3: Probability that convex precoding fails to find a solution vs. the ratio of the number of antennas at the receiver to that at the transmitter.

\[ P_{A \cap S_D}(K, N) = f(N, K), \]  

where

\[ f(N, K) \equiv 2^{1-2N} \sum_{\ell=0}^{2K-1} \binom{2N-1}{\ell}. \]

Figure A.3 shows the probability that the proposed convex precoding scheme fails to find a vector \( x \) which lies in the space spanned by the channel matrix. While for full rank matrices the probability of failure is zero, as the null space of the matrix increases finding a solution becomes less and less probable. However, significant overloads (50-80%) are possible while keeping the probability of failure relatively small. It should be noted that when \( K \) and \( N \) grow to infinity the probability collapses to a step function: interference-free transmission is always possible if the matrix is at least half rank and impossible if the rank is lower than half.

To allow for transmission over singular channels, some of the expressions from Section 2 have to be slightly modified in light of the generalized channel inversion technique:

\[ T_M \equiv \left( \frac{H^* H}{K} \right)^{-1}, \]  

\[ E_M \equiv \left( \frac{H^* H}{K} \right)^{-2}. \]
A. On Convex Vector Precoding for Single-User MIMO Channels

\[
T_D \equiv \frac{H^H}{\sqrt{N}} \left( \frac{HH^H}{N} \right)^{-1}, \quad (A.24)
\]

\[
E_D \equiv \left( \frac{HH^H}{N} \right)^{-1}, \quad (A.25)
\]

and

\[
x \equiv \arg \min_{\tilde{x} \in A \cap S} \tilde{x}^\dagger E \tilde{x}, \quad (A.26)
\]

where \( S \) denotes the span of the relevant matrix for the chosen transmission scheme, \( i.e. \ S_M \) for matched filter detection, or \( S_D \) for distributed detection.

4 The transmitted energy: ensemble theory and the replica method

The technique outlined in Section 2 describes how to minimize the transmitted energy while achieving simple and interference free detection by the receiver. In order to obtain the value of the energy one might first note that, if \( x \) contains \( T \) symbols, the minimum transmitted energy per symbol \( E \) might be written as

\[
E \equiv \min_{x \in A \cap S} T^{-1} x^\dagger E x = - \lim_{\beta \to \infty} \beta^{-1} T^{-1} \ln \sum_{x \in A \cap S} e^{-\beta x^\dagger E x}. \quad (A.28)
\]

The argument of the \( \beta \) limit in (A.28) has the same form as the expression for the Helmholtz free energy of a thermodynamic (\( T \to \infty \)) system with temperature \( 1/\beta \), which can exist in the states \( x \in A \cap S \) and whose energy is dictated by the interaction matrix \( E \) \( [11] \). In the following we take advantage of this fact by taking a thermodynamic approximation, \( i.e. \) we always assume that \( K \) and \( N \) are infinitely large\(^\ast\), yet they have a finite ratio \( \alpha = K/N \). This approximation will allow us to make use of mathematical tools imported from the statistical mechanics literature.

The most fundamental tool of modern statistical mechanics is ensemble theory, according to which all thermodynamic (large system) quantities can

\(^\ast\)In \[6\] we show that this asymptotic channel limit can be a fair approximation for systems with as few as 10 transmitting antennas if the alphabets representing the different information states are convex.
be written as an ensemble average of a suitable microscopic observable. In other words, the observable properties of thermodynamic systems become the average properties of all the microscopic states they can exist in. Sometimes, due to strong disorder in the system, the interactions dictating the properties of the microscopic states are random; each fixed in time, but with a random value due to the random positions of the system’s interacting elements. When the system is large, the number of random elements in the interaction matrix $E$ is enormous. The value of the free energy (A.27) is a result of the sum of the many random interactions contained in $E$, and for a well-behaved statistical distribution, the sum of many random quantities can be represented by their average value multiplied by their number. Therefore, the free energy of a disordered thermodynamic system must be self-averaging over the realizations of the random interactions.

In the context of the MIMO channel described above, the properties of the interaction matrix $E$ are dictated by the particular state the channel $H$ is in, which dictates the system’s observable properties (such as $E$). Hence, as the dimension of the channel increases, we might take a thermodynamic approach and focus just on its average properties.

In the thermodynamic limit the eigenvalue distribution of $E$ is well defined and fully determined by the statistics of the channel matrix $H$. As we shall see in Appendix A, a function which fully describes the eigenvalue distribution of $E$ is its R-transform, denoted $R_E(\cdot)$. Using the replica method of statistical mechanics and assuming the replica symmetric ansatz (further discussed below), we showed in [5] that the energy (A.27) is fully determined by the eigendistribution of $E$ and the alphabet $A$ as follows:

$$E = q \frac{d}{d\chi} R_E (-\chi),$$

where the parameters $q$ and $\chi$, defined as $q + \beta^{-1} \chi \equiv x^t x / T$ for $x$ given by (A.26), are given by the following pair of coupled self-consistent equations:

$$q = \sum_i \frac{p_i}{\sqrt{q R_E'(-\chi)}} \int_c \left| \arg \min_{\xi \in A_i} \left| \frac{\sqrt{q R_E'(-\chi)}}{R_E(-\chi)} - \frac{\xi}{z} \right| z^2 Dz, \right.$$

$$\chi = \sum_i \frac{p_i}{\sqrt{q R_E'(-\chi)}} \int_c \left| \arg \min_{\xi \in A_i} \left| \frac{\sqrt{q R_E'(-\chi)}}{R_E(-\chi)} - \frac{\xi}{z} \right| z Dz. \right.)$$

The index $i$ denotes the different information states, each of which can be represented by elements in the set $A_i$ and occurs in the components of $x$ with probability $p_i$. The expression $R_E'(\chi)$ denotes the first derivative of $R_E(t)$ evaluated at $t = -\chi$. And $Dz \equiv \frac{e^{-z^2}}{\pi} dz$. 

43
For the alphabet relaxation proposed in Section 5, Eqs. (A.30) and (A.31) reduce to

\[ q = 2 + \frac{e^{-\Gamma^2}}{\sqrt{\pi \Gamma}} + \left\{ \frac{1}{\Gamma^2} - 2 \right\} Q\left( \sqrt{2\Gamma} \right), \]  

(A.32)

\[ \chi = \frac{Q\left( \sqrt{2\Gamma} \right)}{R_E(-\chi)}, \]  

(A.33)

where

\[ \Gamma \equiv \frac{R_E(-\chi)}{\sqrt{qR_E'(-\chi)}}, \]  

(A.34)

and the function \( Q(\cdot) \) is defined as

\[ Q(\sigma) \equiv \frac{1}{\sqrt{2\pi}} \int_{\sigma}^{\infty} e^{-t^2/2} dt. \]  

(A.35)

Besides the fundamental postulate of statistical mechanics, the replica method often relies on the so-called replica symmetry ansatz [12], according to which the overlap between any two local minima \( x \) in (A.26) is constant. This important assumption was made in the derivation of Eqs. (A.29)–(A.31). Although this approximation generally yields only a lower bound to the energy, it is commonly used in the analysis of magnetic and neural systems exhibiting quenched interactions (see e.g. [13, 14]) often yielding excellent agreement with exact and numerical results (see e.g. [15]). Indeed replica symmetry has been successfully applied to problems in wireless communications (see e.g. [16–19]) and coding theory (see e.g. [20, 21]). In [6] we showed that the replica symmetric assumption mimics finite size results for the convex alphabet considered in this work. This occurs because convex alphabets can be expected to yield one and only one solution \( x \) which minimizes (A.27).

One should be warned, however, that although replica symmetry yields asymptotically accurate results for convex alphabets, it fails to produce accurate results when the original alphabets are relaxed onto a superlattice. Extended lattice (or otherwise non-convex) alphabets may exhibit complex energy landscapes with many peaks and valleys. This might cause different identical copies (replicas) of the system to become trapped in different energy wells at low temperature. The replica method must then be employed invoking Parisi’s replica symmetry breaking scheme [22]. For a recent and novel analysis of lattice alphabets based on replica symmetry breaking the reader is referred to [8]. For a thorough discussion of replica symmetry, the reader is referred to comprehensive literature on spin glasses [23–25].
5. Results

In order to properly compare, from the point of view of the transmitter, the matched filter and distributed detection schemes proposed in Section 2, we introduce the uncoded spectral efficiency $\epsilon$, which is the number of bits per transmitting antenna. While for matched filter detection (MF) $\epsilon$ always equals 2, for distributed antenna detection (DD) $\epsilon$ equals $2\alpha$.

Figure A.4 shows the energy per transmitted bit (A.27) for both schemes as a function of $\alpha$. As anticipated in Section 3, in the large system limit a vector $x$ cannot be found in the span when the rank of the matrix is less than $1/2$. Figure A.3 shows that this is nearly always the case also for finite systems. When the ratio of antennas at the transmitter to that of the receiver is smaller than 0.5 only the DD scheme allows for the possibility of constructing a typical vector in $\mathcal{S}$; analogously, only the MF scheme might be employed when $\alpha > 2$.

When the ratio of antennas at the receiver to that at the transmitter is between 1 and 1.22 greater $\epsilon$ is achieved by the DD scheme at a lower cost in energy per transmitted bit. However, as we can see in Fig. A.3, in this range DD has a slight probability of failure in finite systems; nevertheless, this probability is small and perhaps tolerable in systems with subsequent error control coding.

In the two remaining regions, namely $0.5 < \alpha < 1$ and $1.22 < \alpha < 2$, the...
A. ON CONVEX VECTOR PRECODING FOR SINGLE-USER MIMO CHANNELS

Figure A.5: Energy per bit vs. spectral efficiency tradeoff for $1 > \alpha > 0.625$. As $\alpha$ gets close to 1 DD is evidently more advantageous as it allows the transmitter to use less energy at no cost in spectral efficiency.

Figure A.6: Energy per bit vs. spectral efficiency tradeoff for $1.22 < \alpha < 1.6$. As $\alpha$ gets close to 1.22 DD is evidently more advantageous as it allows for higher spectral efficiency at no additional per-bit cost.
a tradeoff situation occurs in terms of energy per transmitted bit vs. $\epsilon$. However, one must be careful when $\alpha$ is too close to 0.5 (only DD is viable in realistic finite channels) or too close to 2 (only MF is viable). When these two critical subregions are avoided, the $E$ to $\epsilon$ tradeoff, shown in Figs. A.5 and A.6 is roughly exponential.

Depending on the priorities and resources of transmitter and receiver, other tradeoff situations might be of interest. For instance, looking at Fig. A.4 we can see that, if the receiver has only slightly more than twice as many antennas as the transmitter, then some rate might be gained by turning off roughly half of the receiving antennas, such that $\alpha \in (1, 1.22)$, and switching to DD mode; this gain in rate would come, however, at a moderate probability of failure and also a moderate additional energy penalty per bit for the transmitter.

6 Appendix A. Asymptotics of Random Matrices

Let $P_M(x)$ denote the eigenvalue distribution of the matrix $M$. Let

$$m_M(s) = \int \frac{dP_M(x)}{x-s},$$

(A.36)

which is known as the Stieltjes transform. Then, the R-transform of $P_M(x)$ is

$$R_M(w) = m_M^{\text{inv}}(-w) - \frac{1}{w},$$

(A.37)

with $m^{\text{inv}}(\cdot)$ denoting the inverse function of $m(\cdot)$.

Asymptotics of $\left(\frac{HH^\dagger}{N}\right)^{-1}$

The reader is referred to Appendix B in [5]. There the R-transform of $M \equiv \left(\frac{HH^\dagger}{N}\right)^{-1}$ is shown to be

$$R_M(w) = \frac{2}{1 - \alpha + \sqrt{(1-\alpha)^2 - 4\alpha w}},$$

(A.38)

Asymptotics of $\left(\frac{H^\dagger H}{K}\right)^{-2}$

As $K = \alpha N \to \infty$ (for $\alpha$ finite and positive) the matrix $B \equiv \frac{H^\dagger H}{K}$ has a limiting eigenvalue distribution, namely the Marchenko Pastur distribu-
A. ON CONVEX VECTOR PRECODING FOR SINGLE-USER MIMO CHANNELS

d\!P_B(x) = (1 - \alpha)^+ \delta(x)dx + \frac{\alpha}{2\pi x} \sqrt{\left(x - (1 - 1/\sqrt{\alpha})^2\right)^+ \left(x - (1 + 1/\sqrt{\alpha})^2\right)^+} dx.  \tag{A.39}

As we did in eq. (A.18), for matrix inversion purposes we neglect the contributions from the zero eigenvalues. We then fully characterize this eigenvalue distribution by its Stieltjes transform, which is given by

\[ m_B(s) = \frac{\alpha}{2s} \left( 1 - \frac{1}{\alpha} - s + \sqrt{\frac{1}{\alpha^2} + (s - 1)^2 - 2\frac{1}{\alpha}(1 + s)} \right). \tag{A.40} \]

The eigenvalue distribution of \( \tilde{B}^{-2} \) may also be characterized by its Stieltjes transform, which may be written in terms of the Stieltjes transform of \( B \) as follows:

\[ m_{\tilde{B}^{-2}}(s) = \frac{1}{2s} \left( s^{-1/2} m_B(-s^{-1/2}) - s^{-1/2} m_B(s^{-1/2}) - 2 \right). \tag{A.41} \]

The R-transform of the eigendistribution of \( \tilde{B}^{-2} \) may now be obtained explicitly using eq. (A.37), resulting in the somewhat tedious expression

\[ R_{\tilde{B}^{-2}}(w) = 2 \left( \rho - \sqrt{\kappa + \lambda} - \sqrt{\kappa - \lambda + \rho^2 - \delta - \frac{\epsilon + 8\rho^3 - 24\mu}{4\sqrt{\kappa + \lambda}}} \right)^{-1} - \frac{1}{w}, \tag{A.42} \]

where

\[ \epsilon \equiv \frac{64\alpha^2 w^3}{1 + \alpha + w}, \tag{A.43} \]

\[ \delta \equiv \frac{2\alpha w^2}{3\tau} (5 + 14\alpha + 5\alpha^2), \tag{A.44} \]

\[ \rho \equiv w(1 + 7\alpha + 7\alpha^2 + \alpha^3 + \alpha w), \tag{A.45} \]

\[ \lambda \equiv 2\alpha^2 w^4 \left( (\alpha - 1)^4 + 24\alpha w(\alpha + 1) + 2^{1/3} \psi^{2/3} \right) \frac{3}{3 \cdot 2^{2/3} \psi^{1/3}}, \tag{A.46} \]

\[ \mu \equiv \frac{\alpha \psi^3}{6\tau^2} (5 + \alpha(14 + 5\alpha)) (1 + \alpha (7 + \alpha(7 + \alpha) + w)), \tag{A.47} \]

\[ \kappa \equiv \frac{w^2}{12} \left( \alpha^2 + 1 \right)^2 (3 + \alpha(2 + 3\alpha)) - 2\alpha w(1 + \alpha) (17 + \alpha(38 + 17\alpha)) + 3\alpha^2 w^2 \frac{\tau^2}{\tau^2}, \tag{A.48} \]
and
\[ \tau \equiv (1 + \alpha)(1 + \alpha + w), \quad (A.49) \]
\[ \psi \equiv 2\alpha^3 w^6 \left( 72\alpha w(\alpha - 1)^2(\alpha + 1) + 54\alpha^2 w^2 - (\alpha - 1)^6 \right) + 12\sqrt{3}\sqrt{\alpha^7 w^{13}}\eta, \quad (A.50) \]
\[ \eta \equiv \alpha w(\alpha - 1)^4(31 + \alpha(66 + 31\alpha)) - 2(\alpha - 1)^8(\alpha + 1) + 27\alpha^3 w^3 - 8\alpha^2 w^2(1 + \alpha)(7 + \alpha)(1 + 7\alpha). \quad (A.51) \]
References


Paper B

On Convex Vector Precoding for Multiuser MIMO Broadcast Channels

Rodrigo de Miguel and Ralf R. Müller

Under review for possible publication in IEEE Transactions on Signal Processing
Abstract

We propose different convex alphabet relaxation schemes for vector precoding in MIMO broadcast channels using binary, quaternary and octonary modulation. Expressions are presented for the probability of the different precoding schemes to achieve interference-free communication over singular channels. The energy of transmission is evaluated in the many user limit using the replica method from statistical physics. An alternative channel inversion technique is proposed which makes purely real binary alphabets perform as well as their complex extensions, resulting in reduced complexity in the optimization process. The relevance of the asymptotic analysis is confirmed by finite size simulations.

This material is based upon work partly supported by the Research Council of Norway under Grant No. 171133/V30.

This paper was presented in part at the 2008 IEEE International Zürich Seminar on Communications, Zürich, Switzerland, and the 2008 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France.

R. de Miguel and R. R. Müller are with the Department of Electronics and Telecommunications, The Norwegian University of Science and Technology, Trondheim, Norway, (e-mail: {demiguel,ralf}@iet.ntnu.no).
1 Introduction

In multiple-input/multiple-output (MIMO) channels information is conveyed simultaneously from a group of transmitting antennas to a group of receiving antennas. As these transmissions are not orthogonal, yet they occur over the same physical medium and bandwidth, crosstalk becomes unavoidable. As a result signal processing needs to be done at the receiver and/or transmitter side of the channel if significant data rates are to be achieved. In the context of low cost receivers with limited processing power and battery life, it might be advantageous to shift most of the signal processing to the transmitter side.

In the case of the MIMO-broadcast channel, depicted in Fig. B.1, the receiving antennas are at different locations, which means that they might not jointly process the data they independently receive. The transmitting antennas, however, are collocated and they jointly generate the data streams to be transmitted to each of the single receiving antennas, also known as users. One technique which might be employed by the transmitter in order to keep the users from doing any signal processing is channel inversion before transmission (provided that the transmitter has complete channel state information). Unfortunately, plain channel inversion at the transmitter comes at an increased transmission energy cost. One technique which may be used to contain the transmit power while inverting the channel is non-linear vector precoding (henceforth vector precoding) [1–4]. The vector precoding technique, outlined in Section 2, consists of extending the input alphabets representing different information states; this permits the search for symbols which draw less energy when transmitted with channel inversion.

Müller et al. used the replica method of statistical physics to analyze vector precoding in asymptotic MIMO channels [5]. They used the *replica symmetry ansatz* and proposed the use of both convex and non-convex supersets of the original constellation points. However, recent results [6, 7] show that, contrary to convex alphabets, using non-convex alphabets induces *replica symmetry breaking*, which results in greater transmission energies than anticipated in [5]. In light of this result, optimizationally complex non-convex alphabets are less appealing than convex alphabets, which can achieve good performance while allowing for efficient optimization algorithms [6].

In this work we explore different convex precoding schemes to transmit one, two, or three uncoded bits. We examine all possible ratios of users to transmitting antennas, including those resulting in singular channels, and compute the probability that our methods fail to produce a signal which
B. ON CONVEX VECTOR PRECODING FOR MULTIUSER MIMO BROADCAST CHANNELS

The replica method is used to analyze the large system performance (in terms of transmitted energy vs. uncoded transmission rate) of the different precoding schemes proposed. Quite intuitively, an extended optimization space allows for lower transmission energies than the original constellation points. And as the space is extended further, the transmission energies can be even lower. This suggests the use of both the real and the complex dimension to relax the original alphabets. Yet, when it comes to binary signals, one would think that a single dimension should be enough (indeed, Schober et al. showed that the real part of the matched filter output gives sufficient statistics for multiuser detection of DS-CDMA with BPSK modulation [8]); in this paper we propose an alternative channel inversion technique which allows the precoding of binary signals in real space to result in the same performance as in complex space.

This paper is organized as follows. The vector precoding technique is outlined in Section 2. In Section 3 methods from statistical physics are used to derive the transmitted energy in the many user limit. Sections 4, 5 and 6 present, respectively, alphabet relaxation schemes for binary, quaternary, and octonary modulation. In Section 7 we find the probability of the different relaxation schemes to achieve interference-free transmission. The results are presented in Section 8. Some technical details, including those of the replica method, are relegated to the Appendices.
2 Vector precoding

The MIMO broadcast channel, depicted in Fig. B.1 may be represented by the following vector equality:

\[ r = Ht + n, \]  \hspace{1cm} (B.1)

where \( t \) is the \( N \)-dimensional input to the channel, \( r \) is a vector containing the \( K \) received data streams, \( n \) is a random vector containing additive noise components, and the channel matrix \( H \) is a complex rectangular matrix which can be written as \( H = H_r + jH_i \), where \( H_r \) and \( H_i \) are real random matrices containing independent and identically distributed (i.i.d.) entries with zero mean and variance \( 1/2 \).

The transmitted vector \( t \) is an \( N \)-dimensional linear transformation of the \( K \) information symbols (contained in \( x \)) intended for the \( K \) users, thus we might write

\[ t = Tx. \]  \hspace{1cm} (B.2)

In order to guarantee individual detection by the receivers, the transmitter, who has complete channel state information, might construct \( T \) such that the information symbols in \( x \) can be received interference-free (up to additive noise) by a simple diagonal (yet not necessarily linear) operation \( \hat{\Omega} \) on \( r \):

\[ \hat{\Omega}r = x + \hat{\Omega}n. \]  \hspace{1cm} (B.3)

Using this transmission scheme, the energy per transmitted symbol is

\[ K^{-1}t^\dagger t = K^{-1}x^\dagger Ex, \]  \hspace{1cm} (B.4)

where the energy metric \( E \) is given by

\[ E = T^\dagger T. \]  \hspace{1cm} (B.5)

Channel inversion

The \( K \times K \) matrix \( HH^\dagger \) has rank given by \( \min \{N, K\} \), and its inverse exists only if \( K \leq N \). A channel is said to be overloaded when there are more receiving users than there are antennas at the transmitter, i.e. \( K > N \). In order to allow for the possibility of inverting overloaded channels, the transmitter might employ the generalized channel inversion technique outlined in the following.

When a matrix \( M \) is hermitian, as is \( HH^\dagger \), we might write

\[ M = U\Lambda U^\dagger, \]  \hspace{1cm} (B.6)
where $\mathbf{U}$ is a unitary matrix and $\Lambda \equiv \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_K)$ is a diagonal matrix containing the $K$ eigenvalues of $\mathbf{M}$. We might then define the pseudoinverse of $\mathbf{M}$ as

$$\mathbf{M}^{-1} = \mathbf{U} \Lambda^{-1} \mathbf{U}^{\dagger}, \quad (\text{B.7})$$

where

$$\Lambda^{-1} \equiv \lim_{\epsilon \to 0} \text{diag} \left( \frac{1 - \delta_{\lambda_1,0}}{\lambda_1 + \epsilon}, \frac{1 - \delta_{\lambda_2,0}}{\lambda_2 + \epsilon}, \cdots, \frac{1 - \delta_{\lambda_K,0}}{\lambda_K + \epsilon} \right), \quad (\text{B.8})$$

and $\delta_{ij}$ is the Kronecker delta.

### Complex symbol mapping

If the symbols used to map the data are complex, i.e. $\mathbf{x} \in \mathbb{C}^K$, and the $K$-dimensional complex data vector $\mathbf{x}$ lies in the min $\{N, K\}$-dimensional space spanned by $\mathbf{HH}^{\dagger}$, then the matrix $\mathbf{T}$ and the diagonal operator $\hat{\Omega}$ can be constructed as follows:

$$\mathbf{T} \to \mathbf{T}_c \equiv \frac{\mathbf{H}^{\dagger}}{\sqrt{N}} \left( \frac{\mathbf{HH}^{\dagger}}{N} \right)^{-1}, \quad (\text{B.9})$$

$$\hat{\Omega} \to \hat{\Omega}_c \equiv \frac{1}{\sqrt{N}} \mathbf{I}. \quad (\text{B.10})$$

Then the transmitted energy per symbol is given by

$$K^{-1} \mathbf{t}^{\dagger} \mathbf{t} = K^{-1} \mathbf{x}^{\dagger} \mathbf{E}_c \mathbf{x}, \quad (\text{B.11})$$

where

$$\mathbf{E}_c \equiv \mathbf{T}_c^{\dagger} \mathbf{T}_c = \left( \frac{\mathbf{HH}^{\dagger}}{N} \right)^{-1}. \quad (\text{B.12})$$

### Real symbol mapping

If, on the other hand, only purely real symbols are used to map the data for the $K$ users, i.e. $\mathbf{x} \in \mathbb{R}^K$, then, as long as the real vectors $\mathbf{x}$ lie in the span of $\mathbb{R} \left( \mathbf{HH}^{\dagger} \right)$, the matrix $\mathbf{T} \to \mathbf{T}_R = \mathbf{T}_r + j \mathbf{T}_i$ and the diagonal operator $\hat{\Omega}$ might be constructed as follows:

$$\begin{bmatrix} \mathbf{T}_r \\ \mathbf{T}_i \end{bmatrix} \equiv \frac{1}{\sqrt{N}} \begin{bmatrix} \mathbf{H}^{\dagger} \\ -\mathbf{H}^{\dagger} \end{bmatrix} \left\{ \mathbb{R} \left( \frac{\mathbf{HH}^{\dagger}}{N} \right) \right\}^{-1}, \quad (\text{B.13})$$

$$\hat{\Omega} \to \hat{\Omega}_R \equiv \frac{1}{\sqrt{N}} \mathbb{R}_r \quad (\text{B.14})$$
where $\Re$ is the real-part operator. Under this transmission scheme, the transmitted energy per symbol is given by

$$K^{-1}t^\dagger t = K^{-1}x^\dagger E_Rx,$$  \hfill (B.15)

where

$$E_R \equiv T_R^\dagger T_R = \left\{ \Re \left( \frac{HH^\dagger}{N} \right) \right\}^{-1}. \hfill (B.16)$$

It is important to note that, as opposed to eq. (B.12), the matrix whose pseudoinverse is taken in eq. (B.16) becomes rank deficient only if $K > 2N$.

**Minimizing the transmitted energy**

Although channel inversion by the transmitter keeps the users from having to process any interference, it might come at the cost of a high transmission energy. The goal of the vector precoding technique is minimizing the cost of the channel inversion process, i.e. minimizing eq. (B.4). For this purpose, it is agreed between the transmitter and the users that, although there must be a minimum distance between any two symbols representing different information states, each state might be represented by more than one symbol. This gives the transmitter greater freedom to construct the information vector $x$ with symbols which faithfully represent the intended information, yet they are chosen so as to minimize eq. (B.4).

The information which the transmitter intends for user $k$ is the state $s_k$. The symbols which might represent the state $s_k$ are those contained in the set $A_{s_k}$. Then the $K$-dimensional vector $x$ can be constructed with components $\{x_1 \in A_{s_1}, \ldots, x_K \in A_{s_K}\}$, or, in short $x \in A$, where $A = A_{s_1} \times A_{s_2} \times \cdots \times A_{s_K}$. The transmitter chooses the symbol representation in $A$ which can be transmitted free of interference drawing the least energy, i.e.

$$x \equiv \arg\min_{x \in A \cap S} K^{-1}x^\dagger E\tilde{x}, \hfill (B.17)$$

where $S$ denotes the span of $E^{-1}$.

If the symbol alphabets are discrete, the complexity of (B.17) is exponential in $K$ [9]. However, if the alphabets representing the different information states are convex, then efficient algorithms might be used to find the optimal $x$ [10]. In Sections 4, 5, and 6 we introduce different convex alphabet relaxation schemes for transmitting 1, 2, or 3 uncoded bits to each user.
3 The transmitted energy: ensemble theory and the replica method

The technique outlined in Section 2 describes how to minimize the transmitted energy while achieving interference-free and individual reception by the users. In this section the expression for the transmitted energy per symbol using this transmission technique is presented. First, one might note that the transmitted energy per symbol \( E \) might be written as

\[
E \equiv \min_{x \in A \cap S} K^{-1} x^\dagger E x = \lim_{\beta \to \infty} \beta^{-1} K^{-1} \ln \sum_{x \in A \cap S} e^{-\beta x^\dagger E x}. \tag{B.18}
\]

The argument of the \( \beta \) limit in (B.18) has the same form as the expression for the Helmholtz free energy of a thermodynamic \((K \to \infty)\) system with temperature \(1/\beta\) which can exist in the states \( x \in A \cap S \) and whose energy is dictated by the interaction matrix \( E \). In the following we take advantage of this fact by taking a thermodynamic approximation, i.e. we always assume that \( K \) and \( N \) are infinitely large, yet they have a finite ratio \( \alpha = K/N \). This approximation will allow us to make use of mathematical tools imported from the statistical physics literature.

The most fundamental tool of modern statistical mechanics is ensemble theory, according to which all thermodynamic (large system) quantities can be written as an ensemble average of a suitable microscopic observable. In other words, the observable properties of thermodynamic systems become the average properties of all the microscopic states they can exist in. Some times, due to strong disorder in the system, the interactions dictating the properties of the microscopic states are random; each fixed in time, but with a random value due to the random positions of the system’s interacting elements. When the system is large, the number of random elements in the interaction matrix \( E \) is enormous. The value of the free energy (B.18) is a result of the sum of the many random interactions contained in \( E \) and (for a well-behaved statistical distribution) the sum of many random quantities can be characterized by their average value. This causes the free energy of large systems to be self-averaging over the realizations of the random interactions.

In the context of the MIMO channel described above, the statistical properties of the random interaction matrix \( E \) are dictated by the particular state the channel \( H \) is in, which dictates the system’s observable properties (such as the energy per symbol \( E \)). Hence, as the dimension of the channel increases, we may just focus on its average properties\(^*\).

\(^*\)The validity of this asymptotic approximation to describe realistic channels is ad-
As we shall see in Appendix A, a function which fully describes the eigenvalue distribution of the random matrix $E$ is its R-transform, denoted by $R_E(\cdot)$. Using the replica method of statistical mechanics and assuming the replica symmetric ansatz [12] (further discussed below), we show in Appendix B that the energy per transmitted symbol (B.18) is fully determined by the eigendistribution of $E$ (which is fully determined by the statistics of $H$), and the information symbol alphabets, as follows:

$$\mathcal{E} = \frac{\tau}{2} q \frac{d}{d\chi} R_E(-\chi), \quad (B.19)$$

where $\tau$ equals 1 when $E$ has purely real entries and it equals 2 if the entries are complex †. The parameters $q$ and $\chi$, defined as $q + \beta^{-1} \chi = 2x^\dagger x / \tau K$ for $x$ given by (B.17), are given by the following pair of coupled self-consistent equations:

$$q = \frac{2}{\tau} \sum_i P_i \int_c \left\lfloor \arg \min_{\xi \in A_i} \left| \frac{\sqrt{q R_E'(\chi)}}{R_E(\chi)} - \frac{\zeta}{\zeta} \right| \right)^2 Dz, \quad (B.20)$$

$$\chi = \frac{2}{\tau} \sum_i P_i \int_c \left\lfloor \arg \min_{\xi \in A_i} \left| \frac{\sqrt{q R_E'(\chi)}}{R_E(\chi)} - \frac{\zeta}{\zeta} \right| \right) z^* Dz, \quad (B.21)$$

where the index $i$ denotes the different information states, each of which can be represented by elements in the set $A_i$ and occurs in the components of $x$ with probability $P_i$. The expression $R_E'(\chi)$ denotes the first derivative of $R_E(t)$ evaluated at $t = -\chi$. And $Dz \equiv e^{-|z|^2} dz$.

Eqs. (B.19)-(B.21) allow us to find the value of (B.18) for any unitarily invariant hermitian random matrix $E$ with a converging eigenvalue distribution. As we shall show in Appendix A, for the special case of the matrices $E = E_C$ and $E = E_R$ defined in Section 2 the energy (B.19) may be written as

$$\mathcal{E} = \frac{q R_E'(\chi)}{a R_E^2(\chi)}, \quad (B.22)$$

and its value reduces to the solution of a single parameter ($a$) self-consistent equation:

$$\mathcal{E} = \frac{\sum_i P_i \int_c \left\lfloor \arg \min_{\xi \in A_i} \left| z \sqrt{a \mathcal{E}} - \zeta \right| \right)^2 Dz}{1 - \frac{a\tau}{2} + 2 \sqrt{\frac{a}{\tau} \sum_i P_i \int c \arg \min_{\xi \in A_i} \left| z \sqrt{a \mathcal{E}} - \zeta \right| z^* Dz}, \quad (B.23)$$

†The derivation for the special case of $\tau = 2$ may be found in [5].
An important assumption, known as the replica symmetry ansatz, was made in the derivation of eqs. (B.19)-(B.23). This assumption has been successfully applied to problems in wireless communications (see e.g. [13–16]) and coding theory (see e.g. [17, 18]). And, as we shall see in Section 5, replica symmetry mimics finite size results for the convex alphabets proposed in this work. One should be warned, however, that although replica symmetry yields asymptotically accurate results for convex alphabets, it fails to produce accurate results for alphabets relaxed onto a superlattice. While convex alphabets yield a single energy minimizing state $x$, extended lattice (or otherwise non-convex) alphabets may exhibit complex energy landscapes with many peaks and valleys. This might cause different identical copies (replicas) of the system to become trapped in different energy wells at low temperature. The replica method must then be employed invoking Parisi’s replica symmetry breaking scheme [19]. For a recent and novel analysis of lattice alphabets based on replica symmetry breaking the reader is referred to [6, 7]. For a thorough discussion of replica symmetry, the reader is referred to [20, 21].

4 Convex alphabet relaxation for binary phase shift keying (BPSK)

We consider a source of information consisting of two equiprobable states: $\uparrow$ and $\downarrow$. When no vector precoding is employed the entries in $x$ are usually selected from the unit BPSK alphabets $\mathcal{A}_\uparrow = \{+1\}$ and $\mathcal{A}_\downarrow = \{-1\}$. So long as the minimum distance between two points representing different information states is preserved, points might be added to these sets with the aim of reducing the transmitted energy.

In this section we propose two different convex alphabet extension schemes. While one of these alphabets is purely real, the other allows for complex symbols to map the information states. As outlined in Section 2 when the symbols used to map the information states are purely real then we might use two different channel inversion schemes. While one of them results in the energy metric $\mathcal{E}$, which has complex entries, the other one yields the metric $\mathcal{E}_\tau$, whose entries are purely real. The parameter $\tau$, remember from Section 3 equals 1 when the energy metric $\mathcal{E}$ contains purely real entries, or 2 when its entries are generally complex. Therefore, when the alphabet under consideration contains complex entries we will automatically set $\tau = 2$, whereas for a purely real alphabet it can take any of both values, depending on the channel inversion technique to be employed.
4. Convex alphabet relaxation for binary phase shift keying (BPSK)

**Figure B.2:** Binary real convex relaxation.

**Figure B.3:** Binary complex convex relaxation.

**Binary real convex relaxation**

A trivial way to relax BPSK onto the real line is as follows:

\[ \mathcal{A}_1 = -\mathcal{A}_1 = \{ \xi \text{ s.t. } \xi \geq 1 \}. \]  
(B.24)

This relaxed alphabet, shown in Fig. B.2, yields, after (B.23), the following expression for the energy per transmitted symbol:

\[ E = 2 + \sqrt{\frac{\alpha E}{\pi}} e^{-\frac{1}{2\alpha E}} + \{ \alpha E - 2 \} \frac{Q \left( \sqrt{2/\alpha E} \right)}{2 - \tau \alpha + 2\alpha Q \left( \sqrt{2/\alpha E} \right)}, \]  
(B.25)

where the function \( Q(\cdot) \) is defined as

\[ Q(\sigma) \equiv \frac{1}{\sqrt{2\pi}} \int_{\sigma}^{\infty} e^{-\frac{t^2}{2}} dt. \]  
(B.26)

**Binary complex convex relaxation**

The real relaxation space introduced in Section 4 might be further extended onto the complex plane as follows:

\[ \mathcal{A}_1 = -\mathcal{A}_1 = \{ \xi \in \mathbb{C} \text{ s.t. } \Re \xi \geq 1 \}. \]  
(B.27)
As the information states are now mapped onto complex symbols, we might only use the channel inversion scheme resulting in the energy metric (B.12), which has complex entries. Therefore the parameter $\tau$ will now be equal to 2. Under this precoding scheme, shown in Fig. B.3 eq. (B.23) becomes
\[
\mathcal{E} = \frac{2 + \sqrt{\frac{\alpha \mathcal{E}}{\pi}} e^{-\frac{\pi}{\alpha \mathcal{E}}} + \{\alpha \mathcal{E} - 2\} Q\left(\sqrt{2/\alpha \mathcal{E}}\right) + \alpha \mathcal{E}}{2 + 2\alpha Q\left(\sqrt{2/\alpha \mathcal{E}}\right)}. \tag{B.28}
\]
Notice that, when $\tau = 1$, eq. (B.25) reduces to (B.28).

5 Convex alphabet relaxation for quaternary phase shift keying (QPSK)

We consider now a source of information consisting of four equiprobable states: $\uparrow\uparrow$, $\uparrow\downarrow$, $\downarrow\downarrow$ and $\downarrow\uparrow$. When no vector precoding is employed the entries in $x$ are usually selected from the unit QPSK alphabets $A_{\uparrow\uparrow} = \{1 + j\}$, $A_{\uparrow\downarrow} = \{1 - j\}$, $A_{\downarrow\downarrow} = \{-1 - j\}$ and $A_{\downarrow\uparrow} = \{-1 + j\}$. Just as we did with the BPSK alphabets in Section 4, we might add points to these sets with the aim of reducing the transmitted energy.
6. **Convex alphabet relaxation for octonary modulation**

In [5] they proposed a quaternary extension of the BPSK relaxation scheme introduced in Section 4. The alphabet in Fig. B.2 can be symmetrically extended onto the complex plane as shown in Fig. B.4. Optimization is to be independently carried out on the real and imaginary dimensions. This extended quaternary alphabet is given by

\[
A_{\uparrow\uparrow} = -A_{\downarrow\downarrow} = \{ \xi \text{ s.t. } \Re \xi \geq 1 \land \Im \xi \geq 1 \}, \\
A_{\uparrow\downarrow} = -A_{\downarrow\uparrow} = \{ \xi \text{ s.t. } \Re \xi \geq 1 \land \Im \xi \leq -1 \},
\]

and the energy per transmitted symbol reduces, again, to eq. (B.28).

6 Convex alphabet relaxation for octonary modulation

In this section we consider a source of information consisting of eight equiprobable states, \(\uparrow\uparrow\uparrow, \uparrow\uparrow\downarrow, \uparrow\downarrow\downarrow, \uparrow\downarrow\uparrow, \downarrow\downarrow\uparrow, \downarrow\downarrow\downarrow, \downarrow\uparrow\downarrow, \downarrow\uparrow\uparrow\), and we propose three novel alphabet relaxations for complex octonary modulation.

**Fully symmetric octonary relaxation**

An octonary phase shift keying (8PSK) constellation typically consists of one-point alphabets given by

\[
A_{\uparrow\uparrow\uparrow} = \{ 1 + \sqrt{2} + j \}, \\
A_{\uparrow\uparrow\downarrow} = \{ 1 - (1 + \sqrt{2}) j \}, \\
A_{\uparrow\downarrow\downarrow} = \{ -1 + (1 + \sqrt{2}) j \}, \\
A_{\uparrow\downarrow\uparrow} = \{ -1 - \sqrt{2} - j \}, \\
A_{\downarrow\downarrow\uparrow} = \{ -1 - \sqrt{2} + j \}, \\
A_{\downarrow\downarrow\downarrow} = \{ -1 + \sqrt{2} - j \}, \\
A_{\downarrow\uparrow\downarrow} = \{ -1 + \sqrt{2} + j \}, \\
A_{\downarrow\uparrow\uparrow} = \{ 1 + (1 + \sqrt{2}) j \}.
\]

In order to contain the transmitted energy this alphabet might be symmetrically extended over the complex plane as shown in Fig. B.5. These relaxed symbol sets preserve the minimum distance between any two symbols representing different states, and they are given by

\[
A_{\uparrow\uparrow\uparrow} = -A_{\downarrow\downarrow\uparrow} = \{ \xi \text{ s.t. } \Re \xi \geq 1 \land \Im \xi \geq \Re \xi + \sqrt{2} \}, \\
A_{\uparrow\uparrow\downarrow} = -A_{\downarrow\downarrow\downarrow} = \{ \xi \text{ s.t. } \Re \xi \leq -1 \land \Im \xi \geq -\Re \xi + \sqrt{2} \}, \\
A_{\uparrow\downarrow\downarrow} = -A_{\downarrow\uparrow\downarrow} = \{ \xi \text{ s.t. } \Re \xi \geq 1 \land \Im \xi \leq -\Re \xi - \sqrt{2} \}, \\
A_{\uparrow\downarrow\uparrow} = -A_{\downarrow\uparrow\uparrow} = \{ \xi \text{ s.t. } \Re \xi \leq -1 \land \Im \xi \leq \Re \xi - \sqrt{2} \},
\]

Under this precoding scheme the energy per symbol (B.23) reduces to

\[
\mathcal{E} = \frac{3F_1(\mathcal{E}; \alpha) + 7F_2(\mathcal{E}; \alpha) + \frac{1}{\sqrt{2}} \left( \alpha \mathcal{E} - 4\sqrt{2} - 6 \right) F_3(\mathcal{E}; \alpha) + I_1(\mathcal{E}; \alpha) + I_2(\mathcal{E}; \alpha)}{1 - \alpha - \frac{10}{9} F_1(\mathcal{E}; \alpha) - \frac{2}{9} F_2(\mathcal{E}; \alpha) + \alpha F_3(\mathcal{E}; \alpha) + \alpha I_1(\mathcal{E}; \alpha)},
\]

(B.31)
where

\[ F_1 (\mathcal{E}; \alpha) \equiv \frac{1}{8} \sqrt{\frac{\alpha \mathcal{E}}{\pi}} e^{-\frac{1}{2\alpha \mathcal{E}}} Q \left( 2 + \sqrt{2} \sqrt{\frac{\alpha \mathcal{E}}{\pi}} \right), \quad (B.32) \]

\[ F_2 (\mathcal{E}; \alpha) \equiv \frac{1}{8} \left( 1 + \sqrt{2} \right) \sqrt{\frac{\alpha \mathcal{E}}{\pi}} e^{-\frac{3\sqrt{2}}{4\alpha \mathcal{E}}} Q \left( -\sqrt{2} \sqrt{\frac{\alpha \mathcal{E}}{\pi}} \right), \quad (B.33) \]

\[ F_3 (\mathcal{E}; \alpha) \equiv Q \left( \frac{-\sqrt{2}}{\sqrt{2} \sqrt{\alpha \mathcal{E}}} \right) Q \left( \frac{2 + \sqrt{2}}{\sqrt{2} \sqrt{\alpha \mathcal{E}}} \right), \quad (B.34) \]

\[ I_1 (\mathcal{E}; \alpha) \equiv \int_{\frac{1}{\sqrt{2\pi}}}^{\infty} \left( (1 + 2x^2) \alpha \mathcal{E} - 4 \right) Q \left( \frac{2 + \sqrt{2}}{\sqrt{2} \sqrt{\alpha \mathcal{E}}} + \sqrt{2}x \right) e^{-\frac{x^2}{4\sqrt{\pi}}} dx, \quad (B.35) \]

\[ I_2 (\mathcal{E}; \alpha) \equiv \int_{\frac{1}{\sqrt{2\pi}}}^{\infty} \left( (1 + 2x^2) \alpha \mathcal{E} - 12 - 8\sqrt{2} \right) Q \left( \frac{2 + 2\sqrt{2}}{\sqrt{2} \sqrt{\alpha \mathcal{E}}} - \sqrt{2}x \right) e^{-\frac{x^2}{4\sqrt{\pi}}} dx, \quad (B.36) \]
6. CONVEX ALPHABET RELAXATION FOR OCTONARY MODULATION

\[ I_3 (E; a) \equiv \int_{\frac{1}{\sqrt{a}}}^{\infty} (1 + 2x^2) \left( Q \left( \frac{2}{\sqrt{a} E} + \sqrt{2} x \right) + Q \left( \frac{2 + 2\sqrt{2}}{\sqrt{a} E} - \sqrt{2} x \right) \right) e^{-x^2} dx. \] (B.37)

Octonary stellar relaxation

We propose an octonary constellation whose original one-point alphabets are given by

\[ A_{\uparrow\uparrow\uparrow} = \{2 + 2j\}, \quad A_{\uparrow\uparrow\downarrow} = \{\sqrt{2}\}, \quad A_{\uparrow\downarrow\downarrow} = \{2 - 2j\}, \quad A_{\uparrow\downarrow\uparrow} = \{-\sqrt{2}\}, \quad A_{\downarrow\uparrow\uparrow} = \{-2 - 2j\}, \quad A_{\downarrow\downarrow\uparrow} = \{-\sqrt{2}\}, \quad A_{\downarrow\downarrow\downarrow} = \{-2 + 2j\} \] and \[ A_{\downarrow\uparrow\downarrow} = \{\sqrt{2}\}. \] In order to contain the transmission energy while preserving the minimum distance between any two symbols representing different information states, this alphabet is relaxed as shown in Fig. B.6. The extended symbol sets are given by

\[ A_{\uparrow\uparrow\uparrow} = -A_{\downarrow\uparrow\downarrow} = \{\xi \text{ s.t. } \Re \xi \geq 2 \& \Im \xi \geq 2\}, \]
\[ A_{\uparrow\uparrow\downarrow} = -A_{\downarrow\downarrow\downarrow} = \{\xi \text{ s.t. } \Re \xi \geq \sqrt{2} \& \Im \xi = 0\}, \]
\[ A_{\uparrow\downarrow\downarrow} = -A_{\downarrow\uparrow\downarrow} = \{\xi \text{ s.t. } \Re \xi \geq 2 \& \Im \xi \leq -2\}, \]
\[ A_{\uparrow\downarrow\uparrow} = -A_{\downarrow\downarrow\uparrow} = \{\xi \text{ s.t. } \Re \xi = 0 \& \Im \xi \leq -\sqrt{2}\}. \] (B.38)

and the energy (B.23) becomes

\[ E = \frac{10 + \{\frac{1}{2} a E - 2\} Q (2/\sqrt{a E}) + \{a E - 8\} Q (2\sqrt{2}/a E) + \frac{1}{2} \sqrt{\frac{E}{\pi}} \left\{ 4e^{-\frac{E}{2}} + \sqrt{2} e^{-\frac{E}{2}} \right\}} {2 - 2a + aQ (2/\sqrt{a E}) + 2aQ (2\sqrt{2}/a E)}. \] (B.39)

Octonary triangular relaxation

An alternative octonary constellation is proposed with unperturbed one-point alphabets given by

\[ A_{\uparrow\uparrow\uparrow} = \{2 + \sqrt{3}j\}, A_{\uparrow\uparrow\downarrow} = \{1\}, A_{\uparrow\downarrow\downarrow} = \{2 - \sqrt{3}j\}, \]
\[ A_{\uparrow\downarrow\uparrow} = \{-\sqrt{3}j\}, A_{\downarrow\uparrow\uparrow} = \{-2 - \sqrt{3}j\}, A_{\downarrow\downarrow\downarrow} = \{-1\}, A_{\downarrow\uparrow\downarrow} = \{-2 + \sqrt{3}j\} \] and \[ A_{\downarrow\downarrow\uparrow} = \{\sqrt{3}j\}. \] In order to contain the transmission energy and preserve the minimum distance properties of this triangular constellation, only six of the eight points are allowed to relax as shown in Fig. B.7. The fact
B. ON CONVEX VECTOR PRECODING FOR MULTIUSER MIMO BROADCAST CHANNELS

**Figure B.6:** Octonary stellar relaxation.

**Figure B.7:** Octonary triangular relaxation.
that the two innermost points of the constellation do not undergo any re-
laxation makes this convex precoding scheme easier to implement. The
relaxed symbol sets are given by
\[
A_{\uparrow\uparrow\uparrow} = -A_{\downarrow\downarrow\uparrow} = \left\{ \xi \text{ s.t. } \Re \xi \geq 2 \& \Im \xi \geq \sqrt{3} \right\},
\]
\[
A_{\uparrow\uparrow\downarrow} = -A_{\downarrow\downarrow\downarrow} = \left\{ 1 \right\},
\]
\[
A_{\uparrow\downarrow\downarrow} = -A_{\downarrow\uparrow\downarrow} = \left\{ \xi \text{ s.t. } \Re \xi \geq 2 \& \Im \xi \leq -\sqrt{3} \right\},
\]
\[
A_{\uparrow\downarrow\uparrow} = -A_{\downarrow\uparrow\uparrow} = \left\{ \xi \text{ s.t. } \Re \xi = 0 \& \Im \xi \leq -\sqrt{3} \right\},
\]
and the energy (B.40) reduces to
\[
\mathcal{E} = \frac{18 + \left\{ \alpha \mathcal{E} - 8 \right\} Q \left( 2\sqrt{2/\alpha \mathcal{E}} \right) + \left\{ \frac{3}{2} \alpha \mathcal{E} - 9 \right\} Q \left( \sqrt{6/\alpha \mathcal{E}} \right) + \frac{1}{2} \sqrt{\frac{3 \mathcal{E}}{\pi}} \left\{ 4 e^{-\frac{1}{2}} + 3 \sqrt{3} e^{-\frac{3}{2}} \right\}}{4 - 4 \alpha + 2 \alpha Q \left( 2\sqrt{2/\alpha \mathcal{E}} \right) + 3 \alpha Q \left( \sqrt{6/\alpha \mathcal{E}} \right)}.
\]

7 Interference-Free Transmission over Singular Channels

As discussed in Section 2 a vector \( x \) can be transmitted free of interfer-
ence as long as it lies in the span of \( \mathbf{E}^{-1} \). Because the relaxation regions \( \mathcal{A} = \prod_{i=1}^{K} \mathcal{A}_{ni} \) proposed in the previous sections are convex, it is conceiv-
able that the components of a \( K \)-dimensional vector \( x \in \mathcal{A} \) can be freely
tuned in the two-dimensional relaxation regions until the vector is com-
pletely contained in the \( N \)-dimensional span of \( \mathbf{E}^{-1} \). As we shall see, this
possibility of overloading is a distinct advantage that convex relaxation
schemes offer compared to lattice based schemes such as those proposed
in [1, 3, 22].

Whether or not interference-free transmission is possible when \( K > N \)
boils down to the probability that the precoding space \( \mathcal{A} \) and the span of
the matrix \( \mathbf{E}^{-1} \) intersect. To start our analysis we quote the following result
by Wendel [23]: the probability that the span of a random \( K \times N \) matrix
with i.i.d. real Gaussian entries contains a \( K \)-dimensional vector with all
components strictly positive is given by
\[
w(K, N) = 2^{1-K} \sum_{\ell=0}^{N-1} \binom{K - 1}{\ell}.
\]
This result might be extended to show that, if the random entries in the
\( K \times N \) matrix are complex Gaussian, then the probability that its span con-
tains a $K$-dimensional complex vector with all components in the first quadrant is given by $w(2K, 2N)$ (see [24]).

Using the complex extension of Wendel’s result and noting that the span of $\mathbf{HH}^\dagger$ is the same as that of $\mathbf{H}$, in [25] we showed that $w(2K, 2N)$ is indeed the probability that a vector relaxed in the quaternary space shown in Fig. B.4 can be found in the span of $\mathbf{HH}^\dagger$. This is due to the symmetry of the alphabet, the rotational invariance of the span of $\mathbf{HH}^\dagger$, and the radial invariance of its spanning vectors. Then, the probability $P_Q$ that a quaternary vector relaxed as in Fig. B.4 is transmitted free of interference is

$$P_Q(K, N) = w(2K, 2N).$$  \hspace{1cm} (B.43)

The question now is whether the analysis can be extended further and similar results can be found for the rest of the relaxed alphabets proposed in this paper. Before proceeding, and to get an understanding, let us start by observing that in both the real and complex versions of Wendel’s result, the first argument in the function $f$ corresponds to the number of sign-constrained degrees of freedom the vector exhibits. The second argument indicates the number of dimensions available in the span of the matrix to accommodate such vector.

We shall start by analyzing the probability of interference-free transmission of a $K$-dimensional vector $\mathbf{x}$ using the real binary alphabet shown in Fig. B.2. Due to the symmetry of the alphabet as well as the unitary invariance of $\mathbf{HH}^\dagger$ and the radial invariance of its spanning vectors, we may just consider the probability that a vector with all components strictly positive can be found in the span of $\mathbf{E}^{-1}$. The number of sign-constrained degrees of freedom in such a vector is $K$. However, symbols in this alphabet might be transmitted using either $T_R$ or $T_C$ with resulting energy metrics $E_R$ or $E_C$, respectively. These two energy metrics have different inverses with different spans, and so they should be treated separately. Because $E_R^{-1} = \mathbf{H}_R \mathbf{H}_R^\dagger + \mathbf{H}_I \mathbf{H}_I^\dagger$, when $T_R$ is used the number of degrees of freedom in the span is $2N$ ($\mathbf{H}_R$ and $\mathbf{H}_I$ provide $N$ real dimensions each). Then the probability $P_{B1R}$ that a binary vector relaxed in one dimension as shown in Fig. B.2 can be transmitted free of interference using $T_R$ is

$$P_{B1R}(K, N) = w(K, 2N).$$  \hspace{1cm} (B.44)

However, when $T_C$ is used, although the span of $\mathbf{HH}^\dagger$ offers in principle $N$ complex dimensions, these complex degrees of freedom must be constrained down to accommodate $K$ purely real constraints. Then the total number of real degrees of freedom in the span goes down from $2N$...
to $2N - K$. The probability $P_{B1C}$ that a binary vector relaxed as shown in Fig. B.2 can be transmitted free of interference using $T_C$ is then

$$P_{B1C} (K, N) = w(K, 2N - K).$$ (B.45)

We now turn our attention to the two-dimensional binary convex relaxation in Fig. B.3. Given that the imaginary part is completely unconstrained, the number of sign-constrained degrees of freedom in $x$ is, as before, $K$. However, precisely because the symbols in $x$ need not be one-dimensional the $2N$ real degrees of freedom available in the complex span of $H$ (i.e. the span of $E_C^{-1}$) are unconstrained. Therefore, the probability $P_{B2}$ that a binary vector relaxed as shown in Fig. B.3 can be transmitted free of interference is

$$P_{B2} (K, N) = w(K, 2N).$$ (B.46)

In the case of the octonary stellar relaxation shown in Fig. B.6 each component in $x$ can be one or two-dimensional with probability $1/2$. Each one-dimensional component has one sign-constrained degree of freedom, and each two-dimensional component has two such degrees of freedom. How many there are of each will condition not only how many constraints there are in $x$, but also how constrained the available degrees of freedom in the span of $H$ are. While two-dimensional components don’t collapse the rank of the complex span, each one-dimensional component takes away one degree of freedom. Then we might use combinatorics to obtain the probability $P_{OS}$ that an octonary vector relaxed as shown in Fig. B.6 can be transmitted free of interference:

$$P_{OS} (K, N) = \frac{2^K}{2^K - 1} \sum_{q=1}^{K} \binom{K}{q} w \left( \frac{q}{K} 2K + \frac{K - q}{K} K, 2N - \frac{K - q}{K} K \right).$$ (B.47)

Using the octonary triangular relaxation shown in Fig. B.7 each component in $x$ can be two-dimensional with probability $1/2$, one-dimensional with probability $1/4$, or a point with probability $1/4$. As points have no freedom in any direction, point components in $x$ will each collapse two degrees of freedom from the rank of $H$. Treating the one and two-dimensional regions as before, we might obtain the probability $P_{OT}$ that an octonary vec-
Figure B.8: Probability that convex precoding fails to transmit free of interference vs. the ratio of the number of users to the number of transmitting antennas.

When it comes to the fully symmetric octonary relaxation space shown in Fig. B.5, a similar line of analysis might not be employed due to the non-rectangularity of the precoding regions. Finding the probability that the span of $\mathbf{H}$ contains a vector relaxed in this space is, to our knowledge, still an open problem.

8 Results

In Fig. B.8 we can see a plot of the probability of failing to achieve interference-free transmission. As we can see, depending on the precoding scheme chosen, significant overloads are possible while keeping a relatively small probability of failure, which is perhaps tolerable in systems with subsequent error control coding. As the size of the system increases, the probabilities (B.43)-(B.48) tend to become unit step functions of the channel load $K/N$ such that interference-free transmission is always possible below a certain threshold value.

As the binary relaxation shown in Fig. B.2 is purely real we might choose to precode it with either $\mathbf{T}_C$ or $\mathbf{T}_R$. When $\mathbf{T}_C$ is used, the channel is fully invertible only up to 1 bit/antenna; however, when $\mathbf{T}_R$ is employed, the
channel is fully invertible up to 2 bits/antenna. Also, in Fig. B.8 we can see that precoding with \( T_R \) allows for significant overloads while keeping the probability of failure relatively small.

Precoding the real binary relaxation with \( T_R \) also results in lower transmission energies than when \( T_C \) is used. Indeed the use of \( T_R \) makes optimization over the purely real space perform just as well as over the complex extension shown in Fig. B.3 rendering the complexity brought by the search in the additional dimension useless. Furthermore, precoding with \( T_R \) makes the one-dimensional relaxation for BPSK achieve spectral efficiencies at the same cost as the QPSK relaxation shown in Fig. B.4 this means that transmitting 1 bit per user to \( 2K \) users has the same cost as transmitting 2 bits per user to \( K \) users. These results are shown in Fig. B.9 where the energy per transmitted bit is plotted vs. the uncoded spectral efficiency \( \epsilon \), which is defined as the number of bits per symbol multiplied by the channel load \( K/N \). Although our analysis pertains to the asymptotic limit, finite size simulations show that these results can be used to approximate finite systems with 30 transmitters.

When it comes to using the octonary modulation schemes proposed in Section 6, depending on the ratio of transmitting to receiving elements, it might be more advantageous to use the stellar modulation in Section 6 (S8PSK), or the triangular modulation in Section 6 (T8PSK). The main ad-
vantage that T8PSK presents is that only 6 of the 8 points in the unperturbed constellation are relaxed, which reduces the complexity of the optimization process. Fig. B.10 shows the energy per transmitted bit vs. $\epsilon$ attained for all three octonary modulation schemes proposed in Section 6. In the region below 3 bits per transmitting antenna, where all three schemes guarantee interference-free transmission, T8PSK is most appealing given both its lower energy and lower complexity at the implementation stage. As the system gets overloaded T8PSK is outperformed by S8PSK not only in terms of energy penalty, but also, as shown in Fig. B.8 in terms of robustness to achieve interference-free transmission.

Fig. B.11 allows us to compare T8PSK relaxation with the binary and quaternary relaxations shown in Figs. B.2 and B.4. Uncoded spectral efficiencies of up to $\sim 1.7$ bits/antenna are attained at a lower cost using either the binary or the quaternary scheme. Depending on the chosen precoding scheme, in the vicinity of 1.7 bits/antenna it costs the same to transmit 1 bit per user to $K$ users, 2 bits per user to $K/2$ users, or 3 bits per user to $K/3$ users.
9. Appendices

A. The R-transform

Let $P_M(x)$ denote the eigenvalue distribution of the matrix $M$. Let

$$m_M(s) = \int \frac{dP_M(x)}{x-s}, \quad (B.49)$$

which is known as the Stieltjes transform. Then, the R-transform of $P_M(x)$ is

$$R_M(w) = m_M^{-1}(-w) - \frac{1}{w}, \quad (B.50)$$

with $m^{-1}(\cdot)$ denoting the inverse function of $m(\cdot)$.

It can be verified that

$$m_M^{-1}\left(\frac{1}{s}\right) = -s \left(1 + s m_M(s)\right). \quad (B.51)$$

Let $s = m_M^{-1}(-w)$. Then, we find

$$m_M^{-1}\left(\frac{1}{m_M^{-1}(-w)}\right) = -m_M^{-1}(-w) \left(1 - w m_M^{-1}(-w)\right), \quad (B.52)$$

and

$$\frac{1}{m_M^{-1}(-w)} = m_M^{-1}\left(-m_M^{-1}(-w) \left(1 - w m_M^{-1}(-w)\right)\right). \quad (B.53)$$
With eq. (B.50), we find
\[
\frac{1}{R_M(w) + \frac{1}{w}} = R_{M^{-1}} \left( -wR_M(w) \left( R_M(w) + \frac{1}{w} \right) \right) - \frac{1}{wR_M(w) \left( R_M(w) + \frac{1}{w} \right)},
\]
and
\[
\frac{1}{R_M(w)} = R_{M^{-1}} \left( -R_M(w) \left( 1 + wR_M(w) \right) \right).
\] (B.55)

It is well known \cite{26, 27} that the R-transform of the limiting spectral measure \( P_{E_{C^{-1}}} \) is given by
\[
R_{E_{C^{-1}}}(w) = \frac{1}{1 - \alpha w}.
\] (B.56)
Letting \( M^{-1} = E_{C^{-1}} \), we find form (B.55)
\[
R_{E_{C}}(w) = 1 + \alpha R_{E_{C}}(w) \left( 1 + wR_{E_{C}}(w) \right).
\] (B.57)
Solving (B.57) for the R-transform yields
\[
R_{E_{C}}(w) = f_2(w),
\] (B.58)
where \( f_\tau(w) \) is defined as
\[
f_\tau(w) \equiv \frac{2}{\tau} - \alpha - \sqrt{\left( \frac{2}{\tau} - \alpha \right)^2 - \frac{8}{\tau} \alpha w}.
\] (B.59)

Now, in order to find the R-transform of the spectral measure of \( E_R \) first we take notice that the matrix \( E_{R^{-1}} \) can be written as follows:
\[
E_{R^{-1}} = \frac{H_R H_R^T + H_I H_I^T}{N}.
\] (B.60)
Then we note that the R-transform has the following two properties \cite{26, 27}:
\[
R_{M_1 + M_2}(w) = R_{M_1}(w) + R_{M_2}(w),
\] (B.61)
\[
R_{cM}(w) = cR_M(cw) \quad \forall c \in \mathbb{C}.
\] (B.62)
Using these two properties we might write the R-transform of (B.60) as
\[
R_{E_{R^{-1}}}(w) = \frac{1}{2} R_{H_R H_R^T} \left( \frac{w}{2} \right) + \frac{1}{2} R_{H_I H_I^T} \left( \frac{w}{2} \right).
\] (B.63)
Now recall from Section 2 that $H_r$ and $H_i$ are real random matrices containing independent and identically distributed entries with zero mean and variance $1/2$. Then both $\frac{H_rH_r^T}{N/2}$ and $\frac{H_iH_i^T}{N/2}$ have the same R-transform, which is given by (B.56), implying that

$$R_{E_R}(w) = \frac{1}{1 - \alpha w^2},$$

(B.64)

which in conjunction with (B.55) yields that

$$R_{E_E}(w) = f_1(w).$$

(B.65)

It is now clear by simple substitution how eqs. (B.22) and (B.23) result from (B.19) for $E = E_C$ and $E = E_R$:

$$E = \frac{\tau}{2} \frac{d}{d\chi} \chi R_E(-\chi) = \frac{q_R E(-\chi)}{\alpha R_E^2(-\chi)} = \frac{q}{2 - \alpha + 2\alpha R_E(-\chi)}.$$

(B.66)

### B. The Replica Method

In this appendix we derive eqs. (B.19)-(B.21) which are used to obtain the transmitted energy $E$ per symbol. The derivation follows along the lines of Appendix A in [5], but it is extended to allow for purely real matrices such as $E_R$.

Let us start by recalling eq. (B.18):

$$E = -\lim_{\beta \to \infty} \lim_{K \to \infty} \beta^{-1} K^{-1} \ln \sum_{x \in A \cap S} e^{-\beta x^T E x}.$$  

(B.67)

As indicated in Section 3, the expression in the $\beta$ limit is a thermodynamic free energy (per microscopic degree of freedom in $x$) when $K \to \infty$, so, in order to use tools from statistical mechanics and treat the argument in the $\beta$ limit as a thermodynamic free energy we should write

$$E = -\lim_{\beta \to \infty} \lim_{K \to \infty} \beta^{-1} K^{-1} \ln \sum_{x \in A \cap S} e^{-\beta x^T E x}.$$  

(B.68)

As discussed in Section 3, the free energy is self-averaging in the thermodynamic limit. Furthermore, the eigenproperties of large random matrices, such as $E_C$ and $E_R$, are also postulated to be self-averaging [28]. Hence, we might rewrite (B.18) as follows:

$$E = -\lim_{\beta \to \infty} \frac{1}{K \to \infty} \beta^{-1} K^{-1} \ln Z.$$  

(B.69)
where the overbar denotes the configurational average with respect to the matrix \( \mathbf{E} \) (which results from the channel states \( \mathbf{H} \)), and
\[
\mathcal{Z} \equiv \overline{\sum_{x \in \mathcal{X} \cap \mathcal{S}_H} e^{-\beta x^\dagger \mathbf{E} x}}. \tag{B.70}
\]
is referred to as the partition function.

Finding the average of the logarithm of a partition function is far from trivial. In order to tackle the problem we employ the somewhat mystifying equality [29, sec. 6.8]
\[
\ln U = \lim_{x \to 0} \frac{c U x - c x}{x}, \tag{B.71}
\]
which for convenience we write as
\[
\ln U = \left. \frac{\partial}{\partial n} \ln U^n \right|_{n=0}, \tag{B.72}
\]
where \( c \) is just some constant upon which the probability distribution has no effect.

We might use (B.72) to rewrite the average in (B.69) as
\[
\lim_{K \to \infty} \frac{1}{K} \ln \mathcal{Z} = \lim_{n \to 0} \frac{\partial}{\partial n} \lim_{K \to \infty} \frac{1}{K} \ln \mathcal{Z}^n, \tag{B.73}
\]
which, moving things around, allows us to write (B.69) as
\[
\mathcal{E} = -\lim_{\beta \to \infty} \frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} \lim_{K \to \infty} K^{-1} \ln \mathcal{Z}^n. \tag{B.74}
\]

The task is now reformulated in terms of finding the average of a power of the partition function. This is also far from trivial, unless the power is an integer, in which case we are just dealing with moments of \( \mathcal{Z} \). The parameter \( n \) in (B.74) is, however, a positive real number. In the following we make the key assumption that, although we take \( n \) to be a positive integer, we can nevertheless continually extend the result of the average to \( n = 0 \). This critical assumption together with expression (B.71), is known as the replica trick, and it is the cornerstone of the replica method, which was developed in statistical physics to analyze magnetic glassy systems. The name replica refers to the fact that we are now taking the expectation of the product of \( n \) identical replicas of \( \mathcal{Z} \). We might further rewrite (B.74) as
\[
\mathcal{E} = -\lim_{\beta \to \infty} \frac{1}{\beta} \lim_{n \to 0} \frac{\partial}{\partial n} \mathcal{Z}^n. \tag{B.75}
\]
where

\[
\Xi_n = \lim_{K \to \infty} \frac{1}{K} \ln \prod_{a=1}^{N} Z_a = \lim_{K \to \infty} \frac{1}{K} \ln \sum_{\{x_a \in A \cap S_H\}} \exp \left[ \operatorname{tr} \left( -\beta E \sum_{a=1}^{n} x_a x_a^\dagger \right) \right],
\]  

(B.76)

and the summation notation \(\sum_{\{x_a \in A \cap S_H\}}\) stands for \(\sum_{x_1 \in A \cap S_H} \sum_{x_2 \in A \cap S_H} \cdots \sum_{x_n \in A \cap S_H}\).

When the random matrix \(E\) is Hermitian and unitarily invariant the expectation value in (B.76) can be written in terms of the R-transform \(R_E(w)\) of the eigenvalue distribution of \(E\) as follows [30]:

\[
\Xi_n = \lim_{K \to \infty} \frac{1}{K} \ln \sum_{\{x_a \in A \cap S_H\}} \exp \left[ -K \sum_{a=1}^{n} \int_{0}^{2\lambda_a} R_E(-w) \, dw \right],
\]  

(B.77)

with \(\tau = 1\) when all the entries in \(E\) are purely real (or purely imaginary) and \(\tau = 2\) when the entries in \(E\) are neither purely real nor purely imaginary, and with \(\lambda_1, \ldots, \lambda_n\) being the eigenvalues of the \(n \times n\) dimensional matrix \(\beta Q\) whose elements are \(Q_{ab} = K^{-1} x_a^\dagger x_b\).

From this point we follow along the lines of Appendix A in [5] and obtain expressions (B.19)-(B.21).
References


through correlated channels in the presence of correlated interfer-
ers and noise: A (not so) large n analysis,” IEEE Trans. Inf. Theory,

ration of detection and decoding,” IEEE Trans. Inf. Theory, vol. 50,

2005.


[21] H. Nishimori, Statistical Physics of Spin Glasses and Information Process-
ing (International Series of Monographs on Physics, 111). Oxford Sci-


Paper C

Information and Multiaccess Interference in a Complexity-Constrained Vector Channel

Rodrigo de Miguel, Ori Shental, Ralf R. Müller and Ido Kanter

Abstract

A noisy vector channel operating under a strict complexity constraint at the receiver is introduced. According to this constraint, detected bits, obtained by performing hard decisions directly on the channel’s matched filter output, must be the same as the transmitted binary inputs. An asymptotic analysis is carried out using mathematical tools imported from the study of neural networks, and it is shown that, under a bounded noise assumption, such complexity-constrained channel exhibits a non-trivial Shannon-theoretic capacity. It is found that performance relies on rigorous interference-based multiuser cooperation at the transmitter, and that this cooperation is best served when all transmitters use the same amplitude.
1 Introduction

Multiple-access systems described by vector channels (MIMO, CDMA) are prominent in modern wireless communications. As such, revealing their information-theoretic properties has been a fruitful source of ongoing research (see, e.g., [1, 2]). In these systems, information is conveyed simultaneously from a group of sources to another group of sinks over the same physical medium and bandwidth. These transmissions are not orthogonal and interfere with each other. Detrimental effects such as multiaccess interference and noise can be completely eliminated in theory by adopting optimal detection schemes and sophisticated error-correcting codes [3].

As in many problems of reliable (i.e., errorfree) communication via a detrimental channel, typical information-theoretic investigations of multiaccess systems are often carried out under an upper bounded transmission power or limited bandwidth, but usually no restrictions on complexity are imposed. In the era of pervasive and ubiquitous wireless communication systems, involving only limited computing power, there is an emerging interest in the workings of a stricter complexity-constrained scenario [4].

In this contribution such a reduced complexity setting is introduced and analyzed. This reduced complexity scenario requires that detected bits, sliced at the output of a bank of filters matched to the vector channel, be the same as the transmitted binary inputs. This constraint is analogous to the constraint on single-neuron (or spin) flip metastable states of the Hopfield model for neural networks [5, 6]. We term this scheme Vector Input Preselection (VIP), as transmitted bit combinations are preselected so as to comply with this strict constraint. Such a transmission scheme would result in the appealing use of low-cost receivers, in the context of reduced signal processing and computing. In a sense, this scheme is equivalent to outsourcing part of the detection complexity back to the transmitters. However, although the receivers need only do minimal signal processing, communication does not conclude until they properly interpret the received bits. This can be done using a decoding table identical to the transmitters’ joint coding table. Although the complexity of this task cannot be outsourced back to the transmitters, it is solved by proper memory allocation [3].

The Hopfield model has been used before for developing suboptimal multiuser detectors [7-9]. In this contribution, the Hopfield model is used to compute the Shannon capacity of the noisy VIP channel in the many-users limit, revealing the cost in “information rate” caused by limiting the channel complexity. Interestingly, we find that contrary to intuition, the VIP channel yields a non-trivial capacity.

We analyze the effect that the power distribution among the transmitt-
C. INFORMATION AND MULTIACCESS INTERFERENCE IN A COMPLEXITY-CONSTRAINED VECTOR CHANNEL

ters has on the channel capacity. Although equal power interference is found to be the worst case interference scenario for linear multiuser receivers [2] we find that, for the VIP channel, optimal cooperative interference occurs when all transmitters use equal amplitudes. Indeed we find that cooperation through interference is a key feature of the VIP channel, which is in contrast to systems without input preselection, where higher channel capacity is obtained by interference mitigation at the receiver [10, 2]. It is important to note that the VIP rule employed in this paper is conceptually different from the known technique of multiuser precoding (or pre-equalization) [12–14] in which there is no restriction on the input signaling.

The organization of the present paper is as follows. The basic vector channel model is outlined in Section 2. The asymptotic system analysis is briefly introduced in Section 3. In Section 4 the expression for the asymptotic information capacity of the channel is presented. The results are shown in Section 5 and are further discussed in Section 6. Appendices I and II provide full mathematical details.

2 Channel Model

Consider a synchronous noisy vector channel with $T$ transmitters and $N$ receivers. The channel is characterized by a $N \times T$ random matrix $S$ with independent identically distributed entries with unit variance and zero mean. The matrix $S$ is assumed to be perfectly known at both ends of the channel. The input vector consists of $T$ binary entries which are subject to transformation by the channel matrix $S$. The $N$-length received column vector is projected back into the $T$-dimensional information space by passing through a bank of filters matched to the channel matrix $S$.

Thus overall, an input column vector $i \triangleq (i_1, \ldots, i_T)$ is transformed into an output $o \triangleq (1, \ldots, T)$ as follows:

$$o = \frac{1}{N} S^t SAi + n,$$

where the diagonal matrix $A \triangleq \text{diag}\{a_1, \ldots, a_T\}$ controls the transmission power. The vector $n$ is a $T$-dimensional random vector with arbitrary additive noise components bounded to the interval $(-\kappa, \kappa)$, where the threshold $\kappa$ is a non-negative constant.
Vector Input Preselection

Consider a potential energy surface in $\mathbb{R}^T$ where each $i \in \{-1, 1\}^T$ has its own well of attraction. When an input vector $i$ is perturbed by the channel operator as in (C.1), the resulting state may remain within the initial well or jump to another well, (or to a local maximum between several wells). In the model under consideration, the vector $\varnothing \in \mathbb{R}^T$ is within the well of attraction of $i$ when $\text{sgn}(\varnothing) = i$.

To ease the task of detection at the receiver, the transmitter, having perfect channel state information, carries out a vector input preselection (VIP) process. To meet the VIP rule an input vector $i$ must be basin invariant under channel perturbation (C.1) or, equivalently, satisfy the following equality:

$$i \equiv \text{sgn}\left\{\frac{1}{N}S^\dagger S Ai + n\right\},$$

where the equality and sign function apply to each of the $T$ vector components. Although the VIP process restricts the vectors eligible for transmission to a subset of the $2^T$ possible binary vectors, the receiver can be certain that reception is errorfree by a trivial use of conventional demodulation

$$i = \text{sgn}(\varnothing),$$

and without the use of complicated signal processing. The input vectors which satisfy (C.2) are, in the noiseless limit, equivalent to the metastable states of the Hopfield model.

The VIP rule is effectively outsourcing the complexity of detection from the receiver to the transmitter, which must coordinate its multiple inputs, thus limiting its choice of symbols for transmission. Fig. C.1 shows a diagram of the multiaccess VIP channel, which may describe, for example, a cooperative MIMO channel. An important feature in this scheme of simple detection by input preselection is that throughput bounds are invariant to the exact noise probability distribution, depending only on the noise upper and lower bounds $(−\kappa, \kappa)$.

3 Channel States and the Thermodynamic Limit

Equation (C.2), which represents the condition that a vector $i$ must satisfy in order to remain within its basin of attraction, may be rewritten as follows:

$$\int_{((\kappa-1)\beta^{-1},\kappa)^T} d^T\lambda \delta\left\{T^{-1}S^\dagger S Ai - \left(\beta^{-1}I + \Lambda\right)i\right\} = 1,$$
where $\beta \equiv T/N$, $I$ is the identity matrix, and the $T$ variables of integration are the eigenvalues of the diagonal matrix $\Lambda$. The total number $N$ of input symbols that obey the VIP rule is thus given by

$$N(S; T, \beta, A, \kappa) = \sum_i \int_{((\kappa-1)\beta^{-1}, \infty)^T} d^T \lambda \delta \left\{ T^{-1}S^\dagger SAi - \left( \beta^{-1}I + \Lambda \right)i \right\},$$

where the sum is taken over all $2^T$ possible input vectors that may, in principle, be generated at the transmitter.

The channel matrix $S$ can exist in many different quenched states. We may consider, for simplicity and illustrative purposes, a channel matrix consisting of independent identically distributed binary entries. While microscopically we specify the state of the channel as one of $2^{NT}$ equiprobable states $S$ (henceforth microstates), macroscopically the behavior, or macrostate, of the channel may be expressed as the number of inputs it allows to fulfill the VIP rule. Although the specific set of such inputs depends on the channel microstate, it is only the cardinality of the set that determines a macrostate. Therefore, more than one microstate may correspond to the same macrostate. Let $\Omega(N; T, \beta, A, \kappa)$ be the number of channel microstates that allow exactly and no more than $N$ inputs to satisfy the VIP rule, that is the number of microstates that correspond to macrostate-$N$ for $T$ transmitters, a channel load $\beta$, an amplitude distribution $A$, and a noise bound $\kappa$. There are a total of $2^{NT}$ possible microstates, and $2^T$ possible macrostates,
3. Channel States and the Thermodynamic Limit

Figure C.2: Macrostates-microstates diagram for a noiseless channel represented by a binary matrix, with four transmitters and four receivers. The transmitters all employ equal amplitudes. The most likely macrostate is that which allows six of the sixteen possible inputs to fulfill the VIP constraint.

which entails

\[ \sum_{N=1}^{2^T} \Omega(N; T, \beta, A, \kappa) = 2^{NT}. \quad (C.6) \]

Fig. C.2 shows a macrostates-microstates diagram for the case of a noiseless channel with \( N = T = 4 \) and equal amplitudes. Because all of the microstates are equiprobable we may directly convert the number of microstates into probability dividing by \( 2^{NT} \). For the case shown in Fig. C.2 the most likely macrostate (that is, the one realized by more microstates than any other) is that which allows six of the sixteen possible inputs to fulfill the VIP rule.

The so called basic postulate of statistical mechanics is the equiprobability of microstates in a closed system, and central to ensemble theory is the assumption that all thermodynamic quantities can be written as an ensemble average of a suitable microscopic observable \[15\]. We anticipate that \( N \) grows exponentially with \( T \), and therefore the typical number \( N \) of VIP vectors in the large system is given by the extensive average \( \exp \ln N \), where the overbar denotes the configurational average with respect to all channel microstates \( S \). In the following, the assumption is made that as the dimensions of the system go to infinity at constant ratio
(T, N → ∞; T/N = β), the microstate (probability) distribution converges to its own average and fluctuations away from the most likely macrostate vanish*. This entails that the distribution of N is largely peaked around \( \mathbb{N} \), which allows us to simplify the problem by taking the so called annealed approximation and directly performing the average of N over the quenched channel states S. This assumption is not uncommon in the study of various spin and neural systems exhibiting quenched interactions (see e.g. [5, 6, 16–22]). The annealed approximation is for now accepted, and its validity will be further addressed in Section 6, where it will be discussed and tested against finite size simulations. The self-averaging assumption allows us to write

\[
N(S; T, \beta, A, \kappa) \xrightarrow{T \to \infty} N(T, \beta, A, \kappa) \equiv \mathcal{N}(S; T, \beta, A, \kappa),
\]

where \( \mathcal{N} \) is (an upper bound on) the typical number of valid inputs. \( \mathcal{N} \) may be rewritten as follows:

\[
\mathcal{N}(T, \beta, A, \kappa) = \sum_i \int d^T \lambda \delta \{ T^{-1} S^i S A i - (\beta^{-1} I + \Lambda) i \}. \quad (C.8)
\]

4 Information Capacity

As discussed in Section 3 it is assumed that all channel states will exhibit similar behavior as the channel dimensions become larger. This behavior is characterized by the number \( \mathcal{N} \) of inputs that meet the VIP criterion (C.2). The number of VIP symbols in the large-system limit may be found using the asymptotic mathematical trickery shown in Appendix I (subsection A); the resulting expression is

\[
\mathcal{N}(T, \beta, A, \kappa) = \frac{2 T^2}{4 \pi^2 \beta^2} \exp \left\{ \frac{T}{\beta} g(\beta, A, \kappa) \right\}, \quad (C.9)
\]

where

\[
g(\beta, A, \kappa) = \chi + \frac{1}{2} + \frac{\chi^2 U}{2 \phi} + \frac{1}{2} \ln \frac{\phi}{U} + \beta \int_0^\infty dU \psi(U) \ln Q(\sigma), \quad (C.10)
\]

and the function \( \psi(U) \) is the probability density function for the power of the \( T \) transmitters (squared eigenvalues of \( A \)). The quantities \( \phi \) and \( \chi \) are

*This dominant macrostate is equivalent to what in information theory is known as the set of typical sequences.
the simultaneous solutions to the following coupled equations:

\[
\begin{align*}
1 + \frac{1}{\sqrt{\phi}} \int_{0}^{\infty} U \psi(U) \sqrt{U} \frac{Q'(\sigma)}{Q(\sigma)} &= 0, \\
1 - \frac{\chi}{\sqrt{\phi}} \int_{0}^{\infty} U \psi(U) \sqrt{U} \frac{Q'(\sigma)}{Q(\sigma)} \sigma &= 0
\end{align*}
\]

(C.11),

where \( Q(\sigma) = \frac{1}{\sqrt{2\pi}} \int_{\sigma}^{\infty} dt \exp \left( -\frac{t^2}{2} \right) \), and \( \sigma = \frac{\kappa + \chi \sqrt{\phi}}{\sqrt{\phi}} \).

As all inputs eligible for reliable transmission are a-priori equi-probable, the uncertainty function for the set of valid inputs at the transmitter is equivalent to the Boltzmann entropy. The uncertainty function is known in information theory as Shannon entropy, and in the VIP channel the Shannon entropy at the transmitter yields directly the Shannon capacity, which is the theoretical upper bound on the information transfer rate \( [3] \). The Shannon capacity of an asymptotic VIP channel is given in bits/symbol/transmitter as

\[
C_{\infty} (\beta, A, \kappa) = \lim_{T \to \infty} \frac{\log_2 \mathcal{N} (T, \beta, A, \kappa)}{T}.
\]

(C.12)

5 Results

Effect of the Amplitude Distribution on the Asymptotic Capacity of the VIP Channel

In order to evaluate the performance of the VIP channel when the transmitters send their information with unequal amplitudes, power is assigned to them according to a chi-square distribution, which is often used to model multipath diversity in multiple access channels \( [23] \). The power distribution function \( \psi(U) \) has variance \( 2/r \) and average power \( U \):

\[
\psi(U; r, \overline{U}) = \frac{r}{\overline{U}} \frac{(rU/\overline{U})^{r/2-1}}{2^{r/2} \Gamma(r/2)} \exp \left( -\frac{rU}{2\overline{U}} \right),
\]

(C.13),

where \( \Gamma(\cdot) \) is the Euler gamma function. We find that, regardless of the value of \( \kappa \), the number of VIP inputs \( \mathcal{N} \) is maximized by a power distribution which assigns equal power to each of the \( T \) transmitters. This result is in contrast to the case of linear multiuser receivers, where equal power interference is the worst interference case scenario \( [2] \). Fig. \( \text{C.3} \) shows how the Shannon capacity of the channel is affected by the power distribution among the \( T \) transmitters.
C. INFORMATION AND MULTIACCESS INTERFERENCE IN A COMPLEXITY-CONSTRAINED VECTOR CHANNEL

**FIGURE C.3:** Asymptotic capacity $C_\infty$ as a function of the load $\beta$ for different distributions of power among the transmitters. The curves shown are for a noiseless ($\kappa = 0$) system, and they correspond to different chi-square power distributions with average 1. The inset represents a linear zoom into the $\beta = 0$ to $\beta = 2$ region.

**FIGURE C.4:** Asymptotic capacity $C_\infty$ vs. channel load for different noise bounds. The $T$ transmitters transmit with unit amplitude.
5. Results

**Figure C.5:** Asymptotic capacity $C_\infty$ vs. the noise threshold, for different channel loads. The $T$ transmitters transmit with unit amplitude.

**Figure C.6:** Asymptotic capacity $C_\infty$ vs. channel load for noise bounds greater than the transmitter’s amplitude (taken to be 1 for all transmitters).
Noise Effects on the Asymptotic Capacity of the VIP Channel

Fig. C.4 displays the asymptotic capacity $C_\infty$ vs. the channel load $\beta$ for different noise bounds $\kappa$. We observe that, for noise bounds not greater than the transmitter’s amplitude (henceforth taken to be 1 for all $T$ transmitters), the channel capacity decreases with the channel load and the noise bound. An interesting phenomenon is that when the noise bound is sufficiently large, multiuser interference becomes beneficial, as it may neutralize part of the noise. Fig. C.5 shows a plot of the asymptotic capacity $C_\infty$ vs. the noise threshold $\kappa$ for different channel loads. Two regimes may be identified: for noise thresholds smaller than the transmitting amplitude, an increase in the channel load results in a lower capacity; on the other hand, as the noise threshold becomes slightly larger than 1, multiuser interference plays a constructive role and, as we can see in Fig. C.6, the optimal channel load may adopt a non-trivial value. This indicates that, in the right conditions, the channel capacity per transmitter may be increased by either adding more transmitters, and/or by turning off some receivers. This effect may also appear for noise bounds smaller than 1 when the power distribution is not optimal, as shown in Fig. C.7. In Fig. C.5 we can also observe how, depending on the channel load, when the noise threshold is sufficiently large, reliable communication in the simple VIP channel becomes infeasible at any rate.
6 Discussion

Multiaccess Interference and Cooperation in the VIP Channel

Although the proposed channel model entails low complexity at the receiver end, it requires strict cooperation among transmitters to carry out the vector input preselection. As shown in Fig. 6.3, and contrary to linear multiuser receivers [2], this cooperation is best served when all transmitters send with identical amplitude. This indicates that, in this joint coding scheme, all transmitters are equally important when it comes to information and cooperative interference. While at moderate loads the effect of multiuser cooperation in the channel capacity is most significant, the influence of their power distribution disappears for vanishing channel loads, i.e. when multiaccess interference disappears. Conversely, as the channel load becomes very large, the adverse effects of multiaccess interference are overwhelming and any multiuser cooperation is hopeless.

Although increasing the load $\beta$ of the channel is generally detrimental to the performance bounds, we observe that, for sufficiently large noise bounds, it may be beneficial to have a greater channel load as shown in Fig. 6.6. For instance, when the noise bound $\chi$ equals 1.1, bringing the channel load $\beta$ from 0.1 to 1 increases the capacity per transmitter by a factor of 10. This means that turning off 90% of the receivers would increase the total channel capacity by a factor of 10, or, conversely, that if the number of transmitters is multiplied by 10, then the total channel capacity (capacity per transmitter $\times$ number of transmitters) would increase by a factor of 100.

Consider the case of equal amplitudes and no channel noise; then the only factor hindering VIP performance appears to be the multiaccess interference as contained in the channel load $\beta$. This suggests, in the spirit of [2], introducing some stages of interference cancellation at the receiver as a strategy to seek improved VIP performance. If we introduce $D$ stages of linear parallel interference cancellation (LPIC) the resulting channel model is

$$\boldsymbol{\vartheta} = \sum_{q=0}^{D} \left( \mathbf{I} - \frac{1}{N} \mathbf{S}^\dagger \mathbf{S} \right)^q \frac{1}{N} \mathbf{S}^\dagger \mathbf{S}_i,$$  \hspace{1cm} (C.14)

and its associated VIP rule such that (3) yields errorfree detection is

$$i \equiv \text{sgn} \left\{ \sum_{q=0}^{D} \left( \mathbf{I} - \frac{1}{N} \mathbf{S}^\dagger \mathbf{S} \right)^q \frac{1}{N} \mathbf{S}^\dagger \mathbf{S}_i \right\}.$$

103
As shown in Appendix A, the number of VIP inputs, and hence the asymptotic capacity of this channel may be found as a function of the channel load $\beta$ and the number $D$ of LPIC stages. The surprising result is that performance declines as the number of interference cancellation stages increases. This remarkable behavior, shown in Fig. C.8, reinforces the indication that, although an increased channel load results in a lower capacity, multiaccess interference itself is a key factor in the performance of the VIP channel. If the transmitters are expected to carry out a vector input preselection, they must be allowed to jointly code their input through interference-cooperation. LPIC neutralizes the constructive role interference takes in the vector input preselection process.

**Efficiency vs. Information Rate in the VIP Channel**

Both multiple-input/multiple-output (MIMO) and code-division multiple-access (CDMA) channels may be modeled by eq. (C.1). While in the MIMO case the $T$ transmitters represent transmitting antennae and the $N$ receivers correspond to receiving antennae, in the case of CDMA the $T$ transmitters represent synchronous users sending their bits to a base station, and the number $N$ of receivers represents the spreading factor at the base station, i.e. the number of chips per received symbol. Note, in passing, that for heavily overloaded systems (i.e. $\beta = T/N \to \infty$) the capacity curve decay
of the noiseless case coincides with the capacity of the Hopfield model (see eq. (12) in [5] for an analytical approximation of the capacity curve at large loads).

In the case of CDMA communication a useful measure of efficiency is the spectral efficiency, which is the total capacity per chip, or the total number of bits per chip that can be transmitted reliably. The asymptotic spectral efficiency $\epsilon_\infty$ of the channel is given by the product of the load $\beta$ and the asymptotic capacity $C_\infty$:

$$\epsilon_\infty (\beta, A, \kappa) \equiv \beta C_\infty (\beta, A, \kappa). \quad (C.16)$$

Spectral efficiency of matched filter demodulation without user cooperation is known [24] to monotonically increase with the load, but to be upper bounded by 1.44 bits/chip. As shown in Appendix II it can be easily verified that the spectral efficiency $\epsilon_\infty$ of the cooperative VIP channel is a boundless monotonically increasing function of $\beta$ as shown in Fig. C.9. This means that in order to achieve maximum spectral efficiency the load needs to be infinitely large, which in turn means zero capacity or data transmission rate. Conversely, if the channel were to achieve maximum capacity (which is 1 bit per channel use and user) the bandwidth burnt by spreading would be extremely high, and the spectral efficiency zero.

Drawing an analogy between spectral efficiency and exergy efficiency allows us to make a connection with irreversible thermodynamics. Just like
maximum exergy efficiency (second law efficiency) can only be attained at extremely slow energy conversion rates (cyclical processes), it is only at vanishing data rates that this channel approaches maximum spectral efficiency. This is in synchrony with the inherently irreversible nature of communication processes. The second law of thermodynamics states that in any real (irreversible) process some power will inevitably be dissipated and that is only in ideal (quasistatic) processes that dissipation vanishes and maximum exergy efficiency can be reached.

The Annealed Approximation

As mentioned in Section 3, as the total number of VIP inputs scales exponentially with $T$, the relevant quantity to be computed is the quenched average $\langle \log N \rangle$. We made use of the annealed approximation and computed instead the annealed average $\langle \log N \rangle$ which, as Jensen’s inequality tells us, provides an upper bound to the former. We carried out both annealed and quenched finite size simulations and found that the results obtained with the two methods are essentially identical within error bars for as few as $T = 20$ transmitters. As shown in Fig. C.10 they are also in good agreement with the asymptotic annealed approximation; the agreement appears to weaken for increasing $\beta$ as decreasing $N$ takes the finite size simulations further away from the asymptotic limit.

An asymptotic calculation of $T^{-1} \log N$ would require use of the replica method [25]. Bray and Moore [26, 27] found that for infinite range spin glasses both $T^{-1} \log N$ and $T^{-1} \log \tilde{N}$ ($\tilde{N}$ being the number of metastable states within a infinitesimal energy band) become identically equal for vanishing off-diagonal order parameters in replica space. They also conducted a stability analysis and found that these diagonal solutions are indeed locally stable, which suggests that the number of metastable states is itself self-averaging. The agreement among both finite size averages in the VIP channel, as well as between them and the asymptotic approximation, suggests that a similar conclusion applies to the number of VIP inputs given by eq. (C.2), or at least that the annealed approximation tightly upper bounds the information capacity of the channel.

7 Conclusion

We have examined the asymptotic capacity of a vector channel under a strict input preselection routine. With this routine the complexity of detection is partially outsourced from the receivers back to the transmitters, which must observe a rigorous interference-based multiaccess cooperation
Figure C.10: Asymptotic curves obtained by the asymptotic annealed approximation (solid curves) are tested against both annealed ($T^{-1} \log N$) and quenched ($T^{-1} \log N$) finite size simulations for $T = 20$ transmitters. The solid curves are for noise bounds $\kappa = 0$ (top), $\kappa = 0.5$ (middle), and $\kappa = 0.75$ (bottom). An ensemble of 100 random binary channel matrices was used in the finite size simulations.

In order to validate the analytically-derived results, we evaluated the capacity of a noisy VIP channel with a large yet finite number $T$ of transmitters. Using exhaustive search simulations we verified that convergence is fast, and the asymptotic approximation reasonable for as few as 20 transmitters. Determining the explicit VIP transmissions in a diagrammatic manner (rather than via brute-force enumeration, which becomes infeasible for large $T$), remains an interesting open research question.

8 Appendices

A. Counting VIP Inputs in the Self-Averaging Limit

Consider the VIP channel (C.1):
\[ \mathbf{o} = \frac{1}{N} \mathbf{S}^\dagger \mathbf{S} \mathbf{a}_i + \mathbf{n}. \] (C.17)

If we introduce $D$ stages of linear parallel interference cancellation (LPIC) at the receiver, the channel becomes
\[ \mathbf{o} = \sum_{q=0}^{D} \left( I - \frac{1}{N} \mathbf{S}^\dagger \mathbf{S} \right)^q \frac{1}{N} \mathbf{S}^\dagger \mathbf{S} \mathbf{a}_i + \mathbf{n}, \] (C.18)
and its associated VIP rule is
\[ i \equiv \text{sgn} \left\{ \sum_{q=0}^{D} \left( I - \frac{1}{N} S^\dagger S \right)^q \frac{1}{N} S^\dagger S i + n \right\}. \] (C.19)

Note that if we let \( D = 0 \) (no interference mitigation) we recover the original VIP rule (C.2) from Section II. The VIP rule (C.19) may be restated as follows:

\[ n(S, i; T, \beta, A, \kappa, D) \equiv \int_{((\kappa-1)\beta^{-1}, \infty)^T} d^T \lambda \, \delta \left\{ T^{-1} \sum_{q=0}^{D} \left( I - \frac{1}{N} S^\dagger S A \right)^q S^\dagger S A i - \left( \beta^{-1} + \Lambda \right) i \right\} = 1, \] (C.20)

where the \( T \) variables of integration are the eigenvalues of the diagonal matrix \( A \). The total number of \( N \) of VIP inputs is obtained by summing over all possibilities:

\[ N(S; T, \beta, A, \kappa, D) = \sum_i n(S, i; T, \beta, A, \kappa, D). \] (C.21)

As discussed in Section 3, as the dimensions of the system go to infinity at constant ratio \( (T, N \to \infty; T/N = \beta) \) then the self-averaging property holds, and we may write

\[ N(S; T, \beta, A, \kappa, D) \xrightarrow{T \to \infty} \bar{N}(T, \beta, A, \kappa, D) \equiv \text{bar} \{ N(S; T, \beta, A, \kappa, D) \}, \] (C.22)

where the overbar denotes the configurational average with respect to the channel matrix \( S \). Then, in the large system limit equation (C.21) may be rewritten as follows:

\[ \bar{N}(T, \beta, A, \kappa, D) = \sum_i n(S, i; T, \beta, A, \kappa, D). \] (C.23)

We may rewrite the \( \delta \)-function in (C.20) as a Fourier sum, resulting in

\[ \bar{N}(T, \beta, A, \kappa, D) = \frac{1}{(2\pi)^T} \int_{((\kappa-1)\beta^{-1}, \infty)^T} d^T \lambda \int d^T \omega \times \]
\[ \times \sum_i \exp \left\{ j \beta^{-1} \omega^\dagger i \right\} \exp \left\{ j \omega^\dagger A i \right\} \exp \left\{ -j T^{-1} \sum_{q=0}^{D} \omega^\dagger \left( I - \frac{1}{N} S^\dagger S A \right)^q S^\dagger S A i \right\}. \] (C.24)
The integrals over the $T$ components of $\omega$ run from $-\infty$ to $\infty$. We may then multiply all the components in $\omega$ by the corresponding component in $i$ without altering the result (recall that the components of $i$ are unit-amplitude binary). Hence we may rewrite (C.24) as follows:

$$
N(T, \beta, A, \kappa, D) = \frac{1}{(2\pi)^T} \int (\kappa^{-1} \omega) \cdot \beta \, d\omega \int d\lambda \, \exp \left\{ j\beta^{-1} \omega^\dagger \mathbf{1} \right\} \exp \left\{ j\omega^\dagger \Lambda \mathbf{1} \right\} \sum_i E_i,$$

(C.25)

where $\mathbf{1}$ is the all-1 column vector, $\hat{\omega}$ is a diagonal matrix such that $\omega \equiv \hat{\omega} \mathbf{1}$, and

$$
E_i \equiv \exp \left\{ -jT^{-1} \sum_{q=0}^{D} i^q \omega \left( I - \frac{1}{N} S^\dagger S A \right)^q S^\dagger S A i \right\}.
$$

(C.26)

In order to deal with the expectation in (C.25), the following transform will be useful:

$$
\exp \left\{ - \frac{j}{T} C^\dagger B \right\} = \frac{1}{(2\pi/T)^{\dim B}} \int_{\mathbb{R}^{\dim B}} d c \int_{\mathbb{R}^{\dim B}} d b \times
\exp \left\{ j \frac{T}{2} c^\dagger c \right\} \exp \left\{ - j \frac{T}{2} b^\dagger b \right\} \cos \left\{ \frac{c^\dagger C + b^\dagger C + c^\dagger B - b^\dagger B}{\sqrt{2}} \right\}.
$$

(C.27)

If we let $C \equiv \sum_{q=0}^{D} S \left( I - S^\dagger S A / N \right)^q \hat{\omega} i$, and $B \equiv S A i$, then we may apply the transform (C.27) to the expectation in (C.26):

$$
E_i \equiv \exp \left\{ -jT^{-1} \sum_{q=0}^{D} i^q \omega \left( I - \frac{1}{N} S^\dagger S A \right)^q S^\dagger S A i \right\} = \frac{1}{(2\pi/T)^{\dim B}} \int_{\mathbb{R}^{\dim B}} d c \int_{\mathbb{R}^{\dim B}} d b \exp \left\{ j \frac{T}{2} c^\dagger c \right\} \exp \left\{ - j \frac{T}{2} b^\dagger b \right\} F_i,
$$

(C.28)

where

$$
F_i \equiv \cos \left\{ \frac{D}{\sqrt{2}} \left( c^\dagger + b^\dagger \right) \left( I - S^\dagger S A / N \right)^q \hat{\omega} i + (c^\dagger - b^\dagger) S A i \right\}.
$$

(C.29)

Now two considerations may be made:
1. In order for the integrals in (C.28) to converge, the components of \( b \) and \( c \) must be of order \( T^{-1/2} \). Hence, in the large \( T \) limit any powers of \( b \) and \( c \) larger than quadratic may be neglected.

2. Up to second order in \( x \), the following equalities hold:
   \[
   \cos x = \exp \left( -\frac{x^2}{2} \right) = \exp \left( -\frac{x^2}{2} \right)
   \]  
   (C.30)

Taking these two considerations into account we may write

\[
F_1 \equiv \cos \left\{ \sum_{q=0}^{D} (c^t + b^t) S (I - S^t S / N)^q \hat{\omega}_i + (c^t - b^t) S A_i \right\} = \exp \left\{ - \left( \sum_{q=0}^{D} (c^t + b^t) S (I - S^t S / N)^q \hat{\omega}_i + (c^t - b^t) S A_i \right)^2 / 2 \right\}
\]  
(C.31)

The following results from random matrix theory [28] will be useful to evaluate the expectation (C.31):

\[
(I - S^t S / N)^q = \sum_{\xi=0}^{q} \left( \frac{q}{\xi} \right) \left( \frac{1}{N} \right)^\xi (S^t S)^\xi,
\]  
(C.32)

\[
\overline{S a b^t S^t} = a^t b I,
\]  
(C.33)

\[
\overline{S (S^t S)^\xi a b^t (S^t S)^\eta S^t} = f (T, \beta, \xi + \eta) a^t b I,
\]  
(C.34)

where \( f \) is a scalar function defined as follows:

\[
f(T, \beta, \xi) \equiv \frac{1}{T} \overline{S (S^t S)^\xi S^t} = T^\xi \left( \frac{SS^t}{T} \right)^{\xi+1}.
\]  
(C.35)

In the infinite dimensional limit, the function \( f \) is given by [28]:

\[
f(T, \beta, \xi) = \frac{\beta^{\xi-1} T^{\xi+1}}{\xi + 1} \sum_{i=1}^{\xi+1} \left( \frac{\xi + 1}{i} \right) \left( \frac{\xi + 1}{i - 1} \right) \beta^i.
\]  
(C.36)
Power Distribution in the VIP Channel

In this subsection we analyze the case of no interference cancellation, that is \( D = 0 \). We do, however, allow for an arbitrary amplitude distribution.

Inserting (C.33) into (C.31) yields

\[
F_i = \exp \left\{ - \frac{(c^\dagger + b^\dagger) S\hat{\omega}i + (c^\dagger - b^\dagger) S\hat{A}i}{2} \right\} = \exp \left\{ -\frac{1}{4} \left\{ \omega^\dagger \omega (c^\dagger + b^\dagger) (c + b) + T\overline{U} (c^\dagger - b^\dagger) (c - b) + +2\omega^\dagger A1 (c^\dagger c - b^\dagger b) \right\} \right\},
\]

(C.37)

where \( \overline{U} \) is the average power (squared amplitude) per transmitter. Inserting the expression for \( F_i \) back into (C.28) we obtain

\[
E_i \equiv \exp \left\{ -jT^{-1} \sum_{q=0}^{N} i^\dagger \hat{\omega} \left( \mathbf{I} - \frac{1}{N} S^\dagger \mathbf{S} \right)^q S^\dagger \mathbf{S} \hat{A}i \right\} = \frac{1}{(2\pi/T)^{\beta^{-1}}} \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} d\Phi \exp \left\{ jTP \left( \beta^{-1}\chi - \frac{1}{2} \left( c^\dagger c - b^\dagger b \right) \right) \right\} = 1.
\]

(C.38)

The following two unity expressions may be inserted into (C.38) without altering its result:

\[
\beta^{-1} \int_{-\infty}^{\infty} d\phi \delta \left\{ \phi \beta^{-1} - \frac{1}{2} \left( c^\dagger + b^\dagger \right) (c + b) \right\} = \beta^{-1} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\Phi \exp \left\{ jT\Phi \left( \phi \beta^{-1} - \frac{1}{2} \left( c^\dagger + b^\dagger \right) (c + b) \right) \right\} = 1,
\]

(C.39)

\[
\beta^{-1} \int_{-\infty}^{\infty} d\chi \delta \left\{ \beta^{-1}\chi - \frac{1}{2} \left( c^\dagger c - b^\dagger b \right) \right\} = \beta^{-1} \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} dP \exp \left\{ jTP \left( \beta^{-1}\chi - \frac{1}{2} \left( c^\dagger c - b^\dagger b \right) \right) \right\} = 1.
\]

(C.40)
Inserting (C.39)-(C.40) into (C.38) and rewriting everything in terms of \( \phi \) and \( \chi \) as dictated by the delta functions, yields

\[
E_i = \frac{T^2 \beta^{-2}}{(2\pi / T)^{1 / \beta - 1} (2\pi)^2} \int_{R^{\beta - 1}} d\tau \int_{R^{\beta - 1}} d\phi \int_{-\infty}^{\infty} d\Phi \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} dP \times 
\]

\[
\times \exp \left\{ T^{\beta - 1} \chi \right\} \exp \left\{ jT\Phi \beta^{-1} \chi \right\} \exp \left\{ \frac{1}{2} \phi \omega^\dagger \omega \right\} \exp \left\{ \frac{1}{2} \phi \omega^\dagger \omega \right\} \exp \left\{ jT^\dagger \left( c^\dagger c - b^\dagger b \right) \right\} \times 
\]

\[
\times \exp \left\{ -\frac{T}{2} \left( c^\dagger + b^\dagger \right) (c + b) \right\} \exp \left\{ -\frac{T U}{4} \left( c^\dagger - b^\dagger \right) (c - b) \right\}. 
\]

(C.41)

The \( c \) and \( b \) integrals may be performed analytically, and (C.41) becomes

\[
E_i = \frac{T^2 \beta^{-2}}{(2\pi / T)^{1 / \beta - 1} (2\pi)^2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\Phi \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} dP 
\]

\[
\times \exp \left\{ T^{\beta - 1} \chi \right\} \exp \left\{ j\beta^{-1} \chi \omega^\dagger A1 \right\} \exp \left\{ -\frac{1}{2} \phi \omega^\dagger \omega \right\} \times 
\]

\[
\times \exp \left\{ jT\Phi \beta^{-1} \chi \right\} \exp \left\{ j\Phi \beta^{-1} \chi \omega^\dagger \omega \right\} \exp \left\{ -\frac{1}{2} \phi \omega^\dagger \omega \right\} \exp \left\{ j\Phi \beta^{-1} \chi \omega^\dagger \omega \right\} \times 
\]

\[
\times \exp \left\{ -\frac{T}{2} \left( c^\dagger + b^\dagger \right) (c + b) \right\} \exp \left\{ -\frac{T U}{4} \left( c^\dagger - b^\dagger \right) (c - b) \right\}. 
\]

(C.42)

In the large \( T \) limit the \( P \) and \( \Phi \) integrals are dominated by the term for which the value of the exponent is largest. That term is given by \( \Phi = -j (\phi - \chi^2 U) / 2\phi^2 \) and \( P = -j \chi U / \phi \). Inserting these expressions in (C.42) we may deem the \( P \) and \( \Phi \) variables as integrated, yielding

\[
E_i = 2^{-2} \pi^{2} T^2 \beta^{-2} U^{-T^{\beta - 1} / 2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \exp \left\{ j\beta^{-1} \chi \omega^\dagger A1 \right\} \exp \left\{ -\frac{1}{2} \phi \beta^{-1} \omega^\dagger \omega \right\} \times 
\]

\[
\times \exp \left\{ T^{\beta - 1} \left\{ \chi + \frac{1}{2} + \frac{\chi^2 U}{2\phi^2} + \frac{1}{2} \ln \phi \right\} \right\}. 
\]

(C.43)

Inserting (C.43) back into (C.25) we obtain the following expression for
the number of allowed inputs:

\[
N(T, \beta, A, \kappa) = 2^{-2\pi^{-2}T^2\beta^{-2}} \int d\phi \int d\chi \int dT \omega \int dT \lambda \times
\]
\[
\times \exp \left\{ j\beta^{-1}\omega^\dagger 1 \right\} \exp \left\{ j\omega^\dagger A1 \right\} \exp \left\{ -\frac{1}{2}\phi\beta^{-1}\omega^\dagger\omega \right\} \times
\]
\[
\times \exp \left\{ T\beta^{-1} \left\{ \chi + \frac{1}{2} + \frac{\chi^2\beta\Phi}{2\phi} + \frac{1}{2} \ln \frac{\phi}{\Phi} \right\} \right\}.
\] (C.44)

After performing the \(T\omega\)-integrals, expression (C.44) becomes

\[
N(T, \beta, A, \kappa) = 2^{-\frac{T}{2}}\pi^{-\frac{T}{2}} T^2 \beta^{-\frac{T}{2}} \int d\phi \int d\chi \int dT \lambda \times
\]
\[
\times \phi^{-\frac{T}{2}} \exp \left\{ T\beta^{-1} \left\{ \chi + \frac{1}{2} + \frac{\chi^2\beta\Phi}{2\phi} + \frac{1}{2} \ln \frac{\phi}{\Phi} \right\} \right\} \times
\]
\[
\times \exp \left\{ -\beta \frac{1}{2} j\left( \beta^{-1}1^\dagger + \beta^{-1}\chi A + 1^\dagger A1 \right) \left( \beta^{-1}1 + \beta^{-1}\chi A1 + A1 \right) \right\}.
\] (C.45)

The \(T\lambda\)-integrals may be performed analytically (recall the \(\lambda\) are the components of the vector \(A1\)), yielding

\[
N(T, \beta, A, \kappa) = 2^{-\frac{T}{2}}\pi^{-\frac{T}{2}} T^2 \beta^{-\frac{T}{2}} \int d\phi \int d\chi \times
\]
\[
\times \exp \left\{ T\beta^{-1} \left\{ \chi + \frac{1}{2} + \frac{\chi^2\beta\Phi}{2\phi} + \frac{1}{2} \ln \frac{\phi}{\Phi} + \frac{1}{\beta^{-1}} \frac{1}{T} \sum_{t=1}^{T} \ln Q(\xi_t) \right\} \right\},
\] (C.46)

where

\[
Q(\xi_t) \equiv \frac{1}{\sqrt{2\pi}} \int_{\xi_t}^\infty dx e^{-\frac{x^2}{2}},
\] (C.47)

\[
\xi_t \equiv \frac{\kappa + a_t \chi}{\sqrt{\beta\phi}},
\] (C.48)

and \(a_t\) is the amplitude of the \(t^{th}\) transmitter, i.e. the \(t^{th}\) eigenvalue of \(A\).

The sum inside the exponential in (C.46) may be viewed as an average over the amplitude (or power) distribution of the \(T\) transmitters. Let \(\psi(U)\)
be the normalized power distribution for the transmitters, then as $T$ goes to infinity we may write
\[
\frac{1}{T} \sum_{t=1}^{T} \ln Q(\zeta_t) = \int_{0}^{\infty} dU \psi(U) \ln \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \ e^{-\frac{x^2}{2}} \right). \tag{C.49}
\]

And we may rewrite (C.46) as
\[
\mathcal{N}(T, \beta, A, \kappa) = 2^{T-2\pi^2} T^2 \beta^{-2} \times
\times \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \exp \left\{ T\beta^{-1} \left( \chi + \frac{1}{2} + \frac{\chi^2}{2\phi} + \frac{1}{\beta} \int_{0}^{\infty} dU \psi(U) \ln Q(\sigma) \right) \right\}, \tag{C.50}
\]

where
\[
\sigma \equiv \kappa + \frac{\sqrt{U}\chi}{\sqrt{\beta\phi}}. \tag{C.51}
\]

As $T$ goes to infinity, all that is required to integrate $\phi$ and $\chi$ is maximizing the exponent in (C.50), with respect to both variables. The maximum is given by the simultaneous solution to the equations
\[
\begin{cases}
1 + \frac{\chi}{\phi} + \sqrt{\beta} \int_{0}^{\infty} dU \psi(U) \sqrt{U} Q'(\sigma) Q(\sigma) = 0 \\
1 - \frac{\chi^2}{\phi^2} - \beta \int_{0}^{\infty} dU \psi(U) \frac{Q'(\sigma)}{Q(\sigma)} \sigma = 0 \end{cases}. \tag{C.52}
\]

**Linear Parallel Interference Cancellation in the VIP channel**

In this subsection we restrict the analysis to the case of a uniform amplitude distribution $A = I$, but allow for an arbitrary number $D$ of interference cancellation stages. Once we expand the square inside the expectation (C.31), equations (C.32)-(C.34) are the expressions we need to evaluate $F_i$, which may be rewritten as follows:
\[
F_i = \exp \left[ -\frac{\Xi_D}{4} \left( c^+ + b^+ \right) \left( c + b \right) \omega^i \omega - \frac{\Psi_D}{2} \omega^i 1 \left( c^i c - b^i b \right) - \frac{T}{4} \left( c^i - b^i \right) \left( c - b \right) \right], \tag{C.53}
\]

where $\Xi_D$ and $\Psi_D$ are defined as
\[
\Xi_D(T, \beta) = \sum_{q=0}^{D} \sum_{\xi=0}^{q} \sum_{t=0}^{\xi} \sum_{\eta=0}^{t} \left( \frac{1}{T \beta^{-1}} \right)^{\xi + \eta} f(T, \beta, \xi + \eta), \tag{C.54}
\]

114
\[
\Psi_D(T, \beta) \equiv \sum_{q=0}^{D} \sum_{\xi=0}^{q} \left( \frac{q}{\xi} \right) \left( \frac{-1}{T^\beta} \right)^\xi f(T, \beta, \xi). \quad (C.55)
\]

Inserting the expression for \( F \) back into \((C.28)\) we obtain

\[
E_i = \frac{1}{(2\pi/T)^{1/\beta}} \int_{\mathbb{R}^{1/\beta-1}} d^{1/\beta-1} c \int_{\mathbb{R}^{1/\beta-1}} d^{1/\beta-1} b \exp \left\{ \frac{T}{2} c^\dagger c \right\} \exp \left\{ -\frac{j}{2} c^\dagger b \right\} \times
\]

\[
\times \exp \left[ -\frac{\Xi_D}{4} \left( c^\dagger + b^\dagger \right) (c + b) \omega^\dagger \omega - \frac{\Psi_D}{2} \omega^\dagger \omega 1 \left( c^\dagger c - b^\dagger b \right) - \frac{T}{4} \left( c^\dagger - b^\dagger \right) \left( c - b \right) \right]. \quad (C.56)
\]

Inserting expressions \((C.39)-(C.40)\) into \((C.56)\) yields

\[
E_i = \frac{T^2 \beta^{-2}}{(2\pi/T)^{1/\beta} (2\pi)^2} \int_{\mathbb{R}^{1/\beta-1}} d^{1/\beta-1} b \int_{\mathbb{R}^{1/\beta-1}} d^{1/\beta-1} c \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\Phi \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} dP \times
\]

\[
\times \exp \left\{ T\beta^{-1} \chi \right\} \exp \left\{ jT\Phi\phi\beta^{-1} \right\} \exp \left\{ jT\Phi\beta^{-1} \right\} \times
\]

\[
\times \exp \left\{ -\frac{\Xi_D}{2} \phi\beta^{-1} \omega^\dagger \omega \right\} \exp \left\{ j\Psi_D \beta^{-1} \chi \omega^\dagger 1 \right\} \exp \left\{ \frac{T}{2} \left( c^\dagger c - b^\dagger b \right) \right\} \times
\]

\[
\times \exp \left\{ -\frac{T}{4} \left( c^\dagger - b^\dagger \right) \left( c - b \right) \right\} \exp \left\{ -\frac{jT\Phi}{2} \left( c^\dagger + b^\dagger \right) \left( c + b \right) \right\}. \quad (C.57)
\]

The \( c \) and \( b \) integrals may be performed analytically, and expression \((C.57)\) becomes

\[
E_i = \frac{T^2 \beta^{-2}}{(2\pi/T)^{1/\beta} (2\pi)^2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\Phi \int_{-\infty}^{\infty} d\chi \int_{-\infty}^{\infty} dP \times
\]

\[
\times \exp \left\{ T\beta^{-1} \chi \right\} \exp \left\{ -\frac{\Xi_D}{2} \phi\beta^{-1} \omega^\dagger \omega \right\} \exp \left\{ j\Psi_D \beta^{-1} \chi \omega^\dagger 1 \right\} \times
\]

\[
\times \exp \left\{ T\beta^{-1} \left( j\phi \chi + j\Phi + \ln \frac{2\pi/T}{\sqrt{2j\Phi - 2P + 1} \sqrt{2j\Phi + 2P - 1}} \right) \right\}. \quad (C.58)
\]

In the large \( T \) limit the \( P \) and \( \Phi \) integrals are dominated by the term for which the value of the exponent is largest. That term is given by \( \Phi = -j(\phi - \chi^2)/2\phi^2 \) and \( P = -j\chi/\phi \). Inserting these expressions in \((C.58)\) we
may deem the $P$ and $\Phi$ variables as integrated, yielding

$$E_i = 2^{-2 \pi^2 T^2 \beta^2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \exp\left\{ -\frac{\Xi D}{2} \phi \beta^{-1} \omega^t \omega \right\} \exp\left\{ j \Psi_D \beta^{-1} \chi \omega^t 1 \right\} \times$$

$$\times \exp\left\{ T \beta^{-1} \left( \chi + \frac{1}{2} + \frac{\chi^2}{2\phi} + \frac{1}{2} \ln \phi \right) \right\}. \quad (C.59)$$

Inserting $(C.59)$ back into $(C.25)$ we obtain the following expression for the number of VIP inputs:

$$N(T, \beta, \kappa, D) = 2^{-2 \pi^2 - T^2 \beta^2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \int_{R^T} d^T \omega \int_{((k-1)\beta^{-1},\infty)^T} d^T \lambda \times$$

$$\times \exp\left\{ j \omega^t \Lambda 1 \right\} \exp\left\{ -\frac{\Xi D}{2} \phi \beta^{-1} \omega^t \omega \right\} \exp\left\{ j \Psi_D \beta^{-1} \chi \omega^t 1 \right\} \exp\left\{ j \beta^{-1} \omega^t 1 \right\} \times$$

$$\times \exp\left\{ T \beta^{-1} \left( \chi + \frac{1}{2} + \frac{\chi^2}{2\phi} + \frac{1}{2} \ln \phi \right) \right\}. \quad (C.60)$$

After performing the $T \omega$-integrals, expression $(C.61)$ becomes

$$N(T, \beta, \kappa, D) = 2^{T/2 - 2 \pi^2 - T/2 \beta^2 + T/2 \Xi_D - T/2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \int_{R^T} d^T \lambda \times$$

$$\times \exp\left\{ -\beta \left( 1 + 2 \Psi_D \right) \frac{1}{\sqrt{1 + \chi \Psi_D}} \left( \beta^{-1} \left( 1 + \chi \Psi_D \right) 1 + \Lambda 1 \right) \right\} \times$$

$$\times \phi^{-T/2} \exp\left\{ T \beta^{-1} \left( \chi + \frac{1}{2} + \frac{\chi^2}{2\phi} + \frac{1}{2} \ln \phi \right) \right\}. \quad (C.61)$$

The $T \lambda$-integrals may be performed analytically (recall that $\lambda$ are the components of the vector $\Lambda 1$), yielding

$$N(T, \beta, \kappa, D) = 2^{T/2 - 2 \pi^2 - T/2 \beta^2} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\chi \times$$

$$\times \exp\left\{ \frac{T}{2 \beta} \left( 2 \chi + 1 + \frac{\chi^2}{\phi} + \ln \phi + 2 \beta \ln Q(\xi_D) \right) \right\}, \quad (C.62)$$

where

$$\xi_D \equiv (\chi + \kappa) \sqrt{\frac{1 - \Psi_D}{\phi \beta \sqrt{\Xi_D}}} \quad (C.63)$$

116
As \( T \) goes to infinity, all that is required to integrate \( \phi \) and \( \chi \) is maximizing the exponent in \( (C.62) \), with respect to both variables. The maximum is given by the simultaneous solution to the equations

\[
\begin{align*}
1 - \frac{\chi^2}{\phi} - \beta \frac{Q'(\xi_D)}{Q(\xi_D)} \xi_D &= 0, \\
1 + \frac{\chi}{\phi} + \beta \frac{\xi_D}{\chi + \chi} \frac{Q'(\xi_D)}{Q(\xi_D)} &= 0
\end{align*}
\]  

(C.64)

B. Spectral Efficiency of the CDMA VIP Channel

In this appendix the boundless monotonic behavior of \( \epsilon_\infty \) as a function of \( \beta \) is shown. Inserting eqs. \( (C.9) \) and \( (C.12) \) into \( (C.16) \) we obtain

\[
\epsilon_\infty (\beta, A, \kappa) \equiv \beta C_\infty (\beta, A, \kappa) = \beta + \frac{1}{\ln 2} g (\beta, A, \kappa). \tag{C.65}
\]

For simplicity and without loss of generality, the channel is assumed noiseless \( (\kappa = 0) \) and the average power per user is assumed to be unity \( (U = 1) \). In this case the value of \( \chi \) which solves eq. \( (C.11) \) is \( \chi = -1 \), and eqs. \( (C.10), (C.11) \) and \( (C.13) \) simplify to

\[
g (\beta, A) = -\frac{1}{2} + \frac{1}{2\phi} + \frac{1}{2} \ln \phi + \beta \int_0^\infty dU \psi(U) \ln Q (\sigma), \tag{C.66}
\]

\[
1 - \frac{1}{\phi} - \beta \int_0^\infty dU \psi(U) \frac{Q' (\sigma)}{Q (\sigma)} \sigma = 0, \tag{C.67}
\]

\[
\psi(U; r) = \frac{(rU)^{r/2-1}}{2^{r/2} \Gamma (r/2)} \exp \left\{ -\frac{rU}{2} \right\}, \tag{C.68}
\]

where \( Q(\sigma) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dt \exp \left\{ -\frac{t^2}{2} \right\}, \) and \( \sigma = \frac{-\sqrt{U}}{\sqrt{\phi}}. \)

The asymptotic capacity \( C_\infty \) is upper bounded by 1 so it is clear that eq. \( (C.65) \) approaches zero from above as the channel load \( \beta \) vanishes. Indeed as \( \beta \to 0 \) it is verified that \( \phi \) becomes 1 for any value of \( r \), which in turn means that the asymptotic capacity \( C_\infty \) approaches 1 and the spectral efficiency \( \epsilon_\infty \) approaches zero. However, as \( \beta \) gets arbitrarily large the parameter \( \phi \) becomes linearly proportional to \( \beta \); it is verified that the constant of proportionality is in the interval \( \left( \frac{4}{\pi^2}, \frac{2}{\pi} \right) \) and it is determined by the power distribution parameter \( r \), which represents the number of degrees of
freedom in the chi-square power distribution $\psi(U)$. The constant of proportionality is smallest for $r = 1$ and largest for $r \to \infty$. As $\beta$ gets large, eq. (C.69) becomes

$$g(\beta \to \infty, A) = -\frac{1}{2} + \frac{1}{2\xi\beta} + \frac{1}{2}\ln \xi\beta + \beta \int_0^\infty dU \psi(U) \ln Q(\sigma), \quad (C.69)$$

where $\sigma = \frac{\sqrt{U}}{2\sqrt{\pi\xi}}$, and $\xi$ is a number in the interval $(\frac{\beta}{\pi}, \frac{2}{\pi})$. If (C.69) is inserted into (C.65), the following expansion can be made about $\beta \to \infty$:

$$\epsilon_\infty(\beta \to \infty, A) = \frac{1}{\ln 2} \sqrt{\frac{2}{\pi\xi}} \int_0^\infty dU \psi(U) \sqrt{U} + \frac{\ln \{\beta\xi\} - 1}{2\ln 2} + \vartheta\left(\beta^{-1}\right). \quad (C.70)$$

Expression (C.70) is a monotonous increasing function of $\beta$, which means that $\epsilon_\infty$ never stops growing with the channel load.
References


Part III

Appendices
Appendix 1

A Connected Extension of the Tomlinson-Harashima Lattice

Abstract

In Papers A and B we considered minimizing quadratic forms using alphabets that were both continuous and convex. In this Appendix I shall consider an alphabet set that, although continuous, is non-convex. The purpose of this Appendix is finding expressions for the replica symmetric parameters of this alphabet, which is a connected complex extension of a binary Tomlinson-Harashima lattice.
1. A CONNECTED EXTENSION OF THE TOMLINSON-HARASHIMA LATTICE

Consider $K$ symbols, $x_k \in \tilde{A}_{s_k}$ (with $k = 1, \cdots, K$), contained in a $K$–dimensional vector $x \in \tilde{A}_s$, where $\tilde{A}_s = \tilde{A}_{s_1} \times \tilde{A}_{s_2} \times \cdots \times \tilde{A}_{s_K}$. The components in $x$ are selected from the alphabets $\tilde{A}_\uparrow$ and $\tilde{A}_\downarrow$ with equal probability. These alphabets are given by

\[
\tilde{A}_\uparrow = \left\{ \xi \in \mathbb{C} : \frac{|\xi + 1|}{2} \in \mathbb{N}_{\text{odd}} \& \Im \xi \geq 0 \right\} \cup \left\{ \xi \in \mathbb{C} : \frac{|\xi - 1|}{2} \in \mathbb{N}_{\text{even}} \& \Im \xi \leq 0 \right\}, \quad \tilde{A}_\downarrow = -\tilde{A}_\uparrow. \tag{1.1}
\]

These alphabets, shown in Fig. 1.1, consist of complex spiraloids made of alternating semicircles centered at $+1$ and $-1$, resulting in a connected extension of the Tomlinson-Harashima lattice [1,2].

Consider a positive definite matrix $E$ of dimension $K$. As in Papers A and B, the purpose is to solve the minimization problem posed in Part I-Section 2 of this thesis. That is, for a typical set of symbols $\{s_1, s_2, \cdots, s_k\}$,
find the value of the following expression:

\[ E \equiv \lim_{K \to \infty} \min_{x \in \tilde{A}_s} x^\dagger E x. \] (1.2)

When it comes to solving this minimization problem, we must be aware that the replica symmetric assumption may be problematic. The alphabets in Fig. 1.1 are not convex; this means that the energy landscape may exhibit multiple minima and ergodicity might be broken as discussed in Part I- Section 3 of this thesis. Nevertheless, for the time being and as a first approximation, we might use the result in Papers A and B, where replica symmetry was assumed and this problem solved in terms of \( R_E \) (the R-transform of \( E \)), yielding

\[ E = q\frac{d}{d\chi} R_E (-\chi), \] (1.3)

where the parameters \( q \) and \( \chi \) are given by the following pair of coupled self-consistent equations:

\[ q = \sum_i P_i \int_c \left| \arg \min_{\xi \in \tilde{A}_i} \left| \frac{\sqrt{q R'_E(-\chi)}}{R_E(-\chi)} - \xi \right|^2 \right| \frac{e^{-|z|^2}}{\pi} dz, \] (1.4)

\[ \chi = \sum_i \frac{P_i}{\sqrt{q R'_E(-\chi)}} \Re \int_c \left| \arg \min_{\xi \in \tilde{A}_i} \left| \frac{\sqrt{q R'_E(-\chi)}}{R_E(-\chi)} - \xi \right|^2 \left| z \right|^2 \right| \frac{e^{-|z|^2}}{\pi} dz. \] (1.5)

The index \( i \) denotes the different information states, each of which can be represented by elements in the set \( \tilde{A}_i \) and occurs in the components of \( x \) with probability \( P_i \). The expression \( R'_E(-\chi) \) denotes the first derivative of \( R_E(t) \) evaluated at \( t = -\chi \).

The purpose of this Appendix is to find expressions (1.4) and (1.5) for the specific case of the spiraloid alphabets in Fig. 1.1. As both information states, i.e. \( s = \uparrow \) and \( s = \downarrow \), occur with equal probability we might write \( P_\uparrow = P_\downarrow = 1/2 \). Also, the symmetry between the alphabets \( \tilde{A}_\uparrow \) and \( \tilde{A}_\downarrow \), given by eq. (1.1), allows us to consider just one of the two alphabets and then multiply by two, hence eqs. (1.4) and (1.5) become

\[ q = \int_c \left| \arg \min_{\xi \in \tilde{A}_\uparrow} \left| \frac{\sqrt{q R'_E(-\chi)}}{R_E(-\chi)} - \xi \right|^2 \right| \frac{e^{-|z|^2}}{\pi} dz, \] (1.6)

\[ \chi = \frac{1}{\sqrt{q R'_E(-\chi)}} \Re \int_c \left| \arg \min_{\xi \in \tilde{A}_\uparrow} \left| \frac{\sqrt{q R'_E(-\chi)}}{R_E(-\chi)} - \xi \right|^2 \left| z \right|^2 \right| \frac{e^{-|z|^2}}{\pi} dz. \] (1.7)
By the substitution $z \to x/\sqrt{2}$ and trivial manipulation, eqs. (1.6) and (1.7) become

$$q = \int_c \left| \arg \min_{\xi \in \tilde{\mathcal{A}}_1} |x - \xi \sqrt{\gamma}| \right|^2 \frac{e^{-|x|^2}}{2\pi} dx, \quad (1.8)$$

$$\chi = \frac{\sqrt{\gamma}}{2R_E(-\chi)} \int \Re \left\{ \arg \min_{\xi \in \tilde{\mathcal{A}}_1} |x - \xi \sqrt{\gamma}| x^* \right\} \frac{e^{-|x|^2}}{2\pi} dx, \quad (1.9)$$

where $\gamma$ is given by

$$\gamma = \frac{2R_E^2(-\chi)}{qR_E(-\chi)}. \quad (1.10)$$

Let us define the sets $\tilde{\mathcal{A}}_1^+$ and $\tilde{\mathcal{A}}_1^-$ such that $\tilde{\mathcal{A}}_1 = \tilde{\mathcal{A}}_1^+ + \tilde{\mathcal{A}}_1^-$ as follows:

$$\tilde{\mathcal{A}}_1^+ = \left\{ \xi \in \tilde{\mathcal{A}}_1 : \Im \xi \geq 0 \right\},$$

$$\tilde{\mathcal{A}}_1^- = \left\{ \xi \in \tilde{\mathcal{A}}_1 : \Im \xi < 0 \right\}. \quad (1.11)$$

Then we might rewrite and manipulate eq. (1.8) until we can write it in terms of modified Bessel functions of the first kind:

$$q = \frac{1}{2\pi} \int_{\mathbb{R}} \left( 1 + \arg \min_{\xi \in \tilde{\mathcal{A}}_1^+} |x - \xi \sqrt{\gamma}| \right)^2 \exp \left( -\frac{|x + \sqrt{\gamma}|^2}{2} \right) dx +$$

$$+ \frac{1}{2\pi} \int_{\mathbb{R}} \left( 1 + \arg \min_{\xi \in \tilde{\mathcal{A}}_1^-} |x - \xi \sqrt{\gamma}| \right)^2 \exp \left( -\frac{|x - \sqrt{\gamma}|^2}{2} \right) dx =$$

$$= \frac{1}{2\pi} \int_0^\infty \frac{4i e^{i\phi} + 1}{e^{-\sqrt{\gamma} \phi}} \sum_{i=0}^{\infty} 2i e^{i\phi} + 1 \left[ \int_{\sqrt{\gamma}^{|4i+2|}} \exp \left( -\frac{\gamma + r^2 + 2r \sqrt{\gamma} \cos \phi}{2} \right) rdr \right] d\phi +$$

$$+ \frac{1}{2\pi} \int_0^{2\pi} \sum_{i=0}^{\infty} (4i + 2)e^{i\phi} \left[ \int_{\sqrt{\gamma}^{|4i+4|}} \exp \left( -\frac{\gamma + r^2 - 2r \sqrt{\gamma} \cos \phi}{2} \right) rdr \right] d\phi =$$

$$= \frac{\gamma}{2\pi} \int_0^\infty \frac{2i e^{i\phi} + 1}{e^{-\sqrt{\gamma} \phi}} \sum_{i=0}^{\infty} 2i e^{i\phi} + 1 \left[ \int_{|2i|^2}^{2i+2} \exp \left( -\frac{\gamma (1 + r^2)}{2} \right) rdr \right] d\phi,$$

$$= \frac{\gamma}{2} \sum_{i=0}^{\infty} \int_{|2i|^2}^{2i+2} \exp \left( -\frac{\gamma (1 + r^2)}{2} \right) \left[ (1 + 4i^2) I_0(\gamma r) - 4i I_1(\gamma r) \right] rdr, \quad (1.12)$$

where the modified Bessel functions of the first kind are

$$I_n(x) = \frac{1}{\pi} \int_0^\pi e^{x \cos(\phi)} \cos(n\phi) d\phi. \quad (1.13)$$
Proceeding along the same lines eq. (1.9) may also be written in terms of $I_n$'s:

$$
\chi = \frac{\sqrt{\gamma}}{4\pi R_E (\chi)} \int_{x > 0} \Re \left\{ \arg \min_{\xi \in \tilde{A}_+} |x - \xi \sqrt{\gamma}| \left( x^* + \sqrt{\gamma} \right) \right\} \exp \left( -\frac{|x + \sqrt{\gamma}|^2}{2} \right) dx + \\
+ \frac{\sqrt{\gamma}}{4\pi R_E (\chi)} \int_{x < 0} \Re \left\{ \arg \min_{\xi \in \tilde{A}_-} |x - \xi \sqrt{\gamma}| \left( x^* - \sqrt{\gamma} \right) \right\} \exp \left( -\frac{|x - \sqrt{\gamma}|^2}{2} \right) dx = \\
= \frac{\sqrt{\gamma}}{4\pi R_E (\chi)} \sum_{i=0}^{\infty} 2i \int_{0}^{\pi} \int_{\gamma^{2i+2}}^{\gamma^{2i+2}} (r + \sqrt{\gamma} \cos \phi) \exp \left( -\frac{\gamma + r^2 + 2r \sqrt{\gamma} \cos \phi}{2} \right) r dr d\phi = \\
= \frac{\gamma^2}{2 R_E (\chi)} \sum_{i=0}^{\infty} \int_{[2i+2]}^{[2i+2]} \exp \left( -\frac{\gamma (1 + r^2)}{2} \right) [r I_0 (\gamma r) - I_1 (\gamma r)] r dr. \quad (1.14)
$$
References


Appendix 2

The R-transform of
\[ \gamma \left( \frac{HH^\dagger}{N} + \gamma I \right)^{-1} \]

Abstract

When the goal of vector precoding is minimizing the mean square error (as opposed to minimizing the transmit power, as we did in Papers A and B), then the relevant metric whose eigendistribution must be known is of the form \( \gamma \left( HH^\dagger / N + \gamma I \right)^{-1} \), where \( H \) is the channel matrix. In this Appendix the R-transform of this matrix is derived and presented.
Consider a MIMO channel which can be described by the following equality:

\[ r = Ht + n, \tag{2.1} \]

where \( t \) is the \( N \)-dimensional input to the channel, \( r \) is a vector containing the \( K \) received data streams, \( n \) is a random vector containing additive noise components, and the channel matrix \( H \) is a complex rectangular matrix containing independent and identically distributed (i.i.d.) entries with zero mean and unit variance.

Non-linear vector precoding might be employed in order to minimize the mean square error at the receiver. When this is done, the transmitted vector \( t \) is an \( N \)-dimensional linear transformation of the \( K \) information symbols (contained in \( x \)) intended for the \( K \) receiving elements:

\[ t = Tx, \tag{2.2} \]

where \( T \) is given by \[ T = H^\dagger N \left( \frac{HH^\dagger}{N} + \gamma I \right)^{-1}, \tag{2.3} \]

and \( \gamma \) is a design parameter.

The transmitter searches for the vector \( x \) which minimizes the mean square error. This is a minimization problem similar to that posed in Part I-Section 2 of this thesis, where now the ‘energy’ metric is (see details in [1])

\[ E = \gamma \left( \frac{HH^\dagger}{N} + \gamma I \right)^{-1}. \tag{2.4} \]

If we apply a replica analysis like in Papers A and B, the problem of finding the minimum mean square error boils down to knowing the eigendistribution of \( E \), which is uniquely determined by its R-transform (see Appendices in Papers A and B). In the following I proceed to find the R-transform of \( E \).

It is well known [2,3] that the R-transform of the limiting spectral measure of \( HH^\dagger / N \) is given by

\[ R_{HH^\dagger} (w) = \frac{1}{1 - \alpha w}. \tag{2.5} \]

It is also known that for the identity matrix \( I \)

\[ R_{I}(w) = 1. \tag{2.6} \]
Then we note that the R-transform has the following two properties [2, 3]:

\[ R_{M_1 + M_2}(w) = R_{M_1}(w) + R_{M_2}(w), \]  
\[ R_{cM}(w) = cR_M(cw) \quad \forall c \in \mathbb{C}, \]  

which directly implies that

\[ R_{E^{-1}}(w; \alpha, \gamma) = \frac{1}{\gamma - \alpha w} + 1. \]  

The R-transform of a matrix is related to that of its inverse by the following relation (see Appendix A in Paper B):

\[ \frac{1}{R_M(w)} = R_{M^{-1}}(-R_M(w)(1 + wR_M(w))). \]  

We might use this relation and solve a cubic equation and find that the R-transform of E is

\[ R_E(w; \alpha, \gamma) = \frac{w - 1}{3w} + \frac{1 + \sqrt{3}j \theta(w; \alpha, \gamma)}{6 \cdot 2^{1/3} \alpha w \gamma} - \frac{1 - \sqrt{3}j \phi(w; \alpha, \gamma)}{3 \cdot 2^{2/3} \alpha w \gamma \theta(w; \alpha, \gamma)}, \]  

where

\[ \phi(w; \alpha, \gamma) \equiv \alpha w \gamma^2 (3 - \alpha - \alpha w + 3\gamma) - \alpha^2 \gamma^2, \]  

and

\[ \theta(w; \alpha, \gamma) \equiv \left( \delta(w; \alpha, \gamma) + \sqrt{\delta^2(w; \alpha, \gamma) + 4 \phi^3(w; \alpha, \gamma)} \right)^{1/3}, \]  

and

\[ \delta(w; \alpha, \gamma) \equiv \alpha^2 \gamma^3 (w - 1) (9w - \alpha (w + 2) (2w + 1)) - 9\alpha^2 w \gamma^4 (2w + 1). \]
References


Appendix 3

A Radially Invariant Replica Analysis

Abstract

In this Appendix I consider the minimization of a normalized quadratic form, i.e. a weighted average of the eigenvalues of a matrix. For a quenched matrix, such a quadratic form is uniquely defined by the phase of the vector. Thus, this problem naturally lends itself to a replica analysis where both the replica matrix and the subshells are radially invariant as well. I use this approach to analyze how close typical states drawn from radially invariant alphabets can get to the minimum eigenmode of random matrices with known eigenspectra.
1 Setting Up the Problem

Consider \( K \) symbols, \( \{ x_1 \in A_{s_1}, \ldots, x_K \in A_{s_K} \} \), contained in a \( K \)-dimensional vector \( x \in A_s \), where \( A_s = A_{s_1} \times A_{s_2} \times \cdots \times A_{s_K} \). Consider a positive definite matrix \( E \) of dimension \( K \). Our goal is to find, in the limit \( K \to \infty \), the value of the following:

\[
\min_{x \in A_s} x^\dagger E x. \tag{3.1}
\]

This problem can be rewritten as:

\[
\lim_{K \to \infty} \min_{x \in A_s} \frac{x^\dagger E x}{\| x \|} \equiv \mathcal{E}. \tag{3.2}
\]

Expression \( (3.1) \) is a weighted average of the eigenvalues of \( E \), and it approaches the minimum eigenvalue from above. As shown in Part I of this thesis (Sections 2 and 3), we might tackle this minimization by writing \( (3.1) \) as a free energy and using the replica trick:

\[
\mathcal{E} = - \lim_{\beta \to \infty} \lim_{K \to \infty} \frac{1}{K} \log \sum_{A} e^{-K \beta \sum_{a=1}^{n} \frac{x_a^\dagger E x_a}{\| x_a \|}} = - \lim_{\beta \to \infty} \lim_{n \to 0} \frac{1}{\beta} \partial_n \Xi_n, \tag{3.3}
\]

where

\[
\Xi_n = \lim_{K \to \infty} \frac{1}{K} \log \sum_{x_1 \in A_{s_1}} \sum_{x_2 \in A_{s_2}} \cdots \sum_{x_n \in A_{s_n}} \exp \left( -\beta E \sum_{a=1}^{n} \frac{x_a^\dagger E x_a}{\| x_a \|} \right), \tag{3.4}
\]

and the overbar denotes average with respect to the matrix \( E \).

2 The Expectation in Terms of Eigenproperties

Using a result by Guionnet and Maida \[1\], the expectation in \( (3.4) \) might be rewritten as follows:

\[
\exp \left( -\beta E \sum_{a=1}^{n} \frac{x_a^\dagger E x_a}{\| x_a \|} \right) = \exp \left( -K \sum_{a=1}^{n} \frac{\tau}{2} \int_{0}^{2\pi} R_E(-w) dw \right), \tag{3.5}
\]

where

- \( R_E \) is the R-transform of the eigendistribution of \( E \),
- \( \tau = 1 \) when \( E \) is purely real and \( \tau = 2 \) when the entries in \( E \) are complex.
3. A Radially Invariant Replica Space

Combining (3.4) and (3.5) we might write\( \Xi_n \) as
\[
\Xi_n = \lim_{K \to \infty} \frac{1}{K} \log \sum_{x_1 \in A} \cdots \sum_{x_n \in A} \exp \left( -K \sum_{a=1}^{n} \frac{\tau}{2} \int_{0}^{\lambda_a} R_E(-w)dw \right).
\] (3.6)

As in \[2, 3\], in order to perform the sum inside the logarithm I split the \( Kn \) dimensional space spanned by the replicas into subshells. This subshells, however, are defined invariant in radial space:
\[
S \{ Q \} = \left\{ x_1, x_2, \ldots, x_n \text{ s.t. } x_a^\dagger x_b = Q_{ab} \right\}.
\] (3.7)

We may then rewrite (3.6) as
\[
\Xi_n = \lim_{K \to \infty} \frac{1}{K} \log \int e^{K I\{ Q \}} e^{-K G\{ Q \}} DQ,
\] (3.8)

where
\[
e^{K I\{ Q \}} \equiv \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} \prod_{a=1}^{n} \delta (1 - Q_{aa}) \times
\]
\[
\times \prod_{b > a} \delta \left( \Re \left\{ \frac{x_a^\dagger x_b}{\| x_a \| \| x_b \|} - Q_{ab} \right\} \right) \delta \left( \Im \left\{ \frac{x_a^\dagger x_b}{\| x_a \| \| x_b \|} - Q_{ab} \right\} \right),
\] (3.9)
\[
G \{ Q \} \equiv \sum_{a=1}^{n} \frac{\tau}{2} \int_{0}^{\lambda_a} R_E(-w)dw,
\] (3.10)
\[
DQ \equiv \prod_{a=1}^{n} dQ_{aa} \prod_{b > a} d\Re Q_{ab} d\Im Q_{ab}.
\] (3.11)

By means of a Laplace transformation we might write \( e^{K I\{ Q \}} \) as follows \[3\]:
\[
e^{K I\{ Q \}} = \int_{\mu^2} e^{K \text{tr}(-KQ)} \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} \exp \left( K \sum_{i=1}^{n} \sum_{k=1}^{n} \tilde{Q}_{ik} \frac{x_i^\dagger x_k}{\| x_k \| \| x_i \|} \right) D\tilde{Q},
\] (3.12)
3. A Radially Invariant Replica Analysis

where \( I = (t - j\infty; t + j\infty) \) for some \( t \in \mathbb{R} \), and

\[
D\bar{Q} \equiv \prod_{a=1}^{n} \frac{d\bar{Q}_{aa}}{2\pi j} \prod_{b>a} \frac{d\bar{Q}_{ba}d\bar{Q}_{ba}}{(\pi j)^2}.
\]

Because all vector components are independent and they are drawn from the same probability space, then we might simplify the following expression as follows:

\[
\sum_{x_1 \in A_i} \cdots \sum_{x_n \in A_i} \exp \left( K \sum_{i=1}^{n} \sum_{k=1}^{n} \bar{Q}_{ik} \frac{x_k^*}{\|x_k\|} \frac{x_i}{\|x_i\|} \right) = \exp \left( \sum_{i=1}^{K} \log M_i \{ \bar{Q} \} \right),
\]

where

\[
M_i \{ \bar{Q} \} \equiv \sum_{x_1 \in A_{i_1}} \cdots \sum_{x_n \in A_{i_n}} \exp \left( K \sum_{a=1}^{n} \sum_{b=1}^{n} \bar{Q}_{ab} \frac{x_b^*}{\|x_b\|} \frac{x_a}{\|x_a\|} \right).
\]

Combining (3.14) and (3.12) we obtain

\[
e^{K\mathcal{I}\{\bar{Q}\}} = \int_{B^2} e^{-K[tr(\bar{Q}\bar{Q}) - \frac{1}{k} \sum_{i=1}^{k} \log M_i \{ \bar{Q} \}]} D\bar{Q}.
\]

4 Replica Symmetry

If we take the replica symmetry assumption, we may, without loss of generality, make the following assignments\(^*\):

\[
\bar{Q}_{ab} = \beta^2 f^2 \ \forall a \neq b,
\]

\[
\bar{Q}_{aa} = \beta^2 f^2 - \beta \epsilon \ \forall a,
\]

\[
Q_{ab} = \frac{1}{1 + \chi/\beta} \ \forall a \neq b,
\]

\(^*\)When things are as in [3], the replica symmetric assumption involves the second moment of \( x \). Now it is \( \kappa \) (the square root of the second moment) that is assumed constant in replica space. However, the shells are intentionally defined radially invariant, i.e. invariant in \( \kappa \) space.
\[ Q_{aa} = 1 \ \forall a, \] (3.21)

\[ x_a = \kappa \ \forall a. \] (3.22)

We might then write \( I\{Q\} \) as follows:

\[ I\{Q\} = I\{\chi\} = \frac{1}{K} \log \int_{\mathbb{R}^{2n}} e^{-\kappa \xi \{Q, \tilde{Q}\}} D\tilde{Q}, \] (3.23)

where

\[
\xi \equiv n\beta \left( \beta f^2 - \epsilon + (n - 1) \frac{\beta^2 f^2}{\beta + \chi} \right) + \frac{1}{K} \sum_{i=1}^{N} \log \sum_{x_1 \in A_i} \cdots \sum_{x_n \in A_i} \exp \left( \left| \beta f \sum_{a=1}^{n} \frac{x_a}{\kappa} \right|^2 - \beta \epsilon \sum_{a=1}^{n} \left| \frac{x_a}{\kappa} \right|^2 \right). \] (3.24)

The eigenvalues of the matrix \( Q \) are, \( \frac{\chi + \beta n}{\chi + \beta} \) with multiplicity \( n - 1 \), and \( \frac{\chi + \beta n}{\chi + \beta} \) with multiplicity 1. We might then rewrite (3.10) as follows:

\[ G\{Q\} = G\{\chi\} = \frac{\tau}{2} (n - 1) \int_{0}^{\frac{\beta n}{\chi + \beta}} R_{E}(-w) dw + \frac{\tau}{2} \int_{0}^{\frac{\beta n}{\chi + \beta}} R_{E}(-w) dw. \] (3.25)

From eq. (3.8) we can see that the term which survives the sum is that for which \( I - G \) is maximum. We can set

\[ \partial_{\chi} (I - G) = 0, \] (3.26)

and solve for \( f \) as \( n \to 0, \)

\[ f^2 = \frac{2}{\tau} \frac{\beta}{\chi + \beta} R_{E} \left( -\frac{2}{\tau} \frac{\beta \chi}{\chi + \beta} \right). \] (3.27)

We have taken one derivative (with respect to \( \chi \)) and obtained one equation (that for \( f \)). We can now take the two remaining derivatives of \( I - G \) (with respect to \( \epsilon \) and \( f \)) and then get the two remaining expressions (for \( \chi \) and \( \epsilon \)). First we apply the Hubbard-Stratonovich transform \( e^{\frac{1}{2} |x|^2} = \int_{\mathbb{C}} e^{2R(2\pi z^{*})} Dz \) to
3. A Radially Invariant Replica Analysis

eq (3.24):

\[ \xi = n\beta \left( \beta f^2 - \epsilon + (n-1) \frac{\beta^2 f^2}{\beta + \chi} \right) + \]

\[ - \frac{1}{K} \sum_{i=1}^{K} \log \int_{C} \left( \sum_{x} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)} \right)^n Dz = \]

\[ = n\beta \left( \beta f^2 - \epsilon + (n-1) \frac{\beta^2 f^2}{\beta + \chi} \right) + \]

\[ - \int \log \int_{C} \left( \sum_{x} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)} \right)^n DzdP_s(s), \quad (3.28) \]

where \( Dz \equiv \frac{e^{-|z|^2}}{\pi} dz \). Since \( G \) does not depend on \( f \) or \( \epsilon \), all we have to care about is differentiating \( I \):

\[ \partial_f I = -\partial_f \xi = 0, \quad (3.29) \]

\[ \partial_\epsilon I = -\partial_\epsilon \xi = 0. \quad (3.30) \]

From eqs. (3.29) and (3.30), the following coupled expressions for \( \chi \) and \( \epsilon \) result as \( n \to 0 \):

\[ \frac{\beta \chi}{\beta + \chi} = \frac{1}{f} \int \int_{C} \frac{\sum_{x} R \{ \bar{z}^* \} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}}{\sum_{x} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}} DzdP_s(s), \quad (3.31) \]

\[ 1 = \int \int_{C} \frac{\sum_{x} \frac{\epsilon}{f} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}}{\sum_{x} e^{\beta \left( 2fR \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}} DzdP_s(s). \quad (3.32) \]

Plugging everything back into equation (3.4), we might now obtain an expression for the ‘energy’ \( \mathcal{E} \), obtaining

\[ \mathcal{E} = R_E \left( -\frac{2}{\tau} \chi \right) - \chi \frac{2}{\tau} R_E \left( -\frac{2}{\tau} \chi \right); \quad (3.33) \]

We have three parameters, namely \( c \equiv \frac{2}{\tau} \chi \), \( \epsilon \) and the ‘energy’ \( \mathcal{E} \), and three equations, which are:

\[ c = \sqrt{\frac{2}{\tau}} \int \int_{C} \frac{\sum_{x} R \{ \bar{z}^* \} e^{\beta \left( 2\sqrt{\tau} R_E (-c) R \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}}{\sum_{x} e^{\beta \left( 2\sqrt{\tau} R_E (-c) R \{ \bar{z}^* \} - \epsilon |\bar{z}|^2 \right)}} DzdP_s(s), \quad (3.34) \]
5. **Radially Invariant One-Dimensional Binary Alphabet**

Now consider two states \( \uparrow \) and \( \downarrow \), represented by \( \hat{A}_\uparrow = \mathbb{R}^+ \) and \( \hat{A}_\downarrow = \mathbb{R}^- \), respectively. In this case, eqs. (3.37) and (3.38) reduce to

\[
1 = \sum_s P_s \int_c \left| \arg \min_{\xi \in \hat{A}_s} \frac{\sqrt{2 \tau R'_E(-c)} e^{-|z|^2}}{\pi} \right| dz,
\]

where the index \( s \) denotes the different information states, each of which can be represented by elements in the set \( \hat{A}_s \) and occurs in the components of \( x \) with probability \( P_s \).

5 Radially Invariant One-Dimensional Binary Alphabet

Approaching the minimum eigenvalue of a Marchenko-Pastur matrix

The \( R \)–transform of a Marchenko-Pastur (MP) matrix \( E \) with ratio \( 1/\alpha \) is given by

\[
R_E(w) = \frac{1}{1 - \alpha w},
\]
and its minimum eigenvalue is \((1 - \sqrt{\alpha})^2\). Plugging (3.42) into (3.41) and (3.36) results in
\[
E = \left(1 - \sqrt{\frac{\alpha}{2\tau}}\right)^2.
\] (3.43)

Expression (3.43) is greater than the lowest eigenvalue of the Marchenko-Pastur matrix. It comes infinitely close as \(\alpha \to 0\). It comes closer when \(\tau = 1\) (real matrix) than it does when \(\tau = 2\).

**Approaching the minimum eigenvalue of a complex inverse Marchenko-Pastur matrix**

If \(E\) is the inverse of a complex MP matrix with ratio \(1/\alpha\), its \(R\)–transform is given by (see Appendix A in Paper B)
\[
R_E(w) = \frac{\frac{2}{\tau} - \alpha - \sqrt{\left(\frac{2}{\tau} - \alpha\right)^2 - \frac{8\alpha w}{\tau^2}}}{2\alpha w}
\] where \(\tau = 2\),
\[
(3.44)
\]
and its minimum eigenvalue is \((1 + \sqrt{\alpha})^{-2}\). Plugging (3.44) into (3.41) and (3.36) results, for \(\tau = 2\), in (see Fig. 3.1)
\[
E = \frac{1}{\sqrt{1 - \frac{\alpha^2}{\tau} + (1 - \frac{\alpha}{2}) \sqrt{\alpha(4 - 3\alpha)}}}.
\] (3.45)

**Approaching the minimum eigenvalue of the inverse of the real part of a Marchenko-Pastur matrix**

Consider a matrix \(E\) which is the inverse of the real part of a MP matrix with ratio \(1/\alpha\). The \(R\)– transform of this matrix is given by expression (3.44) evaluated at \(\tau = 1\) (see Appendix A in Paper B). Plugging this expression into (3.41) and (3.36) results, for \(\tau = 1\), in (see Fig. 3.2)
\[
E = \frac{1}{\sqrt{1 + \alpha - \left(\frac{\alpha}{2}\right)^2 + \sqrt{\alpha(4 - \alpha)}}}.
\] (3.46)
6. **Symmetric, Radially Invariant, $m$-PSK Alphabets ($m \geq 2$)**

Consider dividing the complex plane into $m$ equal sectors representing different information states. In this case, eqs. (3.37) and (3.38) reduce to

$$c = \frac{(4 + m) \kappa}{2m} \frac{\epsilon \tau'}{\epsilon \tau}, \quad (3.47)$$

$$1 = \frac{(4 + m) \kappa^2 R'_E(-c)}{2m} \frac{\epsilon^2 \tau}{\epsilon^2 \tau}. \quad (3.48)$$

These two equations may be combined yielding

$$c = \gamma \sqrt{\frac{1}{\tau R'_E(-c)'}}, \quad (3.49)$$

where

$$\gamma \equiv \sqrt{\frac{4 + m}{2m}}. \quad (3.50)$$
3. A Radially Invariant Replica Analysis

Figure 3.2: Approaching the lowest eigenvalue of the inverse of the real part of an MP matrix using typical vectors drawn from a one-dimensional, radially invariant, binary alphabet.

Approaching the minimum eigenvalue of a complex inverse Marchenko-Pastur matrix

Consider a matrix $E$ which is the inverse of a complex MP matrix with ratio $1/\alpha$. The $R-$ transform of this matrix is given by expression (3.44) evaluated at $\tau = 2$. Plugging this expression into (3.49) and (3.36) results, for $\tau = 2$, in (see Fig. 3.3)

$$
E_m = \frac{1}{\sqrt{(1 - \alpha)^2 + 2\alpha\gamma^2 (2 - 2\alpha + \alpha\gamma^2) + 2\sqrt{(2\alpha - 2\alpha^2 + \alpha^2\gamma^2) (\gamma - \alpha\gamma + \alpha\gamma^3)^2}}}.
$$

(3.51)
7. Some Impressions

This energy diverges at a load \( \alpha_{\text{div},m} = \frac{4m}{3m-4} \), and it has a minimum at \( \alpha_{\text{min},m} = \frac{4+m}{3m-4} \). Interestingly, as \( m \to \infty \) the energy converges to†

\[
\mathcal{E}_\infty = \frac{1}{\sqrt{1 - \alpha^2} + \left(1 - \frac{\alpha}{2}\right) \sqrt{\alpha(4-3\alpha)}}. \tag{3.52}
\]

7 Some Impressions

The expression for \( \mathcal{E} \) is always above the minimum eigenvalue, coming infinitely close to it as \( \alpha \to 0 \). The interesting feature is that \( \mathcal{E} \) is not a monotonous function of \( \alpha \), as shown in the three figures.

To explain this feature, it is important to note that expression (3.1) is a weighted average of the eigenvalues of the matrix \( \mathbf{E} \). When \( \alpha \) is zero, then

†Notice this equation is the same as eq. (3.45) for a BPSK alphabet on the real line. This makes sense as \( m \to \infty \) means averaging over infinitely many BPSK lines through the origin.
the Marchenko-Pastur matrix becomes the identity, and the minimization yields 1 necessarily. But as the load $\alpha$ increases, the lowest eigenvalue gets smaller. By giving full radial freedom to the vector components, we can actually take advantage of the existence of the lower eigenvalues which exist at greater values of $\alpha$, resulting in smaller weighted averages. But of course, if $\alpha$ gets much larger, then the largest eigenvalues are also increasing; at some $\alpha$ large enough this effect takes over and the average begins to grow.
References


