Non-Asymptotic Results for Point to Point Channels

by

Austin Daniel Collins

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Author .....
Department of Electrical Engineering and Computer Science
May 23, 2019

Signature redacted

Certified by .....
Yury Polyanskiy
Associate Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Signature redacted

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Non-Asymptotic Results for Point to Point Channels

by

Austin Daniel Collins

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

This thesis demonstrates some non-asymptotic information theoretic results for point to point channels. Non-asymptotic information theory addresses the question: "For a fixed blocklength and fixed probability of error, what is the maximum number of codewords M that I can support?". Compare this to classical (asymptotic) information theory, which answer this question only for blocklengths tending to infinity and probability of error tending to zero. In this sense, non-asymptotic results are more difficult to derive, but are more practically applicable.

First, we look at the multiple input multiple output (MIMO) coherent block fading channel with channel state information available at the receiver, called the MIMO-BF channel. This is perhaps the most well studied model for a wireless communication channel – it captures the setting where two wireless devices are communicating without a dominant line of sight between them, so the signal reflects off many surfaces before reaching the receiver. A typical example of this channel is communication between two cell phones in a city. The MIMO assumption means the transmitter and receive may have multiple antennas – adding multiple antennas can increase achievable rates enormously while costing very little.

This work characterizes the *dispersion* of the MIMO-BF channel. The dispersion is a fundamental channel quantity similar to capacity – it describes the rate penalty incurred for transmitting at a fixed blocklength and error probability. We first prove achievability and converse theorems, together which demonstrate that the dispersion is given by the conditional variance of the information density, minimized over all capacity achieving input distributions. We then give an analytic expression for the dispersion, and describe its implications in terms of the channel parameters. For example, we learn that dispersion scales linearly with the coherence time, while capacity is not a function of the coherence time. We then give an achievability bound to help numerically compute the finite blocklength rates, and demonstrate its application to the MIMO-BF channel.

Secondly, we analyze the MISO case – where the transmitter has many antennas but the receiver has only one, which turns out to be an interesting special case. For this, we first give a theorem characterizing the input distributions that achieve capacity. It turns out that *full rate orthogonal design*-like input distributions achieve capacity, along with the distribution with i.i.d. Gaussian entries. It is shown that these orthogonal design objects are in fact the extremal objects of this channel from

the point of view of dispersion – using them gives better performance then simply sending independent symbols from each antenna at each time step, a result that cannot be seen with only asymptotic analysis. In this way, orthogonal designs appear as the natural space-time coding scheme for the MISO channel.

Finally, we analyze the problem of *variable length list decoding with stop feedback.* This is the following problem: a transmitter sends symbols one by one into a channel until the decoder says "stop". The decoder then outputs a list of L codewords – if the correct codeword is in the list, it succeeds, else it makes an error. Hence there is a tradeoff between stopping time, number of messages, and the probability of error. The question becomes: which indicators can tell the decoder that the correct message is in a set of L messages? We demonstrate for the BEC channel that it is possible to communicate with zero dispersion using a variable length list decoding scheme. However for the BSC, surprisingly you cannot stop in a way the gives zero dispersion, when the list size is sub-exponential relative to the number of messages. Furthermore, we show an application to delayed variable length feedback – i.e. the receiver says "stop" but the transmitter only sees the stop signal after a delay, which gives a more practical way of using stop feedback.

Thesis Supervisor: Yury Polyanskiy
Title: Associate Professor of Electrical Engineering and Computer Science

To my parents,
Bill and Barbara Collins

Acknowledgments

I have learned quite a lot about research and about myself during this Ph.D., through all its ups and downs, and there are many people I am extremely grateful towards. This is of course a small subset of names – its not an exaggeration to say that nearly everyone I've interacted with at MIT has been friendly, open minded, and has improved my graduate experience in some way.

First of all, to my parents Bill and Barbara Collins, and my brother Christian. You've always given me the opportunity to make my own decisions, providing wisdom and support regardless of what I chose to do. I would never have started this PhD without you.

Secondly, I'd like to thank my advisor Yury Polyanskiy. I've been at MIT with you for a long time, and to put it mildly, you've had an enormous impact on the way that I think. From learning how to distill dressed up research and complicated questions down to their intellectual meat, to quickly sketching "physicist" arguments via many approximations, to being an incredible resource for nearly any mathematical concept – your mentorship has made the PhD experience what it was.

I'd like to thank my thesis committee members: Gregory Wornell and Lizhong Zheng – I could not have asked for a better information theory committee at MIT. Thank you for your comments, and for reading another MIT channel coding thesis.

Fourth, I'd like thank the professors who have influenced me over the years. Thanks to Greg Wornell and Guy Bresler, both of whom I was lucky to TA for. Both of you taught me quite a lot about teaching, designing good exam questions, and conveying technical information in general. Thanks to Alan Oppenheim for welcoming me into his group meetings back when my group consisted of only one person – those group meetings taught me a lot about brainstorming research in a maximally open environment, an "intellectual food fight" as Al would say. Finally, thank you to Bobak Nazer for introducing me to information theory back in undergrad, and Prakash Ishwar for the teaching and guidance you provided.

I'd like to thank my friends and lab mates who have been with me over the years, the experience would have been infinitely more difficult if it wasn't for your presence. I'd especially like to thank Ganesh Ajjanagadde, Anish Argarwal, Sarah Cen, Parker Gould, Yuzhou Gu, Igor Kadota, Anuran Makur, Flora Meng, James Noraky, Hajir Roozbehani, Tuhin Sarkar, Sohil Shah, Katie Szeto, Jennifer Tang, Zhi Xu, and everyone else whom I've had the pleasure to interact with on a daily basis – you made LIDS and the 6th floor a wonderful space to be around for so many years.

Finally, I'd like to thank the LIDS staff: Jennifer Donovan, Brian Jones, and Francisco Jaimes for always helping with a smile whenever I needed something, regardless of how small.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

First of all, thanks for deciding to look at my thesis. In this section, we first give an introduction to non-asymptotic information theory, and then briefly describe what we consider to be the main interesting contributions of the thesis.

1.1 Non-Asymptotic Information Theory

Information theory, and more specifically coding theory, centrally deals with the problem of how to map k bits into n channel symbols in order to send them over a noisy channel, then recover those original k bits at the other side. In 1948, Shannon [1] gave us a beautiful description of what is possible in this setting: as n and k both tend to infinity with the ratio k/n being fixed as R, then there is a value C called the *capacity* for which, if $R < C$, communication with an arbitrarily small error is possible, otherwise the probability of error is bounded away from 0. This essentially showed that a communication channel is like a water pipe – there exists a maximal rate C bits per channel use at which information can flow with low probability of error, but beyond that rate, one cannot push information through the channel without incurring enormous errors.

The main uses of this result is 1) to show that you can transmit at a rate strictly larger than zero with arbitrary small probability of error, which is highly non-obvious, and 2) that designers of codes have a benchmark rate of C against which they can compare the performance of their codes. E.g. if a code achieves 50% of capacity, then large rate improvements can be made to the system simply by developing a smarter coding scheme rather than increasing power or getting more bandwidth; however if a code achieves 99% of capacity, it is likely not worth your effort to invest in improving your coding scheme, since you are near the theoretical limit. Compare this to many problems in statistics where we do not have such a demonstrated theoretical limit – researchers are constantly beating each others' algorithms to achieve smaller errors on datasets, but no one knows if these algorithms are 50% or 99% optimal.

Shannon's results were *asymptotic* in nature – they hold when the number of bits and the code blocklength tend to infinity. But what if we ask: which rates are achievable if $k = 100, n = 1000, \epsilon = 10^{-3}$? It is possible that the largest theoretical

rate for these fixed constants may be 90% of capacity, hence a code designer would be mislead, they'd say "why am I only achieving 90% of capacity, how do I get the extra 10%?" when in reality, they already are close to the theoretical limit. Hence capacity is insufficient, or at least inaccurate, for characterizing theoretical rate limits for a fixed blocklength and error probability.

Non-asymptotic information theory addresses this question. The original formulation of non-asymptotic information theory was in terms of *error exponents*, e.g. see Gallager [2, Theorem 5.6.2]. If we define $\epsilon^*(M,n)$ to be the smallest error possible using M codewords and blocklength n, then error exponent results show $\epsilon^*(M,n) \leq \exp(-nE_0)$, i.e. the probability of error decreases with some exponent E_0 as blocklength increases, where E_0 is only a function of the rate $R = \frac{1}{n}\log M$. More recently, Polyanskiy, Poor, and Verdu [3] developed a new framework for looking at non-asymptotic information theory. They instead analyzed the largest number of messages a codebook can support for fixed n and ϵ, denoted by $M^*(n,\epsilon)$. They showed that for a general discrete memoryless channel, we have

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,. \tag{1.1}$$

where Q is the complementary Gaussian CDF. After diving by n to look at rate rather than messages, and taking the blocklength $n \to \infty$, we recover the capacity. Notice that the quantity $\log M^*(n,\epsilon)$ is not directly computable (there are doubly exponential many codebooks in n to search over). To deal with this, upper and lower bounds on $\log M^*(n,\epsilon)$ are proven, with the aim of having a small gap between them. The parameter V above is called the *dispersion* – it is a fundamental channel parameter similar to capacity which measures the gap in rate from capacity incurred by using a fixed n and probability of error ϵ. The dispersion leads to a more refined approximation for the maximal rate, given by the *normal approximation*

$$\frac{1}{n}\log M^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon)\,. \tag{1.2}$$

Simulations show that this approximation is remarkably tight in most cases. The channel dispersion had been established for the most fundamental channel in information theory, e.g. the Binary Symmetric Channel, Binary Erasure Channel, and Additive White Gaussian Noise Channel, but the question of dispersion and finite blocklength performances for more complicated channel models had not yet been studied. This study is largely the contents of this thesis.

1.2 Characterization of the MIMO Coherent Block Fading Channel at Finite Blocklength

The first major contribution of this thesis is to characterize the finite blocklength performance of the Mulitple Input Multiple Output (MIMO) Coherent Block Fading channel (called the MIMO-BF channel). The MIMO-BF channel models wireless

communication between two devices in an environment where there is no dominant line of sight between the transmitter and receiver – hence the signal bounces around before reaching its destination. Telatar [4] gave the most famous result for this channel, showing that the capacity scales as

$$C \approx \min(n_r, n_t) \log(SNR) \tag{1.3}$$

i.e. the minimum of the number of antennas on the transmitter and receiver times the logarithm of the channel SNR. The significance of this result is that one could increase rate linearly simply by scaling up the number of antennas. Scaling up antennas is cheap and easy compared to other methods of increasing rate. This scaling lead to the incorporation of MIMO in all modern telecommunication standards: e.g. 4G, LTE, 802.11 for Wi-Fi, and will play a large role in 5G standard when released. The higher rates provided by good codes and multiple antennas allow for things like Wi-Fi on planes and trains, streaming movies, real time video chatting, and all the future application requiring high data rates that have not yet been invented.

Out of all wireless channel models, the MIMO-BF channel is amongst the most well studied, yet still we do not even know its dispersion. The first major contribution of this thesis is to give achievability and converse theorems that establish the dispersion of this channel. After proving the dispersion via coding theorems, we give a closed form expression, similar to Telatar's expression for capacity (though, certainly less monumental), and discuss its scaling in terms of number of antennas, power, and coherence time. We show how to numerically compute the dispersion of this channel via the $\beta\beta$ bound, the code for which can be found in the SPECTRE package [5].

1.3 MISO Case

The second major contribution of this thesis is the analysis of the Multiple Input Single Output (MISO) case – an especially interesting special case of the MIMO-BF channel. This special case is interesting for the following reason: when we establish the dispersion, we show that it is given in the variational form

$$V = \inf_{P_X : I(X;Y,H)=C} \mathbb{E}\left[\mathrm{Var}\left(i(X;Y,H)|X\right)\right] \tag{1.4}$$

i.e. it is a minimization over capacity achieving input distributions, similar to how capacity is a maximization over distributions. It turns out that when the number of receive antennas is at least 2, then the capacity achieving input distribution is unique, and hence the minimization above is trivial. However, when there is only one receive antenna, the MIMO-BF channel has many interesting capacity achieving input distributions. For example, consider two such capacity achieving input distributions for the MISO channel with two transmit antennas: for $A, B, C, D \sim \mathcal{N}(0, \frac{P}{2})$ i.i.d.,

$$X_G = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \qquad X_A = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}. \tag{1.5}$$

Here, the subscript G denotes the i.i.d. Gaussian input, corresponding to sending independent data from each antenna at each time step, the A denotes the "Alamouti" input, inspired by Alamouti's famous scheme [6], which sends two independent symbols in the first time step, then the orthogonal vector in the second time step. Both of these distributions achieve capacity, but is either a minimizer in (1.4)?

To this end, we first give a characterization theorem of all distributions that achieve capacity in this channel. We show that amongst them are *full rate orthogonal designs*, of which X_A above is an example. We then give a theorem showing that these full rate orthogonal designs are the unique minimizers in (1.4). Note that full rate orthogonal designs do not exist in all dimensions. In dimensions where they exist, this result shows that they are in a sense the extremal objects in the MIMO-BF channel for the MISO case. In dimensions where they do not exist, We give a criteria for finding input distributions that are optimal in the sense of minimizing dispersion: maximize $\mathrm{Var}(\|X\|_F^2)$ subject to X being a capacity achieving input distribution, and explicitly compute its maximizer in the 2×3 and 3×3 cases.

1.4 Variable Length List Decoding

The third contribution we make is to the problem of *variable length list decoding*, which is introduced in this thesis. Up to this point, we have been talking about *fixed* blocklength codes, where the transmitter sends n symbols, and the decoder must estimate the message based on those n symbols. Variable length coding instead requires that the decoder estimates the correct message with n symbols on average – but the total number of symbols sent can vary. Of course, the encoder must have some way of knowing when to stop transmitting, and hence the simplest form of feedback is used – the encoder send symbols until the decoder says "stop", at which point the decoder outputs an estimate of the message. This is called *stop feedback*. In the list decoding setting, the decoder instead outputs a set of L messages, and an error occurs if the correct message is not in the set of messages.

Polyanskiy et al [7] demonstrated that the use of variable length coding with stop feedback can dramatically reduce the gap to capacity for a fixed average blocklength. For example, for a discrete memoryless, we have the expansion

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1)\,, \tag{1.6}$$

whereas with variable length coding, using average blocklength ℓ, we have

$$\log M^*(\ell, \epsilon) = \frac{\ell C}{1-\epsilon} + O(\log \ell) \tag{1.7}$$

I.e. the dispersion and even logarithmic term vanish. The intuition behind this is that a fixed blocklength code requires that the correct codeword be distinguishable with high probability at time n, whereas for a variable length code, if the correct codeword is not yet distinguishable, the system can simply wait a bit longer. This optionality is key to narrow the gap to capacity.

In this thesis, we analyze the variable length coding problem when the decoder uses list decoding. We show that, for the Binary Erasure Channel, it is again possible to kill all terms beside the linear and constant term, similar to (1.7). We give an application of this variable length list decoding scheme to the problem of stop feedback with delay. Namely, suppose we are in the variable length (non list decoding) setting, but when the decoder says "stop", the encoder only sees it after a delay of D symbols. Hence the decoder must say stop *earlier* than it can distinguish the correct message – but when it believes that with D extra symbols, the correct message will become distinguishable. Then we show that for the Binary Symmetric Channel, we cannot manage to stop in a way that gives zero dispersion whenever the list is of size $L = M^{1-\alpha}$, for $\alpha \in (0, 1)$, showing surprisingly that the behavior from (1.7) does not carry over to the BSC case.

1.5 Organization of this Thesis

Chapter 2 focuses on the definition and basic properties of the MIMO-BF channel. First the channel is defined, then the capacity expression and capacity acheiving output distribution are described. Next, a theorem characterizing all input distribution that achieve capacity is proven. Finally, the information density is computed for this channel, which will be instrumental later on.

Chapter 3 gives the achievability and converse theorems for the MIMO-BF channel, showing that they agree up to the $O(\sqrt{n})$ term, thus establishing dispersion. First, an introduction is given to hypothesis testing in the Neyman-Pearson settings, and a number of lemmas are proven which are essential in the achievability and converse proofs. We remark the converse is only a partial converse, but in a benign way – see Section 3.3 for more details.

Chapter 4 discusses the numerical computation of an achievability bound for this channel. A new average probability of error achievability bound, the $\beta\beta$ bound, is stated and proved. Then its computation for the MIMO-BF channel specifically is discussed. The code computation described in this chapter is used in the SPECTRE package [5].

Chapter 5 gives a closed form expression for the dispersion, whereas the dispersion was only given as the variational form (1.4) in Chapter 3. This expression is first computed, then some implications of the expression in terms of number of transmit and receive antennas, coherence time, and power are discussed.

Chapter 6 discussed the special case of the MIMO-BF channel where the receiver has only a single antenna, but the transmitter may have many antennas. First the form of the dispersion is given as a function of the capacity achieving input distribution. Then we introduce objects called *full rate orthogonal design*, and show that they are the minimizers of this dispersion expression uniquely, when they exist. Finally we show that in dimensions where they do not exist, we can use a truncation construction to obtain input distributions that preform strictly better than the i.i.d. Gaussian input. Finally, we give a brute force computation of the optimal input distribution for the $n_t = 2, T = 3$ and $n_t = 3, T = 3$ cases.

Chapter 7 discusses the variable length list decoding, a problem defined here for the first time. Results are given for the Binary Erasure Channel, showing that capacity can be approached quickly. An application the delayed variable length feedback is given. Finally, it is shown that for the Binary Symmetric Channel, we cannot achieve the same fast convergence as in the Binary Erasure Channel.

Chapter 2

The Channel Model and its Capacity Achieving Input Distributions

In this chapter, we begin by defining the main fading channel of interest, the MIMO coherent block fading channel, in Section 2.1. We discuss some classical results for this channel in Section 2.2, then look at the capacity achieving input distributions (caids) for this channel in Section 2.3. Finally, we define the information density for this channel in Section 2.4, which will be useful later on.

Before we define the model, we make a remark on the assumptions that go into specifying a fading channel. Generally, we have the following categories:

- **Channel state information:** receiver (CSIR), transmitter (CSIT), both (CSIRT), or none (noCSI).
- **Fading dynamics:** H generated once then fixed (quasi-static), H generated once for T symbols independently (block fading), H generated independently every time step (fast fading).
- **Fading distribution:** most popular are Rayleigh which models a rich scattering environment, and Rician which models antennas with a dominant line of sight.
- **Availability of antennas:** one transmit and one receive antenna (SISO), one transmit and many receive antennas (SIMO), many transmit and one receive antenna (MISO), many transmit and many receive antennas (MIMO).

This work will consider a rotationally invariant fading process (a generalization of Rayleigh fading), with CSIR, block fading, in the MIMO and MISO settings. Note that some combinations are less realistic than others – for example, fast fading with CSIRT is unrealistic, since the fading coefficients must be estimated, and if they change every time step, there is no time to estimate them.

2.1 The MIMO Coherent Block Fading Channel

The channel model considered in this paper is the frequency-nonselective coherent real block fading (BF) discrete-time channel with multiple transmit and receive antennas (MIMO) (See [8, Section II] for extra background on this model). We will simply refer to it as the *MIMO-BF channel*, which we formally define here. This channel is parameterized by the quantities n_t, n_r, P, T, which are

- $n_t \geq 1$ – the number of transmit antennas
- $n_r \geq 1$ – the number of receive antennas
- $T \geq 1$ – the coherence time of the channel
- $P > 0$ – the power available to the transmitter in decibels

The input-output relation at block j (spanning time instants $(j-1)T+1$ to jT) with $j = 1, \ldots, n$ is given by

$$Y_j = H_j X_j + Z_j \,, \tag{2.1}$$

where

- $\{H_j, j = 1, \ldots\}$ is a $n_r \times n_t$ matrix-valued random fading process.
- X_j is a $n_t \times T$ matrix channel input.
- Z_j is a $n_r \times T$ Gaussian random real-valued matrix with independent entries of zero mean and unit variance.
- Y_j is the $n_r \times T$ matrix-valued channel output.

The process H_j is assumed to be i.i.d. with isotropic distribution P_H, i.e. for any orthogonal matrices $U \in \mathbb{R}^{n_r \times n_r}$ and $V \in \mathbb{R}^{n_t \times n_t}$, both UH and HV are equal in distribution to H. We also assume

$$\mathbb{P}[H \neq 0] > 0 \tag{2.2}$$

to avoid trivialities. We assume coherent demodulation so that the channel state information (CSI) H_j is fully known to the receiver (CSIR).

Note that due to merging channel inputs at time instants $1, \ldots, T$ into one matrix-input, the block-fading channel becomes memoryless. This is slightly different than the phrasing of the MIMO-BF channel in the literature – often the input is simply a dimension n_t vector, rather than an $n_t \times T$ matrix. For example, in the $n_t = T = 2$, $n_r = 1$, a single channel input / output relation is given by

$$[Y_1 \; Y_2] = [H_1 \; H_2] \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} + [Z_1 \; Z_2] \,. \tag{2.3}$$

The reason for such a definition is that, as we will later see, when an input distribution P_X is capacity achieving, all columns are individually i.i.d. Gaussian, however the

joint distribution of the matrix may not be i.i.d. Gaussian. Hence viewing the input as a matrix allows us to identify these capacity achieving input distribution.

We now define a code for this channel:

Definition 1. An $(nT, M, \epsilon, P)_{CSIR}$ average probability of error code of blocklength nT, probability of error ϵ and power-constraint P is a pair of maps: the encoder $f : [M] \to (\mathbb{R}^{n_t \times T})^n$ and the decoder $g : (\mathbb{R}^{n_r \times T})^n \times (\mathbb{R}^{n_r \times n_t})^n \to [M]$ satisfying the probability of error constraint

$$\mathbb{P}[W \neq \hat{W}] \leq \epsilon \,. \tag{2.4}$$

on the probability space

$$W \to X^n \to (Y^n, H^n) \to \hat{W} \,,$$

where the message W is uniformly distributed on $[M]$, $X^n = f(W)$, $X^n \to (Y^n, H^n)$ is as described in (2.1), and $\hat{W} = g(Y^n, H^n)$. In addition the input sequences are required to satisfy the power constraint:

$$\sum_{j=1}^{n} \|X_j\|_F^2 \leq nTP \qquad \mathbb{P}\text{-a.s.} \,,$$

where $\|M\|_F^2 \triangleq \sum_{i,j} M_{i,j}^2$ is the Frobenius norm of the matrix M.

If instead, the probability of error criterion (2.4) is replaced by

$$\max_{w \in \{1,\ldots,M\}} \mathbb{P}\left[W \neq \hat{W} \middle| W = w\right] \leq \epsilon \tag{2.5}$$

then this is called a *maximum* probability of error code. Note that the blocklength is given in increments of T, since the channel input is defined as a matrix with T columns. So for example, $n = 5, T = 2$ corresponds to 10 actual channel uses.

With this, we are interest in the fundamental quantity

$$\log M^*(n, \epsilon, P) = \sup \{M : \exists (nT, M, \epsilon, P)_{CSIR} \text{ code}\} \,. \tag{2.6}$$

In words, the logarithm of the maximum number of messages in a codebook, where each codeword must have blocklength nT and total power bounded by nTP, and have overall probability of error bounded by ϵ. When unclear, we will use M_{avg}^* and M_{max}^* to denote the average and maximum probability of error cases, respectively.

As a remark, note that the usual definition of capacity in terms of this notation is given by

$$C = \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{nT} \log M^*(n, \epsilon, P) \tag{2.7}$$

i.e. first take the limit as blocklength goes to infinity, then as probability of error goes to zero. Note that the base of the logarithm should agree with the base of the

logarithm in the mutual information expression for C.

2.2 Known Results: Capacity and Capacity Achieving Output Distribution

First we review a few known results on the MIMO-BF channel. Since the channel is memoryless, the capacity is given by maximizing the mutual information subject to a second moment constraint on the input distribution, i.e.

$$C = \frac{1}{T} \max_{P_X : \mathbb{E}[\|X\|_F^2] \le TP} I(X; Y, H) \,. \tag{2.8}$$

It was shown by Telatar [9] that whenever distribution of H is isotropic, the input $X \in \mathbb{R}^{n_t \times T}$ with entry i, j given

$$X_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \frac{P}{n_t}\right) , \tag{2.9}$$

is a maximizer. Throughout, we will refer to the distribution with i.i.d. Gaussian entries as in (2.9) is the "Telatar input distribution". If someone had no idea about fading channels, and were told to guess the capacity achieving input distribution, this is probably the distribution you would guess. Plugging this distribution into (2.8) yields the capacity formula

$$C(P) = \frac{1}{2}\mathbb{E}\left[\log\det\left(I_{n_r} + \frac{P}{n_t} HH^T\right)\right] \tag{2.10}$$

$$= \sum_{i=1}^{n_{\min}} \mathbb{E}\left[C_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right] , \tag{2.11}$$

where $C_{AWGN}(P) = \frac{1}{2}\log(1+P)$ is the capacity of the additive white Gaussian noise (AWGN) channel with SNR P, $n_{\min} = \min(n_r, n_t)$ is the minimum of the transmit and receive antennas, and $\{\Lambda_i^2, i = 1, \ldots, n_{\min}\}$ are eigenvalues of HH^T. Note that it is common to think that as $P \to \infty$ the expression (2.11) scales as $n_{\min} \log P$, but this is only true if $\mathbb{P}[\operatorname{rank} H = n_{\min}] = 1$. The main non-trivial step in the proof of (2.10) is that if $X \in \mathbb{R}^{n_t}$ is zero-mean and has covariance matrix Σ, then the entropy of X satisfies $H(X) \le \log\det(2\pi e\Sigma)$, with equality if and only if X is i.i.d. Gaussian with covariance matrix Σ.

The distribution induced by a caid at the channel output (Y, H) is called the *capacity achieving output distribution* (caod). A classical fact is that, while there may be many caids, the caod is unique, e.g. [10, Section 4.4]. This fact gives a convenient way to finding caids – if an input distribution induces the caod through the channel, then it is a caid.

From (2.9) we infer that the caod is given by

$$P^*_{Y,H} \triangleq P_H P^*_{Y|H}\,, \tag{2.12}$$

$$P^*_{Y|H} \triangleq \prod_{j=1}^{T} P^*_{Y^{(j)}|H}\,, \tag{2.13}$$

$$P^*_{Y^{(j)}|H=h} \triangleq \mathcal{N}\left(0, I_{n_r} + \frac{P}{n_t} hh^T\right)\,, \tag{2.14}$$

Here, $Y = [Y^{(1)}, \ldots, Y^{(T)}]$, where $Y^{(j)}$ is j-th column of Y, which, as we specified in (3), is a $n_r \times T$ matrix. I.e. the caod has i.i.d. columns, each column having conditional distribution given by (2.14).

Here, we make a few remarks about what the capacity formula (2.11) tells us:

1. This formula most significantly says that the capacity scales as approximately $n_{\min} \log(1+P)$. This result is enormously significant – we see that adding more antennas at both the transmitter and receiver can boost rate *linearly*. Antennas are generally cheap and harmless, compared to increasing power, which only provides a *logarithmic* increase in rate, and is limited by regulations. Note that the channel model assumes independent fading between each transmit and receive antenna – if $n_{\min}$ is enormous, spatial coupling between antennas can occur, causing the independence assumption to break down, which prevents us from achieving arbitrarily large rates just by adding more antennas.

2. This formula is independent of T, hence from a capacity standpoint, a shorter or longer coherence time does not affect achievable rates.

3. By the identity

$$\det\left(I_{n_r} + \frac{P}{n_t} HH^T\right) = \det\left(I_{n_t} + \frac{P}{n_t} H^T H\right) \tag{2.15}$$

   we see that switching n_r and n_t has no effect on the expression, which is known as *reciprocity* – the channel from the transmitter to the receiver has the same capacity as the channel from the receiver to the transmitter.

4. As we will soon see, when $\text{rank}(H) \leq 1$ a.s., this channel has multiple capacity achieving input distributions. From the point of view of (2.11), all give the same performance. However, we will see that this is not true when looking at more refined asymptotics.

2.3 Capacity Achieving Input Distributions

A very interesting feature of the MIMO-BF channel is that it has many caids in the case where $\text{rank}(H) \leq 1$, whereas many commonly studied channels (e.g. BSC, BEC, AWGN) have a unique caid. While a capacity achieving input distribution is

a random construction for demonstrating the capacity of a channel, the structure of the caid gives insights into the statistics of good codes. For example, a caid for the MIMO-BF channel has i.i.d. Gaussian entries, so one would expect a scheme like V-BLAST (see [11]) that sends independent symbols over each antenna at each time step to be good. For a more formal statement, see [12], where it is shown that for a code to be good, the output distribution it induces must be indistinguishable from the caod.

The following theorem characterizes the set of caids for the MIMO-BF channel. Somewhat surprisingly, for the case of rank-1 H (e.g. for MISO) there are multiple non-trivial jointly Gaussian caids with different correlation structures. For example, space-time block codes can achieve the capacity in the rank 1 case, but do not achieve capacity when the rank is 2 or greater e.g. [13].

Theorem 1.

1. *Every caid X satisfies $\forall a \in \mathbb{R}^{n_t}, b \in \mathbb{R}^T$:*

$$\sum_{i=1}^{n_t}\sum_{j=1}^{T} a_i b_j X_{i,j} \sim \mathcal{N}\left(0, \frac{P}{n_t}\|a\|_2^2\|b\|_2^2\right) . \tag{2.16}$$

   If $\mathbb{P}[\text{rank}\, H \leq 1] = 1$ then condition (2.16) is also sufficient for X to be caid.

2. *Let $X = \begin{pmatrix} R_1 \\ \cdots \\ R_{n_t} \end{pmatrix}$ be decomposed into rows R_i. If X is a caid, then each $R_i \sim \mathcal{N}(0, \frac{P}{n_t} I_T)$ (i.i.d. Gaussian) and*

$$\mathbb{E}[R_i^T R_i] = \frac{P}{n_t} I_T, \qquad i = 1, \ldots, n_t \tag{2.17}$$

$$\mathbb{E}[R_i^T R_j] = -\mathbb{E}[R_j^T R_i], \qquad i \neq j\,. \tag{2.18}$$

   If X is jointly zero-mean Gaussian and $\mathbb{P}[\text{rank}\, H \leq 1] = 1$, then (2.17)-(2.18) are sufficient for X to be caid.

3. *Let $X = (C_1 \ldots C_T)$ be decomposed into columns C_j. If X is a caid, then each $C_j \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ (i.i.d. Gaussian) and*

$$\mathbb{E}[C_i C_i^T] = \frac{P}{n_t} I_{n_t}, \qquad i = 1, \ldots, T \tag{2.19}$$

$$\mathbb{E}[C_i C_j^T] = -\mathbb{E}[C_j C_i^T], \qquad i \neq j\,. \tag{2.20}$$

   If X is jointly zero-mean Gaussian and $\mathbb{P}[\text{rank}\, H \leq 1] = 1$, then (2.19)-(2.20) are sufficient for X to be caid.

4. *When $\mathbb{P}[\text{rank}\, H > 1] > 0$, any caid has pairwise independent rows:*

$$R_i \perp\!\!\!\perp R_j \sim \mathcal{N}\left(0, \frac{P}{n_t} I_T\right) \qquad \forall i \neq j \tag{2.21}$$

and in particular

$$X_{i,j} \perp\!\!\!\perp X_{k,l} \qquad \forall (i,j) \neq (k,l)\,. \tag{2.22}$$

Therefore, among jointly Gaussian X the i.i.d. $X_{i,j}$ is the unique caid.

5. *There exist non-Gaussian caids if and only if* $\mathbb{P}[\operatorname{rank} H \geq \min(n_t, T)] = 0$.

Remark 1. (Special case of rank-1 H) In the MISO case when $n_t > 1$ and $n_r = 1$ (or more generally, $\operatorname{rank} H \leq 1$ a.s.), there is not only a multitude of caids, but in fact they can have non-trivial correlations between entries of X (and this is ruled out by (2.22) for all other cases). As an example, for the $n_t = T = 2$ case, any of the following random matrix-inputs X (parameterized by $\rho \in [-1, 1]$) is a Gaussian caid:

$$X = \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & -\rho\xi_2 + \sqrt{1-\rho^2}\xi_3 \\ \xi_2 & \rho\xi_1 + \sqrt{1-\rho^2}\xi_4 \end{bmatrix}, \tag{2.23}$$

where $\xi_1, \xi_2, \xi_3, \xi_4 \sim \mathcal{N}(0,1)$ i.i.d.. In particular, there are caids for which not all entries of X are pairwise independent.

Remark 2. Another way to state conditions (2.17)-(2.18) is: all elements in a row (resp. column) are pairwise independent $\sim \mathcal{N}(0, \frac{P}{n_t})$ and each 2×2 minor has antipodal correlation for the two diagonals. In particular, if X is a caid, then X^T and any submatrix of X are caids too (for different n_t and T).

Proof. We will rely repeatedly on the following observations:

1. if A, B are two random vectors in $\mathbb{R}^n$ then for any $v \in \mathbb{R}^n$ we have

$$\forall v \in \mathbb{R}^n : v^T A \overset{d}{=} v^T B \quad \iff \quad A \overset{d}{=} B\,. \tag{2.24}$$

   This is easy to show by computing characteristic functions.

2. If A, B are two random vectors in $\mathbb{R}^n$ independent of $Z \sim \mathcal{N}(0, I_n)$, then

$$A + Z \overset{d}{=} B + Z \quad \iff \quad A \overset{d}{=} B\,. \tag{2.25}$$

   This follows from the fact that the characteristic function of Z is nowhere zero.

3. For two matrices $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ we have, $\forall v \in \mathbb{R}^n$:

$$v^T Q_1 v = v^T Q_2 v \iff Q_1 + Q_1^T = Q_2 + Q_2^T\,. \tag{2.26}$$

   This follows from the fact that a quadratic form that is zero everywhere on $\mathbb{R}^n$ must have all coefficients equal to zero.

Part 1 (necessity). Recall that the caod is unique and given by (2.12). Thus an input X is a caid iff for P_H-almost every $h_0 \in \mathbb{R}^{n_r \times n_t}$ we have

$$h_0 X + Z \overset{d}{=} h_0 G + Z\,, \tag{2.27}$$

where G is an $n_t \times T$ matrix with i.i.d. $\mathcal{N}(0, P/n_t)$ entries (for sufficiency, just write $I(X; Y, H) = h(Y|H) - h(Z)$ with $h(\cdot)$ denoting differential entropy). We will argue next that (2.27) implies (under isotropy assumption on P_H) that

$$\forall a \in \mathbb{R}^{n_t}: \quad a^T X \stackrel{d}{=} a^T G. \tag{2.28}$$

From (2.24), (2.28) is equivalent to $\sum_{i,j} a_i b_j X_{i,j} \stackrel{d}{=} \sum_{i,j} a_i b_j G_{i,j}$ for all $b \in \mathbb{R}^{n_t}$.

Let E_0 be a P_H-almost sure subset of $\mathbb{R}^{n_t \times n_r}$ for which (2.27) holds. Let $O(n) = \{U \in \mathbb{R}^{n \times n} : U^T U = U U^T = I_n\}$ denote the group of orthogonal matrices, with the topology inherited from $\mathbb{R}^{n \times n}$. Let $\{U_k\}$ and $\{V_k\}$ for $k \in \{1, 2, \ldots\}$ be countable dense subsets of $O(n_t)$ and $O(n_r)$, respectively. (These exist since $\mathbb{R}^{n^2}$ is a second-countable topological space). By isotropy of P_H we have $P_H[U_k(E_0)V_l] = 1$ and therefore

$$E \stackrel{\triangle}{=} E_0 \cap \bigcap_{k=1,l=1}^{\infty} U_k(E_0)V_l \tag{2.29}$$

is also almost sure: $P_H[E] = 1$, since E is the intersection of countably many almost sure sets. Here, $U_k(E_0)$ denotes the image of E_0 under U_k. By assumption (4), E must contain a non-zero element h_0, for otherwise we would have $P_H[0] = 1$, contradicting (4). Consequently, $h_0 \in U_k(E_0)V_l$ for all k, l, and so $U_k^{-1} h_0 V_l^{-1} \in E_0$ for all k, l. Since for $U \in O(n)$, the map $U \mapsto U^{-1}$ is a bijective continuous transformation of $O(n)$, we have that $\{U_k^{-1}\}$ and $\{V_l^{-1}\}$ are also countable dense subsets of $O(n_t)$ and $O(n_r)$, respectively. From (2.25) and (2.27) along with the definition of E_0, we conclude that

$$U_k^{-1} h_0 V_l^{-1} X \stackrel{d}{=} U_k^{-1} h_0 V_l^{-1} G \qquad \forall k, l.$$

Arguing by continuity and using the density of $\{U_k^{-1}\}$ and $\{V_l^{-1}\}$, this implies also

$$U h_0 V X \stackrel{d}{=} U h_0 V G \qquad \forall U \in O(n_t), V \in O(n_r). \tag{2.30}$$

In particular, for any $a \in \mathbb{R}^{n_t}$ there must exist a choice of U, V such that $U h_0 V$ has the top row equal to $c_0 a^T$ for some constant $c_0 > 0$. Choosing these U, V in (2.30) and comparing distributions of top rows, we conclude (2.28) after scaling by $1/c_0$.

Part 1 (sufficiency). Suppose $\mathbb{P}[\operatorname{rank} H \leq 1] = 1$. Then our goal is to show that (2.28) implies that X is a caid. To that end, it is sufficient to show $h_0 X \stackrel{d}{=} h_0 G$ for all rank-1 h_0. In the special case

$$h_0 = \begin{pmatrix} a^T \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

the claim follows directly from (2.28). Every other rank-1 h_0' can be decomposed as

$h_0' = Uh_0$ for some matrix U, and thus again we get $Uh_0X \stackrel{d}{=} Uh_0G$, concluding the proof.

Parts 2 and 3 (necessity). From part 1 we have that for every a, b we must have $a^TXb \sim \mathcal{N}(0, \|a\|_2^2\|b\|_2^2\frac{P}{n_t})$. Computing expected square we get

$$\mathbb{E}\left[(a^TXb)^2\right] = \frac{P}{n_t}\left(\sum_i a_i^2\right)\left(\sum_j b_j^2\right). \tag{2.31}$$

Thus, expressing the left-hand side in terms of rows R_i as $a^TX = \sum_i a_iR_i$ we get

$$b^T\left\{\mathbb{E}\left[\left(\sum_i a_iR_i\right)^T\left(\sum_i a_iR_i\right)\right]\right\}b = b^T\left(\sum_i a_i^2 I_T\right)b\,,$$

and thus by (2.26) we conclude that for all a:

$$\mathbb{E}\left[\left(\sum_i a_iR_i\right)^T\left(\sum_i a_iR_i\right)\right] = \left(\sum_i a_i^2\right)I_T\,.$$

Each entry of the $T \times T$ matrices above is a quadratic form in a and thus again by (2.26) we conclude (2.17)-(2.18). Part 3 is argued similarly with roles of a and b interchanged.

Parts 2 and 3 (sufficiency). When H is (at most) rank-1, we have from part 1 that it is sufficient to show that $a^TXb \sim \mathcal{N}(0, \|a\|_2^2\|b\|_2^2\frac{P}{n_t})$. When X is jointly zero-mean Gaussian, we have a^TXb is zero-mean Gaussian and so we only need to check its second moment satisfies (2.31). But as we just argued, (2.31) is equivalent to either (2.17)-(2.18) or (2.19)-(2.20).

Part 4. As in Part 1, there must exist $h_0 \in \mathbb{R}^{n_r \times n_t}$ such that (2.30) holds and $\operatorname{rank} h_0 > 1$. Thus, by choosing U, V we can diagonalize h_0 and thus we conclude any pair of rows R_i, R_j must be independent.

Part 5. This part is never used in subsequent parts of the paper, so we only sketch the argument and move the most technical part of the proof to Appendix A. Let $\ell = \max\{r : \mathbb{P}[\operatorname{rank} H \geq r] > 0\}$. Then arguing as for (2.30) we conclude that X is a caid if and only if for any h with $\operatorname{rank} h \leq \ell$ we have

$$hX \stackrel{d}{=} hG\,.$$

In other words, we have

$$\sum_{i,j} a_{i,j}X_{i,j} \stackrel{d}{=} \sum_{i,j} G_{i,j} \qquad \forall a \in \mathbb{R}^{n_t \times T} : \operatorname{rank} a \leq \ell\,. \tag{2.32}$$

If $\ell = \min(n_t, T)$, then rank condition on a is not active and hence, we conclude by (2.24) that $X \stackrel{d}{=} G$. So assume $\ell < \min(n_t, T)$. Note that (2.32) is equivalent to

the condition on characteristic function of X as follows:

$$\mathbb{E}\left[e^{i\sum_{i,j} a_{i,j}X_{i,j}}\right] = e^{-\frac{P}{2n_t}\sum_{i,j} a_{i,j}^2} \qquad \forall a : \operatorname{rank} a \le \ell\,. \tag{2.33}$$

It is easy to find polynomial (in $a_{i,j}$) that vanishes on all matrices of rank $\le \ell$ (e.g. take the product of all $\ell \times \ell$ minors). Then Proposition 47 in Appendix A constructs non-Gaussian X satisfying (2.33) and hence (2.32). □

2.4 Information Density and its Moments

In finite blocklength analysis, a key object of study is the information density, along with its first and second moments. In this section we'll find expressions for these moments, along with showing when the information density is asymptotically normal. First, we give a short description of the information density.

In general, for a channel given by $P_{Y|X}$, and input distribution P_X that induces P_Y, the information density is given by

$$i(x;y) = \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}\,. \tag{2.34}$$

The most intuitive interpretation of this is as the log likelihood ratio in a test of dependence between X and Y, i.e. the binary hypothesis test

$$H_0 : Z \sim P_X P_Y \tag{2.35}$$
$$H_1 : Z \sim P_{XY}\,. \tag{2.36}$$

By the Neyman-Pearson Lemma, we know that the optimal test in the non-Bayesian setting is given by thresholding the log likelihood ratio, i.e. if we observe n sample $Z_1, \ldots, Z_n$ i.i.d., the optimal test is given by

$$\text{if } i(X^n;Y^n) \ge \gamma \text{ output } H_1 \tag{2.37}$$
$$\text{if } i(X^n;Y^n) < \gamma \text{ output } H_0 \tag{2.38}$$

In this way, we can view many decoders in information theory as running M binary dependence tests of this form.

To find the information density for the MIMO-BF channel, it will be convenient to assume that the matrix H is represented as

$$H = U\Lambda V^T\,, \tag{2.39}$$

where U, V are uniformly distributed on $O(n_r)$ and $O(n_t)$ (which follows from the isotropic assumption on H), respectively,[1] and Λ is the $n_r \times n_t$ diagonal matrix with diagonal entries $\{\Lambda_i, i = 1, \ldots, n_{\min}\}$. Joint distribution of $\{\Lambda_i\}$ depends on the fading

[1]Recall that $O(m) = \{A \in \mathbb{R}^{m\times m} : AA^T = A^TA = I_m\}$ is the space of all orthogonal matrices. This space is compact in a natural topology and admits a Haar probability measure.

model. It does not matter for our analysis whether Λ_i's are sorted in some way, or permutation-invariant.

For the MIMO-BF channel, let P_{YH}^* denote the caod (2.12). To compute the information density with respect to P_{YH}^* (for a single T-block of symbols) as defined in (2.12), denote $y = hx + z$ and write an SVD decomposition for matrix h as

$$h = u\lambda v^T ,$$

where $u \in O(n_r)$, $v \in O(n_t)$ and λ is an $n_r \times n_t$ matrix which is zero except for the diagonal entries, which are equal to $\lambda_1, \ldots, \lambda_{n_{\min}}$. Note that this representation is unique up to permutation of $\{\lambda_j\}$, but the choice of this permutation will not affect any of the expressions below. With this decomposition we have:

$$\begin{aligned} i(x; y, h) \triangleq & \frac{T}{2} \log\det\left(I_{n_r} + \frac{P}{n_t} hh^T\right) \\ & + \frac{\log e}{2} \sum_{j=1}^{n_{\min}} \frac{\lambda_j^2 \|v_j^T x\|^2 + 2\lambda_j \langle v_j^T x, \tilde{z}_j\rangle - \frac{P}{n_t}\lambda_j^2 \|\tilde{z}_j\|^2}{1 + \frac{P}{n_t}\lambda_j^2} \end{aligned} \tag{2.40}$$

where we denoted by v_j the j-th column of V, and have set $\tilde{z} = u^T z$, with $\tilde{z}_j$ representing the j-th row of $\tilde{z}$. The definition naturally extends to blocks of length nT additively:

$$i(x^n; y^n, h^n) \triangleq \sum_{j=1}^{n} i(x_j; y_j, h_j) . \tag{2.41}$$

We compute the (conditional) mean of information density to get

$$D_n(x^n) \triangleq \frac{1}{nT} \mathbb{E}\left[i(X^n; Y^n, H^n) | X^n = x^n\right] \tag{2.42}$$

$$= C(P) + \frac{\sqrt{\frac{\eta_2}{2}}}{n_t nT} \sum_{j=1}^{n} (\|x_j\|_F^2 - TP) , \tag{2.43}$$

where we used the following simple fact:

Lemma 2. *Let $U \in \mathbb{R}^{1\times n_t}$ be uniformly distributed on the unit sphere, and $x \in \mathbb{R}^{n_t \times T}$ be a fixed matrix, then*

$$\mathbb{E}[\|Ux\|^2] = \frac{\|x\|_F^2}{n_t} \tag{2.44}$$

Proof. Note that by additivity of $\|Ux\|^2$ across columns, it is sufficient to consider the case $T = 1$, for which the statement is clear from symmetry. □

Remark 3. A simple consequence of Lemma 2 is $\mathbb{E}[\|Hx\|_F^2] = \mathbb{E}[\|H\|_F^2]\frac{\|x\|_F^2}{n_t}$, which follows from considering the SVD of H.

Finally, the following lemma computes the Berry Esseen constant. This is a technical result that will be needed for both the achievability and converse proofs.

Lemma 3. *Fix $x_1, \ldots, x_n \in \mathbb{R}^{n_t \times T}$ and let $W_j = i(x_j; Y_j, H_j)$, where Y_j, H_j are distributed as the output of channel (2.1) with input x_j. Define the Berry-Esseen ratio*

$$B_n(x^n) \triangleq \sqrt{n} \frac{\sum_{j=1}^n \mathbb{E}\left[|W_j - \mathbb{E}[W_j]|^3\right]}{\left(\sum_{j=1}^n \mathrm{Var}(W_j)\right)^{3/2}} . \tag{2.45}$$

Then whenever $\sum_{j=1}^n \|x_j\|_F^2 = nTP$ and $\max_j \|x_j\|_F \le \delta n^{\frac{1}{4}}$ we have

$$B_n(x^n) \le K_1 \delta^2 \sqrt{n} + K_2 n^{1/4} + \frac{K_3}{n^{1/2}}$$

where $K_1, K_2, K_3 > 0$ are constants which only depend on channel parameters but not x^n or n.

Remark 4. Lemma 3 contains the condition that $\max_j \|x_j\|_F \le \delta n^{\frac{1}{4}}$, which is why we need this assumption to hold in the converse statement. In words: in order for the information density to be asymptotically normal, too much power cannot be place in any single time slot. Intuitively, this scaling is the threshold that ruins the "sum of many small independent quantities" in the central limit theorem.

Proof of Lemma 3. We begin with upper bounding the numerator in (2.45), i.e.

$$\sum_{j=1}^n \mathbb{E}\left[|W_j - \mathbb{E}[W_j]|^3\right] . \tag{2.46}$$

The information density is given by

$$i(x; y, h) = \frac{1}{2} \log\det(\Sigma) - \frac{1}{2} \sum_{j=1}^T \|y_j - h x_j\|^2 + \frac{1}{2} \mathrm{tr}\left(y^T \Sigma^{-1} y\right) \tag{2.47}$$

where

$$\Sigma = I_{n_r} + \frac{P}{n_t} H H^T . \tag{2.48}$$

Define $W = i(x; Y, H)$ under the distribution $Y = Hx + Z$. (2.47) reduces to

$$\begin{aligned} W &= \frac{T}{2} \log\det(\Sigma) - \frac{1}{2}\|Z\|_F^2 \\ &\quad + \frac{1}{2} \mathrm{tr}\left(x^T H^T \Sigma^{-1} H x + 2 x^T H^T \Sigma^{-1} Z + Z^T \Sigma^{-1} Z\right) \qquad (2.49) \\ &= c(H, Z) + \frac{1}{2} \mathrm{tr}\left(x^T H^T \Sigma^{-1} H x\right) + \mathrm{tr}\left(x^T H^T \Sigma^{-1} Z\right) \qquad (2.50) \end{aligned}$$

where the scalar random variable

$$c(H, Z) = \frac{T}{2} \log\det(\Sigma) - \frac{1}{2}\|Z\|_F^2 + \frac{1}{2} \mathrm{tr}\left(Z^T \Sigma^{-1} Z\right) \tag{2.51}$$

is the sum of all the terms that do not depend on x. Note that

$$\mathbb{E}\mathrm{tr}\left(x^T H^T \Sigma^{-1} H x\right) = \mathrm{tr}\left(x^T \mathbb{E}\left[H^T \Sigma^{-1} H\right] x\right) \tag{2.52}$$

$$\mathbb{E}\mathrm{tr}\left(x^T H^T \Sigma^{-1} Z\right) = 0\,. \tag{2.53}$$

Therefore, the "centered" information density is

$$\begin{aligned} & W - \mathbb{E}[W] \\ & = c_0(H,Z) - \mathbb{E}[c(H,Z)] \\ & + \frac{1}{2}\mathrm{tr}\left(x^T\left(H^T\Sigma^{-1}H - \mathbb{E}\left[H^T\Sigma^{-1}H\right]\right)x\right) \\ & + \mathrm{tr}\left(x^T H^T \Sigma^{-1} Z\right) && (2.54) \\ & = c_0(H,Z) + \mathrm{tr}\left(x^T A x\right) + \mathrm{tr}\left(x^T B\right) && (2.55) \end{aligned}$$

where

$$A = \frac{1}{2}\left(H^T\Sigma^{-1}H - \mathbb{E}\left[H^T\Sigma^{-1}H\right]\right) \tag{2.56}$$

$$B = H^T\Sigma^{-1}Z \tag{2.57}$$

$$c_0(H,Z) = c(H,Z) - \mathbb{E}[c(H,Z)]\,. \tag{2.58}$$

Hence we can upper bound the centered third moment as

$$\begin{aligned} & \mathbb{E}\left[|W - \mathbb{E}[W]|^3\right] \leq \\ & 3\underbrace{\mathbb{E}\left[|c_0(H,Z)|^3\right]}_{S_1} + 3\underbrace{\mathbb{E}\left[\left|\mathrm{tr}\left(x^T A x\right)\right|^3\right]}_{S_2} + 3\underbrace{\mathbb{E}\left[\left|\mathrm{tr}\left(x^T B\right)\right|^3\right]}_{S_3}\,. \end{aligned} \tag{2.59}$$

We now proceed to upper bound each term individually. First S_2,

$$\begin{aligned} S_2 &= \mathbb{E}\left[\left|\mathrm{tr}\left(x^T A x\right)\right|^3\right] && (2.60) \\ &= \frac{1}{8}\mathbb{E}\left[\left|x^T H^T\Sigma^{-1}Hx - x^T\mathbb{E}\left[H^T\Sigma^{-1}H\right]x\right|^3\right] && (2.61) \\ &\leq \frac{1}{8}\mathbb{E}\left[\left|x^T H^T\Sigma^{-1}Hx + x^T\mathbb{E}\left[H^T\Sigma^{-1}H\right]x\right|^3\right] && (2.62) \\ &\leq \frac{1}{8}\mathbb{E}\left[\left|\frac{2n_t}{P}\|x\|_F^2\right|^3\right] && (2.63) \\ &= \left(\frac{n_t}{P}\right)^3 \|x\|_F^6 && (2.64) \end{aligned}$$

where

- (2.62) follows since $H^T\Sigma^{-1}H$ is PSD, and $\mathbb{E}[H^T\Sigma^{-1}H]$ is also PSD as a non-negative combination of PSD matrices, so that both $x^T H^T\Sigma^{-1}Hx$ and $x^T\mathbb{E}[H^T\Sigma^{-1}H]x$ are non-negative

- (2.63) follows since $H^T\Sigma^{-1}H = VDV^T$ where

$$D = \operatorname{diag}\left(c\left(\Lambda_1^2\right), \ldots, c\left(\Lambda_{n_{\min}}^2\right), 0, \ldots, 0\right) \tag{2.65}$$

and $D \leq \frac{n_t}{P} I_{n_t}$ in the PSD ordering, so

$$x^T H^T \Sigma^{-1} H x \leq \frac{n_t}{P} x^T V V^T x = \frac{n_t}{P} \|x\|_F^2 \tag{2.66}$$

and

$$x^T \mathbb{E}\left[H^T \Sigma^{-1} H\right] x \leq \frac{n_t}{P} x^T \mathbb{E}\left[V V^T\right] x = \frac{n_t}{P} \|x\|_F^2 . \tag{2.67}$$

Now we bound S_3 from (2.59),

$$\begin{aligned}
S_3 &= \mathbb{E}\left[\left|\operatorname{tr}\left(x^T B\right)\right|^3\right] && (2.68)\\
&= \mathbb{E}\left[\left|\operatorname{tr}\left(x^T H^T \Sigma^{-1} Z\right)\right|^3\right] && (2.69)\\
&= \mathbb{E}\left[\left|\sum_{i=1}^{n_t}\sum_{j=1}^{T} \tilde{x}_{ij} Z_{ij} \frac{\Lambda_i}{1+\frac{P}{n_t}\Lambda_i^2}\right|^3\right] && (2.70)\\
&\leq n_t^2 T^2 \sum_{i=1}^{n_t}\sum_{j=1}^{T} \mathbb{E}\left[|\tilde{x}_{ij}|^3 |Z_{ij}|^3 \left|\frac{\Lambda_i}{1+\frac{P}{n_t}\Lambda_i^2}\right|^3\right] && (2.71)\\
&\leq \frac{n_t^2 T^2}{4} \left(\frac{n_t}{P}\right)^{3/2} \|x\|_F^3 && (2.72)
\end{aligned}$$

where

- In (2.70), define $\tilde{x} = V^T x$ and expand the trace.
- (2.71) follows from the triangle inequality, along with $|\sum_{i=1}^n a_i|^3 \leq n^2 \sum_{i=1}^n |a_i|^3$.
- (2.72) we have used $\mathbb{E}[|Z|^3] \leq 2$ for $Z \sim \mathcal{N}(0,1)$ along with the bound

$$\left|\frac{x}{1+ax^2}\right| \leq \frac{1}{2\sqrt{a}} . \tag{2.73}$$

Now notice that

$$\sum_{i=1}^{n_t}\sum_{j=1}^{T} |\tilde{x}_{ij}|^3 \leq \left(\sum_{i=1}^{n_t}\sum_{i=1}^{T} \tilde{x}_{ij}^2\right)^{3/2} \tag{2.74}$$

which can be viewed as the norm inequality $\|a\|_3 \leq \|a\|_2$ for $a \in \mathbb{R}^d$. Finally, we use $\|V^T x\|_F^2 = \|x\|_F^2$ for any orthogonal matrix V.

For the denominator in (2.45), the expression for $\frac{1}{T}\mathrm{Var}(W_j)$ is given in (5.7)-(5.11). Note that the final term (5.11) is non-negative, so we have the lower bound

$$\sum_{j=1}^{n} \mathrm{Var}(W_j) \geq K_1' n + K_2' \sum_{j=1}^{n} \left(\|x_j\|_F^2 - TP\right)^2 \tag{2.75}$$

$$\geq \max\left(nK_1', K_2' \sum_{j=1}^{n} \left(\|x_j\|_F^2 - TP\right)^2\right) \tag{2.76}$$

where

$$K_1' = T^2 \mathrm{Var}\left(C_r(H,P)\right) + T \sum_{i=1}^{n_{\min}} \mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right] \tag{2.77}$$

$$K_2' = T\left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right)^2 . \tag{2.78}$$

Hence $K_1' > 0$ whenever $P > 0$. Note that we use the assumption $\|x^n\|_F^2 = nTP$ freely here, as stated before. The lower bound on the variance (2.76), we obtain the upper bound

$$B_n(x^n) \leq \sqrt{n} \frac{\sum_{j=1}^{n} K_1 \|x_j\|_F^6 + K_2 \|x_j\|_F^3 + K_3}{\left(\max\left(nK_1', K_2' \sum_{j=1}^{n} \left(\|x_j\|_F^2 - TP\right)^2\right)\right)^{3/2}} \tag{2.79}$$

where all constants are non-negative. There are two cases based on which term achieves the max in the dominator. First, suppose

$$nK_1' \geq K_2' \sum_{j=1}^{n} \left(\|x_j\|_F^2 - TP\right)^2 . \tag{2.80}$$

Expanding the square yields

$$K_2' \sum_{j=1}^{n} \|x_j\|_F^4 \leq nK_1' + nT^2P^2K_2' . \tag{2.81}$$

Thus the terms in the numerator are bounded by

$$\sum_{j=1}^{n} \|x_i\|_F^6 \leq \left(\max_{i=1}^{n} \|x_i\|_F^2\right) \sum_{j=1}^{n} \|x_i\|_F^4$$

$$\leq n^{3/2}\delta^2(K_1' + T^2P^2K_2') \tag{2.82}$$

$$\sum_{j=1}^{n} \|x_i\|_F^3 \leq n^{1/4} \sum_{j=1}^{n} \|x_i\|_F^4$$

$$\leq n^{5/4}(K_1' + T^2P^2K_2') \tag{2.83}$$

where (2.82) uses the assumption $\|x_j\|_F \leq \delta n^{\frac{1}{4}}$. Applying this to B_n in (2.79), we see that in this case,

$$B_n(x^n) \leq \sqrt{n}\delta^2 C_1 + n^{1/4} C_2 + \frac{C_3}{n^{1/2}} \tag{2.84}$$

where the constant C_1, C_2, C_3 are non-negative constants.

Now take the case when

$$K_2' \sum_{j=1}^{n} \left(\|x_j\|_F^2 - TP \right)^2 \geq n K_1' \ . \tag{2.85}$$

Note that since $K_1' > 0$, in the case we must also have $K_2' > 0$ for the above inequality to hold. Let a be defined as follows

$$a = \frac{T^2 P^2}{T^2 P^2 + \frac{K_1'}{K_2'}} \ . \tag{2.86}$$

Here $a < 1$ since $K_1'/K_2' > 0$. Applying (2.85) yields

$$a \sum_{j=1}^{n} \|x_j\|_F^4 \geq a \left(n \frac{K_1'}{K_2'} + n T^2 P^2 \right) \tag{2.87}$$

$$\geq n T^2 P^2 \ . \tag{2.88}$$

With this, from (2.79) we obtain the following upper bound

$$B_n(x^n) \leq$$

$$\sqrt{n} \frac{\sum_{j=1}^{n} K_1 \|x_j\|_F^6 + K_2 \|x_j\|_F^3 + K_3}{K_2'^{3/2} \left((1-a) \sum_{j=1}^{n} \|x_j\|_F^4 + a \sum_{j=1}^{n} \|x_j\|_F^4 - n T^2 P^2 \right)^{3/2}} \tag{2.89}$$

$$\leq \sqrt{n} \frac{\sum_{j=1}^{n} K_1 \|x_j\|_F^6 + K_2 \|x_j\|_F^3 + K_3}{K_2'^{3/2} \left((1-a) \sum_{j=1}^{n} \|x_j\|_F^4 \right)^{3/2}} \ . \tag{2.90}$$

where (2.90) uses (2.88). Now, we can upper bound each term in (2.90) as, for the first term,

$$\frac{K_1 \sum_{j=1}^{n} \|x_j\|_F^6}{K_2'^{3/2} \left((1-a) \sum_{j=1}^{n} \|x_j\|_F^4 \right)^{3/2}}$$

$$\leq \frac{K_1 \max_{i=1,\ldots,n} \|x_i\|_F^2}{K_2'^{3/2} (1-a)^{3/2} \left(\sum_{j=1}^{n} \|x_j\|_F^4 \right)^{1/2}} \tag{2.91}$$

$$\leq \frac{K_1 \delta^2 n^{1/2}}{n^{1/2} K_2'^{3/2} (1-a)^{3/2} (T^2 P^2 + n K_1')^{1/2}} \ , \tag{2.92}$$

the second,

$$\frac{K_2 \sum_{j=1}^n \|x_j\|_F^3}{K_2'^{3/2} \left((1-a)\sum_{j=1}^n \|x_j\|_F^4\right)^{3/2}}$$

$$\leq \frac{K_2 n^{1/4}}{K_2'^{3/2}(1-a)^{3/2}\left(\sum_{j=1}^n \|x_j\|_F^4\right)^{3/4}} \tag{2.93}$$

$$\leq \frac{K_2 n^{1/4}}{n^{1/2} K_2'^{3/2}(1-a)^{3/2}(T^2P^2 + nK_1')^{3/4}}, \tag{2.94}$$

and the third,

$$\frac{K_3}{K_2'^{3/2} \left((1-a)\sum_{j=1}^n \|x_j\|_F^4\right)^{3/2}}$$

$$\leq \frac{K_3}{n^{3/2}\left(K_2'(1-a)(T^2P^2 + nK_1')\right)^{3/2}}, \tag{2.95}$$

where in (2.93) we have used $\sum_{i=1}^n a_i^3 \leq n^{1/4}\left(\sum_{i=1}^n a_i^4\right)^{3/4}$ (easily obtained from p-norm inequalities), and both (2.91) and (2.95) use the assumption $\|x_j\|_F \leq \delta n^{\frac{1}{4}}$. Using these bounds in (2.90), we obtain

$$B_n(x^n) \leq \sqrt{n}\delta^2 C_1' + n^{1/4} C_2' + \frac{C_3'}{n^{1/2}} \tag{2.96}$$

where C_1', C_2', C_3' are non-negative constants.

From (2.84) and (2.96), we conclude that

$$B_n(x^n) \leq \sqrt{n}\delta^2 C_1'' + n^{1/4} C_2'' + \frac{C_3''}{n^{1/2}}. \tag{2.97}$$

□

Chapter 3

Coding Theorems

In this chapter, we prove an achievability and converse result second order term in the MIMO-BF channel. The proof techniques for both achievability and converse have quite a few new elements – see the individual sections for more information. To the best of our knowledge, this is the first known channel where the dispersion has been shown to be a minimization over capacity achieving input distributions, as in (3.6). Note that these results appear in [14].

Note that seeing the conditional variance $\mathbb{E}\left[\mathrm{Var}(i(X;Y)|X)\right]$ of the information density is to be expected. A priori, one might expect the variance (unconditional) to give the dispersion. When $\mathcal{X}$ is finite, it turns out these are the same, since

$$\mathrm{Var}(i(X;Y)) = \mathbb{E}\left[\mathrm{Var}(i(X;Y)|X)\right] + \mathrm{Var}\left(\mathbb{E}[i(X;Y)|X]\right) \tag{3.1}$$

and, when P_X is capacity achieving,

$$\forall x \in \mathcal{X}, \quad \mathbb{E}[i(X;Y)|X=x] = D(P_{Y|X=x}||P_Y^*) = C \tag{3.2}$$

so that the second term in (3.1) vanishes, and $\mathrm{Var}(i(X;Y)) = \mathbb{E}[\mathrm{Var}(i(X;Y)|X)]$. However, when $\mathcal{X}$ is not finite, this is no longer the case.

The main result that will be proved in the following sections is the following theorem:

Theorem 4. *For the MIMO-BF channel, there exists an $(nT, M, \epsilon, P)_{CSIR}$ maximal probability of error code with $0 < \epsilon < 1/2$ satisfying*

$$\log M \geq nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + o(\sqrt{n}) \ . \tag{3.3}$$

Furthermore, for any $\delta_n \to 0$ there exists $\delta_n' \to 0$ so that every $(nT, M, \epsilon, P)_{CSIR}$ code with extra constraint that $\max_j \|x^j\|_F \leq \delta_n n^{1/4}$, must satisfy

$$\log M \leq nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + \delta_n'\sqrt{n} \tag{3.4}$$

where

$$C(P) \triangleq \frac{1}{2}\mathbb{E}\left[\log\det\left(I_{n_r} + \frac{P}{n_t}HH^T\right)\right] \tag{3.5}$$

$$V(P) \triangleq \inf_{P_X : I(X;Y|H)=C} \frac{1}{T}\mathbb{E}\left[\mathrm{Var}(i(X;Y,H)|X)\right] \tag{3.6}$$

and $i(x;y,h)$ *is given by* (2.40).

The proof is broken up into Theorem 9 and Theorem 12 for achievability and converse, respectively. Before giving those proofs, first we introduce some notation, definitions, and lemmas concerning hypothesis testing.

3.1 Binary and Composite Hypothesis Testing

Many finite blocklength results are derived by considering an optimal hypothesis between appropriate distributions. A *binary hypothesis test* $P_{Z|W} : \mathcal{W} \to \{0,1\}$ is a test that, given a sample w from a space $\mathcal{W}$, chooses (perhaps non-deterministically) one of two distributions P or Q that could have generated w. $Z = 1$ indicates that the test choose P to be the true distribution, and $Z = 0$ indicates the test chooses Q instead. This is sometimes written as

$$H_0 : W \sim Q \tag{3.7}$$

$$H_1 : W \sim P \tag{3.8}$$

Two types of errors can be made in a binary hypothesis test: we can mistakenly choose P when Q is the actual distribution, or we can choose Q when P is the true distribution. These errors depends on the choice of test $P_{Z|W}$, and in general are asymmetric. Here we will use the convention that we always consider the error when the test chooses P when the actual distribution is Q.

We define $\beta_\alpha(P,Q)$ to be the minimum error probability of all statistical tests $P_{Z|W}$ between distributions P and Q, given that the test chooses P when P is correct with at least probability α. Formally:

$$\beta_\alpha(P,Q) = \inf_{P_{Z|W} : \int_{\mathcal{W}} P_{Z|W} 1|w dP(w) \geq \alpha} \int_{\mathcal{W}} P_{Z|W}(1|w) dQ(w) \tag{3.9}$$

The *Neyman Pearson Lemma* tells us that an optimal test $P^*_{Z|W}$ achieving error β_α exists, and has the form of a ratio test, i.e.

$$\beta_\alpha(P,Q) = Q\left[\frac{dP}{dQ} > \gamma\right] \tag{3.10}$$

Where γ is chosen to satisfy

$$\alpha = P\left[\frac{dP}{dQ} > \gamma\right] \tag{3.11}$$

In a *composite hypothesis test*, P and Q are now parametric families of distributions, $\{P_{\theta_1}\}_{\theta_1 \in \Theta_1}$ and $\{Q_{\theta_2}\}_{\theta_2 \in \Theta_2}$, i.e.

$$H_0 : W \sim Q_{\theta_2} \text{ s.t. } \theta_2 \in \Theta_2 \tag{3.12}$$

$$H_1 : W \sim P_{\theta_1} \text{ s.t. } \theta_1 \in \Theta_1 \tag{3.13}$$

In words: the test sees a sample w and must decide whether the distribution generating that sample was from the P_{θ_1} family or the Q_{θ_2} family. Similar to the binary hypothesis testing case, we denote the minimum error probability of a test $P_{Z|W}$ given that the test chooses H_1 when H_1 is true *for any* $\theta_1 \in \Theta_1$. Formally:

$$\kappa_\tau(\Theta_1, \Theta_2) = \inf_{P_{Z|W} : \inf_{\theta_1 \in \Theta_1}\{\int_{\mathcal{W}} P_{Z|W}(1|w) dP_{\theta_1}(w) \geq \tau\}} \sup_{\theta_2 \in \Theta_2} \int_{\mathcal{W}} P_{Z|W}(1|w) dQ_{\theta_2} \tag{3.14}$$

Our main case of interest will be between the set of distributions $\{P_{Y|X=x}\}_{x \in F}$ and a single distribution Q_Y. We will denote the minimum error probability in the composite hypothesis test in this case as $\kappa_\tau(F, Q_Y)$.

Now that we have the basic definitions, we'll need a few bounds that will be used in the next section. First, we can lower bound a composite hypothesis test in terms of a binary hypothesis test. This is useful because often it is difficult to evaluate κ_τ, but for β_α the Neyman-Pearson lemma gives us the form of the optimal test.

Lemma 5. *For a composite hypothesis test between* $\{P_{Y|X=x}\}_{x \in F}$ *and* Q_Y, *where* $F \subset A = \text{support}(X)$, *for any distribution* $P_{\tilde{X}}$ *such that* $P_{\tilde{X}}[F] > 0$,

$$\kappa_\tau(F, Q_Y) \geq \beta_{\tau P_{\tilde{X}}[F]}(P_{\tilde{X}} \circ P_{Y|X}, Q_Y) \tag{3.15}$$

Here, $P_X \circ P_{Y|X} \triangleq \int P_{Y|X=x} dP_X(x)$.

Proof. Let $P_{Z|Y}$ be any test for the composite hypothesis test between $\{P_{Y|X=x}\}_{x \in F}$ and Q_Y satisfying

$$\inf_{x \in F} \sum_{y \in B} P_{Y|X}(y|x) P_{Z|Y}(1|y) \geq \tau \tag{3.16}$$

Where $Z = 1$ indicates the test chooses $\{P_{Y|X=x}\}_{x \in F}$. Then we use this test $P_{Z|Y}$ for testing P_Y vs Q_Y, where now $Z = 1$ indicates the test chooses P_Y. The corresponding

probability of choosing P_Y when P_Y is correct is (note $P_Y = P_X \circ P_{Y|X}$ by assumption)

$$\sum_{y\in B} P_Y(y)P_{Z|Y}(1|y) = \sum_{y\in B}\left(\sum_{x\in A} P_X(x)P_{Y|X}(y|x)\right)P_{Z|Y}(1|y) \tag{3.17}$$

$$\geq \sum_{y\in B}\left(\sum_{x\in F} P_X(x)P_{Y|X}(y|x)\right)P_{Z|Y}(1|y) \tag{3.18}$$

$$\geq \sum_{x\in F} P_X(x)\left(\inf_{x\in F}\sum_{y\in B} P_{Y|X}(y|x)P_{Z|Y}(1|y)\right) \geq P_X[F]\tau \tag{3.19}$$

Since this hold for all tests $P_{Z|Y}$ for the composite HT, it holds for the test achieving κ_τ. Since $\beta_{\tau P_X[F]}$ lower bounds the $\pi_{1|0}$ error of all test for P_Y vs Q_Y, it lower bounds κ_τ. □

Furthermore, we can lower bound β_α from a binary hypothesis test in terms of the divergence between $D(P||Q)$ using the data processing inequality:

Lemma 6. *For all distributions P, Q s.t. $P \ll Q$, and all $\alpha \in [0,1]$,*

$$\beta_\alpha(P,Q) \geq \exp\left(-\frac{D(P||Q) + h_B(\alpha)}{\alpha}\right) \tag{3.20}$$

Proof. Use the data processing inequality with the kernel $P_{Z|W}$ from our hypothesis test:

$$D(P||Q) \geq d(\alpha||\beta_\alpha) = -h(\alpha) + \alpha\log\frac{1}{\beta} + (1-\alpha)\log\frac{1}{1-\beta} \geq -h(\alpha) + \alpha\log\frac{1}{\beta} \tag{3.21}$$

Where $d(p||q)$ is the divergence between a Bernoulli(p) and Bernoulli(q) distribution. The lemma follows from solving for β_α. □

Finally, we are interested the case when P and Q are product distribution $P = \prod_{i=1}^n P_i$ and $Q = \prod_{i=1}^n Q_i$. When this is the case, with a few regularity conditions we can expand β_α in terms of it's dependence on n by the following lemma from [15, Lemma 14], which we give here

Lemma 7. *Let $P = \prod_{i=1}^n P_i$ and $Q = \prod_{i=1}^n Q_i$ with $P_i \ll Q_i$ be two measures on a measurable space $\mathcal{A}^n$ such that the third moment of $\log\frac{dP}{dQ}$ is bounded, then*

$$\log\beta_\alpha(P,Q) = -nD_n - \sqrt{nV_n}Q^{-1}(\alpha) + o(\sqrt{n}) \tag{3.22}$$

Where

$$D_n = \frac{1}{n}\sum_{i=1}^{n} D(P_i||Q_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\log \frac{dP_i}{dQ_i}\right] \tag{3.23}$$

$$V_n = \frac{1}{n}\sum_{i=1}^{n} V(P_i||Q_i) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\left(\log \frac{dP_i}{dQ_i}\right) \tag{3.24}$$

The proof is an application of the Berry-Esseen Theorem, which quantifies the error in approximating the CDF of a sum of independent random variables by a Gaussian distribution as in the Central Limit Theorem.

3.2 Achievability

In this section, we prove the achievability side of the coding theorem for the MIMO-BF channel. We will rely on the $\kappa\beta$ bound [16, Theorem 25], quoted here:

Theorem 8 ($\kappa\beta$ bound). *Given a channel $P_{Y|X}$ with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$, for any distribution Q_Y on $\mathcal{B}$, any non-empty set $F \subset \mathcal{A}$, and ϵ, τ such that $0 < \tau < \epsilon < 1/2$, there exists and (M, ϵ)-max code satisfying*

$$M \geq \frac{\kappa_\tau(F, Q_Y)}{\sup_{x\in F} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)} \,. \tag{3.25}$$

The art of applying this theorem is in choosing F and Q_Y appropriately. The intuition in choosing these is as follows: although we know the distributions in the collection $\{P_{Y|X=x}\}_{x\in F}$, we do not know which x is actually true in the composite, so if Q_Y is in the "center" of the collection, then the two hypotheses can be difficult to distinguish, making the numerator large. However, for a given x, $P_{Y|X=x}$ vs Q_Y may still be easily to distinguish, making the denominator small. The main principle for applying the $\kappa\beta$-bound is thus: Choose F and Q_Y such that $P_{Y|X=x}$ vs Q_Y is easy to distinguish for any given x, yet the composite hypothesis $Y \sim \{P_{Y|X=x}\}_{x\in F}$ is hard to distinguish from a simple one $Y \sim Q_Y$.

The main theorem of this section gives achievable rates for the MIMO-BF channel, as follows:

Theorem 9. *Fix an arbitrary caid P_X on $\mathbb{R}^{n_t \times T}$ and let*

$$V' \triangleq \frac{1}{T}\mathbb{E}\left[\mathrm{Var}(i(X;Y,H)|X)\right] = \mathbb{E}\left[V_1(X)\right], \tag{3.26}$$

where $V_1(x)$ is introduced in Proposition 18. Then we have

$$\log M^*(nT, \epsilon, P) \geq nTC(P) - \sqrt{nTV'}Q^{-1}(\epsilon) + o(\sqrt{n}) \tag{3.27}$$

with $C(P)$ given by (2.10).

Proof. Let $\tau > 0$ be a small constant (it will be taken to zero at the end). We apply the $\kappa\beta$ bound (3.25) with auxiliary distribution $Q_Y = (P^*_{Y,H})^n$, where $P^*_{Y,H}$ is the caod (2.12), and the set F_n is to be specified shortly. Recall notation $D_n(x^n)$, $V_n(x^n)$ and $B_n(x^n)$ from (2.42), (5.6) and (2.45). For any x^n such that $B_n(x^n) \le \tau\sqrt{n}$, we have from [17, Lemma 14],

$$\begin{aligned} &-\log \beta_{1-\epsilon+\tau}(P_{Y^nH^n|X^n=x^n}, P^{*n}_{YH}) \ge \\ &\quad nTD_n(x^n) + \sqrt{nTV_n(x^n)}Q^{-1}(1-\epsilon-2\tau) - \log\frac{1}{\tau} - K' \end{aligned} \tag{3.28}$$

where K' is a constant that only depends on channel parameters. We mention that obtaining (3.28) from [17, Lemma 14] also requires that $V_n(x^n)$ be bounded away from zero by a constant, which holds since in the expression for $V_n(x^n)$ in Proposition 18, the term (5.8) is strictly positive, term (5.9) will vanish, and terms (5.10) and (5.11) are both non-negative.

Considering (3.28), our choice of the set F_n should not be surprising:

$$\begin{aligned} &F_n \triangleq \\ &\left\{x^n : \|x^n\|_F^2 = nTP, V_n(x^n) \le V' + \tau, \max_j \|x_j\|_F \le \delta n^{\frac{1}{4}}\right\} \end{aligned} \tag{3.29}$$

where $\delta = \delta(\tau) > 0$ is chosen so that Lemma 3 implies $B_n(x^n) \le \tau\sqrt{n}$ for any $x^n \in F_n$. Under this choice from (3.28), (2.43) and Lemma 3 we conclude

$$\begin{aligned} &\sup_{x^n \in F_n} \log \beta_{1-\epsilon+\tau}(P_{Y^nH^n|X^n=x^n}, P^{*n}_{YH}) \le \\ &-nTC(P) + \sqrt{nT(V'+\tau)}Q^{-1}(\epsilon - 2\tau) + K'', \end{aligned} \tag{3.30}$$

where $K'' = K' + \log\frac{1}{\tau}$.

To lower bound the numerator $\kappa_\tau(F_n, P^{*n}_{Y,H})$ we first state two auxiliary lemmas, whose proofs follow. The first, Lemma 10, shows that the output distribution induced by an input distribution that is uniform on the sphere is "similar" (in the sense of divergence) to the n-fold product of the caod.

Lemma 10. *Fix an arbitrary caid P_X and let X^n have i.i.d. components $\sim P_X$. Let*

$$\tilde{X}^n \triangleq \frac{X^n}{\|X^n\|_F}\sqrt{nTP} \tag{3.31}$$

where $\|X^n\|_F = \sqrt{\sum_{t=1}^n \|X_j\|_F^2}$. Then

$$D(P_{Y^nH^n|X^n} \circ P_{\tilde{X}^n} || P^{*n}_{Y,H}) \le \frac{TP\log e}{n_t}\mathbb{E}\left[\|H\|_F^2\right], \tag{3.32}$$

where $P^{*n}_{Y,H}$ is the n-fold product of the caod (2.12).

The second, Lemma 11, shows that a uniform distribution on the sphere has nearly

all of its mass in F_n as $n \to \infty$.

Lemma 11. *With $\tilde{X}^n$ as in Lemma 10 and set F_n defined as in (3.29) (with arbitrary $\tau > 0$ and $\delta > 0$) we have as $n \to \infty$,*

$$\mathbb{P}[\tilde{X}^n \in F_n] \to 1$$

Denote the right-hand side of (3.32) by K_1 and consider the following chain:

$$\begin{aligned}
\kappa_\tau(F_n, Q_{Y^n}) &\geq \exp\left(-\frac{D(P_{Y^n H^n|X^n} \circ P_{\tilde{X}^n} || Q_{Y^n}) + \log 2}{\tau P_{\tilde{X}^n}[F_n]}\right) && (3.33)\\
&\geq \exp\left(-\frac{K_1 + \log 2}{\tau P_{\tilde{X}^n}[F_n]}\right) && (3.34)\\
&= \exp\left(-\frac{K_1 + \log 2}{\tau + o(1)}\right) && (3.35)\\
&\geq K_2(\tau)\,, && (3.36)
\end{aligned}$$

where (3.33) follows from Lemmas 5 and 6 with $P_{\tilde{X}^n}$ as in Lemma 10, (3.34) is from Lemma 10, (3.35) is from Lemma 11, and in (3.36) we introduced a τ-dependent constant K_2.

Putting (3.30) and (3.36) into the $\kappa\beta$-bound we obtain

$$\begin{aligned}
&\log M^*(nT, \epsilon, P) \geq \\
&\quad nTC(P) - \sqrt{nT(V' + \tau)}Q^{-1}(\epsilon - 2\tau) - K'' - K_2(\tau)\,.
\end{aligned}$$

Taking $n \to \infty$ and then $\tau \to 0$ completes the proof. □

Now we prove the two lemmas used in the Theorem.

Proof of Lemma 10. In the case of no-fading ($H_j = 1$) and SISO, this Lemma follows from [18, Proposition 2]. Here we prove the general case. Let us introduce an auxiliary channel acting on X_j as follows:

$$\tilde{Y}_j = H_j \frac{X_j}{\|X^n\|_F}\sqrt{nTP} + Z_j, \qquad j = 1, \ldots, n \tag{3.37}$$

With this notation, consider the following chain:

$$
\begin{aligned}
&D(P_{Y^nH^n|X^n} \circ P_{\tilde{X}^n} || P^{*n}_{Y,H}) \\
&\quad = D(P_{\tilde{Y}^nH^n|X^n} \circ P_{X^n} || P^{*n}_{Y,H}) && (3.38)\\
&\quad = D(P_{\tilde{Y}^nH^n|X^n} \circ P_{X^n} || P_{Y^nH^n|X^n} \circ P_{X^n}) && (3.39)\\
&\quad = D(P_{\tilde{Y}^nH^n|X^n} || P_{Y^nH^n|X^n} | P_{X^n}) && (3.40)\\
&\quad = D(P_{\tilde{Y}^n|H^n,X^n} || P_{Y^n|H^n|X^n} | P_{X^n} P_{H^n}) && (3.41)\\
&\quad = \frac{\log e}{2} \mathbb{E}\left[\left(1 - \frac{\sqrt{nTP}}{\|X^n\|_F}\right)^2 \sum_{t=1}^{n} \|H_j X_j\|_F^2 \right] && (3.42)\\
&\quad = \frac{\log e}{2n_t} \mathbb{E}[\|H\|_F^2] \mathbb{E}\left[\left(\|X^n\|_F - \sqrt{nTP} \right)^2 \right] && (3.43)\\
&\quad = \frac{\log e}{n_t} \mathbb{E}[\|H\|_F^2] (nTP - \sqrt{nTP} \mathbb{E}\left[\|X^n\|_F\right]) && (3.44)
\end{aligned}
$$

where (3.38) is by clear from (3.37), (3.39) follows since P_X is a caid, (3.40)-(3.41) are standard identities for divergence, (3.42) follows since both $\tilde{Y}_j$ and Y_j are unit-variance Gaussians and $D(\mathcal{N}(0,1)\|\mathcal{N}(a,1)) = \frac{a^2 \log e}{2}$, (3.43) is from Lemma 2 (see Remark 3) and (3.44) is just algebra along with the assumption that $\mathbb{E}\left[\|X^n\|_F^2\right] = nTP$.

It remains to lower bound the expectation $\mathbb{E}\left[\|X^n\|_F\right]$. Notice that for any uncorrelated random variables $B_t \geq 0$ with mean 1 and variance 2 we have

$$
\mathbb{E}\left[\sqrt{\frac{1}{n} \sum_{t=1}^{n} B_t } \right] \geq 1 - \frac{1}{n}, \tag{3.45}
$$

which follows from $\sqrt{x} \geq \frac{3x - x^2}{2}$ for all $x \geq 0$ and simple computations. Next consider the chain:

$$
\begin{aligned}
\mathbb{E}[\|X^n\|_F] &= \mathbb{E}\left[\sqrt{ \sum_{i,j} \sum_{t=1}^{n} (X_t)_{i,j}^2 } \right] && (3.46)\\
&\geq \sqrt{\frac{n}{n_t T}} \sum_{i,j} \mathbb{E}\left[\sqrt{ \frac{1}{n} \sum_{t=1}^{n} (X_t)_{i,j}^2 } \right] && (3.47)\\
&= \sqrt{nTP} \left(1 - \frac{1}{n}\right) && (3.48)
\end{aligned}
$$

where in (3.48) we used the fact that for any caid, $\{(X_t)_{i,j}, t = 1, \ldots n\} \sim \mathcal{N}(0, P/n_t)$ i.i.d. (from Theorem 1) and applied (3.45) with $B_t = \frac{(X_t)_{i,j}^2 n_t}{P}$. Putting together (3.44) and (3.48) completes the proof. □

Proof of Lemma 11. Note that since $\|X^n\|_F^2$ is a sum of i.i.d. random variables, we

have $\frac{\|X^n\|_F}{\sqrt{nTP}} \to 1$ almost surely. In addition we have

$$\mathbb{E}\left[\|X_1\|_F^8\right] \le (n_t T)^3 \sum_{i,j} \mathbb{E}\left[(X_1)_{i,j}^8\right] \triangleq K\,,$$

where we used the fact (Theorem 1) that X_1's entries are Gaussian. Then we have from independence of X_j's and Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}[\max_j \|X_j\|_F \le \delta' n^{\frac{1}{4}}] &= \mathbb{P}[\|X_1\|_F \le \delta' n^{\frac{1}{4}}]^n \\ &\ge \left(1 - \frac{K}{\delta'^8 n^2}\right)^n \to 1\end{aligned}$$

as $n \to \infty$. Consequently,

$$\begin{aligned}&\mathbb{P}\left[\max_j \|\tilde{X}_j\|_F \le \delta n^{\frac{1}{4}}\right] \ge \\ &\quad \mathbb{P}\left[\max_j \|X_j\|_F \le \frac{\delta}{2} n^{\frac{1}{4}}\right] - \mathbb{P}\left[\frac{\|X^n\|_F}{\sqrt{nTP}} < \frac{1}{2}\right] \to 1\end{aligned}$$

as $n \to \infty$.

Next we analyze the behavior of $V_n(\tilde{X}^n)$. From Proposition 18 we see that, due to $\|\tilde{X}^n\|_F^2 = nTP$, the term (5.9) vanishes, while (5.10) simplifies. Overall, we have

$$\begin{aligned}&V_n(\tilde{X}^n) = \\ &K + \left(\frac{nTP}{\|X^n\|_F^2}\right)^2 \frac{1}{n} \sum_{j=1}^n \left(\frac{\eta_3 - \eta_4}{n_t} \|X_j\|_F^4 + \eta_4 \|X_j X_j^T\|_F^2\right)\,, \end{aligned} \tag{3.49}$$

where we replaced the terms that do not depend on x^n with K. Note that the first term in parentheses (premultiplying the sum) converges almost-surely to 1, by the strong law of large numbers. Similarly, the normalized sum converges to the expectation (also by the strong law of large numbers). Overall, applying the SLLN in the limit as $n \to \infty$, we obtain:

$$\lim_{n\to\infty} V_n(\tilde{X}^n) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n V_1(\tilde{X}_j) \tag{3.50}$$

$$= \mathbb{E}[V_1(X)] \triangleq V'\,. \tag{3.51}$$

In particular, $\mathbb{P}[V_n(\tilde{X}^n) \le V' + \tau] \to 1$. This concludes the proof of $\mathbb{P}[\tilde{X}^n \in F_n] \to 1$. □

3.3 Converse

Here we state and prove the converse part of Theorem 4. There are two challenges in proving the converse relative to other finite blocklength proofs. First, behavior of the information density (2.40) varies widely as x^n varies over the power-sphere

$$S_n = \{x^n \in (\mathbb{R}^{n_t \times T})^n : \|x^n\|_F^2 = nTP\}. \tag{3.52}$$

Indeed, when $\max_j \|x_j\|_F \geq cn^{\frac{1}{4}}$ the distribution of information density ceases to be Gaussian. In contrast, the information density for the AWGN channel is constant over S_n.

Second, assuming asymptotic normality, we have for any $x^n \in S_n$:

$$\begin{aligned} -\log \beta_{1-\epsilon}(P_{Y^nH^n|X^n=x^n}, P_{Y,H}^{*n}) \approx \\ nC(P) - \sqrt{nV_n(x^n)}Q^{-1}(\epsilon) + o(\sqrt{n})\,. \end{aligned} \tag{3.53}$$

However, the problem is that $V_n(x^n)$ is also non-constant. In fact there exists regions of S_n where $V_n(x^n)$ is abnormally small. Thus we need to also show that no capacity-achieving codebook can live on those abnormal sets.

The main theorem of the section is the following:

Theorem 12. *For any $\delta_n \to 0$ there exists $\delta_n' \to 0$ such that any (n, M, ϵ)-max code with $\epsilon < 1/2$ and codewords satisfying $\max_{1\leq j\leq n} \|x_j\|_F \leq \delta_n n^{\frac{1}{4}}$ has size bounded by*

$$\log M \leq nTC(P) - \sqrt{nTV(P)}Q^{-1}(\epsilon) + \delta_n'\sqrt{n}\,, \tag{3.54}$$

where $C(P)$ and $V(P)$ are defined in (2.11) *and* (3.6)*, respectively.*

Proof. As usual, without loss of generality we may assume that all codewords belong to S_n as defined in (3.52), see [16, Lemma 39]. The maximal probability of error code size is bounded by a meta-converse theorem [16, Theorem 31], which states that for any (n, M, ϵ) code and distribution $Q_{Y^nH^n}$ on the output space of the channel,

$$\frac{1}{M} \geq \inf_{x^n} \beta_{1-\epsilon}(P_{Y^nH^n|X=x^n}, Q_{Y^nH^n})\,, \tag{3.55}$$

where infimum is taken over all codewords. The main problem is to select $Q_{Y^nH^n}$ appropriately. We do this separately for the two subcodes defined as follows. Fix arbitrary $\delta > 0$ (it will be taken to 0 at the end) and introduce:

$$\mathcal{C}_l \triangleq \mathcal{C} \cap \{x^n : V_n(x^n) \leq n(V(P) - \delta)\} \tag{3.56}$$

$$\mathcal{C}_u \triangleq \mathcal{C} \cap \{x^n : V_n(x^n) > n(V(P) - \delta)\}\,. \tag{3.57}$$

To bound the cardinality of $\mathcal{C}_u$, we select $Q_{Y^nH^n} = (P_{Y,H}^*)^n$ to be the n-product of the caod (2.12), then apply the following estimate from [17, Lemma 14], quoted here:

for any $\Delta > 0$ we have

$$\begin{aligned}
&\log \beta_{1-\epsilon}(P_{Y^n H^n|X=x^n}, P_{Y,H}^{*n}) \geq \\
&\quad - nD_n(x^n) - \sqrt{nV_n(x^n)}Q^{-1}\left(1 - \epsilon - \frac{B_n(x^n) + \Delta}{\sqrt{n}}\right) \\
&\quad - \frac{1}{2}\log\frac{n}{\Delta^2}\,,
\end{aligned} \tag{3.58}$$

where D_n, V_n and B_n are given by (2.43), (5.6) and (2.45), respectively. We choose $\Delta = n^{\frac{1}{4}}$ and then from Lemma 3 (which relies on the assumption that $\|x_j\|_F \leq \delta n^{\frac{1}{4}}$) we get that for some constants K_1, K_2 we have for all $x^n \in \mathcal{C}_u$:

$$B_n(x^n) + \Delta \leq K_1\delta_n^2\sqrt{n} + K_2 n^{\frac{1}{4}} + \frac{K_3}{n^{1/2}}\,.$$

From (3.55) and (3.58) we therefore obtain

$$\log|\mathcal{C}_u| \leq nTC(P) - \sqrt{nT(V(P)-\delta)}Q^{-1}(\epsilon - \delta_n'') + \frac{1}{4}\log n\,, \tag{3.59}$$

where $\delta_n'' = K_1\delta_n^2 + K_2 n^{-\frac{1}{4}} \to 0$ as $n \to \infty$.

Next we proceed to bounding $|\mathcal{C}_l|$. To that end, we first state two lemmas. Lemma 13 shows that, if in addition to the power constraint $\mathbb{E}[\|X\|_F^2] \leq TP$, we also required $\mathbb{E}[V_1(X)] \leq V(P) - \delta$, then the capacity of this variance-constrained channel is strictly less than without the latter constraint.

Lemma 13. *Consider the following constrained capacity:*

$$\begin{aligned}
&\tilde{C}(P,\delta) \triangleq \\
&\frac{1}{T}\sup_X \left\{I(X;Y|H) : \mathbb{E}[\|X\|_F^2] \leq TP, \mathbb{E}[V_1(X)] \leq V(P) - \delta\right\}
\end{aligned} \tag{3.60}$$

where $V(P)$ is from (3.6) and $V_1(x)$ is from (5.7). For any $\delta > 0$ there exists $\tau = \tau(P,\delta) > 0$ such that $\tilde{C}(P,\delta) < C(P) - \tau$.

Remark 5. Curiously, if we used constraint $\mathbb{E}[V_1(X)] > V(P) + \delta$ instead of $\mathbb{E}[V_1(X)] \leq V(P) - \delta$ in (3.60), then the resulting capacity equals $C(P)$ regardless of δ.

The following Lemma shows that, with the appropriate choice of an auxiliary distribution Q_{Y^n,H^n}, the expected size of the normalized log likelihood ratio is strictly smaller than capacity, while the variance of that same ratio is upper bounded by a constant (i.e. does not scale with n).

Lemma 14. *Define the auxiliary distribution*

$$Q_{Y|H}(y|h) = \begin{cases} P_{Y|H}^*(y|h) & \|h\|_F^2 > A \\ \tilde{P}_{Y|H}^*(y|h) & \|h\|_F^2 \leq A \end{cases} \tag{3.61}$$

where $A > 1$ is a constant, $P^*_{Y|H}(y|h)$ is the caod for the MIMO-BF channel, and $\tilde{P}^*_{Y|H}(y|h)$ is the caod for the variance-constrained channel in (3.60). Let $Q_{Y,H} = P_H Q_{Y|H}$, and $Q_{Y^n,H^n} = \prod_{i=1}^n Q_{Y,H}$. Then there exists constants $\tau, K > 0$ such that for all $x^n \in \mathcal{C}_l$,

$$C_n \triangleq \frac{1}{nT}\mathbb{E}\left[\log \frac{P_{Y^n,H^n|X^n}}{Q_{Y^n,H^n}}(Y^n, H^n|x^n)\right] \le C(P) - \tau \tag{3.62}$$

$$V_n \triangleq \frac{1}{nT}\mathrm{Var}\left(\log \frac{P_{Y^n,H^n|X^n}}{Q_{Y^n,H^n}}(Y^n, H^n|x^n)\right) \le K \tag{3.63}$$

where $Y_i = H_i x_i + Z_i$, $i = 1, \ldots, n$ is the joint distribution.

Remark 6. The reason we let $Q_{Y|H}$ take on two distributions depending on the value of H is because we do not know the form of $\tilde{P}^*_{Y|H}$, hence we do not explicitly know how it depends on H. This choice of $Q_{Y|H}$ ensures that expectations involving $\tilde{P}^*_{Y|H}$ are finite.

Choose $Q_{Y,H}$ as in Lemma 14, so that the bounds on C_n, V_n from (3.62), (3.63) respectively, hold. Applying [17, Lemma 15] with $\alpha = 1 - \epsilon$ (the statement of this lemma is the contents of (3.64)), we obtain

$$\begin{aligned}
&\log \beta_{1-\epsilon}(P_{Y^n,H^n|X^n=x^n}, \tilde{P}^{*n}_{Y,H}) \\
&\quad \ge -nTC_n - \sqrt{\frac{2nTV_n}{1-\epsilon}} - \log\frac{1-\epsilon}{2} &(3.64)\\
&\quad \ge -nT(C(P) - \tau) - \sqrt{\frac{2nTK}{1-\epsilon}} + \log\frac{1-\epsilon}{2}. &(3.65)
\end{aligned}$$

Therefore, from (3.55) we conclude that for all $n \ge n_0(\delta)$ we have

$$\log|\mathcal{C}_l| \le nT\left(C(P) - \frac{\tau}{2}\right). \tag{3.66}$$

Overall, from (3.59) and (3.66) we get (due to arbitrariness of δ) the statement (3.54). □

Proof of Lemma 13. Introduce the following set of distributions:

$$\mathcal{P}' \triangleq \left\{P_X : \mathbb{E}[\|X\|_F^2] \le TP,\ \mathbb{E}[V_1(X)] \le V - \delta\right\}. \tag{3.67}$$

By Prokhorov's criterion (e.g. [19, Theorem 5.1], tightness implies relative compactness), the norm constraint implies that this set is relatively compact in the topology of weak convergence. So there must exist a sequence of distributions $\tilde{P}_n \in \mathcal{P}'$ s.t. $\tilde{P}_n \xrightarrow{w} \tilde{P}$ and $I(\tilde{X}_n; H\tilde{X}_n + Z|H) \to \tilde{C}(P, \delta)$ where $\tilde{X}_n \sim \tilde{P}_n$. By Skorokhod representation [19, Theorem 6.7], we may assume $\tilde{X}_n \xrightarrow{a.s.} \tilde{X} \sim \tilde{P}$, i.e. there exists random variable $\tilde{X}$ that is the pointwise limit of the $\tilde{X}_n$'s. Notice that for any continuous

bounded function $f(h, y)$ we have

$$\mathbb{E}\left[f(H, H\tilde{X}_n + Z)\right] \to \mathbb{E}\left[f(H, H\tilde{X} + Z)\right],$$

and therefore $P_{\tilde{Y}_n,H} \xrightarrow{w} P_{\tilde{Y},H}$. Assume (to arrive at a contradiction) that $\tilde{C}(P, \delta) = C(P)$, then by the golden formula, cf. [10, Theorem 3.3], we have

$$\begin{aligned} I(\tilde{X}_n; H\tilde{X}_n + Z|H) & \\ &= D(P_{YH|X}\|P^*_{Y,H}|P_{\tilde{X}_n}) - D(P_{\tilde{Y}_n,H}\|P^*_{Y,H}) && (3.68)\\ &= \mathbb{E}\left[D_1(\tilde{X}_n)\right] - D(P_{\tilde{Y}_n,H}\|P^*_{Y,H}) && (3.69)\\ &\le C(P) - D(P_{\tilde{Y}_n,H}\|P^*_{Y,H}), && (3.70) \end{aligned}$$

where $D_1(x)$ is from (2.43). Therefore, we have

$$D(P_{\tilde{Y}_n,H}\|P^*_{Y,H}) \to 0.$$

From weak lower-semicontinuity of divergence [10, Theorem 3.6] we have $D(P_{\tilde{Y},H}\|P^*_{Y,H}) = 0$. In particular, if we denote X^* to have Telatar distribution (2.9), we must have

$$\mathbb{E}\left[\|\tilde{Y}\|_F^2\right] = \mathbb{E}\left[\|H\tilde{X} + Z\|_F^2\right] = \mathbb{E}\left[\|HX^* + Z\|_F^2\right]. \tag{3.71}$$

From Lemma 2 (see Remark 3) we have

$$\mathbb{E}\left[\|Hx\|_F^2\right] = \frac{\mathbb{E}\left[\|H\|_F^2\right]}{n_t}\|x\|_F^2 \tag{3.72}$$

and hence from the independence of Z from (H, X) we get

$$\mathbb{E}\left[\|H\tilde{X} + Z\|_F^2\right] = \frac{\mathbb{E}\left[\|H\|_F^2\right]}{n_t}\mathbb{E}\left[\|\tilde{X}\|_F^2\right] + n_r T,$$

and similarly for the right-hand side of (3.71). We conclude that

$$\mathbb{E}\left[\|\tilde{X}\|_F^2\right] = \mathbb{E}\left[\|X^*\|_F^2\right] = TP.$$

Finally, plugging this fact into the expression for $D_1(x)$ in (2.43) and (3.69) we obtain

$$I(\tilde{X}; H\tilde{X} + Z|H) = \mathbb{E}\left[D_1(\tilde{X}_n)\right] = C(P).$$

That is, $\tilde{X}$ is a caid. But from Fatou's lemma we have (recall that $V_1(x) \ge 0$ since it is a variance)

$$\mathbb{E}\left[V_1(\tilde{X})\right] \le \liminf_{n\to\infty} \mathbb{E}\left[V_1(\tilde{X}_n)\right] \le V(P) - \delta,$$

where the last step follows from $\tilde{P}_n \in \mathcal{P}'$. A caid achieving conditional variance strictly less than $V(P)$ contradicts the definition of $V(P)$, cf. (3.6), as the infimum of $\mathbb{E}\left[V_1(X)\right]$ over all caids. □

Proof of Lemma 14. First we analyze C_n from (3.62). Denote

$$i(x;y,h) = \log \frac{P_{Y|H,X}}{P^*_{Y|H}}(y|h,x) \tag{3.73}$$

$$\tilde{i}(x;y,h) = \log \frac{P_{Y|H,X}}{\tilde{P}^*_{Y|H}}(y|h,x)\,. \tag{3.74}$$

Here, $i(x;y,h)$ is the information density given by (2.40), while $\tilde{i}(x;y,h)$ instead has the caod for the variance-constrainted channel (3.60) in the denominator. Since $Q_{Y|H}$ takes on one of two distributions based on the value of H, conditioning on H in two ways yields

$$C_n = \frac{1}{nT}\mathbb{E}\left[\log \frac{P_{Y^n,H^n|X^n}}{Q_{Y^n,H^n}}(Y^n,H^n|x^n)\right] \tag{3.75}$$

$$= \frac{1}{nT}\sum_{j=1}^{n}\mathbb{E}\left[i(x_j;Y_j,H_j)\big|\|H_j\|_F^2 > A\right]\mathbb{P}[\|H_j\|_F^2 > A] \tag{3.76}$$

$$+ \frac{1}{nT}\sum_{j=1}^{n}\mathbb{E}\left[\tilde{i}(x_j,Y_j,H_j)\big|\|H_j\|_F^2 \le A\right]\mathbb{P}[\|H_j\|_F^2 \le A]\,. \tag{3.77}$$

The H_j's are i.i.d. according to P_H, so we define $p \triangleq \mathbb{P}[\|H_j\|_F^2 > A]$. Using capacity saddle point, (3.76) is bounded by

$$\frac{p}{nT}\mathbb{E}\left[\sum_{j=1}^{n} i(x_j;Y_j,H_j)\middle|\|H_j\|_F^2 > A\right] \le pC(P_{H>A}) \tag{3.78}$$

where $C(P_H)$ denotes the capacity of the MIMO-BF channel with fading distribribution P_H, and $P_{H>A}$ denotes the distribution of H conditioned on $\|H\|_F^2 > A$ (similarly, $P_{H\le A}$ will denote H conditioned on $\|H\|_F^2 \le A$). (3.78) follows from the fact that the information density, i.e. $\log \frac{P_{Y|H,X}}{P^*_{Y|H}}(y|h,x)$, is not a function of P_H, hence changing the distribution P_H does not affect the form of $i(x;y,h)$. Similarly, using Lemma 13, (3.77) is bounded by

$$\frac{1-p}{nT}\mathbb{E}\left[\sum_{j=1}^{n}\tilde{i}(X_j;Y_j,H_j)\middle|\|H_j\|_F^2 \le A\right]$$

$$\le (1-p)\tilde{C}(P_{H\le A}) \tag{3.79}$$

$$= (1-p)(C(P_{H\le A}) - \tau') \tag{3.80}$$

where $\tau' > 0$ is a positive constant, and $\tilde{C}(P_H)$ denotes the solution to the optimization problem (3.60) when the fading distribution is P_H. Putting together (3.78) and

(3.80), we obtain an upper bound on C_n,

$$C_n \leq pC(P_{H>A}) + (1-p)(C(P_{H\leq A}) - \tau'). \tag{3.81}$$

Note that $C(P_H) = \mathbb{E}_{P_H}\left[\log\det(I_{n_r} + P/n_t HH^T)\right]$, so the capacity only depends on P_H through the expectation – the expression inside is not a function of P_H because the i.i.d. Gaussian caid achieves capacity for all isotropic P_H's. Hence, by the law of total expectation, (3.81) simplifies to

$$C_n \leq C(P_H) - (1-p)\tau'. \tag{3.82}$$

Finally, we can upper bound p using Markov's inequality as

$$p = \mathbb{P}[\|H_1\|_F^2 > A] \leq \frac{1}{A} \tag{3.83}$$

since $A > 1$. Applying this bound to (3.82), we obtain

$$C_n \leq C(P_H) - (1-p)\tau' \tag{3.84}$$

$$\leq C(P_H) - \left(1 - \frac{1}{A}\right)\tau'. \tag{3.85}$$

Defining $\tau \triangleq (1 - 1/A)\tau'$ completes the proof of (3.62).

Next we analyze V_n from (3.63). The strategy will be to decompose (3.63) into two terms depending on the value of $\|H\|_F^2$, then show that each term is upper bounded by $A_1 + A_2\sum_{j=1}^n \|x_j\|_F^4$, where A_1, A_2 are constants not depending on x^n. Finally, we will show that $\sum_{j=1}^n \|x_j\|_F^4 = O(n)$ when $x^n \in \mathcal{C}_l$. To this end,

$$V_n = \frac{1}{nT}\mathrm{Var}\left(\log\frac{P_{Y^n,H^n|X^n}}{Q_{Y^n,H^n}}(Y^n, H^n|x^n)\right) \tag{3.86}$$

$$= \frac{1}{nT}\sum_{j=1}^n \mathrm{Var}\left(\log\frac{P_{Y,H|X}}{Q_{Y,H}}(Y_j, H_j|x_j)\right) \tag{3.87}$$

$$\leq \frac{1}{nT}\sum_{j=1}^n \mathbb{E}\left[\left(\log\frac{P_{Y,H|X}}{Q_{Y,H}}(Y_j, H_j|x_j)\right)^2\right] \tag{3.88}$$

where (3.87) follows from the independence of the terms, and (3.88) is from the bound $\mathrm{Var}(X) \leq \mathbb{E}[X^2]$. Again we condition on H in two ways,

$$V_n \leq \frac{p}{nT}\sum_{j=1}^n \mathbb{E}\left[i(x_j; Y_j; H_j)^2 \middle| \|H_j\|_F^2 > A\right] \tag{3.89}$$

$$+ \frac{1-p}{nT}\sum_{j=1}^n \mathbb{E}\left[\tilde{i}(x_j; Y_j, H_j)^2 \middle| \|H_j\|_F^2 \leq A\right]. \tag{3.90}$$

For the first term, (3.89), we know the expression for $i(x;y,h)$ from (2.40), so we simply upper bound $i(x;y,h)^2$. To this end,

$$\begin{aligned} & i(x;y,h)^2 \\ & \leq 2\left(\frac{T}{2}\log\det\left(I_{n_r}+\frac{P}{n_t}hh^T\right)\right)^2 \\ & + 2\left(\frac{\log e}{2}\sum_{j=1}^{n_{\min}}\frac{\lambda_j^2\|v_j^Tx\|^2+2\lambda_j\langle v_j^Tx,\tilde{z}_j\rangle-\frac{P}{n_t}\lambda_j^2\|\tilde{z}_j\|^2}{1+\frac{P}{n_t}\lambda_j^2}\right)^2 \qquad (3.91) \\ & \leq C_1\|h\|_F^2+C_2\|x\|_F^4+C_3(\tilde{z}_j)\|x\|_F^2+C_4(\tilde{z}_j) \qquad (3.92) \end{aligned}$$

where C_1, C_2 are non-negative constants, and $C_3(\tilde{z}_j), C_4(\tilde{z}_j)$ are functions of only $\tilde{z}_j$ that have bounded moments. This follows from:

- Bounding the first term via

$$\left(\frac{T}{2}\log\det\left(I_{n_r}+\frac{P}{n_t}hh^T\right)\right)^2 \leq \log^2(e)\frac{PT^2}{4n_t}n_{\min}\|h\|_F^2 \qquad (3.93)$$

  which can be derived from the basic inequality $\log(1+x)\leq\log(e)\sqrt{x}$.

- Noting that the second term is bounded in h, since for all $\lambda\in\mathbb{R}$,

$$\frac{|\lambda|}{1+\frac{P}{n_t}\lambda^2}\leq\frac{1}{2\sqrt{\frac{P}{n_t}}} \qquad (3.94)$$

$$\frac{\lambda^2}{1+\frac{P}{n_t}\lambda^2}\leq\frac{n_t}{P}. \qquad (3.95)$$

- Noting that all moments of $\|\tilde{z}_j\|^2$ are finite because this is the norm of a standard normal vector.

Therefore, after taking the expectation of (3.92) and summing over all n, we obtain

$$\begin{aligned} & \frac{p}{nT}\sum_{j=1}^{n}\mathbb{E}\left[i(x_j;Y_j;H_j)^2\middle|\|H_j\|_F^2>A\right] \\ & \qquad \leq \frac{1}{nT}\left(C_5+C_6\sum_{j=1}^{n}\|x_j\|_F^4\right) \qquad (3.96) \end{aligned}$$

for some non-negative constants C_5, C_6.

To bound the second term, (3.90), first we split the logarithm as

$$\begin{aligned}
&\mathbb{E}\left[\tilde{i}(x_j;Y_j,H_j)^2\middle|\|H_j\|_F^2\le A\right]\\
&\quad\le 2\mathbb{E}\left[\log\left(P_{Y|H,X}(Y_j|H_j,x_j)\right)^2\middle|\|H_j\|_F^2\le A\right]\\
&\quad+2\mathbb{E}\left[\log\left(\tilde{P}^*_{Y|H}(Y_j|H_j)\right)^2\middle|\|H_j\|_F^2\le A\right] \qquad (3.97)
\end{aligned}$$

The first term in (3.97) is simple to handle, since its expression is given by the definition of the channel,

$$\begin{aligned}
&\mathbb{E}\left[\log\left(P_{Y|H,X}(Y_j|H_j,x_j)\right)^2\middle|\|H_j\|_F^2\le A\right]\\
&\quad=\mathbb{E}\left[\left(-\frac{n_rT}{2}\log(2\pi)-\frac{1}{2}\|Z_j\|_F^2\right)^2\right] \qquad (3.98)\\
&\quad\le\frac{1}{2}n_rT\log^2(2\pi)+\frac{1}{2}n_rT(2+n_rT) \qquad (3.99)\\
&\quad\triangleq K_1 \qquad (3.100)
\end{aligned}$$

i.e. we have a constant upper bound. For the second term in (3.97), notice that $\tilde{P}^*_{Y,H}$ that is inducible through channel, i.e. there exists an input distribution P_X such that $\tilde{P}^*_{Y,H}(y,h)=\mathbb{E}[P_{Y,H|X}(y,h|X)]$. Using this fact, we obtain the bound

$$\begin{aligned}
-\log\tilde{P}^*_{Y|H}(y|h)&=-\log\mathbb{E}[P_{Y|H,X}(y|h,X)] &(3.101)\\
&\le\mathbb{E}[-\log P_{Y|H,X}(y|h,X)] &(3.102)\\
&=\mathbb{E}\left[\frac{n_rT}{2}\log(2\pi)+\frac{1}{2}\|y-hX\|_F^2\right] &(3.103)\\
&\le\frac{n_rT}{2}\log(2\pi)+\|y\|_F^2+TP\|h\|_F^2 &(3.104)
\end{aligned}$$

where (3.102) follows from Jensen's inequality, (3.103) is from the definition of the channel, and (3.104) follows from applying the inequality $\|A+B\|_F^2\le 2\|A\|_F^2+2\|B\|_F^2$ along with $\|hX\|_F^2\le\|h\|_F^2\|X\|_F^2$, then noting that X satisfies $\mathbb{E}[\|X\|_F^2]=TP$. Using this, we can bound the second term in (3.97) via

$$\begin{aligned}
&\mathbb{E}\left[\log\left(\tilde{P}^*_{Y|H}(Y_j|H_j)\right)^2\middle|\|H_j\|_F^2\le A\right] &(3.105)\\
&\le\mathbb{E}\left[\left(\frac{n_rT}{2}\log(2\pi)+\|Y_j\|_F^2|+TP\|H_j\|_F^2\right)^2\middle|\|H_j\|_F^2\le A\right] &(3.106)\\
&\le\mathbb{E}\left[3\frac{n_r^2T^2}{4}\log^2(2\pi)+3\|Y_j\|_F^4\right.\\
&\qquad\left.+3T^2P^2\|H_j\|_F^4\middle|\|H_j\|_F^2\le A\right] &(3.107)\\
&\le K_2+K_3\|x\|_F^4 &(3.108)
\end{aligned}$$

where K_2, K_3 are non-negative constants which do not depend on x, (3.106) is from the above bound (3.104), and (3.108) follows from applying the bound

$$\begin{aligned}
&\mathbb{E}\left[\|Y_j\|_F^4 \big| \|H_j\|_F^2 \leq A\right] \\
&\quad = \mathbb{E}\left[\|H_j x_j + Z_j\|_F^4 \big| \|H_j\|_F^2 \leq A\right] && (3.109)\\
&\quad \leq 8\mathbb{E}\left[\|H_j\|_F^4 \big| \|H_j\|_F^2 \leq A\right] \|x_j\|_F^4 + 16 n_r^2 T^2 && (3.110)\\
&\quad \leq 8A\|x_j\|_F^4 + 16 n_r^2 T^2 . && (3.111)
\end{aligned}$$

Putting together (3.108) and (3.100), we obtain an upper bound on (3.90),

$$\begin{aligned}
&\frac{1-p}{nT}\sum_{j=1}^{n} \mathbb{E}\left[\tilde{i}(x_j; Y_j, H_j)^2 \big| \|H_j\|_F^2 \leq A\right] \\
&\qquad \leq \frac{2(1-p)}{nT}\left(K_3 + K_4 + K_5 \sum_{j=1}^{n} \|x_j\|_F^4\right) . && (3.112)
\end{aligned}$$

Now, since $x^n \in \mathcal{C}_l$ by assumption, we can control the quantity $\sum_{i=1}^{n} \|x_i\|_F^4$ via

$$\begin{aligned}
\sum_{i=1}^{n} \|x_i\|_F^4 &\leq \sum_{i=1}^{n} V_1(x_i) && (3.113)\\
&\leq n(V(P) - \delta) , && (3.114)
\end{aligned}$$

where the first inequality follows from the non-negativity of the terms in $V_1(x)$ given in Proposition 18, and the second inequality is from the definition of $\mathcal{C}_l$. Hence the sum of fourth powers of the $\|x_i\|_F$'s is $O(n)$ on $\mathcal{C}_l$. All together, combining (3.112) and (3.96) yields the following bound on V_n,

$$\begin{aligned}
V_n &\leq \frac{1}{n}\left(K' + K'' \sum_{j=1}^{n} \|x_i\|_F^4\right) && (3.115)\\
&\leq K && (3.116)
\end{aligned}$$

which completes the proof of (3.63). □

Chapter 4

Numerical Computation of Non-Asymptotic Bounds

In this chapter, we discuss one method to numerically computation an achievability bound for the MIMO-BF channel. First, we want to explain why this is necessary. The bounds in Chapter 3 show that the maximum number of supportable codewords at blocklength n and error probability ϵ is upper and lower bounded as

$$\frac{\kappa_\tau(F_n, Q_{Y^nH^n})}{\sup_{x^n \in F_n} \beta_{1-\epsilon+\tau}(P_{Y^nH^n|X^n=x^n}, Q_{Y^nH^n})} \leq M^*(n,\epsilon,P) \leq \frac{1}{\inf_{x^n} \beta_{1-\epsilon}(P_{Y^nH^n|X=x^n}, Q_{Y^nH^n})}. \tag{4.1}$$

These are non-asymptotic bounds, i.e. they hold for all n, ϵ, and P. In order to establish the dispersion, we showed that these bound match up to the $O(\sqrt{n})$ term. Note that matching to the $O(\sqrt{n})$ term is an asymptotic statement, and allow us to use the "normal approximation"

$$\log M^*(n,\epsilon,P) \approx nTC - \sqrt{nV}Q^{-1}(\epsilon) \tag{4.2}$$

to get a more refined approximation for $\log M^*(n,\epsilon,P)$. This approximation is fairly easy to compute, since C and V are given by single letter expressions. However, if one is interested in hard achievability and converse bounds, as some applications may demand, the approximation (4.2) is not sufficient. For example, it may be possible that the constant term, which disappears asymptotically, is in fact large enough to compute with the $O(n)$ term for mild blocklengths. That is where computation of the non-asymptotic bounds comes into play.

To give guarantees based on our achievability and converse proofs, we would have to numerically compute the non-asymptotic upper an lower bounds in (4.1). Often this is difficult, first because κ_τ does not have a nice form, and because the minimization in the denominator over an n dimensional space can be hard to find. Note that the β terms, because of the Neyman-Pearson Lemma, are given by the CDF of an n-fold product distribution. Hence these are not trivial to compute, but at least they can be approximated via monte carlo given enough time.

To solve this problem, in this chapter we derive the $\beta\beta$-bound, which is looser than the $\kappa\beta$ bound, but is more amenable to numerical computation. After deriving it, we show how to apply it to the MIMO-BF channel, and give examples of its computation. We note that this is joint work with Wei Yang, and the results appear in [20].

4.1 $\beta\beta$ Achievability Bound

The $\beta\beta$ bound gives a looser but easier to compute alternative to the $\kappa\beta$ bound. First we state and prove the bound, then give a discussion on its computability.

Theorem 15 ($\beta\beta$ bound). *For all $\epsilon \in (0,1)$, and every input distribution P_X, there exists an (M, ϵ) average probability of error code satisfying*

$$\frac{M}{2} \geq \sup_{0<\tau<\epsilon} \sup_{Q_Y} \frac{\beta_\tau(P_Y, Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)} \tag{4.3}$$

The proof is due to Wei Yang – we give it here for completeness.

Proof. Take $\epsilon \in (0,1)$ and $\tau \in (0, \epsilon)$, and let P_X, Q_Y be two arbitrary measure on the input and output space, respectively. Consider the binary hypothesis test

$$H_0 : (X,Y) \sim P_X Q_Y \tag{4.4}$$
$$H_1 : (x,Y) \sim P_{XY} \tag{4.5}$$

where $P_{XY} = P_X \circ P_{Y|X}$ is the join distribution induced by the channel. I.e. this tests if the pair (X,Y) are dependent, or are independent with product distribution $P_X Q_Y$. Using the Neyman-Pearson Lemma, let $Z(X,Y)$ be the test that correctly outputs H_1 with probability at least $1-\epsilon+\tau$, and has minimal error amongst all tests which output H_1 when in fact H_0 is true, i.e.

$$P_{XY}\left[Z(X,Y)=1\right] \geq 1-\epsilon+\tau \tag{4.6}$$
$$P_X Q_Y\left[Z(X,Y)=1\right] = \beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)\,. \tag{4.7}$$

Note that this is the definition of $\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)$. The encoder employs random coding – each of M codewords $\{C_1, \ldots, C_M\}$ is generated i.i.d. from the distribution P_X. The decoder computes the test $Z(c_j, y)$ for each $j = 1, \ldots, M$, and outputs the smallest j such that $Z(c_j, y) = 1$. If such an index does not exist, the decoder makes an error. Denote the probability of error for a fixed codebook by $P_e(c_1, \ldots, c_M)$. Let W denote the sent codeword. With this encoder and decoder, there are two error events: either an incorrect codeword has $Z(c_j, y) = 1$ for $j < W$ (since the decoder outputs the first index), or if the correct codeword has $Z(c_W, y) = 0$. From this, the

probability of error averaged over all codebooks is given by

$$\mathbb{E}\left[P_e(C_1,\ldots,C_M)\right] \leq \mathbb{P}[Z(C_W,Y)=0] + \mathbb{P}\left[\max_{j<W} Z(C_j,Y)=1\right] \tag{4.8}$$

$$\leq \epsilon - \tau + \mathbb{P}\left[\max_{j<W} Z(C_j,Y)=1\right] \tag{4.9}$$

where the second line follows from (4.6). We handle the second term as follows: Let $\bar{X} \sim P_X$ denote a random variable that is independent from Y, then

$$\mathbb{P}\left[\max_{j<W} Z(C_j,Y)=1\right] = \frac{1}{M}\sum_{j_*=1}^{M} \mathbb{P}\left[\max_{j<j_*} Z(C_j,Y)=1\right] \tag{4.10}$$

$$\leq \frac{1}{M}\sum_{j_*=1}^{M}\sum_{j<j_*} \mathbb{P}\left[Z(\bar{X},Y)=1\right] \tag{4.11}$$

$$\leq \frac{M-1}{2}\mathbb{P}\left[Z(\bar{X},Y)=1\right] \tag{4.12}$$

$$= \frac{M-1}{2}\beta_{1-\epsilon+\tau}(P_{XY},P_XQ_Y)\,, \tag{4.13}$$

where the first line is averaging over W, the second is from the union bound, the third is from summing the series of $j<j_*$, and the fourth is from the definition of $\beta_{1-\epsilon+\tau}(P_{XY},P_XQ_Y)$. Now, choose

$$M = \left\lceil \frac{2\beta_\tau(P_Y,Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY},P_XQ_Y)} \right\rceil\,, \tag{4.14}$$

yielding from (4.13),

$$\mathbb{P}\left[\max_{j<W} Z(C_j,Y)=1\right] \leq \beta_\tau(P_Y,Q_Y)\,. \tag{4.15}$$

Now, note that $\beta_\alpha(P,Q) \leq \alpha$ since the best test is better than the test that ignores the data and outputs P with probability α. With this, we conclude

$$\mathbb{P}\left[\max_{j<W} Z(C_j,Y)=1\right] \leq \tau\,. \tag{4.16}$$

Hence overall, the average probability of error of our random code, from (4.9) is bounded by

$$\mathbb{E}\left[P_e(C_1,\ldots,C_M)\right] \leq \epsilon \tag{4.17}$$

and we conclude that there exists an (M,ϵ) code satisfying (4.14) for arbitrary $\tau \in (0,\epsilon)$ and arbitrary distribution Q_Y. □

Remark 7. The "golden formula" in information theory often refers to the following

equation: for any distribution on the output space Q_Y,

$$I(X;Y) = D(P_{Y|X}||Q_Y|P_X) - D(P_Y||Q_Y)\,. \tag{4.18}$$

An example application is *capacity saddle point* i.e. if P_X^*, P_Y^* are the capacity achieving input and output distribution of a channel, respectively, then for all P_X, Q_Y,

$$D(P_{Y|X}||P_Y^*|P_X) \leq D(P_{Y|X}||P_Y^*|P_X^*) \leq D(P_{Y|X}||Q_Y|P_X^*) \tag{4.19}$$

which allows us to give upper and lower bounds on capacity by cleverly choosing P_X and/or Q_Y. Indeed, in finite blocklength applications, a meta-principle is that

$$D(P||Q) \mapsto -\log \beta_\alpha(P,Q) \tag{4.20}$$

i.e. divergences are "replaced" by β functions. Note that $-\frac{1}{n}\log\beta_\alpha(P^n,Q^n) \to D(P||Q)$ as $n \to \infty$, so this isn't unexpected. In this sense, the $\beta\beta$ bound can be seen as a non-asymptotic analog to the golden formula, it reads:

$$\log\frac{M}{2} \geq \sup_{0<\tau<\epsilon}\sup_{Q_Y}\left(-\log\beta_{1-\epsilon+\tau}(P_{XY},P_XQ_Y) + \log\beta_\tau(P_Y,Q_Y)\right) \tag{4.21}$$

The $\beta\beta$ bound can be seen as an average probability of error analog to the $\kappa\beta$ bound. Indeed, using Lemma 5, we can lower bound bound the $\kappa\beta$ bound by

$$M \geq \frac{\kappa_\tau(F,Q_Y)}{\sup_{x\in F}\beta_{1-\epsilon+\tau}(P_{Y|X=x},Q_Y)} \geq \frac{\beta_{\tau P_X[F]}(P_X\circ P_{Y|X},Q_Y)}{\sup_{x\in F}\beta_{1-\epsilon+\tau}(P_{Y|X=x},Q_Y)} \tag{4.22}$$

The $\beta\beta$ bounds has a few computational advantages. The major difficulty in computing the $\kappa\beta$ bound is computing the supremum in the denominator. When considering the average probability of error in the $\beta\beta$ bound, this supremum does not appear, leaving us to approximate the error in a hypothesis test between two n-fold distributions.

4.2 Application to the MIMO-BF Channel

The art of applying and computing these bounds is in choosing the distributions P_X, Q_Y, and parameter τ such that 1) the quantities are computable, and 2) the bound still remains fairly tight. In this section, we will go through these choices for the MIMO-BF channel where the fading process has i.i.d. Gaussian entries, i.e. the Rayleigh case.

We apply the $\beta\beta$ bound by first choosing the input distribution P_{X^n} to be uniform on the sphere $\{x^n \in \mathbb{C}^{n_t\times nT} : \|x^n\|_F^2 = nTP\}$. This distribution can be represented as

$$P_{X^n} \sim \frac{W^n}{\|W^n\|_F^2}\sqrt{nTP} \tag{4.23}$$

where $W^n \in \mathbb{C}^{n_t \times nT}$ has i.i.d. $\mathcal{CN}(0,1)$ entries. Note that this is the same input distribution as used in the achievability proof. $Q_{Y^nH^n}$ is chosen to be the capacity achieving output distribution – this is the standard choice. Since P_{X^n} can be sampled from using a n-dimensional standard Gaussian, and $Q_{Y^nH^n} = \prod_{i=1}^n P_H Q_{Y|H}$ is the product of two Gaussian distributions, we can easily sample from this distribution. Thus computing the denominator

$$\beta_{1-\epsilon+\tau}(P_{X^nY^nH^n}, P_{X^n}Q_{Y^nH^n}) \tag{4.24}$$

can be done using the standard monte carlo approach.

Note that in the AWGN channel, $\beta_\alpha(P_{X^nY^n}, P_{X^n}Q_{Y^n})$ is constant over all input distributions on the sphere when Q_{Y^n} is the capacity achieving output distribution. Hence to compute this for the AWGN channel, replacing P_{X^n} with $x^n = (\sqrt{P}, \ldots, \sqrt{P})$ can simplify computation. In the MIMO-BF channel, β_α is not longer constant over P_{X^n}, so we cannot use this simplification.

The real challenge is computing $\beta_\tau(P_{Y^nH^n}, Q_{Y^nH^n})$. There is no closed form expression for P_{Y^n}, hence we cannot directly compute the log likelihood ratio via monte carlo. A technique to compute this $\beta_\tau(P_{Y^nH^n}, Q_{Y^nH^n})$ is as follows: Let $P^*_{X^n}$ denote the distribution on $\mathbb{C}^{n_t \times nT}$ that has i.i.d. $\mathcal{CN}(0, P/n_t)$ entries. Then $Q_{Y^nH^n} = P^*_{X^n} \circ P_{Y^nH^n|X^n}$, and

$$P_{Y^nH^n} = P_{X^n} \circ P_{Y^nH^n|X^n} \tag{4.25}$$

$$= P^*_{X^n} \circ P^{(s)}_{Y^nH^n|X^n} \tag{4.26}$$

where $P_X \circ P_{Y|X}$ denotes the output distribution induced by P_X through $P_{Y|X}$, and here $P^{(s)}_{Y^nH^n|X^n}$ denotes the channel

$$Y^n = H^n X^n \frac{\sqrt{nTP}}{\|X^n\|_F} + Z^n \,. \tag{4.27}$$

I.e. we replace the channel by one that forces the input distribution to live on the sphere, and then choose the i.i.d. Gaussian input distribution for this channel. The motivation for doing this is the following – by the data processing inequality for $\beta_\alpha(P, Q)$,

$$\beta_\tau(P_{Y^nH^n}, Q_{Y^nH^n}) \geq \beta_\tau(P^*_{X^n}P^{(s)}_{Y^nH^n|X^n}, P^*_{X^n}P_{Y^nH^n|X^n}) \,. \tag{4.28}$$

i.e. the error in a binary hypothesis test between the true and auxiliary output distributions is lower bounded by the error of the hypothesis test between the joint distributions. Now, the log likelihood ratio of the right hand side can be computed

in closed form, i.e.

$$\log \frac{P^*_{X^n} P^{(s)}_{Y^n H^n|X^n}}{P^*_{X^n} P_{Y^n H^n|X^n}}(x^n, y^n, h^n) = \log \frac{\frac{1}{2\pi} \exp\left(-\left\|y^n - h^n x^n \frac{\sqrt{nTP}}{\|x^n\|}\right\|^2\right)}{\frac{1}{2\pi} \exp\left(-\|y^n - h^n x^n\|^2\right)} \tag{4.29}$$

$$= \log(e) \left(\|y^n - h^n x^n\|^2 - \left\|y^n - h^n x^n \frac{\sqrt{nTP}}{\|x^n\|}\right\|^2 \right). \tag{4.30}$$

Because of this form of the log likelihood ratio, $\beta_\tau(P^*_{X^n} P^{(s)}_{Y^n H^n|X^n}, P^*_{X^n} P_{Y^n H^n|X^n})$ becomes computable via Monte Carlo.

4.2.1 Renyi Divergence Method

There is a nice lower bound on the numerator in terms of the Renyi divergence:

Proposition 16. *For any P, Q such that $P \ll Q$, and any non-negative $\lambda > 1$,*

$$\beta_\tau(P, Q) \geq \tau^{\frac{\lambda}{\lambda-1}} \left(e^{(\lambda-1)D_\lambda(P\|Q)} - (1-\tau)^\lambda\right)^{-\frac{1}{\lambda-1}} \tag{4.31}$$

Where $D_\lambda(P||Q) = \frac{1}{\lambda-1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)}\right)^\alpha\right]$ is the Renyi divergence. This implies

$$\beta_\tau(P, Q) \geq \exp\left(-\frac{D_\lambda(P||Q) + h(\tau)}{\tau}\right) \tag{4.32}$$

Proof. We use the property that the Renyi divergence satisfies the data processing inequality, and use the test between P vs Q from the optimal hypothesis test to obtain

$$D_\lambda(P||Q) \geq d_\lambda(\tau||\beta_\tau) = \frac{1}{\lambda-1} \log\left(\tau^\lambda \beta_\tau^{1-\lambda} + (1-\tau)^\lambda (1-\beta_\tau)^{1-\lambda}\right) \tag{4.33}$$

We lower bound the RHS using $1/(1-\beta_\tau)^{\lambda-1} \geq 1$, then simply solve for β_τ in

$$D_\lambda(P||Q) \geq \frac{1}{\lambda-1} \log\left(\tau^\lambda \left(\frac{1}{\beta_\tau}\right)^{\lambda-1} + (1-\tau)^\lambda\right) \tag{4.34}$$

□

We can use the data processing trick to obtain a closed form upper bound for the Renyi divergence

$$D_\lambda(P_{Y^n H^n}||Q_{Y^n H^n}) \tag{4.35}$$

where again $Q_{Y^n H^n}$ is the caod, and $P_{Y^n H^n}$ induced by P_{X_S} uniform on $\sqrt{nP}S^{n-1}$.

Recall the data (un)processing trick: For two distributions P_Y,Q_Y, if we can view them as outputs of two different channels for the sample input distribution, i.e.

$$P_Y = P_X \circ P_{Y|X} \tag{4.36}$$
$$Q_Y = P_X \circ Q_{Y|X} \tag{4.37}$$

Then the data processing inequality tells us that, for any function f that satisfies data processing,

$$f(P_Y, Q_Y) \leq f(P_X P_{Y|X}, P_X Q_{Y|X}) \tag{4.38}$$

This is useful when the RHS is easy to compute and the LHS is difficult. An example of this is for the MIMO fading channel, both $\beta_\alpha(P_{Y^n H^n}, Q_{Y^n H^n})$ and $D(P_{Y^n H^n}||Q_{Y^n H^n})$ become easier to compute after noticing that each is the output distribution induced by an i.i.d. Gaussian input through

$$P_{Y^n H^n|X^n} \ : \ Y^n = H^n X^n \frac{\sqrt{nTP}}{\|X^n\|_F} + Z^n \tag{4.39}$$
$$Q_{Y^n H^n|X^n} \ : \ Y^n = H^n X^n + Z^n \tag{4.40}$$

Then we compute the Renyi divergence between these distributions,

$$D_\lambda(P_{X^n Y^n H^n}|Q_{X^n Y^n H^n}) = \frac{1}{\lambda - 1} \log \mathbb{E}\left[\left(\frac{P_{Y^n H^n|X^n}}{Q_{Y^n H^n|X^n}}(Y^n, H^n, X^n) \right)^{\lambda - 1} \right]. \tag{4.41}$$

Again, the log likelihood ratio takes a fairly simple form,

$$(\lambda - 1) \log \frac{P_{Y^n H^n|X^n}}{Q_{Y^n H^n|X^n}}(y^n, h^n, x^n) = (\lambda - 1)\frac{1}{2}\left(\|y^n - h^n x^n\|^2 - \left\| y^n - h^n x^n \frac{\sqrt{nP}}{\|x^n\|} \right\|^2 \right). \tag{4.42}$$

Under $P_{X^n Y^n H^n}$, this has distribution

$$\frac{(\lambda - 1)}{2}\left(\left\| Z^n + H^n X^n \left(\frac{\sqrt{nP}}{\|X^n\|} - 1 \right) \right\|^2 - \|Z^n\|^2 \right) \tag{4.43}$$
$$= \frac{\lambda - 1}{2}\left(\|H^n X^n\|^2 \left(\frac{\sqrt{nP}}{\|X^n\|} - 1 \right)^2 + 2 \sum_{i=1}^{n} Z_i H_i X_i \left(\frac{\sqrt{nP}}{\|X^n\|} - 1 \right) \right) \tag{4.44}$$

Denote, for notational simplicity,

$$S_n = \|H^n X^n\|^2 \left(\frac{\sqrt{nP}}{\|X^n\|} - 1 \right)^2 \tag{4.45}$$

And notice that expression (4.44) can be written as

$$\frac{\lambda-1}{2}\left(S_n+2\sqrt{S_n}Z\right) \tag{4.46}$$

Where here, $Z\sim\mathcal{N}(0,1)$. With this, our task is to compute

$$D_\lambda(P_{X^nY^nH^n}|Q_{X^nY^nH^n})=\frac{1}{\lambda-1}\log\mathbb{E}\left[e^{\frac{1}{2}(\lambda-1)\left(S_n+2\sqrt{S_n}Z\right)}\right] \tag{4.47}$$

To compute this, first taking expectations over Z, and either completing the square or noticing that this is simply the MGF, we obtain

$$\mathbb{E}\left[e^{\frac{1}{2}(\lambda-1)\left(S_n+2\sqrt{S_n}Z\right)}\right]=\mathbb{E}\left[e^{\frac{1}{2}(\lambda-1)S_n}\mathbb{E}\left[e^{(\lambda-1)\sqrt{S_n}Z}\middle|S_n\right]\right] \tag{4.48}$$

$$=\mathbb{E}\left[e^{\frac{1}{2}(\lambda-1)S_n+\frac{1}{2}(\lambda-1)^2S_n}\right] \tag{4.49}$$

$$=\mathbb{E}\left[e^{\frac{1}{2}\lambda(\lambda-1)S_n}\right]. \tag{4.50}$$

Plugging in S_n, we must compute

$$\mathbb{E}\left[e^{\frac{1}{2}\lambda(\lambda-1)(1-\frac{\sqrt{nP}}{\|X^n\|})^2\sum_{i=1}^n H_i^2X_i^2}\right] \tag{4.51}$$

Where $H_i\sim\mathcal{N}(0,1)$ for the Rayleigh case. Define

$$a=\lambda(\lambda-1)\left(1-\frac{\sqrt{nP}}{\|X^n\|}\right)^2 \tag{4.52}$$

Then, we're interested in

$$\mathbb{E}\left[e^{\frac{a}{2}\sum_{i=1}^n H_i^2X_i^2}\middle|X^n\right]=\frac{1}{\sqrt{2\pi}^n}\int\exp\left(\frac{a}{2}\sum_{i=1}^n h_i^2x_i^2-\frac{1}{2}\sum_{i=1}^n h_i^2\right)dh^n \tag{4.53}$$

$$=\frac{1}{\sqrt{2\pi}^n}\int\exp\left(-\frac{1}{2}\sum_{i=1}^n h_i^2(1-ax_i^2)\right)dh^n \tag{4.54}$$

which we can numerically approximate. Empirically, it looks like the Renyi divergence beats the non-Renyi divergence method by a decent margin, and about a factor of 2 away from computing $\log\beta_\tau(P_{X^nY^nH^n},Q_{X^nY^nH^n})$, done via monte carlo. The advantage of such a bound is that the Renyi divergence may be easier to compute that $\beta_\tau(P_Y,Q_Y)$, either in closed for or via Monte Carlo. The downside is that when τ is small, for small blocklengths, the bound can be fairly loose. We have used direct computation of $\log\beta_\tau(P_{X^nY^nH^n},Q_{X^nY^nH^n})$, but in cases where this cannot be computed, this Renyi divergence trick may help.

Chapter 5

Analysis of the Dispersion Expression

In Chapter 3, we established that the dispersion is given by

$$V = \min_{P_X\text{-caid}} \mathbb{E}\left[\mathrm{Var}(i(X;Y,H)|X)\right] . \tag{5.1}$$

Furthermore, Theorem 1 gave a characterization of the capacity achieving input distributions: when $\mathrm{rank}(H) > 1$ with positive probability, the caid is uniquely Gaussian, otherwise there are multiple caids. However, this form of the dispersion gives us no insight into the behavior of the channel as a function of the parameters, just as the capacity expression $C = \max_{P_X} I(X;Y,H)$ tells us nothing about how rate behaves as a function of number of antennas. In this chapter, we begin by giving an analytic expression for the dispersion. After that calculation, we look at what we learn from the form of this expression.

5.1 Calculation of the Dispersion

The main theorem of this section is a calculation of the MIMO dispersion expression:

Theorem 17. *Assume that* $\mathbb{P}[\mathrm{rank}\, H > 1] > 0$, *then* $V(P) = V_{iid}(P)$, *where*

$$\begin{aligned} V_{iid}(P) = T\mathrm{Var}&\left(\sum_{i=1}^{n_{\min}} C_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right) \\ &+ \sum_{i=1}^{n_{\min}} \mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right] \\ &+ \left(\frac{P}{n_t}\right)^2 \left(\eta_1 - \frac{\eta_2}{n_t}\right) \end{aligned} \tag{5.2}$$

where $\{\Lambda_i^2, i = 1, \ldots, n_{\min}\}$ *are eigenvalues of* HH^T, $V_{AWGN}(P) = \frac{\log^2 e}{2}\left(1 - \frac{1}{(1+P)^2}\right)$,

and

$$c(\sigma) \triangleq \frac{\sigma}{1+\frac{P}{n_t}\sigma} \tag{5.3}$$

$$\eta_1 \triangleq \frac{\log^2 e}{2} \sum_{i=1}^{n_{\min}} \mathbb{E}\left[c^2(\Lambda_i^2)\right] \tag{5.4}$$

$$\eta_2 \triangleq \frac{\log^2 e}{2} \left(\sum_{i=1}^{n_{\min}} \mathbb{E}\left[c(\Lambda_i^2)\right] \right)^2 \tag{5.5}$$

The proof is given in terms of a series of propositions.

Proposition 18. *Let* $V_n(x^n) \triangleq \frac{1}{nT}\mathrm{Var}(i(X^n;Y^n,H^n)|X^n=x^n)$*, then we have*

$$V_n(x^n) = \frac{1}{n}\sum_{j=1}^{n} V_1(x_j)\,, \tag{5.6}$$

where the function $V_1 : \mathbb{R}^{n_t \times T} \mapsto \mathbb{R}$ *defined as* $V_1(x) \triangleq \frac{1}{T}\mathrm{Var}(i(X;Y,H)|X=x)$ *is given by*

$$V_1(x) = T\mathrm{Var}\left(C_r(H,P)\right) \tag{5.7}$$

$$+ \sum_{i=1}^{n_{\min}} \mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right)\right] \tag{5.8}$$

$$+ \eta_5 \left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t} \right) \tag{5.9}$$

$$+ \eta_3 \left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t} \right)^2 \tag{5.10}$$

$$+ \eta_4 \left(\|xx^T\|_F^2 - \frac{1}{n_t}\|x\|_F^4 \right) \tag{5.11}$$

where $c(\cdot)$ *was defined in* (5.3) *and* $C_r(H,P), \eta_3, \eta_4, \eta_5$ *are given by* (5.12)-(5.15)*.*

Remark 8. Every term in the definition of $V_1(x)$ (except the one with η_5) is non-negative (for η_4-term, see (5.41)). The η_5-term will not be important because for inputs satisfying power-constraint with equality it vanishes. Note also that the first term in (5.15) can alternatively be given as

$$\mathrm{Cov}\left(C_r(H,P), \sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right) \right) = n_t \frac{d}{dP}\mathrm{Var}\left[C_r(H,P)\right]\,.$$

Proof. From (2.40), we have the form of the information density. First note that the information density over n channel uses decomposes into a sum of n independent

terms,

$$i(x^n; Y^n, H^n) = \sum_{j=1}^{n} i(x_j, Y_j, H_j)\,. \tag{5.16}$$

As such, the variance conditioned on x^n also decomposes as

$$\mathrm{Var}(i(x^n; Y^n, H^n)) = \sum_{j=1}^{n} \mathrm{Var}(i(x_j; Y_j, H_j))\,, \tag{5.17}$$

from which (5.6) follows. Because the variance decomposes as a sum in (5.17), we focus on only computing $\mathrm{Var}(i(x; Y, H))$ for a single coherent block. Define

$$f(h) \triangleq TC_r(h, P) \tag{5.18}$$

$$g(x, h, z) \triangleq \frac{\log e}{2} \sum_{k=1}^{n_{\min}} \frac{\Lambda_k^2 \|v_k^T x\|^2 + 2\Lambda_k \langle v_k^T x, \tilde{z}_k \rangle - \frac{P}{n_t}\Lambda_k^2 \|\tilde{z}_k\|^2}{1 + \frac{P}{n_t}\Lambda_k^2} \tag{5.19}$$

so that $i(x; y, h) = f(h) + g(x, h, z)$ in notation from (2.40). With this, the quantity of interest is

$$\mathrm{Var}(i(x, Y, H)) \tag{5.20}$$

$$= \mathrm{Var}(f(H)) + \mathrm{Var}(g(x, H, Z)) + \mathrm{Cov}(f(H), g(x, H, Z)) \tag{5.21}$$

$$= \underbrace{\mathrm{Cov}(f(H), g(x, H, Z))}_{\triangleq T_1} + \underbrace{\mathrm{Var}(f(H))}_{\triangleq T_2}$$

$$+ \underbrace{\mathrm{Var}\left(\mathbb{E}[g(x, H, Z)|H]\right)}_{\triangleq T_3} + \underbrace{\mathbb{E}\left[\mathrm{Var}(g(x, H, Z)|H)\right]}_{\triangleq T_4} \tag{5.22}$$

$$C_r(H, P) \triangleq \frac{1}{2} \log\det\left(I_{n_r} + \frac{P}{n_t} HH^T\right) = \sum_{i=1}^{n_{\min}} C_{AWGN}\left(\frac{P}{n_t}\Lambda_i^2\right) \tag{5.12}$$

$$\eta_3 \triangleq \frac{\log^2 e}{4} \mathrm{Var}\left(\sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right)\right) \tag{5.13}$$

$$\eta_4 \triangleq \frac{\log^2 e}{2n_t(n_t+2)} \left(\mathbb{E}\left[\sum_{i=1}^{n_{\min}} c^2(\Lambda_i^2)\right] - \frac{1}{(n_t-1)} \sum_{i \neq j} \mathbb{E}\left[c\left(\Lambda_i^2\right) c\left(\Lambda_j^2\right)\right]\right) \tag{5.14}$$

$$\eta_5 \triangleq \frac{\log e}{2} \mathrm{Cov}\left(C_r(H, P), \sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right)\right) + \frac{\log^2 e}{T} \sum_{k=1}^{n_{\min}} \mathbb{E}\left[\frac{\Lambda_k^2}{\left(1 + \frac{P}{n_t}\Lambda_k^2\right)^2}\right]\,. \tag{5.15}$$

where (5.22) follows from the identity

$$\begin{aligned}\mathrm{Var}(g(x,H,Z)) &= \mathbb{E}\left[\mathrm{Var}(g(x,H,Z)|H)\right] \\ &\quad + \mathrm{Var}\left(\mathbb{E}[g(x,H,Z)|H]\right) \ . \end{aligned} \tag{5.23}$$

Below we show that T_1 and T_3 corresponds to (5.9), T_2 corresponds to (5.7), T_4 corresponds to (5.8), and T_3 corresponds to (5.10) and (5.11). We evaluate each term separately.

$$\begin{aligned} T_1 &= \mathrm{Cov}(f(H), g(x,H,Z)) && (5.24)\\ &= \mathbb{E}\left[(f(H) - \mathbb{E}[f(H)])(g(x,H,Z) - \mathbb{E}[g(x,H,Z)])\right] && (5.25)\\ &= \frac{\log e}{2}\left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right) \\ &\quad \sum_{k=1}^{n_{\min}} \mathbb{E}\left[(f(H) - \mathbb{E}[f(H)])(c\left(\Lambda_k^2\right) - \mathbb{E}[c\left(\Lambda_k^2\right)])\right] && (5.26)\\ &= \frac{\log e}{2}\left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right) \sum_{k=1}^{n_{\min}} \mathrm{Cov}\left(f(H), c\left(\Lambda_k^2\right)\right) && (5.27) \end{aligned}$$

where (5.26) follows from noting that

$$\mathbb{E}\left[g(x,H,Z)|H\right] = \sum_{k=1}^{n_{\min}} \left(\|V_k^T x\|^2 - \frac{TP}{n_t}\right) c\left(\Lambda_k^2\right) \frac{\log e}{2} \ . \tag{5.28}$$

Now, since V_k is independent from Λ_k by the rotational invariance assumption, we have that $f(H)$ is independent from V_k, since $f(H)$ only depends on H through its eigenvalues, see (5.12). We are only concerned with the expectation over $g(x,H,Z)$ in (5.25), which reduces to

$$\begin{aligned} &\mathbb{E}\left[g(x,H,Z) - \mathbb{E}[g(x,H,Z)]|\Lambda_1, \ldots, \Lambda_{n_{\min}}\right] \\ &= \left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right) \sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right) - \mathbb{E}[c\left(\Lambda_k\right)^2] \frac{\log e}{2} \ , \end{aligned} \tag{5.29}$$

giving (5.26).

Next, T_2 in (5.22) becomes

$$\begin{aligned} T_2 &= \mathrm{Var}(f(H)) && (5.30)\\ &= T^2 \mathrm{Var}\left(\sum_{k=1}^{n_{\min}} C_{AWGN}\left(\frac{P}{n_t}\Lambda_k^2\right)\right) \ . && (5.31) \end{aligned}$$

For T_3 in (5.22), we obtain

$$T_3 = \mathbb{E}\left[\mathrm{Var}(g(x,H,Z)|H)\right] \tag{5.32}$$

$$= \frac{\log^2 e}{4}\mathbb{E}\left[\sum_{k=1}^{n_{\min}} \frac{4\Lambda_k^2\|V_k^T x\|^2 + 2T\left(\frac{P}{n_t}\right)^2\Lambda_k^4}{\left(1+\frac{P}{n_t}\Lambda_k\right)^2}\right] \tag{5.33}$$

$$= \frac{\log^2 e}{2}\sum_{k=1}^{n_{\min}} T\mathbb{E}\left[\frac{2\frac{TP}{n_t}\Lambda_k^2 + T\left(\frac{P}{n_t}\right)^2\Lambda_k^4}{\left(1+\frac{P}{n_t}\Lambda_k\right)^2}\right]$$

$$+ 2\mathbb{E}\left[\frac{\frac{\|x\|_F^2}{n_t}\Lambda_k^2 - \frac{TP}{n_t}\Lambda_k^2}{\left(1+\frac{P}{n_t}\Lambda_k^2\right)^2}\right] \tag{5.34}$$

$$= T\sum_{k=1}^{n_{\min}} V_{AWGN}\left(\frac{P}{n_t}\Lambda_k^2\right)$$

$$+ \log^2(e)\left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right)\mathbb{E}\left[\frac{\Lambda_k^2}{\left(1+\frac{P}{n_t}\Lambda_k^2\right)^2}\right], \tag{5.35}$$

where

- (5.33) follows from taking the variance over $\tilde{Z}$ (recall $\tilde{Z} = U^T Z$ in (2.40)).
- (5.34) follows from Lemma 2 applied to $\mathbb{E}[\|V_k^T x\|^2]$, and adding and subtracting the term

$$\log^2(e)\mathbb{E}\left[\frac{\frac{TP}{n_t}\Lambda_k^2}{\left(1+\frac{P}{n_t}\Lambda_k^2\right)^2}\right]. \tag{5.36}$$

Continuing with T_3 from (5.22),

$$T_3 = \mathrm{Var}\mathbb{E}[g(x,H,Z)|H] \tag{5.37}$$

$$= \mathrm{Var}\left(\frac{\log e}{2}\sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right)\left(\|V_k^T x\|^2 - \frac{TP}{n_t}\right)\right) \tag{5.38}$$

$$= \eta_3\left(\frac{\|x\|_F^2}{n_t} - \frac{TP}{n_t}\right)^2$$

$$+ \frac{\log^2 e}{4}\mathbb{E}\left[\mathrm{Var}\left(\sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right)\|V_k^T x\|^2 \middle| \Lambda_1, \ldots, \Lambda_{n_{\min}}\right)\right] \tag{5.39}$$

where

- (5.38) follows from taking the expectation over $\tilde{Z}$,
- (5.39) follows from applying the variance identity (5.23) with respect to V and $\Lambda_1, \ldots, \Lambda_{n_{\min}}$, as well as recalling (5.13).

We are left to show that the term (5.39) equals (5.11). To that end, define

$$\phi(x) \triangleq \mathbb{E}\left[\mathrm{Var}\left(\sum_{k=1}^{n_{\min}} c\left(\Lambda_k^2\right) \|V_k^T x\|^2 \middle| \Lambda_1, \ldots, \Lambda_{n_{\min}}\right)\right] \tag{5.40}$$

$$= \sum_{k=1}^{n_{\min}} \mathbb{E}[c^2(\Lambda_k^2)]\mathrm{Var}\left(\|V_k^T x\|^2\right)$$

$$+ \sum_{k \neq l}^{n_{\min}} \mathbb{E}\left[c\left(\Lambda_k^2\right) c\left(\Lambda_l^2\right)\right] \mathrm{Cov}(\|V_k^T x\|^2, \|V_l^T x\|^2) \ . \tag{5.41}$$

We will finish the proof by showing

$$\phi(x) = \frac{4}{\log^2 e} \eta_4 \left(\|xx^T\|_F^2 - \frac{1}{n_t}\|x\|_F^4\right) \ .$$

To that end, we first compute moments of V drawn from the Haar measure on the orthogonal group.

Lemma 19. *Let V be drawn from the Haar measure on $O(n)$, then for $i, j, k, l = 1, \ldots, n$ all unique,*

$$\mathbb{E}[V_{ij}^2] = \frac{1}{n} \tag{5.42}$$

$$\mathbb{E}[V_{ij} V_{ik}] = 0 \tag{5.43}$$

$$\mathbb{E}[V_{ij}^2 V_{ik}^2] = \frac{1}{n(n+2)} \tag{5.44}$$

$$\mathbb{E}[V_{ij}^2 V_{kl}^2] = \frac{n+1}{n(n-1)(n+2)} \tag{5.45}$$

$$\mathbb{E}[V_{ij}^4] = \frac{3}{n(n+2)} \tag{5.46}$$

$$\mathbb{E}[V_{ij} V_{ik} V_{lj} V_{lk}] = \frac{-1}{n(n-1)(n+2)} \ . \tag{5.47}$$

Proof of this Lemma is given below.

First, note that the variance $\mathrm{Var}(\|V_k^T x\|^2)$ does not depend on k, since the marginal distribution of each V_k is uniform on the unit sphere. Hence below we only consider

V_1. We obtain

$$\begin{aligned}
&\mathrm{Var}(\|V_1^T x\|^2) \\
&= \mathbb{E}[\|V_1^T x\|^4] - \mathbb{E}^2[\|V_1^T x\|^2] && (5.48)\\
&= \mathbb{E}\left[\left(\sum_{i=1}^{T}\sum_{j=1}^{n_t}\sum_{k=1}^{n_t} V_{j1}V_{k1}x_{ji}x_{ki}\right)^2\right] - \frac{\|x\|_F^4}{n_t^2} && (5.49)\\
&= \mathbb{E}\left[\sum_{j=1}^{n_t}\sum_{k=1}^{n_t}\sum_{l=1}^{n_t}\sum_{m=1}^{n_t} V_{j1}V_{k1}V_{l1}V_{m1}\langle r_j, r_k\rangle\langle r_l, r_m\rangle\right] && (5.50)
\end{aligned}$$

where r_j denotes the j-th row of x. Now it is a matter counting similar terms:

$$\begin{aligned}
\mathbb{E}[\|V_1^T x\|^4] &= \sum_{j=1}^{n_t}\mathbb{E}[V_{j1}^4]\|r_j\|^4 + 2\sum_{j\neq k}^{n_t}\mathbb{E}[V_{j1}^2V_{k1}^2]\langle r_j, r_k\rangle^2 \\
&\quad + \sum_{j\neq k}^{n_t}\mathbb{E}[V_{j1}^2V_{k1}^2]\|r_j\|^2\|r_k\|^2 && (5.51)\\
&= \frac{3}{n_t(n_t+2)}\sum_{j=1}^{n_t}\|r_j\|^4 + \frac{2}{n_t(n_t+2)}\sum_{j\neq k}^{n_t}\langle r_j, r_k\rangle^2 \\
&\quad + \frac{1}{n_t(n_t+2)}\sum_{j\neq k}\|r_j\|^2\|r_k\|^2 && (5.52)\\
&= \frac{1}{n_t(n_t+2)}\left(\|x\|_F^4 + 2\|xx^T\|_F^2\right) && (5.53)
\end{aligned}$$

where

- (5.51) follows from collecting like terms from the summation in (5.50).
- (5.52) uses Lemma 19 to compute each expectation.
- (5.53) follows from realizing that

$$\|x\|_F^4 = \left(\sum_{j=1}^{n_t}\|r_j\|^2\right)^2 = \sum_{j=1}^{n_t}\|r_j\|^4 + \sum_{j\neq k}^{n_t}\|r_j\|^2\|r_k\|^2 \quad (5.54)$$

$$\|xx^T\|_F^2 = \sum_{j=1}^{n_t}\sum_{k=1}^{n_t}\langle r_j, r_k\rangle^2 = \sum_{j=1}^{n_t}\|r_j\|^4 + \sum_{j\neq k}^{n_t}\langle r_j, r_k\rangle^2 \quad (5.55)$$

Plugging this back into (5.48) yields the variance term,

$$\begin{aligned}\mathrm{Var}(\|V_1^T x\|^2) &= \frac{1}{n_t(n_t+2)}\left(\|x\|_F^4 + 2\|xx^T\|_F^2\right) - \frac{\|x\|_F^4}{n_t^2} \\ &= \frac{2}{n_t(n_t+2)}\left(\|xx^T\|_F^2 - \frac{\|x\|_F^4}{n_t}\right) . \end{aligned} \tag{5.56}$$

Now we compute the covariance term from (5.41) in a similar way. By symmetry of the columns of V, we can consider only the covariance between $\|V_1^T x\|^2$ and $\|V_2^T x\|^2$, i.e.

$$\mathrm{Cov}(\|V_1^T x\|^2, \|V_2^T x\|^2) = \mathbb{E}[\|V_1^2 x\|^2 \|V_2^T x\|^2] - \frac{\|x\|_F^4}{n_t^2} . \tag{5.57}$$

Expanding the expectation, we get

$$\mathbb{E}[\|V_1^T x\|^2 \|V_2^T x\|^2] \tag{5.58}$$

$$= \sum_{j,k,l,m} \mathbb{E}[V_{1j}V_{1k}V_{2l}V_{2m}]\langle r_j, r_k\rangle\langle r_l, r_m\rangle \tag{5.59}$$

$$\begin{aligned}&= \sum_{j=1}^{n_t} \mathbb{E}[V_{1j}^4]\|r_j\|^4 + \sum_{j\neq k} \mathbb{E}[V_{1j}^2 V_{2k}^2]\|r_j\|^2\|r_k\|^2 \\ &+ 2\sum_{j\neq k} \mathbb{E}[V_{1j}V_{1k}V_{2j}V_{2k}]\langle r_j, r_k\rangle^2 \end{aligned} \tag{5.60}$$

$$\begin{aligned}&= \frac{1}{n_t(n_t+2)}\sum_{j=1}^{n_t}\|r_j\|^4 + \frac{n_t+1}{(n_t-1)n_t(n_t+2)}\sum_{j\neq k}\|r_j\|^2\|r_k\|^2 \\ &- \frac{2}{(n_t-1)n_t(n_t+2)}\sum_{j\neq k}\langle r_j, r_k\rangle^2 \end{aligned} \tag{5.61}$$

$$= \frac{1}{(n_t-1)n_t(n_t+2)}\left((n_t+1)\|x\|_F^4 - 2\|xx^T\|_F^2\right) . \tag{5.62}$$

With this, we obtain from (5.57),

$$\begin{aligned}&\mathrm{Cov}(\|V_1^T x\|^2, \|V_2^T x\|^2) \\ &= \frac{2}{(n_t-1)n_t(n_t+2)}\left(\frac{\|x\|_F^4}{n_t} - \|xx^T\|_F^2\right) \end{aligned} \tag{5.63}$$

where the steps follow just as in the variance computation (5.51)-(5.53).

Finally, returning to (5.41), using the variance (5.56) and covariance (5.63), we

obtain

$$\phi(x) = \frac{2}{n_t(n_t+2)} \left(\|xx^T\|_F^2 - \frac{\|x\|_F^4}{n_t} \right) \left(\sum_{k=1}^{n_t} \mathbb{E}[c^2(\Lambda_k^2)] - \frac{1}{n_t - 1} \sum_{k \neq l} \mathbb{E}\left[c\left(\Lambda_k^2\right) c\left(\Lambda_l^2\right) \right] \right) . \tag{5.64}$$

Plugging this into (5.39) finishes the proof. □

Proof of Lemma 19. We first note that all entries of V have identical distribution, since permutations of rows and columns leave the distribution invariant. Because of this, we can WLOG only consider $V_{11}, V_{12}, V_{21}, V_{22}$ to prove the lemma.

- (5.42) follows immediately from $\sum_{i=1}^{n} V_{ij}^2 = 1$ a.s.
- Let V_i, V_j be any two distinct columns of V, then (5.43) follows from

$$0 = \mathbb{E}[\langle V_i, V_j \rangle] = n\mathbb{E}[V_{11}V_{21}] \tag{5.65}$$

- For (5.44) and (5.47), let $E_1 = \mathbb{E}[V_{11}^4]$ and $E_2 = \mathbb{E}[V_{11}^2 V_{21}^2]$. The following relations between E_1, E_2 hold,

$$1 = \mathbb{E}\left[\left(\sum_{j=1}^{n} V_{1j}^2 \right)^2 \right] \tag{5.66}$$

$$= nE_1 + n(n-1)E_2 . \tag{5.67}$$

By multiplying V by the orthogonal matrix matrix

$$\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & I_{n-2} \end{bmatrix} \tag{5.68}$$

where I_n is the $n \times n$ identity matrix, we obtain the extra relation

$$E_1 = \mathbb{E}\left[\left(\frac{V_{11}}{\sqrt{2}} + \frac{V_{12}}{\sqrt{2}} \right)^4 \right] \tag{5.69}$$

$$= \frac{1}{2}E_1 + \frac{3}{2}E_2 \tag{5.70}$$

from which we obtain $E_1 = 3E_2$. With this and (5.67), we obtain

$$E_1 = \frac{3}{n(n+2)} \tag{5.71}$$

$$E_2 = \frac{1}{n(n+2)} . \tag{5.72}$$

- For (5.45), take

$$E_3 = \mathbb{E}[V_{11}^2 V_{22}^2] \tag{5.73}$$

$$= \mathbb{E}\left[V_{11}^2\left(1 - \sum_{j\neq 2}^{n} V_{2j}^2\right)\right] \tag{5.74}$$

$$= \frac{1}{n} - \frac{1}{n(n+2)} - (n-2)E_3 \,. \tag{5.75}$$

  Solving for E_3 yields (5.45).

- For (5.47), let V_1, V_2 denote the first and second column of V respectively, and let $E_4 = \mathbb{E}[V_{11}V_{12}V_{21}V_{22}]$, then (5.47) follows from

$$0 = \mathbb{E}[\langle V_1, V_2\rangle^2] \tag{5.76}$$

$$= nE_2 + n(n-1)E_4 \,. \tag{5.77}$$

  Using E_2 from (5.72) and solving for E_4 gives (5.47).

□

The following proposition gives the value of the conditional variance of the information density when input distribution has i.i.d. $\mathcal{N}(0, P/n_t)$ entries. This will turn out to be the operational dispersion in the case where rank $H > 1$.

Proposition 20. *Let $X^n = (X_1, \ldots, X_n)$ be i.i.d. with Telatar distribution* (2.9) *for each entry. Then*

$$\mathbb{E}\left[\mathrm{Var}(i(X^n; Y^n, H^n)|X^n)\right] = nTV_{iid}(P)\,, \tag{5.78}$$

where $V_{iid}(P)$ is the right-hand side of (5.2).

Proof. To show this, we take the expectation of the expression given in Proposition 18 when X^n has i.i.d. $\mathcal{N}(0, P/n_t)$ entries. The terms (5.7) and (5.8) do not depend on X^n, and these give us the first two terms in (5.2). (5.9) vanishes immediately, since $\mathbb{E}[\|X\|_F^2] = TP$ by the power constraint. It is left to compute the expectation over (5.10) and (5.11) from the expression in Proposition 18. Using identities for χ^2 distributed random variables (namely, $\mathbb{E}[\chi_k^2] = k$, $\mathrm{Var}(\chi_k^2) = 2k$), we get:

$$\frac{\eta_3}{n_t^2}\mathrm{Var}(\|X_1\|_F^2) = \frac{\eta_3}{n_t}\left(\frac{P}{n_t}\right)^2 2T \tag{5.79}$$

$$\mathbb{E}[\|X_1\|_F^4] = TP^2\left(T + \frac{2}{n_t}\right) \tag{5.80}$$

$$\mathbb{E}[\|X_1 X_1^T\|_F^2] = n_t T\left(\frac{P}{n_t}\right)^2 (1 + T + n_t) \tag{5.81}$$

$$\mathbb{E}\left[\|X_1 X_1^T\|_F^2 - \frac{\|X_1\|_F^4}{n_t}\right] = T\left(\frac{P}{n_t}\right)^2 (n_t - 1)(n_t + 2)\,. \tag{5.82}$$

Hence, the sum of terms in (5.10) + (5.11) after taking expectation over X^n yields

$$T\left(\frac{P}{n_t}\right)^2\left[2\frac{\eta_3}{n_t}+(n_t-1)(n_t+2)\eta_4\right].$$

Introducing random variables $U_i = c(\Lambda_i^2)$ the expression in the square brackets equals

$$\frac{\log^2 e}{2}\frac{1}{n_t}\left[\mathrm{Var}\left(\sum_i U_i\right)+(n_t-1)\sum_i \mathbb{E}\left[U_i^2\right]\right.$$
$$\left.-\sum_{i\neq j}\mathbb{E}\left[U_iU_j\right]\right]. \tag{5.83}$$

At the same time, the third term in expression (5.2) is

$$\frac{\log^2 e}{2}\frac{1}{n_t}\left[n_t\sum_i \mathbb{E}\left[U_i^2\right]-\left(\sum_i \mathbb{E}\left[U_i\right]\right)^2\right]. \tag{5.84}$$

One easily checks that (5.83) and (5.84) are equal. □

The next proposition shows that, when the rank of H is larger than 1, the conditional variance in (3.6) is constant over the set of caids. Thus we can compute the conditional variance for the i.i.d. $\mathcal{N}(0, P/n_t)$ caid, and conclude that this expression is the minimizer in (3.6).

Proof of Theorem 17. For any caid the term (5.9) vanishes. Let X^* be Telatar distributed. To analyze (5.10) we recall that from (2.22) we have

$$\mathbb{E}\left[\|X\|_F^4\right]=\sum_{i,j,i',j'}\mathbb{E}\left[X_{i,j}^2X_{i',j'}^2\right]=\mathbb{E}\left[\|X^*\|_F^4\right].$$

For the term (5.11) we notice that

$$\|XX^T\|_F^2=\sum_{i,j}\langle R_i,R_j\rangle^2,$$

where R_i is the i-th row of X. By (2.21) from Theorem 1 we then also have

$$\mathbb{E}\left[\|XX^T\|_F^2\right]=\mathbb{E}\left[\|X^*X^{*T}\|_F^2\right].$$

To conclude, $\mathbb{E}\left[V_1(X)\right]=\mathbb{E}\left[V_1(X^*)\right]=V_{iid}(P)$. □

5.2 Intuition about the Dispersion Expression

Just as there was interesting design intuition to gain from Telatar's capacity result, by looking at the dispersion expression found in Theorem 17, we can gain some additional

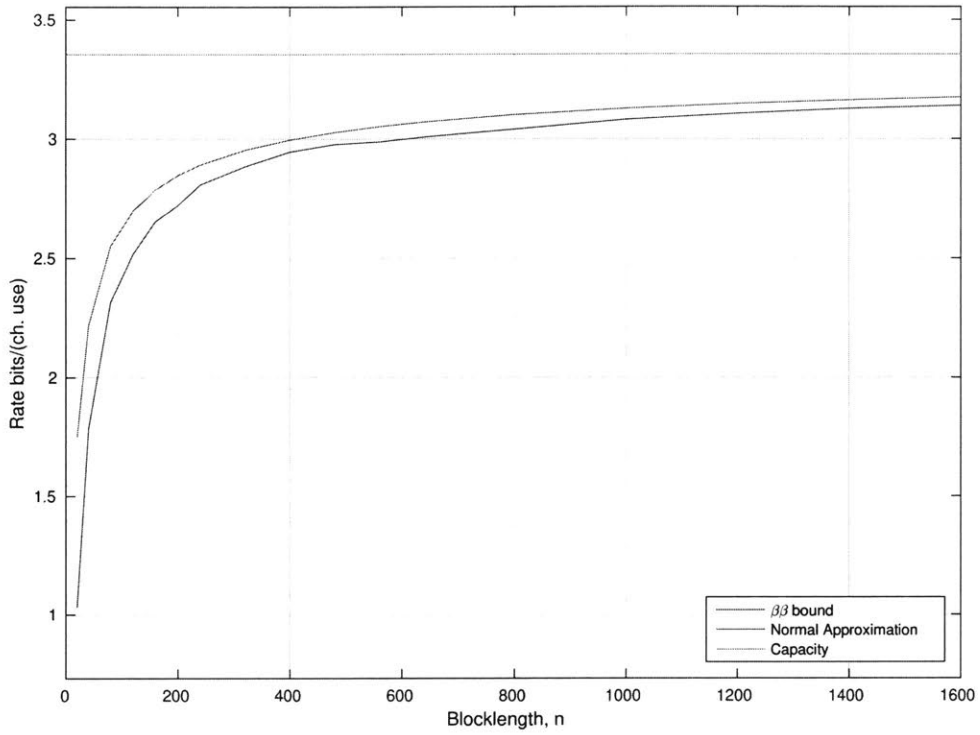

Figure 5-1: Achievability and normal approximation for $n_t = n_r = T = 4$, $P = 0$dB, and $\epsilon = 10^{-3}$.

insights about the MIMO-BF channel.

Figure 5-1 plots the capacity, normal approximation, and $\beta\beta$ achievability bound for the MIMO channel with $n_t = n_r = T = 4$ for the complex case. The details of this computation are given in [21]. The $\beta\beta$ bound was developed by Yang et al [21] and is often more computationally friendly than the $\kappa\beta$ bound. This figure illustrates the gap between achievability and the normal approximation, as well as the gap to capacity. For example, at blocklength 400, we can achieve about 88% of capacity, and at blocklength 1000 we can achieve about about 92% of capacity, given $P = 0$dB and tolerating an error probability of 10^{-3}.

Figure 5-2 shows the dependence of the rate on the coherence time T for the 4×4 MIMO channel. The normal approximation for $T = 1, 20, 80$ is plotted. From (2.10) and (5.2), we know the capacity does not depend on T, but the dispersion depends on T in an affine relationship. Hence, from the dispersion we see that a larger coherence time reduces the maximum transmission rate when the other channel parameters are held fixed. Intuitively, when the coherence time is lower, we are able to average over independent realizations of the fading coefficients in less channel uses. Note that the CSIR assumption implies that we know the channel coefficient perfectly, which may be unrealistic at short coherence times for a practical channel.

We now ask: how does the dispersion depend on the number of transmit and

receive antennas? Figures 5-3 and 5-4 depict the normalized dispersion V/C^2 as a function of the number of antennas. The normalized dispersion gives us a measure of the blocklength needed to achieve a fraction $\eta \in (0,1)$ of capacity, via

$$n \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1-\eta}\right)^2 \frac{V}{C^2}. \tag{5.85}$$

The fading process is chosen to be i.i.d. $\mathcal{N}(0,1)$. Each plot has two curves: one curve with n_r fixed and n_t growing, and the other curve with n_t fixed and n_r growing. In both plots, coherence time is $T = 16$. The difference is that on Fig. 5-3 the *received power* P_r is held fixed (at 20 dB, i.e. P is chosen so that $P_r = 100$), whereas on Fig 5-4 it is the transmit power P that is held fixed (also at 20 dB, i.e. $P = 100$). The relation between P_r and P is as follows:

$$P_r = \frac{P}{n_t}\mathbb{E}[\|H\|_F^2], \tag{5.86}$$

These figures also display the asymptotic limiting values of $\frac{V}{C^2}$ computed via random-matrix theory:

1. When n_r is fixed and $n_t \to \infty$ under fixed *received power* P_r we have

$$C(P_r) = \frac{n_r}{2}\log\left(1+\frac{P_r}{n_r}\right) + o(1) \tag{5.87}$$

$$V(P_r) = \log^2(e)\frac{P_r}{1+\frac{P_r}{n_r}} + o(1)\,. \tag{5.88}$$

2. When n_t is fixed and $n_r \to \infty$ under fixed *received power* P_r we have

$$C(P_r) = \frac{n_t}{2}\log\left(1+\frac{P_r}{n_t}\right) + o(1) \tag{5.89}$$

$$V(P_r) = \log^2(e)\frac{P_r(2+\frac{P_r}{n_t})}{2(1+\frac{P_r}{n_t})^2} + o(1)\,. \tag{5.90}$$

3. When n_r is fixed and $n_t \to \infty$ under fixed *transmitted power* P we have

$$C(P) = \frac{n_r}{2}\log(1+P) + o(1) \tag{5.91}$$

$$V(P) = \log^2(e)\frac{n_r P}{1+P} + o(1)\,. \tag{5.92}$$

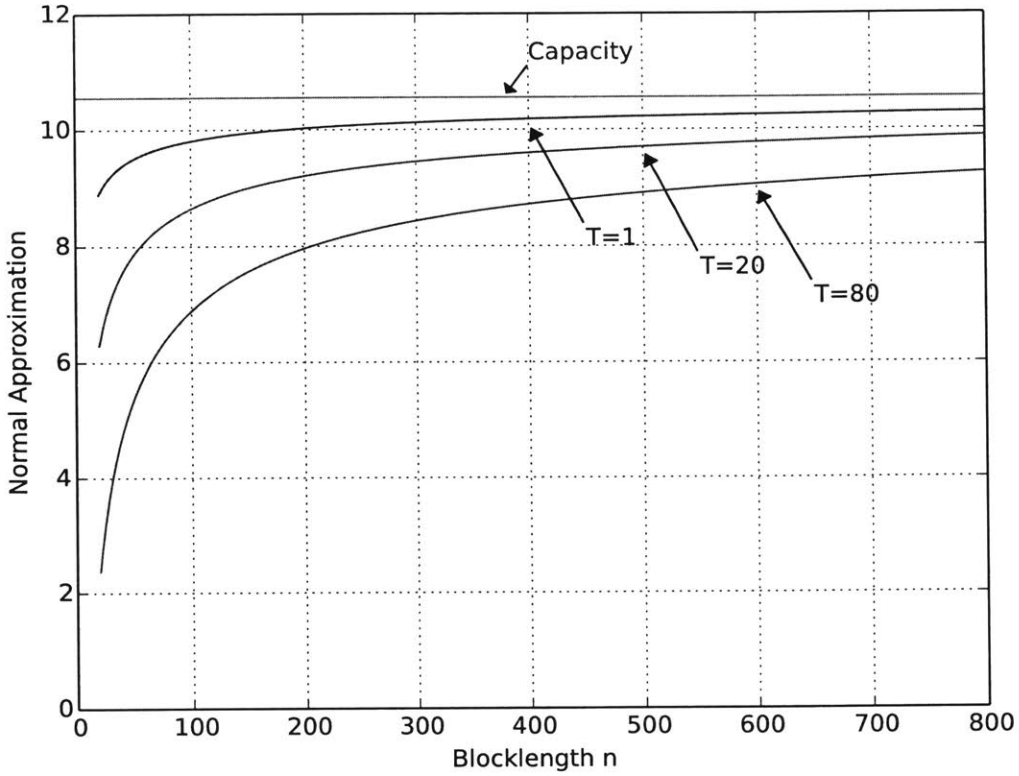

Figure 5-2: The normal approximation for varying coherent times, with $n_t = n_r = 4$, $P = 20dB$, and $\epsilon = 10^{-3}$

4. When n_t is fixed and $n_r \to \infty$ under fixed *transmit power* P we have

$$C(P) = \frac{n_t}{2} \log\left(1 + \frac{n_r P}{n_t}\right) + o(1) \tag{5.93}$$

$$V(P) = \log^2(e)\frac{n_t}{2} + o(1)\,. \tag{5.94}$$

is the same as the capacity of the $n_r \times n_t$ one. Having information about dispersion, we may ask the more refined question: although capacities of the channels are the same, which one has better dispersion (i.e. causes smaller coding latency)?

From approximations (5.88) and (5.90), we can see that the channel dispersion is not symmetric in n_t, n_r. For example, in the setting of Fig. 5-3 we see that the delay penalty in the $n_t \ll n_r$ regime is 58% of the penalty in the $n_r \ll n_t$ regime. Hence, in a two user channel, if user 1 has n_1 antennas and user 2 has $n_2 > n_1$ antennas, then the asymptotic analysis suggest that channel from user 1 to user 2 can support higher rates than the channel from user 2 to user 1 at finite blocklength.

Figure 5-4 shows the scenario where the transmit power is fixed. In this case, the capacity approaches a finite limit when n_r is held fixed and $n_t \to \infty$, but grows logarithmically when n_t is fixed and $n_r \to \infty$, as shown in equations (5.91) and (5.93).

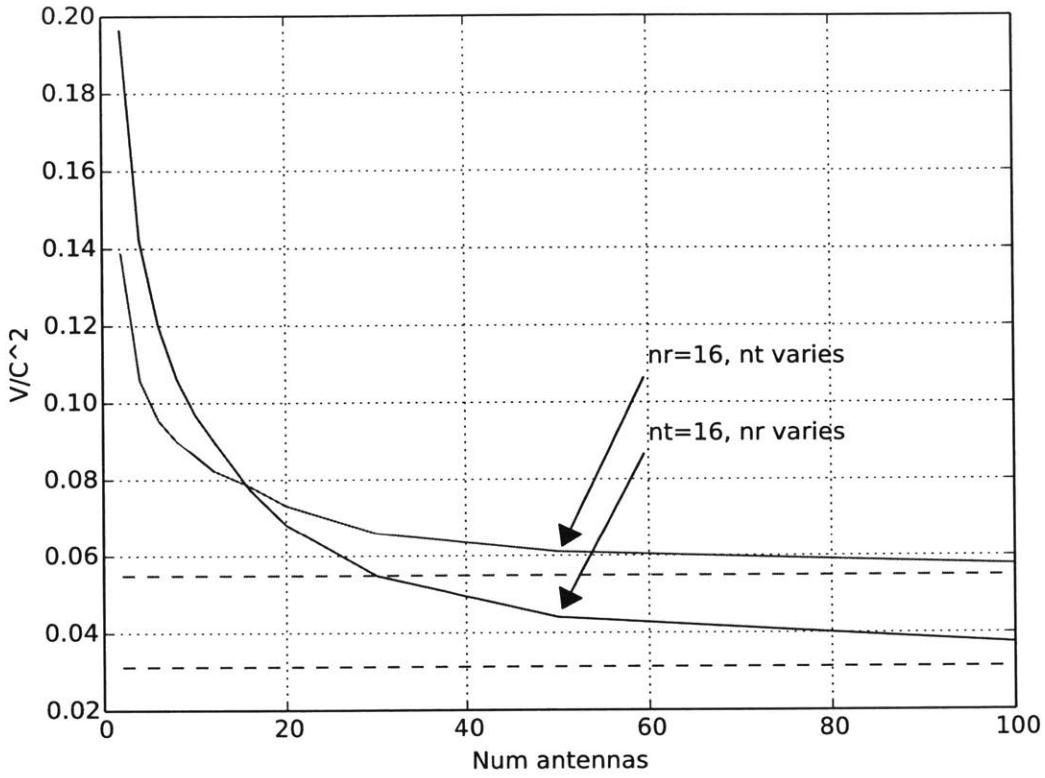

Figure 5-3: Normalized dispersion $\frac{V}{C^2}$ as a function of n_r and n_t. The received power is $P_r = 20dB$ and $T = 16$. Dashed lines are asymptotic values from (5.87)-(5.90).

In this setting, the normalized dispersion approaches a finite limit when n_r is fixed and $n_t \to \infty$, yet it vanishes when n_t is fixed and $n_r \to \infty$. Consequently in this regime, we can always choose the number of receive antennas n_r large enough so that our system can achieve a given fraction of capacity η using blocklength n. The normalized dispersion in this case is proportional to $1/\log^2(n_r)$.

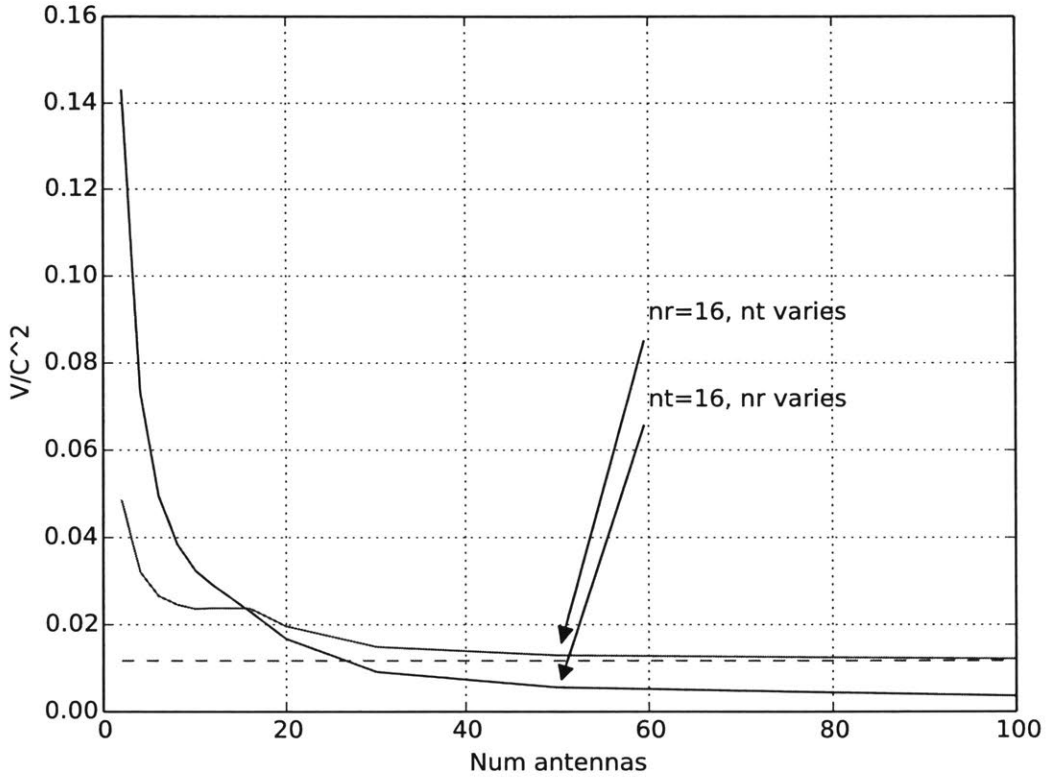

Figure 5-4: Normalized dispersion $\frac{V}{C^2}$ as a function of n_r and n_t. The transmit power is $P = 20dB$ and $T = 16$. Dashed lines are asymptotic values from (5.91)-(5.94).

Chapter 6

The Curious Case of Rank 1

In the case where the fading process has rank 1 almost surely (of which $n_r = 1$, i.e. MISO, is a special case), and interesting phenomenon occurs. As described in Theorem 1, when the fading process has rank 1, suddenly the channel has many non-trivial capacity achieving input distributions. As we saw in Chapter 3, the dispersion is given in the variational form

$$V(P) \triangleq \inf_{P_X : I(X;Y|H)=C} \frac{1}{T}\mathbb{E}\left[\mathrm{Var}(i(X;Y,H)|X)\right] . \tag{6.1}$$

where the minimization is over the set of caids. In this chapter, we'll see that in fact the above function is not constant over the set of caids, and in fact is minimized at non-trivial orthogonal design like input distributions.

6.1 Computation of the Dispersion as a Function of the Input Distribution

To understand the minimal value in (6.1), first we compute the explicit expression in terms of the input P_X. The following proposition gives the expression for the conditional variance in this case, as a function of the caid.

Theorem 21. *When* $\mathbb{P}[\mathit{rank}(H) \leq 1] = 1$, *we have*

$$V(P) = T\mathrm{Var}\left(C_{AWGN}\left(\frac{P}{n_t}\Lambda^2\right)\right) + \mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda^2\right)\right] \tag{6.2}$$

$$+ \left(\frac{P}{n_t}\right)^2 \left(\eta_1 - \frac{\eta_2}{n_t^2 T} v^*(n_t, T)\right) \tag{6.3}$$

where Λ^2 *denotes the distribution of the non-zero eigenvalues of* HH^T, *and*

$$v^*(n_t, T) = \frac{n_t^2}{2P^2} \max_{P_X : I(X;Y,H)=C} \mathrm{Var}(\|X\|_F^2) \tag{6.4}$$

Proof. As in Prop. 20 we need to evaluate the expectation of terms in (5.9)-(5.11). Any caid X should satisfy $\mathbb{E}\left[\|X\|_F^2\right] = TP$ and thus the term (5.9) is zero. The term (5.10) can be expressed in terms of $\mathrm{Var}(\|X\|_F^2)$, but the (5.11) presents a non-trivial complication due to the presence of $\|XX^T\|_F^2$, whose expectation is possible (but rather tedious) to compute by invoking properties of caids established in Theorem 1. Instead, we recall that the sum (5.10)+(5.11) equals (5.39). Evaluation of the latter can be simplified in this case due to constraint on the rank of H. Overall, we get

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Var}(i(X;Y,H)|X)\right] \\
&= T^2\mathrm{Var}\left(C_{AWGN}\left(\frac{P}{n_t}\Lambda_1^2\right)\right) + T\mathbb{E}\left[V_{AWGN}\left(\frac{P}{n_t}\Lambda_1^2\right)\right] \qquad (6.5)\\
&+ \frac{\log^2 e}{4}\mathbb{E}\left[\mathrm{Var}\left(c\left(\Lambda_1^2\right)\left(\|V_1^TX\|^2 - \frac{TP}{n_t}\right)\middle|X\right)\right], \qquad (6.6)
\end{aligned}$$

where $c(\cdot)$ is from (5.3). The last term in (6.6) can be written as

$$\begin{aligned}
&\mathbb{E}\left[c\left(\Lambda_1^2\right)^2\right]\mathbb{E}\left[\left(\|V_1^TX\|^2 - \frac{TP}{n_t}\right)^2\right] \\
&- \mathbb{E}^2\left[c\left(\Lambda_1^2\right)\right]\mathbb{E}\left[\left(\mathbb{E}[\|V_1^TX\|_F^2|X] - \frac{TP}{n_t}\right)^2\right] \qquad (6.7)
\end{aligned}$$

which follows from the identity $\mathrm{Var}(AB) = \mathbb{E}[A^2]\mathbb{E}[B^2] - \mathbb{E}^2[A]\mathbb{E}^2[B]$ for independent A, B. The second term in (6.7) is easily handled since from Lemma 2 we have $\mathbb{E}[\|V_1^TX\|_F^2|X] = \|X\|_F^2/n_t$. To compute the first term in (6.7) recall from Theorem 1 that for any fixed unit-norm v and caid X we must have $v^TX \sim \mathcal{N}(0, P/n_tI_T)$. Therefore, we have

$$\mathbb{E}\left[\left(\|V_1^TX\|^2 - \frac{TP}{n_t}\right)^2\middle|V_1\right] = \frac{2TP^2}{n_t^2}.$$

Putting everything together we get that (6.7) equals

$$\mathbb{E}\left[c\left(\Lambda_1^2\right)^2\right]2T\left(\frac{P}{n_t}\right)^2 - \mathbb{E}\left[c\left(\Lambda_1^2\right)\right]^2\frac{1}{n_t^2}\mathrm{Var}(\|X\|_F^2) \qquad (6.8)$$

concluding the proof. □

6.2 Minimizing the Conditional Variance

Theorem 21 shows that the conditional variance has the form

$$\frac{1}{T}\mathrm{Var}(i(X;Y,H)|X) = V_1(P) - \frac{\chi_2}{n_t^2T}\mathrm{Var}(\|X\|_F^2) \qquad (6.9)$$

where V_1 is independent of the CAID X and χ_2 is a non-negative constant. In this form, the dependence of the dispersion on the input distribution is explicit: in order to minimize the dispersion, we maximize $\mathrm{Var}(\|X\|_F^2)$ over the set of CAIDs. We can expand $\mathrm{Var}(\|X\|_F^2)$ as a sum of covariances which will be easier to deal with. This is captured in the following definition, then use to define V_{min} as the minimal dispersion over the set of CAIDs.

Definition 2. For the MISO channel with n_t transmit antennas and coherence time T we define

$$v^*(n_t, T) \triangleq \max_{P_X: I(X;Y,H)=C} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{T} \sum_{l=1}^{T} \rho_{ikjl}^2 \tag{6.10}$$

where

$$\rho_{ikjl}^2 = \frac{\mathrm{Cov}(X_{ik}^2, X_{jl}^2)}{\mathrm{Var}(X_{11}^2)} \tag{6.11}$$

The notation ρ_{ikjl} is appropriate since whenever X is jointly Gaussian, ρ_{ikjl}^2 is the squared correlation coefficient between X_{ik} and X_{jl}. However, there are non-jointly Gaussian CAIDs where this isn't the case. For instance, when $n_t = T = 2$, for $v \sim \mathrm{Ber}(1/2)$ and A, B i.i.d. $\mathcal{N}(0, P/2)$, the following achieves capacity

$$X = \begin{bmatrix} A & -(-1)^v B \\ B & (-1)^v A \end{bmatrix} \tag{6.12}$$

Here, the correlation coefficient between X_{11} and X_{22} is 0, however (6.11) gives $\rho_{1122}^2 = 1$.

Now we define V_{min}, the minimal dispersion, is given in terms of $v^*(n_t, T)$.

Proposition 22. *The minimal dispersion of an $n_t \times T$ block-fading MISO channel is given by*

$$V_{min} \triangleq \inf_{X\text{-}caid} \frac{1}{T} \mathrm{Var}[i(X;Y,H)|X] = V_1(P) - \frac{2\chi_2 P^2}{n_t^4 T} v^*(n_t, T) \tag{6.13}$$

where V_1 and χ_2 are from (6.9).

Proof. The only term that depends on X in (6.9) is $\mathrm{Var}(\|X\|_F^2)$. We can expand this as a sum of covariance terms:

$$\mathrm{Var}(\|X\|_F^2) = \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{T} \sum_{l=1}^{T} \mathrm{Cov}(X_{ik}^2, X_{jl}^2) \tag{6.14}$$

Writing this in terms of the ρ^2_{ikjl}, and using $\mathrm{Var}(X^2_{ik}) = 2(P/n_t)^2$, (6.11) yields

$$\mathrm{Var}(\|X\|_F^2) = 2\left(\frac{P}{n_t}\right)^2 \sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{T}\sum_{l=1}^{T}\rho^2_{ikjl} \tag{6.15}$$

Maximizing this term over the set of CAIDs gives $2(P/n_t)^2 v^*(n_t, T)$, and plugging this value of $\mathrm{Var}(\|X\|_F^2)$ into the expression for the conditional variance from Proposition 21 gives V_{min} above. □

Intuitively, we see that minimizing dispersion is equivalent to *maximizing* the amount of correlation amongst the entries of X when X is jointly Gaussian. In a sense, this asks for the capacity achieving input distribution having the least amount of randomness.

Next we must characterize $v^*(n_t, T)$. The manifold of CAIDs is not a particularly nice manifold to optimize over, one must account for all the independence constraints on the rows and columns, the covariance constraints on the 2×2 minors, positive definite constraints, etc. Our strategy instead will be to given an upper bound on $v^*(n_t, T)$, then show that for most of the pairs (n_t, T), the upper bound is tight. The crux of the upper bound is the following simple lemma.

Lemma 23. *Let $(A_1, \ldots, A_n)$ and $(B_1, \ldots, B_n)$ be i.i.d. random vectors that may have arbitrary correlation between them, then*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(A_i, B_j) \leq n\sigma^2 \tag{6.16}$$

With equality iff $\sum_{i=1}^n A_i = \sum_{i=1}^n B_i$ almost surely.

Proof. Simply use the fact that covariance is a bilinear function, and apply the Cauchy-Schwarz inequality as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(A_i, B_j) = \mathrm{Cov}\left(\sum_{i=1}^{n} A_i, \sum_{j=1}^{n} B_j\right) \tag{6.17}$$

$$\leq \sqrt{\mathrm{Var}\left(\sum_{i=1}^{n} A_i\right)\mathrm{Var}\left(\sum_{j=1}^{n} B_j\right)} \tag{6.18}$$

$$= \sqrt{(n\mathrm{Var}(A_1))(n\mathrm{Var}(B_1))} \tag{6.19}$$

$$= n\sigma^2 \tag{6.20}$$

We have equality in Cauchy-Schwarz when $\sum_{i=1}^n A_i$ and $\sum_{i=1}^n B_i$ are propositional, and since these sums have the same distribution, the constant of proportionality must be 1, so we have equality in (6.36) iff $\sum_{i=1}^n A_i = \sum_{i=1}^n B_i$ almost surely. □

We will use this lemma shortly to upper bound $\mathrm{Var}(\|X\|_F^2)$. But before stating the main theorem of the section, we review orthogonal designs.

6.3 Orthogonal Designs

6.3.1 Historical Introduction

In this section we will give some relevant background on orthogonal designs, since they will play a large role in this work. For some historical background, Hurwitz was interested the existence of a "composition formula" for positive integers (r, s, n) [22]

$$(x_1^2 + \cdots + x_r^2)(y_1^2 + \cdots + y_s^2) = (z_1^2 + \cdots + z_n^2) \tag{6.21}$$

where the x_i's and y_i's are indeterminantes, and each z_i is a bilinear form in the x_i's and y_i's. For example, such a $(2, 2, 2)$ composition formula is

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2 \tag{6.22}$$

Let $x = (x_1, \ldots, x_r), y = (y_1, \ldots, y_s)$, and $z = (z_1, \ldots, z_n)$, then the condition that z_i is a bilinear form of the x_i's and y_i's means that z can be written as $z = Ay$ where the entries of A $(n \times s)$ are linear combinations of x_i's. In this notation, (6.21) can be restated as

$$z^T z = y^T A^T A y = x^T x y^T y \tag{6.23}$$

Which must hold for all indeterminants in y, so the existence condition reduces to the existence of an $n \times s$ matrix A with entries that are linear combinations of the x_i's such that

$$A^T A = \sum_{i=1}^{r} x_i^2 I_s \tag{6.24}$$

which yields (6.21), since

$$\sum_{i=1}^{n} z_i^2 = \left(\sum_{i=1}^{r} x_i^2\right)\left(\sum_{i=1}^{s} y_i^2\right) \tag{6.25}$$

as desired. So the problem of constructing $n \times s$ matrices A with entries given by $x_1, \ldots, x_r$ satisfying (6.24) is equivalent to finding composition formulas of the form (6.21).

6.3.2 Hurwitz-Radon Families

Definition 3 (Orthogonal Design). A real $n \times n$ orthogonal design of size k is defined to be an $n \times n$ matrix A with entries given by linear forms in $x_1, \ldots, x_k$ and coefficients in $\mathbb{R}$ satisfying

$$A^T A = \left(\sum_{i=1}^{k} x_i^2\right) I_n \tag{6.26}$$

In other words, all columns of A have squared Euclidean norm $\sum_{i=1}^{k} x_i^2$, and all columns are pairwise orthogonal. Orthogonal designs may be represented as the sum $A = \sum_{i=1}^{k} x_i V_i$ where $\{V_1, \ldots, V_k\}$ is a collection of $n \times n$ real matrices satisfying Hurwitz-Radon conditions:

$$\begin{aligned} V_i^T V_i &= I_n, \quad i = 1, \ldots, k \\ V_i^T V_j + V_j^T V_i &= 0 \quad i \neq j \end{aligned} \tag{6.27}$$

The main theorem on Hurwitz-Radon families gives the largest k such that a family satisfying the above conditions exists, as stated in the following theorem from [23,24].

Theorem 24 (Radon-Hurwitz). *There exists a family of $n \times n$ real matrices $\{V_1, \ldots, V_k\}$ satisfying (6.27) iff $k \leq \rho(n)$, where*

$$\rho(2^a b) = 8 \left\lfloor \frac{a}{4} \right\rfloor + 2^{a \bmod 4}, \qquad a, b \in \mathbb{Z},\ b\text{–odd}. \tag{6.28}$$

In particular, $\rho(n) \leq n$ and $\rho(n) = n$ only for $n = 1, 2, 4, 8$.

So the maximal size of a $n \times n$ orthogonal design is the *Hurwitz-Radon number* $\rho(n)$. For a concrete example, note that Alamouti's scheme is created from a Hurwitz-Radon family for $n = k = 2$. Indeed, take the matrices

$$V_1 = I_2, \quad V_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

then Alamouti's orthogonal design can be formed by taking $aV_1 + bV_2$. It turns out that "maximal" Hurwitz-Radon families give capacity achieving input distributions for the MIMO-BF channel, see Proposition 25 for the details.

The definition of orthogonal designs can be generalized to rectangular matrices [25], as follows:

Definition 4 (Generalized Orthogonal Design). A generalized orthogonal design is a $p \times n$ matrix A with $p \geq n$ with entries as linear forms of the indeterminates $\{x_1, \ldots, x_k\}$ satisfying (6.26).

The quantity $R = k/p$ is often called the rate of the generalized orthogonal design. This term is justify by noticing that if p represents a number channel uses and k represents the number of data symbols, then R represents sending k data symbols in p channel uses. In this work, we are only interested in the case $R = 1$ (i.e. $k = p$), called *full-rate orthogonal designs*. Full-rate orthogonal design can be constructed from a Hurwitz-Radon family $\{V_1, \ldots, V_n\}$, each $V_i \in \mathbb{R}^{k \times k}$ by forming the matrix A

$$A = [V_1 x \ \cdots \ V_n x] \tag{6.29}$$

where $x = [x_1, \ldots, x_k]^T$ is the vector of indeterminates. It follows immediately from this construction that (6.26) is satisfied. Theorem 24 allows us to conclude that a generalized full rate $n \times k$ orthogonal design exists if and only if $n \leq \rho(k)$.

The following proposition shows that full rate orthogonal designs correspond to caids in the MIMO-BF channel.

Proposition 25. *Take $n_t = \rho(T)$ and a maximal Hurwitz-Radon family $\{V_i, i = 1, \ldots, n_t\}$ of $T \times T$ matrices (cf. Theorem 24). Let $\xi \sim \mathcal{N}(0, P/n_t I_T)$ be an i.i.d. row-vector. Then the input distribution*

$$X = \left[V_1^T \xi^T \ \cdots V_{n_t}^T \xi^T\right]^T \tag{6.30}$$

achieves capacity for any MIMO-BF channel provided $\mathbb{P}[\operatorname{rank} H \leq 1] = 1$.

Proof. Since $\{V_1, \ldots, V_{n_t}\}$ is a Hurwitz-Radon family, they satisfy (6.27). Form X as in (6.30). Then each row and column is jointly Gaussian, and applying the caid conditions (2.17) and (2.18) from Theorem 1 shows,

$$\mathbb{E}[R_i^T R_i] = V_i^T \mathbb{E}[\xi^T \xi] V_i = \frac{P}{n_t} V_i^T V_i = \frac{P}{n_t} I_T \tag{6.31}$$

$$\begin{aligned} \mathbb{E}[R_i^T R_j] &= V_i^T \mathbb{E}[\xi^T \xi] V_j = \frac{P}{n_t} V_i^T V_j = -\frac{P}{n_t} V_j^T V_i \\ &= -\mathbb{E}[R_j^T R_i] \end{aligned} \tag{6.32}$$

Therefore X satisfies the caid conditions, and hence achieves capacity. □

Remark 9. The above argument implies that if $X \in \mathbb{R}^{n_t \times T}$ is constructed above, then removing the last row of X gives an $(n_t - 1) \times T$ input distribution that also achieves capacity.

6.4 Main Theorem

The main theorem of this section is the following, which summarizes our current knowledge of $v^*(n_t, T)$.

Theorem 26. *For any pair of positive integers n_t, T we have*

$$v^*(T, n_t) = v^*(n_t, T) \leq n_t T \min(n_t, T)\,. \tag{6.33}$$

If $n_t \leq \rho(T)$ or $T \leq \rho(n_t)$ then a full-rate orthogonal design is dispersion-optimal and

$$v^*(n_t, T) = n_t T \min(n_t, T)\,. \tag{6.34}$$

If instead $n_t > \rho(T)$ and $T > \rho(n_t)$ then for a jointly-Gaussian capacity-achieving input X we have[1]

$$\frac{n_t^2}{2P^2} \operatorname{Var}(\|X\|_F^2) < n_t T \min(n_t, T)\,. \tag{6.35}$$

[1]So that in these cases the bound (6.33) is either non-tight, or is achieved by a non-jointly-Gaussian caid.

Finally, if $n_t \leq T$ *and* (6.34) *holds, then* $v^*(n_t', T) = n_t'^2 T$ *for any* $n_t' \leq n_t$ *(and similarly with the roles of* n_t *and* T *switched).*

Theorem 26 states that for dimensions where orthogonal designs exist, the conditional variance (3.6) is minimized if and only if the input is constructed from an orthogonal design as in Proposition 25. The approach is first to prove an upper bound on v^*, then show that conditions for tightness of the upper bound correspond to conditions of the Hurwitz-Radon theorem.

Note that the $\rho(n)$ function is monotonic in even values of n (and is 1 for n odd), and $\rho(n) \to \infty$ along even n. Therefore, for any number of transmit antennas n_t, there is a large enough T such that $n_t \leq \rho(T)$, in which case an $n_t \times T$ full rate orthogonal design achieves the optimal $v^*(n_t, T)$.

We start with a simple lemma, which will be applied with A, B equal to the rows of the capacity achieving input X.

Lemma 27. *Let* $A = (A_1, \ldots, A_n)$ *and* $B = (B_1, \ldots, B_n)$ *each be i.i.d. random vectors from the same distribution with finite second moment* $\mathbb{E}[A_1^2] = \sigma^2 < \infty$. *While* A *and* B *are i.i.d. individually, they may have arbitrary correlation between them. Then*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(A_i, B_j) \leq n\sigma^2 \tag{6.36}$$

with equality iff $\sum_{i=1}^n A_i = \sum_{i=1}^n B_i$ *almost surely.*

Proof. Simply use the fact that covariance is a bilinear function, and apply the Cauchy-Schwarz inequality as follows:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(A_i, B_j) = \mathrm{Cov}\left(\sum_{i=1}^{n} A_i, \sum_{j=1}^{n} B_j\right) \tag{6.37}$$

$$\leq \sqrt{\mathrm{Var}\left(\sum_{i=1}^{n} A_i\right) \mathrm{Var}\left(\sum_{j=1}^{n} B_j\right)} \tag{6.38}$$

$$= \sqrt{(n\mathrm{Var}(A_1))(n\mathrm{Var}(B_1))} \tag{6.39}$$

$$= n\sigma^2 \tag{6.40}$$

We have equality in Cauchy-Schwarz when $\sum_{i=1}^n A_i$ and $\sum_{i=1}^n B_i$ are proportional, and since these sums have the same distribution, the constant of proportionality must be equal to 1, so we have equality in (6.36) iff $\sum_{i=1}^n A_i = \sum_{i=1}^n B_i$ almost surely. □

Proof of Theorem 26. First, we rewrite $v^*(n_t, T)$ defined in (6.4) as

$$v^*(n_t, T) \triangleq$$
$$\frac{n_t^2}{2P^2} \max_{P_X: I(X;Y|H)=C} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \sum_{k=1}^{T} \sum_{l=1}^{T} \mathrm{Cov}(X_{i,k}^2, X_{j,l}^2) \tag{6.41}$$

From here, $v^*(n_t, T) = v^*(T, n_t)$ follows from the symmetry to transposition of the caid-conditions on X (see Theorem 1) and symmetry to transposition of (6.41). From now on, without loss of generality we assume $n_t \leq T$.

For the upper bound, since the rows and columns of X are i.i.d., we can apply Lemma 27 with $A_i = X_{i,k}^2$ and $B_j = X_{j,l}^2$ (and hence $\sigma^2 = 2(P/n_t)^2$) to get

$$\sum_{i,j,k,l} \mathrm{Cov}(X_{i,k}^2, X_{j,l}^2) \leq \sum_{i,j} 2T(P/n_t)^2 = 2n_t^2 T(P/n_t)^2 \,, \tag{6.42}$$

which together with (6.41) yields the upper bound (6.33) (recall that $n_t \leq T$).

Equation (6.42) implies that if X achieves the bound (6.33), then removing the last row of X achieves (6.33) as an $(n_t - 1) \times T$ design. In other words, if (6.33) is tight for $n_t \times T$ then it is tight for all $n_t' \leq n_t$.

Notice that for any X such that any pair $X_{i,k}$,$X_{j,l}$ is jointly Gaussian, we have

$$\frac{n_t^2}{2P^2} \mathrm{Var}(\|X\|_F^2) = \sum_{i,j,k,l} \rho_{ikjl}^2 \,, \tag{6.43}$$

where

$$\rho_{ikjl} \triangleq \frac{n_t}{P} \mathrm{Cov}(X_{ik}, X_{jl}) \,. \tag{6.44}$$

Take $X \in \mathbb{R}^{n_t \times T}$ as constructed in (6.30). By Proposition 25, X is capacity achieving and identity (6.43) clearly holds. In the representation (6.30), the matrix $V_j^T V_i$ contains the correlation coefficients between rows i and j of X, since $\mathbb{E}[(\xi V_j)^T (\xi V_i)] = \frac{P}{n_t} V_j^T V_i$, so

$$\|V_j^T V_i\|_F^2 = \sum_{k=1}^{T} \sum_{l=1}^{T} \rho_{ikjl}^2 \,. \tag{6.45}$$

Therefore we can represent the sum of squared correlation coefficients as

$$\sum_{i,j,k,l} \rho_{ijkl}^2 = \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \|V_j^T V_i\|_F^2 \tag{6.46}$$

$$= \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \mathrm{tr}\left(V_j V_j^T V_i V_i^T\right) \tag{6.47}$$

$$= \mathrm{tr}\left(\left(\sum_{i=1}^{n_t} V_i V_i^T\right)^2\right) \tag{6.48}$$

$$= n_t^2 . T \tag{6.49}$$

Line (6.49) follows since the V_i's are orthogonal by the Hurwitz-Radon condition, so each $V_i V_i^T = I_T$ in the summation in (6.48). Hence the X constructed in (6.30) achieves the upper bound in (6.42) and (6.33).

Next we prove (6.35). Suppose X is a jointly-Gaussian caid saturating the bound (6.42). From Lemma 27, the condition for equality in (6.36) implies that for all $j \in \{1, \ldots, n_t\}$,

$$\|R_j\|_F^2 = \|R_1\|_F^2 \quad a.s. \tag{6.50}$$

where R_j is the j-th row of X for $j = 1, \ldots, n_t$. In particular, this means that every R_j is a linear function of R_1. Consequently, we may represent X in terms of a row-vector $\xi \sim \mathcal{N}(0, P/n_t I)$ as in (6.30), that is $R_j = \xi V_j$ for some $T \times T$ matrices $V_j, j \in [n_t]$. We clearly have

$$\mathbb{E}\left[R_i^T R_j\right] = \frac{P}{n_t} V_i^T V_j .$$

But then the caid constraints (2.17)-(2.18) imply that the matrix A in (6.29) constructed using indeterminates $\{x_1, \ldots, x_{n_t}\}$ and family $\{V_1, \ldots, V_{n_T}\}$ satisfies Definition 4. Therefore, from Theorem 24, (see also [26, Proposition 4]), we must have $n_T \leq \rho(T)$. □

Remark 10. In the case $n_t = T = 2$ it is easy to show that for any non-jointly-Gaussian caid, there exists a jointly-Gaussian caid achieving the same $\mathrm{Var}(\|X\|_F^2)$. Indeed, consider (2.23) with $\rho = \frac{\mathrm{cov}(X_{1,1}^2, X_{2,2}^2) + \mathrm{cov}(X_{1,2}^2, X_{2,1}^2)}{8(P/n_t)^2}$. If this phenomena held in general, we would conclude that (6.34) holds if and only if $n_t \leq \rho(T)$ or $T \leq \rho(n_t)$. As a step towards the proof of the latter, we notice that any caid X achieving equality in (6.42) satisfies

$$XX^T = \frac{\|X\|_F^2}{n_t} I_{n_t} \qquad \text{(a.s.)}, \tag{6.51}$$

which is equivalent to saying $R_i R_j' = 0$ for $i \neq j$. The latter follows from applying (6.50) to rows of UX, where U is an arbitrary orthogonal matrix. Identity (6.51) could be informally stated as "any caid saturating (6.42) is a random full-rate orthogonal design".

In summary, the full-rate orthogonal designs (when those exist) achieve the opti-

mal channel dispersion $V(P)$. Some examples (ξ_j are i.i.d. $\mathcal{N}(0,1)$) for $n_t = T = 4$ and $n_t = 4, T = 3$, respectively, are as follows:

$$X = \sqrt{\frac{P}{4}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \xi_4 \\ -\xi_2 & \xi_1 & -\xi_4 & \xi_3 \\ -\xi_3 & \xi_4 & \xi_1 & -\xi_2 \\ -\xi_4 & -\xi_3 & \xi_2 & \xi_1 \end{bmatrix} \tag{6.52}$$

$$X = \sqrt{\frac{P}{4}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & -\xi_4 \\ -\xi_3 & \xi_4 & \xi_1 \\ -\xi_4 & -\xi_3 & \xi_2 \end{bmatrix}$$

6.5 When Full-Rate Orthogonal Designs do not Exist

For pairs (n_t, T) where $n_t > \rho(T)$, full-rate orthogonal design do not exist. For example $\rho(3) = 1$, so no full-rate orthogonal design exits for $n_t = 2$, $T = 3$. Which caids are minimizer for (3.6) in this case? In general, we do not know the answer and do not even know whether one can restrict the search to jointly-Gaussian caids. But one thing is certain: it is definitely not an i.i.d. Gaussian (Telatar) caid. To show this claim, we will give a method for constructing improved caids.

To that end, suppose that X consists of entries $\pm\xi_j$, $j = 1\ldots, d$, where $\xi_j \overset{i.i.d.}{\sim} \mathcal{N}(0, P/n_t)$. Then we have:

$$\frac{n_t^2}{2P}\mathrm{Var}(\|X\|_F^2) = \sum_{t=1}^{d}(\ell_t)^2\,, \tag{6.53}$$

where ℓ_t is the number of times $\pm\xi_t$ appears in the description of X. By this observation and the remark after Theorem 1 (any submatrix of a caid X is also a caid), we can obtain lower bounds on $v^*(n_t, T)$ for $n_t > \rho(T)$ via the following *truncation construction*:

1. Take $T' > T$ such that $\rho(T') \geq n_t$ and let X' be a corresponding $\rho(T') \times T'$ full-rate orthogonal design with entries $\pm\xi_1, \ldots \pm \xi_{T'}$.

2. Choose an $n_t \times T$ submatrix of X' maximizing the sum of squares of the number of occurrences of each of ξ_j, cf. (6.53).

As an example of this method, by truncating a 4×4 design (6.52) we obtain the following 2×3 and 3×3 submatrices:

$$X = \sqrt{\frac{P}{3}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & \xi_4 \\ -\xi_3 & -\xi_4 & \xi_1 \end{bmatrix} \quad X = \sqrt{\frac{P}{2}} \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 \\ -\xi_2 & \xi_1 & \xi_4 \end{bmatrix} \tag{6.54}$$

Table 6.1: Values for $v^*(n_t, T)$

$n_t \setminus T$	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	8
2		8	10*	16	18	24	26	32
3			21*	36	[39,45]	[46,54]	[57,63]	72
4				64	[68,80]	[80,96]	[100,112]	128
5					[89,125]	[118,150]	[155,175]	200
6						[168,216]	[222,252]	288
7							[301,343]	392
8								512

Note: Table is symmetric about diagonal; intervals $[a, b]$ mark entries for which dispersion-optimal input is unknown. The optimality of entries marked with $*$ is only established in the class of all jointly-Gaussian caids.

By independent methods we were able to show that designs (6.54) are dispersion-optimal out of all jointly Gaussian caids. Note that in these cases (6.34) does not hold, illustrating (6.35).

Our current knowledge about v^* is summarized in Table 6.1. The lower bounds for cases not handled by Theorem 26 were computed by truncating the 8x8 orthogonal design [25, (5)]. Based on the evidence from $2 \times T$ and 3×3 we *conjecture* this construction to be optimal.

From the proof of Theorem 26 it is clear that Telatar's i.i.d. Gaussian is never dispersion optimal, unless $n_t = 1$ or $T = 1$. Indeed, for Telatar's input $\rho_{ikjl} = 0$ unless $(i, k) = (j, l)$. Thus embedding even a single 2×2 Alamouti block into an otherwise i.i.d. $n_t \times T$ matrix X strictly improves the sum (6.41). We note that the value of $\frac{V}{C^2}$ entering (5.85) can be quite sensitive to the suboptimal choice of the design. For example, for $n_t = T = 8$ and $SNR = 20\ dB$ estimate (5.85) shows that one needs

- around 600 channel inputs (that is 600/8 blocks) for the optimal 8×8 orthogonal design, or
- around 850 channel inputs for Telatar's i.i.d. Gaussian design

in order to achieve 90% of capacity. This translates into a 40% longer delay or battery spent in running the decoder.

Thus, curiously even in cases where pure multiplexing (that is maximizing transmission rate) is needed – as is often the case in modern cellular networks – transmit diversity enters the picture by enhancing the finite blocklength fundamental limits. Remember, however, that our discussion pertains only to cases when the transmitter (base-station) is equipped with more antennas than the receiver (user equipment), or when the channel does not have more than one diversity branch.

In cases when full-rate designs do not exist, there have been various suggestions as to what could be the best solution, e.g. [26]. Thus for non full-rate designs the property of minimizing dispersion (such as (6.54)) could be used for selecting the best design for cases $n_t > \rho(T)$.

6.5.1 The 2x3 and 3x3

In this section, we explicitly compute the minimizers in (6.1) for the cases $n_t = 2, T = 3$, and $n_t = 3, T = 3$, where full rate orthogonal designs do not exist. These are both practical regimes, i.e. it is easy to put two or three antennas on a device. While full rate orthogonal designs have been proposed base on heuristics, this section serves as a rigorous justification for space time configurations in the 2×3 and 3×3 case. We begin the with $n_t = 2, T = 3$ case.

Proposition 28. *For $n_t = 2, T = 3$, the distribution*

$$X = \sqrt{\frac{P}{2}} \begin{bmatrix} X_1 & X_2 & X_3 \\ -X_2 & X_1 & X_4 \end{bmatrix} \tag{6.55}$$

where each X_i is i.i.d. $\mathcal{N}(0,1)$ is the optimal jointly Gaussian distribution in the optimization (6.1).

Proof. We do this by brute force. Throughout, let A, B, C, D, E, F be i.i.d. $\mathcal{N}(0,1)$. We know that the first row has i.i.d. jointly Gaussian entries, hence we can write is as (A, B, C). Each column must be independent, hence our input must be in the form

$$X = \begin{bmatrix} A & B & C \\ \rho_{12}B + \rho_{13}C + \rho_1 D & \rho_{21}A + \rho_{23}C + \rho_2 E & \rho_{31}A + \rho_{32}B + \rho_3 F \end{bmatrix}. \tag{6.56}$$

Now we apply the constraints of Theorem 1 to constrain the ρ's. The cross correlation constraints gives the conditions

$$\rho_{21} = -\rho_{12} \tag{6.57}$$
$$\rho_{13} = -\rho_{31} \tag{6.58}$$
$$\rho_{23} = -\rho_{32} \tag{6.59}$$

and entries of the second row must be independent, giving the conditions

$$\rho_{12}\rho_{23} = \rho_{12}\rho_{32} = \rho_{21}\rho_{31} = 0\,. \tag{6.60}$$

We can write out the objective function in terms of the ρ's via

$$\mathrm{Var}(\|X\|_F^2) = 3 + 3 + \rho_{21}^2 + \rho_{31}^2 + \rho_{12}^2 + \rho_{32}^2 + \rho_{13}^2 + \rho_{23}^2\,. \tag{6.61}$$

where the first two terms are from the correlations of the first and second rows with themselves, respectively. The first observation is that we gain no benefit from allocating mass to ρ_1, ρ_2, ρ_3. Secondly, by the constraints in (6.57)-(6.60), at least 4 of the coefficients must be zero. I.e. 3 are forced to be 0 from (6.60), e.g. $\rho_{31}, \rho_{23}, \rho_{32} = 0$, which forces at least one of conditions (6.57)-(6.59) to introduce an additional zero (in the example, $\rho_{13} = 0$. Hence at most 2 of the ρ's can be non-zero in (6.61). Since each ρ^2 can be at most 1, we can take $\rho_{21} = -\rho_{12} = 1$, which gives the maximum value of (6.61), and gives the distribution in the Proposition statement. □

Remark 11. The distribution given in (6.55) is asymmetric, however we can make a symmetric version of this as follows: let A, B, C, D be i.i.d. $\mathcal{N}(0,1)$, and take the distribution

$$X = \sqrt{\frac{P}{3}} \begin{bmatrix} A & B & C \\ \frac{1}{\sqrt{3}}(-B-C+D) & \frac{1}{\sqrt{3}}(A-C-D) & \frac{1}{\sqrt{3}}(A+B+D) \end{bmatrix} \tag{6.62}$$

The it is easily verified that this distribution satisfies the constraints given in the proof of Proposition 28. The reason this configuration does not fall out of the theorem is that we chose to parameterize the second to row to give each entry an independent Gaussian component. This choice of parameterization is general enough to demonstrate optimality of the input (6.55), but does not contain the distribution (6.62).

As a corollary, we obtain the optimal $n_t = T = 3$ input distribution:

Corollary 29. *For $n_t = T = 3$, the following distribution achieves the maximum value in the optimization* (6.1)

$$X = \sqrt{\frac{P}{3}} \begin{bmatrix} X_1 & X_2 & X_3 \\ -X_2 & X_1 & X_4 \\ -X_3 & -X_4 & X_1 \end{bmatrix} \tag{6.63}$$

where X_1, X_2, X_3, X_4 are i.i.d. $\mathcal{N}(0,1)$.

Proof. The objective function $\mathrm{Var}(\|X\|_F^2)$ can be decomposed into an expression only concerning pairs of rows, i.e.

$$\mathrm{Var}(\|X\|_F^2) = \sum_{i=1}^{n_t}\sum_{k=1}^{n_t} \mathrm{Cov}\left(\|R_i\|^2, \|R_k\|^2\right) \tag{6.64}$$

$$= \frac{1}{2}\left(\sum_{i,k=1,2} \mathrm{Cov}\left(\|R_i\|^2, \|R_k\|^2\right) + \sum_{i,k=2,3} \mathrm{Cov}\left(\|R_i\|^2, \|R_k\|^2\right) + \sum_{i,k=1,3} \mathrm{Cov}\left(\|R_i\|^2, \|R_k\|^2\right)\right) \tag{6.65}$$

$$= \frac{1}{2}\left(\mathrm{Var}\left(\|X^{(1,2)}\|^2\right) + \mathrm{Var}\left(\|X^{(2,3)}\|^2\right) + \mathrm{Var}\left(\|X^{(1,3)}\|^2\right)\right) \tag{6.66}$$

where $X^{(i,k)} \in \mathbb{R}^{2\times 3}$ denotes the $n_t = 2, T = 3$ input distribution from taking the i and k-th rows of X. From Proposition 28, we know that the perfectly correlating two elements between rows maximizes $\mathrm{Var}(\|X\|_F^2)$ in the 2×3 case. But the input (6.63) does this for each pair of rows, and hence maximizes each term individually in (6.66), and hence must maximize $\mathrm{Var}(\|X\|_F^2)$. □

Chapter 7

Variable Length List Decoding

The second point to point coding problem we study is the problem of *variable length list decoding with stop feedback*. First we give some background on variable length coding and list decoding, then give motivation for studying the variable length list decoding problem.

The idea of using feedback in communication schemes has been around since Shannon, where he showed the feedback cannot increase channel capacity [27]. Dobrushin '62 [28] showed that for fixed blocklength codes, even the error exponent cannot increase. Burnashev '75 [29] had the idea to use variable length codes along with feedback, where he established upper and lower bounds on the average transmission time of a code with M codewords and probability of error ϵ, showing that the reliability function is given by

$$E_{opt}(R) = C_1 \left(1 - \frac{R}{C}\right) \tag{7.1}$$

where $C_1 = \max_{x,x' \in \mathcal{X}} D(P_{Y|X=x} || P_{Y|X=x'})$. Polyanskiy et al in [7] showed that using feedback with variable length codes can dramatically increase the distance from capacity using a fixed average blocklength, in the sense that the maximal number of codewords for average blocklength ℓ and probability of error ϵ is given by

$$\log M^*(\ell, \epsilon) = \frac{\ell C}{1-\epsilon} + O(1) \tag{7.2}$$

boosting the linear term to $\frac{C}{1-\epsilon}$, and eliminating both the $O(\sqrt{\ell})$ and $O(\log \ell)$ terms. Furthermore, this speed up can be obtained by using only stop feedback rather than full feedback – where the decoder sends only a single bit of feedback to indicate that the encoder should stop sending symbols.

List decoding began with Elias and Wozencraft in [30] and [31] in 1957 and 1958, respectively. Elias' motivation was to show that for large enough L, a code for the BSC chosen at random from the set of all codes preform nearly as well as the best possible code. Subsequently, list decoding became an interesting question to coding theorists – since instead of designing codes where that could tolerate at most $d/2$ bit errors, many more errors could be tolerated when the decoder only must output a

list. Works such as Blinovsky '86 [32] and Elias '91 [33] analyzed how many errors could be tolerated by codes with a fixed list size. Tan '14 [34] gave a finite blocklength result for the performance of list decoding, showing that essentially the same finite blocklength results hold when replacing M with M/L.

Variable length coding allows the system to have *optionality* concerning when to stop. For a fixed blocklength system, the number of messages must be chosen such that at time n, with high probability, the correct codeword is distinguishable for the $M-1$ other codewords. In variable length coding, the correct codeword only needs to be distinguishable in *average* time n. Hence, if the channel conditions are unlucky early on, you can stop later, or if they are good, you can stop sooner. The interesting question unique to variable length coding with stop feedback thus is: how does the system decide when to stop? I.e. in the list decoding setting, how does the system know that, with high probability, the correct codeword is amongst a set of L codewords? Notice that at such a time, we do not expect the correct codeword to be distinguishable from the other $L-1$ incorrect codewords, so there is no "popping out" behavior.

In this chapter, we first define the problem of variable length list decoding with stop feedback. We give a definition that forces the encoder to use random coding – arguing that this capture the nature of the problem, and showing that without this constraint, the problem becomes trivial. We show that for the Binary Erasure Channel (BEC), such a variable length scheme is able to kill all but the linear term and constant term in the expansion of $\log M^*$. We give an application of variable length list decoding to *stop feedback with delay*, where the stop signal is seen by the encoder only after a delay of D time steps. Finally, we argue that, surprisingly, for the Binary Symmetric Channel (BSC), variable length list decoding is not able to eliminate the square root term, when the size of the list satisfies $L = M^{1-\alpha}$ for any $\alpha \in (0,1)$.

7.1 Problem Definition

We define the variable length list decoding problem as follows: an (ℓ, M, L, ϵ) variable length list decodable code with stop feedback (VLLD) for a discrete memoryless channel (DMC) $P_{Y|X}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ consists of

1. A sequence of encoders $f_n : \{1, \ldots, M\} \to \mathcal{X}$, where the channel input at time n is given by $X_n = f_n(W)$, and the message W is the uniformly chosen from the set $\{1, \ldots, M\}$. In this work, we will require that the encoder employs *random coding*, where each $f_n(w) \sim P_X^*$ is i.i.d. from the capacity achieving input distribution.

2. A sequence of decoders $g_n : \mathcal{Y}^n \to \{1, \ldots, M\}^L$, where the decoder at time n gives the "best guess" of the most likely list of L messages. Note that g_n is set-valued.

3. A stopping time $\tau \in \{0, 1, 2, \ldots\}$ measurable on the filtration $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$

satifying the constraint

$$\mathbb{E}[\tau] \leq \ell\,, \tag{7.3}$$

at which time the decoder sends the `stop` signal to the encoder.

4. The probability of error is required to satisfy

$$\mathbb{P}\left[W \notin g_\tau(Y^\tau)\right] \leq \epsilon \tag{7.4}$$

where $Y^n = (Y_1, \ldots, Y_n)$. I.e. an error is made if at time τ, the correct message is not in the set of L messages outputted by the decoder.

With this definition, we define the fundamental quantity $M^*(L, \ell, \epsilon)$, which will be the main object of study:

$$\log M^*(\ell, L, \epsilon) = \sup\{M : \exists(\ell, M, L, \epsilon) - \text{VLLD}\}\,. \tag{7.5}$$

First we discuss the random coding assumption in the definition above.

7.1.1 The Random Coding Assumption

With regards to the random coding definition in point 1 above, notice that the problem becomes trivial if we do not impose this random coding constraint. Indeed, we show the following simple proposition:

Proposition 30. *There exists an (ℓ, M, L, ϵ) code without the random coding constraint satisfying*

$$\log \frac{M}{L} \geq \frac{\ell C}{1-\epsilon} + O(\log \ell) \tag{7.6}$$

Proof. Generate the codebook as follows: form M/L bins of L sequences each (perhaps fewer than L in the final bin). Assign bin $k \in \{1, \ldots, L\}$ a codeword with i.i.d. draws from a distribution P_X, i.e. *all sequences in bin k map to the same codeword.* Then, the encoder and decoder operate just as for an $L = 1$ variable length stop feedback code designed for M/L codewords and having probability of error smaller than ϵ. The decoder produces the bin number, from which it outputs all messages in that bin as the list. As an immediate corollary of Theorem 2 in [7], this scheme achieves

$$\log \frac{M}{L} \geq \frac{\ell C}{1-\epsilon} + O(\log \ell) \tag{7.7}$$

which demonstrates the claim. □

The appealing property of list decoding is that choosing the list size allow us to trade off uncertainty about the correct message for faster stopping. I.e. we can say "suppose we are willing to accept that our message is only one of L, how much earlier

does this relaxation allow us to stop?". For the random coding scheme in Proposition 30, even if we decide to transmit for infinitely many time slots, we would never be able to choose the correct message with high probability, since buckets of L messages are associated with a single codeword. For this reason, we focus on the case when each symbol in the codebook is generated i.i.d. from the capacity achieving input distribution. We conjecture that, instead of this random coding assumption, it is sufficient to require that a coding scheme has good performance for both the unique decoding case and for lists of size $L = M^{1-\epsilon}$.

Conjecture 31. *There does not exist a code for the BSC simulaneously achieving* (7.7) *for unique decoding* ($L = 1$) *and achieving*

$$\log \frac{M}{L} \geq \frac{\mathbb{E}[\tau_L]C}{1-\epsilon} + O(\log \ell) \tag{7.8}$$

for $L = M^{1-\alpha}$ *for any* $\alpha \in (0,1)$, *and* $\mathbb{E}[\tau_L] = \ell - \frac{1}{C}\log L$.

7.2 As a Variable Length Delayed Feedback Scheme

Suppose instead we ask a similar question: given a variable length code with $L = 1$, suppose that when the decoder sends `stop`, the transmitter sees the stop signal only after a delay of D time steps. I.e. the problem set up is as follows: an (ℓ, D, M, ϵ) variable length stop feedback with Delay (VLSF-D) code is defined to be

1. A sequence of encoders $f_n : \{1, \ldots, M\} \to \mathcal{X}$, where the channel input at time n is given by $X_n = f_n(W)$, where W is the uniformly chosen message from the set $\{1, \ldots, M\}$.

2. A sequence of decoders $g_n : \mathcal{Y}^n \to \{1, \ldots, M\}$.

3. A stopping time $\tau \in \{0, 1, 2, \ldots\}$ measurable on the filtration $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ satisfying the constraint $\mathbb{E}[\tau] \leq \ell$, at which time the decoder sends the `stop` signal to the encoder.

4. The probability of error is required to satisfy

$$\mathbb{P}\left[W \neq g_{\tau+D}(Y^{\tau+D})\right] \leq \epsilon \tag{7.9}$$

   where $Y^n = (Y_1, \ldots, Y_n)$. I.e. an error is made if at time $\tau + D$, the encoder outputs the incorrect codeword.

Such a scheme has a clear practical use: feedback is never immediately available at the transmitter, and often is sent on a slower clock than the transmitter is operating at. Hence, most stop feedback schemes will fall into this definition. The question becomes: how much more information can we squeeze out of the system, given the knowledge that the decoder will receive D extra symbols after saying `stop`?

The following proposition shows that the existence of a good VLLD code implies the existence of a good VLSF-D code. Notice that the i.i.d. random codebook is an essential part of the proof.

Proposition 32. *Suppose $P_{Y|X}$ is a channel such that there exists a (ℓ, M, L, ϵ)-VLLD code satisfying*

$$\log \frac{M}{L} \geq \ell C + O(1)\,. \tag{7.10}$$

Then there exists a (ℓ, D, M, ϵ')-VLSF-D code for $\epsilon' > \epsilon$ satisfying

$$\log M \geq (\ell + D)C - \sqrt{DV}Q^{-1}(\epsilon_0) + O(\log D) \tag{7.11}$$

where

$$\epsilon_0 = \frac{\epsilon' - \epsilon}{1 - \epsilon}\,. \tag{7.12}$$

Proof. We construct a VLSF-D code from the VLLD code as follows: the decoder uses the VLLD code to say `stop` when the correct message is in a list of size L with probability at least $1 - \epsilon$, where

$$\log L = DC - \sqrt{DV}Q^{-1}(\epsilon_0) + O(\log D) \tag{7.13}$$

and ϵ_0 is chosen as in (7.12), so that the overall error of the code is at most ϵ'. The VLLD code uses random coding according to the capacity achieving input distribution of the channel, so that the D additional symbols can be viewed as a fixed blocklength code to differentiate L messages in D channel uses, i.e. *the encoder does not need to change its mode of operation.* The decoder decodes this fixed blocklength code, then outputs the message it chooses. Hence we have created a (ℓ, D, M, ϵ')-VLSF-D code, with

$$\log M \geq \log L + \ell C + O(1) \tag{7.14}$$

$$= (\ell + D)\, C - \sqrt{DV}Q^{-1}(\epsilon_0) + O(\log D) \tag{7.15}$$

as claimed. □

In the following sections, we will show that indeed the BEC satisfies the condition (7.10), so this definition is not vacuous. However, we will see that the BSC does not satisfy this condition.

7.3 Posterior Stopping Rule

In this section, we discuss the "ideal" stopping rule: stop when the sum of the largest posterior values exceeds a threshold. Like many situations, it is more difficult to analyze posterior distributions than to (suboptimally) analyze log likelihood ratios.

However, we will see that we can directly analyze this posterior-based stopping rule for the case $L = 1$ using a simple martingale argument, improving the bound and simplifying the achievability scheme given by Polyanskiy et al in [7].

The appealing property of the "stop when the sum of the largest L posteriors exceeds $1 - \epsilon$" stopping rule is that it automatically guarantees that the correct codeword is in the list with at least probability $1 - \epsilon$. This is shown in the following lemma.

Lemma 33. *Stopping at time*

$$\tau = \inf \left\{ t \geq 0 : \max_{S \subset \{1,\ldots,M\}, |S|=L} \sum_{j \in S} P_{W|Y^t}(j|y^t) \geq 1 - \epsilon \right\} \tag{7.16}$$

and outputting

$$W_L = \operatorname*{argmax}_{S \subset \{1,\ldots,M\}, |S|=L} \sum_{j \in S} P_{W|Y^t}(j|y^t) \tag{7.17}$$

at time τ guarantees that $P_e \leq \epsilon$.

Proof. Notice that the probability of error can be expresses in terms of the sum of the largest L posterior distributions:

$$\begin{aligned} P_e &= \mathbb{P}\left[W \notin g_\tau(Y^\tau)\right] && (7.18)\\ &= \mathbb{E}\left[\mathbb{1}_{\{W \notin g_\tau(Y^\tau)\}}\right] && (7.19)\\ &= \mathbb{E}\left[\sum_{j=1}^{M} \mathbb{1}_{\{j \notin g_\tau(Y^\tau)\}} P_{W|Y^\tau}(j|Y^\tau)\right] && (7.20)\\ &= \mathbb{E}\left[\sum_{j \notin W_L} P_{W|Y^\tau}(j|Y^\tau)\right] && (7.21)\\ &\leq \epsilon && (7.22) \end{aligned}$$

where the last line follows directly from the definition of τ. □

It will often be useful to apply Bayes rule to represent the posteriors $P_{W|Y^t}$ differently. To this end, define the quantity

$$S_t^j(y^t) = \log \frac{P_{Y^t|W}(y^t|j)}{Q_{Y^t}(y^t)} \tag{7.23}$$

where $Q_{Y^t} = \prod_{i=1}^{t} Q_Y(y_i)$ is a product distribution, and each Q_Y is a distribution on $\mathcal{Y}$. In this work, Q_Y will often be the capacity achieving output distribution. For a

discrete memoryless channel, (7.23) can be written as a sum of information densities:

$$S_t^j = \sum_{k=1}^{t} i(c_k^j, y_k) \tag{7.24}$$

where $c_i^j = f_i(j)$ is the i-th element of the j-th codeword, and

$$i(x,y) = \log \frac{P_{Y|X}(y|x)}{P_Y(y)} \tag{7.25}$$

is the information density. With this, we can write the posterior as

$$P_{W|Y^t}(j|y^t) = \frac{P_{Y^t|W}(y^t|j)\frac{1}{M}}{\sum_{k=1}^{M} P_{Y^t|W}(y^t|k)\frac{1}{M}} \tag{7.26}$$

$$= \frac{e^{S_t^j}}{\sum_{k=1}^{M} e^{S_t^k}}. \tag{7.27}$$

With this representation, we can write the stopping time (7.16) as

$$\tau = \inf\left\{ t \geq 0 : \max_{S\subset\{1,\ldots,M\},|S|=L} \frac{\sum_{j\in S} e^{S_t^j}}{\sum_{k=1}^{M} e^{S_t^k}} \geq 1-\epsilon \right\}. \tag{7.28}$$

Before proceeding to variable length results, we show a simple argument for the $L = 1$ case using the stopping rule (7.16), that gives a small improvement of $\log \frac{1}{\epsilon} \mapsto \log \frac{1-\epsilon}{\epsilon}$ on the best known bound of Polyanskiy et al [7]. The main contribution is to notice that the martingale M_t (defined below) can make the analysis very simple.

Proposition 34. *There exists an (M, ℓ, ϵ) stop feedback code satisfying*

$$\log M \geq \ell C - \log \frac{1-\epsilon}{\epsilon} - A \tag{7.29}$$

where $A = \max_{x,y} |i(x,y)|$.

Proof. Define the stopping time τ as in (7.16) for $L = 1$, i.e.

$$\tau = \inf\left\{ t : \max_{j\in\{1,\ldots,M\}} P_{W|Y^t}(j|Y^t) \geq 1-\epsilon \right\}. \tag{7.30}$$

When τ occurs, the decoder sends `stop` and outputs $g(Y^\tau) = \operatorname{argmax}_j P_{W|Y^\tau}(j|Y^\tau)$. With this, the probabability of error is bounded by ϵ, as demonstrated in Lemma 33. Now we bound the expected time to stop using this rule. To this end, define τ_1 as

the first time that the posterior for codeword 1 exceeds $1-\epsilon$,

$$\tau_1 = \inf\left\{t : P_{W|Y^t}(1|Y^t) \geq 1-\epsilon\right\} \tag{7.31}$$

$$= \inf\left\{t : Z_t \geq \log\frac{1-\epsilon}{\epsilon}\right\}. \tag{7.32}$$

where

$$Z_t \triangleq \log\frac{e^{S_t^1}}{\sum_{j=2}^{M} e^{S_t^j}} \tag{7.33}$$

and $S_t^j = i(X_j^t; Y^t)$ is the information density for the j-th codeword. Clearly $\mathbb{E}[\tau] \leq \mathbb{E}[\tau_1]$, since the first time that largest posterior exceed $1-\epsilon$ is surely no later than the first time the posterior for the correct codeword exceeds $1-\epsilon$. Lemma 35 below shows that $\mathbb{E}[\tau_1] < \infty$, and Lemma 36 shows that $M_t \triangleq Z_t - tC$ is a submartingale. Hence, by the Optional Stopping Theorem (which we can apply since the increments are bounded: $|Z_t - Z_{t-1}| \leq A$), we have

$$\mathbb{E}[M_0] = -\log(M-1) \tag{7.34}$$

$$= \mathbb{E}[M_\tau] \tag{7.35}$$

$$= \mathbb{E}[Z_\tau] - C\mathbb{E}[\tau] \tag{7.36}$$

so that

$$C\mathbb{E}[\tau] = \log(M-1) + \mathbb{E}[Z_\tau] \tag{7.37}$$

$$\leq \log(M-1) + \log\frac{1-\epsilon}{\epsilon} + A \tag{7.38}$$

since at time τ, we know that Z_t has exceeded $\log\frac{1-\epsilon}{\epsilon}$ for the first time, so it cannot be more than A above this quantity. Hence, choosing the number of codewords to be

$$\log M = \ell C - \log\frac{1-\epsilon}{\epsilon} - A \tag{7.39}$$

guarantees that $\mathbb{E}[\tau] < \ell$. Hence, there exists a code with $\log M$ given by (7.39) and probability of error bounded by ϵ that has average stopping time bounded by ℓ. □

Lemma 35. $\mathbb{E}[\tau_1] < \infty$.

Proof. Can be upper bounded by the time M positively biased random walks cross a threshold, each of which is finite, and therefore the max of M of them is finite. □

Lemma 36. *If $S_t^1 = i(X^t; Y^t)$ and $S_t^j = i(\bar{X}_j^t; Y^t)$, $j = 2, \ldots, M$, where $\bar{X}_j$ is i.i.d. with the same distribution as X but independent from Y, then*

$$M_t = \log\frac{e^{S_t^1}}{\sum_{j=2}^{M} e^{S_t^j}} - tC \tag{7.40}$$

is a submartingale.

Proof. Write

$$\log \frac{e^{S_t^1}}{\sum_{j=2}^{M} e^{S_t^j}} = S_t^1 - \log\left(\sum_{j=2}^{M} e^{S_t^j}\right). \tag{7.41}$$

Now, $e^{S_t^j}$ is a martingale, the sum of martingales is a martingale, and log of a martingale is a supermartingale. Hence

$$S_t^1 - tC - \log\left(\sum_{j=2}^{M} e^{S_t^j}\right) \tag{7.42}$$

is a martingale plus a submartingale, hence is a submartingale. □

7.4 The BEC Case

In this section, we discuss results for the Binary Erasure Channel (BEC). For this, the main result is Theorem 40, which says that there exists an $(\ell, M, L, 0)$ scheme satisfying

$$\log \frac{M}{L} \geq \ell C + O(1) \tag{7.43}$$

just as in the $L = 1$ case. First we give the channel definition and description of how its information density evolves.

The BEC is defined as the discrete memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$, output alphabet $\mathcal{Y} = \{0, e, 1\}$, and transition matrix

$$\begin{bmatrix} 1-\delta & \delta & 0 \\ 0 & \delta & 1-\delta \end{bmatrix}. \tag{7.44}$$

For this channel, the capacity achieving input distribution is uniquely Bernoulli(1/2), and the resulting capacity achieving output distribution is given by

$$P_Y(y) = \begin{cases} \frac{1-\delta}{2} & y = 0, 1 \\ \delta & y = e \end{cases}. \tag{7.45}$$

With this, the information density is given by

$$i(x, y) = \begin{cases} \log 2 & y = x \\ 0 & y = e \\ -\infty & y \neq e \text{ and } y \neq x \end{cases}. \tag{7.46}$$

i.e. when the input and output agree, we learn one bit of information, when the

output is erased, we learn 0 bits, and when the input and output disagree, we can automatically rule out that input codeword. We will keep the notation

$$S_t^j = i(c_j^t, y^t) \tag{7.47}$$

to be the information density between the first t symbols of codeword j and the first t symbols of the output sequence. Note that the correct codeword (assume WLOG that $W = 1$ is sent) will always have $S_t^1 \geq 0$ since the input and output will never disagree. With probability δ an erasure occurs, and all codewords have $S_t^j = S_{t-1}^j$, else an erasure does not occur, in which case all incorrect codewords evolve as

$$S_t^{(j)} = \begin{cases} S_{t-1}^{(j)} + 1 & \text{w.p. } 1/2 \\ -\infty & \text{w.p. } 1/2 \end{cases}, \quad j = 2, \ldots, M \tag{7.48}$$

I.e. with probability δ all partial sums that still have $S_{t-1}^{(j)} \geq 0$ "survive", else they "die out" independently, each with probability $1/2$. This is depicted in Figure 7.4. We will view the process as follows: take M independent Geo(1/2) processes, then insert an erasure into all with probability δ at each time step. This is captures in the following definition.

Let $Z_j \sim \text{Geo}(p)$ i.i.d. for $j = 1, \ldots, M$ (generally, we will take $p = 1/2$ later), and define the process

$$V_t = \sum_{j=1}^{M} \mathbb{1}_{\{Z_j \geq t\}} \tag{7.49}$$

In words, V_t is the number of geometric random variables that have "survived" until time t. Notice that the first time when $V_t = k$ for some $k \leq M$ is the $(M-k)$-th order statistic of the random vector $(Z_1, \ldots, Z_M)$. With this established, we can define the BEC(δ) process, which inserts erasures into the independent geometric random variables above.

Definition 5 (BEC(δ) Process)**.** Let V_t bet the process defined in (7.49) with $p = 1/2$. Construct a Markov chain with state space $\{V_1, V_2, \ldots\}$ as follows: set $\bar{V}_1 = V_1$, and transition according to

$$\mathbb{P}[\bar{V}_i = s | \bar{V}_{i-1} = V_l] = \begin{cases} \delta & s = V_l \\ 1 - \delta & s = V_{l+1} \end{cases} \tag{7.50}$$

for $l \in \{1, \ldots, i-1\}$. Then $\bar{V}_t$ is called the BEC(δ) process.

The following lemma gives an upper bound on expected first time that $V_t \leq k$ for some $k \leq M$.

Lemma 37. *Let V_t be given by* (7.49)*. Define the stopping time*

$$\tau_k = \inf\{t \geq 0 : V_t \leq k\} \tag{7.51}$$

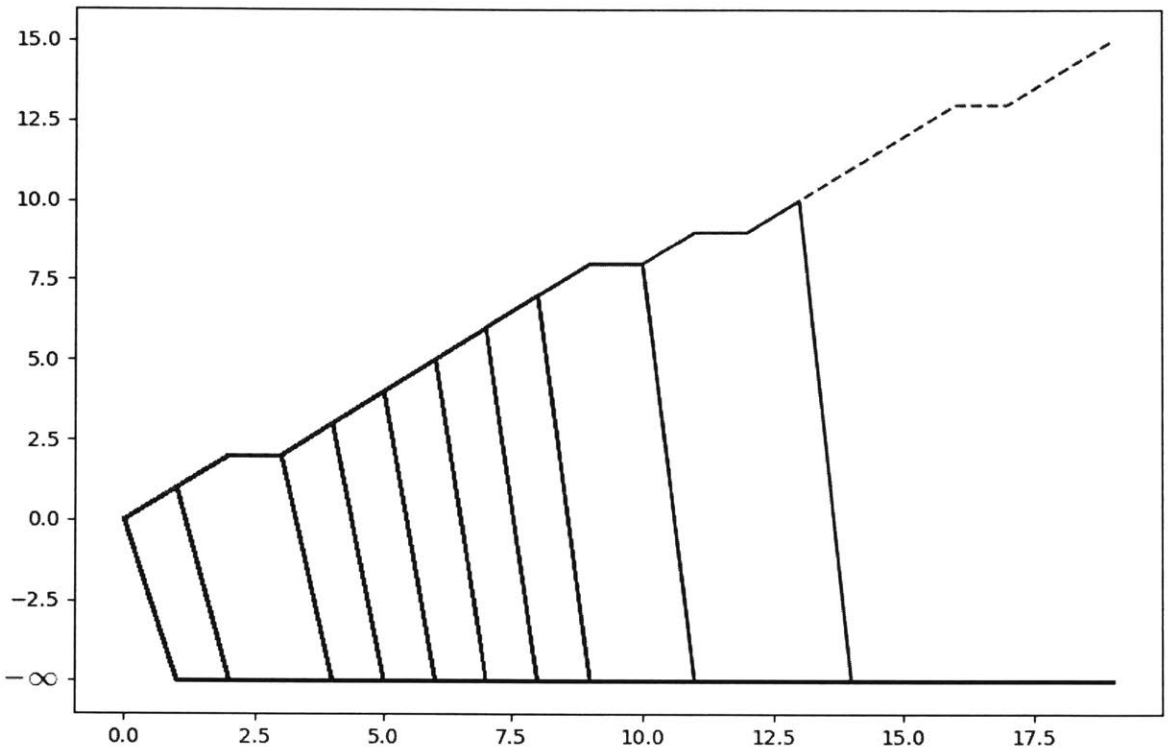

Figure 7-1: Example evolution of the information densities for the BEC, where the x-axis is time and the y axis is the value of the information density. The red curve represents the correct codeword, and the blue curves represent the unsent codewords. At each time step, if there there is no erasure, approximately half of the information densities for the unsent codewords drop to $-\infty$. When an erasure occurs, all information densities increase by 0. Eventually, only the correct codeword survives.

for $0 \leq k \leq M$. *Then*

$$\mathbb{E}[\tau_k] \leq \frac{\log \frac{M}{k} + c}{\log \frac{1}{1-p}} \tag{7.52}$$

where $c \leq 2$ *is a constant independent of all other parameters.*

Proof. First we introduce exponential distributions, and use them to bound the quantities involving geometric distributions. The idea behind this is that the expected order statistics of an exponential distribution has a nice form in terms of harmonic numbers, which we will see shortly.

To this end, consider the following coupling of processes:

$$V_t' = \sum_{i=1}^{M} \mathbb{1}_{\{Z_i' \geq t\}} \tag{7.53}$$

$$V_t = \sum_{i=1}^{M} \mathbb{1}_{\{\lfloor Z_i' \rfloor \geq t\}} \tag{7.54}$$

where $Z_i' \sim \text{Exp}(-\log(1-p))$ i.i.d., so that $\lfloor Z_i' \rfloor \sim \text{Geo}(p)$ i.i.d.. Hence V_t is the same process as defined in (7.49). Since, as a function of t, we have

$$\mathbb{1}_{\{\lfloor Z_i' \rfloor \geq t\}} \leq \mathbb{1}_{\{Z_i' \geq t\}} \ , \tag{7.55}$$

it follows that $V_t \leq V_t'$ almost surely. Consider the stopping times

$$\tau_k = \inf\{t > 0 : V_t \leq k\} \tag{7.56}$$

$$\tau_k' = \inf\{t > 0 : V_t' \leq k\} \ . \tag{7.57}$$

Notice that $V_{\tau_k'} \leq V_{\tau_k'}' \leq k$, and since τ_k' is the first time that $V_t' \leq k$, we have that $\tau_k \leq \tau_k'$ almost surely, and hence

$$\mathbb{E}[\tau_k] \leq \mathbb{E}[\tau_k'] \tag{7.58}$$

The first time that $V_t' \leq k$ is the $(M-k)$-th order statistic of $Z_1', \ldots, Z_M'$. The expectation of the l-th order statistic for an $\text{Exp}(\lambda)$ distribution is given by $\frac{1}{\lambda} \sum_{n=M-l+1}^{M} 1/n$, so we have

$$\mathbb{E}[\tau_k'] = \frac{1}{\lambda} \sum_{n=k+1}^{M} \frac{1}{n} = \frac{1}{\lambda}(H_M - H_k) \tag{7.59}$$

where H_M is the M-th harmonic number. The harmonic numbers are bounded by

$$\frac{1}{2(M+1)} \leq H_M - \log M - \gamma \leq \frac{1}{2M} \tag{7.60}$$

where γ is the Euler-Mascheroni constant ($\gamma \approx 0.577$). Hence

$$H_M \leq \log M + \frac{1}{2M} + \gamma \leq \log M + c \tag{7.61}$$

$$H_k \geq \log k + \frac{1}{2(k+1)} + \gamma \geq \log k \tag{7.62}$$

where $c = 1+\gamma \leq 2$. Applying this to (7.59), we obtain (noting that $\lambda = -\log(1-p)$),

$$\mathbb{E}[\tau'_k] = \frac{\log \frac{M}{k} + c}{\log \frac{1}{1-p}} \ . \tag{7.63}$$

From (7.58), we conclude

$$\mathbb{E}[\tau_k] \leq \frac{\log \frac{M}{k} + c}{\log \frac{1}{1-p}} \ . \tag{7.64}$$

as desired. □

Our next Lemma upper bounds a similar stopping time, but now for the BEC(δ) process.

Lemma 38. *For the BEC(δ) process, define the stopping time*

$$\bar{\tau}_k = \inf\{t \geq 0 : \bar{V}_t \leq k\} \ . \tag{7.65}$$

Then

$$\mathbb{E}[\bar{\tau}_k] \leq \frac{\log \frac{M}{k} + c}{(1-\delta)\log 2} \tag{7.66}$$

where $c \leq 2$.

Proof. From the definition of the BEC(δ) process, define E_t as the number of times the Markov chain failed to transition to a new state before state t was reached. Then E_t has Negative Binomial(t, δ) distribution, and hence

$$\mathbb{E}[E_t] = \frac{t\delta}{1-\delta} \ . \tag{7.67}$$

By definitions (7.51) and (7.65), we can relate the BEC(δ) process stopping time to the process without erasures, via

$$\bar{\tau}_k = \tau_k + E_{\tau_k} \tag{7.68}$$

I.e. V_t is run until time τ_k, then the erasure pattern E_{τ_k} is generated. Hence we have

$$\mathbb{E}\left[\mathbb{E}[E_{\tau_k}|\tau_k]\right] = \mathbb{E}\left[\frac{\tau_k \delta}{1-\delta}\right] \tag{7.69}$$

$$= \frac{\delta}{1-\delta}\mathbb{E}[\tau_k] \; . \tag{7.70}$$

Therefore, using (7.68) and Lemma 37 with $p = 1/2$, we obtain

$$\mathbb{E}[\bar{\tau}_k] = \frac{\mathbb{E}[\tau_k]}{1-\delta} \tag{7.71}$$

$$\leq \frac{\log \frac{M}{k} + c}{(1-\delta)\log 2} \tag{7.72}$$

as claimed. □

We now give a corollary connecting the posterior stopping rule (7.16) to the stopping rule described above. The main point is that the above rule could have easily been discovered using the posterior stopping rule.

Corollary 39. *The stopping rule τ given in* (7.16) *for the BEC also satisfies*

$$\mathbb{E}[\tau] \leq \frac{1}{C}\left(\log \frac{M}{L} + O(1)\right) \tag{7.73}$$

Proof. For the BEC, the sum of the largest L posteriors is given by

$$\max_{S\subset\{1,\ldots,M\},|S|=L} \sum_{j\in S} P_{W|Y^t}(j|y^t) = \max_{S\subset\{1,\ldots,M\},|S|=L} \frac{\sum_{j\in S} e^{S_t^j}}{\sum_{k=1}^{M} e^{S_t^k}} \tag{7.74}$$

$$= \frac{\min(L, \bar{V}_t) e^{S_t^1}}{\bar{V}_t e^{S_t^1}} \tag{7.75}$$

$$= \frac{\min(L, \bar{V}_t)}{\bar{V}_t} \tag{7.76}$$

where $\bar{V}_t$ is given in (7.80). Hence the sum of the largest L posteriors becomes 1 when $\bar{V}_t \leq L$. Lemma 38 shows that the expected first time that $\bar{V}_t$ is less than or equal to k is bounded by (7.66). Hence the first time that the top L posterior sum exceeds $1-\epsilon$ is no larger than the first time the sum is equal to 1, hence we have from Lemma 38,

$$\mathbb{E}[\tau] \leq \frac{1}{C}\left(\log \frac{M}{L} + O(1)\right) \tag{7.77}$$

which shows the claim. □

Now we arrive at the main theorem of this section, which says that for the BEC, we can indeed kill all lower order terms in the expansion of $\log M^*(\ell, M, L, 0)$ for any

list size L. Note that this is a zero error scheme.

Theorem 40. *There exists an $(\ell, M, L, 0)$ VLLD code satisfying*

$$\log \frac{M}{L} \geq \ell C + c_0 \tag{7.78}$$

where $C = (1-\delta)\log 2$ is the BEC capacity, and $c_0 \leq 2 + \log 2$ is a fixed constant.

Proof. Let the encoder map each message $w \in \{1, \ldots, M\}$ to $f_n(w) = X_n(w)$, where $X_n \sim$ Bernoulli(1/2) i.i.d.. The decoder sends `stop` when

$$\tau = \inf\{t \geq 0 : \exists J \subset \{1, \ldots, M\}, |J| \leq L : \forall j \in J^c, i(c_j^t, y^t) = -\infty\}\,. \tag{7.79}$$

In words, the decoder sends `stop` when at most L codewords have information density not equal to $-\infty$. Define

$$\bar{V}_t = \sum_{j=1}^{M} \mathbb{1}_{\{i(c_j^t; Y^t) \geq 0\}} \tag{7.80}$$

then

$$\bar{V}_t = 1 + \sum_{j=1}^{M-1} \mathbb{1}_{\{i(\bar{X}_j^t; Y^t) \geq 0\}} \tag{7.81}$$

since the correct codeword always has $i(c_j^t; Y^t) \geq 0$, and $\bar{X}_j^t$ represents the first t symbols of a randomly generated unsent codeword. Now, the second term is precisely a BEC(δ) process. Applying Lemma 38, noting that τ is the first time that the process with $M-1$ variables is less than or equal to $L-1$, we obtain

$$\mathbb{E}[\tau] \leq \frac{\log \frac{M-1}{L-1} + c}{(1-\delta)\log 2} \tag{7.82}$$

To meet the constraint $\mathbb{E}[\tau] \leq l$, notice that

$$\log \frac{M-1}{L-1} \leq \log \frac{M}{L} + \log 2\,. \tag{7.83}$$

Choose L as

$$\log L = \log M - \ell C + c' \tag{7.84}$$

where $c' = c + \log 2$. This choice of L guarantees $\mathbb{E}[\tau] \leq l$. Hence we have a code with zero error (since the correct codeword never has $S_t^1 = -\infty$) with

$$\frac{M}{L} = \ell C + c' \tag{7.85}$$

codewords and average stopping time $\mathbb{E}[\tau] \leq \ell$, as desired.

□

In Theorem 40, we see that there exists a *zero error* code with only the linear and constant term in this expansion. If we'd like, from this we can form a code where the linear term is boosted to $\frac{C}{1-\epsilon}$, at the expense of no longer being a zero error code.

Corollary 41. *There exists an (ℓ, M, L, ϵ) code for the BEC satisfying*

$$\log \frac{M}{L} \geq \frac{\ell C}{1-\epsilon} + c_0 \tag{7.86}$$

where $c_0 > 0$ is a fixed constant.

Proof. The idea is to use a zero error code for the BEC, but stop at time $t = 0$ with probability ϵ, increasing the error probability but decreasing average stopping time. To this end, construct the new stopping time τ as follows: take an $(\ell', M, L, 0)$ code with $\ell' = \frac{\ell}{1-\epsilon}$ for the BEC that stops at time τ', which we know can be chosen to satisfy, from Theorem 40,

$$\log \frac{M}{L} \geq \ell' C + c_0 \,. \tag{7.87}$$

Our new code stops at time

$$\tau = \begin{cases} \tau' & \text{w.p. } p \\ 0 & \text{else} \end{cases} \tag{7.88}$$

where p is chosen such that the probability of error is

$$P_e \leq p0 + (1-p) = \epsilon \implies p = 1 - \epsilon \,. \tag{7.89}$$

Notice that the average length of this code, by choice of ℓ', is

$$\begin{aligned} \mathbb{E}[\tau] &= p\mathbb{E}[\tau'] + (1-p)0 && (7.90)\\ &= p\ell' && (7.91)\\ &= p\frac{\ell}{1-\epsilon} && (7.92)\\ &= \ell \,. && (7.93) \end{aligned}$$

Hence we have constructed an (ℓ, M, L, ϵ) code satisfying

$$\begin{aligned} \log \frac{M}{L} &\geq \ell' C + c_0 && (7.94)\\ &= \frac{\ell C}{1-\epsilon} + c_0 && (7.95) \end{aligned}$$

as desired. □

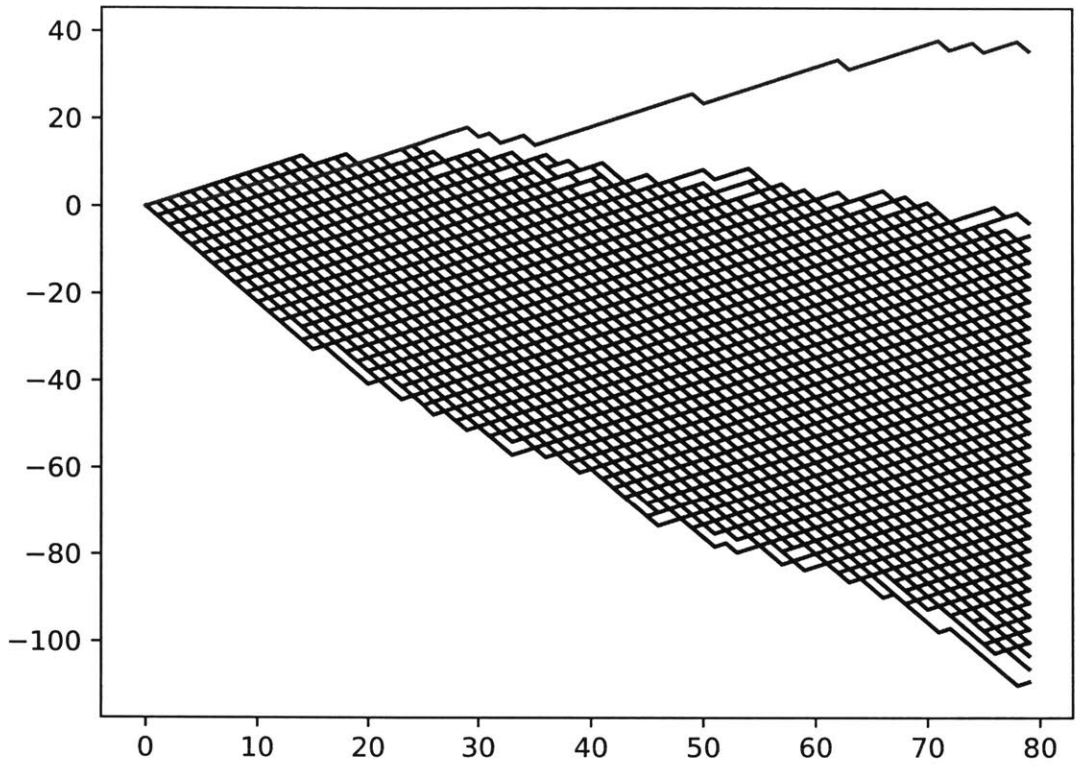

Figure 7-2: Example evolution of the information densities for the BSC, where the x-axis is time and the y-axis is the information density value. The red curve represents the correct codeword, and the blue curves represent the unsent codewords. Eventually, the red curve "pops out" of the collection of blue curves – however in the list decoding problem, we aim to stop before the red curve becomes distinguishable.

7.5 The BSC Case

In this section, we will show that for the BSC, we cannot stop in a way such that the dispersion term vanishes whenever the list has size $L = M^{1-\alpha}$. Note the stark difference in the evolution of the information densities for the BEC and BSC: for the BEC, all values were either equal or zero as in Figure 7.4 – for the BSC information densities move up and down in finite increments. Figure 7.5 gives example trajectories of the BSC information densities. The information density for the correct codeword drifts up with drift $C = D(P_{XY}||P_X P_Y)$ (with P_X being the caid), while the $M-1$ information densities for the unsent codewords drift down with drift $-D(P_X P_Y||P_{XY})$.

The surprising property of the BSC is that the distribution of the posteriors turns out to be nearly deterministic when $L = M^{1-\alpha}$ for some $\alpha \in (0,1)$. Hence, there is no clever way to stop based on the evolution of the posteriors that beats simply ignoring the received sequence. This is shown in the following two sections.

7.5.1 Concentration of the Largest L Posteriors

In this section, we show that the sum of the L largest posteriors concentrates tightly around its mean. This is the contents of the following theorem.

Theorem 42 (Largest L posteriors concentration). *Define the following quantities,*

$$p_t \triangleq \sum_{j \in T_t^L} P_{W|Y^t}(j|Y^t) \tag{7.96}$$

$$\bar{p}_t = \mathbb{P}\left[Y_t^{\bar{\delta}} \geq j_t\right] \tag{7.97}$$

where T_t^L is the set of the largest L posteriors at time t, $Y_t^{\bar{\delta}} \sim Binom(t, \bar{\delta})$, and j_t is such that

$$\mathbb{P}\left[Y_t^{1/2} \geq j_t\right] \leq \frac{L}{M} \leq \mathbb{P}\left[Y_t^{1/2} \geq j_t - 1\right]. \tag{7.98}$$

Then,

$$\mathbb{P}\left[|p_t - \bar{p}_t| \geq t^{-k}\right] \leq c_0 t^{-k_0} \tag{7.99}$$

for some constants $k, k_0, c_0 > 0$.

Proof. First let us introduce notation. Define the quantities

$$S_t^j = i(X_j^t, Y^t) \tag{7.100}$$

$$= t\log(2\delta) + B_t^j \log\frac{1-\delta}{\delta} \tag{7.101}$$

where X_j^t is the first t symbols of the j-th codeword, Y^t is the channel output from codeword 1 (assume $W = 1$ WLOG), $i(x, y)$ is the information density for the BSC, and

$$B_t^j = \begin{cases} \text{Binomial}\,(t, 1-\delta) & j = 1 \\ \text{Binomial}\left(t, \frac{1}{2}\right) & j = 2, \ldots, M \end{cases} \tag{7.102}$$

is a random variable representing the number of flips between Y^t and the first t symbols of the j-th codeword. Note that for the BSC, all S_t^j for $j = 1, \ldots, M$ are independent, which is not true for a general channel. With this definition, we can express the posterior distributions via Bayes rule as

$$P_{W|Y^t}(j|Y^t) = \frac{e^{S_t^j}}{\sum_{k=1}^M e^{S_t^k}} \tag{7.103}$$

for $j = 1, \ldots, M$. Now p_t can be expressed as

$$p_t = \frac{\sum_{j \in T_t^L} e^{S_t^j}}{\sum_{k=1}^{M} e^{S_t^k}} . \tag{7.104}$$

The following arguments will describe the behavior of the numerator and denominator separately.

Lemma 43. *The denominator in* (7.104) *satisfies*

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - M p_{\Gamma_1}\right| \geq M p_{\Gamma_1} t^{-k_0}\right] \leq c t^{-k_1} \tag{7.105}$$

where $\Gamma_1 = \left\{m : \frac{M}{2^t}\binom{t}{m} \geq t^k\right\}$ *for constants* $k, k_0, k_1, c > 0$.

Remark 12. There are two competing behaviors here, in M and in t. I.e. for $M \to \infty$ with fixed t, we have law of large numbers behavior:

$$\frac{1}{M-1}\sum_{j=2}^{M} e^{S_t^j} \approx \mathbb{E}\left[e^{\bar{S}_t}\right] = 1 \tag{7.106}$$

However, for fixed M and $t \to \infty$, we have a softmax type behavior

$$\frac{1}{t}\log\left(\sum_{j=2}^{M} e^{S_t^j}\right) \approx \frac{1}{t}\max_{j=2,\ldots,M} S_t^j . \tag{7.107}$$

Since we can approximate

$$S_t^j \approx t\mu + \sqrt{t\sigma^2} Z_j, \quad j = 2, \ldots, M . \tag{7.108}$$

where $\mu = \mathbb{E}[i(\bar{X}, Y)] \leq 0$ and $\sigma^2 = \mathrm{Var}(i(\bar{X}, Y))$, and the Z_j's are i.i.d. $\mathcal{N}(0,1)$, we see that this tends towards μ.

Hence our argument above in a sense shows that we are operating in the LLN regime.

Remark 13. Note that if we apply Chebyshev's inequality, we see the following:

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - M\right| \geq M\delta\right] \leq \frac{\mathrm{Var}\left(e^{\bar{S}_t}\right)}{M\delta^2} \tag{7.109}$$

$$\leq \frac{\mathbb{E}\left[e^{S_t}\right]}{M\delta^2} \tag{7.110}$$

where in the second line, S_t has the distribution of the information density for the send codeword, and $\bar{S}_t$ has the distribution of the information density for an unset

codeword. If we are interested in times $t \leq \frac{\alpha}{C} \log M$, then

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - M\right| \geq M\delta\right] \leq \frac{1}{M^{1-\alpha\frac{\tilde{C}}{C}}\delta^2} \tag{7.111}$$

where $\tilde{C} = \log \mathbb{E}[e^{i(X;Y)}] > C$. Hence Chebyshev's inequality shows that for times $t \leq \frac{\alpha}{C} \log M$, where $\alpha < C/\tilde{C}$, we have concentration. The more elaborate argument below shows that, for the BSC, we can improve this concentration up to times with $\alpha < 1$.

Proof. We start by rewriting the sum of interest

$$\sum_{j=1}^{M} e^{S_t^j} = \sum_{m=0}^{t} L_m e^{f_t(m)} \tag{7.112}$$

where

$$L_m \triangleq \sum_{j=1}^{M} \mathbb{1}_{\{B_t^j = m\}} \tag{7.113}$$

$$f_t(m) \triangleq t \log 2\delta + m\frac{1-\delta}{\delta} \,. \tag{7.114}$$

Then (7.112) follows from grouping all the occurrences of $S_t^j = f_t(m)$. For now, assume that each $B_t^j \sim \text{Binomial}(t, 1/2)$ i.i.d., i.e. there is no "correct codeword", later in the proof we will show that this approximation is valid. Partition the L_m's into three sets,

$$\Gamma_1 = \{m : \mathbb{E}[L_m] \geq t^k\} \tag{7.115}$$

$$\Gamma_2 = \{m : \mathbb{E}[L_m] \in (t^{-k}, t^k)\} \tag{7.116}$$

$$\Gamma_3 = \{m : \mathbb{E}[L_m] \leq t^{-k}\} \,. \tag{7.117}$$

The idea is that L_m for $m \in \Gamma_1$ exhibits concentration around its mean, and nearly all of the mass is in Γ_1. Applying (7.112), we can bound (7.105) as follows, defining

$p_{\Gamma_1} \triangleq \mathbb{P}\left[Y_t^{\bar{\delta}} \in \Gamma_1\right]$ with Y_t^p denoting the distribution Binomial(t,p), and $\bar{\delta} \triangleq 1-\delta$,

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - Mp_{\Gamma_1}\right| \geq Mp_{\Gamma_1}t^{-k_0}\right] \tag{7.118}$$

$$= \mathbb{P}\left[\left|\sum_{m=0}^{t} L_m e^{f_t(m)} - Mp_{\Gamma_1}\right| \geq Mp_{\Gamma_1}t^{-k_0}\right] \tag{7.119}$$

$$\leq \mathbb{P}\left[\left|\sum_{m\in\Gamma_1} L_m e^{f_t(m)} - Mp_{\Gamma_1}\right| + \sum_{m\in\Gamma_2} L_m e^{f_t(m)} + \sum_{m\in\Gamma_3} L_m e^{f_t(m)} \geq Mp_{\Gamma_1}t^{-k_0}\right] \tag{7.120}$$

$$\leq \underbrace{\mathbb{P}\left[\left|\sum_{m\in\Gamma_1} L_m e^{f_t(m)} - Mp_{\Gamma_1}\right| \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right]}_{1} + \underbrace{\mathbb{P}\left[\sum_{m\in\Gamma_2} L_m e^{f_t(m)} \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right]}_{2}$$

$$+ \underbrace{\mathbb{P}\left[\sum_{m\in\Gamma_3} L_m e^{f_t(m)} \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right]}_{3}. \tag{7.121}$$

Now we bound each term in (7.121) individually.

Term 1 in (7.121): First we apply the union bound to get a statement about binomial distributions:

$$\mathbb{P}\left[\left|\sum_{m\in\Gamma_1} L_m e^{f_t(m)} - Mp_{\Gamma_1}\right| \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right] \tag{7.122}$$

$$\leq \mathbb{P}\left[\exists m \in \Gamma_1 : |L_m - \mathbb{E}[L_m]| \geq \frac{1}{3}t^{-k_0}\mathbb{E}[L_m]\right] \tag{7.123}$$

$$\leq \sum_{m\in\Gamma_1} \mathbb{P}\left[|L_m - \mathbb{E}[L_m]| \geq \frac{1}{3}t^{-k_0}\mathbb{E}[L_m]\right] \tag{7.124}$$

where the first line follows from the fact that $\sum_{m\in\Gamma_1}\mathbb{E}[L_m] = Mp_{\Gamma_1}$, and the second is from the union bound. Note that $L_m \sim$ Binomial$(M, \mathbb{P}[B_t^j = m])$, and the Chernoff bound tells us that if $X \sim$ Binomial(n,p), then for any $\eta > 0$,

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \eta\mathbb{E}[X]\right] \leq e^{-\frac{\eta^2}{3}\mathbb{E}[X]}. \tag{7.125}$$

Applying this to the RHS of (7.124), we obtain

$$\sum_{m\in\Gamma_1} \mathbb{P}\left[|L_m - \mathbb{E}[L_m]| \geq \frac{1}{3}t^{-k_0}\mathbb{E}[L_m]\right] \leq \sum_{m\in\Gamma_1} e^{-\frac{1}{33^2 t^{2k_0}}\mathbb{E}[L_m]} \tag{7.126}$$

$$\leq \sum_{m\in\Gamma_1} e^{-\frac{1}{27t^{2k_0-k}}} \tag{7.127}$$

$$\leq (t+1)e^{-\frac{1}{27}t^{k-2k_0}} \tag{7.128}$$

where we have used that $\mathbb{E}[L_m] \geq t^k$ for $m \in \Gamma_1$ in the second line, and $|\Gamma_1| \leq t+1$ in the third line. Hence term 1 in (7.121) tends to zero exponentially fast whenever $k > 2k_0$.

Term 2 in (7.121)**:** First note that Γ_2 cannot contain very many codewords, since by Markov's inequality,

$$\mathbb{P}\left[\sum_{m\in\Gamma_2} L_m \geq t^{k_0}\right] \leq \frac{(t+1)t^k}{t^{k_0}} \tag{7.129}$$

$$\leq \frac{1}{t^{k_0-k-2}} \tag{7.130}$$

and hence as long as $k_0 > k+2$, this goes to zero. Define A as the event where $\sum_{m\in\Gamma_2} L_m > t^{k_0}$, then term 2 in (7.121) can be bounded by

$$\mathbb{P}\left[\sum_{m\in\Gamma_2} L_m e^{f_t(m)} \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right] \leq \mathbb{P}\left[\sum_{m\in\Gamma_2} L_m e^{f_t(m)} \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\middle| A^c\right] + \mathbb{P}[A] \tag{7.131}$$

$$\leq \mathbb{P}\left[t^{k_0}\max_{m\in\Gamma_2} e^{f_t(m)} \geq \frac{1}{3}Mp_{\Gamma_1}t^{-k_0}\right] + t^{-(k_0-k-2)}. \tag{7.132}$$

In the final line, the quantities inside the $\mathbb{P}[\cdot]$ are deterministic. We now argue that

$$\max_{m\in\Gamma_2} f_t(m) < \log\left(Mp_{\Gamma_1}\right) + O\left(\log t\right). \tag{7.133}$$

which will show that the first term in (7.132) vanishs. To this end, note that we can write

$$f_t(m) = t\left(d\left(\frac{m}{t}\middle|\middle|\frac{1}{2}\right) - d\left(\frac{m}{t}\middle|\middle|\bar{\delta}\right)\right) \tag{7.134}$$

$$\log\mathbb{E}[L_m] = \log M - td\left(\frac{m}{t}\middle|\middle|\frac{1}{2}\right) + O(\log t) \tag{7.135}$$

where $d(p||q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ is the divergence between a Bernoulli(p) and

Bernoulli(q) distribution. Applying the definition of Γ_2 to (7.135), and using the bounds on binomial coefficients, for $k \neq 0, n$,

$$\sqrt{\frac{n}{8k(n-k)}} e^{nh(k/n)} \leq \binom{n}{k} \leq \sqrt{\frac{n}{2\pi k(n-k)}} e^{nh(k/n)}, \tag{7.136}$$

we obtain

$$\log \mathbb{E}[L_m] = \pm k \log t \implies d\left(\frac{m}{t}\middle\|\frac{1}{2}\right) = \frac{1}{t}\log M + O\left(\frac{\log t}{t}\right) \tag{7.137}$$

and hence the LHS of (7.133) is upper bounded by

$$\max_{m\in\Gamma_2} f_t(m) \leq \max_{m\in\Gamma_2} \left(\log M - td\left(\frac{m}{t}\middle\|\bar{\delta}\right) + O(\log t)\right) \tag{7.138}$$

$$\leq \log M - t \min_{m\in\Gamma_2} d\left(\frac{m}{t}\middle\|\bar{\delta}\right) + O(\log t)\,. \tag{7.139}$$

Now we show that the second term in (7.139) scales linearly in t. To this end, we show that there exists a $\bar{\delta}' > \bar{\delta}$ such that $t\bar{\delta}' \in \Gamma_2$. The criteria for membership in Γ_2, as in (7.137), whenever $t \leq \frac{\alpha}{C}\log M$, is that m satisfies

$$d\left(\frac{m}{t}\middle\|\frac{1}{2}\right) = \frac{1}{t}\log M + O\left(\frac{\log t}{t}\right) \tag{7.140}$$

$$\geq \frac{C}{\alpha} + O\left(\frac{\log t}{t}\right) \tag{7.141}$$

Now, $C = d(\bar{\delta}||1/2)$, so since $C/\alpha > C$, any m in Γ_2 must have $m \geq t\bar{\delta}'$, where $\bar{\delta}' = C^{-1}(C/\alpha) > \bar{\delta}$, where the inverse of $C(\bar{\delta})$ is taken on $[1/2, 1]$. Hence we have

$$\min_{m\in\Gamma_2} d\left(\frac{m}{t}\middle\|\bar{\delta}\right) \geq d\left(\bar{\delta}'\middle\|\bar{\delta}\right) > 0\,. \tag{7.142}$$

From this, we conclude that the RHS of (7.133) is upper bounded by

$$\max_{m\in\Gamma_2} f_t(m) \leq \log M - td\left(\bar{\delta}'\middle\|\bar{\delta}\right) + O(\log t) \tag{7.143}$$

The RHS of (7.133) is simpler to bound, via

$$\log Mp_{\Gamma_1} = \log\left(\sum_{m\in\Gamma_1} \mathbb{E}[L_m] e^{f_t(m)}\right) \tag{7.144}$$

$$\geq \max_{m\in\Gamma_1} \left(\log \mathbb{E}[L_m] + f_t(m)\right) \tag{7.145}$$

$$= \max_{m\in\Gamma_1} \left(\log M - td\left(\frac{m}{t}\middle\|\bar{\delta}\right) + O(\log t)\right) \tag{7.146}$$

$$\geq \log M + O(\log t) \tag{7.147}$$

where the first inequality follows from keeping only the largest term in the sum, along with applying $\mathbb{E}[L_m] \geq t^k$ for $m \in \Gamma_1$, and the second follows since $t\bar{\delta} \in \Gamma_1$. We can see this by noting that for any time $t \leq \frac{\alpha}{C} \log M$,

$$\log \mathbb{E}[L_{t\bar{\delta}}] = \log M - tC + O(\log t) \geq tC\left(\frac{1}{\alpha} - 1\right) + O(\log t)\,. \tag{7.148}$$

Combining bound (7.143) and (7.147), we find the first term of (7.132) to vanish, and hence term 2 of (7.121) is bounded by

$$\mathbb{P}\left[\sum_{m\in\Gamma_2} L_m e^{f_t(m)} \geq \frac{1}{3} M p_{\Gamma_1} t^{-k_0}\right] \leq t^{-(k_0-k-2)}\,. \tag{7.149}$$

Term 3 in (7.121)**:** First, notice that the probability that L_m for $m \in \Gamma_3$ is non-zero is small. Indeed,

$$\begin{aligned}
\mathbb{P}[L_m > 0] &= \mathbb{P}[L_m \geq 1] && (7.150)\\
&\leq \mathbb{E}[L_m] && (7.151)\\
&\leq t^{-k} && (7.152)
\end{aligned}$$

where the second line is Markov's inequality, and the third is from the definition of Γ_3. Hence, we can bound term 3 as

$$\begin{aligned}
\mathbb{P}\left[\sum_{m\in\Gamma_3} L_m e^{f_t(m)} \geq \frac{1}{3} M p_{\Gamma_1} t^{-k_0}\right] &\leq \mathbb{P}[\exists m \in \Gamma_3 : L_m > 0] && (7.153)\\
&\leq \sum_{m\in\Gamma_3} \mathbb{P}[L_m > 0] && (7.154)\\
&\leq \frac{|\Gamma_3|}{t^k} && (7.155)\\
&\leq \frac{1}{t^{k-2}} && (7.156)
\end{aligned}$$

where we have used the union bound in the 3rd line, Markov's inequality in the 4th line, and $|\Gamma_3| \leq t + 1 \leq t^2$ in the 5th line.

Finally, note that we can ignore the "correct codeword". Suppose WLOG $W = 1$, then in our original expression i.e. the LHS of (7.105), we can separate out the S_t^1

term,

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - Mp_{\Gamma_1}\right| \geq Mp_{\Gamma_1} t^{-k_0}\right] \tag{7.157}$$

$$\leq \mathbb{P}\left[e^{S_t^1} + \left|\sum_{j=2}^{M} e^{S_t^j} - Mp_{\Gamma_1}\right| \geq Mp_{\Gamma_1} t^{-k_0}\right] \tag{7.158}$$

$$\leq \mathbb{P}\left[e^{S_t^1} \geq \frac{1}{2} Mp_{\Gamma_1} t^{-k_0}\right] + \mathbb{P}\left[\left|\sum_{j=2}^{M} e^{S_t^j} - Mp_{\Gamma_1}\right| \geq \frac{1}{2} Mp_{\Gamma_1} t^{-k_0}\right]. \tag{7.159}$$

Then the second term in (7.159) is bounded by the analysis above $M-1$, and the first term is small, because

$$\mathbb{P}\left[e^{S_t^1} \geq \frac{1}{2} Mp_{\Gamma_1} t^{-k_0}\right] = \mathbb{P}\left[S_t^1 \geq \log(Mp_{\Gamma_1}) + O(\log t)\right] \tag{7.160}$$

$$\leq \mathbb{P}\left[S_t^1 \geq \log M + O(\log t)\right] \tag{7.161}$$

$$\leq \mathbb{P}\left[S_t^1 \geq t\frac{C}{\alpha} + O(\log t)\right] \to 0 \tag{7.162}$$

where the second line used (7.147), and the third uses $\log M \geq t\frac{C}{\alpha}$. The final line goes to zero by the law of large numbers.

Hence, overall we have the bound

$$\mathbb{P}\left[\left|\sum_{j=1}^{M} e^{S_t^j} - Mp_{\Gamma_1}\right| \geq Mp_{\Gamma_1} t^{-k_0}\right] \leq (t+1)e^{-\frac{1}{27}t^{k-2k_0}} + t^{-(k_0-k-2)} + t^{-(k-2)}. \tag{7.163}$$

If e.g. $k = 5/2, k_0 = 1$, then the RHS vanishes as $t \to \infty$. □

Lemma 44. *The denominator in* (7.104) *satisfies*

$$\mathbb{P}\left[\left|\sum_{j \in T_t^L} e^{S_t^j} - Mq\right| \geq Mqt^{-k_0}\right] \leq c_2 t^{-k_1} \tag{7.164}$$

where

$$q = \sum_{m \geq j_*, m \in \Gamma_1} \mathbb{E}[L_m] e^{f_t(m)} \tag{7.165}$$

in the notation of the proof of Lemma 43, and j_* is the solution to, for $Y_t^{1/2} \sim$

Binomial$(t, 1/2)$,

$$\mathbb{P}\left[Y_t^{1/2} \geq j_* + 1\right] \leq \frac{L}{M} \leq \mathbb{P}\left[Y_t^{1/2} \geq j_*\right] \tag{7.166}$$

Proof. The proof is largely the same as Lemma 44, except for one step. Take j_* to be the largest integer value such that

$$\sum_{m \geq j_*} \mathbb{E}[L_m] \geq L\,. \tag{7.167}$$

First, we show that $j_* \in \Gamma_1$.

Lemma 45. *For all time $t \leq \frac{\alpha}{C} \log M$, we have $j_* \in \Gamma_1$ for large enough t.*

Proof. Indeed, for all times $t \leq \frac{\alpha}{C} \log M$,

$$L = M^{1-\alpha} \geq e^{tC\frac{1-\alpha}{\alpha}} \tag{7.168}$$

and hence L is lower bounded by an exponential function in t. Note that $j_* < t/2$, since $L = M^{1-\alpha}$, and $j_* = t/2$ yields $\sum_{m \geq j_*} \mathbb{E}[L_j] \geq M/2$. Hence $\mathbb{E}[L_{j_*}]$ is the largest term in the sum (7.167). Since the sum grows exponentially in t, at least one of the terms must grow exponentially in t, and hence $\mathbb{E}[L_{j_*}]$ certainly grows exponentially in t. Hence $\mathbb{E}[L_{j_*}] \geq t^k$ for large enough t, so $j_* \in \Gamma_1$. □

Because of Lemma 45, we expect that the sum on the LHS of (7.167) concentrates around its mean. To this end, following a similar set of steps as in Lemma 43,

$$\mathbb{P}\left[\left|\sum_{m \geq j_*} L_m - \sum_{m \geq j_*} \mathbb{E}[L_m]\right| \geq t^{-k_1} \sum_{m \geq j_*, m \in \Gamma_1} \mathbb{E}[L_m]\right] \tag{7.169}$$

$$\leq t^{-k'} + \mathbb{P}\left[\left|\sum_{m \geq j_*, m \in \Gamma_1} L_m - \sum_{m \geq j_*, m \in \Gamma_1} \mathbb{E}[L_m]\right| \geq t^{-k_1} \sum_{m \geq j_*, m \in \Gamma_1} \mathbb{E}[L_m]\right] \tag{7.170}$$

$$\leq \sum_{m \geq j_*, m \in \Gamma_1} e^{-\frac{1}{3t^{2k_1}}\mathbb{E}[L_m]} \tag{7.171}$$

$$\leq (t+1)e^{-\frac{1}{3}t^{k-2k_1}} \tag{7.172}$$

$$\leq t^{-k_2} \tag{7.173}$$

where (7.170) follows since $\mathbb{P}\left[\sum_{m \notin \Gamma_1} L_m \geq t^{k_0}\right] \leq t^{-(k_0-k-2)}$ by Markov's inequality, (7.171) follows from Chernoff's bound, and (7.173) follows from $\mathbb{E}[L_m] \geq t^k$ for $m \in \Gamma_1$, just as in (7.149).

Hence we conclude that with probability at least $1 - t^{-k_1}$, by the definition of j_*,

$$\sum_{m \geq j_*} L_m = (1 + t^{-k_1}) \sum_{m \geq j_*} \mathbb{E}[L_m] \geq (1 \pm t^{-k_1}) L \tag{7.174}$$

$$\sum_{m \geq j_* + 1} L_m = (1 + t^{-k_1}) \sum_{m \geq j_* + 1} \mathbb{E}[L_m] \leq (1 \pm t^{-k_1}) L \,. \tag{7.175}$$

Then, with probability at least $(1 - t^{-k_1})^2$, we have

$$\sum_{m \geq j_* + 1} L_m e^{f_t(m)} \leq \sum_{j \in T_t^L} e^{S_t^j} \leq \sum_{m \geq j_*} L_m e^{f_t(m)} \,. \tag{7.176}$$

Noting that

$$\sum_{m \geq j_*} \mathbb{E}[L_m] = M \mathbb{P}\left[Y_t^{1/2} \geq j_* \right] \tag{7.177}$$

we find that j_* satisfies

$$\mathbb{P}\left[Y_t^{1/2} \geq j_* + 1 \right] \leq \frac{L}{M} \leq \mathbb{P}\left[Y_t^{1/2} \geq j_* \right] \tag{7.178}$$

as claimed in (7.166). The concentration of $\sum_{j \in T_t^L} e^{S_t^j}$ now follows from applying the same steps as Lemma 43 to the upper and lower bounds in (7.176).

□

Now we combine lemmas 43 and 44. Define A as the event where the numerator and denominator are both within their high probability bounds, i.e.

$$\left\{ \left| \sum_{j \in T_t^L} e^{S_t^j} - Mq \right| < Mqt^{-k_0} \right\} \bigcap \left\{ \left| \sum_{j=1}^{M} e^{S_t^j} - Mp_{\Gamma_1} \right| < Mp_{\Gamma_1} t^{-k_0} \right\} . \tag{7.179}$$

With this, we can show concentration of p_t. Note that $q/p_{\Gamma_1} = \bar{p}_t$, so that when $t^{-k_0} < 1/4$,

$$\mathbb{P}\left[\left| \frac{\sum_{j \in T_t^L} e^{S_t^j}}{\sum_{j=1}^{M} e^{S_t^j}} - \bar{p}_t \right| > \bar{p}_t 4 t^{-k_0} \right] \leq \mathbb{P}[A^C] \tag{7.180}$$

where in the second line, we have used that, when A occurs and $t^{-k_0} < 1/4$,

$$\bar{p}_t \left(1 - 4t^{-k_0}\right) \leq \bar{p}_t \frac{1 - t^{-k_0}}{1 + t^{-k_0}} \leq p_t \leq \bar{p}_t \frac{1 + t^{-k_0}}{1 + t^{-k_0}} \leq \bar{p}_t \left(1 + 4t^{-k_0}\right) \tag{7.181}$$

From Lemma 43 and Lemma 44, we know that

$$\mathbb{P}\left[A^c\right] \leq (c_1 + c_2) t^{-k_1} \tag{7.182}$$

as claimed. □

7.5.2 Non-Existence of Good Stopping Rules

In the previous section, we showed that the sum of the largest L posteriors concentrates when $L = M^{1-\alpha}$ for any $\alpha \in (0,1)$. In this section, we use this result to show that any stopping rule for the BSC results in a non-zero dispersion term.

Theorem 46. *Let the codebook $f_t(w)$ for $w \in \{1,\dots,M\}, t = 1,2,\dots$ be chosen uniformly at random according to the capacity achieving input distribution, and let τ be any stopping time with $\mathbb{E}[\tau] = \ell$ on filtration $\mathcal{F}_0 = \sigma(\{f_t(w)\|_{t\in\mathbb{Z}_+, w\in[M]})$, and $\mathcal{F} = \sigma(Y_1^t) \vee \mathcal{F}_0$, and such that there exists a decoder $\hat{W} \in \mathcal{F}_\tau$ with probability $\mathbb{P}[W \notin \hat{W}_L] \le \epsilon$, then*

$$\log \frac{M}{L} \le \frac{\ell C}{1-\epsilon} - \sqrt{V\ell \log \ell}\frac{1}{\sqrt{1-\epsilon}} + o(\ell \log \ell) \tag{7.183}$$

where

$$C = \log 2 - h(\delta) \tag{7.184}$$

$$V = \delta(1-\delta)\log\left(\frac{1-\delta}{\delta}\right)^2 \tag{7.185}$$

are the BSC(δ) capacity and dispersion, respectively, and $h(\cdot)$ is the binary entropy function.

Remark 14. Here we demonstrate that the $\frac{\ell C}{1-\epsilon}$ and $\ell \log \ell$ terms can be seen as a consequence of randomizing fixed blocklength codes. Take an (ℓ, M, ϵ) fixed blocklength code for the BSC. We know that such a code can support

$$\log M = \ell C - \sqrt{\ell V} Q^{-1}(\epsilon) + \frac{1}{2}\log \ell + O(1)\,. \tag{7.186}$$

codewords. Suppose we are allowed to randomize the blocklength $\ell \sim \tau$ where $\mathbb{E}[\tau] = \ell$, but τ must be a fixed distribution that does not depend on the codebook or the received symbols. Then, applying the approximation $Q^{-1}(\epsilon) \approx \sqrt{2\log\frac{1}{\epsilon}}$, we see

$$\log \frac{M}{L} \approx \ell C - \sqrt{2V\ell \log \frac{1}{\epsilon}} + \frac{1}{2}\log \ell + O(1)\,. \tag{7.187}$$

Hence, there exists an $(\ell, M, 1/\ell)$ fixed blocklength code satisfying

$$\log \frac{M}{L} \approx \ell C - \sqrt{2V\ell \log \ell} + \frac{1}{2}\log \ell + O(1)\,. \tag{7.188}$$

Now, as done by Polyanskiy [7], with probability $p = \frac{\ell\epsilon' - 1}{\ell - 1}$ choose $\tau = 0$, otherwise

use the $(\ell, M, 1/\ell)$ code above. With this randomization, the probability of error is

$$P_e \leq p + (1-p)\frac{1}{\ell} \tag{7.189}$$

$$= \frac{\ell\epsilon' - 1}{\ell - 1} + \frac{\ell(1-\epsilon')}{\ell - 1}\frac{1}{\ell} \tag{7.190}$$

$$= \epsilon'\frac{\ell - 1}{\ell - 1} \tag{7.191}$$

$$= \epsilon' . \tag{7.192}$$

The average blocklength of this code is now, call this $\mathbb{E}[\tau']$,

$$\mathbb{E}[\tau'] = p\ell + 0(1-p) \tag{7.193}$$

$$= \frac{\ell^2(1-\epsilon')}{\ell - 1} . \tag{7.194}$$

Defining $\ell' = \frac{\ell^2(1-\epsilon')}{\ell-1}$, we see that there exists an (ℓ', M, ϵ') code satisfying

$$\log\frac{M}{L} \geq \frac{\ell' C}{1-\epsilon'} - \sqrt{2V\ell'\log\ell'}\frac{1}{\sqrt{1-\epsilon'}} + o(\ell\log\ell) . \tag{7.195}$$

Hence, at least heuristically, we can achieve (7.195) even if the decoder ignores the codebook and the received sequence, and picked the stopping time τ as above. Note that this argument depends crucially on how $\log\frac{M}{L}$ depends on ϵ.

Proof. We begin by expressing the probability of error in terms of the sum of the largest L posteriors:

$$P_e \geq 1 - \mathbb{E}\left[\sum_{j\in T_L^\tau} P_{W|Y^\tau}(j|Y^\tau)\right] \tag{7.196}$$

where T_L^t indicates the largest L posteriors $P_{W|Y^t}(j|y^t)$ at time t. This lower bound is tight if and only if the decoder g_τ outputs the top L posteriors as the list. Note that regardless of the stopping rule, outputting the messages corresponding to the largest L posteriors always gives a better probability of error than any other list.

The main idea of this argument is the following: we expect the quantity

$$p_t \triangleq \sum_{j\in T_L^t} P_{W|Y^t}(j|Y^t) \tag{7.197}$$

to be nearly deterministic, as shown in the last section, then so too do we expect the probability of error above to evolve deterministically, since the lower bound (7.196) (which is tight when the decoder outputs the largest L posteriors) depends only on

p_t. Rigorously, we have the following chain of inequalities

$$P_e \geq 1 - \mathbb{E}\left[p_\tau\right] \tag{7.198}$$

$$= 1 - \sum_{t=1}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{\tau=t\}} p_t\right] \tag{7.199}$$

$$= 1 - \sum_{t=1}^{\infty} \mathbb{P}\left[|p_t - \bar{p}_t| \leq t^{-k_0}\right] \mathbb{E}\left[\mathbb{1}_{\{\tau=t\}} p_t \middle| |p_t - \bar{p}_t| \leq \bar{p}_t t^{-k_0}\right] \tag{7.200}$$

$$+ \mathbb{P}\left[|p_t - \bar{p}_t| > t^{-k_0}\right] \mathbb{E}\left[\mathbb{1}_{\{\tau=t\}} p_t \middle| |p_t - \bar{p}_t| > t^{-k_0}\right] \tag{7.201}$$

$$\geq 1 - \mathbb{E}\left[\bar{p}_\tau\right] - 2\mathbb{E}\left[\tau^{-k_0}\right] . \tag{7.202}$$

We can interpret the $\mathbb{E}\left[\bar{p}_\tau\right]$ as randomizing between fixed blocklength codes, since $\bar{p}_t$ is the error for a fixed blocklength (t, M, ϵ) code. We will show that such randomization cannot kill the dispersion term. Let ϵ_t be the probability of error for a fixed length code with blocklength t, and suppose we can choose any distribution $\tau \in \{1, 2, \ldots\}$ such that $\mathbb{E}[\tau] = \ell$, how small can $\mathbb{E}[\epsilon_\tau]$ be?

To this end, note that $\bar{p}_t$ satisfies

$$\bar{p}_t = \mathbb{P}\left[Y_t^{\bar{\delta}} \geq \bar{j}_t\right] \tag{7.203}$$

$$\frac{L}{M} = \mathbb{P}\left[Y_t^{1/2} \geq \bar{j}_t\right] \tag{7.204}$$

i.e. take a binary hypothesis test between

$$H_0 : Z \sim Y_t^{1/2} \tag{7.205}$$

$$H_1 : Z \sim Y_t^{\bar{\delta}} \tag{7.206}$$

that outputs H_0 when H_0 is true with probability at least $\bar{p}_t$, then $\bar{p}_t$ gives the minimum type 2 error, i.e. with a slight abuse of notation,

$$\frac{L}{M} = \beta_{\bar{p}_t}\left(Y_t^{\bar{\delta}}, Y_t^{1/2}\right) . \tag{7.207}$$

First note that when $t = 0$, we have that the probability of error is $1 - \frac{1}{M} \approx 1$, i.e. the best that we can do is guess the codeword. Hence our analysis will lower bound $1 - \bar{p}_t$ for $t \geq 1$. From [15, Lemma 14], we have the bound on $\beta_{\bar{p}_t}$, for any $\Delta > 0$,

$$\log \frac{M}{L} = -\log \beta_{\bar{p}_t}\left(Y_t^{\bar{\delta}}, Y_t^{1/2}\right) \tag{7.208}$$

$$\leq tC + \sqrt{tV} Q^{-1}\left(\bar{p}_t + \frac{B + \Delta}{\sqrt{t}}\right) + \frac{1}{2}\log t - \log \Delta \tag{7.209}$$

where C and V are the capacity and dispersion of the BSC, respectively, and $B > 0$

is a fixed constant. Solving for the error $1 - \bar{p}_t$, we obtain

$$1 - \bar{p}_t \geq \Phi\left(\frac{\log \frac{M}{L} - tC - \frac{1}{2}\log \frac{t}{\Delta^2}}{\sqrt{tV}}\right) + \frac{B + \Delta}{\sqrt{t}} \tag{7.210}$$

$$\geq \Phi\left(\frac{\log \frac{M}{L} - tC}{\sqrt{tV}}\right) - \frac{1}{2\sqrt{2\pi tV}} \log \frac{t}{\Delta^2} \tag{7.211}$$

where $\Phi(\cdot)$ denotes the Gaussian CDF, and the second line follows from $\frac{B+\Delta}{\sqrt{t}} \geq 0$ and Taylor's theorem, i.e. defining

$$x = \frac{\log \frac{M}{L} - tC}{\sqrt{tV}}, \tag{7.212}$$

for some $\theta \in \left[x - \frac{1}{2\sqrt{tV}} \log \frac{t}{\Delta^2}, x\right]$, we have

$$\Phi\left(x - \frac{1}{2\sqrt{tV}} \log \frac{t}{\Delta^2}\right) = \Phi(x) - \frac{1}{2\sqrt{tV}} \log \frac{t}{\Delta^2} \varphi(\theta) \tag{7.213}$$

$$\geq \Phi(x) - \frac{1}{2\sqrt{2\pi tV}} \log \frac{t}{\Delta^2} \tag{7.214}$$

where $\varphi(\cdot)$ is the Gaussian pdf, which is maximized at $1/\sqrt{2\pi}$. For large enough M and t, the second term can be made arbitrarily small relative to the first, since the region of interest of the first term is around $t \approx \frac{1}{C} \log \frac{M}{L}$. To this end, choose t large enough so that

$$\frac{1}{2\sqrt{2\pi tV}} \log \frac{t}{\Delta^2} \leq \eta,$$

for some $\eta > 0$ arbitrarily small. Hence we have a lower bound on the probability of error,

$$1 - \bar{p}_t \geq \Phi(x) - \eta. \tag{7.215}$$

Now we return to the question, give any distribution τ on the positive integers such that $\mathbb{E}[\tau] \leq \ell$, how small can $\mathbb{E}[1 - \bar{p}_\tau]$ be? Using (7.215), we have the lower bound

$$\mathbb{E}\left[1 - \bar{p}_\tau\right] \geq \mathbb{E}\left[\Phi\left(\frac{\log \frac{M}{L} - \tau C}{\sqrt{\tau V}}\right)\right] - \eta \tag{7.216}$$

Note that the RHS of (7.215) is a decreasing function of t, as x is monotonically decreasing in t, and that $\Phi(\cdot)$ has an inflection point at 0. Hence we can further lower bound the RHS of (7.216) by the linear function with largest negative slope. The point of this is, if say $1 - \bar{p}_t \geq \max(-at + b, 0)$, then 1) the distribution τ never puts

mass beyond the point $t = b/a$, lest it gains nothing while increasing $\mathbb{E}[\tau]$, and 2)

$$\mathbb{E}[1 - \bar{p}_\tau] \geq \mathbb{E}[\max(-a\tau + b, 0)] = -a\mathbb{E}[\tau] + b \tag{7.217}$$

since we never put mass on the zero part. Hence all distributions τ such that $\mathbb{E}[\tau] = \ell$ give the same lower bound of $-a\ell + b$, eliminating the challenge of optimizing over τ's. To this end, the point of interest t is given by

$$\frac{1 - (1 - \bar{p}_t)}{t} = -\frac{\partial}{\partial t}(1 - \bar{p}_t) \tag{7.218}$$

$$\implies \frac{\bar{p}_t}{t} = \frac{\partial}{\partial t}\bar{p}_t \,. \tag{7.219}$$

Notice that shifting (7.216) by the offset η does not affect the location t where the maximal slope occurs, hence we can ignore η for now. Applying this to the lower bound in (7.216), and we obtain

$$\frac{1}{t}Q\left(\frac{\log\frac{M}{L} - tC}{\sqrt{tV}}\right) = \frac{\partial}{\partial t}Q\left(\frac{\log\frac{M}{L} - tC}{\sqrt{tV}}\right) \tag{7.220}$$

$$= -\varphi\left(\frac{\log\frac{M}{L} - tC}{\sqrt{tV}}\right)\left(-\frac{\log\frac{M}{L} + tC}{2\sqrt{V}t^{3/2}}\right) \tag{7.221}$$

which becomes,

$$Q(x) = \varphi(x)\left(\frac{1}{2}x + \sqrt{t}\frac{C}{\sqrt{V}}\right) \tag{7.222}$$

for x defined as in (7.212). Note that the point t we seek will occur when $x < 0$, since Q is concave on the interval $(-\infty, 0]$, and hence we define $y \triangleq -x$ to work with a positive value. Let $b \triangleq \frac{C}{\sqrt{V}}$ for clarity. Rearranging (7.222) yields

$$\frac{1 - Q(y)}{\varphi(y)} = -\frac{1}{2}y + \sqrt{t}b \,. \tag{7.223}$$

Plugging in $\varphi(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$ yields

$$\sqrt{2\pi}e^{y^2/2} = -\frac{1}{2}y + \sqrt{t}b + \frac{Q(y)}{\varphi(y)} \,. \tag{7.224}$$

When $y \geq 0$, we have constant bounds $Q(y)/\varphi(y)$,

$$0 \leq \frac{Q(y)}{\varphi(y)} \leq \sqrt{\frac{\pi}{2}} \,. \tag{7.225}$$

Using this, first we upper bound y, via

$$\sqrt{2\pi}e^{y^2/2} \leq 0 + \sqrt{t}b + \sqrt{\frac{\pi}{2}}. \tag{7.226}$$

Solving for y yields the upper bound

$$\implies y \leq \sqrt{2\log\frac{\sqrt{t}b + \sqrt{\frac{\pi}{2}}}{\sqrt{2\pi}}}. \tag{7.227}$$

Now, we can also lower bound y via (7.224),

$$\sqrt{2\pi}e^{y^2/2} \geq -\frac{1}{2}\sqrt{2\log\frac{\sqrt{t}b + \sqrt{\frac{\pi}{2}}}{\sqrt{2\pi}}} + \sqrt{t}b + 0 \tag{7.228}$$

$$\geq -\frac{1}{2}\sqrt{t}b + \sqrt{t}b \tag{7.229}$$

$$= \frac{1}{2}\sqrt{t}b \tag{7.230}$$

where in the first line we used (7.227), and the second we used $\log(1+x) \leq x$. Solving for y yields the lower bound

$$y \geq \sqrt{2\log\frac{1}{2}\sqrt{t}b}. \tag{7.231}$$

From (7.227) and (7.231), we conclude that

$$y = (1 + o(1))\sqrt{\log t}. \tag{7.232}$$

Applying the definition of y, this shows that the t from (7.219), which we will call t^*, satisfies

$$\log\frac{M}{L} = t^*C - \sqrt{Vt^*\log t^*}(1 + o(1)). \tag{7.233}$$

Hence, we have the linear lower bound via (7.216) as

$$\mathbb{E}[1 - \bar{\rho}_t] \geq p\Phi\left(\frac{\log\frac{M}{L} - t^*C}{\sqrt{t^*V}}\right) + (1 - p) \tag{7.234}$$

since we have that $(0, 1)$ and $\left(t, \Phi\left(\frac{\log\frac{M}{L} - t^*C}{\sqrt{t^*V}}\right)\right)$ are two points on the line, where p is selected so that $\mathbb{E}[\tau] = \ell$, i.e.

$$pt^* + 0(1 - p) = \ell \implies p = \frac{\ell}{t^*}. \tag{7.235}$$

Plugging this p into (7.234) yields

$$\mathbb{E}[1-\bar{\rho}_t] \geq \frac{\ell}{t^*}\Phi\left(\frac{\log\frac{M}{L} - t^*C}{\sqrt{t^*V}}\right) + 1 - \frac{\ell}{t^*} \tag{7.236}$$

$$= \frac{\ell}{t^*}Q\left(-\sqrt{\log t^*}\right) + 1 - \frac{\ell}{t^*} \tag{7.237}$$

$$\geq 1 - \frac{\ell}{t^*} \tag{7.238}$$

where the second line uses that t^* satisfies (7.233). Setting this equal to ϵ gives the condition on t^*:

$$\epsilon \geq 1 - \frac{\ell}{t^*} \implies t^* \leq \frac{\ell}{1-\epsilon}. \tag{7.239}$$

With this, finally we obtain an upper bound on the number of messages:

$$\log\frac{M}{L} = t^*C - \sqrt{Vt^*\log t^*}(1+o(1)) \tag{7.240}$$

$$\leq \frac{\ell C}{1-\epsilon} + \sqrt{V\ell\log\ell}\frac{1}{\sqrt{1-\epsilon}} + o(\ell\log\ell) \tag{7.241}$$

since the first line is an increasing function in t^* for large enough t^*.

□

Appendix A

Existence of non-Gaussian caids

Proposition 47. *Let $S \subset \mathbb{R}^n$ be such that a) $0 \in S$ and b) there exists a non-zero polynomial in n variables with real coefficients vanishing on S. Then there exists a random variable X taking values in $\mathbb{R}^n$ with the property that its characteristic function $\Psi(t) \stackrel{\triangle}{=} \mathbb{E}\left[e^{i\sum_{k=1}^n t_j X_j}\right], t \in \mathbb{R}^n$ satisfies*

$$\Psi(t) = e^{-\frac{\|t\|_2^2}{2}} \qquad \forall t \in S$$

but there exist a $t_0 \in \mathbb{R}^n$ such that $\Psi(t_0) \neq e^{-\frac{\|t_0\|_2^2}{2}}$ (i.e. $X \not\sim \mathcal{N}(0, I_n)$).

Remark 15. The simplest application of this proposition is the following. Suppose that three random vectors in $\mathbb{R}^3$ have the property that projection onto any (2-dimensional) plane has the joint distribution $\mathcal{N}(0, I_2) \times \mathcal{N}(0, I_2) \times \mathcal{N}(0, I_2)$. Does it imply that the joint distribution of them is $\mathcal{N}(0, I_3) \times \mathcal{N}(0, I_3) \times \mathcal{N}(0, I_3)$? Note that it is easy to argue that joint distribution of any pair of them is indeed $\mathcal{N}(0, I_3) \times \mathcal{N}(0, I_3)$ and thus the only *jointly Gaussian* distribution that satisfies the requirements is indeed the i.i.d. triplet. However, the above proposition shows that the general answer is still negative. Here S is a subset of all $\mathbb{R}^{3\times 3}$ with determinant zero.

Proof. We will slightly extend the argument of [35]. We will assume familiarity with basic commutatitive algebra on the level of [36]. Consider an identity expressing the well-known computation of the Gaussian characteristic function:

$$\frac{1}{\sqrt{2\pi\alpha^2}} \int_{\mathbb{R}} e^{itx - \frac{x^2}{2\alpha^2}} = e^{-\alpha^2 \frac{t^2}{2}} .$$

Setting $\beta = \frac{1}{\alpha^2}$, changing sign of t we get

$$\int_{\mathbb{R}} e^{-itx - \frac{\beta x^2}{2}} dx = \sqrt{\frac{2\pi}{\beta}} e^{-\frac{t^2}{2\beta}} .$$

Differentiating this in β and setting $\beta = \frac{1}{2}$ we get

$$\int_{\mathbb{R}} x^{2k} e^{-itx - \frac{x^2}{4}} dx = p_{2k}(t) e^{-t^2} ,$$

where $p_{2k}(t)$ is some polynomial of degree $2k$ with real coefficients (and involving only even powers of t). For later convenience, we also interchange t and x to get

$$\int_{\mathbb{R}} t^{2k} e^{-itx - \frac{t^2}{4}} \, dt = p_{2k}(x) e^{-x^2} \, . \tag{A.1}$$

(Identity (A.1) also follows from the fact that Hermite polynomials times Gaussian density are eigenfunctions of the Fourier transform.)

Next, suppose that there is a polynomial $q(t_1, \ldots, t_n)$ such that q vanishes on S and each monomial $t_1^{k_1} \cdots t_n^{k_n}$ in q has all $k_1, \ldots, k_n$ even. Then, define the characteristic function

$$\Psi(t_1, \ldots, t_n) \triangleq e^{-\frac{\sum_{k=1}^n t_k^2}{2}} + \epsilon e^{-\frac{\sum_{k=1}^n t_k^2}{4}} q(t_1, \ldots, t_n) \, . \tag{A.2}$$

We will argue that for ϵ sufficiently small, Ψ is a characteristic function of some (obviously non-Gaussian) probability density function f on $\mathbb{R}^n$. By taking the inverse Fourier transform we get that

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\|x\|_2^2}{2}} (1 + \epsilon g(x)) \, .$$

where $e^{-\frac{\|x\|_2^2}{2}} g(x)$ is the inverse Fourier transform of the second term in (A.2). Since $\Psi(t)$ is even in each t_j, we conclude that $f(x)$ is real. Since $q(0) = 0$ (recall that $0 \in S$) we have $\Psi(0) = 1$, and thus $\int_{\mathbb{R}^n} f = 1$. So to prove that f is a valid density function for small ϵ we only need to show that

$$\sup_{x \in \mathbb{R}^n} |g(x)| < \infty \, . \tag{A.3}$$

To that end, notice that applying (A.1) to each monomial $\prod t_j^{2k_j}$ we get

$$\begin{aligned} \int_{\mathbb{R}^n} \left(\prod_{j=1}^n t_j^{2k_j} \right) e^{-i \sum_j t_j x_j - \frac{\|t\|_2^2}{4}} \, dt_1 \cdots dt_n \\ = \left(\prod_j p_{2k_j}(x_j) \right) e^{-\|x\|_2^2} \, . \end{aligned} \tag{A.4}$$

Multiplying the right-hand side by $e^{\frac{\|x\|_2^2}{2}}$ we conclude that contribution of each monomial of q to $\sup_x |g(x)|$ is bounded by

$$\sup_{x \in \mathbb{R}^n} \left| \left(\prod_j p_{2k_j}(x_j) \right) e^{-\frac{\|x\|_2^2}{2}} \right| < \infty \, .$$

Since there are finitely many monomials in q, the proof of (A.3) and of validity of $\Psi(t)$ is done.

We are left to argue that there must necessarily exist polynomial q with required

properties. By assumption there exist some other polynomial q_0 vanishing on S. Consider an inclusion of rings

$$T \triangleq \mathbb{R}[x_1^2, x_2^2, \ldots, x_n^2] \hookrightarrow \mathbb{R}[x_1, \ldots, x_n]$$

where $\mathbb{R}[x_1, \ldots, x_n]$ denotes the ring of polynomials with variables $x_1, \ldots, x_n$ and coefficients in $\mathbb{R}$, and $\hookrightarrow$ denotes an inclusion map. This morphism of rings is obviously finite. Consider ideal (q_0) of $\mathbb{R}[x_1, \ldots, x_n]$ and denote as usual by $(q_0)^c \triangleq (q_0) \cap T$ its contraction. We argue that $(q_0)^c \neq (0)$. Assume otherwise, then we have $(q_0)^c = (0)$ and $\sqrt{(q_0)}^c = (0)$ (since $\sqrt{(0)} = (0)$ as T is an integral domain). Take all minimal primes of (q_0), call these $\{\mathfrak{p}_j\}$, then the radical of (q_0) is the intersection of all prime ideals that contain it, i.e. $\sqrt{(q_0)} = \cap_j \mathfrak{p}_j$. Then, denoting $\mathfrak{q}_j \triangleq \mathfrak{p}_j^c$ we get that $\cap_j \mathfrak{q}_j = (0)$ in T. By "prime-avoidance", cf. [36, Prop. 1.11], we know $(0) \subset \cap_j \mathfrak{q}_j$ implies that $\mathfrak{q}_j \subset (0)$ for some j, hence $\mathfrak{q}_j$ is the zero ideal for some j. This contradicts the "going-up theorem", cf. [36, Corollary 5.9], so we must have $(q_0)^c \neq (0)$, and hence we may take q as an arbitrary non-zero element of $(q_0)^c$. $\square$

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul./Oct. 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[3] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[4] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. trans. telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[5] Spectre, "SPECTRE: Short packet communication toolbox," 2015, GitHub repository. [Online]. Available: https://github.com/yp-mit/spectre

[6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.

[8] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619–2692, 1998.

[9] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.

[10] Y. Polyanskiy and Y. Wu, *Lecture notes on Information Theory*, February 2016. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v4.pdf

[11] P. W. Wolniansky, G. J. Foschini, G. Golden, R. A. Valenzuela *et al.*, "V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *proc. ISSSE*, vol. 98, 1998, pp. 295–300.

[12] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with nonvanishing error probability," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 5–21, 2014.

[13] S. Sandhu and A. Paulraj, "Space-time block codes: A capacity perspective," *IEEE Communications Letters*, vol. 4, no. 12, pp. 384–386, 2000.

[14] A. Collins and Y. Polyanskiy, "Coherent multiple-antenna block-fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 380–405, 2019.

[15] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits.* Princeton University, 2010.

[16] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[17] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2010, available: http://www.princeton.edu/~ypolyans.

[18] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6719–6733, 2015.

[19] P. Billingsley, *Convergence of probability measures.* John Wiley & Sons, 2013.

[20] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "Beta–beta bounds: Finite-blocklength analog of the golden formula," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6236–6256, 2018.

[21] ——, "A beta-beta achievability bound with applications," in *Proc. 2016 IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[22] D. B. Shapiro, *Compositions of quadratic forms.* Walter de Gruyter, 2000, vol. 33.

[23] A. Hurwitz, "Über die komposition der quadratischen formen," *Math. Ann.*, vol. 88, no. 1, pp. 1–25, 1922.

[24] J. Radon, "Lineare scharen orthogonaler matrizen," in *Abh. Sem. Hamburg*, vol. 1. Springer, 1922, pp. 1–14.

[25] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[26] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2468–2503, Oct. 2003.

[27] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[28] R. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Kibernetiki*, vol. 8, pp. 161–168, 1962.

[29] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Problemy peredachi informatsii*, vol. 12, no. 4, pp. 10–30, 1976.

[30] P. Elias, "List decoding for noisy channels," 1957.

[31] J. M. Wozencraft, "List decoding," *Quarterly Progress Report*, vol. 48, pp. 90–95, 1958.

[32] V. M. Blinovsky, "Bounds for codes in the case of list decoding of finite volume," *Problems of Information Transmission*, vol. 22, no. 1, pp. 7–19, 1986.

[33] P. Elias, "Error-correcting codes for list decoding," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 5–12, 1991.

[34] V. Y. Tan and P. Moulin, "Second-order capacities of erasure and list decoding," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 1887–1891.

[35] G. Hamedani and M. Tata, "On the determination of the bivariate normal distribution from distributions of linear combinations of the variables," *American Mathematical Monthly*, pp. 913–915, 1975.

[36] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*. Addison-Wesley Reading, 1969, vol. 2.