Structured Codes in Information Theory: MIMO and Network Applications

Jiening Zhan

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-98
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-98.html

May 11, 2012

Structured Codes in Information Theory: MIMO and Network Applications

by

Jiening Zhan

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

in

Engineering-Electrical Engineering & Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Gastpar, Chair
Professor Anant Sahai
Professor Paul Vojta

Spring 2012

Structured Codes in Information Theory: MIMO and Network Applications

Abstract

Structured Codes in Information Theory: MIMO and Network Applications

by

Jiening Zhan

Doctor of Philosophy in Engineering-Electrical Engineering & Computer Sciences

University of California, Berkeley

Professor Michael Gastpar, Chair

Though random codes are shown to be optimal for point-to-point channels, codes with algebraic structure are found to be useful in many network scenarios. This thesis demonstrates the role of structured codes in several network and MIMO settings. In MIMO channels, structured codes can be used to improve receiver design. Traditional receiver architectures decorrelate the mixed data streams and recover each of them individually. Although optimal when the channel matrix is orthogonal, this can be highly sub-optimal when the channel is near singular. To overcome this limitation, a new architecture that recovers linear combinations of the data streams instead of the data streams individually is proposed. The proposed integer-forcing receiver outperforms traditional linear architectures and achieves the optimal diversity-multiplexing tradeoff. In network information theory, it has been shown that structured codes are useful for computation over multiple-access channels. This thesis considers function computation across general relay networks and proposes a scheme that decouples the physical and network layers. By using lattice codes in the physical layer and network codes in the network layer, the proposed scheme achieves the optimal distortion to within a constant factor. Finally, structured codes can be used to efficiently transmit channel state information when global channel state information is absent in networks. It is shown that sending a function of the channel state information is sufficient and can be much more efficient than sending the full information.

This dissertation is dedicated to my parents.

Acknowledgments

It has been a great pleasure to work with with my advisor, Michael Gastpar, during the course of my graduate school career. Michael taught me that problem solving is not only a science but also an art. His creative approach to research inspired me to try to think outside the box and strive to solve problems that I considered valuable and interesting. His patience and encouragement served as a major source of support during the many stumbling blocks along the way. Thanks for not only providing valuable guidance but also the freedom to work on problems that captivated my interest. Furthermore, special thanks for all your help and feedback with presentations and writing in addition to research.

This thesis also owes much gratitude to Uri Erez, Bobak Nazer, Or Ordentlich, Anant Sahai and Seyong Park. It was a rare opportunity to collaborate with many excellent researchers and definitely something I will miss. Our fruitful discussions not only led to the existence of this thesis but also deepened my understanding of the fundamentals of information theory. Thanks for providing me with such excellent learning opportunities and for your patience in the process. I would also like to thank Paul Vojta for many helpful discussions and for being on my thesis committee and Steven Evans and Kannan Ramchandran for serving on my Qualifying Exam committee.

My groupmates provided invaluable help and support throughout graduate school. I would like to thank Krish Eswaren, Naveen Goela, Nebojsa Milosavljevic, Bobak Nazer, Galen Reeves, Anand Sarwate, and Kelvin So. Special thanks to Bobak Nazer for his guidance in the integer-forcing project and for the countless times he helped me with revisions. Thanks also to Chang Cheng, Venkatesan Ekambaram, Amin Gohari, Pulkit Grover, Sahand Negahban, Dapo Omidiran, Hari Palaiyanur, Sameer Pawar, Gireeja Ranade, Rahul Tandra, Kris Woyach, Hao Zhang, and all the other members of Wireless Foundations for many inspirational discussions inside and outside of research.

I'd like to acknowledge friends outside of Wireless Foundations whose encouragement made even the slow times of research enjoyable. Thanks to Sharon Chang, Andrew Koo, Po-Hsiang Lai, Sonal Singhal, and Ian Tan. Your wise encouragement helped me become more productive and was a source of motivation that led to the completion of this thesis. I'd especially like to thank Ian Tan for taking the non-trivial task of proofreading portions of this thesis.

Finally, I'd like to thank my Mom for her patience and care and my Dad for his humor and wisdom. Thanks for your unconditional love and encouragement, especially throughout the graduate school years.

Contents

Chapter 1

Introduction

Classical information theory generally relied on random coding arguments to characterize the fundamental limits of communication in systems. For example, the capacity achieving codebook of the Additive-White-Noise-Gaussian (AWGN) channel can be constructed by drawing points uniformly from the power constraint sphere. [1, Chapter 10] As a result, there is no particular algebraic structure imposed on the codes. In his seminal paper, Shannon showed that these random coding arguments were sufficient to achieve optimal performance in all single user channels [2]. Furthermore, the same type of codes can be used to achieve the capacity region of several multiple user channels, including the multiple-access channel and special cases of the broadcast channel [1, Chapter 14]. Similarly, Slepian and Wolf used the random coding construction to characterize the optimal rate region for distributed source coding [3]. Codes with algebraic structure were studied in the late sixties, and it was found that although they reduced the encoding and decoding complexity, their performance was, at best, the same as optimal unstructured codes. Generally, though, their performance was usually worse [4]. Consequently, it was suspected that random codes were sufficient to characterize the fundamental limits of communication, and codes with structure were not needed in traditional information theory.

One of the first examples that demonstrated the advantage of codes with structure occurred in the late seventies. Korner and Marton considered the distributed source coding problem when the destination desires to recover a *function* of the sources [5]. In their scenario, Source 1 observes a binary source U, and source 2 observes another binary source U'. Rather than recovering the individual sources U and U' as in the traditional rate-distortion problem, the destination produces an estimate for the mod-2 sum of the source observations: $U \oplus_2 U'$. When the sources are correlated, it was discovered that recovering the mod-2 sum of the observations is more efficient than recovering the individual observations, and the optimal rate region is larger than the Slepian-Wolf region. Furthermore, standard random coding arguments were insufficient here, and linear codes were used instead. Linear codes have the property that the sum (over the underlying finite field) of two codewords is again a codeword. This structural property is crucial for sending the mod-2 sum of the source

observations in this distributed setting.

Lattice codes form the real counterpart of linear codes and extend the linearity property from a finite field into the real field [6]. As a result, they can be used in many wireless settings and form an important class of structured codes. The goal of this thesis is to demonstrate the advantage of lattice codes in three network scenarios. First, we propose a novel, low-complexity, linear receiver architecture for multiple-input-multiple-output (MIMO) channels that achieves significant gain over traditional linear receivers. Our architecture makes use of lattice codes to first recover an equation of the data streams instead of the data streams themselves. This achieves performance close to the optimal joint receiver while adding only slightly more complexity over the traditional linear receiver. Next, we consider function computation across a class of wireless relay networks. By using lattice codes for both channel coding and source quantization, we convert the wireless network problem into a wired network problem, which is well studied in literature [7, 8, 9, 10]. We then develop achievability schemes for the wired network based on the results of the well known multicast problem [11, 12]. Finally, we study the role of structure in channel state estimation. We show that in networks with many relay nodes, lattice codes can be used to transmit a function of the channel state information. In certain cases, this is much more efficient than transmitting the full channel state information.

Before delving into details, we briefly review some situations where lattice codes have demonstrated good performance. In many point-to-point settings, lattice codes have been shown to be a low-complexity alternative that achieves the optimal performance previously attained by random codes:

- **AWGN Channel**: Lattice encoding with Maximum Likelihood decoding was first studied and was shown to attain the capacity of the AWGN channel [13]. Refinements of this scheme were studied in [14, 15]. A nested version of lattice codes was later developed and shown to approach the capacity of point-to-point AWGN channels with lattice decoding [16].

- **Quantization**: Nested lattice codes developed in [16] can also be used for the quantization of Gaussian sources. Furthermore, it is shown that there exists lattices that are good for both coding and quantization [17, 18].

- **Dirty Paper**: When interference is known causally at the transmitter, it can be cancelled by employing precoding techniques [19]. Combing lattice codes and scaling with the MMSE coefficient, the capacity of the dirty paper channel can be achieved [20]. Practical implementations of nested lattice codes for the dirty paper channel are proposed in [21].

- **Wyner-Ziv**: The quantization rate can be reduced by adding correlated side information to the receiver and employing binning techniques. Coset codes have been proposed

as a means to perform efficient binning. Wyner developed a set of coset codes for binning in the situation of lossless compression with side information. Nested lattice codes were shown to be good coset codes in the lossy case by the authors in [22]. These Coset codes can be used for efficient binning in the Wyner-Ziv problem and recover the optimal rate-distortion tradeoff.

Although structured codes can recover the optimal performance previously attained by random codes in many point-to-point scenarios, their value is truly highlighted in network settings. In these multi-user scenarios, lattices are needed to achieve gains previously unattainable using standard random codes:

- **Distributed Source Coding**: A Gaussian version of the Korner-Marton problem was studied in [23]. A nested lattice scheme was used to reconstruct a function of the sources directly without first reconstructing the individual sources. This was shown to be more efficient than recovering the individual sources directly, and the resulting achievable rate region was shown to be larger than the Berger-Tung region in certain regimes.

- **Interference Alignment**: Recent literature showed that the degrees of freedom per user of the interference channel remains constant as the number of users increase [24]. Subspace coding was used, and interference at each destination was aligned so that all interference falls into the same subspace. The linearity property of lattices can be exploited to align the interference [25, 26].

- **Computation over Multiple-Access Channels**: When the destination of a AWGN multiple-access channel recovers a function of the sources, lattice codes have been found to be advantageous and are able to achieve a lower distortion than standard random codes [27].

- **Two-Way Relay Channel**: Lattice codes can be used in this scenario and allow the relay node to recover the sum of the two codewords rather than the individual codewords. Structured schemes have been shown to be more efficient than traditional relaying schemes and achieve within a constant gap of the optimal performance [28, 29].

- **Wireless Network Coding**: In a wired network with many relay nodes, the traditional method of routing has been shown to be insufficient, and coding is needed at the intermediate relays instead [11]. In the case of a wireless network, it is beneficial for the relays to recover only a function of the incoming messages [30, 31, 32, 33].

- **Dirty Paper Multiple-Access Channels**: The dirty paper channel is extended to the case of multiple users who communicate to a common destination in the presence of interference. When the interference is known partially at each transmitter, lattice codes are needed to cancel the interference [34].

- **Secrecy**: Using lattices schemes, multiple users who do not have prior coordination are able to collude against adversaries and eavesdroppers [35].

1.1 Contributions

We provide an overview and a summary of our main results in each scenario:

- **Linear Receiver Design for MIMO:** Linear receivers are often used to reduce the implementation complexity of multiple antenna systems. In a traditional linear receiver architecture, the receive antennas are used to separate out the codewords sent by each transmit antenna, which can then be decoded individually. Although easy to implement, this approach can be highly sub-optimal when the channel matrix is near singular. This paper develops a new linear receiver architecture that uses the receive antennas to create an effective channel matrix with integer-valued entries. Rather than attempting to recover transmitted codewords directly, the decoder recovers integer combinations of the codewords according to the entries of the effective channel matrix. The codewords are all generated using the same linear code, which guarantees that these integer combinations are themselves codewords. If the effective channel is full rank, these integer combinations can then be digitally solved for the original codewords. This thesis focuses on the special case where there is no coding across transmit antennas. In this setting, the integer-forcing linear receiver significantly outperforms traditional linear architectures such as the decorrelator and MMSE receiver. In the high SNR regime, the proposed receiver attains the optimal diversity-multiplexing tradeoff for the standard MIMO channel. It is further shown that in an extended MIMO model with interference, the integer-forcing linear receiver achieves the optimal generalized degrees-of-freedom.

- **Network Function Computation:** In linear function computation, multiple source nodes communicate across a relay network to a single destination whose goal is to recover linear functions of the original source data. For the case when the relay network is a linear deterministic network, a duality relation is established between function computation and broadcast with common messages. Using this relation, a compact, sufficient condition is found describing those cases where the cut-set bound is tight. Then, these insights are used to develop results for the case where the relay network contains Gaussian multiple-access channels. The proposed scheme decouples the physical and network layers. Lattice codes are used for both source quantization and computation in the physical layer. This can be viewed as converting the original Gaussian sources into discrete sources and the Gaussian network into a linear deterministic network. The duality relation is applied to find network codes for computing functions of discrete sources in the network layer. Assuming the original source sequences

are independent Gaussians, the resulting distortion for computing their sum over the Gaussian network is provably within a constant factor of the optimal performance.

- **Channel State Information Estimation:** In networks with many intermediate relay nodes, the assumption of global channel knowledge is optimistic in practice. The fading behavior is typically measured locally at the relay nodes but is not directly known at the destination. One straightforward method is to send the full channel state information to the destination. However, this may be inefficient due to limited power and bandwidth constraints and results in forwarding more information than necessary. Instead, we show that it is sometimes sufficient for the destination to know only a function of the various channel states rather than the full channel state information. We develop a general framework for forwarding a function of the channel state information in relay systems with only local channel knowledge. We apply our framework to several networks and find that *functional forwarding* of channel state information can be much more efficient than full forwarding.

1.2 Outline

The thesis is outlined as follows:

- **Chapter 2:** The standard MIMO model is described and existing receiver architectures are surveyed. The integer-forcing linear receiver, which uses lattice codes to first recover a set of full rank equations of the data streams, is then proposed. It is compared to different architectures and shown to attain the optimal diversity-multiplexing tradeoff.

- **Chapter 3:** The MIMO channel in Chapter 2 is extended to include the case of interference. The integer-forcing architecture is shown to provide an attractive solution to the problem of oblivious interference mitigation. Focusing on the high SNR regime, the generalized degrees of freedom for the integer-forcing linear receiver is characterized and shown to match the optimal joint-receiver.

- **Chapter 4:** The proof of Theorem 3.7 in Chapter 3 requires results in diophantine approximations. Some existing theorems in the diophantine approximations literature, including Dirichlets, Khintchine-Groshev, Minkowski's 1st theorem on successive minimum, and Minkowski's 2nd theorem on successive minima, are first reviewed. Theorem 4.13 is a new result that extends Dirchlets to the case where where a full rank set of linearly independent integer solutions is required. The proof requires several diophantine theorems and makes use of Lagrangian formulations.

- **Chapter 5:** The problem of network function computation is considered. The first section focuses on the deterministic network and shows that computing a single linear function is equivalent to multicast. This insight is then extended beyond a single linear

function to general communication demands. The second section leverages the insights from the deterministic network to characterize the distortion for computing a function of Gaussian sources across a class of Gaussian relay networks.

- **Chapter 6:** The two-stage relay network of interest is first described. A new scheme, *functional forwarding*, where the relays send a function of the channel state information to the destination, is proposed. A series of examples are provided to illustrate that the proposed scheme can be much more efficient than traditional schemes that forward the full channel state information.

Chapter 2

Integer-Forcing Linear Receivers

It is by now well-known that increasing the number of antennas in a wireless system can significantly increase capacity. Since the seminal papers of Foschini and Gans [36] and Telatar [37], multiple-input multiple-output (MIMO) channels have been thoroughly investigated in theory (see [38] for a survey) and implemented in practice [39]. This capacity gain usually comes at the expense of more complex encoders and decoders and a great deal of work has gone into designing low-complexity MIMO architectures. In this chapter, we describe a new low-complexity architecture that can attain significantly higher rates than existing solutions of similar complexity.

We focus on the case where each of the M transmit antennas encodes an independent data stream (see Figure 2.1). That is, there is no coding across the transmit antennas: each data stream $\mathbf{w}_m$ is encoded separately to form a codeword $\mathbf{x}_m$ of length n. Channel state information is only available to the receiver. From the receiver's perspective, the original data streams are coupled in time through encoding and in space (i.e., across antennas) through the MIMO channel. The joint maximum likelihood (ML) receiver simultaneously performs joint decoding across time and receive antennas. Clearly, this is optimal in terms of both rate and probability of error. However, the computational complexity of jointly processing the data streams is high, and it is difficult to implement this type of receiver in wireless systems when the number of streams is large. Instead, linear receivers such as the decorrelator and minimum-mean-squared error (MMSE) receiver are often used as low-complexity alternatives [40].

Traditional linear receivers first separate the coupling in space by performing a linear projection at the front-end of the receiver. In order to illustrate this concept and motivate the proposed new approach, we consider the 2×2 MIMO channel characterized by the following matrix:

$$\mathbf{H} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}. \tag{2.1}$$

The simplest choice of a linear receiver front-end *inverts* the channel matrix. This receiver

is usually referred to as the *decorrelator* in the literature. That is, the receiver first applies the matrix

$$\mathbf{H}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \tag{2.2}$$

to the received signal. Overall, this converts the original channel into a new equivalent channel characterized by the identity matrix and colored Gaussian noise, i.e., to two scalar channels with correlated noise. The linear receiver then proceeds by separately decoding the output of each of these two channels. The well-known drawback of this approach is that the *noise* vector is also multiplied by the linear receiver front-end matrix given in (2.2), which alters the variances of its components. In our simple example, if we assume that the original channel had independent additive white Gaussian noise of unit variance, the equivalent noises after the linear receiver front-end have variances of 2 and 5, respectively. That is, while the receiver front-end has nulled out cross-interference, it has also significantly increased the noise levels.

The integer-forcing linear architecture advocated here is based on the recent insight that if on all transmit antennas, the same linear or lattice code is used, then it is possible to not only decode codewords themselves, but also *integer* linear combinations of codewords directly [31]. Let us denote the codeword transmitted on the first antenna by $\mathbf{x}_1$ and the codeword transmitted on the second antenna by $\mathbf{x}_2$. Then, for the simple example matrix from (2.1), the receiver can decode the integer linear combination $2\mathbf{x}_1+\mathbf{x}_2$ from the first receive antenna and the combination $\mathbf{x}_1+\mathbf{x}_2$ from the second receive antenna. From these (following [31]), it is possible to recover linear equations of the data streams over an appropriate finite field, $2\mathbf{w}_1+\mathbf{w}_2$ and $\mathbf{w}_1+\mathbf{w}_2$. These equations can in turn be digitally solved for the original data streams. The key point in this example is that the noise variances remain unchanged, which increases the effective SNR per data stream. Note that for more general channel matrices beyond the simple example here, it will also be advantageous to first apply an appropriate linear receiver front-end, albeit following principles very different form merely inverting the channel matrix, as we explain in more detail in the sequel.

In this chapter, we first consider the standard MIMO channel and develop a new *integer-forcing* linear receiver architecture that provides multiplexing and diversity gains over traditional linear architectures. Our approach relies on the compute-and-forward framework, which allows linear equations of transmitted messages to be efficiently and reliably decoded over a fading channel [31]. We develop a multiple antenna version of compute-and-forward which employs the antennas at the receiver to rotate the channel matrix towards an effective channel matrix with integer entries. Separate decoders can then recover integer combinations of the transmitted messages, which are finally digitally solved for the original messages. We show that this is much more efficient than using the receive antennas to separate the transmitted codewords and directly decoding each individual codeword. Our analysis uses nested lattice codes originally developed to approach the capacity of point-to-point AWGN and

dirty-paper channels [16, 17, 18, 22] and for which practical implementations were presented in [21] and subsequent works.

In Chapter 3, we generalize the MIMO channel model to include interference [41, 42] and show that the integer-forcing receiver architecture is an attractive approach to the problem of oblivious interference mitigation. By selecting equation coefficients in a direction that depends on both the interference space and the channel matrix, the proposed architecture reduces the impact of interference and attains a non-trivial gain over traditional linear receivers. Furthermore, we show that the integer-forcing receiver achieves the same generalized degrees of freedom as the joint decoder. Our proof uses techniques from Diophantine approximations, which have also recently been used for interference alignment over fixed channels and the characterization of the degrees of freedom for compute-and-forward [43, 44].

In the remainder of the chapter, we start with a formal problem statement in Section 2.1, and then overview the basic existing MIMO receiver architectures and their achievable rates in Section 2.2. In Section 2.3, we present the integer-forcing receiver architecture and a basic performance analysis. We show that the rate difference between the proposed receiver and traditional linear receivers can be arbitrarily large in Section 2.4. We study the outage performance of the integer-forcing linear receiver under a slow fading channel model in Section 2.5. We show that in the case where each antenna encodes an independent data stream, our architecture achieves the same diversity-multiplexing tradeoff as that of the optimal joint decoder. In Chapter 3, we consider the MIMO channel with interference and show that the integer-forcing receiver can be used to effectively mitigate interference. We characterize the generalized degrees-of-freedom for the integer-forcing receiver and find that it is the same as for the joint decoder.

Throughout the chapter, we will use boldface lowercase to refer to vectors, $\mathbf{a} \in \mathbb{R}^M$, and boldface uppercase to refer to matrices, $\mathbf{A} \in \mathbb{R}^{M \times M}$. Let $\mathbf{A}^T$ denote the transpose of a matrix $\mathbf{A}$ and $|\mathbf{A}|$ denote the determinant. Also, let $\mathbf{A}^{-1}$ denote the inverse of $\mathbf{A}$ and $\mathbf{A}^\dagger \triangleq (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ denote the pseudoinverse. The notation $\|\mathbf{a}\| \triangleq \sqrt{\sum_i a_i^2}$ will refer to the ℓ_2-norm of the vector $\mathbf{a}$ while $\|\mathbf{a}\|_\infty \triangleq \max_i |a_i|$ will refer to the ℓ_∞-norm. Finally, we will use $\lambda_{\text{MAX}}(\mathbf{A})$ and $\lambda_{\text{MIN}}(\mathbf{A})$ to refer to the maximum and minimum singular values of the matrix $\mathbf{A}$.

2.1 Problem Statement

The baseband representation of a MIMO channel usually takes values over the complex field. For notational convenience, we will work with the real-valued decomposition of these complex matrices. Recall that any equation of the form $\mathbf{Y} = \mathbf{G}\mathbf{X} + \mathbf{Z}$ over the complex field can be represented by its real-valued representation,

$$\begin{bmatrix} \mathsf{Re}(\mathbf{Y}) \\ \mathsf{Im}(\mathbf{Y}) \end{bmatrix} = \begin{bmatrix} \mathsf{Re}(\mathbf{G}) & -\mathsf{Im}(\mathbf{G}) \\ \mathsf{Im}(\mathbf{G}) & \mathsf{Re}(\mathbf{G}) \end{bmatrix} \begin{bmatrix} \mathsf{Re}(\mathbf{X}) \\ \mathsf{Im}(\mathbf{X}) \end{bmatrix} + \begin{bmatrix} \mathsf{Re}(\mathbf{Z}) \\ \mathsf{Im}(\mathbf{Z}) \end{bmatrix} . \tag{2.3}$$

We will henceforth refer to the $2M \times 2N$ real-valued decomposition of the channel matrix as $\mathbf{H}$. We will use $2M$ independent encoders and $2M$ independent decoders for the resulting real-valued transmit and receive antennas.[1]

Definition 2.1 (Messages)**.** Each of the $2M$ transmit antennas has a length k *data stream* (or message) $\mathbf{w}_m$ drawn independently and uniformly from $\mathcal{W} = \{0, 1, 2, \ldots, q-1\}^k$.

Definition 2.2 (Encoders)**.** Each data stream $\mathbf{w}_m$ is mapped onto a length n channel input $\mathbf{x}_m \in \mathbb{R}^{n \times 1}$ by an *encoder*,

$$\mathcal{E}_m : \mathcal{W} \to \mathbb{R}^n \ .$$

An equal *power allocation* is assumed across transmit antennas

$$\frac{1}{n}\|\mathbf{x}_m\|^2 \leq \mathsf{SNR} \ .$$

Note that equal power constraint per transmit antenna is equivalent to total power constraint when considering the diversity-multiplexing tradeoff. While we formally impose a separate power constraint on each antenna, we note that the performance at high SNR (in terms of the diversity-multiplexing tradeoff) remains unchanged if this is replaced by a sum power constraint over all antennas instead.

Definition 2.3 (Rate)**.** Each of the $2M$ encoders transmits at the same rate

$$R_{\mathrm{TX}} = \frac{k}{n} \log_2 q \ .$$

The *total rate* of the MIMO system is just the number of transmit antennas times the rate, $2MR_{TX}$.

Remark 2.4. Since the transmitters do not have knowledge of the channel matrix, we focus on the case where the $2M$ data streams are transmitted at equal rates. We will compare the integer-forcing receiver against successive cancellation V-BLAST schemes with asymmetric rates in Section 2.5.2.

Definition 2.5 (Channel)**.** Let $\mathbf{X} \in \mathbb{R}^{2M \times n}$ be the matrix of transmitted vectors,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_{2M}^T \end{bmatrix} \ . \tag{2.4}$$

[1]The implementation complexity of our scheme can be decreased slightly by specializing it to the complex field using the techniques in [31]. For notational convenience, we focus solely on the real-valued representation, and do not exploit the constraints on the matrix $\mathbf{H}$.

The *MIMO channel* takes $\mathbf{X}$ as an input, multiplies it by the *channel matrix* $\mathbf{H} \in \mathbb{R}^{2N \times 2M}$ and adds noise $\mathbf{Z} \in \mathbb{R}^{2N \times n}$ whose entries are i.i.d. Gaussian with zero mean and unit variance. The signal $\mathbf{Y} \in \mathbb{R}^{2N \times n}$ observed across the $2N$ receive antennas over n channel uses can be written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z} \, . \tag{2.5}$$

We assume that the channel realization $\mathbf{H}$ is known to the receiver but unknown to the transmitter and remains constant throughout the transmission block of length n.

Definition 2.6 (Decoder)**.** At the receiver, a *decoder* makes an estimate of the messages,

$$\mathcal{D} : \mathbb{R}^{2N \times n} \to \mathcal{W}^{2M} \tag{2.6}$$

$$(\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_{2M}) = \mathcal{D}(\mathbf{y}). \tag{2.7}$$

Definition 2.7 (Achievable Rates)**.** We say that sum rate $R(\mathbf{H})$ is *achievable* if for any $\epsilon > 0$ and n large enough, there exist encoders and a decoder such that reliable decoding is possible

$$\Pr\left((\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_{2M}) \neq (\mathbf{w}_1, \ldots, \mathbf{w}_{2M})\right) \leq \epsilon$$

so long as the total rate does not exceed $R(\mathbf{H})$,

$$2MR_{\mathrm{TX}} \leq R(\mathbf{H}).$$

2.2 Existing Receiver Architectures

Many approaches to MIMO decoding have been studied in the literature. We provide a brief summary of some of the major receiver architectures and the associated achievable rates, including the joint ML receiver, the decorrelator, linear MMSE estimator and the MMSE-SIC estimator.

2.2.1 Joint ML Receivers

Clearly, the best performance is attainable by joint ML decoding across all N receive antennas. This situation is illustrated in Figure 2.1. Let $\mathbf{H}_{\mathcal{S}}$ denote the submatrix of $\mathbf{H}$ formed by taking the columns with indices in $\mathcal{S} \subseteq \{1, 2, \ldots, 2M\}$. If we use a joint ML decoder that searches for the most likely set of transmitted messages vectors $\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_{2M}$, then the following rate is achievable (using Gaussian codebooks at the transmitter):

$$R_{\mathrm{JOINT}}(\mathbf{H}) = \min_{\mathcal{S} \subseteq \{1,2,\ldots,2M\}} \frac{M}{|\mathcal{S}|} \log\det\left(\mathbf{I}_{\mathcal{S}} + \mathsf{SNR}\ \mathbf{H}_{\mathcal{S}}\mathbf{H}_{\mathcal{S}}^T\right) \tag{2.8}$$

where $\mathbf{I}$ is the identity matrix.[2] Note that this is also the capacity of the channel subject to equal rate constraints per transmit antenna. The worst-case complexity of this approach is exponential in the product of the blocklength n and the number of antennas N.

One approach to reduce the complexity of the joint ML decoder is to employ a sphere decoder. Rather than naively checking all possible codewords, the sphere decoder only examines codewords that lie within a ball around the received vector. If the radius of the ball is suitably chosen, this search is guaranteed to return the ML candidate vector. We refer interested readers to [45, 46, 47, 48, 49] for more details on sphere decoding algorithms as well as to [50] for a recent hardware implementation.

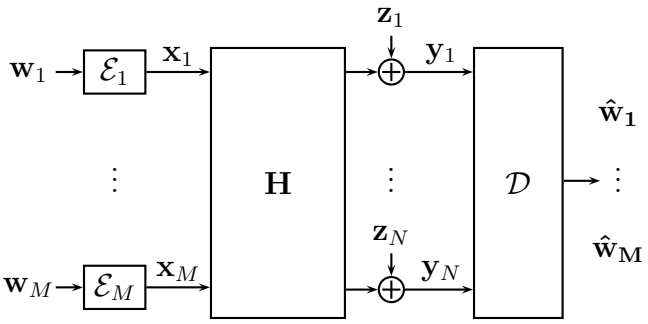

Figure 2.1: MIMO channel with single stream encoding.

2.2.2 Traditional Linear Receivers

Rather than processing all the observed signals from the antennas jointly, one simple approach is to separate out the transmitted data streams using a linear projection and then decode each data stream individually, as shown in Figure 2.2. Given the observed matrix $\mathbf{Y} = \mathbf{HX} + \mathbf{Z}$ from (2.5), the receiver forms the projection

$$\tilde{\mathbf{Y}} = \mathbf{BY} \tag{2.9}$$

$$= \mathbf{BHX} + \mathbf{BZ} \tag{2.10}$$

where $\mathbf{B} \in \mathbb{R}^{2M \times 2N}$. Each row $\tilde{\mathbf{y}}_m^T$ of $\tilde{\mathbf{Y}}$ is treated as a noisy version of $\mathbf{x}_m^T$. In traditional linear receivers, the goal of the projection matrix $\mathbf{B}$ is to separate the incoming data streams. For the decorrelator architecture, we choose the projection to be the pseudoinverse of the channel matrix $\mathbf{B} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$. In the case where $N \geq M$, the resulting channel is interference free. If $\mathbf{H}$ is orthogonal, then the decorrelator architecture can match the performance of a joint ML decoder. As the condition number of $\mathbf{H}$ increases, the performance gap between the decorrelator and the joint decoder increases due to noise amplification (see the example in Section 2.4.2). The performance of the decorrelator can be improved at low

[2] With joint encoding and decoding, a rate of $\frac{1}{2}\log\det\left(\mathbf{I} + \mathsf{SNR}\mathbf{H}\mathbf{H}^T\right)$ is achievable.

SNR using the MMSE architecture which sets $\mathbf{B} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \frac{1}{\mathsf{SNR}}\mathbf{I})^{-1}$. Let $\mathbf{b}_m^T$ be the m^{th} row vector of $\mathbf{B}$ and $\mathbf{h}_m$ the m^{th} column vector of $\mathbf{H}$. The following rate is achievable for the m^{th} data stream using a decorrelator architecture with Gaussian codebooks:

$$R_m(\mathbf{H}) = \frac{1}{2}\log\left(1 + \frac{\mathsf{SNR}\|\mathbf{b}_m^T\mathbf{h}_m\|^2}{\|\mathbf{b}_m\|^2 + \mathsf{SNR}\sum_{i\neq m}\|\mathbf{b}_m^T\mathbf{h}_i\|^2}\right) . \tag{2.11}$$

Since we focus on the case where each data stream is encoded at the same rate, the achievable sum rate is dictated by the worst stream,

$$R_{\text{LINEAR}}(\mathbf{H}) = \min_m 2MR_m(\mathbf{H}) . \tag{2.12}$$

The complexity of a linear receiver architecture is dictated primarily by the choice of decoding algorithm for the individual data streams. In the worst case (when ML decoding is used for each data stream), the complexity is exponential in the blocklength of the data stream. In practice, one can employ low-density parity-check (LDPC) codes to approach rates close to the capacity with linear complexity [51].

The performance of this class of linear receivers can be improved using successive interference cancellation (SIC) [52, 53]. After a codeword is decoded, it may be subtracted from the observed vector prior to decoding the next codeword, which increases the effective signal-to-noise ratio. Let Π denote the set of all permutations of $\{1, 2, \ldots, 2M\}$. For a fixed decoding order $\pi \in \Pi$, let $\pi_m = \{\pi(m), \pi(m+1), \ldots, \pi(2M)\}$ denote the indices of the data streams that have not yet been decoded. Let $\mathbf{h}_{\pi(m)}$ denote the $\pi(m)^{\text{th}}$ column vector of $\mathbf{H}$ and let $\mathbf{H}_{\pi_m}$ be the submatrix consisting of the columns with indices π_m, i.e., $\mathbf{H}_{\pi_m} = [\mathbf{h}_{\pi(m)} \cdots \mathbf{h}_{\pi(2M)}]$. The following rate is achievable in the $\pi(m)^{\text{th}}$, stream using successive interference cancellation:

$$R_{\pi(m)}(\mathbf{H}) = \frac{1}{2}\log\left(1 + \frac{\mathsf{SNR}\|\mathbf{b}_m^T\mathbf{h}_{\pi(m)}\|^2}{\|\mathbf{b}_m\|^2 + \mathsf{SNR}\sum_{i>m}\|\mathbf{b}_m^T\mathbf{h}_{\pi(i)}\|^2}\right) . \tag{2.13}$$

where $\mathbf{b}_m = (\mathbf{H}_{\pi_m}\mathbf{H}_{\pi_m}^T + \frac{1}{\mathsf{SNR}}\mathbf{I})^{-1}\mathbf{h}_{\pi(m)}$ is the projection vector to decode the $\pi(m)^{\text{th}}$ stream after canceling the interference from the $\pi(1), \ldots, \pi(m-1)^{\text{th}}$ streams. For a fixed decoding order π, the achievable sum-rate is given by

$$R_{\text{SIC},1}(\mathbf{H}) = \min_m 2MR_{\pi(m)}(\mathbf{H}) . \tag{2.14}$$

The above scheme is referred to as V-BLAST I (see [54] for more details). An improvement can be attained by selecting the decoding order, and thus the permutation.

In the case of V-BLAST II, the sum rate is improved by choosing the decoding order that maximizes rate of the worst stream,

$$R_{\text{SIC},2}(\mathbf{H}) = \max_{\pi\in\Pi}\min_m 2MR_{\pi(m)}(\mathbf{H}) . \tag{2.15}$$

Hence, V-BLAST I performs worse than V-BLAST II for all channel parameters. We postpone the discussion of V-BLAST III to Section 2.5 where we introduce the outage formulation.

Using ML decoding for each individual data stream, the complexity of the MMSE-SIC architecture is again exponential in blocklength. However, unlike the decorrelator and linear MMSE receiver, not all M streams can be decoded in parallel and delay is incurred as later streams have to wait for earlier streams to finish decoding.

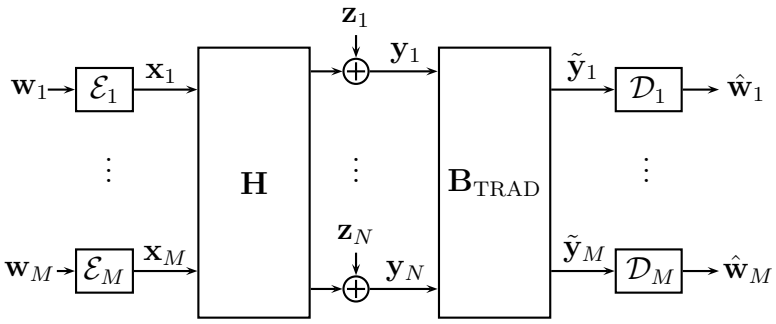

Figure 2.2. A traditional linear receiver. Each of the individual message vectors is decoded directly from the projected channel output. The goal of the linear projection is to approximately invert the channel and cancel the interference from other streams.

2.2.3 Lattice-Reduction Detectors

Another class of linear architectures comes under the name of *lattice-reduction* detectors. It has been shown that lattice reduction can be used to improve the performance of the decorrelator when the channel matrix is near singular [55] and can achieve the receive diversity [56]. Lattice-reduction detectors are symbol-level linear receivers that impose a linear constellation constraint, *e.g.,* a QAM constellation, on the transmitters. The output of the MIMO channel $\mathbf{Y}$ is passed through a linear filter $\mathbf{B}$ to get the resulting output:

$$\widetilde{\mathbf{Y}} = \mathbf{BY} \tag{2.16}$$
$$= \mathbf{BHX} + \mathbf{BZ} \tag{2.17}$$
$$= \mathbf{AX} + \mathbf{BZ} \tag{2.18}$$

where $\mathbf{A} = \mathbf{BH}$ is the effective channel matrix. In lattice reduction, the effective channel matrix is restricted to be unimodular: both its entries and the entries of its inverse must be integers. Let $\mathbf{a}_1^T, \ldots, \mathbf{a}_{2M}^T$ be the row vectors of matrix $\mathbf{A}$. The lattice-reduction detector produces estimates of the symbols of $\mathbf{a}_m^T\mathbf{X}$ from $\widetilde{\mathbf{Y}}$. There are two key differences between

the proposed integer-forcing receiver and the lattice-reduction receiver. First, the integer-forcing receiver operates on the codeword level rather than on the symbol level. Second, the effective channel matrix $\mathbf{A}$ of the integer-forcing receiver is not restricted to be unimodular: it can be any full-rank integer matrix. We compare lattice reduction to the integer receiver in Example 3 by restricting the effective matrix to be unimodular under the integer-forcing architecture. We show that this restriction can result in an arbitrarily large performance gap.

Two other works have developed lattice architectures for joint decoding that can achieve the optimal diversity-multiplexing tradeoff [57, 58].

2.3 Proposed Receiver Architecture

2.3.1 Architecture Overview

Linear receivers such as the decorrelator and the MMSE receiver directly decode the data streams after the projection step. In other words, they use the linear projection matrix $\mathbf{B}$ to invert the channel matrix at the cost of amplifying the noise. Although low in complexity, these approaches are far from optimal when the channel matrix is ill-conditioned. In the integer-forcing architecture, each encoder uses the same linear code and the receiver exploits the code-level linearity to recover equations of the transmitted messages. Instead of inverting the channel, the scheme uses $\mathbf{B}$ to force the effective channel to a full-rank integer matrix $\mathbf{A}$. As in the case of traditional linear receivers, each element of the effective output is then sent to a separate decoder. However, since each encoder uses the same linear code, each decoder can recover an integer linear combination of the codewords. The integer-forcing receiver is free to choose the set of equation coefficients $\mathbf{A}$ to be any full-rank integer matrix. The resulting integer combinations of codewords can be mapped back to a set of full-rank messages over a finite field.[3] Finally, the individual messages vectors are recovered from the set of full-rank equations of message vectors. The details of the architecture are provided in the sequel and an illustration is given in Figure 2.3.

Prior to decoding, our receiver projects the channel output using the $2M \times 2N$ matrix $\mathbf{B}$ to get the effective channel

$$\widetilde{\mathbf{Y}} = \mathbf{BY} \tag{2.19}$$

$$= \mathbf{BHX} + \mathbf{BZ}. \tag{2.20}$$

Each preprocessed output $\tilde{\mathbf{y}}_m$ is then passed into a separate decoder $\mathcal{D}_m : \mathbb{R}^n \to \mathcal{W}$. Decoder

[3]For the scope of the present chapter, we assume that q is prime to ensure invertibility. However, this restriction may be removed as shown in [59].

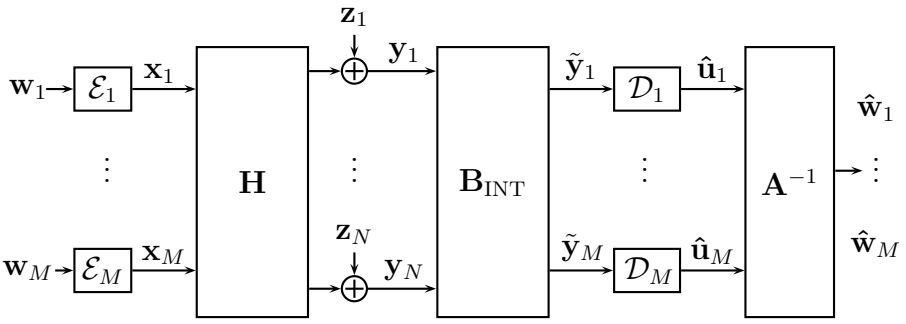

Figure 2.3. The proposed integer-forcing linear receiver. Each decoder first recovers an equation of the transmitted message vectors. These equations are then collected and solved for the individual message vectors. The goal of the linear projection is to create a full-rank, integer-valued effective channel matrix.

m attempts to recover a linear equation of the message vectors

$$\mathbf{u}_m = \left[\sum_{\ell=1}^{2M} a_{m,\ell} \mathbf{w}_\ell \right] \mod q \tag{2.21}$$

for some $a_{m,\ell} \in \mathbb{Z}$. Let $\mathbf{a}_m$ denote the vector of desired coefficients for decoder m, $\mathbf{a}_m = [a_{m1}\ a_{m2}\ \cdots\ a_{m2M}]^T$. We choose $\mathbf{a}_1, \ldots, \mathbf{a}_{2M}$ to be linearly independent.[4] Decoder m outputs an estimate $\hat{\mathbf{u}}_m$ for the equation $\mathbf{u}_m$. We will design our scheme such that, for any $\epsilon > 0$ and n large enough, the desired linear equations are recovered with probability of error satisfying

$$\Pr\left((\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_{2M}) \neq (\mathbf{u}_1, \ldots, \mathbf{u}_{2M})\right) \leq \epsilon\ . \tag{2.22}$$

Let $\mathbf{W} = [\mathbf{w}_1\ \cdots\ \mathbf{w}_{2M}]^T$ denote the matrix of message vectors, $\mathbf{U} = [\mathbf{u}_1\ \cdots\ \mathbf{u}_{2M}]^T$ denote the matrix of linear equations of message vectors and $\mathbf{A} = [\mathbf{a}_1\ \cdots\ \mathbf{a}_{2M}]^T$ denote the integer matrix of equation coefficients. Since $\mathbf{A}$ is full-rank, the original message vectors can be recovered from the set of linear equations by a simple inverse operation:

$$\mathbf{W} = \mathbf{A}^{-1}\mathbf{U} \tag{2.23}$$

In the following subsections, we will provide details on the achievable rate, the choice of the coefficients of the integer matrix $\mathbf{A}$, and the complexity of our architecture.

[4] It is sufficient to consider matrices $\mathbf{B}$ and desired coefficient vectors $\mathbf{a}_m$ that are real-valued decompositions of a complex matrix or vector.

2.3.2 Achievable Rates

We use the compute-and-forward framework developed in [31] to derive the achievable rate of the integer-forcing linear receiver. Let $\mathbf{h}_m^T$ be the m^{th} row vector of $\mathbf{H}$. In the case where $\mathbf{B} = \mathbf{I}$, the channel output to the m^{th} decoder is given by

$$\mathbf{y}_m^T = \mathbf{h}_m^T \mathbf{X} + \mathbf{z}_m^T \tag{2.24}$$

and the rate at which the set of equations $\mathbf{u}_1, \ldots, \mathbf{u}_{2M}$ can be reliably recovered is given in the following theorem. Define $\log^+(x) \triangleq \max\{x, 0\}$.

Theorem 2.8 ([31, Theorem 1])**.** *For any $\epsilon > 0$ and n large enough, there exist fixed encoders and decoders, $\mathcal{E}_1, \ldots, \mathcal{E}_{2M}, \mathcal{D}_1, \ldots, \mathcal{D}_{2M}$, such that all decoders can recover their equations with total probability of error at most ϵ so long as*

$$R_{TX} < \min_{m=1,\ldots,2M} R(\mathbf{H}, \mathbf{a_m}) \tag{2.25}$$

$$R(\mathbf{H}, \mathbf{a}_m) = \frac{1}{2} \log^+ \left(\frac{\mathsf{SNR}}{1 + \mathsf{SNR} \|\mathbf{h}_m - \mathbf{a}_m\|^2} \right) \tag{2.26}$$

for the selected equation coefficients $\mathbf{a}_1, \ldots, \mathbf{a}_{2M} \in \mathbb{Z}^{2M}$.

Remark 2.9. Note that the decoders in Theorem 2.8 are free to choose any equation coefficients that satisfy (2.25). The encoders are completely oblivious to the choice of coefficients.

It is instructive to examine the noise term $1 + \mathsf{SNR}\|\mathbf{h}_m - \mathbf{a}_m\|^2$. The leading 1 corresponds to the additive noise, which has unit variance in our model. The more interesting term $\|\mathbf{h}_m - \mathbf{a}_m\|^2$ corresponds to the "non-integer" penalty since the channel coefficients $\mathbf{h}_m$ are not exactly matched to the coefficients $\mathbf{a}_m$ of the linear equation.

As illustrated in Figure 2.3, we first multiply the channel output matrix $\mathbf{Y}$ by a judiciously chosen matrix $\mathbf{B}$. That is, the effective channel output observed by the m^{th} decoder can be expressed as

$$\tilde{\mathbf{y}}_m^T = \sum_{i=1}^{2M} (\mathbf{b}_m^T \mathbf{h}_i) \mathbf{x}_i^T + \mathbf{b}_m^T \mathbf{Z} \tag{2.27}$$

$$= \tilde{\mathbf{h}}_m^T \mathbf{X} + \tilde{\mathbf{z}}_m^T \tag{2.28}$$

where $\tilde{\mathbf{h}}_m = \mathbf{H}^T \mathbf{b}_m$ is the effective channel to the m^{th} decoder and $\tilde{\mathbf{z}}_m$ is the effective noise with variance $\|\mathbf{b}_m\|^2$. The achievable rate of the integer-forcing linear receiver is given in the next theorem.

Theorem 2.10. *Consider the MIMO channel with channel matrix $\mathbf{H} \in \mathbb{R}^{2N \times 2M}$. Under the integer-forcing architecture, the following rate is achievable:*

$$R < \min_m 2M R(\mathbf{H}, \mathbf{a}_m, \mathbf{b}_m) \tag{2.29}$$

$$R(\mathbf{H}, \mathbf{a}_m, \mathbf{b}_m) = \frac{1}{2} \log^+ \left(\frac{\mathsf{SNR}}{\|\mathbf{b}_m\|^2 + \mathsf{SNR}\|\mathbf{H}^T \mathbf{b}_m - \mathbf{a}_m\|^2} \right)$$

for any full-rank integer matrix $\mathbf{A} \in \mathbb{Z}^{2M \times 2M}$ *and any matrix* $\mathbf{B} \in \mathbb{R}^{2M \times 2N}$.

Proof. Applying Theorem 2.8 with effective channel channel $\tilde{\mathbf{h}}_m = \mathbf{H}^T \mathbf{b}_m$ and effective noise variance $\|\mathbf{b}_m\|^2$, it follows that the receiver can reliably recover the set of linear equations $\mathbf{u}_1, \ldots, \mathbf{u}_{2M}$ where

$$\mathbf{u}_m = \left[\sum_{\ell=1}^{2M} a_{m,\ell} \mathbf{w}_\ell \right] \mod q \ . \tag{2.30}$$

The message vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{2M}$ can be solved in turn by inverting the linear equations, $\mathbf{W} = \mathbf{A}^{-1}\mathbf{U}$. □

Theorem 2.10 provides an achievable rate for the integer-forcing architecture for any preprocessing matrix $\mathbf{B}$ and any full-rank integer matrix $\mathbf{A}$. The remaining task is to select these matrices in such a way as to *maximize* the rate expression given in Theorem 2.10. This turns out to be a non-trivial task. We consider it in two steps. In particular, we first observe that for a fixed integer matrix $\mathbf{A}$, it is straightforward to characterize the optimal preprocessing matrix $\mathbf{B}$. Then, in the next subsection, we discuss the problem of selecting the integer matrix $\mathbf{A}$.

We consider the case when $N \geq M$ and note that given a fixed full-rank integer matrix $\mathbf{A}$, a simple choice for preprocessing matrix is

$$\mathbf{B}_{\text{EXACT}} = \mathbf{H}^\dagger \mathbf{A} \tag{2.31}$$

where $\mathbf{H}^\dagger$ is the pseudoinverse of $\mathbf{H}$. We call this scheme "exact" integer-forcing since the effective channel matrix after preprocessing is simply the full-rank integer matrix $\mathbf{A}$. We also note that choosing $\mathbf{B}_{\text{EXACT}}$ and $\mathbf{A} = \mathbf{I}$ corresponds to the decorrelator. More generally, the performance of exact integer-forcing is summarized in the following corollary.

Corollary 2.11. *Consider the case where* $N \geq M$. *The achievable rate from Theorem 2.10 can be written equivalently as*

$$R < \min_m 2M R(\mathbf{H}, \mathbf{a}_m) \tag{2.32}$$

$$R(\mathbf{H}, \mathbf{a}_m) = \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{\|(\mathbf{H}^T)^\dagger \mathbf{a}_m\|^2} \right) \tag{2.33}$$

for any full-rank integer matrix $\mathbf{A}$ *by setting* $\mathbf{B} = \mathbf{H}^\dagger \mathbf{A}$.

We call the expression in the denominator of (2.33) the "effective noise variance." The achievable rate in (2.32) is determined by the largest effective noise variance,

$$\tilde{\sigma}^2_{\text{EFFECTIVE}} = \max_m \|(\mathbf{H}^T)^\dagger \mathbf{a}_m\|^2 \ . \tag{2.34}$$

Hence, the goal is to choose linearly independent equations $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{2M}$ to minimize the expression $\tilde{\sigma}^2_{\text{EFFECTIVE}}$ in (2.34). The integer-forcing receiver provides the freedom to choose any full-rank integer matrix $\mathbf{A}$. In the remainder of this section, we characterize the optimal linear projection matrix $\mathbf{B}$ for a fixed coefficient matrix $\mathbf{A}$ and provide an equivalent rate expression for Theorem 2.10. We will then use this expression in the Section 2.3.3 to provide insight into how to select the optimal coefficient matrix $\mathbf{A}$.

Corollary 2.12. *The optimal linear projection matrix for a fixed coefficient matrix* $\mathbf{A}$ *is given by*

$$\mathbf{B}_{OPT} = \mathbf{A}\mathbf{H}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1}. \tag{2.35}$$

Remark 2.13. The linear MMSE estimator, given by $\mathbf{B}_{\text{MMSE}} = \mathbf{H}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1}$, is a special case of the integer-forcing receiver with $\mathbf{B}_{\text{OPT}}$ and $\mathbf{A} = \mathbf{I}$.

Remark 2.14. $\lim_{\mathsf{SNR}\to\infty} \mathbf{B}_{\text{OPT}} = \mathbf{B}_{\text{EXACT}}$. Hence, under a fixed channel matrix, exact integer-forcing is optimal as $\mathsf{SNR} \to \infty$.

Proof of Corollary 2.12. Let $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_{2M}]^T$. We solve for each $\mathbf{b}_m$ separately to maximize the achievable rate in Theorem 2.10,

$$\begin{aligned} \mathbf{b}_m &= \text{argmax}_{\mathbf{b}_m} \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{\|\mathbf{b}_m\|^2 + \mathsf{SNR}\|\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m\|^2} \right) \\ &= \text{argmin}_{\mathbf{b}_m} \frac{1}{\mathsf{SNR}} \|\mathbf{b}_m\|^2 + \|\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m\|^2 \,. \end{aligned} \tag{2.36}$$

Define this quantity to be the function $f(\mathbf{b}_m)$ and rewrite as follows:

$$f(\mathbf{b}_m) = \frac{1}{\mathsf{SNR}} \|\mathbf{b}_m\|^2 + \|\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m\|^2 \tag{2.37}$$

$$= \frac{1}{\mathsf{SNR}} \mathbf{b}_m^T\mathbf{b}_m + (\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m)^T(\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m) \tag{2.38}$$

$$= \frac{1}{\mathsf{SNR}} \mathbf{b}_m^T\mathbf{b}_m + \mathbf{b}_m^T\mathbf{H}\mathbf{H}^T\mathbf{b}_m - 2\mathbf{b}_m^T\mathbf{H}\mathbf{a}_m + \mathbf{a}_m^T\mathbf{a}_m \tag{2.39}$$

$$= \mathbf{b}_m^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right) \mathbf{b}_m - 2\mathbf{b}_m^T\mathbf{H}\mathbf{a}_m + \mathbf{a}_m^T\mathbf{a}_m \tag{2.40}$$

$$\tag{2.41}$$

Taking the derivative of f with respect to $\mathbf{b}_m$, we have that

$$\frac{df(\mathbf{b}_m)}{d\mathbf{b}_m} = 2 \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right) \mathbf{b}_m - 2\mathbf{H}\mathbf{a}_m. \tag{2.42}$$

Setting $\frac{df(\mathbf{b}_m)}{d\mathbf{b}_m} = 0$ and solving for $\mathbf{b}_m$, we have that

$$\mathbf{b}_m^T = \mathbf{a}_m^T \mathbf{H}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1}. \tag{2.43}$$

Corollary 2.12 follows since $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_{2M}]^T$. □

Using the optimal linear projection matrix from Corollary 2.12, we derive an alternative expression for the achievable rate in Theorem 2.10.

Theorem 2.15. *The achievable rate from Theorem 2.10 under the optimal projection matrix* $\mathbf{B}_{OPT}$ *from (2.35) can be expressed as*

$$R < \min_m 2M R(\mathbf{H}, \mathbf{a}_m) \tag{2.44}$$

$$R(\mathbf{H}, \mathbf{a}_m) = -\frac{1}{2} \log \mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m \ , \tag{2.45}$$

where $\mathbf{V} \in \mathbb{R}^{2M \times 2M}$ *is the matrix composed of the eigenvectors of* $\mathbf{H}^T\mathbf{H}$ *and* $\mathbf{D} \in \mathbb{R}^{2M \times 2M}$ *is a diagonal matrix with elements*

$$\mathbf{D}_{i,i} = \begin{cases} \frac{1}{\mathsf{SNR}\lambda_i^2 + 1} & i \leq rank(\mathbf{H}) \\ 1 & i > rank(\mathbf{H}) \end{cases} \tag{2.46}$$

and λ_i *is the* i^{th} *singular value of* $\mathbf{H}$.

Proof. Let f be defined as in (2.37) and define $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ to be the singular value decomposition (SVD) of $\mathbf{H}$ with $\mathbf{U} \in \mathbb{R}^{2N \times 2N}, \mathbf{\Sigma} \in \mathbb{R}^{2N \times 2M}$, and $\mathbf{V} \in \mathbb{R}^{2M \times 2M}$. Note that in this SVD representation, $\Sigma_{i,i} = \lambda_i$ and $\Sigma_{i,j} = 0$ for all $i \neq j$. Evaluating f for the m^{th} row $\mathbf{b}_m$ of $\mathbf{B}_{\text{OPT}}$ yields

$$f(\mathbf{b}_m) = \frac{1}{\mathsf{SNR}} \mathbf{b}_m^T \mathbf{b}_m + \mathbf{b}^T \mathbf{H}\mathbf{H}^T \mathbf{b}_m - \mathbf{b}_m^T \mathbf{H}\mathbf{a}_m - \mathbf{a}_m^T \mathbf{H}^T \mathbf{b}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.47}$$

$$= \mathbf{b}_m^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right) \mathbf{b}_m - \mathbf{b}_m^T \mathbf{H}\mathbf{a}_m - \mathbf{a}_m^T \mathbf{H}^T \mathbf{b}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.48}$$

Combining (2.43) and (2.48), it follows that

$$f(\mathbf{b}_m) = \mathbf{b}_m^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right) \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{H}\mathbf{a}_m - \mathbf{b}_m^T \mathbf{H}\mathbf{a}_m - \mathbf{a}_m^T \mathbf{H}^T \mathbf{b}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.49}$$

$$= \mathbf{b}_m^T \mathbf{H}\mathbf{a}_m - \mathbf{b}_m^T \mathbf{H}\mathbf{a}_m - \mathbf{a}_m^T \mathbf{H}^T \mathbf{b}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.50}$$

$$= -\mathbf{a}_m^T \mathbf{H}^T \mathbf{b}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.51}$$

$$= -\mathbf{a}_m^T \mathbf{H}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{H}\mathbf{a}_m + \mathbf{a}_m^T \mathbf{a}_m \tag{2.52}$$

$$= -\mathbf{a}_m^T \mathbf{V}\mathbf{\Sigma}^T \mathbf{U}^T \left(\frac{1}{\mathsf{SNR}} \mathbf{I} + \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T \mathbf{U}^T \right)^{-1} \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \mathbf{a}_m + \mathbf{a}_m^T \mathbf{I} \mathbf{a}_m. \tag{2.53}$$

Since $\mathbf{U}$ is an orthonormal matrix, $\mathbf{U}^{-1} = \mathbf{U}^T$ and (2.53) can be rewritten as follows

$$f(\mathbf{b}_m) = -\mathbf{a}_m^T \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T \left(\frac{1}{\mathsf{SNR}}\mathbf{U}\mathbf{I}\mathbf{U}^T + \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T\right)^{-1} \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{a}_m + \mathbf{a}_m^T\mathbf{I}\mathbf{a}_m \tag{2.54}$$

$$= -\mathbf{a}_m^T \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T(\mathbf{U}^T)^{-1} \left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1} \mathbf{U}^{-1}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{a}_m + \mathbf{a}_m^T\mathbf{I}\mathbf{a}_m \tag{2.55}$$

$$= -\mathbf{a}_m^T \mathbf{V}\mathbf{\Sigma}^T \left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1} \mathbf{\Sigma}\mathbf{V}^T\mathbf{a}_m + \mathbf{a}_m^T\mathbf{I}\mathbf{a}_m. \tag{2.56}$$

Since $\mathbf{V}$ is an orthonormal matrix, $\mathbf{V}^{-1} = \mathbf{V}^T$ and (2.56) can be rewritten as follows

$$f(\mathbf{b}_m) = \mathbf{a}_m^T \left(-\mathbf{V}\mathbf{\Sigma}^T \left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1} \mathbf{\Sigma}\mathbf{V}^T + \mathbf{V}\mathbf{V}^T\right)\mathbf{a}_m \tag{2.57}$$

$$= \mathbf{a}_m^T\mathbf{V} \left(-\mathbf{\Sigma}^T \left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1} \mathbf{\Sigma} + \mathbf{I}\right)\mathbf{V}^T\mathbf{a}_m \tag{2.58}$$

$$= \mathbf{a}_m^T\mathbf{V} \left(\mathbf{I} - \mathbf{\Sigma}^T \left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{\Sigma}\mathbf{\Sigma}^T\right)^{-1} \mathbf{\Sigma}\right)\mathbf{V}^T\mathbf{a}_m \tag{2.59}$$

$$= \mathbf{a}_m^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{a}_m \ . \tag{2.60}$$

Putting everything together, we have that

$$R(\mathbf{H}, \mathbf{a}_m) = -\frac{1}{2}\log \mathbf{a}_m^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{a}_m \ . \tag{2.61}$$

□

2.3.3 Choosing Equations

In the previous section, we explored choices of the preprocessing matrix $\mathbf{B}$ and characterized the optimal $\mathbf{B}$ for a fixed full-rank integer matrix $\mathbf{A}$. Now, we discuss how to select equation coefficients $\mathbf{a}_1, \cdots, \mathbf{a}_{2M}$ to maximize the achievable rate in Theorem 2.10 or, equivalently, Theorem 2.15. In the integer-forcing linear receiver, we are free to recover any full-rank set of linear equations with integer coefficients. However, due to the integer constraint on $\mathbf{A}$, it does not appear to be possible to give a closed-form solution for the best possible full-rank matrix $\mathbf{A}$.

An initially tempting choice for $\mathbf{A}$ might be $\mathbf{A} = \mathbf{I}$. As we noted previously, for this choice of $\mathbf{A}$, selecting $\mathbf{B} = \mathbf{H}^\dagger$ reduces to the decorrelator while selecting $\mathbf{B} = \mathbf{H}^T\left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{H}\mathbf{H}^T\right)^{-1}$ yields the linear MMSE estimator. However, as we show, for most channel matrices, fixing $\mathbf{A} = \mathbf{I}$ is suboptimal.

From Theorem 2.15, the achievable rate under the fixed channel matrix $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_{2M}]^T$ is given by

$$R < \max_{|\mathbf{A}|\neq 0} \min_m \left(-M \log \mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m\right). \tag{2.62}$$

In general, for a fixed SNR and channel matrix, finding the best coefficient matrix $\mathbf{A}$ appears to be a combinatorial problem, requiring an explicit search over all possible full-rank integer matrices. The following lemma shows how the search space can be somewhat reduced.

Lemma 2.16. *To optimize the achievable rate in Theorem 2.15 (or, equivalently, in Theorem 2.10), it is sufficient to check the space of all integer vectors* $\mathbf{a}_m$ *with norm satisfying*

$$\|\mathbf{a}_m\|^2 \leq 1 + \lambda_{MAX}^2 \mathsf{SNR} \ . \tag{2.63}$$

where λ_{MAX} *is the maximum singular value of* $\mathbf{H}$.

Remark 2.17. This lemma thus shows that an exhaustive search only needs to check roughly SNR^M possibilities.

Proof. From (2.62), the achievable rate of the integer-forcing receiver is zero for all $\mathbf{a}_m$ satisfying

$$\mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m \geq 1 \tag{2.64}$$

The left-hand side is lower bounded by

$$\begin{aligned} \mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m &= \|\mathbf{D}^{1/2}\mathbf{V}^T \mathbf{a}_m\|^2 && (2.65)\\ &= \sum_{i=1}^{2M} D_{i,i} |\mathbf{v}_i^T \mathbf{a}_m|^2 && (2.66)\\ &\geq \min_i D_{i,i} \|\mathbf{a}\|^2 && (2.67)\\ &= \frac{1}{1 + \lambda_{\mathrm{MAX}}^2 \mathsf{SNR}} \|\mathbf{a}_m\|^2 && (2.68) \end{aligned}$$

Hence, if $\|\mathbf{a}_m\|^2 \geq 1 + \lambda_{\mathrm{MAX}}^2 \mathsf{SNR}$, then $\mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m \geq 1$. □

To conclude this subsection, we will now explicitly show how and why the choice $\mathbf{A} = \mathbf{I}$ is indeed suboptimal in general. In this context, it is instructive to use Lemma 2.16 to restate (2.62) as

$$R < \max_{\substack{|\mathbf{A}|\neq 0 \\ \|\mathbf{a}_m\|^2 \leq 1 + \lambda_{\mathrm{MAX}}^2 \mathsf{SNR}}} \min_m \left(-M \log \mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m\right). \tag{2.69}$$

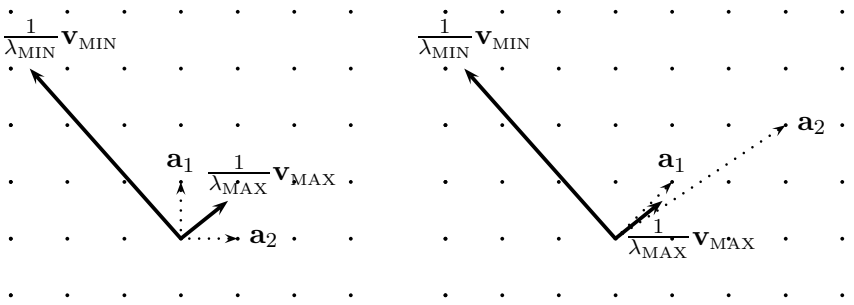

Figure 2.4. The decorrelator (left) fixes the equations to be $\mathbf{a}_1 = [1 \;\; 0]^T$ and $\mathbf{a}_2 = [0 \;\; 1]^T$. The integer-forcing linear receiver (right) allows for any choice of linearly independent equations. Equations should be chosen in the direction of $\mathbf{v}_{\text{MAX}}$ to avoid noise amplification by $\frac{1}{\lambda_{\text{MIN}}}$.

Equation (2.69) suggests that we should choose coefficient vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{2M}$ to be short and in the direction of the maximum right eigenvector of $\mathbf{H}$. To make this concrete, let us study a particular 2×2 real MIMO channel for which the matrix $\mathbf{H}$ has singular values λ_{MIN} and λ_{MAX}, with corresponding right singular vectors $\mathbf{v}_{\text{MIN}}$ and $\mathbf{v}_{\text{MAX}}$, respectively, as illustrated in Figure 2.4. Here, decoder 1 recovers a linear combination of the transmitted message vectors with integer coefficients $\mathbf{a}_1 = [a_{1,1} \;\; a_{1,2}]^T$ and decoder 2 recovers a linear combination with integer coefficients $\mathbf{a}_2 = [a_{2,1} \;\; a_{2,2}]^T$. Using the exact integer-forcing rate from Corollary 2.11, the following rate is achievable

$$R < \min_{m=1,2} \log \left(\frac{\mathsf{SNR}}{\widetilde{\sigma}_m^2} \right) \tag{2.70}$$

where $\widetilde{\sigma}_m^2$ can be interpreted as the effective noise variance for the m^{th} decoder,

$$\widetilde{\sigma}_m^2 = \frac{1}{\lambda_{\text{MIN}}^2} |\mathbf{v}_{\text{MIN}}^T \mathbf{a}_m|^2 + \frac{1}{\lambda_{\text{MAX}}^2} |\mathbf{v}_{\text{MAX}}^T \mathbf{a}_m|^2. \tag{2.71}$$

Since $\frac{1}{\lambda_{\text{MIN}}} \geq \frac{1}{\lambda_{\text{MAX}}}$, (2.71) suggests that $\mathbf{a}_1, \mathbf{a}_2$ should be chosen in the direction of $\mathbf{v}_{\text{MAX}}$ subject to linearly independent constraints to reduce the noise amplification by $\frac{1}{\lambda_{\text{MIN}}}$. In the case of the decorrelator (or MMSE receiver), the equation coefficients are fixed to be $\mathbf{a}_1 = [1 \;\; 0]^T$ and $\mathbf{a}_2 = [0 \;\; 1]^T$. As a result, the noise variance in at least one of the streams will be heavily amplified by $\frac{1}{\lambda_{\text{MIN}}}$ and the rate will be limited by the minimum singular value of the channel matrix. With integer-forcing, we are free to choose *any linearly independent* $\mathbf{a}_1, \mathbf{a}_2$ since we only require that our coefficients matrix $\mathbf{A}$ be invertible. By choosing $\mathbf{a}_1, \mathbf{a}_2$ in the direction $\mathbf{v}_{\text{MAX}}$, we are protected against large noise amplification in the case of near-singular channel matrices.

2.3.4 Complexity

Our architecture has the same implementation complexity as that of a traditional linear receiver with the addition of the matrix search for $\mathbf{A}$. The ideal joint ML receiver aggregates the time and space dimensions and finds the ML estimate across both. As a result, its complexity is exponential in the product of the blocklength and the number of data streams. Our architecture decouples the time and space dimensions by allowing for single-stream decoding. First, we search for the best integer matrix $\mathbf{A}$, which has an exponential complexity in the number of data streams in the worst case. For slow fading channels, this search is only needed once per data frame. Afterwards, our receiver recovers M linearly independent equations of codewords according to $\mathbf{A}$ and then solves these for the original codewords. This step is polynomial in the number of data streams and exponential in the blocklength for an ML decoder. In practice, the decoding step can be considerably accelerated through the use of LDPC codes and the integer matrix search can be sped up via a sphere decoder.

2.4 Fixed Channel Matrices

In this section, we compare the performance of the integer-forcing linear receiver against existing architectures through a series of examples. In Example 1, we compare the performance of different architectures for an ill-conditioned channel matrix and demonstrate that the choice of equation coefficients for the integer-forcing receiver changes with SNR. In Example 2, we compare the performance of the integer-forcing receiver with the decorrelator and show that the decorrelator can perform arbitrarily worse. In Examples 3, we illustrate that the gap between the integer-forcing receiver and lattice reduction can become unbounded. Finally, we show that the gap between the integer-forcing receiver and the joint decoder can be arbitrarily large in Example 4.

2.4.1 Example 1

Consider the 2×2 real MIMO channel with channel matrix

$$\mathbf{H} = \begin{bmatrix} 0.7 & 1.3 \\ 0.8 & 1.5 \end{bmatrix}. \tag{2.72}$$

Figure 2.5 shows the performance of the different architectures. (Recall that we assume equal-rate data streams on both transmit antennas, as in Definition 2.3.) The achievable rates for traditional linear receiver are given by (2.12) and that of the joint receiver is given by (2.8). The decorrelator and the MMSE receiver aim to separate the data streams and cancel the interference from other streams. However, this is difficult since the columns of the channel matrix are far from orthogonal. The integer-forcing architecture attempts to exploit the interference by decoding two linearly independent equations in the direction of

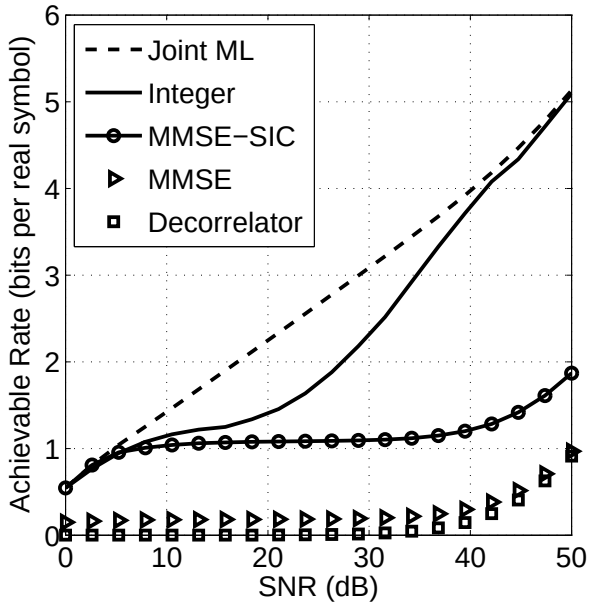

Figure 2.5. Achievable rates for the 2×2 real-valued MIMO channel with fixed channel matrix from (2.72).

the maximum eigenvector $\mathbf{v}_{\text{MAX}} = [0.47 \quad 0.88]^T$. For example, at $\mathsf{SNR} = 30dB$, we choose equation coefficients $\mathbf{a}_1 = [1 \quad 2]^T$ and $\mathbf{a}_2 = [6 \quad 11]^T$, while for $\mathsf{SNR} = 40dB$, we choose equation coefficients $\mathbf{a}_1 = [1 \quad 7]^T$ and $\mathbf{a}_2 = [2 \quad 13]^T$. Thus, for different values of SNR, the optimal equation coefficients generally change.

2.4.2 Example 2: Integer-forcing vs. decorrelator

Consider the 2×2 real MIMO channel with channel matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1+\sqrt{\epsilon} \\ 0 & \epsilon \end{bmatrix} \tag{2.73}$$

where we assume $0 < \epsilon \ll 1$, $\frac{1}{\sqrt{\epsilon}}$ is an integer and $\mathsf{SNR} \gg 1$. We first note that

$$\mathbf{H}^{-1} = \frac{1}{\epsilon} \begin{bmatrix} \epsilon & -(1+\sqrt{\epsilon}) \\ 0 & 1 \end{bmatrix}, \tag{2.74}$$

Using (2.12) with $\mathbf{B} = \mathbf{H}^{-1}$, the achievable rate of the decorrelator is

$$R_{\text{DECORR}} = 2 \min \left\{ \frac{1}{2} \log \left(1 + \frac{\epsilon^2 \mathsf{SNR}}{\epsilon^2 + \epsilon + 2\sqrt{\epsilon} + 1} \right), \frac{1}{2} \log \left(1 + \epsilon^2 \mathsf{SNR} \right) \right\} \tag{2.75}$$

$$\leq \log \left(1 + \epsilon^2 \mathsf{SNR} \right) \tag{2.76}$$

We compare the achievable rate of the decorrelator with the exact integer-forcing rate from Corollary 2.11. The equation coefficients selected by the decoders are

$$\mathbf{a}_1 = [1 \;\; 1]^T \tag{2.77}$$

$$\mathbf{a}_2 = \left[\frac{1}{\sqrt{\epsilon}} \;\; \frac{1}{\sqrt{\epsilon}} + 1\right]^T . \tag{2.78}$$

Using Corollary 2.11, the achievable rate of exact integer-forcing with equations coefficients $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]^T$ is

$$R_{\text{INTEGER}} = 2 \min_{m=1,2} \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{\|(\mathbf{H}^T)^{-1}\mathbf{a}_m\|^2} \right). \tag{2.79}$$

$$= 2 \min \left\{ \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{1 + \frac{1}{\epsilon}} \right), \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{\frac{1}{\epsilon}} \right) \right\} \tag{2.80}$$

$$= \log \left(\frac{\mathsf{SNR}}{1 + \frac{1}{\epsilon}} \right) \tag{2.81}$$

$$\geq \log \left(\frac{\mathsf{SNR}}{\frac{2}{\epsilon}} \right) \tag{2.82}$$

$$= \log \left(\frac{\epsilon \mathsf{SNR}}{2} \right) \tag{2.83}$$

where the inequality follows since $0 < \epsilon \ll 1$.

We compare the two linear architectures to the joint ML decoder whose achievable rate is given by (2.8). For $0 < \epsilon \ll 1$ and $\mathsf{SNR} \gg 1$, the rate of the joint decoder is

$$R_{\text{JOINT}} = \frac{1}{2} \log \det \left(\mathbf{I} + \mathbf{H}\mathbf{H}^T \mathsf{SNR} \right) \tag{2.84}$$

$$= \frac{1}{2} \log \left((1 + \mathsf{SNR})(1 + \epsilon^2 \mathsf{SNR}) + \left(1 + \sqrt{\epsilon}\right)^2 \mathsf{SNR} \right) \tag{2.85}$$

Finally, let us compare the three rates in the setting where $\mathsf{SNR} \to \infty$, and where the parameter ϵ in our channel model tends to zero according[5] to $\epsilon \sim \frac{1}{\sqrt{\mathsf{SNR}}}$. In that special case, we can observe that

$$R_{\text{DECORR}} \sim 1 \tag{2.86}$$

$$R_{\text{INTEGER}} \sim \frac{1}{2} \log(\mathsf{SNR}) \tag{2.87}$$

$$R_{\text{JOINT}} \sim \frac{1}{2} \log(\mathsf{SNR}) \tag{2.88}$$

Hence, the loss from using the decorrelator instead of the integer-forcing receiver becomes unbounded in this regime as $\mathsf{SNR} \to \infty$. Furthermore, the integer-forcing receiver achieves the same scaling as the joint decoder.

[5] Recall that $f(\mathsf{SNR}) \sim g(\mathsf{SNR})$ implies that $\lim_{\mathsf{SNR}\to\infty} \frac{f(\mathsf{SNR})}{g(\mathsf{SNR})} = 1$.

2.4.3 Example 3: Integer-forcing vs. lattice-reduction

In this example, we illustrate the difference between the proposed integer-forcing architecture and the lattice-reduction receiver. First, we note that, unlike the integer-forcing receiver, the lattice-reduction receiver is not required to use a lattice code but it should use a constellation with regular spacing, such as PAM or QAM. However, the key difference is that the effective channel matrix for lattice-reduction receivers is restricted to be unimodular[6] while the effective channel for integer-forcing receivers can be any full-rank integer matrix. In this example, we show that this restriction can result in an arbitrarily large performance penalty. We consider the $M \times M$ MIMO channel with channel matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & -1 & -1 & \cdots & -1 & 2 \end{bmatrix} \tag{2.89}$$

A simple calculation shows that the inverse of this channel matrix is given by

$$\mathbf{H}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{2.90}$$

Since $\mathbf{H}^{-1}$ has non-integer entries, $\mathbf{H}$ is not unimodular. The coefficient matrix that maximizes the achievable rate for the exact integer-forcing receiver from Corollary 2.11 is

$$\mathbf{A}_{\text{INTEGER}} = \mathbf{H} \ ,$$

leading to an effective noise variance in each stream that satisfies

$$\sigma^2_{\text{INTEGER}} = 1 \ . \tag{2.91}$$

By contrast, following the lattice-reduction receiver, we must ensure that the resulting effective channel matrix is unimodular. Using the fact that $\left(\mathbf{H}^T\right)^{-1}$ is a basis for the body-centered cubic (BCC) lattice, it can be shown that the best choice of matrix is

$$\mathbf{A}_{\text{UNIMODULAR}} = \mathbf{I} \ . \tag{2.92}$$

[6] Recall that a matrix is *unimodular* if has integer entries and its inverse has integer entries.

It follows that the effective noise variance in the worst stream is given by

$$\sigma^2_{\text{UNIMODULAR}} = \min_m \| \left(\mathbf{H}^T\right)^{-1} \mathbf{a}_{\text{UNIMODULAR},m} \|^2 \tag{2.93}$$

$$= \max\{M/4, 1\} \ . \tag{2.94}$$

Hence, as the number of antennas becomes large ($M \to \infty$), restricting the effective matrix to be unimodular results in an arbitrarily large loss.

2.4.4 Example 4: Integer-forcing vs. joint ML decoder

Finally, we illustrate the point that the integer-forcing receiver can sometimes be arbitrarily worse than optimal joint decoding. To see this, we consider a 2×2 MIMO channel with the channel matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 0 & \epsilon \end{bmatrix} \tag{2.95}$$

where $0 < \epsilon < 1$. The rate attainable via joint ML decoding is

$$R_{\text{JOINT}} = \log((1 + 2\mathsf{SNR})(1 + \epsilon^2\mathsf{SNR}) - \epsilon^2\mathsf{SNR}^2) \tag{2.96}$$

$$\geq \log(1 + 2\mathsf{SNR}) \ . \tag{2.97}$$

We note that the inverse of the channel matrix is given by

$$\mathbf{H}^{-1} = \begin{bmatrix} 1 & -\frac{1}{\epsilon} \\ 0 & \frac{1}{\epsilon} \end{bmatrix} \ . \tag{2.98}$$

We bound the achievable rate of exact integer-forcing from Corollary 2.11 as follows

$$R_{\text{INTEGER}} = 2 \max_{\mathbf{A}:|\mathbf{A}|\neq 0} \min_{m=1,2} \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{\| \left(\mathbf{H}^T\right)^{-1} \mathbf{a}_m \|^2} \right) \tag{2.99}$$

$$= 2 \max_{\mathbf{A}:|\mathbf{A}|\neq 0} \min_{m=1,2} \frac{1}{2} \log \left(\frac{\mathsf{SNR}}{a_{m,1}^2 + (a_{m,2} - a_{m,1})^2 \frac{1}{\epsilon^2}} \right) \tag{2.100}$$

$$\leq \max_{a_{m,2} \neq a_{m,1}} \log \left(\frac{\mathsf{SNR}}{a_{m,1}^2 + (a_{m,2} - a_{m,1})^2 \frac{1}{\epsilon^2}} \right) \tag{2.101}$$

$$\leq \log\left(\epsilon^2 \mathsf{SNR}\right) \ . \tag{2.102}$$

Let $\epsilon \sim \frac{1}{\sqrt{\mathsf{SNR}}}$ and consider the regime $\mathsf{SNR} \to \infty$. The gap between the optimal joint receiver and the integer-forcing linear receiver can be arbitrarily large. However, as we will see in Section 2.5, the average behavior of the integer-forcing linear receiver is close to the joint decoder under a Rayleigh fading distribution for medium to high SNR.

2.5 Performance for Slow Fading Channels

2.5.1 Model and Definitions

We now demonstrate that integer-forcing receiver nearly matches the performance of the joint ML decoder under a slow fading channel model. Since the integer-forcing receiver can mimic the operation performed by a zero-forcing or MMSE receiver (as well as decode messages via equations), it is not surprising that it offers higher rates. However, these architectures are often coupled with some form of SIC. We will show that the integer-forcing receiver outperforms the following standard SIC architectures:

- V-BLAST I: The receiver decodes and cancels the data streams in a predetermined order, irrespective of the channel realization. Each data stream has the same rate. See (2.14) for the rate expression.
- V-BLAST II: The receiver selects the decoding order separately for each channel realization in such a way as to maximize the effective SNR for the data stream that sees the worst channel. Each data stream has the same rate. See (2.15) for the rate expression.
- V-BLAST III: The receiver decodes and cancels the data streams in a predetermined order. The rate of each data stream is selected to maximize the sum rate. The rate expression is given in Section 2.5.2.

In Sections 2.5.3 and 2.5.4, we compare these schemes through simulations as well as their diversity-multiplexing tradeoffs. For completeness, we also compare integer-forcing to an SIC architecture that allows for both variable decoding order and unequal rate allocation in Appendix A.

We adopt the standard quasi-static Rayleigh fading model where each element of the complex channel matrix is independent and identically distributed according to a circularly symmetric complex normal distribution of unit variance. The transmitter is only aware of the channel statistics while the receiver knows the exact channel realization. As a result, we will have to cope with some outage probability p_{OUT}.

Definition 2.18. Assume there exists an architecture that encodes each data stream at the same rate and achieves sum rate $R(\mathbf{H})$. For a target rate R, then the outage probability is defined as

$$p_{\text{OUT}}(R) = \Pr(R(\mathbf{H}) < R). \tag{2.103}$$

For a fixed probability $p \in (0, 1]$, we define the *outage rate* to be

$$R_{\text{OUT}}(p) = \sup\{R : p_{\text{OUT}}(R) \leq p\}. \tag{2.104}$$

2.5.2 Rate Allocation

We have assumed that each data stream is encoded at the same rate. This is optimal for linear receivers under isotropic fading. However, when SIC is used, rate allocation can be beneficial in an outage scenario. To compare the performance of integer-forcing to SIC with rate allocation, we consider V-BLAST III in this section. V-BLAST III performs SIC with a fixed decoding order and allows for rate allocation among the different data streams. Without loss of generality for Rayleigh fading, if we fix a decoding order, we may take it to be $\pi = (1, 2, \ldots, 2M)$. From (2.13), the achievable rate for stream m is

$$R_{\pi(m)}(\mathbf{H}) = \frac{1}{2} \log \left(1 + \frac{\mathsf{SNR} \|\mathbf{b}_m^T \mathbf{h}_{\pi(m)}\|^2}{\|\mathbf{b}_m\|^2 + \mathsf{SNR} \sum_{i>m} \|\mathbf{b}_m^T \mathbf{h}_{\pi(i)}\|^2} \right). \tag{2.105}$$

Since the streams are decoded in order, the later streams will achieve higher rates on average than the earlier streams. V-BLAST III allocates lower rates to earlier streams and higher rates to later streams. We now generalize Definition 2.18 to include rate allocation.

Definition 2.19. Assume an architecture that achieves rate $R_m(\mathbf{H})$ in stream m. For a target rate R, the outage probability is given by:

$$p_{\text{OUT}}(R) = \min_{\substack{R_1,\ldots,R_{2M} \\ \sum_{m=1}^{2M} R_m \le R}} \Pr \left(\bigcup_{m=1}^{2M} \{R_m(\mathbf{H}) < R_m\} \right). \tag{2.106}$$

For a fixed probability $p \in (0, 1]$, we define the *outage rate* to be:

$$R_{\text{OUT}}(p) = \sup \{R : p_{\text{OUT}}(R) \le p\}. \tag{2.107}$$

2.5.3 Outage Behavior

We now compare the outage rate and probabilities for the receiver architectures discussed above. It is easy to see that the zero-forcing receiver performs strictly worse than the MMSE receiver and V-BLAST I performs strictly worse than V-BLAST II. We have chosen to omit these two architectures from our plots to avoid overcrowding. Figure 2.6 shows the 1 percent outage rate and Figure 2.7 shows the 5 percent outage rate. In both cases, the integer-forcing receiver nearly matches the rate of the joint ML receiver while the MMSE receiver achieves significantly lower performance. The SIC architectures with either an optimal decoding order, V-BLAST II, or an optimized rate allocation, V-BLAST III, improve the performance of the MMSE receiver significantly but still achieve lower rates than the integer-forcing receiver from medium SNR onwards. Our simulations suggest that the outage rate of the integer-forcing receiver remains within a small gap from the outage rate of the joint ML receiver. However, we recall from the example given in Subsection 2.4.4 that it is not true that the integer-forcing receiver is uniformly near-optimal for all fading realizations. Figure

2.8 shows the outage probability for the target sum rate $R = 6$. We note that the integer-forcing receiver achieves the same slope as the joint decoder. In the next subsection, we characterize the diversity-multiplexing tradeoff of the integer-forcing receiver and compare it with the diversity-multiplexing tradeoff of traditional architectures that are considered in [54]. We show that the integer-forcing receiver attains the optimal diversity multiplexing-tradeoff in the case where each transmit antenna sends an independent data stream.

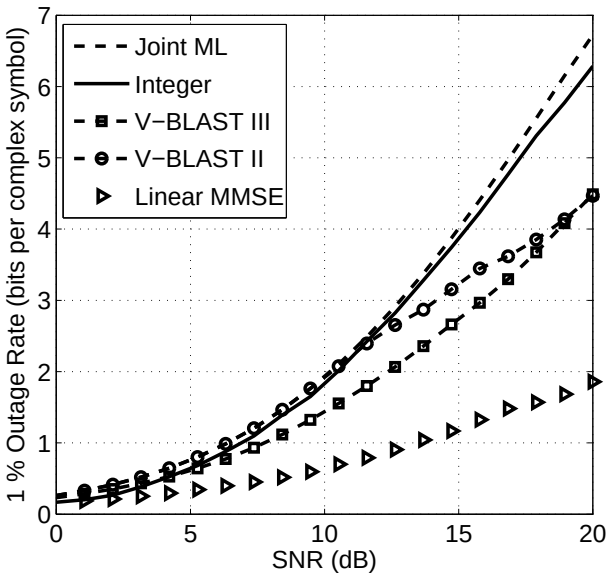

Figure 2.6. 1 percent outage rates for the 2×2 complex-valued MIMO channel with Rayleigh fading.

2.5.4 Diversity-Multiplexing Tradeoff

The diversity-multiplexing tradeoff (DMT) provides a rough characterization of the performance of a MIMO transmission scheme at high SNR [54].

Definition 2.20. A family of codes is said to achieve spatial multiplexing gain r and diversity gain d if the total data rate and the average probability of error satisfy

$$\lim_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} \geq r \tag{2.108}$$

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log P_e(\mathsf{SNR})}{\log \mathsf{SNR}} \leq -d. \tag{2.109}$$

In the case where each transmit antenna encodes an independent data stream[7], the

[7] If joint encoding across the antennas is permitted, then a better DMT is achievable. See [54] for more details.

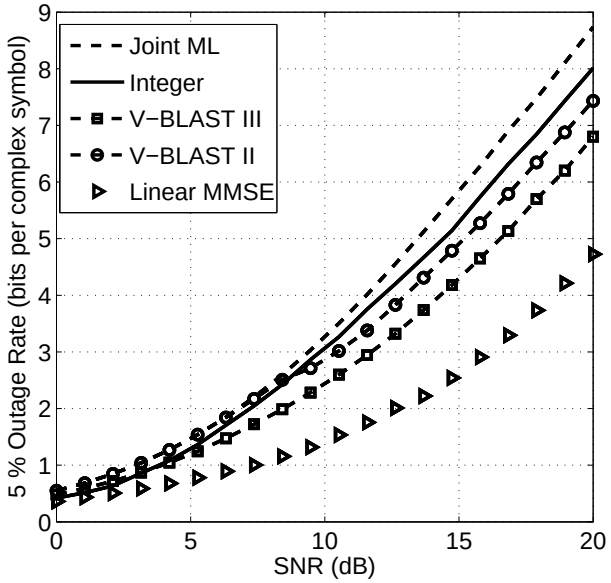

Figure 2.7. 5 percent outage rates for the 2×2 complex-valued MIMO channel with Rayleigh fading.

optimal DMT is

$$d_{\text{JOINT}}(r) = N\left(1 - \frac{r}{M}\right) \tag{2.110}$$

where $r \in [0, M]$ and can be achieved by joint ML decoding [54]. The DMTs achieved by the decorrelator and SIC architectures are as follows [54]:

$$d_{\text{DECORR}}(r) = \left(1 - \frac{r}{M}\right) \tag{2.111}$$

$$d_{\text{V-BLAST I}}(r) = \left(1 - \frac{r}{M}\right) \tag{2.112}$$

$$d_{\text{V-BLAST II}}(r) \leq (N-1)\left(1 - \frac{r}{M}\right) \tag{2.113}$$

$$d_{\text{V-BLAST III}}(r) = \text{piecewise linear curve connecting points } (r_k, n-k) \tag{2.114}$$

$$\text{where } r_0 = 0, r_k = \sum_{i=0}^{k-1} \frac{k-i}{n-i} \quad 1 \leq k \leq n$$

The decorrelator chooses the matrix $\mathbf{B}$ to cancel the interference from the other data streams. As a result, the noise is heavily amplified when the channel matrix is near singular and the performance is limited by the minimum singular value of the channel matrix. In the integer-forcing linear receiver, the effective channel matrix $\mathbf{A}$ is not limited to be the identity matrix but can be any full-rank integer matrix. This additional freedom is sufficient to recover the same DMT as the joint ML decoder.

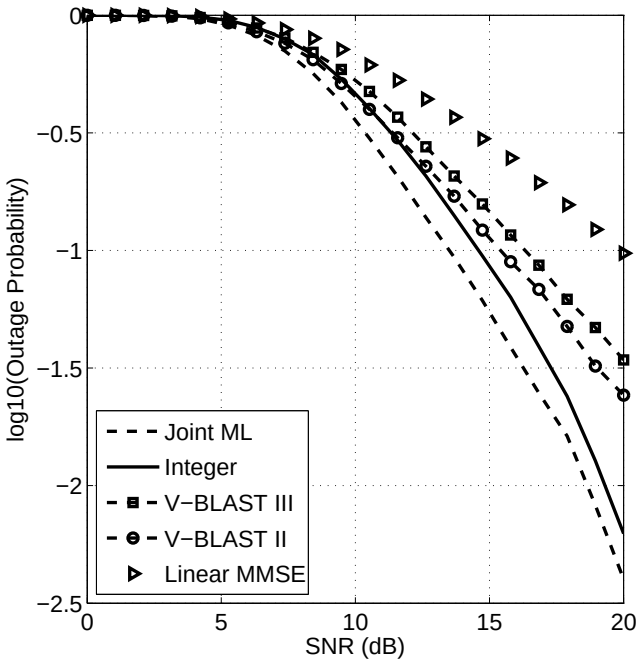

Figure 2.8. Outage probability for the 2×2 complex-valued MIMO channel with Rayleigh fading for a target sum rate of $R = 6$.

Theorem 2.21. *For a MIMO channel with M transmit, $N \geq M$ receive antennas, and Rayleigh fading, the achievable diversity-multiplexing tradeoff for the integer-forcing receiver is given by*

$$d_{\mathit{INTEGER}}(r) = N\left(1 - \frac{r}{M}\right) \tag{2.115}$$

where $r \in [0, M]$.

The proof of Theorem 2.21 is given in Appendix B. Figure 2.9 illustrates the DMT for a 4×4 MIMO channel. The integer-forcing receiver achieves a maximum diversity of 4 while the decorrelator and V-BLAST I achieve can only achieve a diversity of 1 since their performance is limited by the worst data stream. V-BLAST II achieves a higher DMT than V-BLAST I but a lower diversity than the integer-forcing receiver since its rate is still limited by the worst stream after the optimal decoding order is applied. V-BLAST III achieves the optimal diversity at the point $r = 0$ since only one data stream is used in transmission. For values of $r > 0$, the achievable diversity is suboptimal.

2.5.5 Discussion

As noted earlier, receiver architectures based on zero-forcing face a rate penalty when the channel matrix is ill-conditioned. Integer-forcing circumvents this issue by allowing the

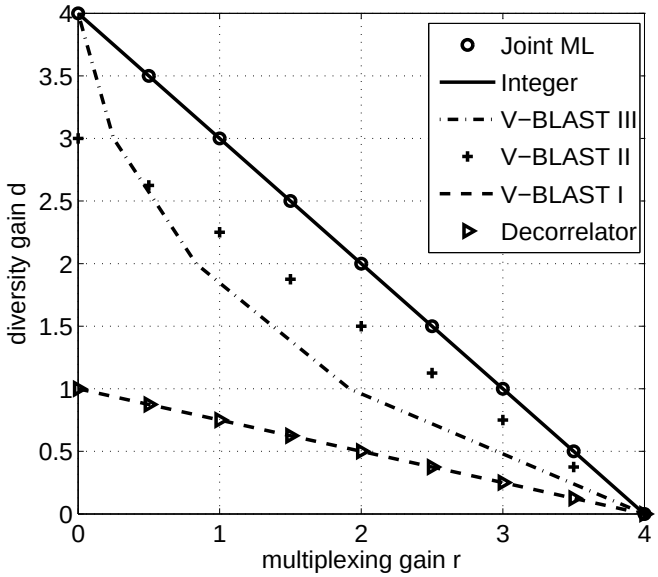

Figure 2.9. Diversity-multiplexing tradeoff for the 4×4 MIMO channel with Rayleigh fading.

receiver to first decode equations whose coefficients are matched with those of the channel. From one perspective, the resulting gains are of a similar nature as those obtained by lattice-reduction receivers. One important difference is that integer-forcing applies a modulo operation at the receiver prior to decoding, which retains the linear structure of the codebook. This allows us to derive closed-form rate expressions analogous to those for traditional linear receiver. Typically, lattice reduction is used at the symbol level followed by a decoding step [55]. While this form of lattice reduction can be used to obtain the full receive diversity [56], it does not seem to suffice in terms of rate.

Another key advantage of integer-forcing is that it completely decouples the spatial aspect of decoding from the temporal aspect. That is, the search for the best integer matrix $\mathbf{A}$ to approximate the channel matrix $\mathbf{H}$ is completed before we attempt to decode the integer combinations of codewords. Thus, apart from the search[8] for the best $\mathbf{A}$, which in a slow-fading environment does not have to be executed frequently, the complexity of the integer-forcing receiver is similar to that of the zero-forcing receiver.

From our outage plots, it is clear that the integer-forcing receiver significantly outperforms the basic MMSE receiver. Moreover, integer-forcing beats more sophisticated SIC-based V-BLAST architectures, even when these are permitted to optimize their rate allocation while integer-forcing is not. We note that it is possible to develop integer-forcing schemes that permit unequal rate allocations [31] as well as a form of interference cancellation [60] but

[8]This search can be considerably sped up in practice through the use of a sphere decoding algorithm.

this is beyond the scope of the present chapter.

Integer-forcing also attains the full diversity-multiplexing tradeoff, unlike the V-BLAST architectures discussed above. Earlier work developed lattice-based schemes that attain the full DMT [57, 58] but, to the best of our knowledge, ours is the first that decouples spatial decoding from temporal decoding. The caveat is that the DMT result presented in the current chapter only applies if there is no spatial coding across transmit antennas, whereas the DMT results of [57, 58] apply in general. Characterizing the DMT of integer-forcing when there is coding across transmit antennas is an interesting subject for future study.

Chapter 3

Mitigating Interference with IF Receivers

We have studied the performance of the integer-forcing (IF) linear receiver under the standard MIMO channel and found that it achieves outage rates close those of the joint ML decoder as well as the same DMT. In this section, we show that integer-forcing architectures are also successful at dealing with a different kind of channel disturbance, namely interference. We assume that the interfering signal is low-dimensional (compared to the number of receive antennas), and we are most interested in the case where the variance of this interfering signal increases (at a certain rate) with the transmit power. We show that the integer-forcing architecture can be used to perform "oblivious" interference mitigation. By oblivious, we mean that the transmitter and receiver are unaware of the codebook of the interferer (if there is one). However, the receiver knows which subspace is occupied by the interference. By selecting equation coefficients in a direction that depends both on the interference space and on the channel matrix, the integer-forcing receiver reduces the impact of interference and attains a significant gain over traditional linear receivers. We will characterize the generalized degrees-of-freedom show that it matches that of the joint ML decoder.

Remark 3.1. Oblivious receivers have been thoroughly studied in the context of cellular systems [61] and distributed MIMO [62].

3.1 Problem Definition

For ease of notation and tractability, the discussion presented in this section is limited to channels whose channel matrix is square, i.e., with equal number of transmit and receive antennas. Recall that the real-valued representation of the $M \times M$ complex-valued MIMO

channel (see Definition 2.5) is given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{3.1}$$

where $\mathbf{Y} \in \mathbb{R}^{2M \times n}$ is the channel output, $\mathbf{H} \in \mathbb{R}^{2M \times 2M}$ is the real-valued representation of the fading matrix, $\mathbf{X}^{2M \times n}$ is the channel input, and the noise $\mathbf{Z} \in \mathbb{R}^{2M \times n}$ has i.i.d. Gaussian entries with unit variance. In this section, we extend the standard MIMO channel to include the case of interference. The generalized model has channel output

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{J}\mathbf{V} + \mathbf{Z} \tag{3.2}$$

where $\mathbf{H}$ is the channel matrix, $\mathbf{X}$ is the channel input, and $\mathbf{Z}$ is the noise, all as in the previous model. An external interferer adds $\mathbf{V} \in \mathbb{R}^{2K \times n}$ in the direction represented by the column space of $\mathbf{J} \in \mathbb{R}^{2M \times 2K}$. We assume that each element of $\mathbf{V}$ is i.i.d. Gaussian with variance INR. We assume that $\mathbf{J}$ is fixed during the whole transmission block and known only to the receiver.

The definition for messages, rates, encoders, and decoders follow along similar lines as those for the standard MIMO channel (see Definitions 2.1, 5.4, 5.5, and 2.3 in Section 2.1).

3.2 Traditional Linear Receivers

As in the case without interference, traditional linear receivers process the channel output $\mathbf{Y}$ by multiplying it by a $2M \times 2M$ matrix $\mathbf{B}$ to arrive at the effective output

$$\widetilde{\mathbf{Y}} = \mathbf{B}\mathbf{Y} \tag{3.3}$$

and recover the message $\mathbf{w}_m$ using only the m^{th} row of the matrix $\widetilde{\mathbf{Y}}$. By analogy to (2.12), the achievable sum rate can be expressed as

$$R_{\text{LINEAR}}(\mathbf{H}, \mathbf{J}, \mathbf{B}) = \min_m 2M R_m(\mathbf{H}, \mathbf{J}, \mathbf{B}). \tag{3.4}$$

where $R_m(\mathbf{H}, \mathbf{J}, \mathbf{B})$ represents the achievable rate for the m^{th} data stream (using Gaussian codebooks),

$$R_m(\mathbf{H}, \mathbf{J}, \mathbf{B}) = \frac{1}{2} \log \left(1 + \frac{\mathsf{SNR} \|\mathbf{b}_m^T \mathbf{h}_m\|^2}{\|\mathbf{b}_m\|^2 + \mathsf{INR} \|\mathbf{J}^T \mathbf{b}_m\|^2 + \mathsf{SNR} \sum_{i \neq m} \|\mathbf{b}_m^T \mathbf{h}_i\|^2} \right),$$

Again, let us discuss several choices of the matrix $\mathbf{B}$. The decorrelator, given by $\mathbf{B} = \mathbf{H}^{-1}$, removes the interference due to other data streams but does not cancel the external interference $\mathbf{J}$ (except in the very special case where the subspace spanned by $\mathbf{J}$ is orthogonal to the subspace spanned by $\mathbf{H}^{-1}$). Alternatively, if we choose $\mathbf{B} = \mathbf{J}^{\perp}$, where $\mathbf{J}^{\perp}$ is the $2K \times 2M$ matrix whose rowspace is orthogonal to the columnspace of $\mathbf{J}$, then the external

interference term is indeed nulled. The resulting output $\mathbf{J}^{\perp}\mathbf{Y}$ can then be processed by a traditional linear receiver. This scheme achieves good performance in high INR regimes but does not perform well in high SNR regimes since the interference due to the other data streams is mostly unresolved. The MMSE receiver improves the performance of the both architectures by choosing $\mathbf{B} = \mathbf{H}\left(\frac{1}{\mathsf{SNR}}\mathbf{I} + \frac{\mathsf{INR}}{\mathsf{SNR}}\mathbf{J}\mathbf{J}^T + \mathbf{H}\mathbf{H}^T\right)^{-1}$. However, since there are $2M$ data streams and the interference is of dimension $2K$, it is impossible to cancel both the interference from other data streams and the external interference with any matrix $\mathbf{B}$. One way out of this conundrum is to reduce the number of transmitted streams to $2M-2K$. For this scenario, the MMSE receiver can be applied to mitigate both the external interference and the interference from other data streams.

Complexity permitting, we can again improve performance by resorting to successive interference cancellation architectures. The achievable rate for V-BLAST I in the standard MIMO channel from (2.14) becomes

$$R_{\mathrm{SIC},1}(\mathbf{H}) = \min_{m} 2M R_{\pi(m)}(\mathbf{H}). \tag{3.5}$$

where

$$R_{\pi(m)}(\mathbf{H}) = \frac{1}{2}\log\left(1 + \frac{\mathsf{SNR}\|\mathbf{b}_m^T\mathbf{h}_{\pi(m)}\|^2}{\|\mathbf{b}_m\|^2 + \mathsf{INR}\|\mathbf{J}^T\mathbf{b}_m\|^2 + \mathsf{SNR}\sum_{i>m}\|\mathbf{b}_m^T\mathbf{h}_{\pi(i)}\|^2}\right) \tag{3.6}$$

and $\mathbf{b}_m = (\frac{1}{\mathsf{SNR}}\mathbf{I} + \mathbf{J}\mathbf{J}^T\frac{\mathsf{INR}}{\mathsf{SNR}} + \mathbf{H}_{\pi_m}\mathbf{H}_{\pi_m}^T)^{-1}\mathbf{h}_{\pi(m)}$. The achievable rate for V-BLAST II follows by maximizing the rate in (3.5) over all decoding orders $\pi \in \Pi$.

3.3 Integer-Forcing Linear Receiver

We apply the integer-forcing linear receiver proposed in Section 2.3 to the problem of mitigating interference (see Figure 3.1). The channel output matrix $\mathbf{Y}$ is first multiplied by a fixed matrix $\mathbf{B}$ to form the matrix $\tilde{\mathbf{Y}}$ whose m^{th} row is the signal fed into the m^{th} decoder. Each such row can be expressed as

$$\tilde{\mathbf{y}}_m^T = \sum_{i=1}^{2M}(\mathbf{b}_m^T\mathbf{h}_i)\mathbf{x}_i^T + \mathbf{b}_m^T\mathbf{J}\mathbf{V} + \mathbf{b}_m^T\mathbf{Z} \tag{3.7}$$

$$= \sum_{i=1}^{2M}\tilde{\mathbf{h}}_m^T\mathbf{X} + \tilde{\mathbf{v}}_m^T + \tilde{\mathbf{z}}_m^T \tag{3.8}$$

where $\tilde{\mathbf{h}}_m = \mathbf{H}^T\mathbf{b}_m$ is the effective channel to the m^{th} decoder, $\tilde{\mathbf{v}}_m$ is the effective interference with variance $\|\mathbf{J}^T\mathbf{b}_m\|^2\mathsf{INR}$, and $\tilde{\mathbf{z}}_m$ is the effective noise with variance $\|\mathbf{b}_m\|^2$. The next theorem and its following remarks generalize Theorem 2.10, Corollary 2.11, and Corollary 2.12 to include the case with interference.

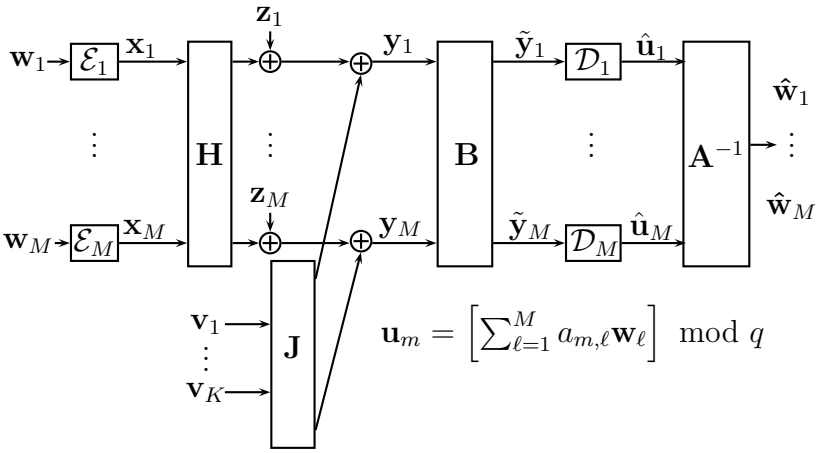

Figure 3.1. Integer-forcing linear receiver for the complex-valued $M \times M$ MIMO channel with external inference of dimension K.

Theorem 3.2. *Consider the MIMO channel with channel matrix $\mathbf{H} \in \mathbb{R}^{2M\times 2M}$ and interference matrix $\mathbf{J} \in \mathbb{R}^{2M\times 2K}$. For any full-rank integer matrix $\mathbf{A} \in \mathbb{Z}^{2M\times 2M}$ and any $2M \times 2M$ matrix $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_{2M}]^T$, the following sum rate is achievable using the integer-forcing linear receiver:*

$$R(\mathbf{H}, \mathbf{J}, \mathbf{A}, \mathbf{B}) < \min_m 2M R(\mathbf{H}, \mathbf{J}, \mathbf{a}_m, \mathbf{b}_m) \tag{3.9}$$

where $R(\mathbf{H}, \mathbf{J}, \mathbf{a}_m, \mathbf{b}_m)$ is given by

$$= \frac{1}{2}\log\left(\frac{\mathsf{SNR}}{\|\mathbf{b}_m\|^2 + \|\mathbf{J}^T\mathbf{b}_m\|^2\mathsf{INR} + \|\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m\|^2\mathsf{SNR}}\right)$$

Remark 3.3. Exact integer-forcing selects $\mathbf{B} = \mathbf{A}\mathbf{H}^{-1}$. The achievable rate can be expressed more concisely as

$$R < \min_m M \log\left(\frac{\mathsf{SNR}}{\|(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2 + \|\mathbf{J}^T(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2\mathsf{INR}}\right). \tag{3.10}$$

Remark 3.4. The optimal projection matrix that maximizes the achievable rate in Theorem 3.2 is given by

$$\mathbf{B}_{\text{OPT}} = \mathbf{A}\mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \mathbf{J}\mathbf{J}^T\frac{\mathsf{INR}}{\mathsf{SNR}} + \mathbf{I}\frac{1}{\mathsf{SNR}}\right)^{-1}. \tag{3.11}$$

3.4 Geometric Interpretation

In the case without interference, the equation coefficients $\mathbf{a}_1, \cdots \mathbf{a}_M$ should be chosen in the direction of the maximum eigenvector of $\mathbf{H}^T\mathbf{H}$ to minimize the effective noise (see Figure

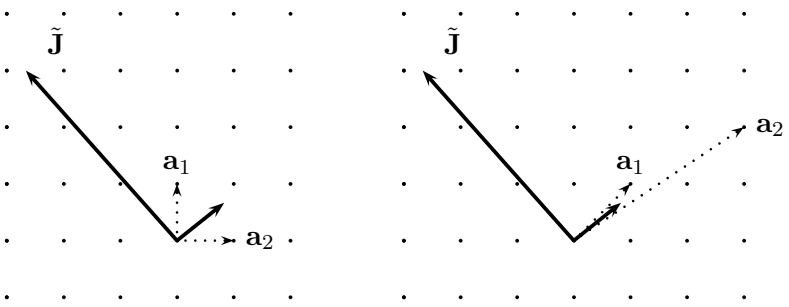

Figure 3.2. The decorrelator (left) fixes the equations to be $\mathbf{a}_1 = [1 \quad 0]^T$ and $\mathbf{a}_2 = [0 \quad 1]^T$. The integer-forcing Linear Receiver (right) allows for any choice of linearly independent equations. Equations should be chosen in the direction orthogonal to $\tilde{\mathbf{J}} = \mathbf{H}^{-1}\mathbf{J}$.

2.4). When the interference is large, the equations coefficients should instead be chosen as close to orthogonal to the effective interference as possible. Consider the (suboptimal) rate expression in (3.10). The "effective" noise variance in the m^{th} stream is

$$\sigma_{\text{EFFEC,m}} = \|(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2 + \|\mathbf{J}^T(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2\mathsf{INR} \; . \tag{3.12}$$

Let λ_{MAX} be the maximum singular value of $\mathbf{H}^{-1}$ and $\tilde{\mathbf{J}} = \mathbf{H}^{-1}\mathbf{J}$. The effective noise variance can be bounded by

$$\sigma_{\text{EFFEC,m}} \leq \lambda_{\text{MAX}}^2\|\mathbf{a}_m\|^2 + \|\tilde{\mathbf{J}}^T\mathbf{a}_m\|^2\mathsf{INR} \; . \tag{3.13}$$

In the high interference regime ($\mathsf{INR} \gg 1$), the equation coefficients should be chosen orthogonal to the direction of the "effective" interference $\tilde{\mathbf{J}}$ to minimize the effective noise variance. This is illustrated in Figure 3.2. In the case of traditional linear receivers, the equation coefficients are fixed to be the unit vectors: $\mathbf{a}_1 = [1 \; 0 \; \cdots 0]^T, \mathbf{a}_2 = [0 \; 1 \; \cdots 0]^T, \ldots, \mathbf{a}_{2M} = [0 \; 0 \; \cdots 1]^T$. As a result, the interference space spanned by $\tilde{\mathbf{J}}$ has significant projections onto at least some of the decoding dimensions $\mathbf{a}_m$. By contrast, in the case of the integer-forcing linear receiver, since $\mathbf{a}_1, \cdots, \mathbf{a}_{2M}$ need only be linearly independent, we can choose all of the decoding dimensions $\mathbf{a}_m$ to be close to orthogonal to $\tilde{\mathbf{J}}$.

3.5 Fixed Channel Example

To illustrate the impact of choosing equation coefficients in a fashion suitable to mitigate external interference, we consider the 2×2 MIMO channel with channel matrix $\mathbf{H}$ and one-dimensional interference space $\mathbf{J}$ given by

$$\mathbf{H} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \mathbf{J} = \frac{1}{3}\begin{bmatrix} L+2 \\ 2L+1 \end{bmatrix} \tag{3.14}$$

where $L \in \mathbb{N}$. In the case of the decorrelator, we invert the channel to arrive at the effective output:

$$\tilde{\mathbf{Y}} = \mathbf{X} + \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} L+2 \\ 2L+1 \end{bmatrix}\mathbf{V} + \tilde{\mathbf{Z}} \tag{3.15}$$

$$= \mathbf{X} + \begin{bmatrix} 1 \\ L \end{bmatrix}\mathbf{V} + \tilde{\mathbf{Z}} \tag{3.16}$$

where $\tilde{\mathbf{Z}} = \mathbf{H}^{-1}\mathbf{Z}$. For $\mathsf{INR} \gg 1$, the effective noise variances scale as

$$\sigma^2_{\text{DECORR},1} \sim \mathsf{INR} \tag{3.17}$$

$$\sigma^2_{\text{DECORR},2} \sim L^2\mathsf{INR} \tag{3.18}$$

Using the integer-forcing linear receiver with the choice of equations $\mathbf{a}_1 = [1\ 0]^T$ and $\mathbf{a}_2 = [-L\ 1]^T$, the effective channel output to the second decoder is

$$\begin{bmatrix} -L & 1 \end{bmatrix}\tilde{\mathbf{Y}} = -L\mathbf{x}_1^T + \mathbf{x}_2^T + \begin{bmatrix} -L & 1 \end{bmatrix}\tilde{\mathbf{Z}} \tag{3.19}$$

where $\mathbf{x}_\ell$ is the codeword sent by the ℓ^{th} antenna. It follows that the effective noise variances are

$$\sigma^2_{\text{INT},1} \sim \mathsf{INR} \tag{3.20}$$

$$\sigma^2_{\text{INT},2} \sim C \tag{3.21}$$

where C is a constant that does not scale with INR. In this example, the integer-forcing linear receiver is able to completely cancel the effect of interference in the second stream by choosing equation coefficients appropriately.

3.6 Generalized Degrees of Freedom

We evaluate the generalized degrees of freedom for the $M \times M$ complex MIMO channel with K-dimensional interference. We specify the interference-to-noise ratio through the parameter α where

$$\alpha = \lim_{\substack{\mathsf{SNR}\to\infty \\ \mathsf{INR}\to\infty}} \frac{\log \mathsf{INR}}{\log \mathsf{SNR}} \tag{3.22}$$

and consider the case where $0 \le \alpha \le 1$. The generalized degrees of freedom are defined as follows (see [63]):

Definition 3.5. (Generalized Degrees-of-Freedom) For a given channel matrix $\mathbf{H}$ and interference matrix $\mathbf{J}$, the generalized degrees-of-freedom of a scheme is

$$d(\mathbf{H}, \mathbf{J}) = \lim_{\substack{\mathsf{SNR}\to\infty \\ \mathsf{INR}=\mathsf{SNR}^\alpha}} \frac{R(\mathsf{SNR}, \mathbf{H}, \mathbf{J})}{\log \mathsf{SNR}} \tag{3.23}$$

where $R(\mathsf{SNR}, \mathbf{H}, \mathbf{J})$ is the achievable sum rate of the scheme.

Definition 3.6. (Rational Independence) We call a matrix $\mathbf{T}^{M\times N}$ rationally independent if for all $\mathbf{q} \in \mathbb{Q}^N \setminus \mathbf{0}$, we have that

$$\mathbf{Tq} \neq \mathbf{0} \tag{3.24}$$

We consider the set of matrices $(\mathbf{H}, \mathbf{J})$ such that $\mathbf{H}^{-1}\mathbf{J}$ is rationally independent. It can be seen that this set has Lebesgue measure one. In the next theorem, we show that for this class of matrices, the integer-forcing linear receiver achieves the same number of generalized degrees of freedom as the joint decoder, and is thus optimal.

Theorem 3.7. *Consider the $M \times M$ complex MIMO channel with K dimensional interference. The integer-forcing linear receiver achieves the generalized degrees of freedom*

$$d_{INT} = M - K\alpha \tag{3.25}$$

for a set of $\mathbf{H}, \mathbf{J}$ that have Lebesgue measure one.

A straightforward derivation shows that the optimal joint decoder with $2M$ streams of data achieves the following generalized degrees of freedom

$$d_{\text{JOINT}} = M - K\alpha \tag{3.26}$$

$$\tag{3.27}$$

for all full-rank channel matrices $\mathbf{H}, \mathbf{J}$. The linear MMSE receiver with $2M$ data streams and the linear MMSE receiver with $2M - 2K$ data streams achieve the following degrees of freedom for all full-rank channel matrices $\mathbf{H}, \mathbf{J}$:

$$d_{\text{MMSE,2M}} = M - M\alpha \tag{3.28}$$

$$d_{\text{MMSE,2M-2K}} = M - K \tag{3.29}$$

When $2M$ data streams are transmitted (on the real-valued representation of the complex MIMO channel), the MMSE receiver does not achieve the optimal number of degrees of freedom since it treats the interference as noise at high SNR while the integer-forcing linear receiver mitigates the interference. When only $2M - 2K$ data streams are transmitted, the linear MMSE receiver can first cancel the interference and then separate the data streams to achieve a degree of freedom of $2M - 2K$. However, this is suboptimal for all regimes $\alpha < 1$ (see Figure 12). A straightforward calculation shows that when the number of transmitted data streams is between $2M - 2K$ and $2M$, the performance is strictly suboptimal in terms of degrees of freedom.

Our proof of Theorem 3.7 uses the following Theorem 4.13 in Chapter 4.

Proof. (Theorem 3.7) To establish this result, we use the (generally suboptimal) choice of the matrix $\mathbf{B}$ that we have referred to as exact integer-forcing, i.e., from (3.10). The achievable rate of this version of the integer-forcing linear receiver is given by

$$R < \max_{\mathbf{A}:|\mathbf{A}|\neq 0} \min_{m} M \log \left(\frac{\mathsf{SNR}}{\|(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2 + \|\mathbf{J}^T(\mathbf{H}^{-1})^T\mathbf{a}_m\|^2\mathsf{INR}} \right) . \tag{3.30}$$

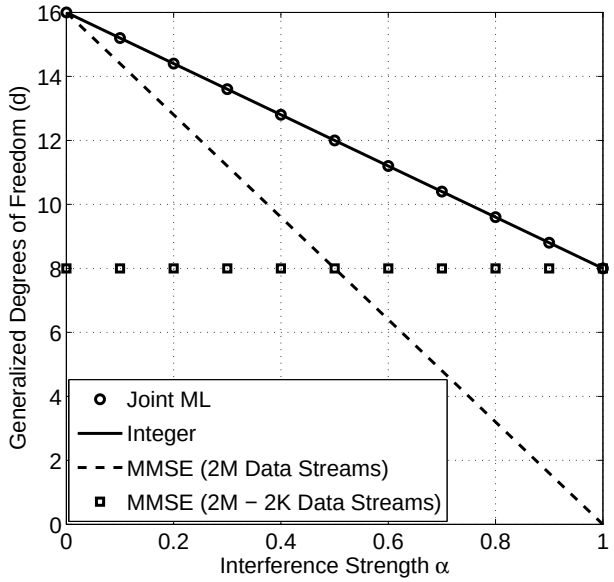

Figure 3.3. Generalized degrees of freedom for the complex-valued 16×16 MIMO channel with 8-dimensional interference ($M = 16, K = 8$).

With $\mathsf{INR} \sim \mathsf{SNR}^{\alpha}$, the effective noise in the worst data stream can be expressed as

$$\sigma^2 = \min_{\mathbf{A}:|\mathbf{A}|\neq 0} \max_{m} \|(\mathbf{H}^{-1})^T \mathbf{a}_m\|^2 + \|\mathbf{J}^T(\mathbf{H}^{-1})^T \mathbf{a}_m\|^2 \mathsf{SNR}^{\alpha} \tag{3.31}$$

$$\leq \min_{\mathbf{A}:|\mathbf{A}|\neq 0} \max_{m} \lambda_{\text{MAX}}^2(\mathbf{H}^{-1})\|\mathbf{a}_m\|^2 + \|\tilde{\mathbf{J}}^T \mathbf{a}_m\|^2 \mathsf{SNR}^{\alpha} \ , \tag{3.32}$$

where $\lambda_{\text{MAX}}(\mathbf{H}^{-1})$ is the maximum singular value of $\mathbf{H}^{-1}$ and we use the shorthand $\tilde{\mathbf{J}} = \mathbf{H^{-1}J}$. Let us partition $\tilde{\mathbf{J}}^T$ into two parts,

$$\tilde{\mathbf{J}}^T = [\mathbf{S}_1, \mathbf{S}_2] \ , \tag{3.33}$$

where $\mathbf{S}_1 \in \mathbb{R}^{2K \times (2M-2K)}$ and $\mathbf{S}_2 \in \mathbb{R}^{2K \times 2K}$. Observe that with probability one, $\tilde{\mathbf{J}}^T$ has rank $2K$ (hence, is full-rank). This implies that we can permute the columns of $\tilde{\mathbf{J}}^T$ in such a way as to ensure that its last $2K$ columns are linearly independent. If we use the same permutation on the coefficients of the vector $\mathbf{a}_m$, our upper bound on the effective noise variance given in (3.32) will remain unchanged. Therefore, without loss of generality, we may assume that $\mathbf{S}_2$ has rank $2K$. We define $\mathbf{T} = -\mathbf{S}_2^{-1}\mathbf{S}_1$. Then, we can write

$$\mathbf{S}_2^{-1}\tilde{\mathbf{J}}^T = [\mathbf{S}_2^{-1}\mathbf{S}_1, \mathbf{S}_2^{-1}\mathbf{S}_2] \tag{3.34}$$

$$= [-\mathbf{T}, \mathbf{I}_{2K}] \ . \tag{3.35}$$

Let the coefficients for the m^{th} equation be given by $\mathbf{a}_m = [\mathbf{q}_m^T, \mathbf{p}_m^T]^T$ where $\mathbf{q}_m \in \mathbb{Z}^{2M-2K}$ and $\mathbf{p}_m \in \mathbb{Z}^{2K}$. We use (3.35) to bound $\|\tilde{\mathbf{J}}^T\mathbf{a}_m\|$:

$$\begin{aligned}\|\tilde{\mathbf{J}}^T\mathbf{a}_m\|^2 &= \|\mathbf{S}_2\mathbf{S}_2^{-1}\tilde{\mathbf{J}}^T\mathbf{a}_m\|^2 &(3.36)\\ &= \|\mathbf{S}_2[-\mathbf{T}, \mathbf{I}_{2K}]\mathbf{a}_m\|^2 &(3.37)\\ &\leq \lambda^2_{\text{MAX}}(\mathbf{S}_2)\|[-\mathbf{T}, \mathbf{I}_{2K}]\mathbf{a}_m\|^2 &(3.38)\\ &= \lambda^2_{\text{MAX}}(\mathbf{S}_2)\|\mathbf{T}\mathbf{q}_m - \mathbf{p}_m\|^2 &(3.39)\end{aligned}$$

where $\lambda_{\text{MAX}}(\mathbf{S}_2)$ the maximum singular value of $\mathbf{S}_2$. Combining (3.39) with (3.32), the effective noise variance is bounded as follows:

$$\sigma^2 \leq \min_{\mathbf{A}:|\mathbf{A}|\neq 0} \max_m \lambda^2_{\text{MAX}}(\mathbf{H}^{-1})\|\mathbf{a}_m\|^2 + \lambda^2_{\text{MAX}}(\mathbf{S}_2)\|\mathbf{T}\mathbf{q}_m - \mathbf{p}_m\|^2\mathsf{SNR}^{\alpha} \ . \tag{3.40}$$

We proceed to bound the quantity $\|\mathbf{T}\mathbf{q}_m - \mathbf{p}_m\|^2$. We decompose $\mathbf{T}$ into its integer and fractional parts:

$$\mathbf{T} = \mathbf{T}_I + \mathbf{T}_F \tag{3.41}$$

where $\mathbf{T}_I$ represents the integer part of $\mathbf{T}$ and $\mathbf{T}_F$ represents the fractional part of $\mathbf{T}$. We define

$$\begin{aligned}\tilde{\mathbf{p}}_m &= \mathbf{p}_m - \mathbf{T}_I\mathbf{q}_m &(3.42)\\ \tilde{\mathbf{a}}_m &= [\mathbf{q}_m^T, \tilde{\mathbf{p}}_m^T]^T &(3.43)\\ \tilde{\mathbf{A}} &= [\tilde{\mathbf{a}}_1 \cdots \tilde{\mathbf{a}}_{2M}]^T &(3.44)\\ \mathbf{G} &= \begin{pmatrix} \mathbf{I}_{2M-2K} & \mathbf{0} \\ -\mathbf{T}_I & \mathbf{I}_{2K} \end{pmatrix} \ . &(3.45)\end{aligned}$$

Since $\mathbf{G}$ is a $2M \times 2M$ lower triangular matrix with non-zero diagonal elements, it has rank $2M$. We note that

$$\mathbf{a}_m = \mathbf{G}^{-1}\tilde{\mathbf{a}}_m \ . \tag{3.46}$$

Since $\mathbf{G}$ is invertible, it follows that if the matrix formed by the coefficient vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{2M}$ is full-rank, then the matrix formed by $\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_{2M}$ is full-rank and vice versa. From (3.41), it have that

$$\begin{aligned}\|\mathbf{T}\mathbf{q}_m - \mathbf{p}_m\| &= \|\mathbf{T}_F\mathbf{q}_m - (\mathbf{p}_m - \mathbf{T}_I\mathbf{q}_m)\| &(3.47)\\ &= \|\mathbf{T}_F\mathbf{q}_m - \tilde{\mathbf{p}}_m\| &(3.48)\end{aligned}$$

and, from (3.46), we have that

$$\begin{aligned}\|\mathbf{a}_m\|^2 &= \|\mathbf{G}^{-1}\tilde{\mathbf{a}}_m\|^2 &(3.49)\\ &\leq \lambda^2_{\text{MAX}}(\mathbf{G}^{-1})\|\tilde{\mathbf{a}}_m\|^2 \ . &(3.50)\end{aligned}$$

From (3.40), (3.48), and (3.50), the effective noise variance can be upper bounded by

$$\sigma^2 \leq \min_{\tilde{\mathbf{A}}:|\tilde{\mathbf{A}}|\neq 0} \max_m \lambda^2_{\text{MAX}}\,(\mathbf{H}^{-1})\lambda^2_{\text{MAX}}(\mathbf{G}^{-1})\|\tilde{\mathbf{a}}_m\|^2 + \lambda^2_{\text{MAX}}(\mathbf{S}_2)\|\mathbf{T}_F\mathbf{q}_m - \tilde{\mathbf{p}}_m\|^2\mathsf{SNR}^{\alpha} \tag{3.51}$$

From Theorem 4.13, there exists a Q' such that for all $Q > Q'$, there exist $2M$ linearly independent vectors:

$$\tilde{\mathbf{a}}_m = [\mathbf{q}_m^T, \tilde{\mathbf{p}}_m^T]^T \in \mathbb{Z}^{2M-2K} \times \mathbb{Z}^{2K} \quad \text{for } m = 1, \ldots, 2M \tag{3.52}$$

satisfying the following two inequalities:

$$\|\mathbf{q}_m\| \leq C(\log Q)^2 Q \tag{3.53}$$

$$\|\mathbf{T}_F\mathbf{q}_m - \tilde{\mathbf{p}}_m\| \leq \frac{C(\log Q)^2}{Q^{\frac{M-K}{K}}} \tag{3.54}$$

where C is some constant independent of Q. For sufficiently large Q, we observe that $\frac{C(\log Q)^2}{Q^{\frac{M-K}{K}}} \leq 1$. Using (3.53) and (3.54), we bound the norm of $\tilde{\mathbf{a}}_m$ as follows:

$$\begin{aligned}
\|\tilde{\mathbf{a}}_m\| &= \sqrt{\|\mathbf{q}_m\|^2 + \|\tilde{\mathbf{p}}_m\|^2} && (3.55)\\
&\leq \|\mathbf{q}_m\| + \|\tilde{\mathbf{p}}_m\| && (3.56)\\
&= \|\mathbf{q}_m\| + \|\tilde{\mathbf{p}}_m + \mathbf{T}_F\mathbf{q}_m - \mathbf{T}_F\mathbf{q}_m\| && (3.57)\\
&\leq \|\mathbf{q}_m\| + \|\mathbf{T}_F\mathbf{q}_m\| + \|\tilde{\mathbf{p}}_m - \mathbf{T}_F\mathbf{q}_m\| && (3.58)\\
&\leq \|\mathbf{q}_m\| + \|\mathbf{T}_F\mathbf{q}_m\| + \frac{C(\log Q)^2}{Q^{\frac{M-K}{K}}} && (3.59)\\
&\leq \|\mathbf{q}_m\| + \|\mathbf{T}_F\mathbf{q}_m\| + 1 && (3.60)\\
&\leq \|\mathbf{q}_m\| + \lambda_{\text{MAX}}(\mathbf{T}_F)\|\mathbf{q}_m\| + 1 && (3.61)\\
&\leq C(\log Q)^2 Q\big(1 + \lambda_{\text{MAX}}(\mathbf{T}_F)\big) + 1 && (3.62)
\end{aligned}$$

where $\lambda_{\text{MAX}}(\mathbf{T}_F)$ is the maximum singular value of $\mathbf{T}_F$.

Combining (3.54) and (3.62), the effective noise variance from (3.51) is bounded by

$$\sigma^2 \leq \lambda^2_{\text{MAX}}(\mathbf{H}^{-1})\lambda^2_{\text{MAX}}(\mathbf{G}^{-1})\Big(C(\log Q)^2 Q\big(1 + \lambda_{\text{MAX}}(\mathbf{T})\big) + 1\Big)^2 + \lambda^2_{\text{MAX}}(\mathbf{S}_2)\mathsf{SNR}^{\alpha}\left(\frac{C(\log Q)^2}{Q^{\frac{M-K}{K}}}\right)^2 \tag{3.63}$$

Let Q scale according to $Q^2 \sim \mathsf{SNR}^{\gamma}$. It follows that

$$\sigma^2 \leq \Theta(\log \mathsf{SNR})\left(\mathsf{SNR}^{\gamma} + \mathsf{SNR}^{\alpha-\gamma\left(\frac{M-K}{K}\right)}\right) . \tag{3.64}$$

Setting $\gamma = \frac{K}{M}\alpha$, we find that the generalized degrees-of-freedom are

$$d_{\text{INT}} = \lim_{\mathsf{SNR}\to\infty} 2M \frac{\frac{1}{2}\log\left(\frac{\mathsf{SNR}}{\sigma^2}\right)}{\log \mathsf{SNR}} \tag{3.65}$$

$$= \lim_{\mathsf{SNR}\to\infty} M \frac{\log\left(\frac{\mathsf{SNR}}{\mathsf{SNR}^{\alpha\frac{K}{M}}}\right)}{\log \mathsf{SNR}} \tag{3.66}$$

$$= M\left(1 - \alpha\frac{K}{M}\right) \tag{3.67}$$

$$= M - K\alpha \ , \tag{3.68}$$

which concludes the proof of Theorem 3.7. □

Chapter 4

Diophantine Approximations

This chapter provides a partial overview of some basic techniques in Diophantine approximations. In the first section, we review some classical results for finding a single integer approximations in Theorems 4.3 - 4.7 and multiple integer approximations in Theorems 4.9 and 4.11. Then, in the second section, we present a new result in Theorem 4.13. Our result shows that finding a set of full rank integer approximations for a matrix is only slightly worse than finding a single approximation in the limit. We note that diophantine approximation techniques have also been useful for other problems in information theory, including alignment interference over fixed channels and charactering the degrees of freedom for compute-and-forward [43, 44].

4.1 Some Classical Results

We first define the unit ball and successive minima in the sequel.

Definition 4.1 (Unit Ball)**.** Let $h : \mathbb{R}^M \to \mathbb{R}$ be a norm. The unit ball with respect to h is denoted bys

$$\mathcal{B}_h = \left\{\mathbf{x} \in \mathbb{R}^M : h(\mathbf{x}) \leq 1\right\} \tag{4.1}$$

the volume of $\mathcal{B}_h$ is denoted by V_h.

Definition 4.2 (Successive Minima)**.** Let $h : \mathbb{R}^M \to \mathbb{R}$ be a norm and $\mathcal{B}_h$ be the unit ball with respect to h. The m^{th} successive minima ϵ_m where $0 < m \leq M$ is given by

$$\epsilon_m = \min\left\{\epsilon : \exists\ m \text{ linearly independent integer points } \mathbf{v}_1, \cdots, \mathbf{v}_m \in \mathbb{Z}^M \cap \epsilon\mathcal{B}_h\right\}$$

In Theorem 4.3 and Corollary 4.4 below, we state Minkowski's upper bound on the first successive minimum.

Theorem 4.3 (Minkowski Successive Minima I)**.** *Let $\mathcal{R}$ be a convex body in $\mathbb{R}^M$ that is symmetric about $\mathbf{0}$ and with volume $V_{\mathcal{R}} > 2^M$. There exists $\mathbf{z} \in \mathbb{Z}^M \backslash \mathbf{0}$ that is contained in $\mathcal{R}$.*

Corollary 4.4. *Let $h : \mathbb{R}^M \to \mathbb{R}$ be a norm and ϵ_1 be the 1^{st} successive minima with respect to h. The following inequality is satisfied*

$$\epsilon_1^M V_h \leq 2^M \tag{4.2}$$

Corollary 4.4 follows directly from Theorem 4.3 by letting $\mathcal{R} = \epsilon_1 \mathcal{B}_h$. Since $\mathcal{B}_h$ is the unit ball with respect to the h norm, $\mathcal{R}$ is a close convex region symmetric about the origin with volume $Vol(\mathcal{R}) = \epsilon_1^M V_h$. In the sequel, we state Blichfeldt's Lemma, which is used in the proof of Theorem 4.3.

Lemma 4.5 (Blichfeldt)**.** *If $\mathcal{R} \subset \mathbb{R}^M$ is a bounded set with volume greater than 1, then there exists $\mathbf{x}, \mathbf{y} \in \mathcal{R}$ such that $\mathbf{x} - \mathbf{y} \in \mathbb{Z}^M$.*

Proof. ([64]) Let C denote the open-half unit hypercube given by

$$C = \left\{ \mathbf{x} \in \mathbb{R}^M : 0 \leq x_i < 1 \text{ for } i = 1, \cdots n \right\}. \tag{4.3}$$

For $\mathbf{z} \in \mathbb{Z}^M$, let $C_{\mathbf{z}}$ represent C translated by $\mathbf{z}$, i.e:

$$C_{\mathbf{z}} = \{\mathbf{z} + \mathbf{x} : \mathbf{x} \in C\} \tag{4.4}$$

and let $\mathcal{R}_{\mathbf{z}}$ be given by

$$\mathcal{R}_{\mathbf{z}} = \mathcal{R} \cap C_{\mathbf{z}} \tag{4.5}$$

We note that the sets $\mathcal{R}_{\mathbf{z}}$ are disjoint and their union equals the space $\mathcal{R}$, i.e:

$$\mathcal{R}_{\mathbf{z}} \cap \mathcal{R}_{\mathbf{z}'} \quad \text{for} \quad \mathbf{z} \neq \mathbf{z}' \tag{4.6}$$

$$\cup_{\mathbf{z} \in \mathbb{Z}^M} \mathcal{R}_{\mathbf{z}} = \mathcal{R} \tag{4.7}$$

Let $\mathcal{R}_{\mathbf{z}} - \mathbf{z} = \{\mathbf{x} - \mathbf{z} : \mathbf{x} \in \mathcal{R}_{\mathbf{z}}\}$. We note that $\mathcal{R}_{\mathbf{z}} - \mathbf{z} \in C$ for all $\mathbf{z} \in \mathbb{Z}^M$ and that $Vol(\mathcal{R}_{\mathbf{z}} - \mathbf{z}) = Vol(\mathcal{R}_{\mathbf{z}})$. Hence, it follows that

$$\sum_{\mathbf{z} \in \mathbb{Z}^n} Vol(\mathcal{R}_{\mathbf{z}} - \mathbf{z}) = \sum_{\mathbf{z} \in \mathbb{Z}^n} Vol(\mathcal{R}_{\mathbf{z}}) = Vol(\mathcal{R}) > 1 \tag{4.8}$$

Since $\mathcal{R}_{\mathbf{z}} - \mathbf{z} \subseteq C$ and $Vol(C) = 1$, (4.8) implies that there exists $\mathbf{z}, \mathbf{z}' \in \mathbb{Z}^M$ with $\mathbf{z} \neq \mathbf{z}'$ such that $\mathcal{R}_{\mathbf{z}} - \mathbf{z} \cap \mathcal{R}_{\mathbf{z}'} - \mathbf{z}' \neq \emptyset$. Hence, there exists $\mathbf{x} \in \mathcal{R}_{\mathbf{z}} - \mathbf{z} \cap \mathcal{R}_{\mathbf{z}'} - \mathbf{z}'$ such that $\mathbf{x} = \mathbf{x}_{\mathbf{z}} - \mathbf{z} = \mathbf{x}_{\mathbf{z}'} - \mathbf{z}'$ for some $\mathbf{x}_{\mathbf{z}'} \in \mathcal{R}_{\mathbf{z}'}$ and $\mathbf{x}_{\mathbf{z}} \in \mathcal{R}_{\mathbf{z}}$. Note that $\mathbf{x}_{\mathbf{z}'}, \mathbf{x}_{\mathbf{z}} \in R$ and that $\mathbf{x}_{\mathbf{z}'} - \mathbf{x}_{\mathbf{z}} = \mathbf{z} - \mathbf{z}' \in \mathbb{Z}^M$. □

Proof. (Theorem 4.3) Let $R' = \left\{\frac{1}{2}\mathbf{x} : \mathbf{x} \in \mathbb{R}^M\right\}$. The volume of R' is bounded as follows

$$Vol(R') = \frac{1}{2^M} Vol(R) > 1 \tag{4.9}$$

Since R' is a scaled version of R, it is also closed, convex and symmetric about the origin. From Lemma 4.5, there exists $\mathbf{x}, \mathbf{y} \in R'$ such that $\mathbf{x} - \mathbf{y} \in \mathbb{Z}^M$. We now show that $\mathbf{x} - \mathbf{y} \in R$. Note that since $R' = \left\{\frac{1}{2}\mathbf{x} : \mathbf{x} \in \mathbb{R}^M\right\}$, we have that $2\mathbf{x}, 2\mathbf{y} \in R$. Since R is symmetric about the origin, we have that $2\mathbf{y} \in R$. Since R is convex, we have that $\frac{1}{2}(2\mathbf{x}) + \frac{1}{2}(-2\mathbf{y}) \in R$. Hence, $\mathbf{x} - \mathbf{y} \in R$. □

Theorem 4.6 (Minkowski Linear Forms)**.** *Let $\mathbf{T} \in \mathbb{R}^{M \times M}$ be a full rank matrix. There exists M integers $a_1, \cdots, a_M$ that satisfy the following inequalities:*

$$\left|\sum_{j=1}^{M} t_{1,j} a_j\right| \leq c_1 \tag{4.10}$$

$$\left|\sum_{j=1}^{M} t_{i,j} a_j\right| < c_i \quad for \quad i = 2, \cdots, M \tag{4.11}$$

as long as

$$c_1 \cdots c_n \geq |det(\mathbf{T})|\,. \tag{4.12}$$

Proof. ([64]) We first consider the case where (4.12) is satisfied with strict inequality. We define the region

$$\mathcal{R} = \left\{\mathbf{x} \in \mathbb{R}^M : \left|\sum_{j=1}^{M} t_{1,j} x_j\right| \leq c_1, \left|\sum_{j=1}^{M} t_{i,j} x_j\right| < c_i \quad \text{for} \quad i = 2, \cdots, M\right\} \tag{4.13}$$

It can be easily shown using elementary calculus that this region is symmetric, convex and has volume

$$V_{\mathcal{R}} = 2^M \left|det(\mathbf{T}^{-1})\right| c_1 \cdots c_M \tag{4.14}$$

$$= \frac{2^M}{|det(\mathbf{T})|} c_1 \cdots c_M \tag{4.15}$$

Using the fact that (4.12) is satisfied with equality, it follows that

$$V_{\mathcal{R}} = \frac{2^M}{|det(\mathbf{T})|} c_1 \cdots c_M \tag{4.16}$$

$$> \frac{2^M}{|det(\mathbf{T})|} |det(\mathbf{T})| \tag{4.17}$$

$$> 2^M \tag{4.18}$$

The result then follows by applying Theorem 4.3. We now consider the case where (4.12) is satisfied with equality. For each ϵ in $0 < \epsilon < 1$, we define the region $\mathcal{R}_\epsilon$ as follows:

$$\mathcal{R}_\epsilon = \left\{ \mathbf{x} \in \mathbb{R}^M : \left| \sum_{j=1}^{M} t_{1,j} x_j \right| \leq c_1 + \epsilon, \left| \sum_{j=1}^{M} t_{i,j} x_j \right| < c_i + \epsilon \quad \text{for} \quad i = 2, \cdots, M \right\} \tag{4.19}$$

and note that $\mathcal{R}_\epsilon \subset \mathcal{R}_{\epsilon'}$ if $\epsilon < \epsilon'$. Let $\mathbf{T}' = \mathbf{T}^{-1}$ and $y_i = \sum_{j=1}^{M} t_{i,j} x_j$. It can be easily seen that the region $\mathcal{R}_\epsilon$ is bounded by some R that is independent of ϵ since

$$|x_i| = \left| \sum_j t'_{i,j} y_i \right| \tag{4.20}$$

$$\leq \sum_j |t'_{i,j}||y_i| \tag{4.21}$$

$$\leq \sum_j |t'_{i,j}| c_i + \epsilon \tag{4.22}$$

$$\leq \sum_j |t'_{i,j}| (c_i + 1) \tag{4.23}$$

$$\leq R \tag{4.24}$$

Since $\mathcal{R}_\epsilon$ is bounded by some R for all $0 < \epsilon < 1$, $\mathcal{R}_\epsilon$ contains only a finite number of integer points. Furthermore, for all $\epsilon > 0, \mathcal{R}_\epsilon$ must contain some non-zero integer point. Hence, there exists some $\mathbf{z} \in \mathbb{Z} \backslash \mathbf{0}$ such that $\mathbf{z} \in \mathcal{R}_\epsilon$ for all ϵ arbitrarily small. The result then follows by taking ϵ to be arbitrarily small. □

Next, we state a foundational result in Diophantine approximations.

Theorem 4.7 (Dirichlet)**.** *For any* $\mathbf{T} \in \mathbb{R}^{K \times (M-K)}$ *and any* $Q \in \mathbb{N}$*, there exists a* $(\mathbf{q}, \mathbf{p}) \in \left(\mathbb{Z}^{M-K} \times \mathbb{Z}^K\right) \backslash \mathbf{0}$ *such that*

$$\|\mathbf{q}\|_\infty \leq Q \tag{4.25}$$

$$\|\mathbf{T}\mathbf{q} - \mathbf{p}\|_\infty < \frac{1}{Q^{\frac{M-K}{K}}} \tag{4.26}$$

Remark 4.8. For the scalar case ($K = 1, M = 2$), Dirichlet's theorem implies that every real number has a sequence of good rational approximations.

Proof. We define the following constants:

$$c_i = Q \quad \text{for} \quad i = 1, \cdots, M - K \tag{4.27}$$

$$c_i = \frac{1}{Q^{\frac{M-K}{K}}} \quad \text{for} \quad i = M - K + 1, \cdots M \tag{4.28}$$

and note that $\Pi_i c_i = 1$. Let the matrix $\mathbf{X} \in \mathbb{R}^{M\times M}$ be given as follows

$$\mathbf{X} = \begin{bmatrix} \mathbf{T} & -\mathbf{I}_K \\ \mathbf{I}_{M-K} & \mathbf{0} \end{bmatrix}$$

Since exchanging rows of a matrix only affects the sign of its determinant, we have that

$$|\det(\mathbf{X})| = \left|\det\left(\begin{bmatrix} \mathbf{I}_{M-K} & \mathbf{0} \\ \mathbf{T} & -\mathbf{I}_K \end{bmatrix}\right)\right| . \tag{4.29}$$

Now we use the fact that the determinant of a lower triangular matrix is just the product of its diagonal entries,

$$|\det(\mathbf{X})| = 1. \tag{4.30}$$

The proof then follows by Theorem 4.6. □

Theorem 4.9 (Khintchine-Groshev)**.** *Fix a decreasing function $\Psi : \mathbb{N} \to \mathbb{R}^+$. If*

$$\sum_{q=1}^{\infty} q^{M-K-1}\Psi(q)^K < \infty, \tag{4.31}$$

then for almost all $\mathbf{T} \in \mathbb{R}^{K\times(M-K)}$ with $|T_{i,j}| \leq 1$ for all i, j, there are only a finite number of solutions $(\mathbf{q}, \mathbf{p}) \in \mathbb{Z}^{M-K} \times \mathbb{Z}^K$ to the following inequality

$$\|\mathbf{Tq} - \mathbf{p}\|_\infty < \Psi(\|\mathbf{q}\|_\infty) \tag{4.32}$$

Remark 4.10. As a result of Theorem 4.9,

$$\|\mathbf{Tq} - \mathbf{p}\|_\infty < \frac{1}{\|\mathbf{q}\|_\infty \log \|\mathbf{q}\|_\infty} \tag{4.33}$$

has infinitely many solutions for almost all $\mathbf{T}$; while

$$\|\mathbf{Tq} - \mathbf{p}\|_\infty < \frac{1}{\|\mathbf{q}\|_\infty \log \left(\|\mathbf{q}\|_\infty\right)^2} \tag{4.34}$$

has infinitely many solutions for almost no $\mathbf{T}$.

The proof of Theorem 4.9 follows along the same lines as the Borel-Cantelli Lemma and can be found in [64, 65].

Theorem 4.11 (Minkowski II)**.** *Let $h : \mathbb{R}^M \to \mathbb{R}$ be a norm and $\epsilon_1, \cdots \epsilon_M$ be the successive minima with respect to h. The following inequality is satisfied*

$$\epsilon_1 \cdots \epsilon_M V_h \leq 2^M \tag{4.35}$$

Remark 4.12. Since $\epsilon_1 \leq \epsilon_2 \cdots \leq \epsilon_M$, Theorem 4.11 implies Theorem 4.3.

The proof of Theorem 4.11 can be found in [64] and involves many properties of convex bodies. We now state our result on diophantine approximations.

4.2 A New Result: Full-Rank Approximations

Theorem 4.13. *Let $\mathbf{T} \in \mathbb{R}^{2K\times(2M-2K)}$ be rationally independent and assume $|t_{i,j}| \leq 1$ for all i, j. There exists a $Q' \in \mathbb{N}$ such that all for $Q > Q'$, there exist M linearly independent integer vectors $(\mathbf{q}_1, \mathbf{p}_1), \cdots (\mathbf{q}_M, \mathbf{p}_M) \in \mathbb{Z}^{M-K} \times \mathbb{Z}^K$ that satisfy*

$$\|\mathbf{q}_m\| \leq CQ(\log Q)^2 \tag{4.36}$$

$$\|\mathbf{T}\mathbf{q}_m - \mathbf{p}_m\| \leq \frac{C(\log Q)^2}{Q^{\frac{M-K}{K}}} \tag{4.37}$$

where C is a constant that is independent of Q.

Proof. For any vector $\mathbf{v} \in \mathbb{Z}^{2M}$, we denote the first $2M - 2K$ components by $\mathbf{q}$ and the remaining $2K$components by $\mathbf{p}$, and will thus write

$$\mathbf{v} = \begin{bmatrix}\mathbf{q}\\ \mathbf{p}\end{bmatrix}. \tag{4.38}$$

From the statement of the theorem, $\mathbf{T} = [\mathbf{t}_1 \cdots \mathbf{t}_{2K}]^T$ is a (rationally independent) $2K \times (2M - 2K)$ real-valued matrix with $|t_{i,j}| \leq 1$ for all i, j. For a fixed $\mathbf{T}$, we define the semi-norms f, g as follows:

$$f(\mathbf{v}) = \|\mathbf{T}\mathbf{q} - \mathbf{p}\| \tag{4.39}$$

$$g(\mathbf{v}) = \|\mathbf{q}\| . \tag{4.40}$$

For a fixed $Q \geq 2M$, we let λ_1 denote the minimum value of $f(\mathbf{v})$ under the constraint $g(\mathbf{v}) \leq Q$,

$$\lambda_1 = \min_{\substack{\mathbf{v}\in\mathbb{Z}^{2M}\setminus\{\mathbf{0}\}\\ g(\mathbf{v})\leq Q}} f(\mathbf{v}) \tag{4.41}$$

$$= \min_{\substack{\mathbf{q}\in\mathbb{Z}^{2M-2K}\\ \|\mathbf{q}\|\leq Q}} \min_{\substack{\mathbf{p}\in\mathbb{Z}^{2K}\\ [\mathbf{q}^T,\mathbf{p}^T]^T\neq\mathbf{0}}} \|\mathbf{T}\mathbf{q} - \mathbf{p}\|, \tag{4.42}$$

$\mathbf{v}_1 \in \mathbb{Z}^{2M}$ denote an integer vector that achieves λ_1

$$\mathbf{v}_1 = \operatorname*{argmin}_{\substack{\mathbf{v}\in\mathbb{Z}^{2M}\setminus\{\mathbf{0}\}\\ g(\mathbf{v})\leq Q}} f(\mathbf{v}), \tag{4.43}$$

$\mathbf{q}_1 \in \mathbb{Z}^{2M-2K}$ be the first $2M - 2K$ components of $\mathbf{v}_1$, and μ_1 be the value of $\mathbf{v}$ evaluated by g,

$$\mu_1 = g(\mathbf{v}_1) \tag{4.44}$$

$$= \|\mathbf{q}_1\| . \tag{4.45}$$

We note that for large enough Q, $\lambda_1 < 1$ and $\mu_1 > 0$.

From now on, we assume that Q is sufficiently large so that $\lambda_1 < 1$ and $\mu_1 > 0$. Based on the seminorms f and g, we define the function $h : \mathbb{R}^{2M} \to \mathbb{R}$ as follows:

$$h(\mathbf{v}) = \left(f^2(\mathbf{v}) + \frac{\lambda_1^2}{\mu_1^2} g^2(\mathbf{v}) \right)^{1/2} \tag{4.46}$$

$$= \left(\|\mathbf{T}\mathbf{q} - \mathbf{p}\|^2 + \frac{\lambda_1^2}{\mu_1^2} \|\mathbf{q}\|^2 \right)^{1/2}. \tag{4.47}$$

In the sequel, we show that h is a norm. We define the $2M \times 2M$ matrix $\mathbf{\Gamma}$ as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{T} & -\mathbf{I}_{2K} \\ \frac{\lambda_1}{\mu_1}\mathbf{I}_{2M-2K} & \mathbf{0} \end{bmatrix}$$

Note that we can rewrite the function h using $\mathbf{\Gamma}$:

$$h(\mathbf{v}) = \|\mathbf{\Gamma}\mathbf{v}\|. \tag{4.48}$$

Since exchanging rows of a matrix only affects the sign of its determinant, we have that

$$|\det(\mathbf{\Gamma})| = \left| \det\left(\begin{bmatrix} \frac{\lambda_1}{\mu_1}\mathbf{I}_{2M-2K} & \mathbf{0} \\ \mathbf{T} & -\mathbf{I}_{2K} \end{bmatrix} \right) \right|. \tag{4.49}$$

Now we use the fact that the determinant of a lower triangular matrix is just the product of its diagonal entries,

$$|\det(\mathbf{\Gamma})| = \left(\frac{\lambda_1}{\mu_1} \right)^{2M-2K}. \tag{4.50}$$

Since $\mathbf{T}$ is rationally independent, it follows that $\lambda_1 > 0$. Since $\mu_1 > 0$ by assumption, we have that $\frac{\lambda_1}{\mu_1} > 0$. Since $\mathbf{\Gamma}$ is full-rank and thus injective, h is a norm.

Let V_h be the volume of the h-unit ball. Let $\mathbf{u} = \mathbf{\Gamma}\mathbf{v}$. It follows that:

$$V_h = \int_{\{\mathbf{v}:\|\mathbf{\Gamma}\mathbf{v}\| \le 1\}} d\mathbf{v} \tag{4.51}$$

$$= \int_{\{\mathbf{u}:\|\mathbf{u}\| \le 1\}} |\det(\mathbf{\Gamma}^{-1})| d\mathbf{u} \tag{4.52}$$

$$= \frac{1}{|\det(\mathbf{\Gamma})|} \int_{\{\mathbf{u}:\|\mathbf{u}\| \le 1\}} d\mathbf{u} \tag{4.53}$$

$$= \frac{1}{|\det(\mathbf{\Gamma})|} V_{2M} \tag{4.54}$$

$$= \left(\frac{\mu_1}{\lambda_1} \right)^{2M-2K} V_{2M}, \tag{4.55}$$

where V_{2M} is the volume of the unit ball with respect to the Euclidean norm (in $2M$ dimensional space).

Let $\epsilon_1, \ldots, \epsilon_{2M}$ be the successive minima with respect to h (see Definition 4.2). Let $\mathbf{y}_1, \ldots, \mathbf{y}_{2M} \in \mathbb{Z}^{2M}$ be linearly independent integer points that achieve the successive minima, i.e:

$$h(\mathbf{y}_i) = \epsilon_i. \tag{4.56}$$

Using Minkowski's 2nd Theorem on successive minima (Theorem 4.11), we have that:

$$V_h \prod_{i=1}^{2M} \epsilon_i \leq 2^{2M}, \tag{4.57}$$

where V_h is the volume of the h-unit ball. Using (4.55), we have that:

$$\left(\frac{\mu_1}{\lambda_1}\right)^{2M-2K} V_{2M} \prod_{i=1}^{2M} \epsilon_i \leq 2^{2M}. \tag{4.58}$$

Rewriting the above, we have that

$$\left(\frac{\mu_1}{\lambda_1}\right)^{2M-2K} \prod_{i=1}^{2M} \epsilon_i \leq C, \tag{4.59}$$

where C is a constant that depends only on $2M$. Rearranging (4.59), we arrive at the following:

$$\epsilon_{2M} \leq C \left(\frac{\lambda_1}{\epsilon_1} \cdots \frac{\lambda_1}{\epsilon_{2M-2K}}\right) \left(\frac{1}{\epsilon_{2M-2K+1} \cdots \epsilon_{2M-1}}\right) \left(\frac{1}{\mu_1^{2M-2K}}\right) \tag{4.60}$$

$$= C \left(\frac{\lambda_1}{\epsilon_1} \cdots \frac{\lambda_1}{\epsilon_{2M-2K}}\right) \left(\frac{\lambda_1}{\epsilon_{2M-2K+1}} \cdots \frac{\lambda_1}{\epsilon_{2M-1}}\right) \left(\frac{1}{\mu_1^{2M-2K} \lambda_1^{2K-1}}\right) \tag{4.61}$$

$$= C \left(\frac{\lambda_1}{\epsilon_1} \cdots \frac{\lambda_1}{\epsilon_{2M-1}}\right) \left(\frac{1}{\mu_1^{2M-2K} \lambda_1^{2K-1}}\right) \tag{4.62}$$

$$= C \left(\frac{\lambda_1}{\epsilon_1} \cdots \frac{\lambda_1}{\epsilon_{2M-1}}\right) \left(\frac{\lambda_1}{\mu_1^{2M-2K} \lambda_1^{2K}}\right). \tag{4.63}$$

For all $\mathbf{v} \in \mathbb{Z}^{2M} \setminus \{0\}$, we have that $h(\mathbf{v}) \geq \lambda_1$. To see this, we can consider the case where $\|\mathbf{q}\| < \mu_1$ and $\|\mathbf{q}\| \geq \mu_1$ separately. When $\|\mathbf{q}\| \geq \mu_1$, h can be bounded as follows

$$h(\mathbf{v}) = \left(\|\mathbf{T}\mathbf{q} - \mathbf{p}\|^2 + \frac{\lambda_1^2}{\mu_1^2} \|\mathbf{q}\|^2\right)^{1/2} \tag{4.64}$$

$$\geq \frac{\lambda_1}{\mu_1} \|\mathbf{q}\| \tag{4.65}$$

$$\geq \lambda_1. \tag{4.66}$$

We now consider the case where $\|\mathbf{q}\| < \mu_1$. We first bound h as follows

$$h(\mathbf{v}) = \left(\|\mathbf{T}\mathbf{q} - \mathbf{p}\|^2 + \frac{\lambda_1^2}{\mu_1^2}\|\mathbf{q}\|^2 \right)^{1/2} \tag{4.67}$$

$$\geq \|\mathbf{T}\mathbf{q} - \mathbf{p}\|. \tag{4.68}$$

Recall that $\mu_1 = \|\mathbf{q}_1\|$ and $\lambda_1 = \min_{\mathbf{p}\in\mathbb{Z}^{2K}} \|\mathbf{T}\mathbf{q}_1 - \mathbf{p}\|$. Assume that there exists a $\mathbf{q}$ with $\|\mathbf{q}\| < \|\mathbf{q}_1\|$ such that

$$\min_{\mathbf{p}} \|\mathbf{T}\mathbf{q} - \mathbf{p}\| < \min_{\mathbf{p}} \|\mathbf{T}\mathbf{q}_1 - \mathbf{p}\| \tag{4.69}$$

$$= \lambda_1, \tag{4.70}$$

then the definition of $\mathbf{q}_1$ in (4.43) is violated. Hence, in this case, $h(\mathbf{v}) \geq \|\mathbf{T}\mathbf{q} - \mathbf{p}\| \geq \lambda_1$.

Since $\epsilon_j = h(\mathbf{y}_j)$ for some $\mathbf{y}_j \in \mathbb{R}^{2M}$, it follows that

$$\epsilon_j \geq \lambda_1 \quad \text{for all} \quad j = 1, \ldots, 2M. \tag{4.71}$$

Combining the above with (4.63), it follows that

$$\epsilon_{2M} \leq C \frac{\lambda_1}{\mu_1^{2M-2K}\lambda_1^{2K}}. \tag{4.72}$$

From the definition of h, ϵ_{2M}, and $\mathbf{y}_1 \cdots \mathbf{y}_{2M}$, we have that

$$h(\mathbf{y}_j) \leq \epsilon_{2M} \quad \text{for} \quad j = 1 \cdots 2M. \tag{4.73}$$

By the construction of h (see (4.46)), we have that:

$$f(\mathbf{y}_j) \leq h(\mathbf{y}_j) \leq \epsilon_{2M} \leq C \frac{\lambda_1}{\lambda_1^{2K}\mu_1^{2M-2K}} \tag{4.74}$$

$$\frac{\lambda_1}{\mu_1} g(\mathbf{y}_j) \leq h(\mathbf{y}_j) \leq \epsilon_{2M} \leq C \frac{\lambda_1}{\lambda_1^{2K}\mu_1^{2M-2K}}, \tag{4.75}$$

for $j = 1, \cdots, 2M$. The above equations imply that

$$f(\mathbf{y}_j) \leq C \frac{\lambda_1}{\lambda_1^{2K}\mu_1^{2M-2K}} \tag{4.76}$$

$$g(\mathbf{y}_j) \leq C \frac{\mu_1}{\lambda_1^{2K}\mu_1^{2M-2K}}, \tag{4.77}$$

for $j = 1, \ldots, 2M$.

Recall that λ_1, μ_1 are defined with respect to a fixed Q. We now show that for all sufficiently large Q,

$$\lambda_1(Q)^{2K}\mu_1(Q)^{2M-2K} \geq \frac{1}{\log(\mu_1(Q))^2}. \tag{4.78}$$

We define the function $\Psi : \mathbb{Z} \to \mathbb{R}$ as follows

$$\Psi(q) = 1 \quad \text{for} \quad q = 1 \tag{4.79}$$

$$\Psi(q) = \frac{1}{q^{\frac{2M-2K}{2K}} (\log q)^{\frac{2}{2K}}} \quad \text{for} \quad q > 1. \tag{4.80}$$

We note that with this choice of Ψ, it follows that

$$\sum_q q^{2M-2K-1} \Psi(q)^{2K} < \infty. \tag{4.81}$$

Applying Theorem 4.9, we have that for rationally independent $\mathbf{T} \in \mathbb{R}^{2K \times (2M-2K)}$, there are only a finite number of integer solutions $[\mathbf{q}^T, \mathbf{p}^T]^T \in \mathbb{Z}^{2M-2K} \times \mathbb{Z}^{2K}$ that satisfy the following condition:

$$\|\mathbf{T}\mathbf{q} - \mathbf{p}\|_\infty < \frac{1}{\|\mathbf{q}\|_\infty^{\frac{2M-2K}{2K}} (\log \|\mathbf{q}\|_\infty)^{\frac{2}{2K}}}. \tag{4.82}$$

We rewrite this condition as follows

$$\|\mathbf{q}\|_\infty^{2M-2K} \|\mathbf{T}\mathbf{q} - \mathbf{p}\|_\infty^{2K} < \frac{1}{(\log \|\mathbf{q}\|_\infty)^2}. \tag{4.83}$$

We first fix an rationally independent $\mathbf{T} \in \mathbb{R}^{2K \times (2M-2K)}$. Recall from (4.43) that $\mathbf{q}_1(Q)$ is the integer vector that achieves $\lambda_1(Q)$ for a given Q. Clearly, $\{\|\mathbf{q}_1(Q)\|_\infty\}_Q$ is a non-decreasing integer sequence (in Q). Assume that $\|\mathbf{q}_1(Q)\|_\infty$ is unbounded as $Q \to \infty$. By Theorem 4.9, we know that there are only a finite number of integers $\mathbf{q}$ that satisfy the condition in (4.83). Let $\mathbf{q}'$ be the integer with the largest L_∞ norm that satisfies the condition in (4.83). This suggests that for all Q where $\|\mathbf{q}_1(Q)\|_\infty > \|\mathbf{q}'\|_\infty$, $\mathbf{q}_1(Q)$ does not satisfying the condition in (4.83). Since $\{\|\mathbf{q}_1(Q)\|_\infty\}_Q$ is an unbounded non decreasing sequence, there exists some Q' such that for all $Q > Q'$, $\mathbf{q}_1(Q)$ does not satisfy the condition in (4.83). Note that (4.78) follows since any $\mathbf{q}, \mathbf{p}$ that satisfies

$$\|\mathbf{q}\|_\infty^{2M-2K} \|\mathbf{T}\mathbf{q} - \mathbf{p}\|_\infty^{2K} \geq \frac{1}{(\log \|\mathbf{q}\|_\infty)^2} \tag{4.84}$$

also satisfies

$$\|\mathbf{q}\|^{2M-2K} \|\mathbf{T}\mathbf{q} - \mathbf{p}\|^{2K} \geq \frac{1}{(\log \|\mathbf{q}\|)^2}. \tag{4.85}$$

Finally, for any rationally independent $\mathbf{T} \in \mathbb{R}^{2K \times (2M-2K)}$, we prove that sequence $\{\|\mathbf{q}_1(Q)\|_\infty\}_Q$ is unbounded as $Q \to \infty$. We prove this by contradiction. That is, assume that there exists

some $C \in \mathbb{Z}_+$ such that $\|\mathbf{q}_1(Q)\|_\infty \leq C$ for all Q. This implies that $\mathbf{q}_1(Q)$ takes only a finite set of values. Hence, there exists a C' such that

$$\min_{\mathbf{p}\in\mathbb{Z}^{2K}} \|\mathbf{T}\mathbf{q}_1(Q) - \mathbf{p}\|_\infty \geq C' \tag{4.86}$$

for all Q. However, by definition of $\lambda_1(Q)$ and Dirichlet's theorem (Theorem 4.7) we have that

$$\min_{\mathbf{p}\in\mathbb{Z}^{2K}} \|\mathbf{T}\mathbf{q}_1(Q) - \mathbf{p}\|_\infty \leq \frac{1}{Q^{\frac{2M-2K}{2K}}} \tag{4.87}$$

for all $Q \in \mathbb{N}$. This results in a contradiction with our assumption. Therefore, (4.78) is proved.

Dirichlet's Theorem (Theorem 4.7) is defined in terms of the ℓ_∞ norm and λ_1 is defined in terms of the ℓ_2 norm. Using the fact that

$$\lambda_1 = \min_{\substack{\mathbf{q}\in\mathbb{Z}^{2M-2K} \\ \|\mathbf{q}\|\leq Q}} \min_{\substack{\mathbf{p}\in\mathbb{Z}^{2K} \\ [\mathbf{q}^T,\mathbf{p}^T]^T\neq\mathbf{0}}} \|\mathbf{T}\mathbf{q} - \mathbf{p}\| \tag{4.88}$$

$$\leq \sqrt{2K} \min_{\substack{\mathbf{q}\in\mathbb{Z}^{2M-2K} \\ \|\mathbf{q}\|\leq Q}} \min_{\substack{\mathbf{p}\in\mathbb{Z}^{2K} \\ [\mathbf{q}^T,\mathbf{p}^T]^T\neq\mathbf{0}}} \|\mathbf{T}\mathbf{q} - \mathbf{p}\|_\infty, \tag{4.89}$$

and (4.87), we have that

$$\lambda_1 \leq \frac{\sqrt{2K}}{Q^{\frac{2M-2K}{2K}}}. \tag{4.90}$$

By definition, we have that

$$\mu_1 \leq Q. \tag{4.91}$$

Using (4.76), (4.78), (4.90), and (4.91), and assuming that Q is sufficiently large, we bound $f(\mathbf{y}_j)$ as follows:

$$f(\mathbf{y}_j) \leq C\frac{\lambda_1}{\lambda_1^{2K}\mu^{2M-2K}} \tag{4.92}$$

$$\leq C\lambda_1(\log\mu_1)^2 \tag{4.93}$$

$$\leq C'\frac{(\log\mu_1)^2}{Q^{\frac{2M-2K}{2K}}} \tag{4.94}$$

$$\leq C'\frac{(\log Q)^2}{Q^{\frac{2M-2K}{2K}}}, \tag{4.95}$$

where C' is a constant that does not depend on Q. Similarly, we can bound $g(\mathbf{y}_j)$ as follows:

$$g(\mathbf{y}_j) \leq C\mu_1(\log\mu_1)^2 \leq C'Q(\log Q)^2, \tag{4.96}$$

which concludes the proof. □

Chapter 5

Network Function Computation

In this chapter, we consider linear function computation over multi-hop wired and wireless networks. Currently, function computation over wired and wireless networks remain as two separate areas of research. Function computation in wired networks has been studied in [8, 9, 10, 66, 67, 68] and general results for network with arbitrary topologies were discovered. While wired networks contain noiseless bit pipes, wireless networks include effects such as noisy superposition at the receivers and broadcast at the transmitters. These effects make it difficult to characterize the performance of function computation over networks with complicated topologies. Most existing literature has been focused on networks with simple topologies such as a single multiple-access channel [31, 69, 70]. Hence, a natural question that arises is how far the current state of the art techniques can be used to understand function computation over multi-hop wireless networks.

We integrate the study of wired network computation and wireless network computation, and show a connection between the two problems. More precisely, we develop coding strategies for wireless networks by first turning them into wired networks. In addition, we develop new results in both wired and wireless setting and provide a distinction between the set of problems that are solvable using cut-set upper bounds and those are not solvable. It is well known that cut-set upper bounds provide a tight converse for many interesting scenarios in wired networks with arbitrary topologies, including multicast [11, 12, 71] and broadcast settings [71]. Meanwhile, the two-unicast problem whose capacity is strictly tighter than cut-set is still open except for special cases [72]. Hence, the tightness of the cut-set bound can be a criteria to identify the problems that are "solvable" with the current state of the art.

In the first part of this chapter, we develop an understanding of computation over wired networks via a duality relation with broadcast networks. Duality between multiple-access and broadcast channels was first discovered in single-hop scalar and MIMO wireless networks [73, 74]. In [74], the goal was to study MIMO broadcast problems using solutions from multiple-access MIMO channels. Since the computation capacity of the Gaussian multiple-access channel is unknown, this elegant duality relationship cannot be extended to wireless

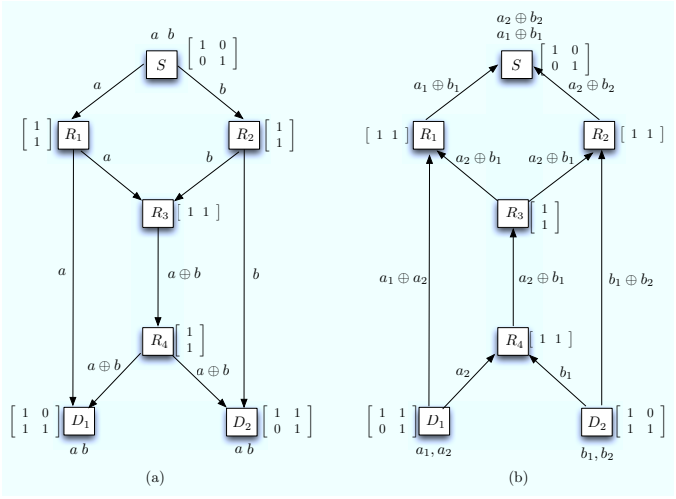

Figure 5.1. As shown in [7][Theorem 5], multicast is dual to function computation for orthogonal networks. In the multicast setting in (a), each destination wants messages a, b. In the function computation setting in (b), the destination wants $a_1 \oplus b_1, a_2 \oplus b_2$. The same LTI code (with matrix transpose), can be used for both problems

networks. Instead, a duality relation between multicast and computing the XOR of the data packets over wired multiple-access networks was found in [7]. Figure 5.1 illustrates the relation using the standard butterfly example. In the multiple-access network, source 1 transmits a_1, a_2, source 2 transmits b_1, b_2, and the receiver recovers $a_1 \oplus b_1, a_2 \oplus b_2$. In the multicast network, the source transmits a, b and both destinations recover a, b. It is shown that the same linear code works for both problems. In essence, transmitting a common message in the broadcast network is dual to computing a function in the multiple-access network. We generalize this duality relation to arbitrary linear deterministic networks using the algebraic network coding approach based on characterizing the transfer functions of both networks [71, 75, 76]. Using this connection, we identify the scenarios under which cut-set bounds are tight for communicating over the broadcast network and computing over the multiple-access network.

In the second part of this chapter, we apply the results from wired networks to wireless networks. The main idea is to extract insights from the linear deterministic model introduced in [77]. However, the problem considered in [77] is concerned with recovering the individual messages. Since only the rate of information flow matters and structure in the messages is not needed, it was sufficient to use codes without any algebraic structure. By contrast, in computation problems where a particular algebraically structured function of the messages is communicated, it is crucial to keep the algebraic structure of the messages. We use nested lattice codes both for channel coding [16, 31] and source quantization [17, 18, 23, 78, 79]. After applying nested-lattice codes, the wireless network problem can be reduced to a wired network problem. The duality relation can then be used to compute functions

of discrete sources over linear deterministic networks. Using this approach, we characterize the distortion for sending the sum of Gaussian sources across a class of relay networks to within a constant factor of the optimal performance. Furthermore, we discuss the problems where the cut-set bounds are approximately tight, and give an intuition based on linear deterministic models.

Our approach can be understood as separating the overall network problem into two different layers, a physical layer and a network layer, but with different paradigms. First, the current physical layer requires each node in the network to decode a message which has an explicit source node. However, in our scheme each node receives a noisy superposition of the incoming signals and may decode a function of the incoming messages as suggested in [32]. Second, the current network layer performs routing where the outgoing data packets are a subset of the incoming data packets. In our scheme, as in standard network coding, the relays can create a new packet by taking a linear combination of its incoming data packets. Overall, our scheme separates the physical and the network layers of the network and allows for computation in both layers. These new paradigms can potentially provide a different dimension to the design of distributed sensor networks.

5.1 Computation in Deterministic Networks

Our primary problem of interest is linear computation over linear deterministic multiple-access networks as described in Sections 5.1.1 and 5.1.2 below. Instead of solving the computation problem directly, we define the dual broadcast problem in Sections 5.1.3 and 5.1.4 and prove the equivalence of the two problems in Theorem 5.17. The duality connection between computation over multiple-access networks and communication over broadcast networks provides a means to convert the original problem to one that is well studied in the literature [11, 71, 80]. In Theorem 5.27, we find a compact expression to describe the situations under which cut-set bounds are tight by first considering the broadcast scenario and then applying the solution to the computation situation.

5.1.1 Linear deterministic Multiple-Access Network

A linear deterministic multiple-access network $\mathcal{N}_{\text{DET-MAC}}$ is represented by a set of nodes $\{N_i\}_{i\in\Omega}$. A given node N_i contains $b_{i,\text{in}}$ input links and $b_{i,\text{out}}$ output links. Node N_i inputs $X_i \in \mathbb{F}_p^{|b_{i,\text{in}}|}$ into the network and receives outputs $Y_i \in \mathbb{F}_p^{|b_{i,\text{out}}|}$ from the network. We assume that the underlying finite field is $\mathbb{F}_p$ with prime p. We divide the set of all node indices into source nodes $\mathcal{S}$, relay nodes $\mathcal{R}$ and a single destination node $\mathcal{D}$. We let the source node indices be $\mathcal{S} = \{1, \ldots, m\}$. We assume that $Y_i = \mathbf{0}$ for all $i \in \mathcal{S}$ and $X_i = \mathbf{0}$ for $i \in \mathcal{D}$. Hence, the source nodes do not receive any information from the network and the destination node does not input any information into the network. The channel matrix from N_i to N_j

is denoted by $H_{i,j} \in \mathbb{F}_p^{|b_{j,\text{out}}| \times |b_{i,\text{in}}|}$. The output of node Y_j is given by

$$Y_j = \sum_{i \in \Omega} H_{i,j} X_i \tag{5.1}$$

with all operations over $\mathbb{F}_p$.

Definition 5.1 (Cut)**.** We call a subset $\Gamma \subseteq \Omega$ a cut. The channel matrix for cut Γ is denoted by the matrix H_{Γ,Γ^c} and the information-flow value of cut Γ is given by

$$C_\Gamma^{\text{DET-MAC}} = \max_{p(x_\Omega)} I(X_\Gamma; Y_{\Gamma^c} | X_{\Gamma^c}). \tag{5.2}$$

Remark 5.2. It can be easily shown that in the linear deterministic network with input-output structure given by (5.1), $\max_{p(x_\Omega)} I(X_\Gamma; Y_{\Gamma^c} | X_{\Gamma^c}) = \text{rank}(H_{\Gamma,\Gamma^c}) \log_2 p$.

5.1.2 Computation over Multiple-Access Networks

We consider sending ℓ linear functions of discrete sources across the deterministic multiple-access network.

Definition 5.3 (Source Information)**.** Node N_i for $i \in \mathcal{S}$ observes information $\mathbf{U}_i = (U_{i,1}^{k_1}, \cdots U_{i,\ell}^{k_\ell})$ where $U_{i,j}$ are drawn i.i.d uniformly from the prime-sized finite field $\mathbb{F}_p$. We assume that $\mathbf{U}_i = 0$ for all $i \in \mathcal{S}^c$.

Definition 5.4 (Encoders)**.** At time t, node N_i uses encoder $\mathcal{E}_{i,t}$ to map its received signals Y_i^{t-1} and information $\mathbf{U}_i$ to $X_{i,t}$:

$$\mathcal{E}_{i,t} : \mathbb{F}_p^{k_1} \times \cdots \times \mathbb{F}_p^{k_\ell} \times \mathbb{F}_p^{b_{i,\text{out}} \times (t-1)} \to \mathbb{F}_p^{b_{i,\text{in}}} \tag{5.3}$$

$$X_{i,t} = \mathcal{E}_{i,t}(\mathbf{U}_i, Y_i^{t-1}) \tag{5.4}$$

for $t = 1, \ldots, n$. The symbol $X_{i,t}$ is input into the network.

Definition 5.5 (Decoder)**.** The destination observes Y_i^n for $i \in \mathcal{D}$ and reconstructs $\widehat{V}_1^k, \ldots, \widehat{V}_\ell^k$ using decoder $\mathcal{G}$:

$$\mathcal{G} : \mathbb{F}_p^{b_{i,\text{out}} \times n} \to \mathbb{F}_p^{k_1} \times \cdots \times \mathbb{F}_p^{k_\ell} \tag{5.5}$$

$$\left(\widehat{V}_1^k, \cdots, \widehat{V}_\ell^k\right) = \mathcal{G}\left(Y_i^n\right) \tag{5.6}$$

where $\widehat{V}_j^k$ is an estimate for the linear function $V_j^{k_1} = \sum_{i=1}^m \alpha_{j,i} U_{j,i}^{k_j}$ with coefficients $\alpha_{i,j} \in \mathbb{F}_p$ and all operations over $\mathbb{F}_p$.

Definition 5.6 (Computation Rates)**.** The computation rate tuple $(R_1, \cdots R_\ell)$ where $R_i = \frac{k_i}{n} \log_2 p$ is achievable if for any $\epsilon > 0$, there exist n, encoders $\{\mathcal{E}_{i,t}\}_{t=1}^{n}$ $\forall i \in \Omega$ and a decoder $\mathcal{G}$ such that

$$\Pr\left(\left(\widehat{V}_1^{k_1}, \cdots, \widehat{V}_\ell^{k_\ell}\right) \neq \left(V_1^{k_1}, \cdots, V_\ell^{k_\ell}\right)\right) \leq \epsilon. \tag{5.7}$$

We find it useful to define the computation demands of the multiple-access network, which will be used to establish the duality connection with the broadcast network.

Definition 5.7 (Computation Demands)**.** The set $\mathcal{Q} = \{\mathcal{Q}_1, \ldots \mathcal{Q}_\ell\}$ specifies the computation demands of the network. The element $\mathcal{Q}_j \subseteq \mathcal{S}$ denotes the non-zero coefficient indices of function $V_j^{k_j}$:

$$\mathcal{Q}_j = \{i \in \{1, \ldots, m\} : \alpha_{j,i} \neq 0\} \tag{5.8}$$

Remark 5.8. $\mathcal{Q} = \{\{1, \ldots, m\}\}$ corresponds to case where the destination recovers a single linear function: $V_1^{k_1} = \sum_{i=1}^{m} \alpha_{1,i} U_{1,i}^{k_j}$ with $\alpha_{1,i} \neq 0$ for all i.

Remark 5.9. $\mathcal{Q} = \{\{1\}, \{2\}, \ldots, \{m\}\}$ corresponds to case where the destination recovers independent information from each source node: $U_{1,1}^{k_1}, U_{2,2}^{k_2}, \ldots U_{m,m}^{k_m}$.

Figure 5.1 shows that sending a single linear function is connected to multicast, and the same linear code can be used for both cases. We extend this relation to arbitrary linear deterministic networks and general computation demands. We first define the dual broadcast network and describe its communication demands in 5.1.3, 5.1.4. Then, we describe duality relation in Section 5.1.5.

5.1.3 Dual Broadcast Network

Consider a linear deterministic multiple-access network $\mathcal{N}_{\text{DET-MAC}}$ defined in Section 5.1.1 with nodes $\{N_i\}_{i\in\Omega}$, source nodes $\mathcal{S}_{\text{MAC}}$, destination node $\mathcal{D}_{\text{MAC}}$, and channel matrices $\{H_{\text{MAC},i,j}\}$. The dual broadcast network $\mathcal{N}_{\text{DET-BC}}$ is represented by the same set of nodes Ω but with all the links reversed. The channel matrices of $\mathcal{N}_{\text{DET-BC}}$ are given by

$$H_{\text{BC},i,j} = H_{\text{MAC},j,i}^{T} \quad \text{for all} \quad i, j \in \Omega \tag{5.9}$$

As a result of the reversal, the source nodes of the multiple-access network becomes the destination nodes of the broadcast network: $\mathcal{D}_{\text{BC}} = \mathcal{S}_{\text{MAC}} = \{1, \ldots, m\}$. Similarly, the destination node of the multiple-access network becomes the source node of the dual broadcast network: $\mathcal{S}_{\text{BC}} = \mathcal{D}_{\text{MAC}}$.

Similar to case of the multiple-access network, the value of cut $\Gamma \subseteq \Omega$ of a linear deterministic broadcast network $\mathcal{N}_{\text{DET-BC}}$ is given by $\mathcal{C}_\Gamma^{\text{DET-BC}} = \max_{p(x_\Omega)} I(X_\Gamma; Y_{\Gamma^c} | X_{\Gamma^c})$. Furthermore, from Remark 5.2 and (5.9), the cut values of a pair of dual networks $(\mathcal{N}_{\text{DET-MAC}}, \mathcal{N}_{\text{DET-BC}})$ satisfy the following condition:

$$\mathcal{C}_\Gamma^{\text{DET-MAC}} = \mathcal{C}_{\Gamma^c}^{\text{DET-BC}} \quad \forall \quad \Gamma \subseteq \Omega. \tag{5.10}$$

5.1.4 Communication Across Broadcast Networks

Consider sending ℓ linear functions across the multiple-access network $\mathcal{N}_{\text{DET-MAC}}$ with computation demands $\mathcal{Q} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_\ell\}$ as described in Section 5.1.2. We re-map this to the problem of transmitting ℓ messages across the dual broadcast network $\mathcal{N}_{\text{DET-BC}}$ with communication demands dictated by $\mathcal{Q}$. We first provide some definitions and then show the equivalence of these two problems in the next Section.

Definition 5.10 (Messages)**.** Node N_i for $i \in \mathcal{S}_{\text{BC}}$ has a set of ℓ independent messages $\mathbf{W}_i = \{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$ where message $\mathbf{w}_j$ is drawn independently and uniformly from $\mathcal{W} = \{0, \ldots, p^{k_i} - 1\}$. We assume that $\mathbf{W}_i = 0$ for all $i \in \mathcal{S}_{\text{BC}}^c$.

Definition 5.11 (Encoders)**.** At time t, Node N_i uses encoder $\mathcal{E}_{i,t}$ to map it's received signals Y_i^{t-1} and message $\mathbf{W}_i$ to $X_{i,t}$:

$$\mathcal{E}_{i,t} : \mathbb{F}_p^{k_1} \times \cdots \times \mathbb{F}_p^{k_\ell} \times \mathbb{F}_p^{b_{i,\text{out}} \times (t-1)} \to \mathbb{F}_p^{b_{i,\text{in}}} \tag{5.11}$$

$$X_{i,t} = \mathcal{E}_{i,t}(\mathbf{W}_i, Y_i^{t-1}) \tag{5.12}$$

for $t = 1, \ldots, n$.

Definition 5.12 (Decoder)**.** At destination $i \in \mathcal{D}_{\text{BC}}$, the output Y_i^n is received and decoder $\mathcal{G}_i$ produces an estimate for all messages $\mathbf{w}_j$ such that $j \in \mathcal{T}_i$ where $\mathcal{T}_i = \{j : i \in \mathcal{Q}_j\}$:

$$\mathcal{G}_i : \mathbb{F}_p^{b_{i,\text{out}} \times n} \to \times_{j \in \mathcal{T}_i} \mathbb{F}_p^{k_j} \tag{5.13}$$

$$(\hat{\mathbf{w}}_{j,i} \text{ s.t } j \in \mathcal{T}_i) = \mathcal{G}\left(Y_i^n\right) \tag{5.14}$$

Definition 5.13 (Achievable Rates)**.** The rate tuple $(R_1, \ldots, R_\ell)$ where $R_i = \frac{k_i}{n} \log p$ is are achievable if for any $\epsilon > 0$, there exist n, encoders $\{\mathcal{E}_{i,t}\}_{t=1}^n \, \forall i \in \Omega$ and decoders $\mathcal{G}_i \; \forall i \in \mathcal{D}_{\text{BC}}$ such that

$$P\left(\bigcup_{i=1}^{m} \bigcup_{j \in \mathcal{T}_i} \{\hat{\mathbf{w}}_{j,i} \neq \mathbf{w}_j\}\right) \leq \epsilon. \tag{5.15}$$

Similar to the computation demand for a multiple-access network, we find it useful to define the communication demand for a broadcast network.

Definition 5.14 (Communication Demands)**.** The set $\mathcal{P} = \{\mathcal{P}_1, \ldots \mathcal{P}_\ell\}$ dictates the communication demands of the broadcast network. The element $\mathcal{P}_j \subseteq \mathcal{D}_{\text{BC}}$ contains the destinations that desire to recover message $\mathbf{w}_j$, i.e destination i recovers messages $\mathbf{w}_j$ if $i \in \mathcal{P}_j$.

Remark 5.15. In the multicast situation, $\mathcal{P} = \{\{1, \ldots, m\}\}$.

Remark 5.16. In the case with no common message, $\mathcal{P} = \{\{1\}, \ldots, \{m\}\}$ since each destination only recovers a message uniquely sent to it by the source.

5.1.5 Duality Relation

We extend the relation observed in the butterfly network in Figure 5.1 to arbitrary linear deterministic networks general computation demands. We show that a linear time-invariant code used for computation across $\mathcal{N}_{\text{DET-MAC}}$ also can be used for communication in the dual $\mathcal{N}_{\text{DET-BC}}$. As a result, the problem of function computation is equivalent to the problem of broadcast and the solution of one problem can be used to solve the other problem. This approach is useful since broadcast networks are well studied in the literature and cases where cut-set is tight have been characterized [11, 71].

Since we limit our discussion to linear time-invariant codes, each encoder $\mathcal{E}_i$ and decoder $\mathcal{G}$ can be written as matrices over the underlying finite field $\mathbb{F}_p$. As observed in [71], the finite field $\mathbb{F}_p$ has to be extended to achieve the multicast capacity. However, this approach may lead to a mismatch in function computation since the relays operate in the extended field while the destination recovers a linear function over the original field. Instead, we extend the underlying field $\mathbb{F}_p$ to the rational function field $\mathbb{F}_p[z]$, which can be viewed as a z-transform [81]. Extending the original field size is then equivalent increasing the memory along the lines of [82]. Therefore, the encoders and decoder can be simply written as matrices $\mathbf{K}_i(z)$ where each elements of the matrices comes from $\mathbb{F}_p[z]$. For notational simplicity, we write $\mathbf{K}_i$ instead of $\mathbf{K}_i(z)$.

Theorem 5.17. *Consider a pair of dual linear deterministic networks $\mathcal{N}_{DET\text{-}MAC}, \mathcal{N}_{DET\text{-}BC}$ with demands $\mathcal{Q}$ and $\mathcal{P}$ respectively where $\mathcal{P} = \mathcal{Q}$. If the linear time-invariant code $\{\mathbf{K}_i\}_{i\in\Omega}$ achieves computation rates $(R_1, \cdots, R_\ell)$ for $\mathcal{N}_{DET\text{-}MAC}$ with any $\{\alpha_{i,j}\}_{i,j}$ then $\left\{\mathbf{K}_i^T\right\}_{i\in\Omega}$ achieves the same rates for $\mathcal{N}_{DET\text{-}BC}$.*

Proof. See Appendix C □

Remark 5.18. Without loss of generality, we can consider the case where $\alpha_{i,j} = 1$ if $\alpha_{i,j} \neq 0$ since we can precode and replace $U_{i,j}$ by $\alpha_{i,j}U_{i,j}$.

Corollary 5.19. *Consider a pair of dual linear deterministic networks $\mathcal{N}_{DET\text{-}MAC}, \mathcal{N}_{DET\text{-}BC}$ with demands $\mathcal{Q}$ and $\mathcal{P}$ respectively where $\mathcal{P} = \mathcal{Q}$. If the cut-set is achievable using linear time-invariant codes in $\mathcal{N}_{DET\text{-}MAC}$, then cut-set is achievable in $\mathcal{N}_{DET\text{-}BC}$.*

Proof. Let $\{\mathbf{K}_i\}_{i\in\Omega}$ be a linear time-invariant code that achieves computation rates $R_1, \ldots, R_\ell$ on $\mathcal{N}_{\text{DET-MAC}}$ with computation demands $\mathcal{Q}$. By assumption, the computation rates $R_1, \ldots, R_\ell$ meet the cut-set bounds for $\mathcal{N}_{\text{DET-MAC}}$. By Theorem 5.17, $\left\{\mathbf{K}_i^T\right\}_{i\in\Omega}$ is a linear time invariant code that achieves rates $R_1, \ldots, R_\ell$ for $\mathcal{N}_{\text{DET-BC}}$ for communication demands $\mathcal{P} = \mathcal{Q}$. By (5.10), the rates $R_1, \ldots, R_\ell$ also meets the cut-set bounds for $\mathcal{N}_{\text{DET-BC}}$. □

Remark 5.20. Theorem 5.17 implies that the solution of multicast can be used for computing a single linear function: $V_1 = \sum_{i=1}^{m} \alpha_{1,i}U_i$.

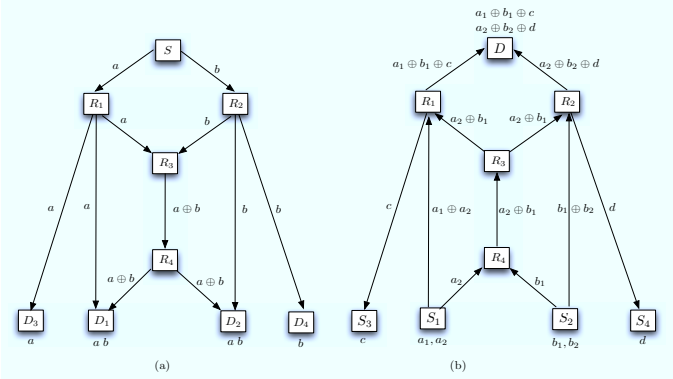

Figure 5.2. In the network in (a), more message destinations are added in addition to the original butterfly network in 5.1. The same linear time invariant code achieves the rates given by the cut-set bounds in both the broadcast network in (a) and the dual multiple-access network in (b)

Remark 5.21. It is well known that cutset is achievable using linear time-invariant codes in the multicast scenario [12]. Combining this with Remark 5.20 and Corollary 5.19, cutset is tight for sending a single linear function across $\mathcal{N}_{\text{DET-MAC}}$. Hence, any computation rate satisfying the following is achievable:

$$R \leq \min_{i \in \mathcal{S}} \min_{\Gamma \subseteq \Omega : i \in \Gamma} C_{\Gamma}^{\text{DET-MAC}} \tag{5.16}$$

We illustrate the duality concept through a series of simple examples in the next Section. The examples show that the same linear code can be used for both function computation and broadcast.

5.1.6 Examples of Dual Networks

I **Broadcast with Multicast Users**. The dual networks in Figure 5.2 is an extension of the butterfly networks in Figure 5.1 with additional users. The multiple-access network in (b) has four source nodes. Source 1 observes a_1, a_2, source 2 observes b_1, b_2, source 3 observes c and source 4 observes d where each element is drawn i.i.d uniformly from $\mathbb{F}_2$. The destination desires to recover two linear functions: $a_1 \oplus b_1 \oplus c$, $a_2 \oplus b_2 \oplus d$. Figure 5.2 provides a linear time invariant code that achieves computation rate pairs $(R_1, R_2) = (1, 1)$, which satisfy the cut-set bounds. The source node in the dual broadcast network in (a) observes symbols a, b. Destination 3 desires to recover a, destination 2 and 3 desire to recover a, b, and destination 4 desires to recover b. Figure 5.2 shows that the same linear time invariant code used in the multiple-access network can be used in the dual broadcast network to achieves rates $(R_1, R_2) = (1, 1)$.

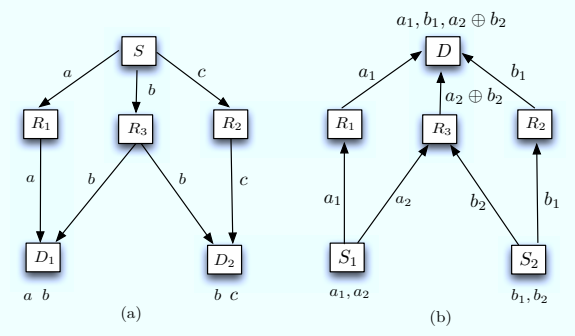

Figure 5.3. (a) is a two-user broadcast network with a common message and (b) is it's dual multiple-access network. It is shown that cut-set is achievable using the same linear time invariant code in both cases.

II **Two-User Broadcast with a Common Message**. The multiple-access network in Figure 5.3 (a) contains two source nodes. Source 1 observes a_1, a_2 and source 2 observes b_1, b_2 where each element is drawn i.i.d uniformly from $\mathbb{F}_2$. The destination desires to recover $a_1, b_1, a_2 \oplus b_2$. The computation rate tuple $(R_1, R_2, R_3) = (1, 1, 1)$ is achievable using linear time invariant codes and meets the cut-set upper bounds. Figure 5.3 (b) is a two-user broadcast network with a common message. The source observes a, b, c and sends a to destination 1, c to destination 2, and b to both destinations. The same linear code used for the multiple-access network can be used for the broadcast network to achieve the rates given by the cut-set upper bounds. So far, we have considered two examples where cut-set bounds are tight. Next, we examine a situation where this is not true.

III **Three-User Broadcast with a Common Message**. Consider the dual linear deterministic network pairs in Figure 5.4. The broadcast network has three destinations and transfer functions:

$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{H}_{S,D_2} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \mathbf{H}_{S,D_3} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{5.17}$$

The source of the broadcast network in Figure 5.4 (a) observes a, b drawn i.i.d uniformly from $\mathbb{F}_2$ and desires to send a to all destinations and b to only destination 3. The rate tuple $(R_1, R_2) = (1, 1)$ satisfies the cut-set upper bounds. However, it is not achievable since both destinations 1 and 2 desire to recover a but the source communicates with these two destinations through different input links. Hence, if a is transmitted on both input links then b cannot be sent. We show in Lemma D.1 in Appendix D that cut-set upper bounds are not tight for this example. Figure 5.4 provides a linear time invariant code that achieves rates $(R_1, R_2) = (1, 0)$ in the broadcast network and computation rates $(R_1, R_2) = (1, 0)$ for the dual multiple-access network in (b). Hence, duality

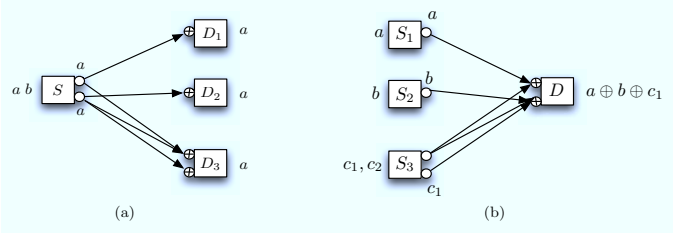

Figure 5.4. Cut-set is not always tight as shown in the three-user broadcast network with a common message in (a). However, duality between computation over multiple-access networks and communication in broadcast networks still holds.

continues to hold even when cut-set is not tight.

5.1.7 Universal Cut-Set Tightness

The duality relation between function computation and broadcast in linear deterministic networks was illustrated in Examples 1-3 in the previous section. Cut-set was shown to be tight in Examples 1 and 2 but not tight in Example 3. In this section, we characterize the scenarios under which the cut-set bounds are tight. We first provide the notion of universal tightness in Definition 5.22. We then focus on the broadcast network and find the communication demands under which cut-set is universally tight. Finally, the duality relation is used to apply the results from broadcast to function computation.

Definition 5.22 (Universally Tight)**.** We call the cutset bound "universally tight" under communication demands $\mathcal{P}$ if it is achievable for all broadcast networks with computation demands $\mathcal{P}$. Similarly, cutset bound is "universally tight" under computation demands $\mathcal{Q}$ if it is achievable for all multiple-access networks with computation demands $\mathcal{Q}$.

Remark 5.23. When cut-set is universally tight under communication demands $\mathcal{P}$, then every network with communication demands $\mathcal{P}$ achieves cut-set. When the cut-set is not universally tight under communication demands $\mathcal{P}$, then there exist a network with communication demands $\mathcal{P}$ for which the cut-set bound is not achievable.

Definition 5.24 (Equivalent Communication Demands)**.** The communication demands $\mathcal{P}$ and $\mathcal{P}'$ are equivalent if there exists an bijective mapping $f : \mathcal{P} \to \mathcal{P}'$ such that the cardinality of arbitrary unions and intersections are the preserved:

$$\left|\bigcup_{\gamma\in\Psi}\bigcap_{\gamma\in\Phi}\mathcal{P}_\gamma\right| = \left|\bigcup_{\gamma\in\Psi}\bigcap_{\gamma\in\Phi} f(\mathcal{P}_\gamma)\right| \tag{5.18}$$

for all $\Psi, \Phi \subseteq \mathcal{P}$.

Definition 5.25 (Sub-demands)**.** We call $\mathcal{P}_{\text{SUB}}$ a sub-demand of the communication demands $\mathcal{P}$ if it is the resulting communication demand after removing users and/or messages from the broadcast network with communication demands $\mathcal{P}$.

Remark 5.26. Let $\mathcal{P} = \{\{1,2\},\{2,3\},\{1,3\}\}$. We note that demands $\{\{1\},\{3\},\{1,3\}\}$ and $\{\{2,3\},\{1,3\}\}$ are sub-demands of $\mathcal{P}$ but $\{\{1,2\},\{2,3\},\{3\}\}$ is not a sub-demand.

Theorem 5.27. *The set of all communication demands under which cut-set is universally tight are equivalent to the following demands and their sub-demands:*

$$\mathcal{P}^{(\ell)}_{TIGHT,1} = \{\{1\} \cup \{\ell,\ldots,m\},\ldots,\{\ell-1\} \cup \{\ell,\ldots,m\},\{\ell,\ldots,m\}\} \quad \text{for} \quad \ell = 2,\ldots,m \tag{5.19}$$

$$\mathcal{P}_{TIGHT,2} = \{\{1\},\{2\},\ldots,\{m\}\} \tag{5.20}$$

$$\mathcal{P}_{TIGHT,3} = \{\{1,3,\ldots,m\},\{2,3,\ldots,m\},\{1,2,3,\ldots,m\}\} \tag{5.21}$$

Furthermore, linear time invariant codes are sufficient to achieve cut-set for communication the demands in (D.1) *to* (D.3)*.*

Theorem 5.27 states that there are essentially only two types of communication demands for which cut-set is universally tight since $\mathcal{P}_{\text{TIGHT},2}$ for m users is a subset of $\mathcal{P}^{(m+1)}_{\text{TIGHT},1}$ for $m+1$ users. The first type, $\mathcal{P}^{(\ell)}_{\text{TIGHT},1}, \mathcal{P}_{\text{TIGHT},2}$, is broadcast with multicast users as shown in Example 5.2. The second type, $\mathcal{P}_{\text{TIGHT},3}$, is two-user broadcast with a common message and multicast users as shown in Example 5.3. Theorem 5.27 implies that for all other communication demands except those given in (D.1) to (D.3), there exists a broadcast network under which cut-set is not achievable. Example 3 illustrates such a case.

Proof. See Appendix D. □

5.1.8 Discussion

In this section, we studied function computation in linear deterministic multiple-access networks. We first considered the dual broadcast problem and then applied the solutions of broadcast to function computation. We characterized the communication demands under which cut-set upper bounds are tight for all broadcast networks. By the duality connection, this maps into a set of computation demands under which cut-set upper bounds are tight for function computation over all multiple-access networks. In the broadcast setting, we examined only the case where the destinations demand a subset of the messages and leave the general case where the destinations may be interested in functions of messages to future work.

5.2 Sum Computation in Networks of Gaussian MACs

Based on the insight from the liner deterministic model in Section 5.1, we consider function computation over wireless multi-hop networks. The proposed approach converts the original problem into one with discrete sources and a linear deterministic network and applies the duality relation from Section 5.1 to solve the deterministic network problem. We illustrate the scheme in Figure 5.5 (a). Here, the goal is to transmit the sum of Gaussian sources U, U' across the multiple-access network. In the network, all links are orthogonal with capacity $\log |\mathbb{F}_p|$ except for the wireless Gaussian MAC from S_1, S_2 to D. This example differs from the butterfly network in Figure 5.1 in two main aspects: Gaussian sources and a Gaussian MAC from S_1, S_2 to R_1. The Gaussian MAC is converted into a linear deterministic MAC over $\mathbb{F}_p$ using nested lattice codes for physical layer computation [31]. The Gaussian sources are mapped into finite field symbols u_1, u_2 and u_1', u_2' using nested-lattice based quantization [23, 78]. After the conversion, the goal is to compute the finite-field sum $u_1 \oplus_p u_1', u_2 \oplus_p u_2'$ over the linear deterministic network in (b). Applying the duality relation in Theorem 5.17, the solution from the dual broadcast network in (c) is used for function computation in (b). Since the linear time-invariant code achieves the cut-set upper bounds for multicast, it also achieves the cut-set upper bounds for computing the finite-field sum. Hence, the finite field sum $u_1 \oplus u_1', u_2 \oplus u_2'$ is recovered with optimal rate in (b). The destination then forms an estimate for $U + U'$ based on $u_1 \oplus u_1', u_2 \oplus u_2'$.

Using the approach above, we characterize the distortion for sending the sum of m Gaussian sources across a class of relay networks in Theorem 5.34. We show that it is within a constant fraction of the optimal performance in Theorem 5.35 and Corollary 5.36. In the sequel, we first provide the problem setup in Sections 5.2.1 and 5.2.2 and illustrate the gain from performing computation over the physical and network layers through examples in Section 5.2.3.

5.2.1 Channel Model

A network of Gaussian multiple-access channels $\mathcal{N}_{\text{GAUSS-MAC}}$ is represented by a set of nodes $\{N_i\}_{i\in\Omega}$. As in the deterministic network, we divide the set of all node indices into those for source nodes $\mathcal{S}$, relay nodes $\mathcal{R}$ and a single destination node $\mathcal{D}$, and let the source node indices be given by $\mathcal{S} = \{1, \ldots, m\}$. We assume that $Y_i = \mathbf{0}$ for all $i \in \mathcal{S}$ and $X_i = \mathbf{0}$ for $i \in \mathcal{D}$. In our network model, we assume that node N_i communicates with node N_j and $N_{j'}$ on orthogonal links for all $j' \neq j$. A given node N_i has b_i input links into the network. At time t, Node N_i inputs signal $(X_{i,1}[t], \ldots, X_{i,b_i}[t]) \in \mathbb{R}^{b_i}$, satisfying the symmetric power constraint:

$$\frac{1}{n}\sum_{t=1}^{n} X_{i,j}^2[t] \leq \frac{\mathsf{SNR}}{b_i} \quad \forall\ j \tag{5.22}$$

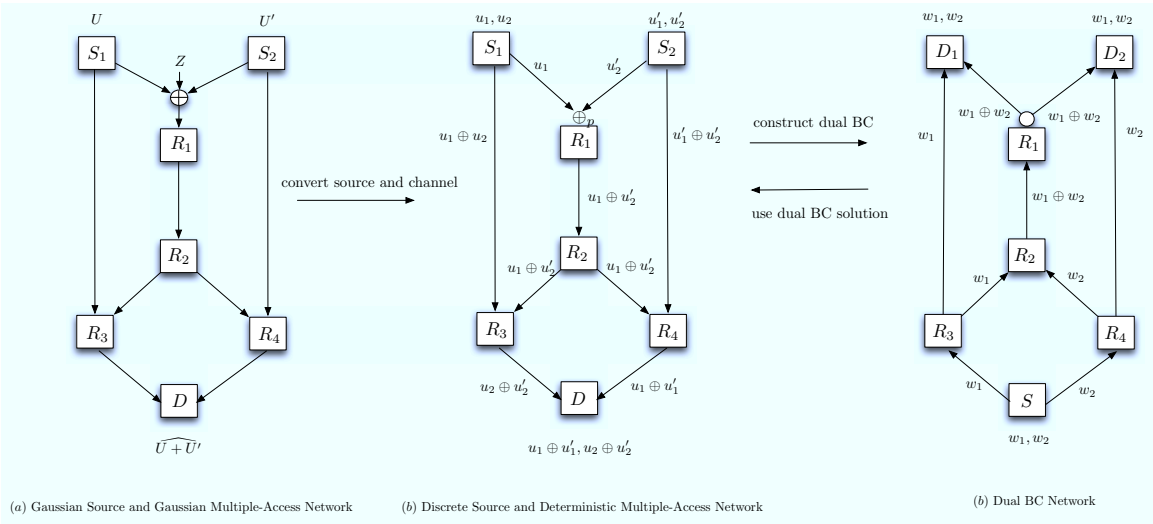

Figure 5.5. The goal is to send the sum of Gaussian sources U, U' across the multiple-access network in (a). All links are orthogonal with capacity $\log |\mathbb{F}_p|$ except for the wireless Gaussian MAC from S_1, S_2 to D. The Gaussian sources are converted to discrete sources through quantization and the Gaussian MAC is converted into a linear deterministic MAC through physical layer computation as shown in (b). The goal is to transmit the finite field sum of discrete source symbols across a linear deterministic network. The duality from Theorem 5.17 is applied and the solution from the dual broadcast network is used.

and receives output $Y_i[t] \in \mathbb{R}$ given by

$$Y_j[t] = \sum_{i \in \Omega} h_{i,j} X_i[t] + Z_j[t] \tag{5.23}$$

where $h_{i,j} \in \mathbb{R}$ is the fading coefficient on the link from node N_i to node N_j and $\{Z_j[t]\}_{t=1}^n$ is an independent Gaussian process with unit variance.

Definition 5.28 (Cut)**.** We call a subset $\Gamma \subseteq \Omega$ a cut. The value of cut Γ is given by

$$C_\Gamma^{\text{GAUSS-MAC}} = \max_{p(x_\Omega): \mathbb{E}[X_{i,j}^2] \leq \frac{\mathsf{SNR}}{b_i}} I(X_\Gamma; Y_{\Gamma^c} | X_{\Gamma^c}). \tag{5.24}$$

We define the degree of the network, which represents the maximum number of users in any given multiple-access channel in the network. This will be useful in characterizing the achievable distortion.

Definition 5.29 (Degree)**.** The degree of the MAC with output $Y_j = \sum_{i \in \Omega} h_{i,j} X_i + Z_i$ is given by

$$d_j = \sum_{i \in \Omega} \mathbf{1}\left\{h_{i,j} \neq 0\right\} \tag{5.25}$$

The degree of the network $\mathcal{N}_{\text{GAUSS-MAC}}$ is $d = \max_{j \in \Omega} d_{\text{j}}$ and represents the maximum degree over all MACs in the network.

5.2.2 Computation over Networks of Gaussian MACs

We provide the details to the problem of sending the sum of Gaussian sources across a class of relay networks with Gaussian MACs described in the previous section.

Definition 5.30 (Source Information)**.** Node N_i where $i \in \mathcal{S}$ observes a length k sequence $U_i^k = (U_{i,1}, \cdots U_{i,k})$ where $U_{i,j} \sim$ i.i.d $\mathcal{N}(0,1)$. We assume that $U_i^k = 0$ for all $i \in \mathcal{S}^c$.

Definition 5.31 (Encoders)**.** At time t, node N_i encodes its received signal $\{Y_i[j]\}_{j=1}^{t-1}$ and information U_i^k into b_i length n codewords using the mapping:

$$\mathcal{E}_{i,t} : \mathbb{R}^k \times \mathbb{R}^{(t-1)} \to \mathbb{R}^{b_i} \tag{5.26}$$

$$X_i[t] = \mathcal{E}_{i,t}\left(U_i^k, \{Y_i[j]\}_{j=1}^{t-1}\right) \tag{5.27}$$

We assume that $n = qk$ for some $q \in \mathbb{Q} \backslash \{0\}$. Each $\{X_{i,j}[t]\}_{t=1}^n$ must satisfy the power constraint given in (5.22).

Definition 5.32 (Decoder)**.** The destination observes $\{Y_i[t]\}_{t=1}^n$ for $i \in \mathcal{G}$ and recovers the estimate $\hat{V}^k$ for the sum of the source observations $V^k = \sum_{i=1}^m U_i^k$ using decoder $\mathcal{G}$:

$$\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^k \tag{5.28}$$

$$\hat{V}^k = \mathcal{G}(\{Y_i[t]\}_{t=1}^n). \tag{5.29}$$

The quality of the estimate $\hat{V}^k$ is measured by the standard squared error distortion:

$$D = \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}\left[\left(V_i - \hat{V}_i\right)^2\right]. \tag{5.30}$$

Definition 5.33 (Achievable Distortion)**.** A distortion D is achievable if all for $\epsilon > 0$ and large n, there exists encoders $\{\mathcal{E}_{i,t}\}_{t=1}^n$ $\forall i \in \Omega$ satisfying power constraint SNR and a decoder $\mathcal{G}$ that outputs an estimate $\hat{V}^k$ such that

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{E}\left[\left(V_i - \hat{V}_i\right)^2\right] \leq D + \epsilon. \tag{5.31}$$

5.2.3 Illustrative Examples

Our approach separates the physical and network layers and performs computation over the physical layer using nested lattice codes and computation over the network layer using network codes. We illustrate the gain from performing computing in both layers through two simple examples.

I **Network Layer Computation** We show the importance of computation in the network layer instead of routing by considering the example in Figure 5.6. In this two-hop network, each Gaussian point to point channel has signal to noise ratio SNR. The goal is to transmit the sum of independent Gaussian sources U and U' to the destination. Each point-to-point Gaussian channel can be trivially converted to a finite field link through the use of channel codes. The Gaussian sources can also be converted to finite field symbols through lattice-based quantization. The destination can reconstruct an estimate for the real sum $U + U'$ based on the finite field sum $U_Q \oplus_p U'_Q$ as shown in [23, 78]. The relay receives the quantized symbols U_Q, U'_Q. It is advantageous to compute and retransmit the finite field sum $U_Q \oplus_p U'_Q$ as shown in Figure 5.6 (a). However, if routing is used instead, then the relay has to time share between sending U_Q, U'_Q as shown in (b). This results in inefficiency of channel usage and causes the Gaussian symbols U, U' to be estimated individually instead. We compare the achievable distortions of the different approaches with the cut-set lower bounds:

$$D_{\text{COMP}} = \frac{4\sigma^2}{1+\mathsf{SNR}}, \quad D_{\text{ROUTING}} = \frac{2\sigma^2}{\sqrt{1+\mathsf{SNR}}}, \quad D_{\text{CUT-SET}} \geq \frac{\sigma^2}{1+\mathsf{SNR}} \tag{5.32}$$

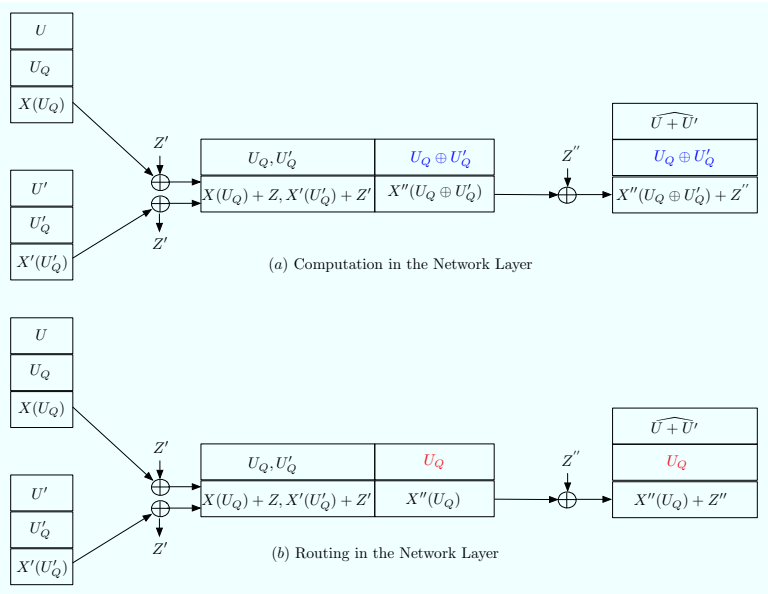

Figure 5.6. Each Gaussian point-to-point channel is converted into a finite field link. The Gaussian sources U, U' are quantized into finite-field symbols U_Q, U'_Q. In (a), the relay receives U_Q, U'_Q and computes $U_Q \oplus_p U'_Q$ for retransmission. Based on the finite field sum $U_Q \oplus_p U'_Q$, an estimate for the $U + U'$ is reconstructed. The achievable distortion D_{COMP} in (5.32) is within a constant gap of the cut-set lower bounds $D_{\text{cut-set}}$ in (5.32). If routing is performed instead as in (b), the relay has to time share between sending symbols U_Q, U'_Q and an estimate for each Gaussian source U, U' is formed. As a result, the achievable distortion D_{ROUTING} in (5.32) can be arbitrarily large from the cut-set lower bounds.

and find that D_{COMP} is within a constant fraction of $D_{\text{CUT-SET}}$ while D_{ROUTING} can be an arbitrary fraction larger as $\mathsf{SNR} \to \infty$.

II **Physical Layer Computation**. We show the importance of physical layer computation in transmitting the sum of independent Gaussian sources across a Gaussian MAC with per user signal to noise ratio SNR as shown in Figure 5.7 and in [27]. As in Example I, the Gaussian sources can be converted to finite field symbols through lattice-based quantization and the destination can reconstruct an estimate for the real sum $U + U'$ based on the finite field sum $U_Q \oplus_p U'_Q$. The key in this example is the conversion of the Gaussian MAC to a finite-field channel. Algebraically structured lattice codes are used to communicate the finite-field sum $U_Q \oplus_p U'_Q$ across the Gaussian MAC in (a). This process can be viewed as converting the Gaussian MAC into a linear deterministic MAC. If the individual quantized symbols U_Q, U'_Q are sent directly instead as in (b), this can be viewed as converting the Gaussian MAC into two finite field links as shown. However, the rate on each link is approximately half the rate of the deterministic MAC in (a). Furthermore, this scheme causes the estimates for U and U' to be formed separately since both the quantized source symbols are recovered by the destination. We compare the achievable distortion of the different approaches with the cutset lower bounds:

$$D_{\text{COMP}} = \frac{4\sigma^2}{\mathsf{SNR}}, \quad D_{\text{NO-COMP}} = \frac{2\sigma^2}{\sqrt{1+\mathsf{SNR}}}, \quad D_{\text{CUT-SET}} \geq \frac{\sigma^2}{1+\mathsf{SNR}}. \tag{5.33}$$

and find that D_{COMP} is within a constant fraction of $D_{\text{CUT-SET}}$ while $D_{\text{NO-COMP}}$ can be an arbitrary fraction larger as $\mathsf{SNR} \to \infty$.

5.2.4 Upper and Lower Bounds on Distortion

Previously, the gain from computation in the physical and the network layer seen in two simple networks. Here, we apply our approach to a network of Gaussian MACs and characterize the achievable distortion for computing the sum of Gaussian sources across the network in Theorem 5.34. The achievable distortion is compared with the cut-set lower bounds in Theorem 5.35 and found to be within a constant ratio of the optimal performance in Corollary 5.36.

Theorem 5.34. *The achievable distortion for sending the sum of Gaussian sources with variance σ^2 across the network $\mathcal{N}_{\text{GAUSS-MAC}}$ with nodes Ω and degree d satisfies*

$$D_{\text{ACHIEVABLE}} \leq \sigma^2 m^2 2^{2q\alpha} \max_{i \in \mathcal{S}} 2^{-2q\left(\min_{\Gamma: i \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}}\right)} \tag{5.34}$$

where $q = \frac{n}{k}$ is the number of channel uses per source symbol, $C_\Gamma^{\text{GAUSS-MAC}}$ is the value of cut Γ given in Definition 5.28, and α is the constant:

$$\alpha = |\Omega|((d+1)\log(d+2) + 2\log d + 1). \tag{5.35}$$

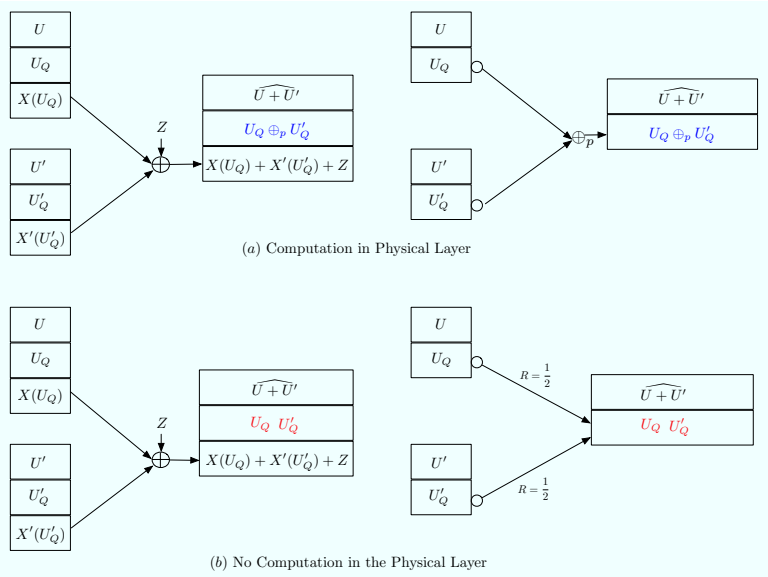

Figure 5.7. Similar to example I, the Gaussian sources are quantized and the destination recovers an estimate for $U + U'$ is based on the finite field sum $U_Q \oplus_p U'_Q$. In (a), computation is performed in the physical layer using nested lattice codes. This can be viewed as converting the Gaussian MAC is converted into a linear deterministic MAC. The achievable distortion D_{COMP} in (5.33) is within a constant gap of the cut-set lower bound $D_{\text{cut-set}}$ in (5.33). Instead of performing computation in the physical layer, both symbols U_Q, U'_Q are sent to the destination in (b). This can be viewed as converting the Gaussian MAC into two finite field links each with approximately only half the sum rate of the deterministic MAC in (a). The achievable distortion $D_{\text{NO-COMP}}$ in (5.33) can be arbitrarily large from the cut-set lower bounds.

The proof of Theorem 5.34 is given in Sections 5.2.7. Next, we provide the cut-set lower bound on distortion.

Theorem 5.35. *The optimal distortion D_{OPT} for sending the sum of Gaussian sources with variance σ^2 across the network $\mathcal{N}_{GAUSS\text{-}MAC}$ satisfies the following cut-set bound:*

$$D_{OPT} \geq \sigma^2 \max_{i \in \mathcal{S}} 2^{-2q\left(\min_{\Gamma \subseteq \Omega : i \in \Gamma} C_{\Gamma}^{GAUSS\text{-}MAC}\right)}. \tag{5.36}$$

Proof. See Appendix E. □

Corollary 5.36. *The ratio between the achievable distortion $D_{ACHIEVABLE}$ and the optimal distortion D_{OPT} is bounded by $\frac{D_{ACHIEVABLE}}{D_{OPT}} \leq 2^{2(\log m + q\alpha)}$ where α is the constant in (5.35).*

Proof. Follows directly from Theorem 5.34 and Theorem 5.35. □

Corollary 5.36 shows that the ratio between the achievable and optimal distortion can be bounded by the number and the degree of the network independent from the network topology

Remark 5.37. The extension beyond Gaussian sources is feasible. Theorem 5.35 can be generalized to non-Gaussian sources using Shannon's lower bound [1] and Theorem 5.34 can be generalized as long as the source is contained inside a ball of an appropriate radius in L_2 sense.

The rest of this Section is devoted to providing the underlying tools and the proof of Theorem 5.34, which involves channel coding and source quantization. First, nested-lattice code constructions are given in Section 5.2.5. The use of nested-lattices for channel coding is shown in Section 5.2.6 and for source quantization in Section 5.2.7. It is shown that the Gaussian sources are mapped to symbols on the finite field using nested-lattice based quantization. The Gaussian network is converted into a linear deterministic network using nested-lattice channel codes. The finite-field sum of the quantized source symbols are transmitted across the converted linear deterministic network. An estimate for the sum of the Gaussian sources is reconstructed based on the finite-field sum of the quantized points.

5.2.5 Code Construction: Nested Lattices

Nested-lattice codes were proposed in [16, 17, 18, 22] and used for computation in [31] and quantization in [23, 78]. We use a different nested-lattice construction than those in [31]. Instead of scaling and rotating the coarse lattice to generate the fine lattice as in [31], we concatenate the generator matrices as shown in (5.37). We find this construction provides eases the analysis and provides a simpler mapping between the nested-lattice and the finite field.

We first show the construction of nested-lattices. We use the same notation as that in [31] and use $\cdot$ to denote multiplication in the finite field $\mathbb{F}_p$. A pair of nested lattices (Λ_C, Λ_F) is constructed as follows:

$$\Lambda_C = \left\{p^{-1}(\mathbf{G}_1 \cdot \mathbf{w}_1) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}\right\} \tag{5.37}$$

$$\Lambda_F = \left\{p^{-1}(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus_p \mathbf{G}_2 \cdot \mathbf{w}_2) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2}\right\} \tag{5.38}$$

where $\mathbb{F}_p$ is a finite-field with p prime, $\mathbf{G}_1 \in \mathbb{F}_p^n \times \mathbb{F}_p^{k_1}, \mathbf{G}_2 \in \mathbb{F}_p^n \times \mathbb{F}_p^{k_2}$ with each element drawn i.i.d uniformly from $\mathbb{F}_p$. If p, n, k_1, k_2 are chosen to scale appropriately, the matrix $[\mathbf{G}_1, \mathbf{G}_2]$ becomes full rank with high probability and the lattices Λ_C, Λ_F are simultaneously good for covering, quantization and AWGN coding (see [17] for definitions and proofs).

Some definitions based on those in [16] are given in the sequel followed by key properties of the nested lattices.

Definition 5.38 (Coset)**.** Given a lattice Λ, a coset of Λ in $\mathbb{R}^n$ is any translated version of it. For example, for any $\mathbf{x} \in \mathbb{R}^n$, the set $\{\lambda + \mathbf{x} : \lambda \in \Lambda\}$ is a coset of Λ.

Definition 5.39 (Fundamental Voronoi Region)**.** The fundamental Voronoi region $\mathcal{V}$ of Λ is a subset of $\mathbb{R}^n$ that contains the minimum Euclidean norm coset representatives of the cosets of Λ. Every $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as

$$\mathbf{x} = \lambda + \mathbf{r} \tag{5.39}$$

with $\lambda \in \Lambda, \mathbf{r} \in \mathcal{V}$, where $\lambda = Q_{\mathcal{V}}(\mathbf{x})$ is a nearest neighbor of $\mathbf{x}$ in Λ, and $\mathbf{r} = \mathbf{x} - Q_{\mathcal{V}}(\mathbf{x})$ is the error.

Lemma 5.40. $|\Lambda_F \cap \mathcal{V}_C| = p^{k_2}$

Proof. Using the fact that we can rewrite any $\mathbf{x} \in \mathbb{R}^n$ as $\mathbf{x} = p\mathbf{z} + \mathbf{r}$ where $\mathbf{z} \in \mathbb{Z}^n$ and $\mathbf{r} \in \mathbb{F}_p^n$, we have that

$$\Lambda_F = \left\{p^{-1}(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}^{k_2}\right\} \tag{5.40}$$

$$= \left\{p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 + \mathbf{G}_2 \cdot \mathbf{w}_2 + p\mathbf{z}\right) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}^{k_2}\right\} \tag{5.41}$$

$$= \left\{p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 + \mathbf{G}_2 \cdot \mathbf{w}_2\right) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}^{k_2}\right\} \tag{5.42}$$

$$= \left\{p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1\right) + \mathbb{Z}^n + p^{-1}\left(\mathbf{G}_2 \cdot \mathbf{w}_2\right) : \mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}^{k_2}\right\} \tag{5.43}$$

$$= \Lambda_C + \left\{p^{-1}\left(\mathbf{G}_2 \cdot \mathbf{w}_2\right) : \mathbf{w}_2 \in \mathbb{F}^{k_2}\right\} \tag{5.44}$$

The above suggest that Λ_F is composed of p^{k_2} cosets of Λ_C. We show that these cosets are unique. Assume there exists $\mathbf{c}, \mathbf{c}' \in \Lambda_F$ with $\mathbf{c} = \mathbf{c}'$. It can be shown that

$$\mathbf{c} - \mathbf{c}' = p^{-1}\left(\left(\mathbf{G}_1 \cdot \left(\mathbf{w}_1 - \mathbf{w}_1'\right)\right) \oplus \left(\mathbf{G}_2\left(\mathbf{w}_2 - \mathbf{w}_2'\right)\right)\right) + \mathbf{z} \tag{5.45}$$

where $\mathbf{z} \in \mathbb{Z}^n$. Since the $n \times (k_1 + k_2)$ matrix $[\mathbf{G}_1, \mathbf{G}_2]$ is full rank and

$$p^{-1} \left(\mathbf{G}_1 \cdot (\mathbf{w}_1 - \mathbf{w}_1') \oplus \mathbf{G}_2 \cdot (\mathbf{w}_2 - \mathbf{w}_2') \right) \in \left\{ 0, \frac{1}{p}, \cdots, \frac{p-1}{p} \right\}^n \tag{5.46}$$

it follows that $\mathbf{c} - \mathbf{c}' = \mathbf{0}$ implies that

$$[\mathbf{w}_1 - \mathbf{w}_1', \mathbf{w}_2 - \mathbf{w}_2']^T \mod p = \mathbf{0} \tag{5.47}$$

Hence, we have

$$\mathbf{c} = \mathbf{c}' \Rightarrow \mathbf{w}_2 = \mathbf{w}_2' \tag{5.48}$$

and it follows that

$$\mathbf{w}_2 \neq \mathbf{w}_2' \Rightarrow \mathbf{c} \neq \mathbf{c}' \tag{5.49}$$

This shows that each $\mathbf{w}_2 \in \mathbb{F}_p^{k_2}$ generates a unique coset. As a result, $|\Lambda_F \cap \mathcal{V}_C| = p^{k_2}$. □

Definition 5.41. By construction, for each $\mathbf{c} \in \Lambda_F$, we have that

$$\mathbf{c} = p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \right) + \mathbf{z} \tag{5.50}$$

for some $\mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2}$, and $\mathbf{z} \in \mathbb{Z}^n$. We define the mapping $w_2 : \Lambda_F \cap \mathcal{V}_C \to \mathbb{F}_p^{k_2}$ where

$$w_2(\mathbf{c}) = \mathbf{w}_2. \tag{5.51}$$

Lemma 5.42. *The mapping w_2 is a group isomorphism from $(\Lambda_F \cap \mathcal{V}_C, \mod \Lambda_C)$ to $(\mathbb{F}_p, \mod p)$.*

Proof. We first show that w_2 is an injection. Combining this fact with Lemma 5.40, it follows that w_2 is a bijection. We assume that there exists $\mathbf{c}, \mathbf{c}' \in \Lambda_F \cap \mathcal{V}_C$ with $\mathbf{c} \neq \mathbf{c}'$ and $\mathbf{w}_2 = \mathbf{w}_2'$. It follows that

$$\begin{aligned}
\mathbf{c} - \mathbf{c}' &= p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \right) + \mathbf{z} - p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2' \right) + \mathbf{z}' && (5.52)\\
&= p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \right) - p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2' \right) + \mathbf{z} + \mathbf{z}' && (5.53)\\
&= p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \right) + p^{-1} \left(\mathbf{G}_1 \cdot -\mathbf{w}_1' \oplus \mathbf{G}_2 \cdot -\mathbf{w}_2' \right) + \mathbf{z} + \mathbf{z}' && (5.54)\\
&= p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_1 \cdot -\mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_2 \cdot -\mathbf{w}_2' + p\mathbf{z}'' \right) + \mathbf{z} + \mathbf{z}' && (5.55)\\
&= p^{-1} \left(\mathbf{G}_1 \cdot (\mathbf{w}_1 \oplus -\mathbf{w}_1') \oplus \mathbf{G}_2 \cdot (\mathbf{w}_2 \oplus -\mathbf{w}_2') \right) + \mathbf{z}'' + \mathbf{z} + \mathbf{z}' && (5.56)\\
&= p^{-1} \left(\mathbf{G}_1 \cdot (\mathbf{w}_1 \oplus -\mathbf{w}_1') \right) + \mathbf{z}'' + \mathbf{z} + \mathbf{z}' && (5.57)
\end{aligned}$$

Hence, $\mathbf{c} - \mathbf{c}' \in \Lambda_C$. We define the set Γ as follows

$$\Gamma = \left\{ \mathbf{c} - \mathbf{c}' : \mathbf{c}, \mathbf{c}' \in \Lambda_F \cap \mathcal{V}_C, \mathbf{c} \neq \mathbf{c}' \right\}. \tag{5.58}$$

We show that $\Gamma \cap \Lambda_C = \{\mathbf{0}\}$. Without loss of generality, we assume that there exists $\mathbf{c}, \mathbf{c}' \in \Lambda_F \cap \mathcal{V}_C$ with $\mathbf{c} \neq \mathbf{c}'$ and $\mathbf{c} - \mathbf{c}' \in \Lambda_C \backslash \{\mathbf{0}\}$. We can rewrite $\mathbf{c}$ in two different ways:

$$\mathbf{c} = \mathbf{0} + \mathbf{c} \tag{5.59}$$
$$\mathbf{c} = \mathbf{c} - \mathbf{c}' + \mathbf{c}' \tag{5.60}$$

where $\mathbf{0}, \mathbf{c} - \mathbf{c}' \in \Lambda_C$ with $\mathbf{c} - \mathbf{c}' \neq \mathbf{0}$ and $\mathbf{c} \in \mathcal{V}_C$. This contradicts the definition of the fundamental voronoi region. Hence, the mapping w_2 is injective.

So far, we have shown that w_2 is a bijection. We show that the following are true.

$$\forall\ \mathbf{c} \in \Lambda_F, w_2(\mathbf{c}) = w_2(\mathbf{c} \mod \Lambda_C) \tag{5.61}$$
$$\forall\ \mathbf{c}, \mathbf{c}' \in \Lambda_F, w_2(\mathbf{c} + \mathbf{c}') = w_2(\mathbf{c}) \oplus w_2(\mathbf{c}'). \tag{5.62}$$

We first show (5.61). Given $\mathbf{c} \in \Lambda_F$, we can rewrite it as follows:

$$\mathbf{c} = \lambda_C + (\mathbf{c} \mod \Lambda_C) \tag{5.63}$$

for some $\lambda_C \in \Lambda_C$. By construction, it follows that

$$\lambda_C = p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1\right) + \mathbf{z} \tag{5.64}$$
$$\mathbf{c} \mod \Lambda_C = p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}' \tag{5.65}$$

for some $\mathbf{w}_1, \mathbf{w}_1' \in \mathbb{F}_p^{k_1}, \mathbf{w}_2' \in \mathbb{F}_p^{k_2}, \mathbf{z}, \mathbf{z}' \in \mathbb{Z}^n$. It follows that

$$\begin{aligned}
\mathbf{c} &= \lambda_C + (\mathbf{c} \mod \Lambda_C) && (5.66)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1\right) + \mathbf{z} + p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}' && (5.67)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1\right) + p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z} + \mathbf{z}' && (5.68)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 + \mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z} + \mathbf{z}' && (5.69)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2' + p\mathbf{z}''\right) + \mathbf{z} + \mathbf{z}' && (5.70)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot (\mathbf{w}_1 \oplus \mathbf{w}_1') \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}'' + \mathbf{z} + \mathbf{z}' && (5.71)
\end{aligned}$$

for some $\mathbf{z}'' \in \mathbb{Z}^n$. Hence, we have that $w_2(\mathbf{c}) = \mathbf{w}_2' = w_2(\mathbf{c} \mod \Lambda_C)$. We now show (5.62). Given $\mathbf{c}, \mathbf{c}' \in \Lambda_F$, it follows that

$$\mathbf{c} = p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2\right) + \mathbf{z} \tag{5.72}$$
$$\mathbf{c}' = p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}' \tag{5.73}$$

for some $\mathbf{w}_1, \mathbf{w}_1' \in \mathbb{F}_p^{k_1}, \mathbf{w}_2, \mathbf{w}_2' \in \mathbb{F}_p^{k_2}, \mathbf{z}, \mathbf{z}' \in \mathbb{Z}^n$. We note that $\mathbf{c} + \mathbf{c}' \in \Lambda_F$. It follows that

$$\begin{aligned}
\mathbf{c} + \mathbf{c}' &= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2\right) + \mathbf{z} + p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}' && (5.74)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2\right) + p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z} + \mathbf{z}' && (5.75)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2' + p\mathbf{z}''\right) + \mathbf{z} + \mathbf{z}' && (5.76)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2'\right) + \mathbf{z}'' + \mathbf{z} + \mathbf{z}' && (5.77)\\
&= p^{-1}\left(\mathbf{G}_1 \cdot (\mathbf{w}_1 \oplus \mathbf{w}_1') \oplus \mathbf{G}_2 \cdot (\mathbf{w}_2 \oplus \mathbf{w}_2')\right) + \mathbf{z}'' + \mathbf{z} + \mathbf{z}' && (5.78)
\end{aligned}$$

Hence, $w_2\left(\mathbf{c} + \mathbf{c}'\right) = \mathbf{w}_2 \oplus \mathbf{w}_2'$ □

5.2.6 Proof of Theorem 3: Channel Coding

An achievable computation rate (as given in Definition 5.6) for sending the finite-field sum of discrete sources (as described in Definition 5.3) across a network of Gaussian MACs is given in Lemma 5.43 below. The idea is to use nested-lattice channel codes to converted the Gaussian network $\mathcal{N}_{\text{GAUSS-DET}}$ into a linear deterministic network $\mathcal{N}_{\text{DET-MAC}}$ in Section 5.1.1. Applying the duality relation, network coding is used to transmit the finite-field sum of discrete sources across the linear deterministic network as in Section 5.1.

Lemma 5.43. *Consider a network of Gaussian MACs $\mathcal{N}_{GAUSS\text{-}MAC}$ with discrete source observations $U_1^k, \ldots, U_m^k$ where each $U_{i,j}$ is drawn i.i.d uniformly from some prime-sized finite field $\mathbb{F}_p$. For all $\epsilon > 0$ and n, p large, there exists encoders $\{\mathcal{E}_{i,t}\}_{t=1}^n \forall i \in \Omega$ and a decoder $\mathcal{G}$ that produces an estimate $\hat{V}^k \in \mathbb{F}_p^k$ such that*

$$\Pr(\hat{V}^k \neq U_1^k \oplus_p \cdots \oplus_p U_m^k) < \epsilon \tag{5.79}$$

as long as the computation rate $R = \frac{k}{n} \log p$ satisfies

$$R < \min_{i \in \mathcal{S}} \min_{\Gamma \subseteq \Omega : i \in \Gamma} C_\Gamma^{GAUSS\text{-}MAC} - \alpha \tag{5.80}$$

where $\alpha = |\Omega|((d+1)\log(d+2) + 2\log d + 1)$.

Proof. The Gaussian network $\mathcal{N}_{\text{GAUSS-MAC}}$ can be converted into a linear deterministic network $\mathcal{N}_{\text{DET-MAC}}$ such that each cut value in $\mathcal{N}_{\text{DET-MAC}}$ is within α of the corresponding cut in $\mathcal{N}_{\text{GAUSS-MAC}}$. Lemma 5.43 then follows by applying Corollary 5.19. We first show the conversion of a single m-user Gaussian MAC into a linear deterministic MAC. The Gaussian MAC contains only source and destination nodes, i.e $\Omega = \mathcal{S} \cup \mathcal{D}$ where $\mathcal{S} = \{1, \ldots, m\}$ and $\mathcal{D} = \{m+1\}$. The channel gain from source node i to the destination node is given by h_i. Without loss of generality, we assume that $h_i > 0$ and $\mathsf{SNR} = 1$. We first assume that the channel gains satisfy the following inequalities:

$$h_1^2 \geq 1, \quad h_i^2 - h_{i-1}^2 \geq (m - (i-2))h_{i-1}^2 + h_{i-2}^2 + \cdots + h_1^2 + 1 \quad \text{for} \quad i > 1 \tag{5.81}$$

We use a superposition nested-lattice scheme similar to that in [31] but with nested generator matrices as described in Section 5.2.5 instead. We choose m nested lattice pairs $(\Lambda_{C_i}, \Lambda_{F_i})$ given by

$$\Lambda_{C_i} = \left\{ p^{-1} \left(\mathbf{G}_{i,1} \cdot \mathbf{w}_1 \right) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_{1,i}} \right\} \tag{5.82}$$

$$\Lambda_{F_i} = \left\{ p^{-1} \left(\mathbf{G}_{i,1} \cdot \mathbf{w}_1 \oplus_p \mathbf{G}_{i,2} \cdot \mathbf{w}_2 \right) + \mathbb{Z}^n : \mathbf{w}_1 \in \mathbb{F}_p^{k_{1,i}}, \mathbf{w}_2 \in \mathbb{F}_p^{k_i} \right\}. \tag{5.83}$$

where each element of $\mathbf{G}_{i,1}$ and $\mathbf{G}_{i,2}$ is drawn i.i.d uniformly from $\mathbb{F}_p$. The variances for the lattices $\Lambda_{C_i}, \Lambda_{F_i}$ are set to be:

$$\sigma^2(\Lambda_{C_i}) = h_i^2 - h_{i-1}^2, \quad \sigma^2(\Lambda_{F_i}) = (m - (i-2))h_{i-1}^2 + h_{i-2}^2 + \cdots + h_1^2 + 1 \tag{5.84}$$

We define the set of one-to-one mappings $w_{2,i} : \Lambda_{F_i} \cap \mathcal{V}_{C_i} \to \mathbb{F}_p^{k_i}$ for $i = 1, \ldots, m$ along the lines of Definition 5.51. User i splits message $\mathbf{w}_i$ into i parts:

$$\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)}, \cdots, \mathbf{w}_i^{(i)} \quad \text{where} \quad \mathbf{w}_i^{(j)} \in \mathbb{F}^{k_j} \quad \text{for} \quad j = 1, \ldots, i \tag{5.85}$$

The j^{th} part of message $\mathbf{w}_i$ is mapped to a lattice point using the mapping $w_{2,j}$: $\mathbf{c}_i^{(j)} = w_{2,j}^{-1}(\mathbf{w}_i^{(j)})$. The lattice points are added together, dithered with $\mathbf{d}_i^{(j)}$ that is drawn uniformly from $\mathcal{V}_{C_j}$ to get the resulting vector:

$$\mathbf{x}_i^{(j)} = \left(\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j}\right) \tag{5.86}$$

User i transmits the signal $\mathbf{x}_i$ given by

$$\mathbf{x}_i = \frac{1}{h_i} \sum_{j=1}^{i} \mathbf{x}_i^{(j)} = \frac{1}{h_i} \sum_{j=1}^{i} \left(\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j}\right) \tag{5.87}$$

The destination receives:

$$\mathbf{y} = \sum_{i=1}^{m} h_i \mathbf{x}_i + \mathbf{z} \tag{5.88}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{i} \mathbf{x}_i^{(j)} + \mathbf{z} \tag{5.89}$$

$$= \sum_{j=1}^{m} \sum_{i=j}^{m} \mathbf{x}_i^{(j)} + \mathbf{z} \tag{5.90}$$

$$= \sum_{j=1}^{m} \sum_{i=j}^{m} \left(\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j}\right) + \mathbf{z} \tag{5.91}$$

To recover $\mathbf{w}_m^{(m)}$, the destination computes:

$$\left(\mathbf{y} - \mathbf{d}_m^{(m)}\right) \mod \Lambda_{C_m} = \left(\sum_{j=1}^{m} \sum_{i=j}^{m} \left(\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j}\right) + \mathbf{z} - \mathbf{d}_m^{(m)}\right) \mod \Lambda_{C_m} \tag{5.92}$$

$$= \left(\mathbf{c}_m^{(m)} + \sum_{j=1}^{m-1} \sum_{i=j}^{m} \left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j} + \mathbf{z}\right) \mod \Lambda_{C_m} \tag{5.93}$$

Let the noise $\mathbf{z}^{(m)} = \sum_{j=1}^{m-1}\sum_{i=j}^{m}\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j} + \mathbf{z}$. The destination further computes:

$$Q_{\Lambda_{F_m}}\left((\mathbf{y} - \mathbf{d}_m^{(m)}) \mod \Lambda_{C_m}\right) \mod \Lambda_{C_m} = Q_{\Lambda_{F_m}}\left((\mathbf{c}_m^{(m)} + \mathbf{z}^{(m)}) \mod \Lambda_{C_m}\right) \mod \Lambda_{C_m} \tag{5.94}$$

$$= Q_{\Lambda_{F_m}}\left(\mathbf{c}_m^{(m)} + \mathbf{z}^{(m)}\right) \mod \Lambda_{C_m} \tag{5.95}$$

We define the event: $\mathcal{E}^{(m)} = \left\{Q_{\Lambda_{F_m}}\left(\mathbf{c}_m^{(m)} + \mathbf{z}^{(m)}\right) \mod \Lambda_{C_m} = \mathbf{c}_m^{(m)} \mod \Lambda_{C_m}\right\}$. Under event $\mathcal{E}_d$, we can reliably recover the message $\mathbf{w}_m^{(m)}$ by the inverse mapping w_2^{-1}:

$$\mathbf{w}_m^{(m)} = w_{2,m}^{-1}(\mathbf{c}_m^{(m)} \mod \Lambda_{C_m}) \tag{5.96}$$

as long as $\frac{k_m}{n}\log m \leq \frac{1}{2}\log(\text{SINR}_m)$ where $\text{SINR}_m = \frac{\sigma^2(\Lambda_{C_m})}{\sigma^2(\Lambda_{F_m})} = \frac{h_m^2 - h_{m-1}^2}{2h_{m-1}^2 + h_{m-2}^2 + \cdots + h_1^2 + 1}$

If message $w_m^{(m)}$ is successfully decoded, then $\mathbf{x}_m^{(m)}$ is subtracted from the received signal $\mathbf{y}$. The resulting signal is given by

$$\mathbf{y}^{(m-1)} = \mathbf{y} - \mathbf{x}_m^{(m)} = \sum_{j=1}^{m-1}\sum_{i=j}^{m}\mathbf{x}_i^{(j)} + \mathbf{z} \tag{5.97}$$

The receiver then computes

$$\left(\mathbf{y}^{(m-1)} - \mathbf{d}_{m-1}^{(m-1)} - \mathbf{d}_m^{(m-1)}\right) \mod \Lambda_{C_{m-1}} \tag{5.98}$$

$$= \left(\sum_{j=m-1}^{m}\left(\mathbf{x}_j^{(m-1)} - \mathbf{d}_j^{(m-1)}\right) + \sum_{j=1}^{m-2}\sum_{i=j}^{m}\mathbf{x}_i^{(j)} + \mathbf{z}\right) \mod \Lambda_{C_{m-1}} \tag{5.99}$$

$$= \left(\sum_{j=m-1}^{m}\mathbf{c}_j^{(m-1)} + \sum_{j=1}^{m-2}\sum_{i=j}^{m}\mathbf{x}_i^{(j)} + \mathbf{z}\right) \mod \Lambda_{C_{m-1}} \tag{5.100}$$

Let the noise $\mathbf{z}^{(m-1)} = \sum_{j=1}^{m-2}\sum_{i=j}^{m}\left(\mathbf{c}_i^{(j)} + \mathbf{d}_i^{(j)}\right) \mod \Lambda_{C_j} + \mathbf{z}$. The destination further computes:

$$Q_{\Lambda_{F_{m-1}}}\left((\mathbf{y}^{(m-1)} - \sum_{j=m-1}^{m}\mathbf{d}_j^{(m-1)}) \mod \Lambda_{C_{m-1}}\right) \mod \Lambda_{C_{m-1}} \tag{5.101}$$

$$= Q_{\Lambda_{F_{m-1}}}\left((\sum_{j=m-1}^{m}\mathbf{c}_j^{(m-1)} + \mathbf{z}^{(m-1)}) \mod \Lambda_{C_{m-1}}\right) \mod \Lambda_{C_{m-1}} \tag{5.102}$$

$$= Q_{\Lambda_{F_{m-1}}}\left(\sum_{j=m-1}^{m}\mathbf{c}_j^{(m-1)} + \mathbf{z}^{(m-1)}\right) \mod \Lambda_{C_{m-1}} \tag{5.103}$$

Let $\mathcal{E}^{(m-1)}$ denote the event:

$$\mathcal{E}^{(m-1)} = \left\{ Q_{\Lambda_{F_{m-1}}} \left(\sum_{j=m-1}^{m} \mathbf{c}_j^{(m-1)} + \mathbf{z}^{(m-1)} \right) \bmod \Lambda_{C_{m-1}} = \left(\mathbf{c}_{m-1}^{(m-1)} + \mathbf{c}_m^{(m-1)} \right) \bmod \Lambda_{C_{m-1}} \right\} \tag{5.104}$$

Under event $\mathcal{E}^{(m)} \cap \mathcal{E}^{(m-1)}$, we can reliably recover the equation

$$\mathbf{w}_{m-1}^{(m-1)} \oplus \mathbf{w}_d^{(m-1)} = w_2^{-1} \left(\left(\mathbf{c}_{m-1}^{(m-1)} + \mathbf{c}_m^{(m-1)} \right) \mod \Lambda_{C_{m-1}} \right) \tag{5.105}$$

with computation rate $\frac{k_{m-1}}{n} \log p \leq \frac{1}{2} \log(\text{SINR}_{m-1})$ where $\text{SINR}_{m-1} = \frac{h_{m-1}^2 - h_{m-2}^2}{3h_{m-2}^2 + h_{m-3}^2 + \cdots + h_1^2 + 1}$.

Each of the remaining $m-2$ layers of the superposition code is decoded in a similar manner. As a result, the Gaussian MAC is converted to a linear deterministic MAC with source nodes $\mathcal{S}_{\text{DET}} = \{1, \ldots, m\}$ and destination node $\mathcal{D}_{\text{DET}} = \{m+1\}$. The transfer function from source node i to the destination is given by

$$\mathbf{H}_i = [\mathbf{I}_{R_1} \ldots \mathbf{I}_{R_i}, \mathbf{0} \ldots \mathbf{0}]^T \tag{5.106}$$

where R_i is the rate for the i^{th} layer of the superposition lattice code and is given by $R_i = \frac{1}{2} \log(\text{SINR}_i)$ with $\text{SINR}_i = \frac{h_i^2 - h_{i-1}^2}{(m-(i-2))h_{i-1}^2 + h_{i-2}^2 + \cdots + h_1^2 + 1}$. The overall transfer function of the linear deterministic multiple-access channel is given by

$$\mathbf{H} = [\mathbf{H}_1 \ldots \mathbf{H}_m]. \tag{5.107}$$

Let C_j^{DET} represent the value of the cut $\Gamma = \{i : i \in \mathcal{S}_{\text{DET}}, i \leq j\}$ in the linear deterministic network . It follows that:

$$C_j^{\text{DET}} = \sum_{i=1}^{j} R_i = \frac{1}{2} \sum_{i=1}^{j} \log(\text{SINR}_i) \tag{5.108}$$

At each step, it can be shown that $\frac{1}{n}\mathbb{E}[\|\mathbf{z}^{(i)}\|^2] \leq \sigma^2(\Lambda_{F_i})$ for $i = 1 \ldots m$. From [16, Theorem 5], it follows that

$$\Pr\left(\mathcal{E}^{(m)} \cap \mathcal{E}^{(m-1)} \cap \cdots \cap \mathcal{E}^{(1)}\right) = 1 - \sum_{i=1}^{m} \Pr(\mathcal{E}^{(i)}) \to 1 \quad \text{as } n \to \infty. \tag{5.109}$$

We now remove the assumptions on the channel gains in (5.81). Let $r_1 = \min_{h_i^2 \leq 1} i$ and define r_j for $j > 1$ recursively as follows:

$$r_j = \min_{h_i^2 \geq (m+2)h_{r_{j-1}}^2} i \quad \text{for } j > 1 \tag{5.110}$$

We terminate when r_j cannot be defined and let $r_{\text{MAX}} = \max_i r_i$. For each $j \in \{1, \cdots, m\}$, we define $t_j = \arg\max_{i:r_i \leq j} i$. Using a r_{MAX} layered superposition code, it can be shown that the cut values of the linear deterministic MAC in (5.108) become

$$C_j^{\text{DET}} = \sum_{i=1}^{t_j} \frac{1}{2} \log\left(\text{SINR}_{r_i}\right) \quad \text{for} \quad j = 1, \ldots, r_{\text{MAX}} \tag{5.111}$$

where $\text{SINR}_{r_i} \geq \frac{h_{r_i}^2 - h_{k_{r_i-1}}^2}{m h_{k_{r_i-1}}^2 + 1}$. The cut values of the linear deterministic MAC from (5.111) are bounded as follows

$$\begin{aligned}
C_j^{\text{DET}} &\geq \frac{1}{2} \log\left(h_{r_1}^2\right) + \sum_{i=2}^{t_j} \frac{1}{2} \log\left(\frac{h_{r_i}^2 - h_{r_{i-1}}^2}{m h_{r_{i-1}}^2 + 1}\right) && (5.112)\\
&\geq \frac{1}{2} \log\left(h_{r_1}^2\right) + \sum_{i=2}^{t_j} \frac{1}{2} \log\left(\frac{h_{r_i}^2 \frac{(m+1)}{(m+2)}}{(m+1) h_{r_{i-1}}^2}\right) && (5.113)\\
&= \frac{1}{2} \log\left(h_{r_{t_j}}^2 \left(\frac{1}{m+2}\right)^{t_j - 1}\right) && (5.114)\\
&\geq \frac{1}{2} \log\left(h_{r_{t_j}}^2 \left(\frac{1}{m+2}\right)^{m}\right) && (5.115)
\end{aligned}$$

Let C_j^{GAUSS} be the value of the cut $\Gamma = \{i : i \in \mathcal{S}, i \leq j\}$ in the Gaussian MAC. It follows that

$$\begin{aligned}
C_j^{\text{GAUSS}} &= \frac{1}{2} \log\left(1 + \left(\sum_{i=1}^{j} h_i\right)^2\right) && (5.116)\\
&\leq \frac{1}{2} \log\left(1 + \left(\sum_{i=1}^{t_j} \sum_{r_{i-1} \leq i < r_i} h_i + \sum_{i=k_{t_j}}^{j} h_i\right)^2\right) && (5.117)\\
&\leq \frac{1}{2} \log\left(1 + \left(j\sqrt{m+2} h_{r_{t_j}}\right)^2\right) && (5.118)\\
&\leq \frac{1}{2} \log\left(1 + m^2(m+2) h_{r_{t_j}}^2\right) && (5.119)\\
&\leq \frac{1}{2} \log\left(2m^2(m+2) h_{r_{t_j}}^2\right) && (5.120)
\end{aligned}$$

The gap between C_j^{GAUSS} and C_j^{DET} can be bounded as follows:

$$C_j^{\text{GAUSS}} - C_j^{\text{DET}} \leq \frac{1}{2}\log\left(2m^2(m+2)h_{r_{t_j}}^2\right) - \frac{1}{2}\log\left(h_{r_{t_j}}^2\left(\frac{1}{m+2}\right)^m\right) \tag{5.121}$$

$$= \log\left(2m^2(m+2)^{m+1}\right) \tag{5.122}$$

$$= (m+1)\log(m+2) + 2\log m + 1 \tag{5.123}$$

From the structure of the linear deterministic channel in (5.106), (5.107), it is sufficient to consider only the gaps $C_j^{\text{GAUSS}} - C_j^{\text{DET}}$ for $j = 1 \dots m$. We repeat the described conversion for every Gaussian MAC in the network. For a fixed set $\Gamma \subseteq \Omega$, we consider the cut between Γ and Γ^c. Since the network contains only Gaussian MACs, it is sufficient to consider the MIMO channel with input $\mathbf{x}_\Gamma$, output $\mathbf{y}_{\Gamma^c}$ and channel matrix $\mathbf{H}_{\Gamma,\Gamma^c}$. It follows that $\mathbf{y}_{\Gamma^c} = \mathbf{H}_{\Gamma,\Gamma^c}\mathbf{x}_\Gamma + \mathbf{z}_{\Gamma^c}$. Furthermore, the MIMO channel can be decomposed into a set of MACs with outputs:

$$y_j = \sum_{i:i\in\Gamma} h_{i,j}x_{i,j} + z_j \text{ for } j \in \Gamma^c \tag{5.124}$$

where $h_{i,j}$ are the channel coefficients. Let d_j be degree of the MAC with output y_j. The gap between $C_\Gamma^{\text{GAUSS-MAC}}$ and $C_\Gamma^{\text{DET-MAC}}$ is given by

$$C_\Gamma^{\text{GAUSS-MAC}} - C_\Gamma^{\text{DET-MAC}} \leq \sum_{j\in\Gamma^c}(d_j+1)\log(d_j+2) + 2\log d_j + 1 \tag{5.125}$$

$$\leq |\Gamma^c|((d+1)\log(d+2) + 2\log d + 1) \tag{5.126}$$

$$\leq |\Omega|((d+1)\log(d+2) + 2\log d + 1) \tag{5.127}$$

□

5.2.7 Proof of Theorem 5.34: Source Quantization

We illustrate the source quantization scheme[1] based on [23, 78] for the achievable distortion in Theorem 5.34 . First, the sources are quantized using a nested lattice code. As in (5.37), a pair of nested lattices (Λ_F, Λ_C) with the following construction is chosen:

$$\Lambda_C = \left\{p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1\right) + \mathbb{Z}^k : \mathbf{w}_1 \in \mathbb{F}_p^{k_1'}\right\} \tag{5.128}$$

$$\Lambda_F = \left\{p^{-1}\left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2\right) + \mathbb{Z}^k : \mathbf{w}_1 \in \mathbb{F}_p^{k_1'}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2'}\right\} \tag{5.129}$$

[1] If only quantization is concerned, other schemes may outperform the proposed approach. However, in the absence of a bijection between the quantization points and the finite field elements, it is unclear how to transmit the quantized points through networks with arbitrary topologies.

where each element of $\mathbf{G}_1 \in \mathbb{F}_p^{k \times k_1'}$ and $\mathbf{G}_2 \in \mathbb{F}_p^{k \times k_2'}$ is drawn i.i.d uniformly from $\mathbb{F}_p$. The lattice variances are chosen to be:

$$\sigma^2(\Lambda_F) = \frac{D}{m}\left(\frac{m\sigma^2}{m\sigma^2 - D}\right), \quad \sigma^2(\Lambda_C) = m\sigma^2\left(\frac{m\sigma^2}{m\sigma^2 - D}\right). \tag{5.130}$$

Each Gaussian vector $\mathbf{u}_i$ is first quantized and dithered with vector $\mathbf{d}_i$:

$$\mathbf{c}_i = Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) \mod \Lambda_C \quad \text{for} \quad i = 1 \cdots m \tag{5.131}$$

where the dithers $\mathbf{d}_1, \ldots, \mathbf{d}_m$ are drawn i.i.d uniformly from the voronoi region of the coarse lattice $\mathcal{V}_C$. We define the one-to-one mapping $w_2 : \Lambda_F \cap \mathcal{V}_C \to \mathbb{F}_p^{k_2'}$ as in Definition 5.51 and map each quantized point $\mathbf{c}_i$ to a point on the finite field $\mathbb{F}^{k_2'}$ using the mapping w_2. Source node i transmits $w_2(\mathbf{c}_i)$ across the Gaussian network $\mathcal{N}_{\text{GAUSS-MAC}}$ and the destination recovers the mod sum: $w_2(\mathbf{c}_1) \oplus \cdots \oplus w_2(\mathbf{c}_m)$. Applying Lemma 5.43, this can be reliably recovered as long as

$$\frac{k_2'}{k} \log p < \frac{n}{k} \min_{i \in \mathcal{S}} \min_{\Gamma \subseteq \Omega : i \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \alpha \tag{5.132}$$

where $C_\Gamma^{\text{GAUSS-MAC}}$ is the value of cut Γ in $\mathcal{N}_{\text{GAUSS-MAC}}$ and α is the constant in (5.35). If $w_2(\mathbf{c}_1) \oplus_p \cdots \oplus_p w_2(\mathbf{c}_m)$ is reliably recovered, the destination maps it back to a point in $\Lambda_F \cap \mathcal{V}_C$:

$$w_2^{-1}(w_2(\mathbf{c}_1) \oplus_p \cdots \oplus_p w_2(\mathbf{c}_m)) = (\mathbf{c}_1 + \cdots + \mathbf{c}_m) \mod \Lambda_C \tag{5.133}$$

Let $\beta = \frac{m\sigma^2 - D}{m\sigma^2}$. The destination recovers an estimate $\hat{\mathbf{v}}$ for $\mathbf{v} = \sum_{i=1}^{m} \mathbf{u}_i$ as follows:

$$\hat{\mathbf{v}} = \beta \left(\left(\sum_{i=1}^{m} \mathbf{c}_i \mod \Lambda_c - \sum_{i=1}^{m} \mathbf{d}_i \right) \mod \Lambda_C \right) \tag{5.134}$$

$$= \beta \left(\left(\sum_{i=1}^{m} \mathbf{c}_i - \sum_{i=1}^{m} \mathbf{d}_i \right) \mod \Lambda_C \right) \tag{5.135}$$

$$= \beta \left(\left(\sum_{i=1}^{m} (\mathbf{c}_i - \mathbf{d}_i) \right) \mod \Lambda_C \right) \tag{5.136}$$

$$= \beta \left(\left(\sum_{i=1}^{m} (Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) \mod \Lambda_c - \mathbf{d}_i) \right) \mod \Lambda_C \right) \tag{5.137}$$

$$= \beta \left(\left(\sum_{i=1}^{m} (Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - \mathbf{d}_i) \right) \mod \Lambda_c \right) \tag{5.138}$$

$$= \beta \left(\left(\sum_{i=1}^{m} (\mathbf{u}_i + Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - (\mathbf{u}_i + \mathbf{d}_i)) \right) \mod \Lambda_C \right) \tag{5.139}$$

$$= \beta \left(\left(\mathbf{v} + \sum_{i=1}^{m} (Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - (\mathbf{u}_i + \mathbf{d}_i)) \right) \mod \Lambda_C \right) \tag{5.140}$$

We define the event $\mathcal{E}_k$ where

$$\mathcal{E}_k = \left\{ \left(\mathbf{v} + \sum_{i=1}^{m} (Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - (\mathbf{u}_i + \mathbf{d}_i)) \right) \mod \Lambda_C = \mathbf{v} + \sum_{i=1}^{m} (Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - (\mathbf{u}_i + \mathbf{d}_i)) \right\} \tag{5.141}$$

The noise $Q_{\Lambda_F}(\mathbf{u}_i + \mathbf{d}_i) - (\mathbf{u}_i + \mathbf{d}_i)$ has the same distribution as $-\mathbf{d}_i$ and is independent

of $\mathbf{u}_i$ [22, 23]. Using this fact, we have that

$$\frac{1}{k}\mathbb{E}\left[\|\mathbf{v}+\sum_{i=1}^{m}\left(Q_{\Lambda_F}(\mathbf{u}_i+\mathbf{d}_i)-(\mathbf{u}_i+\mathbf{d}_i)\right)\|^2\right] = \frac{1}{k}\mathbb{E}\left[\|\mathbf{v}\|^2\right] \tag{5.142}$$

$$+\frac{1}{k}\sum_{i=1}^{m}\mathbb{E}\left[\|Q_{\Lambda_F}(\mathbf{u}_i+\mathbf{d}_i)-(\mathbf{u}_i+\mathbf{d}_i)\|^2\right] \tag{5.143}$$

$$= \frac{1}{k}\mathbb{E}\left[\|\mathbf{v}\|^2\right]+\sum_{i=1}^{m}\frac{1}{k}\mathbb{E}\left[\|-\mathbf{d}_i\|^2\right] \tag{5.144}$$

$$= m\sigma^2+m\frac{D}{m}\left(\frac{m\sigma^2}{m\sigma^2-D}\right) \tag{5.145}$$

$$= m\sigma^2\left(\frac{m\sigma^2}{m\sigma^2-D}\right) \tag{5.146}$$

$$= \sigma^2(\Lambda_C) \tag{5.147}$$

Along the same lines as in [16, 23], it can be shown that $\Pr(\mathcal{E}_k)\to 1$ as $k\to\infty$. Under event $\mathcal{E}_k$, the estimate becomes: $\hat{\mathbf{v}}=\beta\mathbf{v}+\beta\sum_{i=1}^{m}\left(Q_{\Lambda_F}(\mathbf{u}_i+\mathbf{d}_i)-(\mathbf{u}_i+\mathbf{d}_i)\right)$. By conditioning on the events $\mathcal{E}$ and $\mathcal{E}^c$, the achievable distortion can be bounded as follows:

$$D_{\text{ACHIEVABLE}} = D_{\text{ACHIEVABLE},\mathcal{E}}\Pr(\mathcal{E})+D_{\text{ACHIEVABLE},\mathcal{E}^c}\Pr(\mathcal{E}^c) \tag{5.148}$$

$$\leq \frac{1}{k}\mathbb{E}\left[\|\mathbf{v}-\hat{\mathbf{v}}\|^2|\mathcal{E}\right]\Pr(\mathcal{E})+m\sigma^2\Pr(\mathcal{E}^c) \tag{5.149}$$

$$= \frac{1}{k}\mathbb{E}\left[\|(1-\beta)\mathbf{v}-\beta\sum_{i=1}^{m}\left(Q_{\Lambda_F}(\mathbf{u}_i+\mathbf{d}_i)-(\mathbf{u}_i+\mathbf{d}_i)\right)\|^2\right]+m\sigma^2\Pr(\mathcal{E}^c) \tag{5.150}$$

$$= (1-\beta)^2\frac{1}{k}\mathbb{E}\left[\|\mathbf{v}\|^2\right] \tag{5.151}$$

$$+\beta^2\sum_{i=1}^{m}\frac{1}{k}\mathbb{E}\left[\|\left(Q_{\Lambda_F}(\mathbf{u}_i+\mathbf{d}_i)-(\mathbf{u}_i+\mathbf{d}_i)\right)\|^2\right]+m\sigma^2\Pr(\mathcal{E}^c) \tag{5.152}$$

$$= (1-\beta)^2m\sigma^2+\beta^2m\frac{D}{m}\left(\frac{m\sigma^2}{m\sigma^2-D}\right)+m\sigma^2\Pr(\mathcal{E}^c) \tag{5.153}$$

$$= (1-\beta)^2m\sigma^2+\beta^2D\left(\frac{m\sigma^2}{m\sigma^2-D}\right)+m\sigma^2\Pr(\mathcal{E}^c) \tag{5.154}$$

Since $\beta=\frac{m\sigma^2-D}{m\sigma^2}$, the achievable distortion from (5.154) becomes

$$D_{\text{ACHIEVABLE}} = (1-\beta)^2m\sigma^2+\beta^2D\left(\frac{m\sigma^2}{m\sigma^2-D}\right)+\epsilon \tag{5.155}$$

$$= D+\epsilon \tag{5.156}$$

where ϵ can be made arbitrarily small as $k \to \infty$. From the nested lattice constructions and (5.132), we have that

$$\frac{k_2'}{k} \log p = \frac{1}{2} \log \left(\frac{\sigma^2 (\Lambda_C)}{\sigma^2 (\Lambda_F)} \right) \tag{5.157}$$

$$= \frac{1}{2} \log \left(\frac{m^2 \sigma^2}{D} \right) \tag{5.158}$$

$$\leq q \min_{i \in \mathcal{S}} \min_{\Gamma \subseteq \Omega : i \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \alpha \tag{5.159}$$

where α is given by (5.35). Hence, (5.159) implies that any distortion D satisfying

$$D \geq m^2 \sigma^2 2^{-2\alpha} \max_{i \in \mathcal{S}} 2^{-2q \min_{\Gamma \subseteq \Omega : i \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}}} \tag{5.160}$$

is achievable.

5.3 Extension to Asymmetric Linear Functions

So far, this chapter considered computing symmetric linear functions of Gaussian sources across Gaussian networks. We reduced this problem to computing the sum of discrete sources over linear deterministic networks, which is shown to achieve the cut-set bounds in Remark 5.21. Hence, in the Gaussian case, cut-set is approximately tight and the achievable distortion can be characterized to within a constant ratio of the optimal performance.

The proposed approach can also be applied to the problem of sending an asymmetric function of Gaussian sources across the two-user Gaussian network. Here, source 1 observes U_1, source 2 observes U_2 and the destination desires to estimate $U_1 + \beta U_2$ where $\beta > 1$. The corresponding deterministic problem is sending the finite-field sum with an additional private message across the linear deterministic multiple-access network. To see this, consider the case of sending $4U_1 + U_2$. First, U_2 and U_2 can be conceptually written as $0.b_{11}b_{12}\cdots$ and $0.b_{21}b_{22}\cdots$ respectively as in [78]. The asymmetric function $4U_1 + U_2$ can be thought as $b_{11}b_{12}.b_{13} \oplus b_{21}\ b_{14} \oplus b_{22} \cdots$. Here, we can see the destination wants to decode the private messages b_{11}, b_{12} and the finite-field sums, $b_{13} \oplus b_{21}, b_{14} \oplus b_{22}, \ldots$. It is shown in Theorem 5.27 that cut-set is tight for this scenario. Hence, applying a similar approach as that for the symmetric sum, cut-set is approximately tight in the Gaussian case and the achievable distortion is within a constant ratio of the cut-set bounds. We provide the details in the sequel.

5.3.1 Asymmetric Functions over Gaussian Networks

We consider a two-user Gaussian network $\mathcal{N}_{\text{GAUSS-MAC}}$ with source nodes $\mathcal{S} = \{1, 2\}$. Source 1 observes $U_1 \sim \mathcal{N}(0, \sigma^2)$, Source 2 observes an independent $U_2 \sim \mathcal{N}(0, \sigma^2)$, and the destination recovers an asymmetric linear function: $U_1 + \gamma U_2$. Without loss of generality, the space

of functions to can be limited to $\gamma > 1$. In Theorem 5.44 and 5.45 below, the achievable distortion is characterized to within a constant gap of the optimal performance.

Theorem 5.44. *Consider sending the linear function $U_1 + \gamma U_2$ with $\gamma > 1$ across the two-user Gaussian network $\mathcal{N}_{MAC}$. The achievable distortion satisfies*

$$D_{ACHIEVABLE} \leq \max\left\{2^{2q\alpha+4}2^{-2qC_1}, (1+\gamma^2)\sigma^2 2^{2q\alpha+9}2^{-2qC_2}\right\} \tag{5.161}$$

where $C_i = \min_{\Gamma: i\in\Gamma} C_\Gamma^{GAUSS\text{-}MAC}$ for $i = 1, 2$ and α is the constant in (5.35).

Theorem 5.45. *The optimal distortion D_{OPT} for sending $U_1 + \gamma U_2$ across the network $\mathcal{N}_{GAUSS\text{-}MAC}$ satisfies the following cut-set bound:*

$$D_{OPT} \geq \sigma^2 \max\left\{2^{-2qC_1}, \gamma^2 2^{-2qC_2}\right\} \tag{5.162}$$

where $C_i = \min_{\Gamma\subseteq\Omega: i\in\Gamma} C_\Gamma^{GAUSS\text{-}MAC}$ for $i = 1, 2$.

Proof. Follows along the same lines as Theorem 5.35. □

Remark 5.46. The ratio between the achievable distortion and the cut-set lower bounds is bounded by a constant that depends only on q, the number of nodes, and the degree of the network. The ratio is an independent of the network topology.

Proof. (Theorem 5.44). We first state the counterpart of Lemma 5.43 for asymmetric functions in Lemma 5.47. This Lemma provides an achievable rate for sending a finite field sum with a private message across the Gaussian network $\mathcal{N}_{\text{GAUSS-MAC}}$.

Lemma 5.47. *Consider a two-user Gaussian network $\mathcal{N}_{GAUSS\text{-}MAC}$ with discrete source observations $U_{1,1}^{k_1}$ at source 1 and $U_{2,1}^{k_1}, U_{2,2}^{k_2}$ at source 2 where each $U_{i,j}$ is drawn i.i.d uniformly from some prime-sized finite field $\mathbb{F}_p$. For all $\epsilon > 0$ and n, p large, there exists encoders $\{\mathcal{E}_{i,t}\}_{t=1}^n \forall i \in \Omega$ and a decoder $\mathcal{G}$ that produces an estimates $\widehat{V}_1^{k_1} \in \mathbb{F}_p^{k_1}, \widehat{V}_2^{k_2} \in \mathbb{F}_p^{k_1}$ such that*

$$\Pr\left(\left\{\widehat{V}^{k_1} \neq U_{1,1}^{k_1} \oplus_p U_{1,1}^{k_1}\right\} \cup \left\{\widehat{V}^{k_2} \neq U_{2,2}^{k_2}\right\}\right) < \epsilon \tag{5.163}$$

as long as the computation rates $R_1 = \frac{k_1}{n}\log p, R_2 = \frac{k_2}{n}\log p$ satisfy

$$R_1 < \min_{\Gamma\subseteq\Omega: 1\in\Gamma} C_\Gamma^{GAUSS\text{-}MAC} - \alpha \tag{5.164}$$

$$R_1 + R_2 < \min_{\Gamma\subseteq\Omega: 2\in\Gamma} C_\Gamma^{GAUSS\text{-}MAC} - \alpha \tag{5.165}$$

where α is given by (5.35).

Proof. Follows along the same lines as that for Lemma 5.43. □

We now construct the nested-lattices used for source quantization. Let $\Lambda_C \subset \Lambda_{C'} \subset \Lambda_{C''} \subset \Lambda_F$ be a set of 4 nested lattices from the following construction:

$$\Lambda_C = \left\{ p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \right) + \mathbb{Z}^k : \mathbf{w}_1 \in \mathbb{F}_p^{k_1'} \right\} \tag{5.166}$$

$$\Lambda_{C'} = \left\{ p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \right) + \mathbb{Z}^k : \mathbf{w}_1 \in \mathbb{F}_p^{k_1'}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2'} \right\} \tag{5.167}$$

$$\Lambda_{C''} = \left\{ p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_3 \cdot \mathbf{w}_3 \right) + \mathbb{Z}^k : \mathbf{w}_1 \in \mathbb{F}_p^{k_1'}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2'}, \mathbf{w}_3 \in \mathbb{F}_p^{k_3'} \right\} \tag{5.168}$$

$$\Lambda_F = \left\{ p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_3 \cdot \mathbf{w}_3 \oplus \mathbf{G}_4 \cdot \mathbf{w}_4 \right) + \mathbb{Z}^k : \mathbf{w}_i \in \mathbb{F}_p^{k_i'} \text{ for } i = 1, \ldots, 4 \right\} \tag{5.169}$$

where $\mathbf{G}_1 \in \mathbb{F}^{k \times k_1'}, \mathbf{G}_2 \in \mathbb{F}^{k \times k_1' + k_2'}, \mathbf{G}_3 \in \mathbb{F}^{k \times k_1' + k_2' + k_3'}, \mathbf{G}_4 \in \mathbb{F}^{k \times k_1' + k_2' + k_3' + k_4'}$ and each element is drawnly i.i.d from $\mathbb{F}_p$. If $p, k, k_1', k_2', k_3', k_4'$ are chosen to scale appropriately, the matrix $[\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4]$ becomes full rank with high probability and the lattices $\Lambda_C, \Lambda_{C'}, \Lambda_{C''}, \Lambda_F$ are simultaneously good for covering, quantization and AWGN coding (see [17] for definitions and proofs).

The lattices are scaled to have the following variances:

$$\sigma^2(\Lambda_C) = 2^4 (1+\gamma^2) \sigma^2 \left(\frac{(1+\gamma^2)\,\sigma^2}{(1+\gamma^2)\,\sigma^2 - D} \right) \tag{5.170}$$

$$\sigma^2(\Lambda_{C'}) = 2^4 \sigma^2 \left(\frac{(1+\gamma^2)\,\sigma^2}{(1+\gamma^2)\,\sigma^2 - D} \right) \tag{5.171}$$

$$\sigma^2(\Lambda_{C''}) = \sigma^2 \left(\frac{(1+\gamma^2)\,\sigma^2}{(1+\gamma^2)\,\sigma^2 - D} \right) \tag{5.172}$$

$$\sigma^2(\Lambda_F) = \frac{D}{2} \left(\frac{(1+\gamma^2)\,\sigma^2}{(1+\gamma^2)\,\sigma^2 - D} \right) \tag{5.173}$$

Before describing the source quantization scheme, we first provide some definitions and lemmas regarding the lattice constructions.

Lemma 5.48. $|\Lambda_F \cap \mathcal{V}_{C'}| = p^{k_3+k_4}$, $|\Lambda_{C''} \cap \mathcal{V}_C| = p^{k_2+k_3}$

Proof. Follows along similar lines as Lemma 5.40 □

Definition 5.49. Let $\mathbf{c} \in \Lambda_F \cap \mathcal{V}_{C'}$. By construction, we have that

$$\mathbf{c} = p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1 \oplus \mathbf{G}_2 \cdot \mathbf{w}_2 \oplus \mathbf{G}_3 \cdot \mathbf{w}_3 \oplus \mathbf{G}_4 \cdot \mathbf{w}_4 \right) + \mathbf{z} \tag{5.174}$$

for some $\mathbf{w}_1 \in \mathbb{F}_p^{k_1}, \mathbf{w}_2 \in \mathbb{F}_p^{k_2}, \mathbf{w}_3 \in \mathbb{F}_p^{k_3}, \mathbf{w}_4 \in \mathbb{F}_p^{k_4}, \mathbf{z} \in \mathbb{Z}^n$. We define the mapping: $\phi : \Lambda_F \cap \mathcal{V}_{C'} \to \mathbb{F}_p^{k_3+k_4}$ where

$$\phi(\mathbf{c}) = (\mathbf{w}_3, \mathbf{w}_4) \tag{5.175}$$

Let $\mathbf{c}' \in \Lambda_{C''} \cap \mathcal{V}_C$. By construction, we have that

$$\mathbf{c}' = p^{-1} \left(\mathbf{G}_1 \cdot \mathbf{w}_1' \oplus \mathbf{G}_2 \cdot \mathbf{w}_2' \oplus \mathbf{G}_3 \cdot \mathbf{w}_3' \right) + \mathbf{z}' \tag{5.176}$$

for some $\mathbf{w}_1' \in \mathbb{F}_p^{k_1}, \mathbf{w}_2' \in \mathbb{F}_p^{k_2}, \mathbf{w}_3' \in \mathbb{F}_p^{k_3}, \mathbf{z}' \in \mathbb{Z}^n$. We define the mapping: $\phi' : \Lambda_{C''} \cap \mathcal{V}_C \to \mathbb{F}_p^{k_2+k_3}$ where

$$\phi'(\mathbf{c}') = (\mathbf{w}_2', \mathbf{w}_3') \tag{5.177}$$

Lemma 5.50. *The function ϕ is a group isomorphism from $(\Lambda_F \cap \mathcal{V}_{C''}, \mod \Lambda_{C''})$ to $(\mathbb{F}_p^{k_3+k_4}, \mod p)$ and ϕ' is a group isomorphism from $(\Lambda_{C''} \cap \mathcal{V}_C, \mod \Lambda_C)$ to $(\mathbb{F}_p^{k_2+k_3}, \mod p)$.*

Proof. Follows along similar lines as Lemma 5.42 □

Lemma 5.51. $\{\mathbf{x} + \mathbf{y} : \mathbf{x}, \mathbf{y} \in \mathcal{V}_{C''}\} \subseteq \mathcal{V}_{C'}$

Proof. Follows from the fact that $\Lambda_{C''}, \Lambda_{C'}$ are both simultaneous good for quantization and AWGN coding and that $\Lambda_{C'} = 2(2\sigma(\Lambda_{C''}))^2$. □

We provide the details to the source quantization scheme. The vectors $\mathbf{u}_1$, $\gamma \mathbf{u}_2$ are first dithered and then quantized using the fine lattice Λ_F to get the following quantization points:

$$\mathbf{c}_1 = Q_F(\mathbf{u}_1 + \mathbf{d}_1) \tag{5.178}$$
$$\mathbf{c}_2 = Q_F(\gamma \mathbf{u}_1 + \mathbf{d}_2) \tag{5.179}$$

where the dithers $\mathbf{d}_1, \mathbf{d}_2 \sim$ are drawn i.i.d uniformly from $\mathcal{V}_F$. We define:

$$\mathbf{c}_{2,r} = \mathbf{c}_2 \mod \Lambda_{C''} \tag{5.180}$$
$$\mathbf{c}_{2,q} = Q_{C''}(\mathbf{c}_2) \tag{5.181}$$
$$\mathbf{c}_{2,q}^{'} = \mathbf{c}_{2,q} \mod \Lambda_C \tag{5.182}$$
$$\mathbf{c}_{1,r} = \mathbf{c}_1 \mod \Lambda_{C''} \tag{5.183}$$

We note that $\mathbf{c}_{1,r}, \mathbf{c}_{2,r} \in \Lambda_F \cap \mathcal{V}_{C''} \subset \Lambda_F \cap \mathcal{V}_{C'}$. Source 1 transmits $\mathbf{w}_{1,r}$ where

$$\mathbf{w}_{1,r} = \phi(\mathbf{c}_{1,r}). \tag{5.184}$$

Source 2 transmits $\mathbf{w}_{2,r}, \mathbf{w}_{2,q}$, given by

$$\mathbf{w}_{2,r} = \phi(\mathbf{c}_{2,r}) \tag{5.185}$$
$$\mathbf{w}_{2,q} = \phi'(\mathbf{c}_{2,q}^{'}). \tag{5.186}$$

Applying Lemma 5.47, $\mathbf{w}_{1,r} \oplus_p \mathbf{w}_{2,r}, \mathbf{w}_{2,q}$ can be reliably recovered as long as

$$\frac{k_3' + k_4'}{k} \log p < \frac{n}{k} \min_{\Gamma \subseteq \Omega : 1 \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \gamma \tag{5.187}$$

$$\frac{k_2' + k_3'}{k} + \frac{k_3' + k_4'}{k} \log p < \frac{n}{k} \min_{\Gamma \subseteq \Omega : 2 \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \alpha \tag{5.188}$$

where $C_\Gamma^{\text{GAUSS-MAC}}$ is the value of cut Γ in $\mathcal{N}_{\text{GAUSS-MAC}}$ and α is the constant in (5.35). Mapping back, we have that:

$$\phi^{-1}(\mathbf{w}_{1,r} \oplus_p \mathbf{w}_{2,r}) = (\mathbf{c}_{1,r} + \mathbf{c}_{2,r}) \mod \Lambda_{C'} \tag{5.189}$$

$$\phi'^{-1}(\mathbf{w}_{2,q}) = \mathbf{c}_{2,q} \mod \Lambda_C \tag{5.190}$$

Let $\beta = \frac{(1+\gamma^2)\sigma^2 - D}{(1+\gamma^2)\sigma^2}$. The destination computes:

$$((\mathbf{c}_{1,r} + \mathbf{c}_{2,r}) \mod \Lambda_{C'} + \mathbf{c}_{2,q} \mod \Lambda_C - \mathbf{d}_1 - \mathbf{d}_2) \mod \Lambda_C \tag{5.191}$$

$$= \beta\,((\mathbf{c}_{1,r} + \mathbf{c}_{2,r}) \mod \Lambda_{C'} + \mathbf{c}_{2,q} - \mathbf{d}_1 - \mathbf{d}_2) \mod \Lambda_C \tag{5.192}$$

$$\overset{(a)}{=} \beta\,((\mathbf{c}_{1,r} + \mathbf{c}_{2,r}) + \mathbf{c}_{2,q} - \mathbf{d}_1 - \mathbf{d}_2) \mod \Lambda_C \tag{5.193}$$

$$= \beta\,(\mathbf{c}_{1,r} + \mathbf{c}_2 - \mathbf{d}_1 - \mathbf{d}_2) \mod \Lambda_C \tag{5.194}$$

$$= \beta\,(\mathbf{c}_1 + \mathbf{c}_2 - \mathbf{d}_1 - \mathbf{d}_2) \mod \Lambda_C \tag{5.195}$$

$$= \beta\,(\mathbf{u}_1 + \gamma \mathbf{u}_2 + Q_F(\mathbf{u}_1 + \mathbf{d}_1) - (\mathbf{u}_1 + \mathbf{d}_1) + Q_F(\mathbf{u}_2 + \mathbf{d}_2) - (\mathbf{u}_2 + \mathbf{d}_2)) \mod \Lambda_C \tag{5.196}$$

where (a) follows by Lemma 5.51. Using a similar argument as in the symmetric case, any distortion D satisfying the following inequalities can be shown to be achievable:

$$\frac{1}{2} \log \left(\frac{2^4 \sigma^2 \frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}}{\frac{D}{2}\left(\frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}\right)} \right) \leq q \left(\min_{\Gamma \subseteq \Omega : 1 \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \gamma \right) \tag{5.197}$$

$$\frac{1}{2} \log \left(\frac{2^4 (1+\gamma^2)\sigma^2 \left(\frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}\right)}{\sigma^2 \left(\frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}\right)} \right) + \tag{5.198}$$

$$\frac{1}{2} \log \left(\frac{2^4 \sigma^2 \left(\frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}\right)}{\frac{D}{2}\left(\frac{(1+\gamma^2)\sigma^2}{(1+\gamma^2)\sigma^2 - D}\right)} \right) \leq q \left(\min_{\Gamma \subseteq \Omega : 2 \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}} - \gamma \right) \tag{5.199}$$

Rearranging, it can be shown that

$$D \geq \sigma^2 2^{q\alpha + 5} 2^{-2C_1} \tag{5.200}$$

$$D \geq (1+\gamma^2)\alpha^2 2^{q\gamma + 9} 2^{-2C_2} \tag{5.201}$$

$$\tag{5.202}$$

where $C_i = \min_{\Gamma: i \in \Gamma} C_\Gamma^{\text{GAUSS-MAC}}$ for $i = 1, 2$ and α is the constant in (5.35). □

Using the same logic, the problem of sending an asymmetric function of three Gaussian sources corresponds to the three user deterministic computation problem with private messages. However, as shown in Theorem 5.27, cut-set is not tight for this deterministic scenario and does not lead to a constant ratio in the Gaussian case.

5.4 Discussion

We studied linear function computation in both linear deterministic and Gaussian networks. In the first part of this chapter, we developed a framework for computing functions of discrete sources over linear deterministic multiple-access networks [77]. We observed a duality relation between broadcast with common messages and multiple-access with computation that extends the well-known broadcast multiple-access duality to various communication demands. The duality relation allows the recasting of a multiple-access network computation problem into a broadcast network problem. This is useful since broadcast problems are well studied and solutions to various cases have been developed. We applied the duality relationship to develop new results regarding computation over multiple-access networks. We focused on characterizing scenarios under which cut-set upper bounds is tight since it is the most common information theoretic bound used in networks. We considered broadcast networks with various message demands and found that there are only two set of demands under which cut-set is tight for all networks.

In the second part of this chapter, we extracted the deterministic model insights and applied it to Gaussian networks. We considered computing the sum of Gaussian sources over a class of relay networks with Gaussian multiple-access channels and proved that the achievable distortion is within a constant ratio of the optimal distortion given by the cut-set lower bounds. Our scheme separates the physical and network layers and uses nested lattices codes for computation over the physical layer and network codes in the network layer. As a result of the separation, we reduced the original problem into one of computing discrete sources over linear deterministic networks and can apply the framework in the first part of the chapter.

In this work, duality and cut-set bounds were proposed as conceptual tools to study function computation over networks. Thus, the natural question is the extent to which these tools can be generalized. The set of computation demands in multiple-access networks can be expanded by generalizing the communication demands of broadcast networks to include cases where the destinations are interested in functions of the messages. In future work, it would be interesting to develop new achievability schemes and converses for cases where the cut-set is shown to be not tight.

Chapter 6

Functional Forwarding of Channel State Information

The lack of global channel state information often results in a significant reduction in capacity in wireless networks. It is well known that non-coherent, fast fading, point-to-point channels lose a fraction of their capacity at high SNR [83], [84] [85]. The knowledge of channel state information at the basestation can increase the sum rate nontrivially in the wireless downlink infrastructure with multiple antennas [86], [87]. In the interference channel, the multiplexing gain is significantly larger when channel state information is known both at the transmitter and the receiver [24, 88]. The optimal multiplexing gain is still achievable even in the case when the transmitter does not know the channel state information but learns it via feedback from the receiver [89]. However, the achievable multiplexing gain decreases significantly when channel state information is completely absent at the transmitter [90]. These results suggest that it is crucial to learn channel state information in networks.

In large networks involving many intermediate relays, the cost of forwarding channel state information is non-trivial. One network that has been studied in the literature is the uplink infrastructure with basestation cooperation [91, 92, 93]. Here, mobiles send their information to a set of nearby basestations, which first process the received information then jointly forward it to the remote central processor. In the case where the channel state information is not known globally, the basestations measure the channel states of their local links through the use of pilot signals at the mobiles but the central processor does not have access to channel state information directly. One obvious strategy is for the basestations to forward the entire channel state information to the receiver. However, this can be inefficient when there is a large number of mobiles and basestations present in the network.

In this chapter we propose a scheme called *functional forwarding* in which the nodes in the wireless network, rather than sending full information, send only the function of the CSI needed at the decoder. Our research is motivated by the fact that full CSI is often not needed. Instead, a function of the CSI is sufficient. In recent work [27], it was shown that it is sometimes much more efficient to communicate only a function of the information, rather

than the full information. In this chapter, we adapt this approach to the particular problem of efficiently forwarding CSI. Hence, by contrast to [27], we are not interested in an error free forwarding nor in a distorted version, but we have to forward just enough CSI to make decoding successful at the desired rate.

The rest of this chapter is organized as follows. In Section II, we present our channel model, the two-stage fading relay network. We develop a general framework for characterizing the achievable rate for lossless functional forwarding in Section III. We first state the general achievable rate for functional forwarding in Theorem 1 and then consider a series of examples to compare the performance of functional forwarding and full forwarding of CSI. In Section IV, we extend our general framework to the lossy case and consider a Gaussian network example.

6.1 Channel Model

Though the concept of functional forwarding of CSI is more widely applicable, in the present chapter we restrict attention to one particular network topology. We consider the two-stage relay network with N information sources (and we will sometimes refer to them as *users*), M relays, and a single destination (see Figure 6.1). Each source chooses a message w_j uniformly from the set $\mathcal{W}_j = \{1, 2, .., M_j\}$. Each message is mapped into a length n codeword:

$$\mathcal{E}_j : \mathcal{W}_j \to \mathcal{X}_j^n \qquad \text{for} \quad j = 1, ..., N \tag{6.1}$$

Let $X_j[i]$ be the channel input from source j at time i. In the first stage of the network, which we will refer to as the broadcast (BC) stage, the transmitted codewords are broadcasted to the M relays through the channel characterized by:

$$Q(y_1, ..., y_M | x_1, ..., x_N, \mathbf{H}) \tag{6.2}$$

where $\mathbf{H}$ is an $M \times N$ matrix from alphabet $\mathcal{H}$ denoting the channel state information. We will find it convenient to denote the row vectors of $\mathbf{H}$ by $\mathbf{h}_m^T$, for $m = 1, 2, \ldots, M$. In the present chapter, we adopt a fast-fading model where the matrix $\mathbf{H}$ changes over time. We use i to denote the (discrete) time index and will write $\mathbf{H}[i]$ for the matrix at time i. More precisely, $\mathbf{H}[i]$ is drawn i.i.d. each time instant according to some distribution $P_{\text{CSI}}(\mathbf{H})$. We assume $\mathbf{h}_m$ is known locally at relay m but is not known globally. The transmitters and destination know only the distribution of the CSI.

In the second stage of our network, referred to as the multiple-access (MAC) stage, the relays communicate to the destination through a multiple-access channel. We allow $\ell \in \mathbb{N}$ uses of the MAC per use of the broadcast channel, meaning that we study the case where the bandwidth of the multiple access section of our network is an integer multiple of the bandwidth of the broadcast section. Each relay encodes its observation y_m^n and channel state information $\mathbf{h}_m^n$ into a length ℓn codeword:

$$\mathcal{R}_m : \mathcal{H}_m^n \times \mathcal{Y}_m^n \to \mathcal{X}_{r,m}^{\ell n} \quad \text{for} \quad m = 1, 2, \ldots, M \tag{6.3}$$

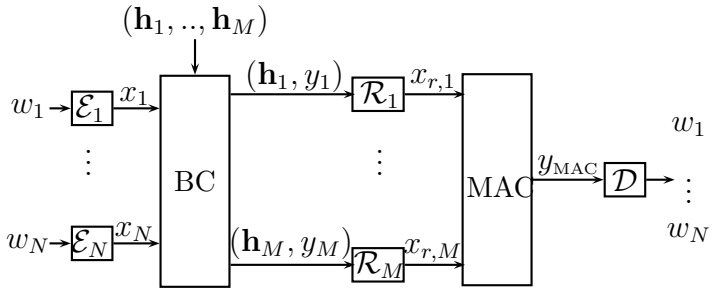

Figure 6.1. Two-Stage Fading Relay Network. The first stage is the broadcast (BC) stage and the second stage is the multiple-access (MAC) stage. CSI is known locally at the relays but is not known at the destination and the transmitters. We refer to the N encoding nodes labeled $\mathcal{E}_j$ as *information sources* or *users*, and to the single decoding node labeled $\mathcal{D}$ as the *destination.*

The final destination receives Y_{MAC} and decodes the original transmitted messages:

$$\mathcal{D} : \mathcal{Y}_{\text{MAC}}^{\ell n} \to \mathcal{W}_1 \times \ \cdots \ \times \mathcal{W}_N \tag{6.4}$$

$$(\hat{w}_1, ..., \hat{w}_M) = D(y_{\text{MAC}}^{\ell n}) \tag{6.5}$$

We require that the messages be reliably recovered:

$$Pr((\hat{w}_1, ..., \hat{w}_M) \neq (w_1, ..., w_M)) \leq \epsilon \tag{6.6}$$

for all $\epsilon > 0$ for n large.

6.2 Functional Forwarding

In this section, we describe our framework for the relays to send a function of the channel state information to the destination. We first give some key definitions used in our proposed strategy. Next, we describe functional forwarding in detail and provide its achievable rate for the two-stage relay network. Finally, we discuss the choice of forwarding function and show that functional forwarding can be much more effective than full forwarding through a series of examples.

6.2.1 Definitions

The central element in the strategies discussed in this chapter is the so-called "forwarding" function. We define this function as follows.

Definition 6.1. (Forwarding Function): Let f_F be a fixed many-to-one function and $\mathcal{U}$ be an alphabet:

$$f_F : \mathcal{Y}_1 \times \mathcal{H}_1 \times \cdots \times \mathcal{Y}_M \times \mathcal{H}_M \to \mathcal{U} \tag{6.7}$$

As shown in the above definition, the value of the forwarding function depends on the relay observations and channel state information: $U = f_F(y_1, \mathbf{h}_1, \cdots, y_M, \mathbf{h}_M)$. Clearly, in the considered network model, this value is not known to any individual relay. Rather, the forwarding function is sent to the destination in a distributed fashion using a computation code, which we now proceed to define formally.

Definition 6.2. (Computation Code): Given joint sequences $\{Y_1, \mathbf{h}_1, ..., Y_M, \mathbf{h}_M\}^k$ of type P and a fixed multiple-access channel, a (k, n, ϵ) $f_{F,P}$ computation code consists of

M encoders:

$$\mathcal{R}_m : \mathcal{Y}_m^k \times \mathcal{H}_m^k \to \mathcal{X}_{r,m}^n \tag{6.8}$$

such that

$$X_{r,m}^n = \mathcal{R}_m(Y_j^k, \mathbf{h}_j^k); \tag{6.9}$$

a forwarding function

$$U^k = f_{\mathrm{F,P}}(Y_1^k, \mathbf{h}_1^k, ..., Y_M^k, \mathbf{h}_M^k); \tag{6.10}$$

and a decoder

$$\mathcal{D} : \mathcal{Y}_{\mathrm{MAC}}^n \to \mathcal{U}^k; \tag{6.11}$$

that outputs an estimate $\hat{U}^k$ where

$$\hat{U}^k = \mathcal{D}(Y_{\mathrm{MAC}}^n) \tag{6.12}$$

$$Pr(U^k \neq \hat{U}^k) \leq \epsilon. \tag{6.13}$$

Definition 6.3. (Computation Rate): An $f_{F,P}$ computation rate $\kappa = \frac{k}{n}$ is achievable if for all $\epsilon \in (0, 1)$, there exists a $(\kappa n, n, \epsilon)$ $f_{F,P}$ computation code for all n greater than some $n_0 \in \mathbb{Z}$.

When the particular function is clear from context, we will often merely refer to *computation rate* in order to keep the terminology simple. The inverse of the computation rate represents the number of channel uses required for the destination of the multiple-access channel to reliably recover one instance of the forwarding function. For a fixed input distribution, broadcast channel, and channel state distribution, we find it useful to parametrize the multiple-access part of the two-stage fading network by its computation rate for a chosen forwarding function.

6.2.2 Proposed Scheme

We now proceed to describe our proposed strategy. Each user has a message w_j that is drawn uniformly from the set all of all messages $\mathcal{W}_j = \{1, \dots, M_j\}$ and constructs a random codebook where each element of each codeword is drawn according to distribution $P_j(X)$. Each encoder maps its message to a length n codeword:

$$\mathcal{E}_j : \mathcal{W}_j \to \mathcal{X}_j^n \quad \text{for } j = 1 \cdots N \tag{6.14}$$

The codewords are then transmitted to the relays through the broadcast stage of the network with the channel matrix from Equation (6.2). At time i, relay m observes $y_m[i]$, which is a noisy combination of the transmitted signals $x_1[i] \cdots x_N[i]$. Relay m has knowledge of its own channel state information vector $\mathbf{h}_m[i]$. We let $\mathbf{x}[i] = [x_1[i] \cdots x_M[i]]^T$ be the set of transmitted symbols at time instance i, $\mathbf{H}[i] = [\mathbf{h}_1[i] \cdots \mathbf{h}_M[i]]^T$ be the set of channel state information and $\mathbf{y}[i] = [y_1[i] \cdots y_M[i]]^T$ be the set of relay observations. We define the induced distribution for the set of relay observations and channel state information.

Definition 6.4. For the two-stage relay network under a fixed input distribution of the form $P_1(x_1)P_1(x_2) \cdots P_N(x_N)$, the induced distribution P_{IND} is defined as follows:

$$P_{\text{IND}}(\mathbf{y}, \mathbf{H}) = \sum_{x_1, \cdots, x_N} Q(\mathbf{y}|\mathbf{H}, \mathbf{x}) P_{\text{CSI}}(\mathbf{H}) P_1(x_1) \cdots P_N(x_N) \tag{6.15}$$

where Q is the channel distribution for the broadcast part of the network and P_{CSI} is the distribution for the channel state information.

We first choose a forwarding function $U = f_{F,P_{\text{IND}}}(y_1, \mathbf{h}_1, \cdots, y_M, \mathbf{h}_M)$. The relays will send U to the destination in a distributed fashion. Using a (k, n, ϵ) $f_{F,P_{\text{IND}}}$ computation code, each relay encodes its first k observations y_m^k and channel state information $\mathbf{h}_m^k$ into a length ℓn codeword:

$$\mathcal{R}_m : \mathcal{H}_m^k \times \mathcal{Y}_m^k \to \mathcal{X}_{r,m}^{\ell n} \quad \text{for} \quad m = 1, 2, \dots, M \tag{6.16}$$

where $\ell \in \mathbb{Z}_+$ is the bandwidth expansion of the MAC channel. The destination observes $\mathbf{y}_{\text{MAC}}$, which is a noisy combination of the transmitted signals from each relay, and performs decoding in two stages. In the first stage, the forwarding function is recovered:

$$\mathcal{D}_1 : \mathcal{Y}_{\text{MAC}}^{\ell n} \to \mathcal{U}^k \tag{6.17}$$

$$\hat{U}^k = \mathcal{D}_1(y_{\text{MAC}}^{\ell n}) \tag{6.18}$$

In the second stage the the original messages are recovered from the forwarding function:

$$\mathcal{D}_2 : \mathcal{U}^k \to \mathcal{W}_1 \times \quad \cdots \quad \times \mathcal{W}_N \tag{6.19}$$

$$(\hat{w}_1, ..., \hat{w}_M) = D_2(u^k) \tag{6.20}$$

We require that the messages be reliably recovered:

$$Pr((\hat{w}_1, ..., \hat{w}_M) \neq (w_1, ..., w_M)) \leq \epsilon \tag{6.21}$$

for all $\epsilon > 0$ for n large.

6.2.3 Achievable Rate

For the two stage relay network with broadcast channel characterized by $Q(\mathbf{y}|\mathbf{H},\mathbf{x})$ and channel state information distribution $P_{\text{CSI}}(\mathbf{H})$, we give the achievable rate for functional forwarding in the following theorem.

Theorem 6.5. *Consider the two-stage relay network under a fixed input distribution $P_1(x_1)\cdots P_N(x_N)$. For a given forwarding function $U = f_{F,P_{IND}}(Y_1,\mathbf{h}_1,...,Y_M,\mathbf{h}_M)$, let κ_U be an achievable computation rate for the multiple-access channel. The set of rates $(R_1\cdots R_N)$ are achievable if it satisfies the following inequalities:*

$$\sum_{i\in\mathcal{S}} R_i \leq \min\{\kappa_U\ell, 1\} I(X_{\mathcal{S}};U|X_{\mathcal{S}^c}) \quad \forall \quad \mathcal{S}\subseteq\{1,...,N\}$$

where ℓ is the bandwidth expansion of the multiple-access part of the network.

Remark 6.6. Full forwarding corresponds to the case where $U = (Y_1,\mathbf{h}_1,\cdots,Y_M,\mathbf{h}_M)$ in Theorem 6.5.

The proof of Theorem 1 is given in Appendix F.

6.2.4 Forwarding Functions

We observe that finding the optimal forwarding function f_F to maximize the achievable rate in Theorem 1 is non-trivial since the forwarding function appears in both the mutual information term and the computation rate κ_U. Furthermore, finding the optimal forwarding function involves first finding the computation rate for a general class of functions over a set of MACs, which is generally an open problem [27]. In this section, we briefly discuss criteria for selecting a good forwarding function and show that it is important to exploit the structure of the MAC.

Consider the single user two-stage relay network. From Theorem 1, the achievable rate using functional forwarding is given by

$$R = \min\{\kappa_U\ell, 1\}\, I(X;U) \tag{6.22}$$

where $U = f_{F,P_{\text{IND}}}(Y_1,\mathbf{h}_1,...,Y_M,\mathbf{h}_M)$ is the selected forwarding function and κ_U is the MAC's $f_{F,P_{\text{IND}}}$ computation rate. One good candidate for U is the sufficient statistic for X given the relay observations and fading information $(Y_1,\mathbf{h}_1,...,Y_M,\mathbf{h}_M)$. For this choice of forwarding function, no information is lost if the destination knows U rather than the full channel state information and relay observations. When the structure of the MAC is "perfectly matched" to the sufficient statistic (in a sense that will become clear through the examples in the sequel), then choosing the forwarding function to be the sufficient statistic is exactly optimal.

Example 1. Binary Network with OR MAC

We consider a single user binary network with an OR MAC (see Figure 6.2). The first stage consists of a binary broadcast channel with input: $X \in \{0,1\}$ and fading: $h_m \sim$ i.i.d $\mathcal{B}\left(\frac{1}{2}\right)$. Relay m observes $Y_m = h_m X$ for $m = 1, 2$. The MAC in the second stage has binary inputs $X_{r,m} \in \{0,1\}$ and output $Y_{\text{MAC}} = X_{r,1} \vee X_{r,2}$ (where $\vee$ is the OR function). We assume a bandwidth expansion $\ell = 2$. The forwarding function $U_{\text{SUFF}} = (U_{\text{SUFF},1}, U_{\text{SUFF},2})$ is chosen to be the sufficient statistic

$$U_{\text{SUFF},1} = Y_1 \vee Y_2 \tag{6.23}$$
$$U_{\text{SUFF},2} = h_1 \vee h_2 \tag{6.24}$$

The relays use uncoded transmission to first send U_1 and then U_2 across the MAC in a distributed fashion. The relays first transmit $X_{r,m}[i] = h_m[i]Y_m[i]$ for $i = 1, .., n$ and then $X_{r,m}[i] = h_m[i-n]$ for $i = n+1, \ldots, 2n$. The destination receives

$$Y_{\text{MAC}}[i] = Y_1[i] \vee \ldots \vee Y_M[i] \quad \text{for} \quad i = 1, \ldots, n$$
$$Y_{\text{MAC}}[i] = h_1[i-n] \vee \ldots \vee h_M[i-n] \quad \text{for} \quad i = n+1, \ldots, 2n$$

Thus, via uncoded transmission, we can establish that an achievable computation rate is given by $\kappa_U = \frac{1}{2}$. We note that since the structure of the MAC is perfectly matched to the sufficient statistic, the final destination receives $(U_{\text{SUFF},1}, U_{\text{SUFF},2})$. Plugging the achievable computation rate into Theorem 1, the achievable rate for functional forwarding is given by

$$R = \frac{1}{2} 2I(X; U_{\text{SUFF},1}, U_{\text{SUFF},2}) \tag{6.25}$$
$$= I(X; Y_1, Y_2, h_1, h_2) \tag{6.26}$$
$$= 1 - \left(\frac{1}{2}\right)^2 \tag{6.27}$$

where the last inequality follows by choosing the input distribution $X \sim \mathcal{B}\left(\frac{1}{2}\right)$. We evaluate the cutset upper bound [1]:

$$C_{\text{Cut-Set}} \tag{6.28}$$
$$= \max_{P(x, x_{r,1}, x_{r,2})} \min\left\{I(X; Y_1, Y_2, h_1, h_2), 2I(X_{r,1}, X_{r,2}; Y_{\text{MAC}})\right\} \tag{6.29}$$
$$= \max_{P(x)} I(X; Y_1, Y_2, h_1, h_2) \tag{6.30}$$
$$= 1 - \left(\frac{1}{2}\right)^2 \tag{6.31}$$

and find that functional forwarding is optimal.

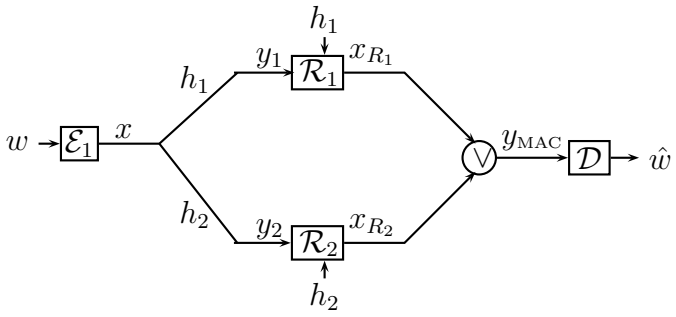

Figure 6.2. The binary network considered in Example 1, with two relays and the binary OR MAC

Forwarding the sufficient statistic is exactly optimal in this network since the sufficient statistic preserves information perfectly and the MAC is exactly matched with the sufficient statistic. However, when the MAC at hand is not matched to the sufficient statistic, then the latter is no longer the best forwarding function. This fact is illustrated in the next example.

Example 2. Binary Network with XOR MAC

We consider the two stage relay network shown in Figure 6.2. The binary broadcast channel in the first stage is the same as that in Example 1 and the sufficient statistic is given by Equation (6.23). The MAC in the second stage is an XOR MAC with inputs $X_{r,m} \in \{0,1\}$ for all $m = 1, 2$ and output $Y_{\text{MAC}} = X_{r,1} \oplus_2 X_{r,2}$. We consider the matched bandwidth case where $\ell = 1$. An achievable computation rate for the sufficient statistic is $\kappa_{\text{SUFF}} = 0.286$ (derived from Example 7 in [27]) and the functional forwarding achievable rate if the sufficient statistic is forwarded is given by

$$R_{\text{SUFF}} = \kappa_{\text{SUFF}} I(X; U_{\text{SUFF}}) \tag{6.32}$$
$$= 0.214 \tag{6.33}$$

where $X \sim \mathcal{B}(\frac{1}{2})$. Here, the achievable rate for functional forwarding is not optimal since the OR MAC cannot compute the XOR function efficiently. Hence, rather than forwarding the sufficient statistic, we forward the XOR function instead. We choose the forwarding function $U_{\text{XOR}} = (U_{\text{XOR},1}, U_{\text{XOR},2})$ to be the XOR of the inputs:

$$U_{\text{XOR},1} = Y_1 \oplus_2 Y_2 \tag{6.34}$$
$$U_{\text{XOR},2} = h_1 \oplus_2 h_2 \tag{6.35}$$

Using uncoded transmission separately for each component of U_{XOR} and using two channel uses, we find that an achievable computation rate for forwarding U_{XOR} is $\kappa_{\text{XOR}} = \frac{1}{2}$ and thus, using Theorem 1 with input distribution $X \sim \mathcal{B}\left(\frac{1}{2}\right)$, we find that the following rate is achievable:

$$R_{\text{XOR}} = \kappa_{\text{XOR}} I(X; U_{\text{XOR}}) \tag{6.36}$$
$$= 0.25 \tag{6.37}$$

where $X \sim \mathcal{B}(\frac{1}{2})$. This example shows that it is not always optimal to send the sufficient statistic. Rather, it is crucial to consider the structure of the MAC when choosing the forwarding function.

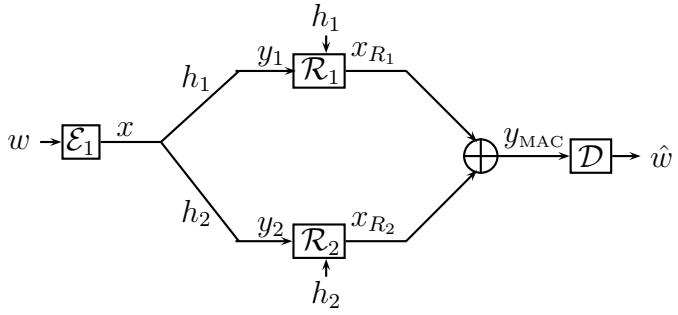

Figure 6.3. The binary network considered in Example 2, with two relays and binary XOR MAC

6.2.5 Single-User Examples

In this section, we compare the performance of functional forwarding versus full forwarding through a series of examples. Full forwarding corresponds to sending the entire set of relay observations and channel state information. Although our choices for the forwarding functions are not always optimal, we show that functional forwarding can be much more efficient than full forwarding.

We consider the single user two-stage binary network (see Figure 6.4) with different multiple-access channels. The broadcast network in the first stage is same as that of Example 1 but we now consider M relays instead of just two. In the second stage, the relays communicate to the destination through a multiple-access channel.

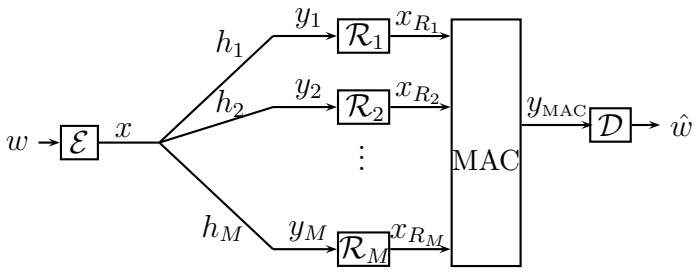

Figure 6.4. The Binary Relay Figure considered in Example 1-3 for varying multiple-access channels (MAC)

Example 1 (Continued).

The capacity region of this multiple-access channel is given by

$$\mathcal{C} = \left\{ (R_1 \cdots R_M) \in \mathbb{R}_+^M : \sum_{m=1}^{M} R_m = 1 \right\} \tag{6.38}$$

and the sum capacity $C_{\text{MAC,SUM}} = 1$. We choose the forwarding function to be the sufficient statistic: $U_{\text{SUFF}} = (U_{\text{SUFF},1}, U_{\text{SUFF},2})$ where

$$U_{\text{SUFF},1} = y_1 \vee y_2 \cdots \vee y_M \tag{6.39}$$
$$U_{\text{SUFF},2} = h_1 \vee h_2 \cdots \vee h_M \tag{6.40}$$

We observe that exactly as in the case of two users discussed before, uncoded transmission performs well here, attaining a computation rate of $\kappa_U = \frac{1}{2}$. We compare this to the strategy where the full information is forwarded to the destination from M' relays. From Theorem 1 (and also Remark 1), it can be shown that the overall achievable rate is given by:

$$R_{\text{FULL}}(M') = \frac{C_{\text{SUM, MAC}}}{H(Y_1, h_1, \ldots, Y_{M'}, h_{M'})} I(X; Y_1, h_1 \cdots Y_{M'}, h_{M'}) \tag{6.41}$$

We choose M' to maximize the overall sum transmission rate:

$$M' = \arg\max_{m \le M} R_{\text{FULL}}(m) \tag{6.42}$$

Figure 6.6 compares the performance of the different schemes. We note that functional forwarding is exactly optimal while full forwarding is highly suboptimal when the number of relays becomes large. In this case, the total amount of CSI becomes exceedingly large, and thus the MAC from the relays to the destination becomes the main bottleneck.

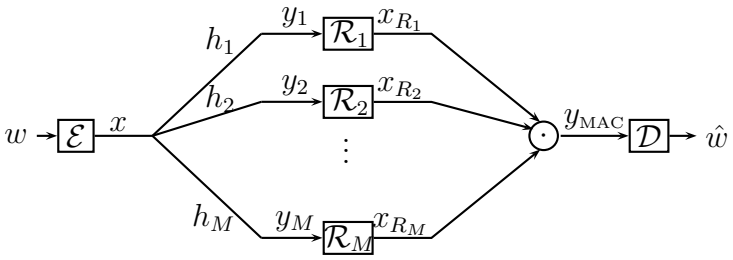

Figure 6.5. The binary network consider in Example 1 (continued), with M relays and multiplying MAC

In Example 1, the sufficient statistic is perfectly matched with the MAC and forwarding the sufficient statistic is optimal. In the next example, we show that even in the case where the MAC is not perfectly matched with the sufficient statistic, forwarding the sufficient statistic still gives a nontrivial gain over full forwarding.

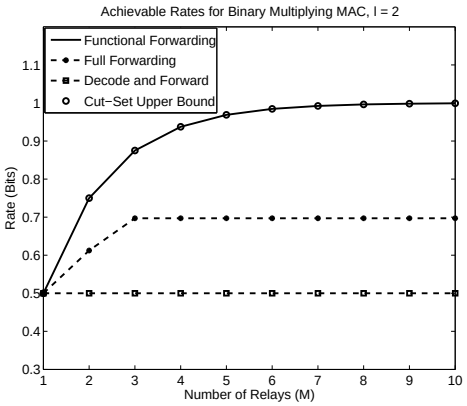

Figure 6.6. Achievable rates for binary network with multiplying MAC with from Example 1 (continued). In this network, functional forwarding is exactly optimal.

Example 3.

Consider the two-stage binary network where the MAC has inputs $X_{r,m} \in \{0, ..., 4M-1\}$ for all $m = 1, .., M$ and output $Y_{\text{MAC}} = \sum_{m=1}^{M} X_{r,m} + Z \mod 4M$ where $Z \sim \mathcal{B}\left(\frac{1}{2}\right)$. This is a symmetric MAC with capacity $C_{\text{MAC,SUM}} = \log(4M) - 1$. As in Example 2, we choose the forwarding function to be the sufficient statistics $U_{\text{SUFF}} = (U_{\text{SUFF},1}, U_{\text{SUFF},2})$ given by Equation (6.39). Let $V_1 = \sum_{m=1}^{M} Y_m \mod 4M$ and $V_2 = \sum_{m=1}^{M} h_m \mod 4M$. Since $U_{\text{SUFF},1}, U_{\text{SUFF},2}$ can be recovered from (V_1, V_2), we use the linear the computation code developed in [27] to send (V_1, V_2) at computation rate $\kappa = \frac{C_{\text{MAC,SUM}}}{H(V_1,V_2)}$. In Figure 6.7, we compare the performance of different relaying strategies when the MAC has matched bandwidth ($\ell = 1$). We find that functional forwarding outperforms other strategies for $M \geq 2$ and is optimal for $M \geq 4$.

6.2.6 Multi-User Example

The previous examples consisted of single user networks. Here, we consider the effect of functional forwarding in the two stage binary network with two users and M relays (see Figure 6.8). At time i, each user transmits $X_j[i] \in \{0, 1\}$ for $j = 1, 2$ and the relays observe:

$$Y_m[i] = h_{m,1}[i]X_1[i] \oplus h_{m,2}[i]X_2[i] \tag{6.43}$$

where $h_{m,1}, h_{m,2} \sim$ i.i.d $\mathcal{B}(\frac{1}{2})$. The relays use a computation code to send the forwarding function to the destination. We consider a binary XOR MAC with inputs $X_{r,m} \in \{0, 1\}$ for all $m = 1, ..., M$ and output $Y_{\text{MAC}} = \oplus_m^M X_{r,m}$. This is a symmetric MAC with capacity $C_{MAC,SUM} = 1$. We assume a bandwidth expansion $\ell = 4$. The forwarding function $U =$

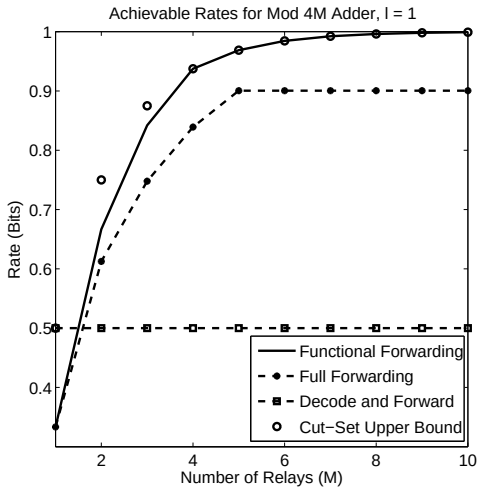

Figure 6.7. Achievable rates for binary network with mod 4M adder MAC from Example 3. In this case, functional forwarding is optimal when the number of relays is greater than 3.

(U_1, U_2) is chosen as follows:

$$U_1 = \mathbf{H}^T \mathbf{Y} \tag{6.44}$$

$$U_2 = \mathbf{H}^T \mathbf{H} \tag{6.45}$$

where $\mathbf{H}$ is the channel matrix and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]$ consists of the relay observations. We note that all operations are over the binary field. Using uncoded transmission, an achievable computation rate for the forwarding function is $\kappa_U = 0.2$. Using Theorem 1, the achievable sum rate with functional forwarding is given by

$$R = 0.8\mathbb{E}\left[rank\left(\mathbf{H}^T\mathbf{H}\right)\right] \tag{6.46}$$

Figure 6.9 shows the performance of different relaying schemes. We observe that although the considered version of functional forwarding is not optimal it outperforms full forwarding when there are three or more relays.

6.3 Extension to a Gaussian Network

In this section, we examine the performance of functional forwarding in a Gaussian Network (see Figure 6.10). We first describe the Gaussian channel model then show that functional forwarding can be much more efficient than full forwarding.

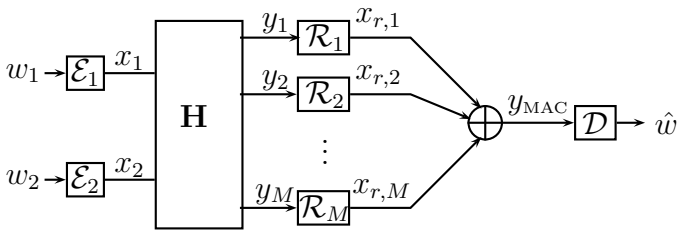

Figure 6.8: Two user binary network with binary OR MAC

6.3.1 Channel Model

We consider the two-stage Gaussian relay network that consists of a fading broadcast channel followed by a non-fading multiple-access channel (see Figure 6.10). The source chooses a message w uniformly from the set $\mathcal{W} = \left\{1, 2, ..., 2^{nR}\right\}$ and encodes into into a length n codeword $\mathbf{x}$:

$$\mathcal{E}_m : \mathcal{W} \to \mathbb{R}^n \tag{6.47}$$

We assume the typical power constraint is satisfied at the source:

$$\frac{1}{n}\sum_{i=1}^{n} x^2[i] \leq \mathsf{SNR}_s \tag{6.48}$$

At time i, the source transmits $x[i]$ and relay m observes:

$$y_m[i] = h_m[i]x[i] + z_m[i] \tag{6.49}$$

The fading coefficients are drawn independently from a Gaussian distribution: $h_m[i]$ i.i.d $\sim \mathcal{N}(0,1)$ and $\{z_m[i]\}_i$ is a white Gaussian process with unit variance. The fading coefficient $h_m[i]$ is assumed to be known at relay m but unknown at the destination as well as at the other relays. We assume ℓ uses of the MAC are allowed per use of the broadcast channel and find it interesting to consider the case of a bandwidth expansion. Relay m encodes its observation y_m^n and channel state information h_m^n into a length ℓn codeword $x_{r,m}^{\ell n}$:

$$\mathcal{R}_m : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{\ell n} \tag{6.50}$$

We assume a power constraint of SNR_r at the relays:

$$\frac{1}{\ell n}\sum_{i=1}^{\ell n} x_{r,m}^2[i] \leq \mathsf{SNR}_r \tag{6.51}$$

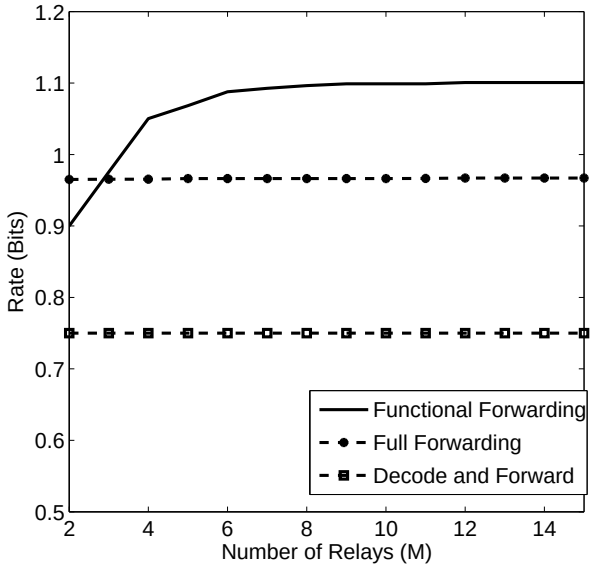

Figure 6.9. Achievable sum-rate for two-user binary network with binary XOR MAC, $\ell = 4$. Functional forwarding outperforms full forwarding in this case.

At time i, relay m transmits signal $x_{r,m}[i]$ to the destination. The destination observes a linear sum of the relay transmissions with additive white Gaussian noise:

$$y_{\text{MAC}}[i] = \sum_{m=1}^{M} x_{r,m}[i] + z_{\text{MAC}}[i], \tag{6.52}$$

where $\{z_{\text{MAC}}[i]\}_i$ is a Gaussian process with unit variance. The destination decodes the message:

$$\mathcal{D} : \mathbb{R}^{\ell n} \to \mathcal{W} \tag{6.53}$$

$$\hat{w} = \mathcal{D}(y_{\text{MAC}}^{\ell n}). \tag{6.54}$$

We require that the message be reliably recovered:

$$Pr(\hat{w} \neq w) \leq \epsilon \tag{6.55}$$

for all $\epsilon > 0$ for n large.

6.3.2 Forwarding Function

Let $\mathbf{y} = [y_1 \ \cdots \ y_M]^T$ represent the vector of relay observations, $\mathbf{h} = [h_1 \ \cdots \ h_M]^T$ be the vector of the channel state information and $\mathbf{z} = [z_1 \ \cdots \ z_M]^T$ be the vector of noise noise. It

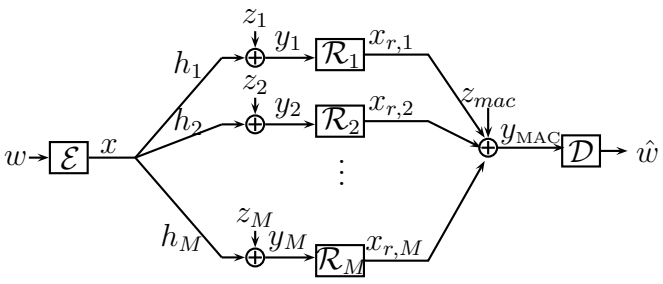

Figure 6.10: Single User Gaussian Network

follows that:

$$\mathbf{y} = \mathbf{h}x + \mathbf{z} \tag{6.56}$$

A sufficient statistic can be formed by projecting the relays' received signal onto the direction of the fading vector:

$$\mathbf{h}^T\mathbf{y} = \|\mathbf{h}\|^2 x + \tilde{z} \tag{6.57}$$

where $\tilde{z} \sim \mathcal{N}(0, \|\mathbf{h}\|^2 N)$. This reduces the broadcast channel to a point-to-point Gaussian channel with fading coefficient $\|\mathbf{h}\|^2$. We observe that $(\mathbf{h}^T\mathbf{y}, \|\mathbf{h}\|^2)$ is the sufficient statistic for the transmitted signal given the relay outputs and fading coefficients and have the following Markov Chain:

$$x \to (\mathbf{y}, \mathbf{h}) \to \left(\mathbf{h}^T\mathbf{y}, \|\mathbf{h}\|^2\right) \tag{6.58}$$

The vector of channel state information $\mathbf{h}$ is not needed to decode the transmitted signal but rather its norm $\|\mathbf{h}\|^2$ is sufficient. Given fading coefficients $\mathbf{h}$ and relay observations $\mathbf{y}$, we select the forwarding function to be the sufficient statistic:

$$U = \mathbf{h}^T\mathbf{y} \tag{6.59}$$

$$V = \|\mathbf{h}\|^2 \tag{6.60}$$

The relays use a computation code to first send U and then V across the multiple access channel to the destination. Note that U is linear in $h_m y_m$ and V is linear in h_m^2. Recently, it was shown in [27] that lattice codes can efficiently compute linear functions of Gaussian random variables over the Gaussian MAC. Our computation code is a modified version of that in [27].

6.3.3 Achievable Rates

The following theorem gives the achievable rate for the Gaussian network using functional forwarding.

Theorem 6.7. *Consider the Gaussian, two-stage relay channel with a bandwidth expansion of ℓ. For any $\ell_1, \ell_2 \in \mathbb{Z}_+$ such that $\ell_1 + \ell_2 = \ell$, the following rate is achievable*

$$R = \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{\hat{V}^2\mathsf{SNR}_s}{\mathbb{E}[(V-\hat{V})^2|\hat{V}]\mathsf{SNR}_s + \hat{V} + D_2}\right)\right] \tag{6.61}$$

where $\hat{V} = \mathbb{E}[V|V+Z]$, V is Chi-Squared with M degrees of freedom and Z is zero-mean Gaussian with variance D_1 given by

$$D_1 = (2\mathsf{SNR}_s + 1)\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_1} \tag{6.62}$$

and the constant D_2 is given by

$$D_2 = 2\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_2} \tag{6.63}$$

In the case where $\ell_1 = \ell_2 = 1$, our computation code involves only amplify and forward. In general, we use a modified version of the scheme from [27] that uses lattice codes from [16, 17]. The proof is given in Appendix G. In the following Corollary, we provide a lower bound on the achievable rate.

Corollary 6.8. *The functional forwarding achievable rate for the Gaussian two-stage network from Theorem 6.7 can be lower bounded as follows*

$$R \geq \frac{1}{2}\mathbb{E}\left[\log\left(\frac{V^2\mathsf{SNR}_s}{D_1\mathsf{SNR}_s + M + D_2}\right)\right] - 1 \tag{6.64}$$

$$\tag{6.65}$$

where V is Chi-Squared with M degrees of freedom and the constants:

$$D_1 = (2\mathsf{SNR}_s + 1)\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_1} \tag{6.66}$$

$$D_2 = 2\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_2} \tag{6.67}$$

Proof. Form Theorem 6.7, the achievable rate of the Gaussian Network is given by:

$$R = \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{\hat{V}^2\mathsf{SNR}_s}{\mathbb{E}[(V-\hat{V})^2|\hat{V}]\mathsf{SNR}_s + \hat{V}1 + D_2}\right)\right] \tag{6.68}$$

where $\hat{V} = \mathbb{E}[V|V+Z]$, V is Chi-Squared with M degrees of freedom and Z is zero-mean Gaussian with D_1 given by

$$D_1 = (2\mathsf{SNR}_s + 1)\left(\frac{N}{\mathsf{SNR}_r}\right)^{\ell_1} \tag{6.69}$$

and the constant D_2 is given by

$$D_2 = 2\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_2} \tag{6.70}$$

We further bound this mutual information as follows:

$$R = \frac{1}{2}\mathbb{E}\left[\log\left(1+\frac{\hat{V}^2\mathsf{SNR}_s}{\mathbb{E}[(V-\hat{V})^2|\hat{V}]\mathsf{SNR}_s+\hat{V}+D_2}\right)\right] \tag{6.71}$$

$$\geq \frac{1}{2}\mathbb{E}\left[\log\left(1+\frac{V^2\mathsf{SNR}_s}{\mathbb{E}[(V-\hat{V})^2|\hat{V}]\mathsf{SNR}_s+\hat{V}+D_2}\right)\right]-1 \tag{6.72}$$

$$\geq \frac{1}{2}\mathbb{E}\left[\log\left(\frac{V^2\mathsf{SNR}_s}{\mathbb{E}[(V-\hat{V})^2|\hat{V}]\mathsf{SNR}_s+\hat{V}+D_2}\right)\right]-1 \tag{6.73}$$

$$\geq \frac{1}{2}\mathbb{E}\left[\log\left(\frac{V^2\mathsf{SNR}_s}{D_1\mathsf{SNR}_s+M+D_2}\right)\right]-1 \tag{6.74}$$

□

In the next section, we compare the achievable rates for the different relaying techniques.

6.3.4 Example: Scaling Illustration

In this section, we characterize the performance of different relaying techniques when a large number of relays is in the network. We compare the performance of functional forwarding to full forwarding using compressed and forward at the relays and decode and forward (see [94] for a description of compressed-and-forward and decode-and-forward). We parametrize $\mathsf{SNR}_s, \mathsf{SNR}_r$ and ℓ with respect to the number of relays M as follows:

$$\mathsf{SNR}_s = \Theta(M^\alpha) \tag{6.75}$$

$$\mathsf{SNR}_r = \Theta(1) \tag{6.76}$$

$$\ell = \Theta(M^\beta) \tag{6.77}$$

where $\alpha, \beta > 0$[1]. We note that P_r, N are fixed to be constants and we assume that $\frac{P_r}{N} > 1$. Our regime of interest represents a regime of high SNR and a large number of relay nodes. It is well known that the presence of channel state information is crucial under high SNR. We show that forwarding full channel state information is inefficient when there is a large number of relays.

[1]The notation $f(n) = \Theta(g(n))$ means that there exist constants $C, C' > 0$ such that $f(n) \leq Cg(n)$ and $g(n) \leq C'f(n)$

From the standard cut-set bound [1, Theorem 14.10.1], the achievable rate for the Gaussian network must satisfy the following inequalities:

$$\begin{aligned} R &\leq \min\left\{\mathbb{E}\left[\frac{1}{2}\log\left(1+\|\mathbf{h}\|^2\mathsf{SNR}_s\right)\right], \frac{\ell}{2}\log\left(1+M^2\mathsf{SNR}_r\right)\right\} \\ &\leq \mathbb{E}\left[\frac{1}{2}\log\left(1+\|\mathbf{h}\|^2\mathsf{SNR}_s\right)\right] && (6.78) \\ &\leq \frac{1}{2}\log\left(1+\mathbb{E}[\|\mathbf{h}\|^2]\mathsf{SNR}_s\right) && (6.79) \\ &\leq \frac{1}{2}\log\left(1+M\mathsf{SNR}_s\right) && (6.80) \\ &\leq \frac{1}{2}\log(MM^{\alpha})+\Theta(1) \\ &= \frac{(1+\alpha)}{2}\log M+\Theta(1) && (6.81) \end{aligned}$$

We let R_{UPPER} denote an upper bound to the Gaussian network

$$R_{\text{UPPER}} = \frac{(1+\alpha)}{2}\log M+\Theta(1) \tag{6.82}$$

The achievable rate using decode and forward is given by

$$R_{\text{DECODE}} = \mathbb{E}\left[\frac{1}{2}\log\left(1+\|h_1\|^2\mathsf{SNR}_s\right)\right] \tag{6.83}$$

We define r_{DECODE} to be the ratio of R_{DECODE} and R_{UPPER} in our scaling regime of interest:

$$r_{\text{DECODE}} = \lim_{M\to\infty}\frac{R_{\text{DECODE}}}{R_{\text{UPPER}}} \tag{6.84}$$

From Equations (6.82), (6.83), it follows that:

$$\begin{aligned} r_{\text{DECODE}} &\leq \lim_{M\to\infty}\frac{\frac{1}{2}\log(M^{\alpha})+\Theta(1)}{\frac{(1+\alpha)}{2}\log M+\Theta(1)} && (6.85) \\ &= \frac{\alpha}{1+\alpha} && (6.86) \end{aligned}$$

Compared to the upper bound, we note that the decode and forward rate is suboptimal. This suggest that it is important for the destination to know the channel state information and relay observations from multiple relays before recovering the user's message.

Using the rate of functional forwarding given in Theorem 2, we evaluate the performance of functional forwarding compared against the upper bound as follows:

$$\begin{aligned} r_{\text{FUNCTION}} &= \lim_{M\to\infty} \frac{R_{\text{FUNCTION}}}{R_{\text{UPPER}}} \\ &\geq \frac{\frac{1}{2}\log\left(\frac{M^2 M^\alpha}{M^\alpha\left(\frac{P_r}{N}\right)^{M^\beta}+M}\right)-\Theta(1)}{\frac{(1+\alpha)}{2}\log M+\Theta(1)} \\ &= 1 \end{aligned}$$

We note that functional forwarding is optimal in our regime of interest. Finally, we compare the performance of full forwarding as a function of the upper bound. Using compress and forward to send the fading coefficients and relay observations individually, it can be shown that

$$\begin{aligned} r_{\text{FULL}} &= \lim_{M\to\infty} \frac{R_{\text{FULL}}}{R_{\text{UPPER}}} \\ &\geq \frac{\min\left\{\frac{1+\alpha}{2},\frac{\beta+\alpha}{2}\right\}\log M+\Theta(1)}{\frac{(1+\alpha)}{2}\log M+\Theta(1)} \\ &= \min\left\{1,\frac{\beta+\alpha}{1+\alpha}\right\} \end{aligned}$$

The scaling results are summarized in Table 6.1 and displayed in Figure 11. We note that functional forwarding is optimal in our scaling regime of interest while full forwarding is suboptimal when the MAC does not have enough bandwidth (in the case where $\beta < 1$). In this regime, there are M fading coefficients to be sent over the MAC but only M^β channel uses. Since sending multiple coefficients at once causes interference, full forwarding allows only M^β of the relays to send their information.

Table 6.1: Scaling results $r = \frac{R}{R_{\text{UPPER}}}$ for Single User Gaussian Network

Functional Forwarding	1
Full Forwarding	$\min\left\{1,\frac{\beta+\alpha}{1+\alpha}\right\}$
Decode and Forward	$\frac{\alpha}{1+\alpha}$

6.4 Conclusion

We propose a framework to forward a function of the channel state information for the two-stage fading network. We applied our framework to a series of examples and showed

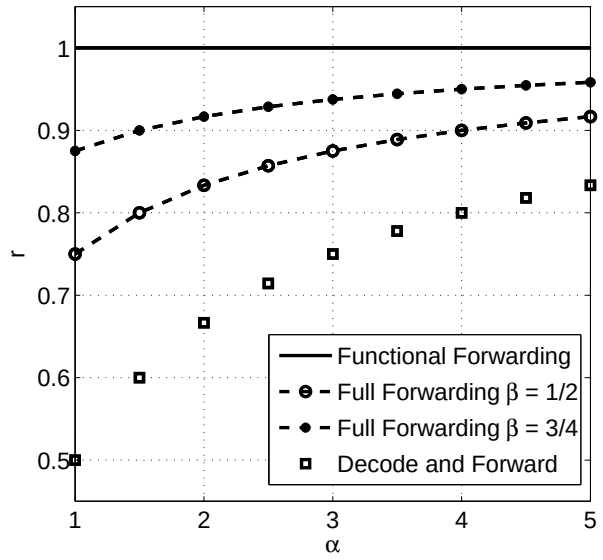

Figure 6.11. Achievable rates as a fraction of the upper bound for various relaying schemes for the Gaussian Network in Figure 6.10

that functional forwarding of channel state information can be much more efficient than full forwarding.

Chapter 7

Conclusion

In this thesis, we demonstrated the role of structured codes in information theory and focused on three MIMO and network settings. First, while considering a MIMO channel, it was shown that structured lattice codes can improve linear receiver design. Traditional linear receivers recover the individual data streams transmitted across the MIMO channel. Although these types of architectures are low in complexity, they experience a high performance loss compared to the theoretically optimal joint receiver. To bridge this performance gap, the proposed integer-forcing linear receiver instead recovers linear equations of the data streams. Standard random coding arguments are insufficient since equations of data streams are recovered, and lattice codes are used for their algebraic structure. The proposed receiver architecture achieves much better performance than traditional linear receivers at the cost of only slightly higher complexity. Second, we leveraged lattice codes to connect problems of computation over wireless networks to those of computation across wireline networks. By using lattice codes for both source quantization and source coding, the wireless network problem is converted into a deterministic wireline network problem. Tools from computation over wireline networks can then be applied. As a result, we characterized the distortion for transmitting the sum of Gaussian sources across a class of wireless relay-networks to within a constant gap of the optimal distortion. Finally, lattice codes can be used to transmit a function of the channel state information in relay networks where only partial channel state information is available. This is shown to be much more efficient than transmitting the full channel state information.

It is known that lattice codes achieve the optimal performance in many point-to-point settings and provide non-trivial gains over standard random codes in many network settings. This thesis further highlights the advantages of lattice codes through three network scenarios. A natural question that arises is the development of algebraically structured codes beyond lattice codes and their application to wireless network scenarios. Currently, lattice codes with similar encoding and decoding constructions are applied universally across many network settings. Although they are shown to be advantageous over random codes in many network scenarios, their optimality still remains to be proven for most cases. An interesting direction

for future research would be the development of algebraically structured codes specifically suited for particular network scenarios. Furthermore, the formal definition of a structured code and a random code remain to be determined. An interesting question would be to characterize the difference between the two and categorize problems based on the type of coding needed.

Bibliography

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. Wiley Series in Telecommunications and Signal Processing, Wiley-Interscience, 2 ed., July 2006.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.

[3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471 – 480, jul 1973.

[4] R. Ahlswede, "Group codes do not achieve shannon's channel capacity for general discrete channels," *The Annals of Mathematical Statistics*, vol. 42, pp. 224–240, February 1971.

[5] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, pp. 219 – 221, mar 1979.

[6] J. H. Conway, N. J. A. Sloane, and E. Bannai, *Sphere-packings, lattices, and groups*. New York, NY, USA: Springer-Verlag New York, Inc., 1987.

[7] R. Koetter, M. Effros, T. Ho, and M. Medard, "Network codes as codes on graphs," in *Conference on Information Sciences and Systems (CISS)*, Mar 2004.

[8] A. Ramamoorthy and M. Langberg, "Communicating the sum of sources over a network," *ArXiv e-prints*, Jan. 2010.

[9] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Linear Codes, Target Function Classes, and Network Computing Capacity," *ArXiv e-prints*, Dec. 2011.

[10] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network coding for computing: Cut-set bounds," *IEEE Transactions on Information Theory*, vol. 57, pp. 1015 –1030, feb. 2011.

[11] R. Ahlswede, N. Cai, S. yen Robert Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[12] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 371 –381, feb. 2003.

[13] R. de Buda, "Some optimal codes have structure," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 893–899, 1989.

[14] T. Linder, C. Schlegal, and K. Zeger, "Corrected proof of de buda's theorem [lattice channel codes]," *IEEE Transactions on Information Theory*, vol. 39, pp. 1735 –1737, sep 1993.

[15] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the awgn channel," *IEEE Transactions on Information Theory*, vol. 44, pp. 273 –278, jan 1998.

[16] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, pp. 2293–2314, October 2004.

[17] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, pp. 3401–3416, October 2005.

[18] R. Zamir, "Lattices are everywhere," in *4th Annual Workshop on Information Theory and its Applications, UCSD*, (La Jolla, CA), February 2009.

[19] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, May 1983.

[20] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3820–3833, 2005.

[21] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, pp. 3417–3432, October 2005.

[22] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, pp. 1250–1276, June 2002.

[23] D. Krithivasan and S. Pradhan, "Lattices for distributed source coding: Jointly gaussian sources and reconstruction of a linear function," *IEEE Transactions on Information Theory*, vol. 55, pp. 5628 –5651, dec. 2009.

[24] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k - user interference channel," *IEEE Transactions on Information Theory*, vol. 54, pp. 3425 –3441, aug. 2008.

[25] G. Bresler, A. Parekh, and D. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 56, pp. 4566 –4592, sept. 2010.

[26] B. Nazer, S. Jafar, M. Gastpar, and S. Vishwanath, "Ergodic interference alignment," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 1769 –1773, 28 2009-july 3 2009.

[27] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, pp. 3498 –3516, oct. 2007.

[28] M. Wilson, K. Narayanan, H. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Transactions on Information Theory*, vol. 56, pp. 5641 –5654, nov. 2010.

[29] W. Nam, S.-Y. Chung, and Y. Lee, "Capacity of the gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Transactions on Information Theory*, vol. 56, pp. 5488 –5494, nov. 2010.

[30] B. Nazer and M. Gastpar, "Computing over multiple-access channels with connections to wireless network coding," in *Information Theory, 2006 IEEE International Symposium on*, pp. 1354 –1358, july 2006.

[31] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, pp. 6463–6486, October 2011.

[32] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, pp. 438 –460, march 2011.

[33] U. Niesen, B. Nazer, and P. Whiting, "Computation Alignment: Capacity Approximation without Noise Accumulation," *ArXiv e-prints*, Aug. 2011.

[34] T. Philosof, R. Zamir, U. Erez, and A. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5006–5035, 2011.

[35] X. He and A. Yener, "Providing Secrecy With Structured Codes: Tools and Applications to Two-User Gaussian Channels," *ArXiv e-prints*, July 2009.

[36] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, March 1998.

[37] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, November - December 1999.

[38] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, June 2000.

[39] "IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks– specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput," *IEEE Std 802.11n-2009*, pp. c1–502, 29 2009.

[40] K. Kumar, G. Caire, and A. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Transactions on Information Theory*, vol. 55, pp. 4398 – 4418, October 2009.

[41] D. Bliss, K. Forsythe, A. Hero III, and A. Yegulalp, "Environmental issues for MIMO capacity," *IEEE Trans. Signal Processing*, vol. 50, pp. 2128 – 2142, November 2002.

[42] J. Winters, J. Salz, and R. Gitlin., "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, pp. 1740 –1751, Feb/Mar/Apr 1994.

[43] A. Motahari, S. Oveis-Gharan, and A. Khandani, "Real interference alignment with real numbers," *IEEE Transactions on Information Theory*, Submitted August 2009. Also available at [arXiv:0908.1208].

[44] U. Niesen and P. Whiting, "The degrees of freedom of compute-and-forward," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011)*, (St. Petersburg, Russia), August 2011.

[45] E. Viterbo and J. Boutros, "A universal lattice decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 1639–1642, July 1999.

[46] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, pp. 2201–2214, August 2002.

[47] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, pp. 2389–2402, October 2003.

[48] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2806 – 2818, August 2005.

[49] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1474–1484, April 2005.

[50] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1566 – 1577, July 2005.

[51] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, pp. 619 – 637, February 2001.

[52] M. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization and successive decoding achieves the total capacity of the gaussian multiple-access channel," in *Conference Record of the 31st Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA , USA), pp. 1405 – 1409, November 1997.

[53] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *URSI International Symposium on Signals, Systems, and Electronics (ISSSE 98)*, (Pisa, Italy), Sept.-Oct. 1998.

[54] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[55] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM 2002)*, (Taipei, Taiwan), November 2002.

[56] M. Taherzadeh, A. Mobasher, and A. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Transactions on Information Theory*, vol. 53, pp. 4801–4805, December 2007.

[57] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 968–985, June 2004.

[58] J. Jalden and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs and system models," *IEEE Transactions on Information Theory*, vol. 56, pp. 4765–4780, October 2010.

[59] O. Ordentlich, J. Zhan, U. Erez, M. Gastpar, and B. Nazer, "Practical code design for compute-and-forward," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011)*, (St. Petersburg, Russia), August 2011.

[60] J. Zhan, B. Nazer, , O. Ordentlich, U. Erez, and M. Gastpar, "Integer-forcing architectures for mimo: Distributed implementation and sic integer-forcing architectures for MIMO: Distributed implementation and SIC," in *Conference Record of the 44th Asilomar Conference on Signals, Systems, and Computers*, (Monterey, CA), November 2010.

[61] A. Sanderovich, S. Shamai, , H. V. Poor, and Y. Steinberg, "Uplink macro diversity of limited backhaul cellular network," *IEEE Transactions on Information Theory*, vol. 55, pp. 3457–3478, August 2009.

[62] A. Sanderovich, S. Shamai, and Y. Steinberg, "Distributed MIMO receiver – achievable rates and upper bounds," *IEEE Transactions on Information Theory*, vol. 55, pp. 4419–4438, October 2009.

[63] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, pp. 5534–5562, December 2008.

[64] J. W. S. Cassels, *An Introduction to Diophantine Approximations*. Cambridge University Press, 1957.

[65] V. I. Bernik and M. M. Dodson, "Metric Diophantine approximation on manifolds," 1999.

[66] R. Appuswamy and M. Franceschetti, "Computing linear functions by linear coding over networks," *ArXiv e-prints*, Feb. 2011.

[67] M. Langberg and A. Ramamoorthy, "Communicating the sum of sources in a 3-sources/3-terminals network; revisited," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1853 –1857, june 2010.

[68] H.-I. Su and A. El Gamal, "Distributed lossy averaging," *IEEE Transactions on Information Theory*, vol. 56, pp. 3422 –3437, july 2010.

[69] B. Nazer and M. Gastpar, "The case for structured random codes in network capacity theorems," *European Transactions on Telecommunications, Special Issue on New Directions in Information Theory*, vol. 19, pp. 455–474, June 2008.

[70] R. Soundararajan and S. Vishwanath, "Communicating linear functions of correlated gaussian sources over a mac," *IEEE Transactions on Information Theory*, vol. 58, pp. 1853 –1860, march 2012.

[71] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, pp. 782–795, 2003.

[72] C.-C. Wang and N. B. Shroff, "Beyond the butterfly - a graph-theoretic characterization of the feasibility of network coding with two simple unicast sessions," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 121 –125, june 2007.

[73] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658 – 2668, oct. 2003.

[74] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of gaussian multiple-access and broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 768 – 783, may 2004.

[75] M. Kim and M. Medard, "Algebraic network coding approach to deterministic wireless relay networks," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 1518 –1525, 29 2010-oct. 1 2010.

[76] S. Y. Park and A. Sahai, "An algebraic mincut-maxflow theorem," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 608 –612, 31 2011-aug. 5 2011.

[77] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, pp. 1872 –1905, april 2011.

[78] M. Maddah-Ali and D. Tse, "Interference neutralization in distributed lossy source coding," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 166 –170, june 2010.

[79] A. Wagner, "On distributed compression of linear functions," *IEEE Transactions on Information Theory*, vol. 57, pp. 79 –94, jan. 2011.

[80] V. Prabhakaran, S. Diggavi, and D. Tse, "Broadcasting with degraded message sets: A deterministic approach," in *Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing*, 2007.

[81] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & systems (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[82] S.-Y. Li and R. Yeung, "On convolutional network coding," in *Information Theory, 2006 IEEE International Symposium on*, pp. 1743 –1747, july 2006.

[83] L. Zheng and D. Tse, "Communication on the grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *Information Theory, IEEE Transactions on*, vol. 48, pp. 359 –383, feb 2002.

[84] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2426–2467, 2003.

[85] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 50, pp. 3095 – 3110, dec. 2004.

[86] N. Ravindran and N. Jindal, "Multi-User Diversity vs. Accurate Channel State Information in MIMO Downlink Channels," *ArXiv e-prints*, July 2009.

[87] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates with Downlink Training and Channel State Feedback," *ArXiv e-prints*, Nov. 2007.

[88] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over mimo x channels: Interference alignment, decomposition, and performance analysis," *Information Theory, IEEE Transactions on*, vol. 54, pp. 3457 –3470, aug. 2008.

[89] J. Thukral and H. Bölcskei, "Interference alignment with limited feedback," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pp. 1759–1763, June 2009.

[90] Y. Zhu and D. Guo, "The degrees of freedom of isotropic mimo interference channels without state information at the transmitters," *Information Theory, IEEE Transactions on*, vol. 58, pp. 341 –352, jan. 2012.

[91] A. Sanderovich, O. Somekh, H. Poor, and S. Shamai, "Uplink macro diversity of limited backhaul cellular network," *Information Theory, IEEE Transactions on*, vol. 55, pp. 3457 –3478, aug. 2009.

[92] O. Somekh, B. Zaidel, and S. Shamai, "Sum rate characterization of joint multiple cell-site processing," *Information Theory, IEEE Transactions on*, vol. 53, pp. 4473 –4497, dec. 2007.

[93] B. Nazer, A. Sanderovich, M. Gastpar, and S. Shamai, "Structured superposition for backhaul constrained cellular uplink," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 1530 –1534, 28 2009-july 3 2009.

[94] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *Information Theory, IEEE Transactions on*, vol. 51, pp. 3037 – 3063, sept. 2005.

[95] J. Lagarias, H. Lenstra Jr., and C. Schnorr, "Korkin-zolotarev bases and successive minima of a lattice and its reciprocal lattice," *Combinatorica*, vol. 10, pp. 333–348, December 1990.

[96] S. Lang, *Metric Diophantine Approximation on Manifolds*. New York, NY: Springer-Verlag, 1995.

[97] I. Aliev and M. Henk, "Successive minima and best simultaneous Diophantine approximations," *Monatshefte Für Mathematik*, vol. 147, pp. 95–101, 2006. Available at [arXiv:math/0503365v1].

[98] M. Taherzadeh, A. Mobasher, and A. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Transactions on Information Theory*, vol. 53, pp. 4567–4582, December 2007.

[99] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *Information Theory, IEEE Transactions on*, vol. 49, pp. 951 – 963, april 2003.

Appendix A

Integer-Forcing vs. V-Blast IV

Recall that V-BLAST II performs decoding in the optimal order and V-BLAST III allows for rate allocation. In this appendix, we introduce V-BLAST IV, which allows for both rate allocation and an optimized decoding order. Under V-BLAST IV, the data streams are decoded with respect to the ordering

$$\pi^* = \text{argmax}_{\pi \in \Pi} \min_m 2M R_{\pi(m)}(\mathbf{H}). \tag{A.1}$$

where $R_{\pi(m)}(\mathbf{H})$ is given by (3.6). We compare the behavior of V-BLAST IV to that of the integer-forcing linear receiver in Figures A.1, A.2, A.3. The results show although V-BLAST IV achieves good performance for low to medium SNR, the integer-forcing linear achieves higher outage rates and lower outage probabilities in the medium to high SNR regime.

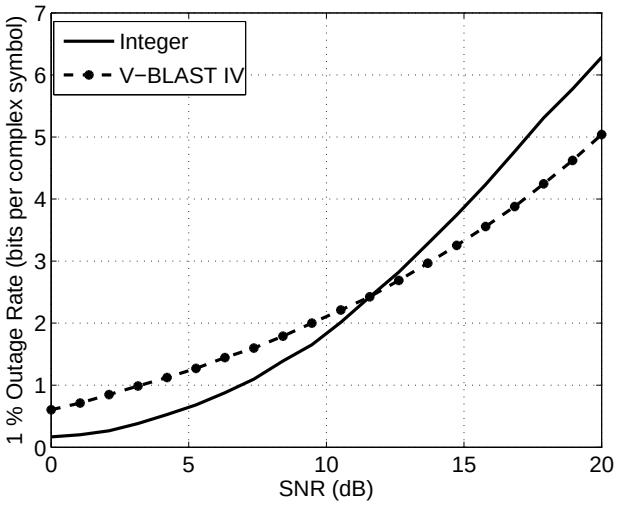

Figure A.1. 1 percent outage rates for the 2×2 complex-valued MIMO channel with Rayleigh fading.

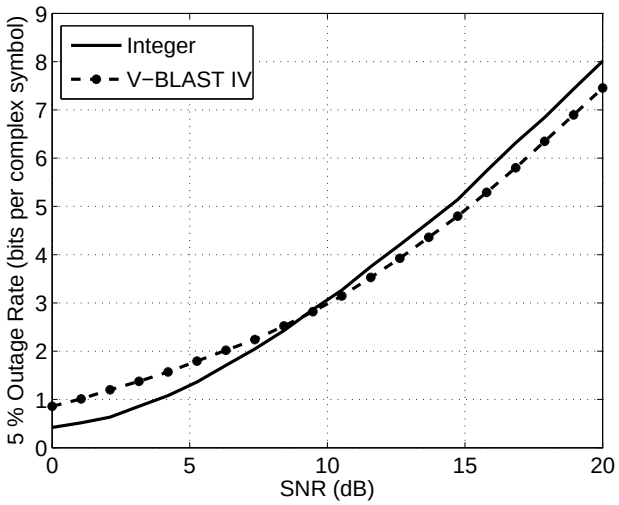

Figure A.2. 5 percent outage rates for the 2×2 complex-valued MIMO channel with Rayleigh fading.

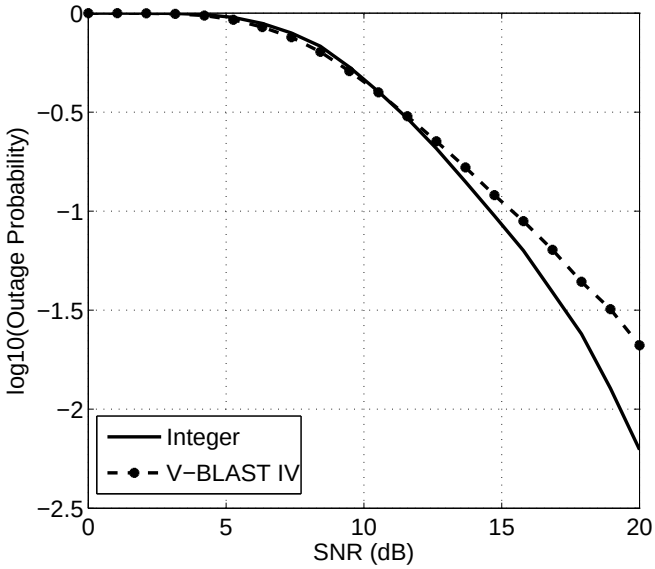

Figure A.3. Outage probability for the 2×2 complex-valued MIMO channel with Rayleigh fading for a target sum rate of $R = 6$.

Appendix B

Proof of Theorem 2.21

In order to establish Theorem 2.21, we need a few key facts about lattices.

Definition B.1 (Lattice)**.** A lattice $\Lambda \subset \mathbb{R}^{2M}$ is a set of points that satisfy the following properties:

$$\text{i) } \mathbf{0} \in \Lambda \tag{B.1}$$

$$\text{ii) if } \mathbf{x}, \mathbf{y} \in \Lambda \text{ then } \mathbf{x} + \mathbf{y} \in \Lambda. \tag{B.2}$$

We call the rank-L matrix $\mathbf{G}$ a generator matrix for Λ if

$$\Lambda = \left\{ \mathbf{G}\mathbf{d} : \mathbf{d} \in \mathbb{Z}^{2M} \right\} \tag{B.3}$$

We use the definition of dual lattices from [95].

Definition B.2 (Dual Lattice)**.** Given a lattice $\Lambda \subset \mathbb{R}^{2M}$ with a rank-L generator matrix $\mathbf{G}$, the dual lattice Λ^* has generator matrix $\left(\mathbf{G}^T\right)^\dagger$,

$$\Lambda^* = \left\{ \left(\mathbf{G}^T\right)^\dagger \mathbf{d} : \mathbf{d} \in \mathbb{Z}^{2M} \right\} \ . \tag{B.4}$$

To prove Theorem 2.21, we consider *successive minima* for the involved lattices, a standard concept from the Diophantine approximation literature (see e.g. [64, 96, 97]), defined as follows.

Definition B.3 (Successive Minima)**.** Let $\mathcal{B} = \left\{ \mathbf{x} \in \mathbb{R}^{2M} : \|\mathbf{x}\| \leq 1 \right\}$ be the unit ball. Given a lattice $\Lambda \subset \mathbb{R}^{2M}$ with a rank-L generator matrix, the m^{th} successive minimum $\epsilon_m(\Lambda)$ where $1 \leq m \leq L$ is given by

$$\epsilon_m(\Lambda) = \{\min \epsilon : \exists\ m \text{ linearly independent lattice points } \mathbf{v}_1, \ldots, \mathbf{v}_m \in \Lambda \cap \epsilon\mathcal{B}\}$$

Remark B.4. Definition B.3 implies that $\epsilon_1(\Lambda) \leq \epsilon_2(\Lambda) \leq \cdots \leq \epsilon_L(\Lambda)$ for any lattice Λ.

The following basic property linking the successive minima of a lattice with those of its dual lattice is key to our proof.

Lemma B.5 ([95, Proposition 3.3]). *Let $\Lambda \subset \mathbb{R}^{2M}$ be an arbitrary lattice with a rank-L generator matrix and Λ^* be its dual lattice. The successive minima for Λ and Λ^* satisfy the following inequality:*

$$\epsilon_\ell^2(\Lambda^*)\epsilon_1^2(\Lambda) \leq \frac{m^2(m+3)}{4} \text{ for } m = 1, 2, \ldots, L \ . \tag{B.5}$$

Finally, we also need the following result concerning a random Gaussian lattice.

Lemma B.6 ([98, Lemma 3]). *Let $\mathbf{H} \in \mathbb{R}^{2N\times 2M}$ be the real-valued decomposition of a $N \times M$ complex Gaussian matrix with i.i.d. Rayleigh entries. Let $\Lambda = \left\{\mathbf{Hd} : \mathbf{d} \in \mathbb{Z}^{2M}\right\}$ be the lattice generated by $\mathbf{H}$. Then*

$$\Pr(\epsilon_1(\Lambda) \leq s) = \begin{cases} \gamma s^{2N}, & M < N, \\ \delta s^{2N} \max\left\{-(\ln s)^{N+1}, 1\right\}, & M = N. \end{cases}$$

where γ and δ are constants independent of s.

We now provide the proof of Theorem 2.21. Let $R = r \log \mathsf{SNR}$ be the target rate where $r \in [0, M]$. For a fixed set of equations $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_{2M}]^T$ and a fixed preprocessing matrix $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_{2M}]^T$, the outage probability is given by

$$\begin{aligned} p_{\text{OUT}}(r, \mathbf{A}, \mathbf{B}) &= \Pr\left(R(\mathbf{H}, \mathbf{A}, \mathbf{B}) < r \log \mathsf{SNR}\right) \\ &= \Pr\left(\min_m R(\mathbf{H}, \mathbf{a}_m, \mathbf{b}_m) < \frac{r}{2M} \log \mathsf{SNR}\right) \\ &= \Pr\left(\max_m \|\mathbf{b}_m\|^2 + \mathsf{SNR}\|\mathbf{H}^T\mathbf{b}_m - \mathbf{a}_m\|^2 > \mathsf{SNR}^{1-\frac{r}{M}}\right) \end{aligned}$$

For a fixed set of equations $\mathbf{A}$, we are free to choose any projection matrix $\mathbf{B}$, resulting in the following bound:

$$\begin{aligned} p_{\text{OUT}}(r, \mathbf{A}) &= \min_{\mathbf{B}} p_{\text{OUT}}(r, \mathbf{A}, \mathbf{B}) \\ &\leq p_{\text{OUT}}(r, \mathbf{A}, \mathbf{A}\mathbf{H}^\dagger) \\ &= \Pr\left(\max_m \left\|\left(\mathbf{H}^T\right)^\dagger \mathbf{a}_m\right\|^2 > \mathsf{SNR}^{1-\frac{r}{M}}\right) \end{aligned}$$

We then choose the best set of full-rank equations by optimizing (B.6) over all integer matrices $\mathbf{A} \in \mathbb{Z}^{2M\times 2M}$ with non-zero determinant:

$$p_{\text{OUT}}(r) = \min_{\mathbf{A}:|\mathbf{A}|>0} p_{\text{OUT}}(r,\mathbf{A}) \tag{B.6}$$

$$\leq \min_{\mathbf{A}:|\mathbf{A}|>0} \Pr\left(\max_m \left\|\left(\mathbf{H}^T\right)^\dagger \mathbf{a}_m\right\|^2 > \mathsf{SNR}^{1-\frac{r}{M}}\right) \tag{B.7}$$

$$= \Pr\left(\min_{\mathbf{A}:|\mathbf{A}|>0} \max_m \left\|\left(\mathbf{H}^T\right)^\dagger \mathbf{a}_m\right\|^2 > \mathsf{SNR}^{1-\frac{r}{M}}\right) \tag{B.8}$$

We use properties of dual lattices to bound (B.8). For a fixed $\mathbf{H}$, let Λ_{CHANNEL} be the lattice generated by $\mathbf{H}$ and Λ_{DUAL} be the dual lattice generated by $\left(\mathbf{H}^T\right)^\dagger$,

$$\Lambda_{\text{CHANNEL}} = \left\{\mathbf{Hd} : \mathbf{d} \in \mathbb{Z}^{2M}\right\} \tag{B.9}$$

$$\Lambda_{\text{DUAL}} = \left\{\left(\mathbf{H}^T\right)^\dagger \mathbf{d} : \mathbf{d} \in \mathbb{Z}^{2M}\right\}. \tag{B.10}$$

Using the definition of successive minima (Definition B.3), it follows that

$$\min_{\mathbf{A}:|\mathbf{A}|>0} \max_m \left\|\left(\mathbf{H}^T\right)^\dagger \mathbf{a}_m\right\|^2 = \max_{m=1,\cdots,2M} \epsilon_m(\Lambda_{\text{DUAL}}) \tag{B.11}$$

$$= \epsilon_{2M}(\Lambda_{\text{DUAL}})\ . \tag{B.12}$$

We now express (B.8) in terms of the successive minima of Λ_{DUAL},

$$p_{\text{OUT}}(r) \leq \Pr\left(\min_{\mathbf{A}:|\mathbf{A}|>0} \max_m \left\|\left(\mathbf{H}^T\right)^\dagger \mathbf{a}_m\right\|^2 > \mathsf{SNR}^{1-\frac{r}{M}}\right) \tag{B.13}$$

$$= \Pr\left(\epsilon_{2M}^2(\Lambda_{\text{DUAL}}) > \mathsf{SNR}^{1-\frac{r}{M}}\right) \tag{B.14}$$

Using Lemma B.5, we can bound the successive minima of Λ_{DUAL} in terms of the successive minima of Λ_{CHANNEL},

$$\epsilon_{2M}^2(\Lambda_{\text{DUAL}}) \leq \frac{2M^3+3M^2}{\epsilon_1^2(\Lambda_{\text{CHANNEL}})}\ . \tag{B.15}$$

Combining (B.14) and (B.15), the outage probability is upper bounded by

$$p_{\text{OUT}}(r) \leq \Pr\left(\frac{2M^3+3M^2}{\epsilon_1^2(\Lambda_{\text{CHANNEL}})} > \mathsf{SNR}^{1-\frac{r}{M}}\right) \tag{B.16}$$

$$= \Pr\left(\epsilon_1^2(\Lambda_{\text{CHANNEL}}) < \frac{2M^3+3M^2}{\mathsf{SNR}^{1-\frac{r}{M}}}\right) \tag{B.17}$$

This probability can in turn be upper bounded using Lemma B.6. For large SNR, we find that

$$p_{\mathrm{OUT}}(r) \leq \frac{\max\{\gamma,\delta\}\,(2M^3+3M^2)^N\,(\ln \mathsf{SNR})^{N+1}}{\mathsf{SNR}^{N\left(1-\frac{r}{M}\right)}}.$$

where γ, δ are constants independent of SNR. The achievable diversity for multiplexing gain r is thus

$$\begin{aligned} d(r) &= \lim_{\mathsf{SNR}\to\infty} \frac{-\log \rho_{\mathrm{OUT}}(r)}{\mathsf{SNR}} && \text{(B.18)} \\ &\geq \lim_{\mathsf{SNR}\to\infty} \frac{N\left(1-\frac{r}{M}\right)\mathsf{SNR}}{\mathsf{SNR}} - \frac{o(\mathsf{SNR})}{\mathsf{SNR}} && \text{(B.19)} \\ &= N\left(1-\frac{r}{M}\right) && \text{(B.20)} \end{aligned}$$

Appendix C

Proof of Theorem 5.17

We fix a broadcast network $\mathcal{N}_{\text{DET-BC}}$ with nodes Ω and communication demands $\mathcal{P}$. We assume that the linear time invariant code $\{\mathbf{K}_i\}_{i\in\Omega}$ achieves the rate tuple $(R_1, \cdots R_\ell)$. For clarity purposes, we use the notation $N_{\text{BC},S}$ to denote the source node, $N_{\text{BC},D_1}, \ldots, N_{\text{BC},D_m}$ denote the m destination nodes and N_i for $i = 1, \ldots, r$ to denote the r relay nodes. Let G_{BC} be the overall transfer function for the broadcast network $\mathcal{N}_{\text{DET-BC}}$ from the input of the source node $N_{\text{BC},S}$ to the output of the destination nodes $N_{\text{BC},D_1}, \ldots, N_{\text{BC},D_m}$. Since $X_{\text{BC},D_i} = \emptyset$, it follows that

$$G_{\text{BC}} = [G^T_{\text{BC},S,D_1} \cdots G^T_{\text{BC},S,D_m}]^T \tag{C.1}$$

where G_{BC,S,D_i} is the transfer function from node $N_{\text{BC},S}$ to node N_{BC,D_i}. Let $\mathcal{N}_{\text{DET-MAC}}$ be the dual multiple-access channel with source nodes $N_{\text{MAC},S_1}, \cdots N_{\text{MAC},S_m}$ and destination node $N_{\text{MAC},D}$. We assume that $\mathcal{N}_{\text{MAC}}$ applies the linear time invariant code $\left\{\mathbf{K}_i^T\right\}_{i\in\Omega}$. Let G_{MAC} be the overall transfer function for the dual multiple-access network from source nodes $N_{\text{MAC},S_1}, \ldots, N_{\text{MAC},S_m}$ to the destination node $N_{\text{MAC},D}$. Since $Y_{\text{MAC},S_i} = \emptyset$ for all i, it follows that

$$G_{\text{MAC}} = [G_{\text{MAC},S_1,D} \cdots G_{\text{MAC},S_m,D}]^T \tag{C.2}$$

where $G_{\text{MAC},S_i,D}$ is the transfer function from N_{MAC,S_i} to the destination node $N_{\text{MAC},D}$. We show that $G^T_{\text{BC}} = G_{\text{MAC}}$. From (C.1) and (C.2), it is sufficient to show that $G^T_{\text{BC},S,D_i} = G_{\text{MAC},S_i,D}$ for all $i = 1, \ldots, m$. From [76], the transfer function G_{BC,S,D_i} is given by

$$G_{\text{BC},S,D_i} = [H_{\text{BC},1,D_i}K_1 \cdots H_{\text{BC},r,D_i}K_r] \left(I - \begin{bmatrix} H_{\text{BC},1,1}K_1 & \cdots & H_{\text{BC},r,1}K_r \\ \vdots & & \vdots \\ H_{\text{BC},1,r}K_1 & \cdots & H_{\text{BC},r,r}K_r \end{bmatrix} \right)^{-1} \begin{bmatrix} H_{\text{BC},S,1} \\ \vdots \\ H_{\text{BC},S,r} \end{bmatrix} \tag{C.3}$$

(C.4)

and the transfer function $G_{\text{MAC},S_i,D}$ is given by

$$G_{\text{MAC},S_i,D} = [H_{\text{MAC},1,D}K_1^T \cdots H_{\text{MAC},q,D}K_r^T] \left(I - \begin{bmatrix} H_{\text{MAC},1,1}K_1^T & \cdots & H_{\text{MAC},r,1}K_r^T \\ \vdots & & \vdots \\ H_{\text{MAC},1,r}K_1^T & \cdots & H_{\text{MAC},r,r}K_r^T \end{bmatrix} \right)^{-1} \tag{C.5}$$

$$\begin{bmatrix} H_{\text{MAC},S_i,1} \\ \vdots \\ H_{\text{MAC},S_i,r} \end{bmatrix} + H_{\text{MAC},S_i,D} \tag{C.6}$$

We note that

$$H_{\text{MAC},S_i,D} = H_{\text{BC},S,D_i}^T \quad \text{for all } i = 1 \cdots m \tag{C.7}$$

$$H_{\text{MAC},S_i,j} = H_{\text{BC},j,D_i}^T \quad \text{for all } i = 1 \cdots m, j = 1 \cdots r \tag{C.8}$$

$$H_{\text{MAC},i,j} = H_{\text{BC},j,i}^T \quad \text{for all } i = 1 \cdots r, j = 1 \cdots r \tag{C.9}$$

$$H_{\text{MAC},j,D} = H_{\text{BC},S,j}^T \quad \text{for all } j = 1 \cdots r \tag{C.10}$$

Hence, we can rewrite

$$G_{\text{MAC},S_i,D} = [H_{\text{BC},S,1}^T K_1^T \cdots H_{\text{BC},S,r}^T K_r^T] \left(I - \begin{bmatrix} H_{\text{BC},1,1}^T K_1^T & \cdots & H_{\text{BC},1,r}^T K_r^T \\ \vdots & & \vdots \\ H_{\text{BC},r,1}^T K_1^T & \cdots & H_{\text{BC},r,r}^T K_r^T \end{bmatrix} \right)^{-1} \tag{C.11}$$

$$\begin{bmatrix} H_{\text{BC},1,D_i}^T \\ \vdots \\ H_{\text{BC},r,D_i}^T \end{bmatrix} + H_{\text{BC},S,D_i}^T$$

Taking the transpose of G_{BC,S,D_i}, we have that

$$G_{\text{BC},S,D_i}^T = [H_{\text{BC},1,D_i} \cdots H_{\text{BC},q,D_i}] \tag{C.12}$$

$$\left(I - \begin{bmatrix} K_1 H_{\text{BC},1,1} & \cdots & K_1 H_{\text{BC},q,1} \\ \vdots & & \vdots \\ K_q H_{\text{BC},1,q} & \cdots & K_q H_{\text{BC},q,q} \end{bmatrix} \right)^{-1} \begin{bmatrix} K_1 H_{\text{BC},1,S} \\ \vdots \\ K_q H_{\text{BC},q,S} \end{bmatrix} + H_{\text{BC},S,D_i}$$

We define the matrices $\mathbf{H}_S, \mathbf{H}_D, \mathbf{H}, \mathbf{K}$ as follows:

$$\mathbf{H}_S = [H_{\text{BC},S,1}^T \ldots H_{\text{BC},S,q}^T]^T, \quad \mathbf{H}_D = [H_{\text{BC},1,D} \ldots H_{\text{BC},q,D}]^T \tag{C.13}$$

$$\mathbf{H} = \begin{bmatrix} H_{\text{BC},1,1} & \cdots & H_{\text{BC},q,1} \\ \vdots & & \vdots \\ H_{\text{BC},1,q} & \cdots & H_{\text{BC},q,q} \end{bmatrix}, \quad \mathbf{K} = diag\left(\mathbf{K}_1, \ldots, \mathbf{K}_q\right)$$

Using the notation in (C.13), (C.5) and (C.12) can be rewritten as follows

$$G_{\text{MAC},S_i,D} = \mathbf{H}_D^T \mathbf{K}(\mathbf{I} - \mathbf{HK})^{-1}\mathbf{H_S} + H_{\text{BC},S,D_i} \tag{C.14}$$

$$G_{\text{BC},S,D_i}^T = \mathbf{H}_D^T (\mathbf{I} - \mathbf{KH})^{-1}\mathbf{KH_S} + H_{\text{BC},S,D_i} \tag{C.15}$$

From simple linear algebra, it can be shown that $\mathbf{K}(\mathbf{I} - \mathbf{HK})^{-1} = (\mathbf{I} - \mathbf{KH})^{-1}\mathbf{K}$. Hence, $G_{\text{MAC},S_i,D} = G_{\text{BC},S,D_i}^T$.

Let $\mathbf{K}_S$ be the preprocessing matrix and $\mathbf{K}_{D_1} \cdots \mathbf{K}_{D_m}$ be the postprocessing matrices for $\mathcal{N}_{\text{DET-BC}}$. The end-to-end transfer function to destinations $D_1 \dots D_m$ are given by $G_{\text{BC, end-end}} = [\mathbf{K}_{S,D_i}^T \dots \mathbf{K}_{S,D_m}^T]^T G_{\text{BC}} \mathbf{K}_S$. By assumption, rates $R_1, \cdots R_\ell$ are achievable for $\mathcal{N}_{\text{DET-BC}}$. By using the same linear code (with matrix transpose), the end-to-end transfer function for $\mathcal{N}_{\text{DET-MAC}}$ is given by

$$G_{\text{MAC, end-end}} = \mathbf{K}_S^T G_{\text{MAC}} [\mathbf{K}_{S,D_i}^T \dots \mathbf{K}_{S,D_m}^T]^T \tag{C.16}$$

$$= \mathbf{K}_S^T G_{\text{BC}}^T [\mathbf{K}_{S,D_i} \dots \mathbf{K}_{S,D_m}] \tag{C.17}$$

$$= G_{\text{MAC, end-end}}^T \tag{C.18}$$

Since we want to compute ℓ functions under demands $\mathcal{Q} = \mathcal{P}$, computation rates $R_1, \cdots, R_\ell$ are achievable.

Appendix D

Proof of Theorem 5.27

We first show that cut-set is universally tight for communication demands:

$$\mathcal{P}^{(\ell)}_{\text{TIGHT},1} = \{\{1\} \cup \{\ell, \ldots, m\}, \ldots, \{\ell - 1\} \cup \{\ell, \ldots, m\}, \{\ell, \ldots, m\}\} \quad \text{for } \ell = 2, \ldots, m \tag{D.1}$$

$$\mathcal{P}_{\text{TIGHT},2} = \{\{1\}, \{2\}, \ldots, \{m\}\} \tag{D.2}$$

$$\mathcal{P}_{\text{TIGHT},3} = \{\{1, 3, \ldots, m\}, \{2, 3, \ldots, m\}, \{1, 2, 3, \ldots, m\}\} \tag{D.3}$$

The proof for $\mathcal{P}^{(\ell)}_{\text{TIGHT},1}, \mathcal{P}_{\text{TIGHT},2}$ are given in [71, Theorem 10]. In the case where $m = 2$, demands $\mathcal{P}_{\text{TIGHT},3}$ corresponds to the two-user broadcast network with a common message. The tightness of cut-set for this case has been shown. For $m > 2$, the additional users desire to recover all the messages. The proof $\mathcal{P}_{\text{TIGHT},3}$ when $m > 2$ follows along the same lines as that $\mathcal{P}^{(\ell)}_{\text{TIGHT},1}$.

We now show that cut-set is not universally tight for all other communication demands outside of those given in (D.1) - (D.3). We first state Lemma D.1, which provides three communication demands for which cut-set is not universally tight.

Lemma D.1. *The cut-set bound is not universally tight for the communication demands:*

$$\mathcal{P}_{\textit{NOT-TIGHT},1} = \{\{1, 2, 3\}, \{3\}\} \tag{D.4}$$

$$\mathcal{P}_{\textit{NOT-TIGHT},2} = \{\{1, 2\}, \{3\}\} \tag{D.5}$$

$$\mathcal{P}_{\textit{NOT-TIGHT},3} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\} \tag{D.6}$$

Proof. We provide broadcast channel examples for each communication demand and show that cut-set is not achievable in each example.

Consider the broadcast channel with source S and destinations D_1, D_2, D_3 and channel

matrices:

$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{D.7}$$
$$\mathbf{H}_{S,D_2} = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{D.8}$$
$$\mathbf{H}_{S,D_3} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{D.9}$$

We show that under communication demand $\mathcal{P}_{\text{NOT-TIGHT},1}$, the rate $R_1 = 1, R_2 = 1$ is not achievable. We use a slightly modified version of the proof in [80]. We have the following set of inequalities:

$$\begin{aligned} nR_1 \log p &= H(W_1) && \text{(D.10)} \\ &= H(W_1) - H(W_1|Y_1^n) + H(W_1|Y_1^n) && \text{(D.11)} \\ &= I(W_1; Y_1^n) + H(W_1|Y_1^n) && \text{(D.12)} \\ &\overset{(a)}{=} I(W_1; Y_1^n) + n\epsilon && \text{(D.13)} \\ &\leq H(Y_1^n) - H(Y_1^n|W_1) + n\epsilon && \text{(D.14)} \\ &\leq n \log p - H(Y_1^n|W_1) + n\epsilon && \text{(D.15)} \end{aligned}$$

where (a) follows by Fano's inequality. Along the same lines as the above, it can be shown that

$$nR_1 \log p \leq n \log p - H(Y_2^n|W_1) + n\epsilon \tag{D.16}$$

We have the following set of inequalities

$$\begin{aligned} nR_2 \log p &= H(W_2) && \text{(D.17)} \\ &= H(W_2|W_1) && \text{(D.18)} \\ &= H(W_2|W_1) - H(W_2|Y_3^n, W_1) + H(W_2|Y_3^n, W_1) && \text{(D.19)} \\ &= I(W_2; Y_3^n|W_1) + H(W_2|Y_3^n, W_1) && \text{(D.20)} \\ &\overset{(a)}{=} I(W_2; Y_3^n|W_1) + n\epsilon && \text{(D.21)} \\ &= I(W_2; Y_1^n, Y_2^n, Y_3^n|W_1) + n\epsilon && \text{(D.22)} \\ &= H(Y_1^n, Y_2^n, Y_3^n|W_1) + H(Y_1^n, Y_2^n, Y_3^n|W_1, W_2) + n\epsilon && \text{(D.23)} \\ &\overset{(b)}{=} H(Y_1^n, Y_2^n, Y_3^n|W_1) + n\epsilon && \text{(D.24)} \\ &= H(Y_1^n, Y_2^n|W_1) + H(Y_3^n|Y_1^n, Y_2^n, W_1) + n\epsilon && \text{(D.25)} \\ &\overset{(c)}{=} H(Y_1^n, Y_2^n|W_1) + n\epsilon && \text{(D.26)} \\ &\leq H(Y_1^n|W_1) + H(Y_2^n|W_2) + n\epsilon && \text{(D.27)} \end{aligned}$$

where (a) follows by Fano's inequality, (b) follows by since the channel is deterministic, (c) follows since Y_3^n is a deterministic function of Y_1^n, Y_2^n. Combining (D.15), (D.16), and (D.27), it follows that

$$nR_2 \leq n - nR_1 + n - nR_1 + n\epsilon' \tag{D.28}$$

Hence, we have the condition $2R_1 + R_2 \leq 2$ and the rate $R_1 = 1, R_2 = 1$ is not achievable.

We consider the broadcast channel with source S and destinations D_1, D_2, D_3 and channel matrices given by

$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{D.29}$$
$$\mathbf{H}_{S,D_2} = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{D.30}$$
$$\mathbf{H}_{S,D_3} = \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{D.31}$$

Along the same lines as that for the previous example, it can be shown that under communication demand $\mathcal{P}_{\text{NOT-TIGHT},2}$, the rates $R_1 = 1, R_2 = 1$ are not achievable.

Consider the broadcast channel with 3 destinations, no relays, and channel matrices:

$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{D.32}$$
$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{D.33}$$
$$\mathbf{H}_{S,D_1} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{D.34}$$

We show that the cut-set bound is not tight under communication demand $\mathcal{P}_{\text{NOT-TIGHT},3}$. It can be easily shown that rates $R_1 = 1, R_2 = 1, R_3 = 1$ is not achievable. We have the following set of inequalities:

$$\begin{aligned} n(R_1 + R_2)\log p &= H(W_1, W_2) && \text{(D.35)} \\ &= I(W_1, W_2; Y_1^n) + H(W_1, W_2|Y_1^n) && \text{(D.36)} \\ &\overset{(a)}{\leq} I(W_1, W_2; Y_1^n) + n\epsilon_n && \text{(D.37)} \\ &\leq H(Y_1^n) - I(Y_1^n|W_1, W_2) + n\epsilon_n && \text{(D.38)} \\ &\leq 2n\log p - H(Y_1^n|W_1, W_2) + n\epsilon_n && \text{(D.39)} \end{aligned}$$

where (a) follows by Fano's inequality. Similarly, it can be shown that

$$n(R_2 + R_3)\log p \leq 2n\log p - H(Y_2^n|W_2, W_3) + n\epsilon_n \tag{D.40}$$

By applying Fano's inequality, it can be shown that

$$\begin{aligned}
nR_3 &= H(W_3) && \text{(D.41)}\\
&= H(W_3|W_2, W_1) && \text{(D.42)}\\
&= I(W_3; Y_3^n|W_2, W_1) + H(W_3|Y_3^n, W_2, W_1) && \text{(D.43)}\\
&\leq (W_3; Y_3^n|W_2, W_1) + n\epsilon && \text{(D.44)}\\
&\leq I(W_3; Y_1^n, Y_2^n, Y_3^n|W_1, W_2) && \text{(D.45)}\\
&= H(Y_1^n, Y_2^n, Y_3^n|W_1, W_2) - H(Y_1^n, Y_2^n, Y_3^n|W_1, W_2, W_3) && \text{(D.46)}\\
&= H(Y_1^n, Y_2^n, Y_3^n|W_1, W_2) && \text{(D.47)}\\
&= H(Y_1^n, Y_2^n|W_1, W_2) + H(Y_3^n|W_1, W_2, Y_1^n, Y_2^n) && \text{(D.48)}\\
&= H(Y_1^n, Y_2^n|W_1, W_2) && \text{(D.49)}\\
&\leq H(Y_1^n|W_1, W_2) + H(Y_2^n|W_1, W_2) && \text{(D.50)}\\
&\leq 2n - n(R_1 + R_2) + 2n - n(R_2 + R_3) && \text{(D.51)}
\end{aligned}$$

where the last step follows by (D.35), (D.40). Hence, we have the resulting condition:

$$R_1 + 2R_2 + 2R_3 \leq 4 \tag{D.52}$$

and note that $(1, 1, 1)$ is not achievable. □

We now show that all other communication demands outside of $\mathcal{P}^{(\ell)}_{\text{TIGHT},1}, \mathcal{P}_{\text{TIGHT},2}, \mathcal{P}_{\text{TIGHT},3}$ contain one of the demands Lemma D.1 as a sub-demand. This can easily shown by exhaustive search in the case where $m = 3$. We consider the number of users $m > 4$.

For a given $m \in \mathbb{N}$, we let $\mathcal{M} = \{1, \ldots, m\}$ and $\mathcal{T}(\mathcal{M})$ denote it's power set minus the null element $\emptyset$. We define the following sets:

$$\Theta_m = \mathcal{T}\left(\mathcal{T}(\mathcal{M})\right) \tag{D.53}$$

$$\Phi_m = \{\mathcal{P} \in \Theta_m : \mathcal{P}_{\text{NOT-TIGHT},i} \text{ for any } i = 1, 2, 3 \text{ is equiv. to or is a sub-demand of } \mathcal{P}\} \tag{D.54}$$

$$\Psi_m = \left\{\mathcal{P} \in \Theta_m : \mathcal{P} \text{ is equiv. to or is a sub-demand of } \mathcal{P}^{(\ell)}_{\text{TIGHT},1}, \mathcal{P}_{\text{TIGHT},2}, \mathcal{P}_{\text{TIGHT},3}\right\} \tag{D.55}$$

We observe that Θ_m is the set of all communication demands for m users. The goal is show that Θ_m can be partitioned into Φ_m and Ψ_m. It is sufficient to show that $\Phi_m \cup \Psi_m = \Theta_m$ and $\Phi_m \cap \Psi_m = \emptyset$. It can be easily shown that $\Phi_m \cap \Psi_m = \emptyset$, and we focus on showing that $\Phi_m^c \subseteq \Psi_m$. We define the sets Υ_ℓ to be a subset of Θ_m where the number of messages is ℓ:

$$\Upsilon_\ell = \{\mathcal{P} \in \Theta_m : |\mathcal{P}| = \ell\} \tag{D.56}$$

There are $2^m - 1$ elements in Υ_1. It can be easily seen that each element in Υ_1 belongs to Ψ_m. We consider the set Υ_2. We search exhaustively and eliminate the set of demands that

contain $\mathcal{P}_{\text{NON-TIGHT},1}, \mathcal{P}_{\text{NON-TIGHT},2}$ as a sub-demand. We find that the following non-equivalent demands are in $\Upsilon_2 \cap \Phi_m^c$:

$$\mathcal{P}_{2,1,j} = \{\{1\} \cup \{3,\ldots,j\}, \{2\} \cup \{3,\ldots,j\}\} \quad \text{for } j = 2,\ldots,m \tag{D.57}$$
$$\mathcal{P}_{2,2,j} = \{\{1,2\} \cup \{3,\ldots,j\}, \{2\} \cup \{3,\ldots,j\}\} \quad \text{for } j = 2,\ldots,m \tag{D.58}$$

We consider Υ_3. To find the demands in this set that is in Φ_m^c, we need only consider $\mathcal{P} \in \Upsilon_3$ such that $\mathcal{P}_{2,1,j}$ or $\mathcal{P}_{2,2,j}$ is a sub-set of $\mathcal{P}$. The resulting non-equivalent message demands in $\Upsilon_3 \cap \Phi_m^c$ are

$$\mathcal{P}_{3,1,j} = \{\{1\} \cup \{3,\ldots,j\}, \{2\} \cup \{3,\ldots,j\}, \{1,2\} \cup \{3,\ldots,j\}\} \quad \text{for } j = 3,\ldots,m \tag{D.59}$$
$$\mathcal{P}_{3,2,j} = \{\{1\} \cup \{4,\ldots,j\}, \{2\} \cup \{4,\ldots,j\}, \{3\} \cup \{4,\ldots,j\}\} \quad \text{for } j = 3,\ldots,m \tag{D.60}$$
$$\mathcal{P}_{3,3,j} = \{\{1,3\} \cup \{4,\ldots,j\}, \{2,3\} \cup \{4,\ldots,j\}, \{3\} \cup \{4,\ldots,j\}\} \quad \text{for } j = 4,\ldots,m \tag{D.61}$$

We consider Υ_4. To find the demands in this set that is in Φ_m^c, we need only consider $\mathcal{P} \in \Upsilon_4$ such that $\mathcal{P}_{3,1,j}, \mathcal{P}_{3,2,j}$ or $\mathcal{P}_{3,3,j}$ is a sub-set of $\mathcal{P}$. The resulting non-equivalent message demands in $\Upsilon_4 \cap \Phi_m^c$ are given by:

$$\mathcal{P}_{4,1,j} = \{\{1\} \cup \{5,\ldots,j\}, \ldots, \{4\} \cup \{5,\ldots,j\}\} \quad \text{for } j = 4,\ldots,k \tag{D.62}$$
$$\mathcal{P}_{4,2,j} = \{\{1,4\} \cup \{5,\ldots,j\}, \ldots, \{3,4\} \cup \{5,\ldots,j\}, \{4\} \cup \{5,\ldots,j\}\} \text{ for } j = 4,\ldots,k \tag{D.63}$$

Using induction on ℓ for $\ell > 4$, it can be shown that the resulting configurations in $\Upsilon_\ell \cap \Phi_m^c$ are given by

$$\mathcal{P}_{\ell,1,j} = \{\{1\} \cup \{\ell+1,\ldots,j\}, \ldots, \{\ell\} \cup \{\ell+1,\ldots,j\}\} \quad \text{for } j = \ell,\ldots,k \tag{D.64}$$
$$\mathcal{P}_{\ell,2,j} = \{\{1,\ell\} \cup \{\ell+1,\ldots,j\}, \ldots, \{\ell-1,\ell\} \cup \{\ell+1,\ldots,j\}, \{\ell\} \cup \{\ell+1,\ldots,j\}\} \tag{D.65}$$
$$\text{for } j = \ell,\ldots,m \tag{D.66}$$

We observe that $\Upsilon_\ell \cap \Phi_m^c \subseteq \Psi_m$ for each ℓ. Since $\cup_\ell \Upsilon_\ell = \Theta_m$ and $\Upsilon_\ell \cap \Upsilon_{\ell'} = \emptyset$ for all $\ell \neq \ell'$, it can be concluded that $\Phi_m^c \subseteq \Psi_m$.

Appendix E

Proof of Theorem 5.35

For a fixed cut $\Gamma \subseteq \Omega$ of the network $\mathcal{N}_{\text{GAUSS-MAC}}$, we let $U_\Gamma = \{U_i \ \forall \ i \in \Gamma\}$, $X_\Gamma = \{X_i \ \forall \ i \in \Gamma\}$, and $Y_\Gamma = \{Y_i \ \forall \ i \in \Gamma\}$.

Definition E.1 (Rate Distortion Function)**.** For a given cut $\Gamma \subseteq \Omega$, the rate distortion function $R_\Gamma(\cdot)$ is defined as follows:

$$R_\Gamma(D) = \min_{p(\hat{V}|U_\Omega):\mathbb{E}\left[(V-\hat{V})^2|U_\Gamma\right]\leq D} I(U_\Gamma; \hat{V}|U_{\Gamma^c}) \tag{E.1}$$

Remark E.2. $R_\Gamma(\cdot)$ is non increasing and convex (see Lemma 10.4.1 in [1]).

Theorem E.3. *Consider transmitting the sum of m Gaussian sources with variance σ^2 across $\mathcal{N}_{GAUSS\text{-}MAC}$. If there exists source encoders $\{\mathcal{E}_{i,t}\}_{t=1}^n \forall i \in \Omega$ satisfying power constraint* SNR *and a decoder $\mathcal{G}$ that achieves distortion D, then the following must be satisfied:*

$$R_\Gamma(D) < C_\Gamma^{GAUSS\text{-}MAC} \quad \textit{for all} \quad \Gamma \subseteq \Omega \tag{E.2}$$

where $R_\Gamma(\cdot)$ is given in Definition E.1 and $C_\Gamma^{GAUSS\text{-}MAC}$ is given in Definition 5.28.

Proof. For a given cut Γ, we form two super nodes: $N_\Gamma = \{N_i\}_{i\in\Gamma}$, and $N_{\Gamma^c} = \{N_i\}_{i\in\Gamma^c}$ and assume N_Γ knows the information $\{U_{\Gamma,j}\}_1^n = \{U_{i,j} \ \forall \ i \in \Gamma\}_1^n$ and N_{Γ^c} knows the information $\{U_{\Gamma^c,j}\}_1^n = \{U_{i,j} \ \forall \ i \in \Gamma^c\}_{j=1}^n$. The encoders for node N_Γ and N_{Γ^c} are given by $\{\mathcal{E}_{\Gamma,t}\}_{t=1}^n$ and $\{\mathcal{E}_{\Gamma^c,t}\}_{t=1}^n$ where

$$X_{\Gamma,t} = \mathcal{E}_{\Gamma,t}(\{U_{\Gamma,j}\}_1^k, \{Y_{\Gamma,j}\}_1^{t-1}) \ \text{ for } \ t = 1, \ldots, n \tag{E.3}$$

$$X_{\Gamma^c,t} = \mathcal{E}_{\Gamma,t}(\{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) \ \text{ for } \ t = 1, \ldots, n \tag{E.4}$$

The decoder for N_{Γ^c} is given by $\mathcal{G}$ and produces an estimate $\{\hat{V}_j\}_1^k = \mathcal{G}(\{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^n)$

for the sum $\{V_j\}_1^k$ where $V_j = \sum_{i=1}^m U_{i,j}$. We bound the mutual information as follows:

$$I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k | \{U_{\Gamma^c,j}\}_1^k) \leq I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k | \{U_{\Gamma^c,j}\}_1^k) \tag{E.5}$$

$$+ I(\{U_{\Gamma,j}\}_1^k ; \{Y_{\Gamma^c,j}\}_1^n | \{\hat{V}_j\}_1^k, \{U_{\Gamma^c,j}\}_1^k) \tag{E.6}$$

$$\leq I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k, \{Y_{\Gamma^c}\}_1^n | \{U_{\Gamma^c,j}\}_1^k) \tag{E.7}$$

$$\leq I(\{U_{\Gamma,j}\}_1^k ; \{Y_{\Gamma^c,j}\}_1^n | \{U_{\Gamma^c,j}\}_1^k) \tag{E.8}$$

$$+ I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k | \{Y_{\Gamma^c}\}_1^n, \{U_{\Gamma^c,j}\}_1^k) \tag{E.9}$$

We have the Markov chain: $\{U_{\Gamma,j}\}_1^k \to (\{Y_{\Gamma^c}\}_1^n, \{U_{\Gamma^c}\}_1^n) \to \{\hat{V}_j\}_1^k$ since $\{\hat{V}_j\}_1^k = \mathcal{G}(\{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^n)$. Hence, $I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k | \{Y_{\Gamma^c,j}\}_1^n, \{U_{\Gamma^c,j}\}_1^k) = 0$ and (E.9) becomes

$$I(\{U_{\Gamma,j}\}_1^k ; \{\hat{V}_j\}_1^k | \{U_{\Gamma^c,j}\}_1^k) \leq I(\{U_{\Gamma,j}\}_1^k ; \{Y_{\Gamma^c,j}\}_1^n | \{U_{\Gamma^c,j}\}_1^k) \tag{E.10}$$

$$= \sum_{t=1}^n I(\{U_{\Gamma,t}\}_1^k ; Y_{\Gamma^c,t} | \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) \tag{E.11}$$

$$= \sum_{t=1}^n H(Y_{\Gamma^c,t} | \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) \tag{E.12}$$

$$- H(Y_{\Gamma^c,t} | \{U_{\Gamma,j}\}_1^k, \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) \tag{E.13}$$

Using the fact that $X_{\Gamma^c,t}$ is a deterministic function of $\{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}$, we have that

$$H(Y_{\Gamma^c,t} | \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) = H(Y_{\Gamma^c,t} | \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}, X_{\Gamma^c,t}) \tag{E.14}$$

$$\leq H(Y_{\Gamma^c,t} | X_{\Gamma^c,t}) \tag{E.15}$$

Using the fact that conditioning reduces entropy, it follows that

$$H(Y_{\Gamma^c,t} | \{U_{\Gamma,j}\}_1^k, \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}) \geq H(Y_{\Gamma^c,t} | \{U_{\Gamma,j}\}_1^k, \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}, X_{\Gamma^c,t}, X_{\Gamma,t}) \tag{E.16}$$

Using the fact that $Y_{\Gamma^c,t}$ depends only on the current symbol $X_{\Gamma^c,t}, X_{\Gamma,t}$, (E.16) becomes

$$H(Y_{\Gamma^c,t} | \{U_{\Gamma,j}\}_1^k, \{U_{\Gamma^c,j}\}_1^k, \{Y_{\Gamma^c,j}\}_1^{t-1}, X_{\Gamma^c,t}, X_{\Gamma,t}) = H(Y_{\Gamma^c,t} | X_{\Gamma^c,t}, X_{\Gamma,t}) \tag{E.17}$$

Combining (E.13), (E.15), (E.16), (E.17), we have that

$$I(\{U_{\Gamma,j}\}_1^k;\{\hat{V}_j\}_1^k|\,\{U_{\Gamma^c,j}\}_1^k) \le \sum_{t=1}^n H(Y_{\Gamma^c,t}|X_{\Gamma^c,t}) - H(Y_{\Gamma^c,t}|X_{\Gamma,t},X_{\Gamma^c,t}) \tag{E.18}$$

$$= \sum_{t=1}^n I(X_{\Gamma,t};Y_{\Gamma^c,t}|X_{\Gamma^c,t}) \tag{E.19}$$

After introducing a time-sharing variable Q distributed uniformly on $\{1,\dots,n\}$, it can be shown that (E.19) becomes

$$I(\{U_{\Gamma,j}\}_1^k;\{\hat{V}_j\}_1^k|\,\{U_{\Gamma^c,j}\}_1^k) \le \sum_{t=1}^n I(X_{\Gamma,t};Y_{\Gamma^c,t}|X_{\Gamma^c,t}) \tag{E.20}$$

$$\le nI(X_{\Gamma,Q};Y_{\Gamma^c,Q}|X_{\Gamma^c,Q}) \tag{E.21}$$

$$= n\frac{1}{n}\sum_{t=1}^n I(X_{\Gamma,Q};Y_{\Gamma^c,Q}|X_{\Gamma^c,Q},Q=t) \tag{E.22}$$

$$= nI(X_{\Gamma,Q};Y_{\Gamma^c,Q}|X_{\Gamma^c,Q},Q) \tag{E.23}$$

$$= nH(Y_{\Gamma^c,Q}|X_{\Gamma^c,Q},Q) - nH(Y_{\Gamma^c,Q}|X_{\Gamma,Q},X_{\Gamma^c,Q},Q) \tag{E.24}$$

$$\le nH(Y_{\Gamma^c,Q}|X_{\Gamma^c,Q}) - nH(Y_{\Gamma^c,Q}|X_{\Gamma,Q},X_{\Gamma^c,Q},Q) \tag{E.25}$$

$$= nH(Y_{\Gamma^c,Q}|X_{\Gamma^c,Q}) - nH(Y_{\Gamma^c,Q}|X_{\Gamma,Q},X_{\Gamma^c,Q}) \tag{E.26}$$

Since $U_{i,1},\dots U_{i,k}$ is an i.i.d sequence, the left hand side of (E.26) can be shown to be

$$I(\{U_{\Gamma,j}\}_1^k;\{\hat{V}_j\}_1^k|\,\{U_{\Gamma^c,j}\}_1^k) = H(\{U_{\Gamma,j}\}_1^k\,|\,\{U_{\Gamma^c,j}\}_1^k) - H(\{U_{\Gamma,j}\}_1^k\,|\,\{U_{\Gamma^c,j}\}_1^k,\{\hat{V}_j\}_1^k) \tag{E.27}$$

$$= \sum_{t=1}^k H(U_{\Gamma,t}|\,\{U_{\Gamma,j}\}_1^{t-1},U_{\mathcal{S}^c}^{1:k}) \tag{E.28}$$

$$- H(U_{\Gamma,t}|\,\{U_{\Gamma,j}\}_1^{t-1},\{U_{\Gamma^c,j}\}_1^k,\{\hat{V}_j\}_1^k) \tag{E.29}$$

$$= \sum_{t=1}^k H(U_{\Gamma,t}|U_{\Gamma^c,t}) - H(U_{\Gamma,t}|\,\{U_{\Gamma,j}\}_1^{t-1},\{U_{\Gamma^c,j}\}_1^k,\{\hat{V}_j\}_1^k) \tag{E.30}$$

$$\ge \sum_{t=1}^n H(U_{\Gamma,t}|U_{\Gamma^c,t}) - H(U_{\Gamma,t}|U_{\Gamma,t},U_{\Gamma^c,t},\hat{V}_t) \tag{E.31}$$

$$\ge \sum_{t=1}^k I(U_{\Gamma,t};\hat{V}_t|U_{\Gamma^c,t}) \tag{E.32}$$

Using rate distortion function in definition E.1, the fact that it is non-increasing and convex,

it follows that

$$\sum_{t=1}^{k} I(U_{\Gamma,t}; \hat{V}_t | U_{\Gamma^c,t}) \geq \sum_{t=1}^{k} R_\Gamma(E[(V_t - \hat{V}_t)^2 | U_{\Gamma^c,t}]) \tag{E.33}$$

$$\geq kR_\Gamma\left(\frac{1}{k}\sum_{t=1}^{k} E[(V_t - \hat{V}_t)^2 | U_{\Gamma^c,t}]\right) \tag{E.34}$$

$$\geq kR_\Gamma(D) \tag{E.35}$$

The result follows by combining (E.26), (E.32), (E.35). □

We now evaluate the expression in rate distortion function in definition (E.1). First, we show that we can relax the constraint set to all the set of all $\hat{V}$ that are jointly Gaussian with U_Ω. Let $\hat{V}$ be a random variable and let $\hat{V}_G$ be a Gaussian random variance with the same covariance structure as $\hat{V}$. Define the vector $[\alpha_i, \beta_{i,1}, \cdots \beta_{i,|\Gamma^c|}]$ such that

$$[\alpha_i, \beta_{i,1}, \ldots, \beta_{i,|\Gamma^c|}] = LLSE[U_i | \hat{V}_G, U_{\Gamma^c}] \tag{E.36}$$

for all $i \in \Gamma$. It follows that

$$I(U_\Gamma; W_G | U_{\Gamma^c}) = h(U_\Gamma | U_{\Gamma^c}) - h(U_\Gamma | W_G, U_{\Gamma^c}) \tag{E.37}$$

$$= h(U_\Gamma | U_{\Gamma^c}) - h\left(U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma^c | U_{\Gamma^c}, \hat{V}_G\right) \tag{E.38}$$

$$\overset{(a)}{=} h(U_\Gamma | U_{\Gamma^c}) - h\left(U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma^c\right) \tag{E.39}$$

$$\overset{(b)}{\leq} h(U_\Gamma | U_{\Gamma^c}) - h\left(U_i - \alpha_i \hat{V} - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma^c\right) \tag{E.40}$$

$$\leq h(U_\Gamma | U_{\Gamma^c}) - h\left(U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma^c | U_{\Gamma^c}, \hat{V}\right) \tag{E.41}$$

$$= h(U_\Gamma | U_{\Gamma^c}) - h(U_\Gamma | U_{\Gamma^c}, \hat{V}) \tag{E.42}$$

$$= I(U_\Gamma; \hat{V} | U_{\Gamma^c}) \tag{E.43}$$

We first justify (a). Since $U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j$ is independent of $\hat{V}_G, U_{\Gamma^c}$ for all $i \in \Gamma$ and they are jointly Gaussian, we have that any linear combination of $U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma$ is independent of $\hat{V}_G, U_{\Gamma^c}$. Hence, $U_i - \alpha_i \hat{V}_G - \sum_{j=1}^{|\Gamma^c|} \beta_{i,j} U_j \ \forall\ i \in \Gamma$ is independent of $\hat{V}_G$. The

inequality in (b) follows since $\hat{V}$ and $\hat{V}_G$ have the same covariance structure and Gaussian maximizes entropy.

We rewrite $\hat{V} = \sum_{i\in\Gamma} \alpha_i U_i + \sum_{i\in\Gamma^c} \alpha_i U_i + \gamma Z$ where $Z \sim \mathcal{N}(0,1)$ is independent of U_Ω. The rate distortion function can be rewritten as:

$$R_\Gamma(D) = \min_{\mathbb{E}\left[(V-\sum_{i\in\Gamma}\alpha_i U_i+\sum_{i\in\Gamma^c}\alpha_i U_i+\gamma Z)^2|U_\Gamma\right]\le D} I\left(U_\Gamma; \sum_{i\in\Gamma}\alpha_i U_i + \sum_{i\in\Gamma^c}\alpha_i U_i + \gamma Z | U_{\Gamma^c}\right) \tag{E.44}$$

Using the independent of $U_1,\dots,U_m,Z$, the mutual information in (E.44) is given by

$$I(U_\Gamma; \sum_{i\in\Gamma}\alpha_i U_i + \sum_{i\in\Gamma^c}\alpha_i U_i + \gamma Z|U_{\Gamma^c}) = h(\sum_{i\in\Gamma}\alpha_i U_i + \sum_{i\in\Gamma^c}\alpha_i U_i + \gamma Z|U_{\Gamma^c}) \tag{E.45}$$

$$- h(\sum_{i\in\Gamma}\alpha_i U_i + \sum_{i\in\Gamma^c}\alpha_i U_i + \gamma Z|U_\Omega) \tag{E.46}$$

$$= h(\sum_{i\in\Gamma}\alpha_i U_i + \gamma Z|U_{\Gamma^c}) - h(\gamma Z|U_\Omega) \tag{E.47}$$

$$= h(\sum_{i\in\Gamma}\alpha_i U_i + \gamma Z) - h(\gamma Z) \tag{E.48}$$

$$= \frac{1}{2}\log\left(\frac{\sum_{i\in\Gamma}\alpha_i^2\sigma^2+\gamma^2}{\gamma^2}\right). \tag{E.49}$$

Recalling that $V = \sum_{i\in\Omega} U_i$, the distortion constraint is given by

$$\mathbb{E}[(V - \sum_{i\in\Gamma}\alpha_i U_i + \sum_{i\in\Gamma^c}\alpha_i U_i + \gamma Z)^2|U_\Gamma] = \mathbb{E}[(\sum_{i\in\Gamma}(1-\alpha_i)U_i + \sum_{i\in\Gamma^c}(1-\alpha_i)U_i + \gamma Z)^2|U_\Gamma] \tag{E.50}$$

$$= \mathbb{E}[(\sum_{i\in\Gamma}(1-\alpha_i)U_i + \sum_{i\in\Gamma^c}(1-\alpha_i)U_i + \gamma Z)^2|U_\Gamma] \tag{E.51}$$

$$= (\sum_{i\in\Gamma}(1-\alpha_i)U_i)^2 + \mathbb{E}[(\sum_{i\in\Gamma^c}(1-\alpha_i)U_i + \gamma Z)^2|U_\Gamma] \tag{E.52}$$

$$= (\sum_{i\in\Gamma}(1-\alpha_i)U_i)^2 + \sum_{i\in\Gamma^c}(1-\alpha_i)^2\sigma^2 + \gamma^2 \tag{E.53}$$

The rate distortion function can be rewritten as

$$\min_{\alpha_i,\gamma:\left(\sum_{i\in\Gamma}(1-\alpha_i)U_i\right)^2+\sum_{i\in\Gamma^c}(1-\alpha_i)^2\sigma^2+\gamma^2\le D} \frac{1}{2}\log\left(\frac{\sum_{i\in\Gamma}\alpha_i^2\sigma^2+\gamma^2}{\gamma^2}\right) \tag{E.54}$$

It can be easily shown that the maximizing parameters are given as follows:

$$\alpha_i = 0 \quad \text{for} \quad i \in \Gamma^c, \quad \alpha_i = 1 - \frac{D}{|\mathcal{S} \cap \Gamma|\sigma^2} \quad \text{for} \quad i \in \Gamma^c, \quad \gamma^2 = D\left(1 - \frac{D}{|\mathcal{S} \cap \Gamma|\sigma^2}\right) \tag{E.55}$$

With these parameter values, it follows that

$$R_\Gamma(D) = \frac{1}{2}\log\left(\frac{|\mathcal{S} \cap \Gamma|\sigma^2}{D}\right) \tag{E.56}$$

The result follows by considering only the cuts where $|\mathcal{S} \cap \Gamma| = 1$.

Appendix F

Proof for Theorem 6.5

We first consider the $N = 2$ case. Fix an input distribution $P(x_1, x_2) = p_1(x_1)p_2(x_2)$. General M_1 independent codewords $X_1^n(i)$ for $i = \{1, ..., M_1\}$ of length n, where each element is chosen i.i.d from $\sim \prod_{i=1}^n p_1(x_i)$. Similarly, generate M_2 independent codewords of length n, where each element is chosen i.i.d from $\prod_{i=1}^n p_2(x_i)$. User 1 chooses his message w_1 uniformly from $\{1 \cdots M_1\}$ and user 2 chooses his message w_2 uniformly from $\{1 \cdots M_2\}$. User 1 encodes its message into codeword $X_1^n(w_1)$ and broadcasts it to the relays. Similarly, user 2 encodes his message into codeword $X^n(w_2)$ and broadcasts it to the relays. Relay m observes Y_m^n and has knowledge of its local fading information h_m^n.

Let P_{IND} be the induced distribution on the joint sequence $\{Y_1, \mathbf{h}_1 \cdots Y_M, \mathbf{h}_M\}^n$ from definition 4. Fix a forwarding function $U = f_{F,P_{\text{IND}}}(Y_1, \mathbf{h}_1, \cdots, Y_M, \mathbf{h}_M)$. Let κ_U be the achievable computation rate (for the MAC) for U. The relays use a computation code to jointly forward k instances of the forwarding function $U^k = f((Y_1^k, h_1^k), ..., (Y_M^k, h_M^k))$ to the receiver over ℓn uses of the MAC. The computation code outputs an estimate $\hat{U}^k$ for U^k. We define the event T as follows:

$$T = \left\{U^k \neq \hat{U}^k\right\} \tag{F.1}$$

From Definition 4, U^k can be recovered losslessly at the destination if

$$\frac{k}{n} \leq \min(\kappa_U \ell, 1) \tag{F.2}$$

In the rest of our proof, we fix $\frac{k}{n} = \min(\kappa_U \ell, 1)$ and scale k, n.

We fix an $\epsilon_0 > 0$. There exists a k_0 such that for all $k \geq k_0$, there exists a $(k, \ell n, \epsilon)$ reliable computation code that outputs an estimate $\hat{U}^k$ such that

$$P(T) \leq \epsilon_0 \tag{F.3}$$

The destination recovers the messages w_1, w_2 based on the estimate $\hat{U}^k$ using jointly typical decoding. Fix $\delta > 0$. Let $\mathcal{A}_\delta^{(k)}$ be the set of sequences $\left\{(x_1^k, x_2^k, u^k)\right\}$ that are jointly

typical with respect to $p(x_1, x_2, u)$. We define the events $E_{i,j}, \hat{E}_{i,j}$ for $i \in \{1 \cdots M_1\}, j \in \{1 \cdots M_2\}$ as follows:

$$E_{i,j} = \left\{(X_1^k(i), X_2^k(j), U^k) \in \mathcal{A}_\delta^{(k)}\right\} \tag{F.4}$$

$$\hat{E}_{i,j} = \left\{(X_1^k(i), X_2^k(j), \hat{U}^k) \in \mathcal{A}_\delta^{(k)}\right\} \tag{F.5}$$

We can bound the probability of $\hat{E}_{i,j}$ in terms of the probability of $E_{i,j}$ and T as follows:

$$P(\hat{E}_{i,j}) = P(\hat{E}_{i,j} \cap T^c) + P(\hat{E}_{i,j} \cap T) \tag{F.6}$$

$$\leq P(\hat{E}_{i,j} \cap T^c) + P(T) \tag{F.7}$$

$$\overset{(a)}{=} P(E_1 \cap T^c) + P(T) \tag{F.8}$$

$$\leq P(E_{i,j}) + P(T) \tag{F.9}$$

where (a) follows from the fact under event T, $E_{i,j} = \hat{E}_{i,j}$. Since we are considering average probability of error and our codebook is constructed in a symmetric manner, we can assume that message $w_1 = 1, w_2 = 1$ were transmitted. The average error probability can be bounded as follows:

$$P_{\text{ERROR}} = P(\hat{E}_{1,1}^c) + P(\cup_{(i,j)\neq(1,1)} \hat{E}_{i,j}) \tag{F.10}$$

$$\leq P(E_{1,1}^c) + P(T) + P(\cup_{(i,j)\neq(1,1)} E_{i,j}) + P(T) \tag{F.11}$$

$$\leq P(E_{1,1}^c) + \sum_{(i,j)\neq(1,1)} P(E_{i,j}) + 2P(T) \tag{F.12}$$

$$\leq P(E_{1,1}^c) + \sum_{j\neq 1} P(E_{1,j}) + \sum_{i\neq 1} P(E_{i,1}) \tag{F.13}$$

$$+ \sum_{i\neq 1, j\neq 1} P(E_{i,j}) + 2P(T) \tag{F.14}$$

From the joint AEP, there exists a k_1 such that for all $k \geq k_1$, $P(E_{1,1}) \leq \epsilon_1$. From the proof of the channel coding Theorem (8.7.1 in [1]), we have the following bounds:

$$P(E_{i,1}) \leq 2^{-kI(X_1;Y|X_2)-3\delta} \quad \text{for} \quad i = 2 \cdots M_1 \tag{F.15}$$

$$P(E_{1,i}) \leq 2^{-kI(X_2;Y|X_1)-3\delta} \quad \text{for} \quad j = 2 \cdots M_1 \tag{F.16}$$

$$P(E_{i,j}) \leq 2^{-kI(X_1,X_2;Y)-4\delta} \quad \text{for} \quad i \neq 1, j \neq 1 \tag{F.17}$$

We fix an $\epsilon_2 > 0$. There exists a k_2 such that for all $k \geq k_2$:

$$\sum_{i \neq 1, j \neq 1} P(E_{i,j}) \leq \epsilon_2 \tag{F.18}$$

$$\sum_{(1,j)} P(E_{1,j}) \leq \epsilon_2 \tag{F.19}$$

$$\sum_{(i,1)} P(E_{i,1}) \leq \epsilon_2 \tag{F.20}$$

if the following conditions are satisfied:

$$\frac{\log M_1}{n} < \frac{k}{n} I(X_1; U|X_2) \tag{F.21}$$

$$\frac{\log M_2}{n} < \frac{k}{n} I(X_1; U|X_2) \tag{F.22}$$

$$\frac{\log M_1}{n} + \frac{\log M_2}{n} < \frac{k}{n} I(X_1, X_2; U) \tag{F.23}$$

We choose $k \geq \max(k_0, k_1, k_2)$ and $n = \frac{k}{\min\{\kappa_U \ell, 1\}}$. Our probability of error becomes:

$$P_{\text{ERROR}} \leq \epsilon_0 + \epsilon_1 + \epsilon_2 \tag{F.24}$$

Finally, we choose $\epsilon_0, \epsilon_1, \epsilon_2$ to be arbitrarily small and then δ to be arbitrarily small.

The extension to the general N user case follows using the same techniques.

Appendix G

Proof for Theorem 6.7

Our scheme consists of two sets of codes: an outer code for transmitting the message and an inner code for sending the forwarding function. A random codebook construction is performed at the source and a computation code is used by the relays. We decompose the overall scheme into three stages: outer encoding, computation coding, and outer decoding. In general, we assume that there is a bandwidth expansion of ℓ between the MAC and the broadcast channel. For simplicity, we will first consider the case where $\ell = 4$.

G.0.1 Stage I: Message Encoding

The user selects his message w uniformly from the set $\mathcal{W} = \left\{1, ..., 2^{nR}\right\}$. It constructs an i.i.d random codebook $\mathcal{C}$ where each element of each codeword is drawn according to distribution $\mathcal{N}(0, \mathsf{SNR}_s - \delta)$. The user encodes its message into a length n codeword:

$$\mathcal{E} : \mathcal{W} \to \mathbb{R}^n \tag{G.1}$$

$$X^n(w) = \mathcal{E}(w) \tag{G.2}$$

The user then broadcasts his codeword to the relays. At time i, relay m observes $Y_m[i]$ where

$$Y_m[i] = \sum_{j=1}^{N} h[i]X[i] + Z[i] \tag{G.3}$$

and has knowledge of its local channel coefficients $\mathbf{h}_m[i]$.

G.0.2 Stage II: Forwarding the Sufficient Statistic

The relays desire to communicate the forwarding function to the destination over the MAC. The forwarding function consists of two components:

$$U = \sum_{m=1}^{M} h_m Y_m \tag{G.4}$$

$$V = \sum_{m=1}^{M} h_m^2 \tag{G.5}$$

Given a bandwidth expansion of $\ell = 4$, we allocate $\ell_1 n = 2n$ channel uses to send $\mathbf{u}$ and $\ell_2 n = 2n$ channel uses to send V^n. We first describe the code for forwarding U^n. Our scheme consists of 2 iterations with n channel uses in each iteration. Our computation code relies on the existence of a sequence of lattices which are good for coding as demonstrated in [17].

Lemma G.1. *There exists a sequence of Good Lattices Λ_n with Voronoi regions $\mathcal{V}_{0,n}$ and second moments SNR_r such that given an length n i.i.d Gaussian sequence $\mathbf{z}$ with variance $\sigma_z^2 < \mathsf{SNR}_r$ and an $\epsilon > 0$, $\exists$ n_0 such that for all $n \geq n_0$.*

$$Pr(\mathbf{z} \in \mathcal{V}_{0,n}) \geq 1 - \epsilon \tag{G.6}$$

The above lemma states that an i.i.d Gaussian sequence with second moment strictly less than the second moment of the lattice falls into the Voronoi region with high probability. See [17] for proof.

We first consider the case where $\ell_1 = 2$. Let $\mathbf{s}_m[i] = h_m[i]Y_m[i]$ for $i = 1 \cdots n$ and its variance $\sigma_s^2 = 2\mathsf{SNR}_s + 1$. The first iteration involves only uncoded transmission. Relay m sends

$$\mathbf{x}_m^{(1)} = \sqrt{\mathsf{SNR}_r \sigma_s^2}\mathbf{s}_m \tag{G.7}$$

The destination receives

$$\mathbf{y}^{(1)} = \sum_{m=1}^{M} \mathbf{x}_m^{(1)} + \mathbf{z}^{(1)} \tag{G.8}$$

and forms the linear estimate

$$\hat{\mathbf{u}}^{(1)} = \sqrt{\frac{\sigma_s^2}{\mathsf{SNR}_r}}\mathbf{y}^{(1)} \tag{G.9}$$

Let $\mathbf{q}^{(1)} = \hat{\mathbf{u}}^{(1)} - \mathbf{u}$ be the estimation error from the first iteration. It can be easily seen that

$$\mathbf{q}^{(1)} = \sqrt{\frac{\sigma_s^2}{\mathsf{SNR}_r}}\mathbf{z}^{(2)} \tag{G.10}$$

Let $D^{(1)}$ be the variance of $\mathbf{q}^{(1)}$. It follows that

$$D^{(1)} = \frac{\sigma_s^2}{\mathsf{SNR}_r} N \tag{G.11}$$

For the second iteration, we choose a sequence of Lattices Λ_n according to Lemma G.1. Let $\mathbf{d}_1, ..., \mathbf{d}_M$ be independent dither vectors drawn uniformly from $\mathcal{V}_{0,n}$. User m transmits $\mathbf{x}_m^{(2)} = [\gamma \mathbf{s}_m + \mathbf{d}_m] \bmod \Lambda_n$.

The channel output is given by

$$\mathbf{y}^{(2)} = \sum_{i=1}^{M} \mathbf{x}_m^{(2)} + \mathbf{z}^{(2)} \tag{G.12}$$

The receiver computes

$$\mathbf{r} = \mathbf{y}^{(2)} - \left(\sum_{m=1}^{M} \mathbf{d}_m + \gamma \hat{\mathbf{u}}^{(1)} \right) \tag{G.13}$$

$$\mathbf{t} = \mathbf{r} \bmod \Lambda_k \tag{G.14}$$

$$= \left[\sum_{i=1}^{M} \mathbf{x}_j + \mathbf{z}^{(2)} - \sum_{i=1}^{M} (\mathbf{d}_j + \gamma \mathbf{s}_j) - \gamma \mathbf{q}^{(1)} \right] \bmod \Lambda_k \tag{G.15}$$

$$= \left[\mathbf{z}^{(2)} - \gamma \mathbf{q}^{(1)} \right] \bmod \Lambda_k \tag{G.16}$$

and updates the estimate and estimation noise from the first iteration:

$$\hat{\mathbf{u}}^{(2)} = \beta \mathbf{t} + \hat{\mathbf{u}}^{(1)} \tag{G.17}$$

$$\mathbf{q}^{(2)} = \beta \mathbf{t} + \mathbf{q}^{(1)} \tag{G.18}$$

where $\mathbf{q}^{(2)} = \hat{\mathbf{u}}^{(2)} - \mathbf{u}$. Define the event

$$\mathcal{T}_u = \left\{ [\mathbf{z}^{(2)} - \gamma \mathbf{q}^{(1)}] \bmod \Lambda_n = \mathbf{z}^{(2)} - \gamma \mathbf{q}^{(1)} \right\} \tag{G.19}$$

Under event $\mathcal{T}_u$, the updated estimate and estimation error becomes

$$\hat{\mathbf{u}}^{(2)} = \beta \mathbf{z}^{(2)} + (1 - \beta\gamma) \mathbf{q}^{(1)} + \mathbf{u} \tag{G.20}$$

$$\mathbf{q}^{(2)} = \beta \mathbf{z}^{(2)} + (1 - \beta\gamma) \mathbf{q}^{(1)} \tag{G.21}$$

$$= \beta \mathbf{z}^{(2)} + (1 - \beta\gamma) \sqrt{\frac{D^{(1)}}{N}} \mathbf{z}^{(1)} \tag{G.22}$$

$$\tag{G.23}$$

Let β, γ_o be chosen as follows:

$$\gamma_o = \sqrt{\frac{\mathsf{SNR}_r - 1}{D^{(1)}}} \tag{G.24}$$

$$\beta = \frac{D^{(1)}\gamma}{1 + D^{(1)}\gamma^2} \tag{G.25}$$

For any fixed $\epsilon > 0$, the following variance of the estimation error is achievable under event $\mathcal{T}$:

$$E[(\mathbf{q}^{(2)}[i])^2] = \sigma_s^2 \left(\frac{1}{\mathsf{SNR}_r}\right)^2 + \epsilon \tag{G.26}$$

by choosing γ arbitrarily close to γ_o from below. Finally, note that with our choice of γ, the following condition is satisfied

$$1 + \gamma^2 D^{(1)} < \mathsf{SNR}_r \tag{G.27}$$

Hence, we can ensure that the event $\mathcal{T}_u$ occurs with high probability. For all $\epsilon > 0$, Lemma 1 guarantees that there exists n_0 such that for all $n > n_0$

$$P(\mathcal{T}_u) \geq 1 - \epsilon \tag{G.28}$$

Using the same type of scheme, we send $\mathbf{v}$ over the remaining $2n$ uses of the MAC. Similar to the case in forwarding $\mathbf{u}$, we can show that under an event $\mathcal{T}_v$ that occurs with high probability for arbitrarily long block lengths, the estimation noise $\hat{\mathbf{v}} - \mathbf{v}$ is i.i.d Gaussian.

G.0.3 Stage III: Message Decoding

We perform jointly typical decoding at the destination based on estimates $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. Fix a fixed $\epsilon_1, \epsilon_2 > 0$, let $\mathbf{w}_u$ and $\mathbf{w}_v$ be the estimation errors $\hat{\mathbf{u}} - \mathbf{u}$ and $\hat{\mathbf{v}} - \mathbf{v}$ under events $\mathcal{T}_u$ and $\mathcal{T}_v$. Hence, it follows that $\mathbf{w}_u$ and $\mathbf{w}_v$ are i.i.d Gaussian noises with zero mean and respective variances D_u and D_v given by

$$D_u = (2\mathsf{SNR}_s + N)\left(\frac{1}{\mathsf{SNR}_r}\right)^2 + \epsilon_1 \tag{G.29}$$

$$D_v = 2\left(\frac{1}{\mathsf{SNR}_r}\right)^2 + \epsilon_2 \tag{G.30}$$

We define length n vectors $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ where

$$\tilde{U}^n = U^n + W_u^n \tag{G.31}$$

$$\tilde{V}^n = V^n + W_v^n \tag{G.32}$$

We calculate the average decoding error over all random codebook constructions and show that this error can be made arbitrarily small for long blocklengths. Let $\mathcal{A}_\delta$ be the set of jointly typical sequences $\left\{\left(X^n, \tilde{U}^n, \tilde{V}^n\right)\right\}$ with respect to $p(x, \tilde{u}, \tilde{v})$. We define the events

$$E_0 = \left\{\frac{1}{n}\sum_{i=1}^{n} X_i(w) < P\right\} \tag{G.33}$$

$$G = \left\{\left(X^n(w), \hat{U}^n, \hat{V}^n\right) \overset{a.s}{=} \left(X^n(w), \tilde{U}^n, \tilde{V}^n\right)\right\} \tag{G.34}$$

$$E_1 = \left\{\left(X^n(w), \hat{U}^n, \hat{V}^n\right) \notin \mathcal{A}_\delta\right\} \tag{G.35}$$

$$\tilde{E}_1 = \left\{\left(X^n(w), \tilde{U}^n, \tilde{V}^n\right) \notin \mathcal{A}_\delta\right\} \tag{G.36}$$

$$E_2 = \left\{\exists(w' \neq w) : \left(X^n(w'), \hat{U}^n, \hat{V}^n\right) \in \mathcal{A}_\delta\right\} \tag{G.37}$$

$$\tilde{E}_2 = \left\{\exists(w' \neq w) : \left(X^n(w'), \tilde{U}^n, \tilde{V}^n\right) \in \mathcal{A}_\delta\right\} \tag{G.38}$$

By the law of large numbers for all $\epsilon_o > 0$, there exists n_0 such that $Pr(E_0) < \epsilon_o$.

We bound the probability of $\hat{E}_1$ as follows

$$P(E_1) = P(E_1 \cap G) + P(E_1 \cap G^c) \tag{G.39}$$

$$\leq P(E_1 \cap G) + P(G^c) \tag{G.40}$$

$$= P(\tilde{E}_1 \cap G) + P(G^c) \tag{G.41}$$

$$\leq P(\tilde{E}_1) + P(G^c) \tag{G.42}$$

Similarly, the probability of $\hat{E}_2$ can be bounded as follows

$$P(E_2) \leq P(\tilde{E}_2) + P(G^c) \tag{G.43}$$

For all $\delta > 0$, there exists a n_1 such that for all $n \geq n_1$, $P(\tilde{E}_1) \leq \delta$. From the proof of the channel coding [1, Theorem 8.7.1], there exists a n_2 such that for all $n \geq n_2$, $P(\tilde{E}_2) \leq \delta$ if $R \leq I(X; \tilde{U}, \tilde{V}) - 3\delta$.

From the construction of the computation code, it follows that

$$P(G) = P(\mathcal{T}_u \cap \mathcal{T}_v) \tag{G.44}$$

$$P(G^c) = P(\mathcal{T}_u^c \cup \mathcal{T}_v^c) \tag{G.45}$$

$$\leq P(\mathcal{T}_u^c) + P(\mathcal{T}_v^c) \tag{G.46}$$

There exists n_3 such that for all $n \geq n_3$,

$$P(G^c) \leq P(\mathcal{T}_u^c) + P(\mathcal{T}_v^c) \tag{G.47}$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} \tag{G.48}$$

For a fixed δ, ϵ, we choose $n = \max(n_0, n_1, n_2, n_3)$ to ensure that

$$P(E_0) \leq \epsilon \tag{G.49}$$
$$P(\tilde{E}_1) \leq \delta + \epsilon \tag{G.50}$$
$$P(\tilde{E}_2) \leq \delta + \epsilon \tag{G.51}$$

We then make δ, ϵ arbitrarily small by choosing n large enough.

Finally, we find a lower bound on the mutual information $I(X; \tilde{U}, \tilde{V})$. We first rewrite $\tilde{U}$ as follows:

$$\tilde{U} = U + W_u \tag{G.52}$$
$$= VX + \sum_{m=1}^{M} h_m z_m + W_u \tag{G.53}$$
$$= \hat{V}X + (V - \hat{V})X + \sum_{m=1}^{M} h_m z_m + W_u \tag{G.54}$$

where $\hat{V} = \mathbb{E}[V|V + W_v]$ is the MMSE estimate of V given $V + W_v$. Define

$$J = (V - \hat{V})X + \sum_{m=1}^{M} h_m z_m + W_u \tag{G.55}$$

We observe that $\widehat{V}$ is independent of X and

$$Cov(X, J|\widehat{V}) = 0 \ \ a.s \tag{G.56}$$

By assuming that the noise J is Gaussian (Theorem 1 in [99]), we arrive at a lower bound of the mutual information

$$I(X; \tilde{U}, \hat{V}) \geq \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{|\widehat{V}|^2 \mathsf{SNR}_s}{Var(J|\widehat{V})}\right)\right] \tag{G.57}$$
$$\geq \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{|\widehat{V}|^2 \mathsf{SNR}_s}{\mathbb{E}[(V - \hat{V})^2|\hat{V}]\mathsf{SNR}_s + \hat{V}\sigma^2 + D_u}\right)\right] \tag{G.58}$$

In the case where $\ell > 4$, we iterate the computation scheme for forwarding $\mathbf{u}$ and $\mathbf{v}$. For example, in the case where $\ell_1 = 3$, the relays forward

$$\mathbf{x}_m = [\tau \mathbf{s} + \mathbf{d}_m] \mod \Lambda \tag{G.59}$$

in the 3rd iteration and update upon the estimate $\hat{\mathbf{u}}^{(2)}$ from the second iteration. For general ℓ_1, ℓ_2, the estimation noise variance becomes:

$$D_u = (2\mathsf{SNR}_s + 1)\left(\frac{N}{\mathsf{SNR}_r}\right)^{\ell_1} + \epsilon_1 \tag{G.60}$$

$$D_v = 2\left(\frac{1}{\mathsf{SNR}_r}\right)^{\ell_2} + \epsilon_2 \tag{G.61}$$

with high probability for arbitrarily long block lengths.