APPROACHING THE ERGODIC CAPACITY OF WIRELESS NETWORKS WITH LATTICE CODES

by

Ahmed Monier Hindy

APPROVED BY SUPERVISORY COMMITTEE:

Aria Nosratinia, Chair

Naofal Al-Dhahir

John P. Fonseka

Murat Torlak

Dedicated to my beloved father,

may his soul rest in peace.

APPROACHING THE ERGODIC CAPACITY OF WIRELESS NETWORKS WITH LATTICE CODES

by

AHMED MONIER HINDY, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

August 2017

ACKNOWLEDGMENTS

It has been my pleasure to have been one of the graduate students of Aria Nosratinia. Over the past five years Aria has been a great supervisor and a wonderful mentor. Without his patience, knowledge and endless support, this dissertation would have never seen daylight. Thanks for bringing the best out of me. I am also lucky enough to have had the best colleagues one can hope for: Ahmed Abotabl, Mohamed Fadel, Hussein Saad, Noha Helal and Fan Zhang. I have learned a lot from our discussions and interactions, more importantly thanks for the countless smiles you have drawn on my face during times of hardship. I am also thankful to my dissertation committee professors Naofal Al-Dhahir, John Fonseka and Murat Torlak for their support and helpful comments throughout my journey. I would have never been able to pursue a PhD without the support from professors at UT Dallas, Nile University and Alexandria University. I strongly believe that your mentorship has played a major role in shaping my future. Thanks to my professors Mohammed Nafie, Amr El-Keyi, Tamer ElBatt, Hesham El Gamal, Ahmed Sultan and Hossam Khairy. I am also grateful to my friends in Dallas who made me feel home from day one: Mohamed Mokhtar, Ahmed ElSamadouny, Ahmed Gamal Helmy, Ahmed Omar, Mohammed Zinati, Ahmed Hesham Mehanna, Joseph Beshay, Ahmed Ragheb, Ayman Fawzy, Yasser Abdelhamid, Mostafa Sayed, Mahmoud Elgenedy, Ahmed ElShafie, Yahia Ramadan, Mahmoud Essam and Fatemeh Saki. I would also like to thank my friends at home for their constant support throughout my PhD journey, Karim Abdullah, Karim Ashraf, Karim ElShazly, Ahmed Essam, Ahmed Helmy, Maysara Sabry, Ahmed Safaa and Ahmed Sameh. I am also indebted to my friends Ahmed Elmoslimany, Ahmed Ahmedin, Ahmed Arafa, Mohamed Khalifa, Osama Gamal, Islam Elbakoury, Eman Naguib, Ghada Hatem and Omar Mehanna. Thank you for the memorable times we have spent together.

I am indebted to my lovely wife Noha for her ongoing love and care, and I would also love to congratulate her for accomplishing this dissertation as she has been a great part of it. Thank you my son, Yaseen, for helping me throw away all my worries as soon as I hold you in my arms.

My endless love to my brother, Mohamed, and my sister, Eman, and their beautiful families for your ongoing love, care and support through my life. Most importantly, I owe any success I achieve in life to my father and mother, the most amazing parents I could dream of. I love you Mom, I know you believe I can be the best person in the entire universe, I promise I will keep trying to be so but please forgive me because I am not. Dad, it was one of my life goals to wave to you on my PhD hooding ceremony, and give you a very big hug afterwards to show you how grateful I am to you. Although I won't be able to do that, I am sure you are watching me now so I want to thank you for everything you did for me. I am lucky enough to have my God father be my real father. Thank you God for all these blessings you have surrounded me with.

June 2017

APPROACHING THE ERGODIC CAPACITY OF WIRELESS NETWORKS

WITH LATTICE CODES

Ahmed Monier Hindy, PhD
The University of Texas at Dallas, 2017

Supervising Professor: Aria Nosratinia, Chair

Despite significant progress in the area of lattice coding and decoding, their operation under ergodic fading has been mostly unexplored. In this dissertation, lattice coding and decoding are studied for several ergodic fading scenarios, and their performance is analyzed. Specifically, the multiple-input-multiple-output (MIMO) point-to-point channel, the multiple-access channel (MAC), the dirty paper channel, the broadcast channel and the interference channel are studied under stationary and ergodic fading, with channel state information available only at the receiver (CSIR), or channel state also available at the transmitter (CSIT). For the point-to-point channel, the case of noisy channel state information at the receiver is also considered. Motivated by practical considerations, the proposed decoding rules are only a function of channel statistics and do not depend on the instantaneous realizations of the channel. When the channel state information is available at all communication nodes, it is shown that lattice codes achieve the capacity of the MIMO point-to-point channel as well as the K-user broadcast channel. When channel state information is available at the receiver, it is shown that the gap to capacity is a constant that diminishes with the number of receive antennas, even at finite signal-to-noise ratio (SNR) under Rayleigh fading. Single-input-single-output (SISO) channels are also considered, where the decoding process is simpler and the gap to capacity is shown to be bounded by a constant for a wide range of fading distributions. The same conclusion follows for the MAC. Additionally, an alternative decoding approach is presented for block-fading SISO point-

to-point channels that are drawn from a discrete distribution, where channel-matching decision regions are proposed. The gap to capacity is shown to be a constant that diminishes under mild conditions. The fading MIMO dirty paper channel with CSIR is also studied, where a lattice coding/decoding scheme achieves a constant gap to capacity. An inner bound for the dirty paper channel is also developed using Gaussian codebooks in conjunction with random binning. Results are extended to MIMO broadcast channels with CSIR, and are compared to newly developed outer bounds for the broadcast channel. Finally, the two-user fading interference channel is studied under the ergodic strong regime. The capacity region of this channel is calculated using Gaussian codebooks. In addition, a lattice coding/decoding scheme is proposed, and its achievable rate region is computed, whose gap to capacity is shown to be small.

.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In practical applications, structured codes are favored due to computational complexity issues; lattice codes are an important class of structured codes that has gained special interest in the last few decades. An early attempt to characterize the performance of lattice codes in the additive white Gaussian noise (AWGN) channel was made by de Buda [1]; a result that was later corrected by Linder et al. [2]. Subsequently, Loeliger [3] showed the achievability of $\frac{1}{2}\log(\text{SNR})$ with lattice coding and decoding. Urbanke and Rimoldi [4] showed the achievability of $\frac{1}{2}\log(1+\text{SNR})$ with maximum-likelihood decoding. Erez and Zamir [5] demonstrated that lattice coding and decoding achieve the capacity of the AWGN channel using a method involving common randomness via a dither variable and minimum mean-square error (MMSE) scaling at the receiver. Subsequently, Erez et al. [6] proved the existence of lattices with good properties that achieve the performance promised in [5]. El Gamal et al. [7] showed that lattice codes achieve the capacity of the AWGN MIMO channel, as well as the optimal diversity-multiplexing tradeoff under quasi-static fading. Prasad and Varanasi [8] developed lattice-based methods to approach the diversity of the MIMO channel with low complexity. Dayal and Varanasi [9] developed diversity-optimal codes for Rayleigh fading channels using finite-constellation integer lattices and maximum-likelihood decoding. Zhan et al. [10] introduced integer-forcing linear receivers as an efficient decoding approach that exploits the linearity of lattice codebooks. Ordentlich and Erez [11] showed that in conjunction with precoding, integer-forcing can operate within a constant gap to the MIMO channel capacity. So far lattice coding results have mostly addressed channel coefficients that are either constant or quasi-static. Vituri [12] study the performance of lattice codes with unbounded power constraint under regular fading channels. Recently, Luzzi and Vehkalahti [13] showed that a class of lattices belonging to a family of division algebra codes achieve rates within a constant gap to the ergodic capacity at all SNR, where the gap depends on the algebraic properties of the code as well as the antenna configuration.

Unfortunately, the constant gap in [13] can be shown to be quite large at many useful antenna configurations, in addition to requiring substantial transmit power to guarantee any positive rate. Liu and Ling [14] showed that polar lattices achieve the capacity of the i.i.d. SISO fading channel. Campello et al. [15] also proved that algebraic lattices achieve the ergodic capacity of the SISO fading channel.

Going beyond the point-to-point channel, Song and Devroye [16] investigated the performance of lattice codes in the Gaussian relay channel. Nazer and Gastpar [17] introduced the compute-and-forward relaying strategy based on the decoding of integer combinations of interfering lattice codewords from multiple transmitters. Compute-and-forward was also an inspiration for the development of integer-forcing [10]. Özgür and Diggavi [18] showed that lattice codes can operate within a constant gap to the capacity of Gaussian relay networks. Ordentlich et al. [19] proposed lattice-based schemes that operate within a constant gap to the sum capacity of the K-user MAC, and the sum capacity of a class of K-user symmetric Gaussian interference channels. In [20], Lin et al. proposed a version of the lattice coding and decoding scheme in [7] for the fading MIMO dirty paper channel, where a decoding rule that depends on the channel realizations is used. The proof in [7], originally derived for quasi-static MIMO channels, uses the *Minkowski-Hlawka Theorem* to prove the existence of a codebook with negligible error probability for a given channel state. However, the existence of a universal codebook that achieves the same error probability over all channel states is not guaranteed and hence the achievable rates in [20] remain under question. Recently, versions of the fading dirty paper channel have been studied where the signal and dirt incur different fading processes. Bergel et al. [21] proposed a lattice coding scheme for the fading dirty paper channel with non-causal noisy channel knowledge, which is tight at small enough error variance. Recently, lattice codes have been proposed for Gaussian interference channels. This is in part motivated by [22], showing that a naive extension of the Han-Kobayashi scheme with Gaussian codes is suboptimal for the Gaussian interference channel with more than two users, and that linear codes (more specifically lattice codes) outperform Gaussian codes in such case. Following this result, Jafar and Vishwanath [23]

showed that the generalized Degrees-of-Freedom of the K-user symmetric Gaussian interference channel is achieved via lattice codes.

This dissertation investigates the rates achieved by lattice coding and decoding under stationary and ergodic conditions in multiple-antenna wireless networks. A brief outline of related results on the ergodic capacity of these channels is as follows. The ergodic capacity of the Gaussian fading channel was established by McEliece and Stark [24]. Goldsmith and Varaiya [25] extended the result to the full CSIT case. The capacity of the ergodic MIMO channel was established by Telatar [26] and Foschini and Gans [27]. Channels with imperfect channel state information have been studied in the literature. Medard [28] computed bounds on the achievable rates for a given channel estimation error variance at the receiver. The capacity region of the ergodic MIMO MAC was found by Shamai and Wyner [29]. Li and Goldsmith derived the capacity of the ergodic fading broadcast channel with channel state information at all nodes [30]. The surveys by Biglieri et al. [31] and Goldsmith et al. [32] provide a broader view of the fading channels literature.

On the other hand, several works have addressed extensions of Costa's "Writing on dirty paper" [33]. Erez et al. [34] generalized Costa's work to non-Gaussian states via dithered lattice codes and minimum mean-square error (MMSE) estimation. Recently, Renyi and Shamai [35] considered a variation where the state is multiplied by a time-varying channel coefficient that is known exclusively at the receiver. Weingarten et al. showed that dirty paper coding achieves the capacity region of the Gaussian MIMO broadcast channel [36].

More generally, the problem of channels with state has been subject to extensive research in the literature. Shannon studied the discrete-memoryless channel with state known causally at the transmitter [37]. Gel'fand and Pinsker [38] studied the discrete-memoryless channel with the state known non-causally at the transmitter. Costa [33] solved the same problem when both the state and the noise are additive and Gaussian, and showed that the capacity is unaffected by the additive state (dirt). Jafar [39] studied various channels with state, where he compared the impact on the capacity of causal and non-causal channel knowledge at the transmitter.

Several results exist for the ergodic fading broadcast channel with receive channel state information. Li and Goldsmith derived the capacity of the ergodic fading broadcast channel with channel state information at all nodes [30]. Tuninetti and Shamai [40] derived an inner bound for the rate region that is based on random coding and joint decoding at one of the receivers. Tse and Yates [41] derived inner and outer bounds for the fading broadcast channel. The bounds provided are solutions of integrals. Jafarian and Vishwanath [42] computed outer bounds that are valid when the fading coefficients are drawn from a discrete distribution. Jafar and Goldsmith [43] proved that increasing the number of transmit antennas in an isotropic fading SISO broadcast channel with CSIR does not increase the capacity.

For the two-user Gaussian interference channel, the best known inner bound to date is that of Han and Kobayashi [44], which meets the capacity region under strong interference. The capacity region of the strong interference channel was also derived in [45]. A notable result due to Etkin et al. [46], showed that a special case of the inner bound in [44] is within one bit of capacity. Most of the results on the Gaussian interference channel to date focus on scenarios without channel fading. Recently, a few papers addressed stationary and ergodic channel fading coefficients. In [47] Sankar et al. considered the two-user ergodic fading interference channel with channel state information available at all nodes, where Gaussian codes were shown to achieve the sum capacity under some scenarios, namely the *ergodic very strong* (a mix of strong and weak fading states satisfying specific strict fading average conditions) and *uniformly strong*, (every cross fading coefficient is larger in magnitude than its corresponding direct fading coefficient) interference regimes. Farsani [48] also studied a variant of the latter problem with partial channel state information at the transmitters.

The contributions of this dissertation are as follows.

- **MIMO Point-to-point Channel with CSIR.** In this dissertation we propose a lattice coding and decoding strategy and analyze its performance for a variety of MIMO ergodic channels, showing that the gap to capacity is small at both high and low SNR. The fading

processes in this paper are finite-variance stationary and ergodic. First, we present a lattice coding scheme for the MIMO point-to-point channel under isotropic fading, whose main components include the class of nested lattice codes proposed in [5] in conjunction with a time-varying MMSE matrix at the receiver. The proposed decision regions are spherical and depend only on the channel distribution, and hence the decision regions remain unchanged throughout subsequent codeword transmissions.[1] The relation of the proposed decoder with Euclidean lattice decoding is also discussed. The rates achieved are within a constant gap to the ergodic capacity for a broad class of fading distributions. Under Rayleigh fading, a bound on the gap to capacity is explicitly characterized which vanishes as the number of receive antennas grows. Similar results are also derived for the fading K-user MIMO MAC. The proposed scheme provides useful insights on the implementation of MIMO systems under ergodic fading. First, the results reveal that structured codes can achieve rates within a small gap to capacity. Moreover, channel-independent decision regions approach optimality when the number of receive antennas is large. Furthermore, for the special case of SISO channels the gap to capacity is characterized for all SNR values and over a wide range of fading distributions. Unlike [13], the proposed scheme achieves positive rates at low SNR where the gap to capacity vanishes. At moderate and high SNR, the gap to capacity is bounded by a constant that is independent of SNR and only depends on the fading distribution. In the SISO channel under Rayleigh fading, the gap is a diminishing fraction of the capacity as the SNR increases.

Moreover, an alternative decoding approach is proposed for MIMO block-fading channels drawn from a discrete distribution. We show that when the decision regions are designed to match the channel realizations, the gap to capacity is a constant that depends on the channel coherence length as well as the fading distribution size. Special cases where the gap to capacity vanishes are also provided.

[1] Although the decision regions are designed independently of the channel realizations, the received signal is multiplied by an MMSE matrix prior to decoding the signal, and hence channel knowledge at the receiver remains necessary for the results in this paper.

- **MIMO Point-to-point Channel with CSIT.** Under fading and in the presence of CSIT, the most straight forward capacity approaching schemes employ separable coding, i.e., coding independently and in parallel over different fading states of the channel [25, 32]. Unfortunately, separable coding causes significant delay and requires large memory at both the transmitter and receiver. Also, separable encoding in practice forces a low-probability fading state to either cause huge delays, small block-length effects, or the symbols in that fading state to be ignored with associated rate loss. As a result, achieving the ergodic capacity of block-fading channels without separable coding (i.e., with coding across states) remains an important theoretical and practical question.[2] In this dissertation we show that non-separable lattice coding achieves the ergodic capacity of the block fading SISO channel. At the transmitter, the symbols of the codeword are permuted across time using a linear permutation matrix. Time-varying MMSE scaling is used at the receiver, followed by a decoder that is universal for all fading realizations drawn from a given fading distribution. Hence, the codebook design and decision regions are fixed across different transmissions; the only channel-dependent blocks are the permutation and MMSE scaling functions. We first highlight the main ideas of the proposed scheme in the context of a heuristic channel model whose behavior approximates the ergodic fading channel. We then generalize the solution to all fading distributions whose realizations are robustly typical, and to continuous distributions via a bounding argument. The results are then extended to MIMO block-fading channels.

- **Point-to-point Channel with imperfect CSIR.** Under imperfect channel state information at the receiver, bounds on the achievable rates using lattice coding and decoding are derived for a given channel estimation error variance, similar to the approach adopted in Medard's work [28]. Nested lattice codes are used whereas the decision regions proposed

[2]It was pointed out in [31] that under maximum likelihood decoding the ergodic capacity of point-to-point channels with CSIT can be attained using Gaussian signaling without separable coding. However, the same result is not necessarily true for non-Gaussian (structured) codebooks.

depend on the statistics of the fading channel and not the actual channel realizations. This offers a notable advantage from the perspective of decoding complexity, since the decision regions remain fixed for a given fading distribution. For a wide range of fading distributions, the gap between the rates achieved using lattice codebooks compared to Gaussian codebooks is a constant that does not depend on the channel estimation error variance nor the input power. As a byproduct, we use the Generalized Degrees-of-Freedom [46] as a performance metric for channels with estimation error, where it is shown that the lattice scheme achieves the same GDoF as Medard's Gaussian-input scheme.

- **Ergodic MIMO Multiple-Access Channel.** The point-to-point scheme under CSIR is extended to the K-user ergodic multiple-access channel. Similar to the point-to-point case, the rates achieved are within a constant gap to the sum capacity for a class of fading distributions. Under Rayleigh fading, this gap vanishes when the number of antennas at the receiver is large: a result that may have direct applications to massive MIMO systems [49].

- **MIMO Dirty paper channel with CSIR.** A modified version of Costa's dirty paper problem is studied. In the model under study, both the input signal and the state are multiplied by the same time-varying channel coefficient, which is known only at the receiver. Explicit capacity expressions of the problem are non-trivial since the knowledge of the states is distributed among the transmitter and the receiver (the transmitter has non-causal knowledge of the dirt while the receiver has knowledge of the channel). Following similar steps to the work in [34], a lattice coding and decoding scheme is proposed for the fading dirty paper channel under study. As will be shown, both the encoding process and decoding process involve significantly reduced complexity, compared to the random binning and maximum likelihood decoding techniques used in Costa's results [33]. Moreover, the state distribution is allowed to be more general, and is not necessarily Gaussian. Under a wide range of fading distributions, the gap to capacity is shown to be a constant

that does not depend on either the signal or state power. The results are also extended to the MIMO dirty paper channels, where it is shown that the gap to capacity diminishes with the number of receive antennas.

Moreover, the results are applied to a class of broadcast channels with ergodic fading and receive channel state information. Under certain SNR regimes, the rate region achieved by the proposed schemes are close to the capacity region.

- **MIMO Broadcast channel with CSIR.** One advantage of the dirty paper model is its straightforward extension to the broadcast channel with CSIR, where each receiver decodes a signal contaminated by interference stemming from the same source, and hence the desired signal and interference undergo the same fading process. Hence, we apply the dirty paper channel results to a two-user MIMO BC with different fading dynamics, where the fading process is stationary and ergodic for one receiver and quasi-static for the other receiver. In addition, the case where both users experience ergodic fading processes that are independent of each other is also studied. Unlike conventional broadcast channel techniques, the proposed scheme does not require any receiver to know the codebooks of the interference signals. Performance is compared with a version of dirty paper coding under non-causal CSIT. For the cases under study, the lattice coding scheme achieves rates very close to dirty paper coding over most of the rate region.

- **Broadcast channel with CSIT.** The K-user ergodic broadcast channel with full CSI is also studied, using a variant of the point-to-point scheme. The proposed approach is based on lattice coding and decoding in conjunction with separable coding. It is shown that this approach achieves the capacity region of the broadcast channel.

- **Ergodic Strong Interference Channel.** In this paper, the two-user ergodic fading Gaussian interference channel is considered under *ergodic strong* conditions, i.e., a mix of strong and weak fading states where the cross-fading coefficient is statistically stronger

than the direct-fading coefficient, where channel state information is available only at the receivers. To the best of our knowledge, this is the first result on ergodic strong interference channels in the literature, which obviously includes the uniformly strong and ergodic very strong interference channels as special cases. First, an outer bound of the capacity region is established, which can be achieved via Gaussian codes. Moreover, a lattice coding and decoding scheme is proposed, which is based on nested lattice codes at the transmitters and MMSE scaling in conjunction with lattice decoding at the receivers. The decoding rule proposed is universal for all realizations of a given fading distribution, and hence offering significant advantages in terms of computational complexity.

Notation

Boldface lowercase letters denote column vectors and boldface uppercase letters denote matrices. The set of real numbers, complex numbers and positive integers are denoted $\mathbb{R}, \mathbb{C}, \mathbb{Z}^+$, respectively. $\boldsymbol{A}^T, \boldsymbol{A}^H$ denote the transpose and Hermitian transpose of matrix $\boldsymbol{A}$, respectively. a_i denotes element i of $\boldsymbol{a}$. $\boldsymbol{A}(i,:)$ is row i in $\boldsymbol{A}$. a_i denotes element i of $\boldsymbol{a}$. $\boldsymbol{\delta}_k^{(n)}$ is a length n column vector whose entry k is one, with all other entries being zero. $\boldsymbol{A} \succeq \boldsymbol{B}$ indicates that $\boldsymbol{A} - \boldsymbol{B}$ is positive semi-definite. $\det(\boldsymbol{A})$ and $\mathrm{tr}(\boldsymbol{A})$ denote the determinant and trace of $\boldsymbol{A}$, respectively. $\mathcal{B}(q)$ is an n-dimensional ball of radius q and the volume of shape $\mathcal{A}$ is denoted $\mathrm{Vol}(\mathcal{A})$. log, ln represent logarithms in base 2 and the natural logarithm, respectively. We define $j = \sqrt{-1}$. Real and imaginary parts of a complex number are shown with superscripts ^{R} and ^{I}, respectively. $\mathbf{1}_n$ is the length n all ones column vector. $\mathbb{P}, \mathbb{E}$ denote the probability and expectation operators, respectively. κ^+ denotes $\max\{\kappa, 0\}$. $|\mathcal{A}|$ denotes the number of elements in set $\mathcal{A}$.

CHAPTER 2

PRELIMINARIES

2.1 Lattice Codes

A lattice Λ is a discrete subgroup of $\mathbb{R}^n$ which is closed under reflection and real addition. The fundamental Voronoi region $\mathcal{V}$ of the lattice Λ is defined by

$$\mathcal{V} = \big\{ \boldsymbol{s} : \arg\min_{\boldsymbol{\lambda}\in\Lambda} ||\boldsymbol{s} - \boldsymbol{\lambda}|| = \mathbf{0} \big\}. \tag{2.1}$$

The *second moment per dimension* of Λ is defined as

$$\sigma_\Lambda^2 = \frac{1}{n\mathrm{Vol}(\mathcal{V})} \int_{\mathcal{V}} ||\boldsymbol{s}||^2 d\boldsymbol{s}, \tag{2.2}$$

and the *normalized second moment* $G(\Lambda)$ of Λ is

$$G(\Lambda) = \frac{\sigma_\Lambda^2}{\mathrm{Vol}^{\frac{2}{n}}(\mathcal{V})}, \tag{2.3}$$

where $G(\Lambda) > \frac{1}{2\pi e}$ for any lattice in $\mathbb{R}^n$.

Every $\boldsymbol{s} \in \mathbb{R}^n$ can be uniquely written as $\boldsymbol{s} = \boldsymbol{\lambda} + \boldsymbol{e}$ where $\boldsymbol{\lambda} \in \Lambda$, $\boldsymbol{e} \in \mathcal{V}$. The quantizer is then defined by

$$Q_\mathcal{V}(\boldsymbol{s}) = \boldsymbol{\lambda}\,, \quad \text{if } \boldsymbol{s} \in \boldsymbol{\lambda} + \mathcal{V}. \tag{2.4}$$

Define the modulo-Λ operation corresponding to $\mathcal{V}$ as follows

$$[\boldsymbol{s}] \,\mathrm{mod}\Lambda \triangleq \boldsymbol{s} - Q_\mathcal{V}(\boldsymbol{s}). \tag{2.5}$$

The mod Λ operation also satisfies

$$\big[\boldsymbol{s} + \boldsymbol{t}\big] \,\mathrm{mod}\Lambda = \big[\boldsymbol{s} + [\boldsymbol{t}] \,\mathrm{mod}\Lambda\big] \,\mathrm{mod}\Lambda \quad \forall \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^n. \tag{2.6}$$

The lattice Λ is said to be nested in Λ_1 if $\Lambda \subseteq \Lambda_1$.

We employ the class of nested lattice codes proposed in [5]. The transmitter constructs a codebook $\mathcal{L}_1 = \Lambda_1 \cap \mathcal{V}$, whose rate is given by

$$R = \frac{1}{n} \log \frac{\mathrm{Vol}(\mathcal{V})}{\mathrm{Vol}(\mathcal{V}_1)}\,. \tag{2.7}$$

The coarse lattice Λ has an arbitrary second moment P_x and is good for covering and quantization, whereas the fine lattice Λ_1 is good for AWGN coding, where both are construction-A lattices [3, 5]. The existence of such lattices has been proven in [6].

A lattice Λ is good for covering if

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\mathrm{Vol}(\mathcal{B}_n(R_c))}{\mathrm{Vol}(\mathcal{B}_n(R_f))} = 0, \tag{2.8}$$

where the covering radius R_c is the radius of the smallest sphere spanning $\mathcal{V}$ and R_f is the radius of the sphere whose volume is equal to $\mathrm{Vol}(\mathcal{V})$. In other words, for a good nested lattice code with second moment P_x, the Voronoi region $\mathcal{V}$ approaches a sphere of radius $\sqrt{nP_x}$. A lattice Λ is good for quantization if

$$\lim_{n\to\infty} G(\Lambda) = \frac{1}{2\pi e}\,. \tag{2.9}$$

A key ingredient of the lattice coding scheme in [5] is using common randomness (dither) $\boldsymbol{d}$ in conjunction with the lattice code at the transmitter. $\boldsymbol{d}$ is also known at the receiver, and is drawn uniformly over $\mathcal{V}$.

Lemma 1. *[5, Lemma 1] If $\boldsymbol{t} \in \mathcal{V}$ is independent of $\boldsymbol{d}$, then $\boldsymbol{x}$ is uniformly distributed over $\mathcal{V}$ and independent of the lattice point $\boldsymbol{t}$.*

Lemma 2. *[50, Theorem 1]. An optimal lattice quantizer with second moment σ_Λ^2 is white, and the autocorrelation of its dither $\boldsymbol{d}_{opt}$ is given by $\mathbb{E}[\boldsymbol{d}_{opt}\boldsymbol{d}_{opt}^T] = \sigma_\Lambda^2 \boldsymbol{I}_n$.*

Note that the optimal lattice quantizer is a lattice quantizer with the minimum $G(\Lambda)$. Since the proposed class of lattices is good for quantization, the autocorrelation of $\boldsymbol{d}$ approaches that of $\boldsymbol{d}_{\mathrm{opt}}$ as n increases.

For a more comprehensive review on lattice codes see [51].

2.2 Robust Typicality

Let x be a random variable with finite support $\{\alpha_1, \ldots, \alpha_\chi\}$ according to a probability distribution $\mathbb{P}$, and let $\boldsymbol{x} = [x_1, \ldots, x_n]$ be a sequence of independent samples/realizations of x. Denote

by n_k the number of occurrences of the outcome α_k. Orlitsky and Roche [52] introduced the notion of *robust typicality*, where the sequence $\boldsymbol{x}$ is said to be δ-robustly typical for $\delta > 0$ if for all $k \in \{1, \ldots, \chi\}$,

$$|n_k - n\mathbb{P}_k| \leq \delta n \mathbb{P}_k. \tag{2.10}$$

where $\mathbb{P}_k$ is a short-hand notation for $\mathbb{P}(\alpha_k)$. In words, (2.10) implies that for small values of δ a sequence is δ-robustly typical if the number of occurrences of α_k is not too far from np_k for all k. The following lemma, proved in [52], reveals that when the random sequence $\boldsymbol{x}$ drawn from $\mathbb{P}$ is long enough, then it is robustly typical with high probability.

Lemma 3. *[52, Lemma 17] The probability of a sequence $\boldsymbol{x}$ of length n being not δ-robustly typical is upper bounded by*

$$\mathbb{P}(\boldsymbol{x} \notin T_\delta) \leq \sum_{k=1}^{\chi} \mathbb{P}\big(|n_k - n\mathbb{P}_k| > \delta n \mathbb{P}_k\big) \leq 2\chi e^{-\delta^2 \mu n/3}, \tag{2.11}$$

where T_δ is the set of typical sequences defined for the probability distribution $\mathbb{P}$, and $\mu \triangleq \min \mathbb{P}_k$ is the smallest non-zero probability in $\mathbb{P}$.

CHAPTER 3

ERGODIC POINT-TO-POINT CHANNEL

3.1 Point-to-point Channel with CSIR

3.1.1 MIMO Channel with Isotropic Fading[1]

Consider a MIMO point-to-point channel with N_t transmit antennas and N_r receive antennas. The received signal at time instant i is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x}_i + \boldsymbol{w}_i, \tag{3.1}$$

where $\boldsymbol{H}_i$ is an $N_r \times N_t$ matrix denoting the channel coefficients at time i. The channel is zero-mean with strict-sense stationary and ergodic time-varying gain. Moreover, $\boldsymbol{H}$ is isotropically distributed, i.e., $\mathbb{P}(\boldsymbol{H}) = \mathbb{P}(\boldsymbol{HV})$ for any unitary matrix $\boldsymbol{V}$ independent of $\boldsymbol{H}$. We first consider real-valued channels; the extension to complex-valued channels will appear later in this section. The receiver has instantaneous channel knowledge, whereas the transmitter only knows the channel distribution. $\boldsymbol{x}_i \in \mathbb{R}^{N_t}$ is the transmitted vector at time i, where the codeword

$$\boldsymbol{x} \triangleq [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_n^T]^T \tag{3.2}$$

is transmitted throughout n channel uses and satisfies $\mathbb{E}[||\boldsymbol{x}||^2] \leq nP_x$. The noise $\boldsymbol{w} \in \mathbb{R}^{N_r n}$ defined by $\boldsymbol{w}^T \triangleq [\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_n^T]^T$ is a zero-mean i.i.d. Gaussian noise vector with covariance $\boldsymbol{I}_{N_r n}$, and is independent of the channel realizations. For convenience, we define the SNR per transmit antenna to be $\rho \triangleq P_x/N_t$.

Theorem 1. *For the ergodic fading MIMO channel with isotropic fading, any rate R satisfying*

$$R < -\frac{1}{2} \log \det \left(\mathbb{E}\left[(\boldsymbol{I}_{N_t} + \rho \boldsymbol{H}^T \boldsymbol{H})^{-1} \right] \right) \tag{3.3}$$

is achievable using lattice coding and decoding.

[1] © 2017 IEEE. Reprinted, with permission, from A. Hindy and A. Nosratinia, Lattice Coding and Decoding for Multiple-Antenna Ergodic Fading Channels, IEEE Transactions on Communications, May 2017

Proof. Encoding: Nested lattice codes $\Lambda \subseteq \Lambda_1$ are used. The transmitter emits a lattice point $\boldsymbol{t} \in \Lambda_1$ that is dithered with $\boldsymbol{d}$ which is drawn uniformly over $\mathcal{V}$. Λ has a second moment P_x and is good for covering and quantization. Λ_1 is good for AWGN coding. Both are construction-A lattices [3, 5]. The dithered codeword is then as follows

$$\boldsymbol{x} = \big[\boldsymbol{t} - \boldsymbol{d}\big] \bmod \Lambda \; = \boldsymbol{t} - \boldsymbol{d} + \boldsymbol{\lambda}\,, \tag{3.4}$$

where $\boldsymbol{\lambda} = -Q_{\mathcal{V}}(\boldsymbol{t}-\boldsymbol{d}) \in \Lambda$ from (2.5). The coarse lattice $\Lambda \in \mathbb{R}^{N_t n}$ has a second moment ρ. The codeword is composed of n vectors $\boldsymbol{x}_i$, each of length N_t as shown in (3.2), which are transmitted throughout n channel uses.

Decoding: The received signal can be expressed in the form $\boldsymbol{y} = \boldsymbol{H}_s\boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{H}_s$ is a block-diagonal matrix whose diagonal block i is $\boldsymbol{H}_i$. The received signal $\boldsymbol{y}$ is multiplied by a matrix $\boldsymbol{U}_s \in \mathbb{R}^{N_r n \times N_t n}$ and the dither is removed as follows

$$\begin{aligned}
\boldsymbol{y}' \triangleq& \boldsymbol{U}_s^T\boldsymbol{y} + \boldsymbol{d} \\
=& \boldsymbol{x} + (\boldsymbol{U}_s^T\boldsymbol{H}_s - \boldsymbol{I}_{N_t n})\boldsymbol{x} + \boldsymbol{U}_s^T\boldsymbol{w} + \boldsymbol{d} \\
=& \boldsymbol{t} + \boldsymbol{\lambda} + \boldsymbol{z},
\end{aligned} \tag{3.5}$$

where

$$\boldsymbol{z} \triangleq (\boldsymbol{U}_s^T\boldsymbol{H}_s - \boldsymbol{I}_{N_t n})\boldsymbol{x} + \boldsymbol{U}_s^T\boldsymbol{w}, \tag{3.6}$$

and $\boldsymbol{t}$ is independent of $\boldsymbol{z}$, according to Lemma 1. The matrix $\boldsymbol{U}_s$ that minimizes $\mathbb{E}\big[||\boldsymbol{z}||^2\big]$ is then a block-diagonal matrix whose diagonal block i is the $N_t \times N_r$ MMSE matrix at time i given by

$$\boldsymbol{U}_i = \rho(\boldsymbol{I}_{N_r} + \rho\boldsymbol{H}_i\boldsymbol{H}_i^T)^{-1}\boldsymbol{H}_i. \tag{3.7}$$

From (3.6),(3.7), the equivalent noise at time i, i.e., $\boldsymbol{z}_i \in \mathbb{R}^{N_t}$, is expressed as

$$\begin{aligned}
\boldsymbol{z}_i =& \Big(\rho\boldsymbol{H}_i^T(\boldsymbol{I}_{N_r} + \rho\boldsymbol{H}_i\boldsymbol{H}_i^T)^{-1}\boldsymbol{H}_i - \boldsymbol{I}_{N_t}\Big)\boldsymbol{x}_i + \rho\boldsymbol{H}_i^T(\boldsymbol{I}_{N_r} + \rho\boldsymbol{H}_i\boldsymbol{H}_i^T)^{-1}\boldsymbol{w}_i \\
=& -(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}_i^T\boldsymbol{H}_i)^{-1}\boldsymbol{x}_i + \rho\boldsymbol{H}_i^T(\boldsymbol{I}_{N_r} + \rho\boldsymbol{H}_i\boldsymbol{H}_i^T)^{-1}\boldsymbol{w}_i,
\end{aligned} \tag{3.8}$$

where (3.8) holds from the matrix inversion lemma, and $\boldsymbol{z} \triangleq [\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_n^T]^T$. Naturally, the distribution of $\boldsymbol{z}$ conditioned on $\boldsymbol{H}_i$ (which is known at the receiver) varies across time. To avoid complications arising from known but variable channel gains, we ignore the instantaneous channel knowledge, i.e., the receiver considers $\boldsymbol{H}_i$ a random matrix after equalization. The following lemma highlights certain geometric properties of $\boldsymbol{z}$ in the $N_t n$-dimensional space.

Lemma 4. *Let Ω be a sphere defined by*

$$\Omega \triangleq \{\boldsymbol{v} \in \mathbb{R}^{N_t n} \,:\, ||\boldsymbol{v}||^2 \leq (1+\epsilon) tr(\bar{\boldsymbol{\Sigma}})\}, \tag{3.9}$$

where $\bar{\boldsymbol{\Sigma}} \triangleq \rho\, \mathbb{E}\big[(\boldsymbol{I}_{N_t n} + \rho \boldsymbol{H}_s^T \boldsymbol{H}_s)^{-1}\big]$. Then, for any $\epsilon > 0$ and $\gamma > 0$, there exists $n_{\gamma,\epsilon}$ such that for all $n > n_{\gamma,\epsilon}$,

$$\mathbb{P}\big(\boldsymbol{z} \notin \Omega\big) < \gamma. \tag{3.10}$$

Proof. See Appendix A.1. □

We apply a version of the ambiguity decoder proposed in [3] defined by the spherical decision region Ω in (3.9).[2] The decoder chooses $\hat{\boldsymbol{t}} \in \Lambda_1$ if the received point falls inside the decision region of the lattice point $\hat{\boldsymbol{t}}$, but not in the decision region of any other lattice point.

Error Probability: As shown in [3, Theorem 4], on averaging over the set of all good construction-A fine lattices $\mathcal{C}$ of rate R, the probability of error can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e < \mathbb{P}(\boldsymbol{z} \notin \Omega) + (1+\delta)\,\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V}_1)} = \mathbb{P}(\boldsymbol{z} \notin \Omega) + (1+\delta) 2^{nR}\,\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.11}$$

for any $\delta > 0$, where (3.11) follows from (2.7). This is a union bound involving two events: the event that the noise vector is outside the decision region, i.e., $\boldsymbol{z} \notin \Omega$ and the event that the post-equalized point is in the intersection of two decision regions, i.e., $\big\{\boldsymbol{y}' \in \{\boldsymbol{t}_1 + \Omega\} \cap \{\boldsymbol{t}_2 + \Omega\}\big\}$, where $\boldsymbol{t}_1,\ \boldsymbol{t}_2 \in \Lambda_1$ are two distinct lattice points. Owing to Lemma 4, the probability of the first event vanishes with n. Consequently, the error probability can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e < \gamma + (1+\delta) 2^{nR} \frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.12}$$

[2] Ω satisfies the condition in [3] of being a bounded measurable region of $\mathbb{R}^{N_t n}$, from (3.9).

for any $\gamma, \delta > 0$. For convenience define $\boldsymbol{\Psi} = \rho\bar{\boldsymbol{\Sigma}}^{-1}$. The volume of Ω is given by

$$\mathrm{Vol}(\Omega) = (1+\epsilon)^{\frac{N_t n}{2}} \mathrm{Vol}\big(\mathcal{B}_{N_t n}(\sqrt{N_t n \rho})\big) \det\left(\boldsymbol{\Psi}^{\frac{-1}{2}}\right). \tag{3.13}$$

The second term in (3.12) is bounded by

$$\begin{aligned} &(1+\delta)2^{nR}(1+\epsilon)^{N_t n/2} \frac{\mathrm{Vol}(\mathcal{B}_{N_t n}(\sqrt{N_t n\rho}))}{\mathrm{Vol}(\mathcal{V})} \det\left(\boldsymbol{\Psi}^{\frac{-1}{2}}\right) \\ &= (1+\delta)2^{-N_t n\left(-\frac{1}{N_t n}\log\left(\frac{\mathrm{Vol}(\mathcal{B}_{N_t n}(\sqrt{N_t n\rho}))}{\mathrm{Vol}(\mathcal{V})}\right)+\xi\right)}, \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} \xi &\triangleq \frac{-1}{2}\log(1+\epsilon) - \frac{1}{2N_t n}\log\det(\boldsymbol{\Psi}^{-1}) - \frac{1}{N_t}R \\ &= \frac{-1}{2}\log(1+\epsilon) - \frac{1}{2N_t}\log\det\left(\mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^T\boldsymbol{H})^{-1}\big]\right) - \frac{1}{N_t}R. \end{aligned} \tag{3.15}$$

From (2.8), since the lattice Λ is good for covering, the first term of the exponent in (3.14) vanishes. From (3.14), whenever ξ is a positive constant we have $\mathbb{P}_e \to 0$ as $n \to \infty$, where ξ is positive as long as

$$R < -\frac{1}{2}\log\det\left(\mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^T\boldsymbol{H})^{-1}\big]\right) - \frac{1}{2}\log(1+\epsilon) - \epsilon',$$

where ϵ, ϵ' are positive numbers that can be made arbitrarily small by increasing n. From (3.5), the outcome of the decoding process in the event of successful decoding is $\hat{\boldsymbol{t}} = \boldsymbol{t} + \boldsymbol{\lambda}$, where the transformation of $\boldsymbol{t}$ by $\boldsymbol{\lambda} \in \Lambda$ does not involve any loss of information. Hence, on applying the modulo-Λ operation on $\hat{\boldsymbol{t}}$

$$[\hat{\boldsymbol{t}}] \bmod \Lambda \;=\; [\boldsymbol{t} + \boldsymbol{\lambda}] \bmod \Lambda \;=\; \boldsymbol{t}, \tag{3.16}$$

where the second equality follows from (2.6) since $\boldsymbol{\lambda} \in \Lambda$. Since the probability of error in (3.12) is averaged over the set of lattices in $\mathcal{C}$, there exists at least one lattice that achieves the same (or less) error probability. Following in the footsteps of [5, 7], the existence of a sequence of covering-good coarse lattices with second moment ρ that are nested in Λ_1 can be shown. The final step required to conclude the proof is extending the result to Euclidean lattice decoding, which is provided in the following lemma.

Lemma 5. *The error probability of the Euclidean lattice decoder given by*[3]

$$\hat{\boldsymbol{t}} = \left[arg \min_{\boldsymbol{t} \in \Lambda_1} ||\boldsymbol{y}' - \boldsymbol{t}'||^2 \right] \ mod \Lambda \tag{3.17}$$

is upper-bounded by that of the ambiguity decoder in (3.9)*.*

Details of the proof of Lemma 5 is provided in Appendix A.2, whose outline is as follows. For the cases where the ambiguity decoder declares a valid output ($\boldsymbol{y}'$ lies exclusively within a unique decision sphere), both the Euclidean lattice decoder and the ambiguity decoder with spherical regions would be identical, since a sphere is defined by the Euclidean metric. However, when the ambiguity decoder fails to declare an output (ambiguity or atypical received sequence), the Euclidean lattice decoder still yields a valid output, and hence is guaranteed to achieve the same (or better) error performance, compared to the ambiguity decoder. This concludes the proof of Theorem 1. □

We extend the results to complex-valued channels using a similar technique to that in [17, Theorem 6].

Theorem 2. *For the ergodic fading MIMO channel with complex-valued channels $\tilde{\boldsymbol{H}}$ that are known at the receiver, any rate R satisfying*

$$R < -\log\det\left(\mathbb{E}\left[(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\right]\right) \tag{3.18}$$

is achievable using lattice coding and decoding.

Proof. A sketch of the components of the proof specific to complex-valued channels is provided. Other parts follow the proof of Theorem 1 for real-valued channel coefficients.

Encoding: Since the channel is complex-valued, *two* independent codewords are selected from the same nested lattice code $\Lambda_1' \supseteq \Lambda'$ where Λ' has a second moment $\rho/2$. The transmitted signal is the combination of the two dithered lattice codewords,

$$\tilde{\boldsymbol{x}} = \left[\boldsymbol{t}' - \boldsymbol{d}'\right] \mathrm{mod}\Lambda' + j\left[\boldsymbol{t}'' - \boldsymbol{d}''\right] \mathrm{mod}\Lambda', \tag{3.19}$$

[3]The Euclidean decoder in (3.17) does not involve the channel realizations, unlike that in [7, 13].

with $\mathbb{E}\big[||\tilde{\boldsymbol{x}}||^2\big] \leq nP_x$, and the dithers $\boldsymbol{d}', \boldsymbol{d}''$ are independent. Note that the independence of the dithers is crucial for successful encoding.

Decoding: The MMSE matrix at time i is given by

$$\tilde{U}_i = \rho \tilde{\boldsymbol{H}}_i^H (\rho \tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H + \boldsymbol{I}_{N_r})^{-1}. \tag{3.20}$$

Following MMSE scaling and dither removal similar to (3.5), the real and imaginary equivalent channels at the receiver are as follows

$$\begin{aligned} \tilde{\boldsymbol{y}}'^R &= \boldsymbol{t}' + \boldsymbol{\lambda}' + \tilde{\boldsymbol{z}}^R, \\ \tilde{\boldsymbol{y}}'^I &= \boldsymbol{t}'' + \boldsymbol{\lambda}'' + \tilde{\boldsymbol{z}}^I, \end{aligned} \tag{3.21}$$

where $\boldsymbol{\lambda}'$,$\boldsymbol{\lambda}'' \in \Lambda'$. $\tilde{\boldsymbol{z}}^R$ and $\tilde{\boldsymbol{z}}^I$ are the equivalent noise components over the real and imaginary channels, respectively. The autocorrelation of each is given by

$$\boldsymbol{\Sigma}_{ii} = \mathbb{E}[\boldsymbol{z}_i^R (\boldsymbol{z}_i^R)^H] = \mathbb{E}[\boldsymbol{z}_i^I (\boldsymbol{z}_i^I)^H] = \frac{\rho}{2} \mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\big]. \tag{3.22}$$

The decoder then recovers the lattice points $\boldsymbol{t}'$ and $\boldsymbol{t}''$ independently over the real and imaginary domains. Hence, the complex-valued channel is transformed to two parallel real-valued channels at the receiver, where reliable rates can be achieved on each channel as long as

$$\check{R} < -\frac{1}{2} \log\det\Big(\mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\big]\Big), \tag{3.23}$$

and double the rate in (3.23) is the achievable rate in (3.18) for the complex-valued MIMO channel. □

We compare the achievable rate in (3.18) with the ergodic capacity for isotropic MIMO channels, given by [26]

$$C = \mathbb{E}\big[\log\det(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})\big]. \tag{3.24}$$

Corollary 1. *The gap Δ between the rate of the lattice scheme* (3.18) *and the ergodic capacity in* (3.24) *for the $N_t \times N_r$ ergodic fading MIMO channel is upper bounded by*

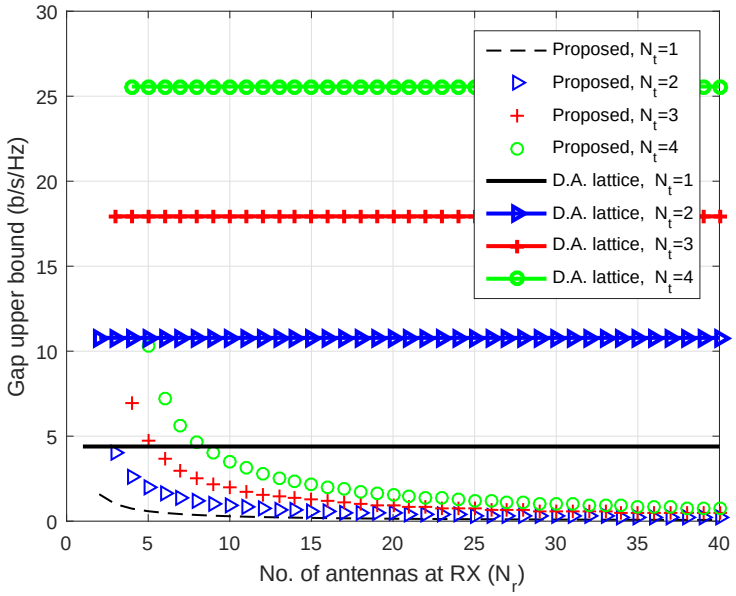

Figure 3.1. Gap to capacity under Rayleigh fading MIMO vs. division algebra lattices, for all $\rho \geq 1$.

- $N_r \geq N_t$ *and* $\rho \geq 1$*: For any channel for which all elements of* $\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\big] < \infty$

$$\Delta < \log\det\Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}]\big)\,\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\big]\Big). \tag{3.25}$$

- $N_r > N_t$ *and* $\rho \geq 1$*: When* $\tilde{\boldsymbol{H}}$ *is i.i.d. complex Gaussian with zero mean and unit variance,*

$$\Delta < N_t \,\log\big(1 + \frac{N_t + 1}{N_r - N_t}\big). \tag{3.26}$$

- $N_t = 1$ *and and* $\rho < \frac{1}{\mathbb{E}[||\tilde{\boldsymbol{h}}||^2]}$*: When* $\mathbb{E}\big[||\tilde{\boldsymbol{h}}||^4\big] < \infty$*,*

$$\Delta < 1.45\,\mathbb{E}\big[||\tilde{\boldsymbol{h}}||^4\big]\,\rho^2. \tag{3.27}$$

Proof. See Appendix A.3. □

The expression in (3.26) for Rayleigh fading is depicted in Figure 3.1 for a number of antenna configurations. The gap-to-capacity vanishes with N_r for any $\rho \geq 1$. This result has two crucial implications. First, under certain antenna configurations, lattice codes approximate the capacity

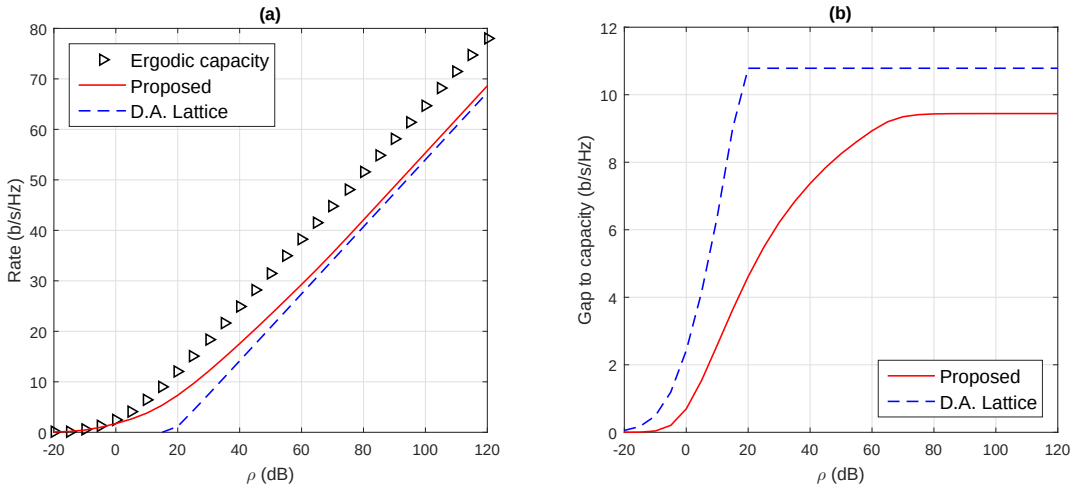

Figure 3.2. Rates achieved by the proposed scheme vs. ergodic capacity under i.i.d. Rayleigh fading with $N_t = N_r = 2$.

at finite SNR. Moreover, channel-independent decision regions approach optimality for large N_r. The results are also compared with that of the class of division algebra lattices proposed in [13] (denoted D.A. lattice), whose gap-to-capacity is both larger and insensitive to N_r. For the square MIMO channel with $N_t = N_r = 2$, the throughput of the proposed lattice scheme is plotted in Figure 3.2 and compared with that of [13]. The gap to capacity is also plotted, which show that for the proposed scheme the gap also saturates when $N_t = N_r$.

Remark 1. *Division algebra codes in [13] guarantee non-zero rates only above a per-antenna SNR threshold that is no less than* $21N_t - 1$ *when* $N_t < N_r$ *and* $\mathbb{E}[\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}] = \boldsymbol{I}_{N_t}$*, e.g., an SNR threshold of* 10 *dB for a* 1×2 *channel. Our results guarantee positive rates at all SNR; for the single-input multiple-output (SIMO) channel at low SNR the proposed scheme has a gap on the order of* ρ^2*. Since at* $\rho \ll 1$ *we have* $C \approx \rho \mathbb{E}\left[||\tilde{h}||^2\right] \log e$*, the proposed scheme can be said to asymptotically achieve capacity at low SNR. Our results also show the gap diminishes to zero with increasing the number of receive antennas under Rayleigh fading.*

3.1.2 SISO channel

When each node has one antenna, we find tighter bounds on the gap to capacity for a wider range of fading distributions. Without loss of generality let $\mathbb{E}[|\tilde{h}|^2] = 1$. The gap to capacity is given by

$$\Delta = \mathbb{E}\big[\log\left(1+\rho|\tilde{h}|^2\right)\big] + \log\left(\mathbb{E}[\frac{1}{1+\rho|\tilde{h}|^2}]\right). \tag{3.28}$$

In the following, we compute bounds on this gap for a wide range of fading distributions, at both high and low SNR values.

Corollary 2. *When $N_t = N_r = 1$, the gap to capacity Δ is upper bounded as follows*

- $\rho < 1$*: For any fading distribution where* $\mathbb{E}\big[|\tilde{h}|^4\big] < \infty$*,*

$$\Delta < 1.45\,\mathbb{E}\big[|\tilde{h}|^4\big]\,\rho^2. \tag{3.29}$$

- $\rho \geq 1$*: For any fading distribution where* $\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big] < \infty$*,*

$$\Delta < 1 + \log\Big(\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big]\Big). \tag{3.30}$$

- $\rho \geq 1$*: Under Nakagami-m fading with* $m > 1$*,*

$$\Delta < 1 + \log\big(1 + \frac{1}{m-1}\big). \tag{3.31}$$

- $\rho \geq 1$*: Under Rayleigh fading,*

$$\Delta < 0.48 + \log\big(\log(1+\rho)\big). \tag{3.32}$$

Proof. See Appendix A.4. □

Although the gap depends on the SNR under Rayleigh fading, Δ is a vanishing fraction of the capacity as ρ increases, i.e., $\lim_{\rho\to\infty} \frac{\Delta}{C} = 0$. Simulations are provided to give a better view of Corollary 2. First, the rate achieved under Nakagami-m fading with $m = 2$ and the corresponding gap to capacity are plotted in Figure 3.3. The performance is compared with that of the division algebra lattices from [13]. Similar results are also provided under Rayleigh fading in Figure 3.4.

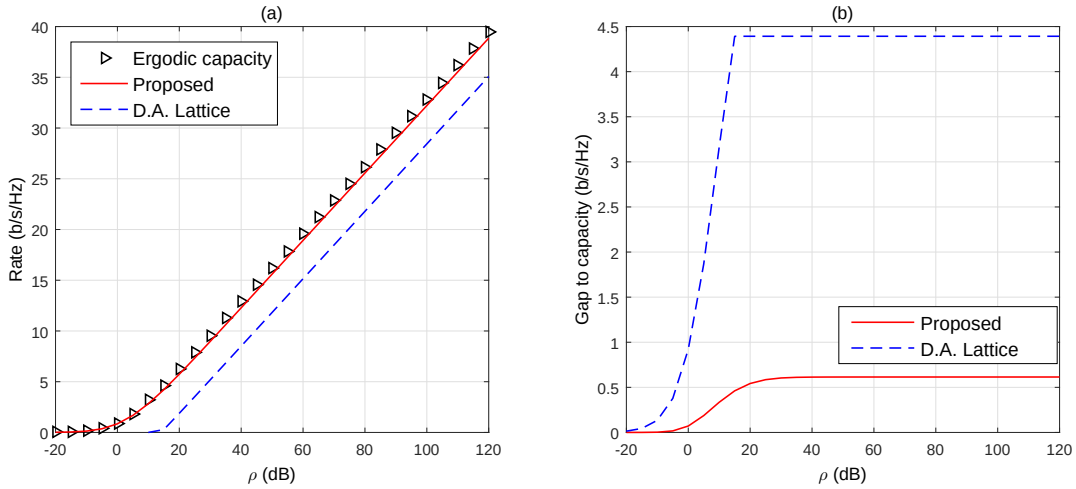

Figure 3.3. (a) Rates achieved vs. division algebra lattices for SISO Nakagami-m fading channels with $m = 2$. (b) Comparison of the gap to capacity.

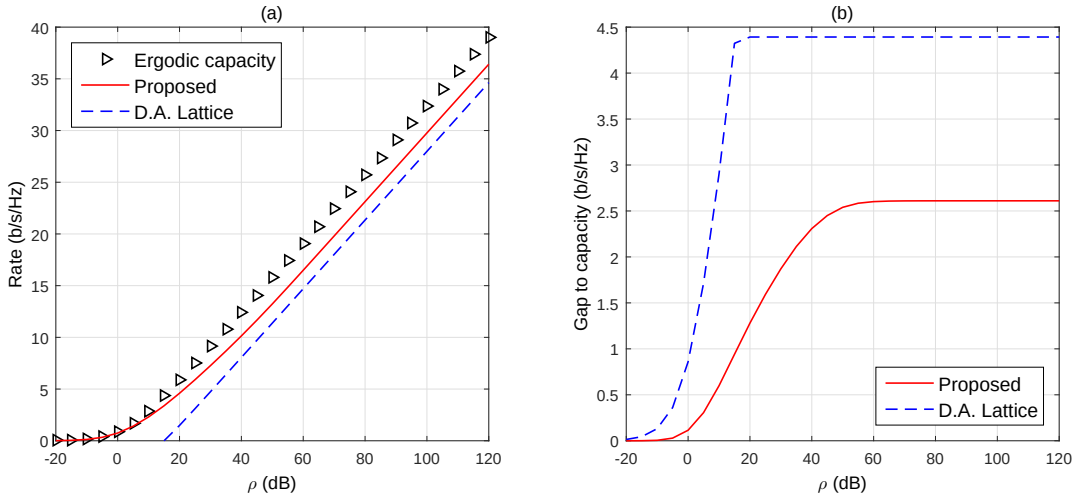

Figure 3.4. (a) Rates achieved vs. division algebra lattices for SISO Rayleigh fading channels. (b) Comparison of the gap to capacity.

Remark 2. *A closely related problem appears in [7], where lattice coding and decoding were studied under quasi-static fading for MIMO channels with CSIR, and a realization of the class of construction-A lattices in conjunction with channel-matching decision regions (ellipsoidal shaped) was proposed. Unfortunately, this result by itself does not apply to ergodic fading because the application of the Minkowski-Hlawka Theorem [3, Theorem 1], on which the existence argument of [7] depend,s only guarantees the existence of a lattice for each channel realization, and is silent about the existence of a* universal *single lattice that is suitable for all channel realizations. This universality requirement is the key challenge under ergodic fading. The essence of the proposed lattice scheme in this section is approximating the ergodic fading channel (subsequent to MMSE equalization) with a non-fading additive-noise channel with lower SNR* $\rho' \triangleq \alpha\rho$*, where* $\alpha \leq 1$*. The distribution of the (equivalent) additive noise term,* $\boldsymbol{z}$*, in the approximate model depends on the fading distribution but not on the realization, which allows fixed decision regions for all fading realizations. The* SNR penalty factor α *incurred from this approximation for the special case of* $N_t = N_r = 1$ *is given by*

$$\alpha = \mathbb{E}\big[\frac{|\tilde{h}|^2}{\rho|\tilde{h}|^2+1}\big] \,/\, \mathbb{E}[\frac{1}{\rho|\tilde{h}|^2+1}]. \tag{3.33}$$

As shown in the gap analysis in this section, the loss caused by this approximation is small under most settings.

3.1.3 MIMO Channel with Block Fading[4]

Although the lattice scheme presented in the previous section approaches the ergodic capacity under several fading scenarios, one still hopes to tighten the gap to capacity, as well as extend the results to non-isotropic fading. In this section we propose an alternative decoding approach for the lattice coding scheme that approaches the capacity under numerous scenarios, including

[4]© 2017 IEEE. Reprinted, with permission, from A. Hindy and A. Nosratinia, On the Universality of Lattice Codes for a Class of Ergodic Fading Channels, in IEEE International Symposium on Information Theory (ISIT), June 2017

MIMO channels with non-isotropic fading. For ease of exposition we limit our study to real-valued point-to-point channels under block fading, i.e., The channel coherence length is b with n' independent fading blocks, such that $n = n'b$. The *temporal* covariance of each antenna pair link is then given by

$$\boldsymbol{\Sigma_h} = \sigma_h^2 \begin{bmatrix} \mathbf{1_b} & \mathbf{0_b} & \dots & \mathbf{0_b} \\ \mathbf{0_b} & \mathbf{1_b} & \dots & \mathbf{0_b} \\ & & \ddots & \\ \mathbf{0_b} & \mathbf{0_b} & \dots & \mathbf{1_b} \end{bmatrix}. \tag{3.34}$$

Before we proceed, we present a more general form of the capacity expression in (3.24) for non-isotropic real-valued channels, given by [26]

$$C = \frac{1}{2} \max_{\mathrm{tr}(\boldsymbol{K_x}) \leq N_t \rho} \mathbb{E}_H \big[\log \det(\boldsymbol{I}_{N_r} + \boldsymbol{H} \boldsymbol{K_x} \boldsymbol{H}^T) \big], \tag{3.35}$$

where $\boldsymbol{K_x}$ is the covariance matrix of each super-symbol $\boldsymbol{x}_i$.

Lemma 6. *Consider a MIMO channel $\boldsymbol{y} = \boldsymbol{H}_s \boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{H}_s \triangleq diag\big(\boldsymbol{H}_1, \dots, \boldsymbol{H}_n\big)$, and $\boldsymbol{H}_1, \dots, \boldsymbol{H}_n$ are realizations of a stationary and ergodic process, and is only known at the receiver. Then there exists at least one lattice codebook that achieves all rates satisfying*

$$R < \frac{1}{2} \mathbb{E} \big[\log \det(\boldsymbol{I}_{N_t} + \rho \boldsymbol{H}^T \boldsymbol{H}) \big]. \tag{3.36}$$

Proof. The encoding process is identical to that in (3.2). The received signal is also multiplied by the MMSE equalization matrix $\boldsymbol{U}_i$ and the dither is removed as shown in (3.5),(3.7).

Following these steps a different version of the ambiguity decoder is applied, defined by an ellipsoidal decision region $\Omega \in \mathbb{R}^{N_t n}$,

$$\Omega \triangleq \Big\{ \boldsymbol{v} \in \mathbb{R}^{N_t n} \ : \ \boldsymbol{v}^T \boldsymbol{\Sigma}_s^{-1} \boldsymbol{v} \leq (1+\gamma) N_t n \Big\}, \tag{3.37}$$

where $\boldsymbol{\Sigma}_s$ is a block-diagonal matrix, whose diagonal block i, $\boldsymbol{\Sigma}_i$, is given by

$$\boldsymbol{\Sigma}_i \triangleq \rho \big(\boldsymbol{I}_{N_t} + \rho \boldsymbol{H}_i^T \boldsymbol{H}_i \big)^{-1}. \tag{3.38}$$

The volume of Ω is then

$$\mathrm{Vol}(\Omega) = (1+\gamma)^{\frac{N_t n}{2}} \mathrm{Vol}\big(\mathcal{B}_{N_t n}(\sqrt{N_t n \rho})\big) \prod_{i=1}^{n} \det(\boldsymbol{\Psi}_i)^{\frac{-1}{2}}, \tag{3.39}$$

where $\boldsymbol{\Psi}_i \triangleq \boldsymbol{I}_{N_t} + \rho \boldsymbol{H}_i^T \boldsymbol{H}_i$.

Error Probability: Similar to the proof of Theorem 1, on averaging over the set of all fine lattices $\mathcal{C}$ of rate R, the probability of error can be bounded by

$$\begin{aligned} \frac{1}{|\mathcal{C}|} \sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e &< \mathbb{P}(\boldsymbol{z} \notin \Omega) + (1+\delta) \frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V}_1)} \\ &= \mathbb{P}(\boldsymbol{z} \notin \Omega) + (1+\delta) 2^{nR} \frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \end{aligned} \tag{3.40}$$

for any $\delta > 0$.

The following lemma addresses the error event represented by the first term.

Lemma 7. *For any $\epsilon' > 0$, there exists $n_{\epsilon'}$ such that for all $n > n_{\epsilon'}$, $P(\boldsymbol{z} \notin \Omega) < \epsilon'$.*

Proof. See Appendix A.5. □

Consequently, the error probability can be bounded by

$$\epsilon'' \triangleq \frac{1}{|\mathcal{C}|} \sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e < \epsilon' + (1+\delta) 2^{nR} \frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.41}$$

for any $\gamma, \delta > 0$. For convenience define $\boldsymbol{\Psi}_i = \rho \boldsymbol{\Sigma}_i^{-1}$.

The second term in (3.41) is then

$$\epsilon_{\text{avg}} \triangleq 2^{-n\left(-R + \frac{1}{2n}\sum_{i=1}^{n} \log\det(\boldsymbol{\Psi}_i) - \epsilon'''\right)}, \tag{3.42}$$

where

$$\epsilon''' \triangleq \frac{1}{n} \log\big(\frac{\mathrm{Vol}(\mathcal{B}_{N_t n}(\sqrt{N_t n \rho}))}{\mathrm{Vol}(\mathcal{V})}\big) + \log(1+\gamma)^{\frac{N_t}{2}} + \frac{1}{n}\log(1+\delta) \tag{3.43}$$

From (2.8), the first term in (3.43) vanishes, and so do the second and third terms as n increases. Since the probability of error averaged over the codebooks in $\mathcal{C}$ is bounded by

$$\epsilon'' \triangleq \epsilon' + \epsilon_{\text{avg}}, \tag{3.44}$$

there exists at least one codebook that achieves $R < \frac{1}{2n}\sum_{i=1}^{n}\log\det(\mathbf{\Psi}_i)$, which converges to (3.36). The existence of a nested lattice that achieves the average performance follows from the proof of Theorem 1. Following in the footsteps of [7], it can be shown that the ellipsoidal decision region in (3.37) achieves the same (or higher) error probability as the *Euclidean lattice decoder*, given a fixed transmission rate. The Euclidean lattice decoder is given by

$$\hat{\boldsymbol{t}} = \left[\arg\min_{\boldsymbol{t}'\in\Lambda_1} ||\Sigma^{\frac{-1}{2}}(\boldsymbol{y}' - \boldsymbol{t}')||^2\right] \mathrm{mod}\Lambda\,. \tag{3.45}$$

This concludes the proof of Lemma 6. □

Note that Lemma 6 does not imply the rate in (3.36) is *universally* achievable, since it does not guarantee the existence of a single codebook that achieves this rate. The following theorem establishes the rates achieved using a universal codebook.

Theorem 3. *For a stationary and ergodic block-fading $N_t \times N_r$ MIMO channel with coherence interval b whose fading coefficients are drawn from a discrete distribution with finite support $|\mathcal{H}|$, lattice codes achieve rates within a constant gap $\Delta \triangleq \frac{N_t N_r}{b}\log|\mathcal{H}|$ bits per channel use of the ergodic capacity.*

Proof. Lemma 6 assures the existence of one codebook in $\mathcal{C}$ that achieves the rate in (3.36) with error probability that is less than ϵ''. We now show that if we allow a *multiplicative* increase in the error probability, numerous codebooks in $\mathcal{C}$ can support the rate R in (3.36).

Lemma 8. *For the channel in Lemma 6, at least $\frac{\kappa-1}{\kappa}|\mathcal{C}|$ codebooks in $\mathcal{C}$ achieve the rate R in* (3.36) *with at most $\kappa\epsilon''$ error probability, for any $\kappa \in \mathbb{Z}^+$ where $\kappa < |\mathcal{C}|$.*

Proof. We expurgate codebooks from $\mathcal{C}$ as follows. First, arrange the codebooks in *descending order* of the error probability (with respect to the channel defined in Lemma 6). Then, throw away the first $\frac{1}{\kappa}|\mathcal{C}|$ codebooks. For the $\frac{\kappa-1}{\kappa}|\mathcal{C}|$ remaining codebooks, the error probability of *each*

codebook is then bounded by $\epsilon' + \kappa\epsilon_{\text{avg}}$, from (3.44). This can be explained as follows[5]

$$\begin{aligned}
|\mathcal{C}|\epsilon_{\text{avg}} &= \sum_{\ell=1}^{|\mathcal{C}|/\kappa} \epsilon_\ell \; + \epsilon_{1+|\mathcal{C}|/\kappa} \; + \sum_{\ell=2+|\mathcal{C}|/\kappa}^{|\mathcal{C}|} \epsilon_\ell \\
&\geq \sum_{\ell=1}^{|\mathcal{C}|/\kappa} \epsilon_\ell + \; \epsilon_{1+|\mathcal{C}|/\kappa} \\
&\geq (1 + \frac{1}{\kappa}|\mathcal{C}|)\, \epsilon_{1+|\mathcal{C}|/\kappa}\,, \qquad (3.46)
\end{aligned}$$

where ϵ_ℓ is the error probability incurred by codebook $\mathcal{C}_\ell$. Hence,

$$\epsilon_{1+|\mathcal{C}|/\kappa} \leq \frac{\kappa}{1+\frac{\kappa}{|\mathcal{C}|}}\, \epsilon_{\text{avg}} \; < \; \kappa\, \epsilon_{\text{avg}}. \qquad (3.47)$$

Since $\epsilon_{l+|\mathcal{C}|/\kappa} \leq \epsilon_{1+|\mathcal{C}|/\kappa}$ for any $\ell > 1$, each of the last $\frac{\kappa-1}{\kappa}\,|\mathcal{C}|$ codebooks in $\mathcal{C}$ have error probability that does not exceed $\epsilon' + \kappa\epsilon < \kappa\epsilon''$. □

To summarize, Lemma 8 shows that given a channel matrix $\boldsymbol{H}_s$, a constant fraction of all codebooks in $\mathcal{C}$ achieves the rate in (3.36). For example, if $\kappa = 100$, then at least 99% of the codebooks in $\mathcal{C}$ incur no more than $100\epsilon''$ error probability, where ϵ'' can be made arbitrarily small by increasing n. Note that the proof technique in Lemma 8 is not limited to lattice codes, and can be used whenever error calculations are averaged over an ensemble of codebooks.

Now, assume we have a block-fading MIMO channel with n' independent blocks, where each fading element is drawn from a discrete distribution with finite support of size $|\mathcal{H}|$. The number of all possible matrices $\boldsymbol{H}_s$ is then $|\mathcal{H}|^{Mn'}$, where $M \triangleq N_t N_r$. We aim at answering the following question: under what rates can a single codebook in $\mathcal{C}$ achieve vanishing error probability for all possible channel matrices $\boldsymbol{H}_s$?

With a slight abuse of notation, denote by $\mathcal{C}_j$ the set of codebooks that achieve at most $\kappa\epsilon''$ error probability for the channel matrix $\boldsymbol{H}_s^{(j)}$ indexed by j. Recall the cardinality of each of

[5]Recall that the first term in (3.40) is independent of the codebook. Hence, it remains constant for all codebooks.

these sets is $\frac{\kappa-1}{\kappa}|\mathcal{C}|$. The event that no codebook is universal over all possible fading vectors can be represented by either of the following equivalent set equalities

$$\begin{aligned}\mathcal{C}_1 \cap \mathcal{C}_2 \cap \ldots \cap \mathcal{C}_{|\boldsymbol{H}_s|^{Mn'}} &= \phi, \\ \mathcal{C}_1^c \cup \mathcal{C}_2^c \cup \ldots \cup \mathcal{C}_{|\boldsymbol{H}_s|^{Mn'}}^c &= \mathcal{C},\end{aligned} \tag{3.48}$$

where $\mathcal{A}^c$ denotes the complement of the set $\mathcal{A}$. Hence, it can be shown via a union bound that

$$|\mathcal{C}_1^c \cup \ldots \cup \mathcal{C}_{|\boldsymbol{H}_s|^{Mn'}}^c| \leq \frac{1}{\kappa}|\mathcal{C}|\,|\mathcal{H}|^{Mn'}. \tag{3.49}$$

Hence, from (3.48) a universal codebook with negligible error probability is guaranteed to exist, as long as $\kappa > |\mathcal{H}|^{Mn'}$. On substituting in (3.41), (3.42),

$$\begin{aligned}\mathbb{P}_e &< \epsilon' + \kappa\epsilon \\ &< \epsilon' + 2^{-n\left(-R+\frac{1}{2}\mathbb{E}\left[\log\det(\boldsymbol{I}_{N_t}+\rho\boldsymbol{H}^T\boldsymbol{H})\right]-\frac{Mn'}{n}\log|\mathcal{H}|-\epsilon'''\right)},\end{aligned} \tag{3.50}$$

and reliable rates can be achieved as long as

$$R < \frac{1}{2}\mathbb{E}\big[\log\det(\boldsymbol{I}_{N_t}+\rho\boldsymbol{H}^T\boldsymbol{H})\big] - \frac{M}{b}\log|\mathcal{H}| - \epsilon''', \tag{3.51}$$

where ϵ''' can be made arbitrarily small by increasing n. Since the second term in (3.50) decreases with n, there exists $n_\epsilon \in \mathbb{Z}^+$ such that for all $n > n_\epsilon$, $\mathbb{P}_e$ in (3.50) satisfies $\mathbb{P}_e < 2\epsilon'$. The final step to complete the proof is showing that the number of possible channel matrices $\boldsymbol{H}_s$ does not exhaust $|\mathcal{C}|$, otherwise $\kappa > |\mathcal{H}|^{Mn'}$ cannot be guaranteed. From the lattice construction in [6, Section III], there exists at least q^n generator matrices that generate unique lattices, where q is the size of the prime field from which the lattice is drawn. Since $\sqrt{n}/q \to 0$ as $n \to \infty$, a lower bound on $|\mathcal{C}|$ is $n^{n/2}$. Since the number of possible channels cannot exceed $|\mathcal{H}|^{Mn}$ where $|\mathcal{H}|^M$ is fixed, there exists $\check{n} \in \mathbb{Z}^+$ such that for all $n > \check{n}$, $n^{n/2} > |\mathcal{H}|^{Mn}$. Hence, $\kappa > |\mathcal{H}|^{Mn'}$ is guaranteed for large enough n.

The previous result concludes the gap to capacity for channels with *isotropic fading*, whose optimal input signal covariance is $\rho\boldsymbol{I}_{N_t}$. The extension to general channel distributions is

straightforward. Let $\boldsymbol{K_x}^*$ denote the optimal input covariance matrix, i.e., $\boldsymbol{K_x}^* = \arg\max\ C$ given in (3.35). The transmitted codeword is then

$$\check{\boldsymbol{x}} \triangleq [\boldsymbol{K_x}^{*\frac{1}{2}}\boldsymbol{x}_1^T, \ldots, \boldsymbol{K_x}^{*\frac{1}{2}}\boldsymbol{x}_n^T]^T, \tag{3.52}$$

where $\boldsymbol{x}$ is drawn from a nested lattice code whose coarse lattice $\Lambda \in \mathbb{R}^{N_t n}$ has a unit second moment. Hence, the received signal can be expressed by

$$\begin{aligned}\boldsymbol{y}_i &= \boldsymbol{H}_i \boldsymbol{K_x}^{*\frac{1}{2}} \boldsymbol{x}_i + \boldsymbol{w}_i \\ &\triangleq \check{\boldsymbol{H}}_i \boldsymbol{x}_i + \boldsymbol{w}_i,\end{aligned} \tag{3.53}$$

where $\check{\boldsymbol{H}}_i \triangleq \boldsymbol{H}_i \boldsymbol{K_x}^{*\frac{1}{2}}$. Given the equivalent channel in (3.53) the following rates can be shown to be achievable

$$\begin{aligned}R &< \frac{1}{2}\mathbb{E}\Big[\log\det\left(\boldsymbol{I}_{N_t} + \check{\boldsymbol{H}}^T\check{\boldsymbol{H}}\right)\Big] - \frac{M}{b}\log|\mathcal{H}| \\ &= \frac{1}{2}\mathbb{E}\Big[\log\det\left(\boldsymbol{I}_{N_r} + \check{\boldsymbol{H}}\check{\boldsymbol{H}}^T\right)\Big] - \frac{M}{b}\log|\mathcal{H}| \\ &= \frac{1}{2}\mathbb{E}\Big[\log\det\left(\boldsymbol{I}_{N_r} + \boldsymbol{H}\boldsymbol{K_x}^*\boldsymbol{H}^T\right)\Big] - \frac{M}{b}\log|\mathcal{H}|,\end{aligned}$$

which is the optimal value of the expression in (3.35). This concludes the proof of Theorem 3. □

One can regard the gap Δ is a rate penalty that is incurred by the universality constraint. Numerical examples of Δ in bits per channel use are as follows,

- Case 1: For the binary fast fading channel with $N_r = N_t = 1$, $b = 1$ and $\mathcal{H} = \{0, a\}$ for $a \neq 0$, $\Delta < 1$.

- Case 2: For a 2×2 MIMO block-fading channel with $b = 20$ and $|\mathcal{H}| \leq 10^3$, $\Delta < 2$.

- Case 3: For any $N_t \times N_r$ MIMO block-fading channel with $n' = \mathcal{O}(n^{1-\delta})$ and $0 < \delta < 1$, $\Delta \to 0$ as $n \to \infty$.

Note that in Case 1 the single-antenna lattice scheme in Section 3.1.1 achieves no more than 1 bit per channel use,[6] whereas the gap to capacity of the MIMO scheme in [13] is more than 4 and 10 bits per channel use for Case 1 and Case 2 respectively.[7] Interestingly, the mod-Λ lattice scheme achieves the ergodic capacity in Case 3, which is analogous to the case where the codeword transmission is *interleaved* over b channel realizations such that both b and n' are sufficiently large [53].

3.2 MIMO Channel with CSIT[8]

For the band-limited Additive White Gaussian Noise (AWGN) channel, approaching capacity with manageable complexity has been extensively studied [54, 55, 56, 57, 58, 59, 60, 61]. Under fading and in the presence of CSIT, the most straight forward capacity approaching schemes employ separable coding, i.e., coding independently and in parallel over different fading states of the channel [25, 32]. Unfortunately, separable coding causes significant delay and requires large memory at both the transmitter and receiver. Also, separable encoding in practice forces a low-probability fading state to either cause huge delays, small block-length effects, or the symbols in that fading state to be ignored with associated rate loss. As a result, achieving the ergodic capacity of block-fading channels *without* separable coding (i.e., with coding across states) remains an important theoretical and practical question.[9] This section shows that non-separable lattice coding achieves the ergodic capacity of the block fading SISO channel. At the transmitter, the symbols of the codeword are permuted across time using a linear permutation matrix.

[6]The fading distribution in Case 1 is not regular, and hence a constant gap is not guaranteed.

[7]In Case 2, for a channel with $\mathbb{E}[\boldsymbol{H}^T\boldsymbol{H}] = \boldsymbol{I}_{N_t}$, the minimum power (per antenna) required to achieve positive rates is 0 dB using the proposed scheme, compared to 16 dB for the scheme in [13].

[8]

[9]It was pointed out in [31] that under maximum likelihood decoding the ergodic capacity of point-to-point channels with CSIT can be attained using Gaussian signaling *without* separable coding. However, the same result is not necessarily true for non-Gaussian (structured) codebooks.

Time-varying Minimum Mean-Square Error (MMSE) scaling is used at the receiver, followed by a decoder that is universal for all fading realizations drawn from a given fading distribution. Hence, the codebook design and decision regions are fixed across different transmissions; the only channel-dependent blocks are the permutation and MMSE scaling functions.

We first highlight the main ideas of the proposed scheme in the context of a heuristic channel model whose behavior approximates the ergodic fading channel. We then generalize the solution to all fading distributions whose realizations are robustly typical, and to continuous distributions via a bounding argument. The results are then extended to MIMO block-fading channels.

First, we start with the SISO case. Consider a real-valued single antenna point-to-point channel with block-fading and i.i.d. Gaussian noise. The received signal is given by $y_i = h_i x_i + w_i$, where communication occurs over n channel uses. For convenience, we represent the entire received symbols in a single equation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w}, \tag{3.54}$$

where $\boldsymbol{H}$ is an $n \times n$ diagonal matrix whose diagonal entries are h_i, which are drawn from a discrete distribution of finite support $\mathcal{H}$ similar to the channel model in Section 3.1.3, with channel coherence length is b with n' independent fading blocks, such that $n = n'b$. Both the transmitter and receiver have full knowledge of the channel state information. The noise $\boldsymbol{w} \in \mathbb{R}^n$ is a zero-mean i.i.d. Gaussian vector with covariance $\boldsymbol{I}_n$ and is independent of $\boldsymbol{H}$. The vector $\boldsymbol{x} \in \mathbb{R}^n$ is the transmitted codeword satisfying an average power constraint $\mathbb{E}\big[||\boldsymbol{x}||^2\big] \leq n\rho$.

Under CSIT, the ergodic capacity of the real-valued point-to-point channel is given by [25]

$$C = \frac{1}{2}\mathbb{E}_h\Big[\log\big(1 + h^2\rho^*(h)\big)\Big], \tag{3.55}$$

where $\rho^*(h)$ denotes the channel-dependent *waterfilling* power allocation [25], which satisfies $\mathbb{E}_h\big[\rho^*(h)\big] = \rho$.[10] To highlight the essential ideas of the proposed technique, we first address the problem over a heuristic channel model, and then extend to arbitrary channel distributions under SISO and MIMO.

[10]Waterfilling power allocation is provided in more detail in Section 3.2.4.

3.2.1 The Random Location Channel

Assume a heuristic channel model, whose channel coefficients h_i take on values from set $\mathcal{H} \triangleq \{\mathfrak{h}_1, \ldots, \mathfrak{h}_{|\mathcal{H}|}\}$, arranged in ascending order of the magnitudes. Let $\boldsymbol{h} \triangleq [h_1, \ldots, h_n]$ denote a realization of the channel coefficient values incurred by an arbitrary codebook transmission. The transmitter knows *non-causally* a shuffled version of $\boldsymbol{h}$, i.e., the transmitter knows that $\mathfrak{h}_k$ occurs exactly n_k times, where $\sum_{k=1}^{|\mathcal{H}|} n_k = n$, however, the locations of these coefficients are only known *causally* at the transmitter and receiver. This channel model approximates the stationary ergodic fading process so as to simplify the counting of the number of occurrences of each fading state, and is denoted *the random location channel*. We acknowledge that in this model, the n fading coefficients are not fully independent; in that sense this model approximates, but is not precisely the same as, the usual ergodic fading model.

Theorem 4. *For the random location channel defined above, the rates*

$$R < \frac{1}{2n} \sum_{i=1}^{n} \log\left(1 + h_i^2 \rho^*(h_i)\right) \tag{3.56}$$

are achievable using lattice codes, where $\rho^*(h)$ represents the waterfilling power allocation strategy for the channel drawn from $\mathcal{H}$.

Proof. Encoding: Nested lattice codes are used where $\Lambda \subseteq \Lambda_1$. The transmitter emits a lattice point $\boldsymbol{t} \in \Lambda_1$ that is dithered with $\boldsymbol{d}$ which is drawn uniformly over $\mathcal{V}$. The dithered codeword is as follows

$$\begin{aligned} \boldsymbol{x} =& \left[\boldsymbol{t} - \boldsymbol{d}\right] \bmod \Lambda \\ =& \boldsymbol{t} - \boldsymbol{d} + \boldsymbol{\lambda}, \end{aligned} \tag{3.57}$$

where $\boldsymbol{\lambda} = -Q_{\mathcal{V}}(\boldsymbol{t} - \boldsymbol{d}) \in \Lambda$ from (2.5). The coarse lattice $\Lambda \in \mathbb{R}^n$ has a second moment ρ. The codeword is then multiplied by two cascaded matrices as follows

$$\boldsymbol{x}' = \boldsymbol{D}\,\boldsymbol{V}\,\boldsymbol{x}, \tag{3.58}$$

where $\boldsymbol{V}$ is a *permutation matrix* and $\boldsymbol{D}$ is a diagonal matrix with $D_{ii} = \sqrt{\rho^*(h_i)/\rho}$, where $\rho^*(h_i)$ is the optimal waterfilling power allocation for the fading coefficient $|h_i|$, as given in [25]. As a short-hand, we rewrite $\rho^*(h_i)$ as ρ_i^*. Note that $\mathbb{E}[\sum_{i=1}^n D_{ii}^2] = 1$. Hence, $\frac{1}{n}\mathbb{E}[||\boldsymbol{x}'||^2] < \rho$ is also guaranteed.

Decoding: The received signal $\boldsymbol{y}$ is multiplied by a receiver matrix $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ cascaded with an inverse permutation $\boldsymbol{V}^T$ and the dither is removed as follows

$$\begin{aligned}\boldsymbol{y}' =& \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{y} + \boldsymbol{d}\\ =& \boldsymbol{x} + (\boldsymbol{V}^T\boldsymbol{U}\boldsymbol{H}\boldsymbol{D}\boldsymbol{V} - \boldsymbol{I}_n)\boldsymbol{x} + \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{w} + \boldsymbol{d}\\ =& \boldsymbol{t} + \boldsymbol{\lambda} + (\boldsymbol{V}^T\boldsymbol{U}\boldsymbol{H}\boldsymbol{D}\boldsymbol{V} - \boldsymbol{I}_n)\boldsymbol{x} + \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{w},\\ =& \boldsymbol{t} + \boldsymbol{\lambda} + \boldsymbol{z},\end{aligned} \tag{3.59}$$

where

$$\boldsymbol{z} \triangleq (\boldsymbol{V}^T\boldsymbol{U}\boldsymbol{H}\boldsymbol{D}\boldsymbol{V} - \boldsymbol{I}_n)\boldsymbol{x} + \boldsymbol{V}^T\boldsymbol{U}\boldsymbol{w}, \tag{3.60}$$

and $\boldsymbol{z}$ is independent of $\boldsymbol{t}$ from Lemma 1.

The receiver matrix $\boldsymbol{U}$ is chosen to be the MMSE matrix given by

$$\boldsymbol{U} = \rho\boldsymbol{H}\boldsymbol{D}(\boldsymbol{I}_n + \rho\boldsymbol{H}^2\boldsymbol{D}^2)^{-1}. \tag{3.61}$$

$\boldsymbol{U}$ is diagonal, where $U_{ii} = \rho D_{ii}h_i/(1 + \rho D_{ii}^2 h_i^2)$. Now, the diagonal elements of $\boldsymbol{U}$ are

$$U_{ii} = \frac{\sqrt{\rho\rho_i^*}h_i}{1 + \rho_i^* h_i^2}. \tag{3.62}$$

With a slight abuse of notation, let $\boldsymbol{\pi}$ be a permutation function that orders channel realization such that $(h_{\boldsymbol{\pi}(1)}, h_{\boldsymbol{\pi}(2)}, \ldots, h_{\boldsymbol{\pi}(n)})$ represent the channel coefficients arranged in ascending order of the magnitudes. We pick the permutation matrix $\boldsymbol{V}$ such that $\boldsymbol{H}_{\boldsymbol{\pi}} \triangleq \boldsymbol{V}^T\boldsymbol{H}\boldsymbol{V}$, where the diagonal entries of $\boldsymbol{H}_{\boldsymbol{\pi}}$ are $h_{\boldsymbol{\pi}(i)}$. The structure of $\boldsymbol{V}$ is given in Appendix A.6. From (3.60) and (3.61), z_i are given by [11]

$$z_i = \frac{-1}{\rho_{\boldsymbol{\pi}(i)}^* h_{\boldsymbol{\pi}(i)}^2 + 1}x_i + \frac{\sqrt{\rho\rho_{\boldsymbol{\pi}(i)}^*}h_{\boldsymbol{\pi}(i)}}{\rho_{\boldsymbol{\pi}(i)}^* h_{\boldsymbol{\pi}(i)}^2 + 1}w_{\boldsymbol{\pi}(i)}. \tag{3.63}$$

[11] Since waterfilling dedicates more power to channels with larger magnitude, $h_i^2 \geq h_j^2$ implies $\rho_i^* h_i^2 \geq \rho_j^* h_j^2$ [25].

Now, we apply a version of the ambiguity decoder defined by an ellipsoidal decision region, given by

$$\Omega_1 \triangleq \{\boldsymbol{s} \in \mathbb{R}^n \ : \ \boldsymbol{s}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{s} \leq (1+\epsilon)n\}, \tag{3.64}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal elements are given by

$$\Sigma_{ii} = \frac{\rho}{\rho^*_{\boldsymbol{\pi}(i)} h^2_{\boldsymbol{\pi}(i)} + 1}. \tag{3.65}$$

The decoder chooses $\hat{\boldsymbol{t}} \in \Lambda_1$ if and only if the received point falls exclusively within the decision region of the lattice point $\hat{\boldsymbol{t}}$, i.e., $\boldsymbol{y}' \in \hat{\boldsymbol{t}} + \Omega_1$.

Probability of error: On averaging over the set of all fine lattices $\mathcal{C}$ of rate R, the probability of error can be bounded by

$$\frac{1}{|\mathcal{C}|} \sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e = \mathbb{P}(\boldsymbol{z} \notin \Omega_1) + (1+\delta) 2^{nR} \frac{\mathrm{Vol}(\Omega_1)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.66}$$

for any $\delta > 0$. Following in the footsteps in Appendix A.5, it can be shown that $\mathbb{P}(\boldsymbol{z} \notin \Omega_1) < \gamma$ for any $\gamma > 0$ for large n. Consequently, the error probability can be bounded by

$$\frac{1}{|\mathcal{C}|} \sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e < \gamma + (1+\delta) 2^{nR} \frac{\mathrm{Vol}(\Omega_1)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.67}$$

for any $\gamma, \delta > 0$. The volume of Ω_1 is given by

$$\mathrm{Vol}(\Omega_1) = (1+\epsilon)^{\frac{n}{2}} \mathrm{Vol}\big(\mathcal{B}(\sqrt{n\rho})\big) \Big(\prod_{i=1}^{n} \frac{1}{\rho^*_i h_i^2 + 1} \Big)^{\frac{1}{2}}. \tag{3.68}$$

The second term in (3.67) is then bounded by

$$\begin{aligned} &(1+\delta) 2^{nR} (1+\epsilon)^{n/2} \Big(\prod_{i=1}^{n} \frac{1}{\rho^*_i h_i^2 + 1} \Big)^{\frac{1}{2}} \frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n\rho}))}{\mathrm{Vol}(\mathcal{V})} \\ = (1+\delta) &2^{-n\Big(-\frac{1}{n} \log\left(\frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n\rho}))}{\mathrm{Vol}(\mathcal{V})} \right) + \xi \Big)}, \end{aligned} \tag{3.69}$$

where

$$\begin{aligned} \xi \triangleq & \frac{-1}{2} \log(1+\epsilon) - \frac{1}{2n} \log \Big(\prod_{i=1}^{n} \frac{1}{\rho^*_i h_i^2 + 1} \Big) - R \\ = & \frac{-1}{2} \log(1+\epsilon) + \frac{1}{2n} \sum_{i=1}^{n} \log(1 + \rho^*_i h_i^2) - R. \end{aligned} \tag{3.70}$$

From (2.8), since the lattice Λ is good for covering, the first term of the exponent in (3.69) vanishes. From (3.69), whenever ξ is a positive constant we have $\lim_{n\to\infty} \mathbb{P}_e = 0$. Hence, positive ξ can be achieved as long as

$$R < \frac{1}{2n}\sum_{i=1}^{n} \log\left(1 + h_i^2 \rho_i^*\right)) - \frac{1}{2}\log(1+\epsilon) - \epsilon', \tag{3.71}$$

where ϵ, ϵ' are positive numbers that can be made arbitrarily small by increasing n. The remainder of the proof follows that in Section 3.1.3. This concludes the proof of Theorem 4. □

3.2.2 Channel with Discrete Fading

Now, we are ready to address the ergodic fading channel whose channel coefficients are drawn from an arbitrary discrete distribution. Unlike the random location channel discussed earlier, the number of occurrences of α_k in any given block length is not fixed.

Theorem 5. *Lattice codes achieve the ergodic capacity of the fast fading channel with channel state information available at all nodes.*

Proof. The proof appears in Appendix A.7. In a nutshell, we follow a *best effort* approach in designing the permutation matrix $\boldsymbol{V}$. In order to account for the ordering errors, we use a fixed decision region $\tilde{\Omega}_1$ that is slightly larger than $\Omega_1^{(p)}$ (the decision region resulting from perfect channel ordering, which is non-realizable due to the causality of the channel knowledge). However, when the channel is robustly typical, the total number of ordering errors is negligible at large n, and hence the rate loss incurred by using larger decision regions vanishes. □

The extension of Theorem 5 to complex-valued channels is straightforward, using techniques similar to that in Theorem 2. The channel would then be ordered with respect to the magnitude, $|\tilde{h}_i|$.

3.2.3 Extension to Continuous-valued Fading

In order to extend the arguments to continuous-valued fading channels, we assume the channel is block-faded with independent fading, and that the fading distribution possesses a finite second moment. We note that in the case of full CSI, the information density contributed by each transmission is a strictly increasing function of the absolute value of the fading coefficient. First, let $\tilde{g} \triangleq |h|^2\rho_h/\rho$ denote the squared channel gain times the normalized waterfilling power allocation for such gain. Thus, we can partition the continuous values $\tilde{g}$ into $L+1$ sets $G_\ell \triangleq [g_{\ell-1}, g_\ell]$, where $g_0 \triangleq 0$ and $g_{L+1} = \infty$. For any sequence of channel gains $\tilde{g}$ drawn from a continuous distribution, we quantize $\tilde{g}$ to the lower limit of the bracket G_i into which it belongs, producing a discrete valued sequence. The independence of the continuous-valued fading realizations guarantee the independence of the discrete-valued counterparts, and hence robust typicality would still apply. We show that the rate R supported by such sequence is within a gap to capacity that can be bounded as follows

$$
\begin{aligned}
C - R =& \mathbb{E}[\log(1+\rho\tilde{g})] \; - \; \mathbb{E}[\log(1+\rho g)] \\
=& \; \mathbb{E}[\log(1+\frac{1+\rho\tilde{g}}{1+\rho g})|\tilde{g} \le g_L]\mathbb{P}(g \le g_L) + \mathbb{E}[\log(\frac{1+\rho\tilde{g}}{1+\rho g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \\
<& \max\left\{ \log(\frac{1+\rho g_\ell}{1+\rho g_{\ell-1}}) \right\}_{\ell=1}^{L}, + \mathbb{E}[\log(\frac{1+\rho\tilde{g}}{1+\rho g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \\
<& \max\left\{ \log(1+\frac{\rho(g_\ell - g_{\ell-1})}{1+\rho g_{\ell-1}}) \right\}_{\ell=1}^{L}, + \mathbb{E}[\log(1+\frac{\rho(\tilde{g} - g_L)}{1+\rho g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \\
<& \max\left\{ \log\left(1+\rho(g_\ell - g_{\ell-1})\right) \right\}_{\ell=1}^{L}, + \mathbb{E}[\log(1+\frac{\rho(\tilde{g} - g_L)}{1+\rho g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \qquad (3.72) \\
<& \gamma_1 + \mathbb{E}[\log(1+\frac{\tilde{g} - g_L}{g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \\
=& \gamma_1 + \mathbb{E}[\log(\frac{\tilde{g}}{g_L})|\tilde{g} > g_L]\mathbb{P}(\tilde{g} > g_L) \\
<& \gamma_1 + c\,(\frac{\mathbb{E}[\tilde{g}|\tilde{g} > g_L]}{g_L} - 1)\mathbb{P}(\tilde{g} > g_L) \qquad (3.73) \\
<& \gamma_1 + c\,(\frac{\mathbb{E}[\tilde{g}]}{g_L\mathbb{P}(\tilde{g} > g_L)} - 1)\mathbb{P}(\tilde{g} > g_L) \qquad (3.74) \\
<& \gamma_1 + \frac{c\,\mathbb{E}[\tilde{g}]}{g_L} \; \triangleq \; \gamma_1 + \gamma_2\,, \qquad (3.75)
\end{aligned}
$$

where $c \triangleq \log e$, and $\gamma_1 \triangleq \max\left\{\log\left(1+\rho(g_\ell - g_{\ell-1})\right)\right\}_{\ell=1}^{L}$. (3.73) follows since $\log_e(x) < x-1$ for all $x > 0$ and (3.74) follows from the law of total expectation. γ_1 vanishes when $\max\{g_\ell - g_{\ell-1}\}_{i=1}^{L} \ll \frac{1}{\rho}$, while γ_2 vanishes when $g_L \gg \mathbb{E}[\tilde{g}]$. Note that a necessary condition for γ_2 to vanish is that the second moment of $\tilde{g}$ is finite.

Note that the gap can be made tighter when the distribution of $\tilde{g}$ has a vanishing tail. For instance, when $\tilde{g}$ is exponential,

$$\begin{aligned} C - R &< \gamma_1 + c\,(\frac{\mathbb{E}[\tilde{g}|\tilde{g} > g_L]}{g_L} - 1)\mathbb{P}(\tilde{g} > g_L) \\ &< \gamma_1 + c\,(\frac{\mathbb{E}[\tilde{g} + g_L]}{g_L} - 1)\mathbb{P}(\tilde{g} > g_L) \qquad (3.76) \\ &< \gamma_1 + \frac{c\,\mathbb{E}[\tilde{g}]}{g_L}\, e^{-\frac{g_L}{\mathbb{E}[\tilde{g}]}}, \qquad (3.77) \end{aligned}$$

which vanishes exponentially with g_L. Note that (3.76) follows since $\tilde{g}$ is exponentially distributed and hence memoryless.

The argument for the bounding above can be summarized is as follows. There are $L+1$ quantization bins, we bound the total rate loss due to quantization by the rate loss in each of the bins. The first L terms bound the amount of loss in rate by the input-output information density at the highest versus the lowest channel gain in each bracket $G_1, \ldots, G_L$. This strategy will not work for the final bin because the channel gain in G_{L+1} is unbounded, therefore instead we use the total rate contributed by the bin G_{L+1} as a bound. Fortunately, this term also vanishes at large g_L since the probability of occurrence of such fading values is small enough.

3.2.4 Extension to MIMO Channels

The result in Theorem 5 can be extended to an $N_t \times N_r$ MIMO channel with full CSIT. The received signal at time instant i is given by

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x}_i + \boldsymbol{w}_i, \qquad (3.78)$$

where we denote by $\boldsymbol{H}_i \in \mathbb{R}^{N_r \times N_t}$ the channel-coefficient matrix at time i, with a slight abuse of notation.

Theorem 6. *Lattice codes achieve the ergodic capacity of the block fading channel in* (3.78) *with channel state information available at all nodes.*

Proof. The proof can be summarized as follows. Since $\boldsymbol{H}_i$ is known perfectly, the transmitter and receiver can transform the MIMO channel into $\mathcal{S} \triangleq \min\{N_t, N_r\}$ SISO parallel channels via *singular value decomposition*, whose individual capacities can be achieved as shown in Section 3.2.2. Let the singular value decomposition of $\boldsymbol{H}_i$ be $\boldsymbol{H}_i = \boldsymbol{B}_i \boldsymbol{L}_i \boldsymbol{F}_i^T$, where $\boldsymbol{B}_i \in \mathbb{R}^{N_r \times N_r}$, $\boldsymbol{F}_i \in \mathbb{R}^{N_t \times N_t}$ are orthonormal matrices representing the left and right eigenvalue matrices of $\boldsymbol{H}_i$, respectively. $\boldsymbol{L}_i$ is an $N_r \times N_t$ rectangular diagonal matrix with $\mathcal{S}$ non-zero values on the main diagonal. Hence, at the receiver, the received signal is spatially equalized as follows

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{B}_i^T \boldsymbol{y}_i, \tag{3.79}$$

and at the transmitter, the signal is spatially precoded such that

$$\boldsymbol{x}_i = \boldsymbol{F}_i \tilde{\boldsymbol{x}}_i. \tag{3.80}$$

From (3.78),(3.80),(3.79), $\tilde{\boldsymbol{y}}_i$ can be represented by

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{L}_i \tilde{\boldsymbol{x}}_i + \tilde{\boldsymbol{w}}_i, \tag{3.81}$$

where $\tilde{\boldsymbol{w}}_i \triangleq \boldsymbol{B}_i^T \boldsymbol{w}_i \in \mathbb{R}^{N_r}$ is i.i.d. Gaussian, since $\boldsymbol{B}_i$ is orthonormal. Each element in $\tilde{\boldsymbol{y}}_i \in \mathbb{R}^{N_r}$ is then

$$\tilde{y}_i^{(s)} = \ell_i^{(s)} \tilde{x}_i^{(s)} + \tilde{w}_i^{(s)}, \qquad s = 1, \ldots, \mathcal{S}, \tag{3.82}$$

where $\ell_i^{(1)}, \ell_i^{(2)}, \ldots, \ell_i^{(\mathcal{S})}$ represent the singular values of $\boldsymbol{H}_i$ in descending order. The received signal in (3.82) is nothing but a set of $\mathcal{S}$ parallel channels, whose individual capacities can be achieved similar to that in Section 3.2.2 via transmitting $\mathcal{S}$ simultaneous lattice codebooks across space. The final step is allocating the power optimally, which can be achieved via waterfilling over time and space [53, Section 8.2.3]. Assuming that the joint probability distribution of $\ell^{(1)}, \ldots, \ell^{(\mathcal{S})}$ is known, the power of stream s at time instant i is given by

$$P_i^{(s)} = \{c - \frac{1}{(\ell_i^{(s)})^2}\}^+, \tag{3.83}$$

where c is chosen such that

$$c \triangleq \sum_{s=1}^{\mathcal{S}} \mathbb{E}\big[\{c - \frac{1}{(\ell_i^{(s)})^2}\}^+\big] = P, \tag{3.84}$$

where P is the average power constraint for the MIMO system. This concludes the proof of Theorem 6. □

3.3 Point-to-point Channel with imperfect CSIR[12]

Channels with imperfect channel state information have been studied extensively in the literature. Medard [28] computed bounds on the achievable rates for a given channel estimation error variance at the receiver. Hassibi and Hochwald [62] computed the optimal training time, power and signaling that maximize the throughput of MIMO channels. Most of the results to date concentrate on fading channels with Gaussian inputs. The performance of linear codes under imperfect channel state information has been largely unexplored. In [63], Pappi et al. [63] derived an expression for the compute-and-forward rates for a non-fading model with fixed error values. The extension to time-varying channels is not straightforward.

In this section, we compute bounds on the achievable rates using lattice coding and decoding as a function of channel estimation error variance, similar to Medard's approach [28]. Nested lattice codes are used where the decision regions proposed depend on the statistics of the fading channel and not the actual channel realizations. This offers a notable advantage from the perspective of decoding complexity, since the decision regions are not a function of channel realizations. For a wide range of fading distributions, the gap between the rates achieved using lattice codebooks compared to Gaussian codebooks is a constant that does not depend on the channel estimation error variance nor the input power. As a byproduct, we calculate the *Generalized Degrees-of-Freedom* [46] for channels with estimation error, where it is shown that the lattice scheme achieves the same GDoF as Medard's Gaussian-input scheme.

[12] © 2016 IEEE. Reprinted, with permission, from A. Hindy and A. Nosratinia, Lattice codes under imperfect channel state information, in International Symposium on Information Theory and its Applications (ISITA), November 2016

3.3.1 Achievable scheme

Consider a single antenna point-to-point channel with stationary and ergodic time-varying gain and i.i.d. Gaussian noise. Real-valued channels are considered. The received signal is given by $y_i = h_i x_i + w_i$. The vector $\boldsymbol{x} \in \mathbb{R}^n$ is the transmitted codeword with average power ρ. The noise $\boldsymbol{w} \in \mathbb{R}^n$ is zero-mean i.i.d. Gaussian with covariance $\boldsymbol{I}_n$ and is independent of h.

In order to account for channel estimation errors, the actual channel coefficients h_i are decomposed into two components $h_i = \hat{h}_i + \tilde{h}_i$, where $\hat{h}_i$ is the channel estimate available at the receiver, and $\tilde{h}_i$ is the zero-mean channel estimation error with variance σ_h^2, which is independent of $\hat{h}_i$.

The following theorem is the main result of this section.

Theorem 7. *For the ergodic fading point-to-point channel with partial channel state information at the receiver, any rate*

$$R < -\frac{1}{2}\log\left(\mathbb{E}\left[\frac{1}{1+\frac{\rho\hat{h}^2}{1+\rho\sigma_h^2}}\right]\right) \tag{3.85}$$

is achievable using lattice coding and decoding.

Proof. Encoding: The emitted lattice codeword is given by

$$\begin{aligned}\boldsymbol{x} &= [\boldsymbol{t} - \boldsymbol{d}] \bmod \Lambda, \\ &= \boldsymbol{t} + \boldsymbol{\lambda} - \boldsymbol{d},\end{aligned} \tag{3.86}$$

where $\boldsymbol{\lambda} = -Q_{\mathcal{V}}(\boldsymbol{t} - \boldsymbol{d}) \in \Lambda$. The dither $\boldsymbol{d}$ is uniform over $\mathcal{V}$, and $\boldsymbol{t}$ is a lattice point drawn from $\Lambda_1 \supseteq \Lambda$. As shown previously, the dither guarantees that $\boldsymbol{x}$ and $\boldsymbol{t}$ are independent.

Decoding: The received signal is multiplied by an equalization matrix $\boldsymbol{U}$ and the dither is removed as follows

$$\begin{aligned}\boldsymbol{y}' =& \boldsymbol{U}\boldsymbol{y} + \boldsymbol{d} \\ =& \boldsymbol{x} + (\boldsymbol{U}\boldsymbol{H} - \boldsymbol{I}_n)\boldsymbol{x} + \boldsymbol{U}\boldsymbol{w} + \boldsymbol{d} \\ =& \boldsymbol{t} + \boldsymbol{\lambda} + \boldsymbol{z},\end{aligned} \tag{3.87}$$

where

$$\boldsymbol{z} \triangleq (\boldsymbol{U}\hat{\boldsymbol{H}} + \boldsymbol{U}\tilde{\boldsymbol{H}} - \boldsymbol{I}_n)\boldsymbol{x} + \boldsymbol{U}\boldsymbol{w}, \tag{3.88}$$

and $\boldsymbol{z}$ is independent of $\boldsymbol{t}$ from Lemma 1.

The matrix $\boldsymbol{U}$ that minimizes $\mathbb{E}[\boldsymbol{z}^T\boldsymbol{z}]$ is the MMSE matrix given by

$$\boldsymbol{U} = \rho\hat{\boldsymbol{H}}\big(\rho\hat{\boldsymbol{H}}^2 + (1+\sigma_h^2)\boldsymbol{I}_n\big)^{-1}. \tag{3.89}$$

$\boldsymbol{U}$ is diagonal, with the following diagonal elements

$$U_{ii} = \frac{\rho\hat{h}_i}{\rho\hat{h}_i^2 + \rho\sigma_h^2 + 1}\,. \tag{3.90}$$

From (3.88) and (3.90)

$$z_i = \frac{\rho\hat{h}_i\tilde{h}_i - \rho\sigma_h^2 - 1}{\rho\hat{h}_i^2 + \rho\sigma_h^2 + 1}x_i + \frac{\rho\hat{h}_i}{\rho\hat{h}_i^2 + \rho\sigma_h^2 + 1}w_i. \tag{3.91}$$

Obviously, z_i depends on the channel estimates $\hat{h}_i$, causing the covariance of the noise to vary over time. In order to simplify decoding, we ignore the instantaneous channel knowledge subsequent to MMSE scaling. We then apply a version of the ambiguity decoder with a spherical decision region Ω as follows

$$\Omega \triangleq \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v}^T\boldsymbol{v} \leq (1+\epsilon)n\mathbb{E}\big[\frac{\rho(\sigma_h^2\rho+1)}{\rho\hat{h}^2 + \rho\sigma_h^2 + 1}\big]\}, \tag{3.92}$$

where ϵ is an arbitrary positive constant.

Probability of error: The probability of error can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i\in\mathcal{C}}\mathbb{P}_e = \mathbb{P}(\boldsymbol{z}\notin\Omega) + (1+\delta)2^{nR}\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.93}$$

for any $\delta > 0$. Following in the footsteps of Lemma 4, it can be shown that $\mathbb{P}(\boldsymbol{z}\notin\Omega) < \gamma$ for any $\gamma > 0$ at large n. Consequently, the error probability can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i\in\mathcal{C}}\mathbb{P}_e < \gamma + (1+\delta)2^{nR}\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{3.94}$$

for any $\gamma, \delta > 0$. The volume of Ω is given by

$$\mathrm{Vol}(\Omega) = (1+\epsilon)^{\frac{n}{2}} \big(\frac{\sigma_z^2}{\rho}\big)^{\frac{n}{2}} \mathrm{Vol}\big(\mathcal{B}(\sqrt{n\rho})\big). \tag{3.95}$$

The second term in (3.94) is then bounded by

$$\begin{aligned} &(1+\delta) 2^{nR} (1+\epsilon)^{\frac{n}{2}} \big(\frac{\sigma_z^2}{\rho}\big)^{\frac{n}{2}} \frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n\rho}))}{\mathrm{Vol}(\mathcal{V})} \\ &= (1+\delta) 2^{-n\Big(-\frac{1}{n}\log\big(\frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n\rho}))}{\mathrm{Vol}(\mathcal{V})}\big)+\xi\Big)}, \end{aligned} \tag{3.96}$$

where

$$\xi \triangleq \frac{-1}{2}\log(1+\epsilon) - \frac{1}{2}\log\Big(\mathbb{E}\big[\frac{\sigma_h^2\rho+1}{\rho\hat{h}^2+\rho\sigma_h^2+1}\big]\Big) - R\,. \tag{3.97}$$

From (2.8), since the lattice Λ is good for covering, the first term of the exponent in (3.96) vanishes. From (3.96), whenever ξ is a positive constant we have $\lim_{n\to\infty} \mathbb{P}_e = 0$. Hence,

$$R < -\frac{1}{2}\log\Big(\mathbb{E}\big[\frac{1}{1+\frac{\rho\hat{h}^2}{\rho\sigma_h^2+1}}\big]\Big) - \frac{1}{2}\log(1+\epsilon) - \epsilon', \tag{3.98}$$

is achievable where ϵ, ϵ' can be made arbitrarily small by increasing n. The remainder of the proof follows that of Theorem 10. This completes the proof of Theorem 7. □

For complex-valued channels, the results can be extended in a manner similar to that in Theorem 2, stated below without proof.

Theorem 8. *The rate achieved for the complex-valued ergodic fading channel with imperfect channel state information is given by*

$$R < -\log\Big(\mathbb{E}\big[\frac{1}{1+\frac{\rho|\hat{h}|^2}{\rho\sigma_h^2+1}}\big]\Big). \tag{3.99}$$

In [28], it was shown that using Gaussian inputs, the following rates are achievable

$$R_G \le \mathbb{E}\Big[\log\big(1+\frac{\rho|\hat{h}|^2}{1+\rho\sigma_h^2}\big)\Big]. \tag{3.100}$$

We compute the gap between the lattice coding rates in (3.99) and Medard's rates with Gaussian inputs (3.100). Without loss of generality let $\mathbb{E}[|\hat{h}|^2] = 1$.

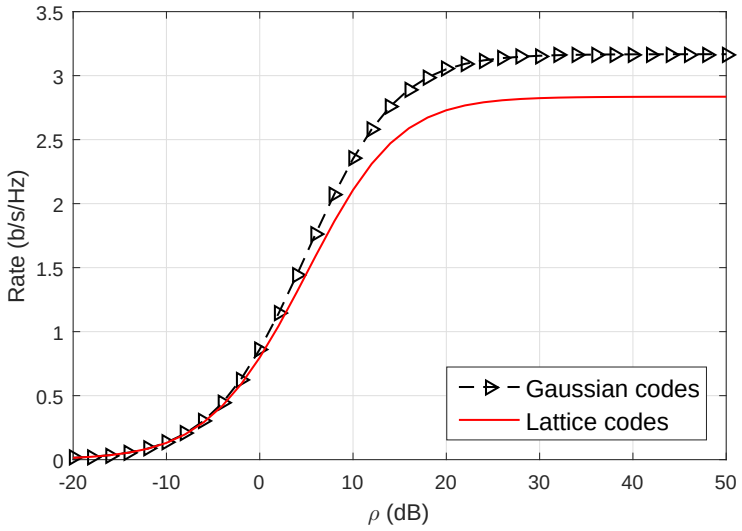

Figure 3.5. Rates under Nakagami-m fading with $m = 2$ and channel estimation error variance $\sigma_h^2 = 0.1$.

Corollary 3. *When* $\sigma_h^2 \leq \frac{\rho-1}{\rho}$ *The gap* Δ *between the rate expressions in* (3.99) *and* (3.100) *is upper bounded by*

- *When* $\mathbb{E}\left[\frac{1}{|\hat{h}|^2}\right] < \infty$*:* $\Delta < 1 + \mathbb{E}\left[\frac{1}{|\hat{h}|^2}\right]$.
- *Nakagami-m fading with* $m > 1$*:* $\Delta < 1 + \log\left(1 + \frac{1}{m-1}\right)$.
- *Rayleigh fading:* $\Delta < 0.48 + \log\left(\log\left(1 + \frac{1}{\sigma_h^2}\right)\right)$.

Proof. See Appendix A.8. □

In Figure 3.5, the rates in (3.99) and (3.100) are plotted as a function of SNR under Nakagami-m fading with $m = 2$ and $\sigma_h^2 = 0.1$. The rates are plotted under Rayleigh fading in Figure 3.6.

3.3.2 Generalized Degrees-of-Freedom Approach

The rates in (3.99), (3.100) are significantly impacted by the self interference resulting from the channel estimation error, causing both rate expressions to be upper-bounded by a constant $\mathbb{E}\left[\log(1 + \frac{\hat{h}}{\sigma_h^2})\right]$ that does not depend on the SNR. However, the previous observation

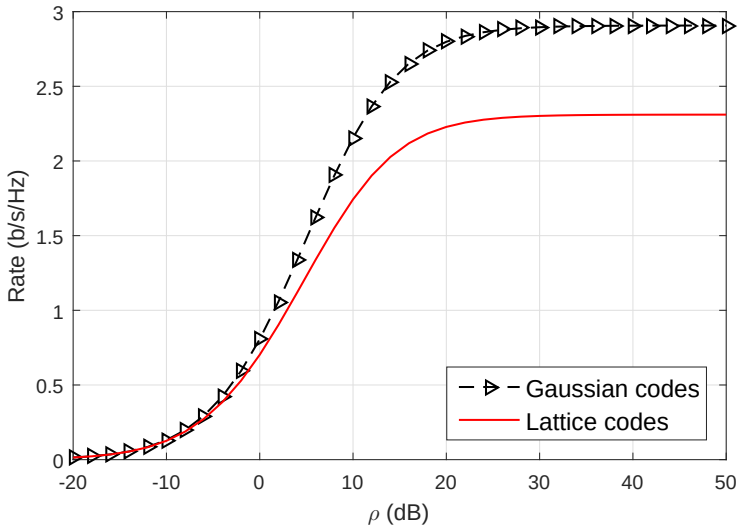

Figure 3.6. Rates under Rayleigh fading and channel estimation error variance $\sigma_h^2 = 0.1$.

overlooks a crucial aspect of practical systems: the channel estimation error variance is itself a function of the operating SNR [62]. Hence, a new figure of merit that captures the dependence of σ_h^2 on the SNR is needed. This is reminiscent of the analysis of the capacity region of the interference channel in [46], where the GDoF was introduced to measure the performance for a fixed ratio of the logarithms of the self interference to the desired signal power at high SNR. We follow a similar approach for channels with estimation errors at the receiver, where a modified version of the GDoF metric is used, as follows

$$\mathcal{D}(\alpha) \triangleq \lim_{\mathrm{SNR},\mathrm{INR}\to\infty;\alpha} \frac{R(\mathrm{INR},\mathrm{SNR})}{\mathbb{E}[\log \mathrm{SNR}]}\,, \tag{3.101}$$

where $\alpha \triangleq \frac{\log \mathrm{INR}}{\mathbb{E}[\log \mathrm{SNR}]}$, $\mathrm{INR} = \rho\sigma_h^2$ and $\mathrm{SNR} = \rho|\hat{h}|^2$. We assume $\mathbb{E}[\log \mathrm{SNR}] > \log \mathrm{INR}$.

Corollary 4. *Under Rayleigh fading and Nakagami-m fading with $m > 1$, the GDoF of the lattice scheme and Medard's Gaussian-input scheme is the same for a given α, given by*

$$\mathcal{D}(\alpha) = 1 - \alpha\,. \tag{3.102}$$

Proof. See Appendix A.9. □

CHAPTER 4

ERGODIC MULTIPLE-ACCESS CHANNEL[1]

4.1 MIMO Multiple-Access Channel

Consider a K-user MIMO MAC with N_r receive antennas and N_{t_k} antennas at transmitter k. The received signal at time i is given by

$$\tilde{\boldsymbol{y}}_i^* = \tilde{\boldsymbol{H}}_{1,i}^* \tilde{\boldsymbol{x}}_{1,i}^* + \tilde{\boldsymbol{H}}_{2,i}^* \tilde{\boldsymbol{x}}_{2,i}^* + \ldots + \tilde{\boldsymbol{H}}_{K,i}^* \tilde{\boldsymbol{x}}_{K,i}^* + \tilde{\boldsymbol{w}}_i, \tag{4.1}$$

where $\tilde{\boldsymbol{H}}_1^*, \ldots, \tilde{\boldsymbol{H}}_K^*$ are stationary and ergodic processes with zero-mean and complex-valued coefficients. The noise $\tilde{\boldsymbol{w}}$ is circularly-symmetric complex Gaussian with zero mean and unit variance, and user k has a total power constraint $N_{t_k}\rho_k^*$. An achievable strategy for the K-user MIMO MAC is independent encoding for each antenna, i.e., user k demultiplexes its data to N_{t_k} data streams, and encodes each independently and transmits it through one of its antennas. The channel can then be analyzed as a single-input multiple-output (SIMO) MAC with $L \triangleq \sum_{k=1}^{K} N_{t_k}$ virtual users. The received signal is then given by

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{h}}_{1,i}\tilde{x}_{1,i} + \tilde{\boldsymbol{h}}_{2,i}\tilde{x}_{2,i} + \ldots + \tilde{\boldsymbol{h}}_{L,i}\tilde{x}_{L,i} + \tilde{\boldsymbol{w}}_i, \tag{4.2}$$

where $\tilde{\boldsymbol{h}}_{\nu(k)+1,i}, \ldots, \boldsymbol{h}_{\nu(k)+N_{t_k},i}$ denote the N_{t_k} column vectors of $\tilde{\boldsymbol{H}}_{k,i}^*$, and $\nu(k) \triangleq \sum_{j=1}^{k-1} N_{t_j}$. The virtual user ℓ in (4.2) has power constraint ρ_l, such that

$$\rho_{\nu(k)+1} + \ldots + \rho_{\nu(k)+N_{t_k}} = N_{t_k}\rho_k^*, \quad k = 1, 2, \ldots, K. \tag{4.3}$$

The MAC achievable scheme largely depends on the point-to-point lattice coding scheme proposed in Section 3.1.1, in conjunction with successive decoding. For the L-user SIMO MAC, there are $L!$ distinct decoding orders, and the rate region is the *convex hull* of the $L!$ corner points. We define the one-to-one function $\pi(\ell) \in \{1, 2, \ldots, L\}$ that depicts a given decoding order. For example, $\pi(1) = 2$ means that the codeword of user two is the first codeword to be decoded.

[1] © 2017 IEEE. Reprinted, with permission, from A. Hindy and A. Nosratinia, Lattice Coding and Decoding for Multiple-Antenna Ergodic Fading Channels, IEEE Transactions on Communications, May 2017

Theorem 9. *For the L-user SIMO MAC with ergodic fading and complex-valued channel coefficients, lattice coding and decoding achieve the following rate region*

$$R_{MAC} \triangleq Co\Bigg(\bigcup_{\pi} \Big\{ (R_1,\ldots,R_L) : R_{\pi(\ell)} \leq -\log\Big(\mathbb{E}\big[\frac{1}{1+\rho_{\pi(\ell)} \tilde{\boldsymbol{h}}_{\pi(\ell)}^H \tilde{\boldsymbol{F}}_{\pi(\ell)}^{-1} \tilde{\boldsymbol{h}}_{\pi(\ell)}} \big] \Big) \Big\} \Bigg), \tag{4.4}$$

where

$$\tilde{\boldsymbol{F}}_{\pi(\ell)} \triangleq \boldsymbol{I}_{N_r} + \sum_{j=\ell+1}^{L} \rho_{\pi(j)} \tilde{\boldsymbol{h}}_{\pi(j)} \tilde{\boldsymbol{h}}_{\pi(j)}^H, \tag{4.5}$$

and $Co(\cdot)$ represents the convex hull *of its argument, and the union is over all permutations $\big(\pi(1),\ldots,\pi(L)\big)$.*

Proof. For ease of exposition we first assume the received signal is real-valued in the form $\boldsymbol{y}_i = \sum_{\ell=1}^{L} \boldsymbol{h}_{\ell,i}\boldsymbol{x}_{\ell,i} + \boldsymbol{w}_i$.

Encoding: The transmitted lattice codewords are given by

$$\boldsymbol{x}_\ell = [\boldsymbol{t}_\ell - \boldsymbol{d}_\ell] \bmod \Lambda^{(\ell)} \quad \ell = 1,2,\ldots,L, \tag{4.6}$$

where each lattice point $\boldsymbol{t}_l$ is drawn from $\Lambda_1^{(\ell)} \supseteq \Lambda^{(\ell)}$, and the dithers $\boldsymbol{d}_\ell$ are independent and uniform over $\mathcal{V}^{(\ell)}$. The second moment of $\Lambda^{(\ell)}$ is ρ_ℓ. Note that since transmitters have different rates and power constraints, each transmitter uses a different nested pair of lattices. The independence of the dithers across different users is necessary so as to guarantee the L transmitted codewords are independent of each other.

Decoding: The receiver uses time-varying MMSE equalization and successive cancellation over L stages, where in the first stage $\boldsymbol{x}_{\pi(1)}$ is decoded in the presence of $\boldsymbol{x}_{\pi(2)},\ldots,\boldsymbol{x}_{\pi(L)}$ as noise, and then $\boldsymbol{h}_{\pi(1),i}\boldsymbol{x}_{\pi(1),i}$ is subtracted from $\boldsymbol{y}_i$ for $i = 1,\ldots,n$. Generally, in stage ℓ, the receiver decodes $\boldsymbol{x}_{\pi(\ell)}$ from $\boldsymbol{y}_{\pi(\ell)}$, where $\boldsymbol{y}_{\pi(\ell),i} \triangleq \boldsymbol{y}_i - \sum_{j=1}^{\ell-1} \boldsymbol{h}_{\pi(j),i}\boldsymbol{x}_{\pi(j),i}$. Note that at stage ℓ the codewords $\boldsymbol{x}_{\pi(1)},\ldots,\boldsymbol{x}_{\pi(\ell-1)}$ had been canceled-out in previous stages, whereas $\boldsymbol{x}_{\pi(\ell+1)},\ldots,\boldsymbol{x}_{\pi(L)}$ are treated as noise. The MMSE vector at time i, $\boldsymbol{u}_{\pi(\ell),i}$, is given by

$$\boldsymbol{u}_{\pi(\ell),i} = \rho_{\pi(\ell)} \big(\boldsymbol{I}_{N_r} + \sum_{j=\ell}^{L} \rho_{\pi(j)} \boldsymbol{h}_{\pi(j),i} \boldsymbol{h}_{\pi(j),i}^T \big)^{-1} \boldsymbol{h}_{\pi(\ell),i}\,, \tag{4.7}$$

and the equalized signal at time i is expressed as follows

$$y'_{\pi(\ell),i} = \boldsymbol{u}^T_{\pi(\ell),i}\boldsymbol{y}_{\pi(\ell),i} + d_{\pi(\ell),i} = t_{\pi(\ell),i} + \lambda_{\pi(\ell),i} + z_{\pi(\ell),i}, \tag{4.8}$$

where $\boldsymbol{\lambda}_{\pi(\ell)} \in \Lambda^{(\pi(\ell))}$, and

$$z_{\pi(\ell),i} = \big(\boldsymbol{u}^T_{\pi(\ell),i}\boldsymbol{h}_{\pi(\ell),i} - 1\big)x_{\pi(\ell),i} + \sum_{j=\ell+1}^{L} \boldsymbol{u}^T_{\pi(\ell),i}\boldsymbol{h}_{\pi(j),i}x_{\pi(j),i} + \boldsymbol{u}^T_{\pi(\ell),i}\boldsymbol{w}_i. \tag{4.9}$$

Similar to the point-to-point scheme in Section 3.1.1, we ignore the instantaneous channel state information subsequent to the MMSE equalization step. In order to decode $\boldsymbol{x}_{\pi(\ell)}$ at stage ℓ, we apply an ambiguity decoder defined by a spherical decision region

$$\Omega^{(\pi(\ell))} \triangleq \Big\{\boldsymbol{v} \in \mathbb{R}^n \ : \ ||\boldsymbol{v}||^2 \leq (1+\epsilon)n\rho_{\pi(\ell)}\,\mathbb{E}\Big[\frac{1}{1+\rho_{\pi(\ell)}\boldsymbol{h}^T_{\pi(\ell)}\boldsymbol{F}^{-1}_{\pi(\ell)}\boldsymbol{h}_{\pi(\ell)}}\Big]\,\boldsymbol{I}_n\Big\}. \tag{4.10}$$

where ϵ is an arbitrary positive constant.

Error Probability: For an arbitrary decoding stage ℓ, the probability of error is bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}} \mathbb{P}_e^{(\pi(\ell))} < \mathbb{P}(\boldsymbol{z}_{\pi(\ell)} \notin \Omega^{(\pi(\ell))}) + (1+\delta)2^{n\check{R}_{\pi(\ell)}}\,\frac{\mathrm{Vol}(\Omega^{(\pi(\ell))})}{\mathrm{Vol}(\mathcal{V}^{(\pi(\ell))})}, \tag{4.11}$$

for some $\delta > 0$. Following in the footsteps of the proof of Lemma 4, it can be shown that $\mathbb{P}(\boldsymbol{z}_{\pi(\ell)} \notin \Omega^{(\pi(\ell))}) < \gamma$, where γ vanishes with n; the proof is therefore omitted for brevity. From (4.10),

$$\mathrm{Vol}\big(\Omega^{(\pi(\ell))}\big) = (1+\epsilon)^{\frac{n}{2}}\mathrm{Vol}\big(\mathcal{B}_n(\sqrt{n\rho_{\pi(\ell)}})\big)\,\Big(\mathbb{E}\Big[\frac{1}{1+\rho_{\pi(\ell)}\boldsymbol{h}^T_{\pi(\ell)}\boldsymbol{F}^{-1}_{\pi(\ell)}\boldsymbol{h}_{\pi(\ell)}}\Big]\Big)^{\frac{n}{2}}. \tag{4.12}$$

The second term in (4.11) is then bounded by

$$(1+\delta)2^{-n\Big(-\frac{1}{n}\log\big(\frac{\mathrm{Vol}(\mathcal{B}_n(\sqrt{n\rho_{\pi(\ell)}}))}{\mathrm{Vol}(\mathcal{V}^{(\pi(\ell))})}\big)+\xi\Big)}, \tag{4.13}$$

where

$$\xi = -\frac{1}{2}\log\Big(\mathbb{E}\Big[\frac{1}{1+\rho_{\pi(\ell)}\boldsymbol{h}^T_{\pi(\ell)}\boldsymbol{F}^{-1}_{\pi(\ell)}\boldsymbol{h}_{\pi(\ell)}}\Big]\Big) - \check{R}_{\pi(\ell)} - \frac{1}{2}\log(1+\epsilon). \tag{4.14}$$

The first term of the exponent in (4.13) vanishes since $\Lambda^{(\pi(\ell))}$ is covering-good. Then, the error probability vanishes when

$$\check{R}_{\pi(\ell)} < -\frac{1}{2}\log\Big(\mathbb{E}\big[\frac{1}{1+\rho_{\pi(\ell)}\boldsymbol{h}_{\pi(\ell)}^T\boldsymbol{F}_{\pi(\ell)}^{-1}\boldsymbol{h}_{\pi(\ell)}}\big]\Big) \tag{4.15}$$

for all $\ell \in \{1,2,\ldots,L\}$. The achievable rate region can then be extended to complex-valued channels, such that

$$R_{\pi(\ell)} < -\log\Big(\mathbb{E}\big[\frac{1}{1+\rho_{\pi(\ell)}\tilde{\boldsymbol{h}}_{\pi(\ell)}^H\tilde{\boldsymbol{F}}_{\pi(\ell)}^{-1}\tilde{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big),\ \ \ell = 1,..,L. \tag{4.16}$$

This set of rates represents one corner point of the rate region. The whole rate region is characterized by the convex hull of the $L!$ corner points that represent all possible decoding orders, as shown in (4.4). This concludes the proof of Theorem 9. □

Returning to the MIMO MAC model in (4.1), it is straightforward that the rate achieved by user k would then be

$$R_k^* = \sum_{j=1}^{N_{t_k}} R_{\nu(k)+j}, \tag{4.17}$$

where R_j are the rates given in (4.16). Now we compare $R_{\text{sum}} \triangleq \sum_{k=1}^{K} R_k^*$ with the sum capacity of the MIMO MAC model in (4.1). We focus our comparison on the case where the channel matrices have i.i.d. complex Gaussian entries and all users have the same number of transmit antennas as well as power budgets, i.e., $N_{t_k} = N_t$, $\rho_k^* = \rho$ for all $k \in \{1,2,\ldots,K\}$. The optimal input covariance is then a scaled identity matrix [26] and the sum capacity is given by [32]

$$C_{\text{sum}} = \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_r} + \sum_{k=1}^{K}\rho\tilde{\boldsymbol{H}}_k^*\tilde{\boldsymbol{H}}_k^{*H}\big)\Big]. \tag{4.18}$$

Corollary 5. *For the K-user fading MIMO MAC in* (4.1)*, when $\tilde{\boldsymbol{H}}_k^*$ is i.i.d. complex Gaussian and $N_r > KN_t$, the gap between the sum rate of the lattice scheme and the sum capacity at $\rho \geq 1$ is upper bounded by*

$$\Delta < \sum_{\ell=1}^{N_tK}\log\big(1+\frac{\ell+1}{N_r-\ell}\big). \tag{4.19}$$

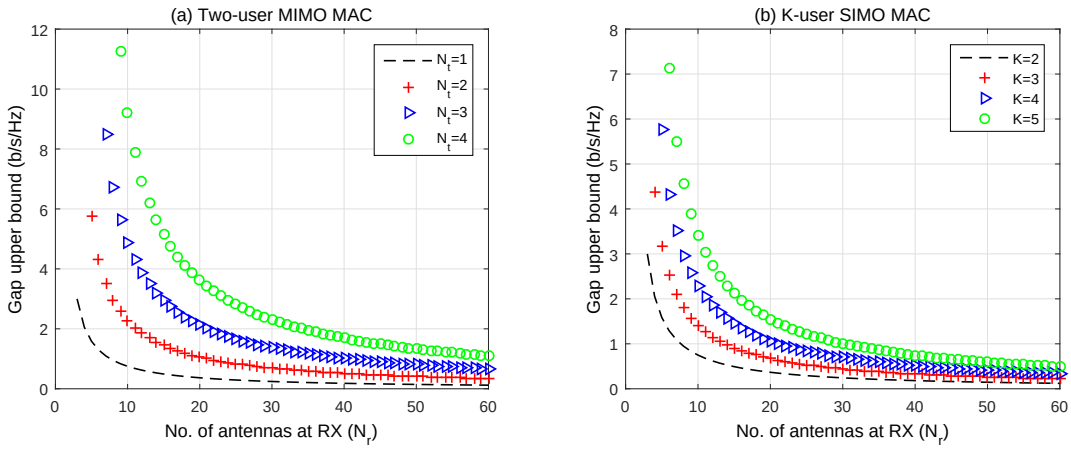

Figure 4.1. The upper bound on the gap to sum capacity of the MIMO MAC vs. N_r.

Proof. See Appendix B.1. □

Similar to Corollary 1, the gap to capacity vanishes at finite SNR as N_r grows, i.e., $\Delta \to 0$ as $N_r \to \infty$. The expression in (4.19) is plotted in Figure 4.1 for $K = 2$, as well as for the K-user SIMO MAC.

4.2 SISO Multiple-Access Channel

For the two-user MAC with $N_r = N_t = 1$, the rate region in (4.4) can be expressed by [2]

$$R_1 < -\gamma_1\,,$$

$$R_2 < -\gamma_2\,,$$

$$(\gamma_4 - \gamma_2)R_1 + (\gamma_3 - \gamma_1)R_2 < (\gamma_1\gamma_2 - \gamma_3\gamma_4), \tag{4.20}$$

where

$$\gamma_1 = \log\left(\mathbb{E}\Big[\frac{1}{1+\rho_1|\tilde{h}_1|^2}\Big]\right), \quad \gamma_2 = \log\left(\mathbb{E}\Big[\frac{1}{1+\rho_2|\tilde{h}_2|^2}\Big]\right),$$
$$\gamma_3 = \log\left(\mathbb{E}\Big[\frac{1}{1+\frac{\rho_1|\tilde{h}_1|^2}{1+\rho_2|\tilde{h}_2|^2}}\Big]\right), \gamma_4 = \log\left(\mathbb{E}\Big[\frac{1}{1+\frac{\rho_2|\tilde{h}_2|^2}{1+\rho_1|\tilde{h}_1|^2}}\Big]\right).$$

[2] Unlike the two-user MAC *capacity region*, the sum rate does not necessarily have a unit slope.

For the case where all nodes are equipped with a single antenna, we characterize the gap to sum capacity of the two-user MAC for a wider range of distributions and over all SNR values. For ease of exposition we assume $\tilde{h}_1$ and $\tilde{h}_2$ are identically distributed with $\mathbb{E}[|\tilde{h}_1|^2] = \mathbb{E}[|\tilde{h}_2|^2] = 1$. Δ is then given by

$$\Delta \triangleq \mathbb{E}\big[\log\left(1+\rho|\tilde{h}_1|^2+\rho|\tilde{h}_2|^2\right)\big] + \log\Big(\mathbb{E}\big[\frac{1+\rho|\tilde{h}_1|^2}{1+\rho|\tilde{h}_1|^2+\rho|\tilde{h}_2|^2}\big]\mathbb{E}\big[\frac{1}{1+\rho|\tilde{h}_1|^2}\big]\Big). \tag{4.21}$$

Corollary 6. *The gap to capacity of the two-user MAC given in* (4.21) *is upper-bounded as follows*

- $\rho < \frac{1}{2}$*: For any fading distribution where* $\mathbb{E}\big[|\tilde{h}_1|^4\big] < \infty$*,*

$$\Delta < 1.45\left(1+2\mathbb{E}\big[|\tilde{h}_1|^4\big]\right)\rho^2. \tag{4.22}$$

- $\rho \geq \frac{1}{2}$*: For any fading distribution where* $\mathbb{E}\big[\frac{1}{|\tilde{h}_1|^2}\big] < \infty$*,*

$$\Delta < 2 + \log\Big(\mathbb{E}\big[\frac{1}{|\tilde{h}_1|^2}\big]\Big). \tag{4.23}$$

- $\rho \geq \frac{1}{2}$*: Under Nakagami-m fading with* $m > 1$*,*

$$\Delta < 2 + \log\big(1+\frac{1}{m-1}\big). \tag{4.24}$$

- $\rho \geq \frac{1}{2}$*: Under Rayleigh fading,*

$$\Delta < 1.48 + \log\big(\log(1+\rho)\big). \tag{4.25}$$

Proof. See Appendix B.2. □

In Figure 4.2, the sum rate of the two-user MAC lattice scheme is compared with the sum capacity under Nakagami-m fading with $m = 2$, as well as under i.i.d. Rayleigh fading. It can be shown that the gap to sum capacity is small in both cases. We plot the rate region under Rayleigh fading at $\rho = -6$ dB per user in Figure 4.3. The rate region is shown to be close to the capacity region, indicating the efficient performance of the lattice scheme at low SNR.

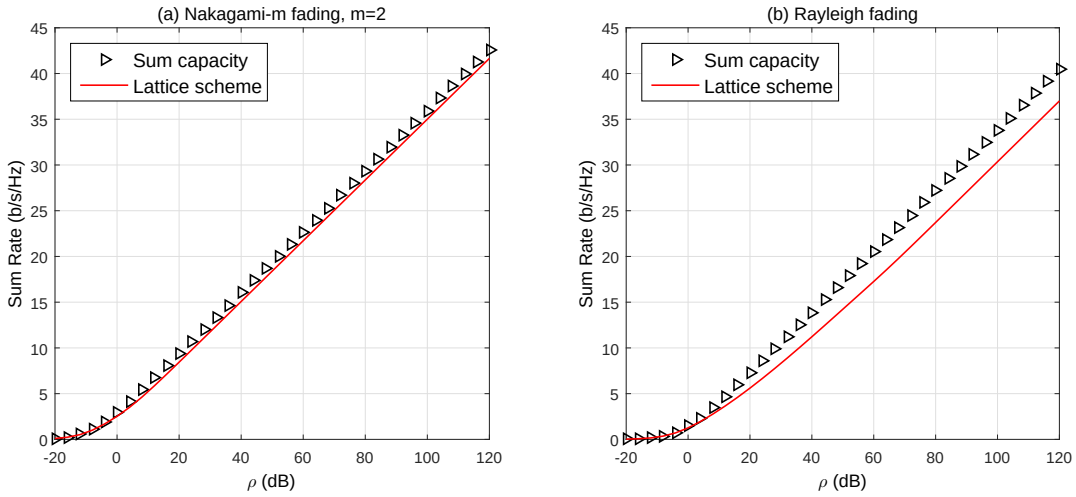

Figure 4.2. The two-user MAC sum rate vs. sum capacity.

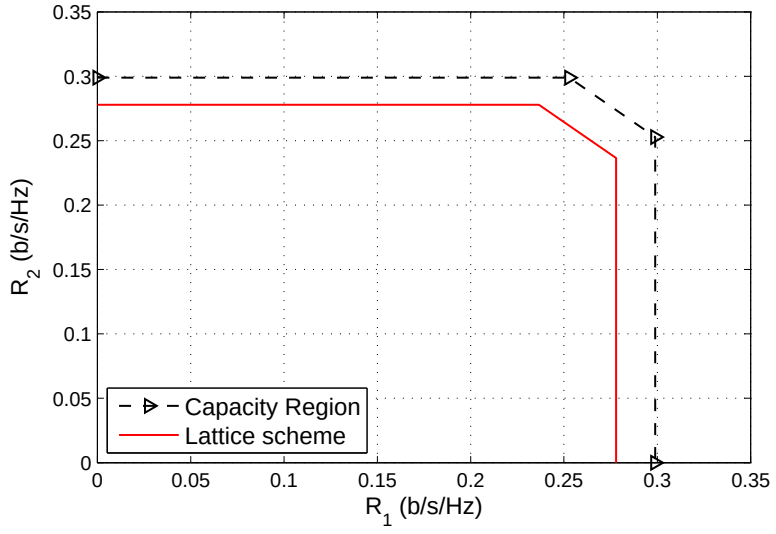

Figure 4.3. The two-user MAC rate region vs. ergodic capacity at $\rho = -6\ dB$ under Rayleigh fading.

CHAPTER 5

ERGODIC DIRTY PAPER AND BROADCAST CHANNELS[1]

Costa's work on the dirty paper channel [33] has been an influential result in the realm of information theory. Recently, variants of the dirty paper channel have been addressed in the literature. Vaze and Varanasi [64] studied the fading MIMO dirty paper channel with partial channel state information at the transmitter, where a scheme was derived that is optimal in the high-SNR limit. For the same setting Bennatan andErgodic Fading MIMO Dirty Paper and Broadcast Channels: Capacity Bounds and Lattice Strategies Burshtein [65] proposed a numeric approach that achieves capacity under certain design constraints. The results in [65] were also applied to the fading MIMO broadcast channel. For the single-antenna fading dirty paper channel with receiver channel knowledge, Zhang et al. [66] showed that a variant of Costa's scheme is optimal at both high and low SNR.[2]

This chapter addresses the fading MIMO dirty paper channel in the absence of CSIT, where the dirt is white, stationary and ergodic. The desired signal and dirt undergo the same fading state, which represents the case where the sources of the desired signal and interference are co-located at the same node. We show that dirty paper coding is within a constant gap to ergodic capacity for all SNR and all dirt power. This improves on the result in [66] since the gap to capacity is computed analytically for all SNR and all antenna configurations. Moreover, a lattice coding and decoding scheme is proposed, where the class of nested lattice codes proposed in [5] are used at the transmitter, and the decision regions are universal for almost all realizations of a given fading distribution.

---

[1] © 2017 IEEE. Reprinted, with permission, from A. Hindy and A. Nosratinia, Ergodic Fading MIMO Dirty Paper and Broadcast Channels: Capacity Bounds and Lattice Strategies, IEEE Transactions on Wireless Communications, to appear

[2] In [20], Lin et al. proposed a version of the lattice coding and decoding scheme in [7] for the fading MIMO dirty paper channel, where a decoding rule that depends on the channel realizations is used. The proof in [7], originally derived for quasi-static MIMO channels, uses the *Minkowski-Hlawka Theorem* to prove the existence of a codebook with negligible error probability for a given channel state. However, the existence of a universal codebook that achieves the same error probability over all channel states is not guaranteed and hence the achievable rates in [20] remain under question.

One advantage of the fading dirty paper model is its straightforward extension to the broadcast channel with CSIR, where the desired signal and interference stem from the same source and therefore undergo the same fading process. We apply the dirty paper channel results to a two-user MIMO broadcast channel with different fading dynamics, where the fading process is stationary and ergodic for one receiver and quasi-static for the other receiver. In addition, the case where both users experience ergodic fading processes that are independent of each other is also studied. Unlike conventional broadcast channel techniques, the proposed scheme does not require any receiver to know the codebooks of the interference signals. Performance is compared with a version of dirty paper coding under non-causal CSIT.

5.1 Dirty Paper Channel: System Model

Consider a MIMO point-to-point channel with Gaussian noise and N_t, N_r antennas at the transmitter and receiver sides, respectively. The fading process is stationary and ergodic, where the random channel matrix is denoted by $\boldsymbol{H}$. The received signal is impeded not only by Gaussian noise, but also by another channel impediment $\boldsymbol{s}$ that experiences the same fading as the information-bearing signal $\boldsymbol{x}$, as follows

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x}_i + \boldsymbol{H}_i \boldsymbol{s}_i + \boldsymbol{w}_i, \tag{5.1}$$

where the channel coefficient matrices $\boldsymbol{H}_i \in \mathbb{C}^{N_r \times N_t}$ at time $i = 1, \ldots, n$ denote realizations of the random matrix $\boldsymbol{H}$. Moreover, $\boldsymbol{H}$ is zero mean and isotropically distributed, i.e., $P(\boldsymbol{H}) = P(\boldsymbol{HV})$ for any unitary matrix $\boldsymbol{V}$ independent of $\boldsymbol{H}$. The unordered eigenvalues of the Hermitian random matrix $\boldsymbol{H}^H\boldsymbol{H}$, denoted by $\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2$, are also random, and their distribution is characterized by the distribution of $\boldsymbol{H}$. The receiver has instantaneous channel knowledge, whereas the transmitter only knows the channel distribution. $\boldsymbol{x}_i \in \mathbb{C}^{N_t}$ is the transmitted vector at time i, where the codeword

$$\boldsymbol{x} \triangleq [\boldsymbol{x}_1^H \, \boldsymbol{x}_2^H, \ldots \boldsymbol{x}_n^H]^H \tag{5.2}$$

is transmitted throughout n channel uses and satisfies $\mathbb{E}[||\boldsymbol{x}||^2] \leq nP_x$. The noise $\boldsymbol{w} \in \mathbb{C}^{N_r n}$ defined by $\boldsymbol{w} \triangleq [\boldsymbol{w}_1^H, \ldots, \boldsymbol{w}_n^H]^H$ is a circularly-symmetric zero-mean i.i.d. Gaussian noise vector with covariance $P_w \boldsymbol{I}_{N_r n}$, and is independent of the channel coefficients. $\boldsymbol{s} \in \mathbb{C}^{N_t n}$, where $\boldsymbol{s} \in [\boldsymbol{s}_1^H, \ldots, \boldsymbol{s}_n^H]^H$ represents the state (dirt) that is independent of $\boldsymbol{H}$ and $\boldsymbol{w}$ and is known non-causally at the transmitter. Unless otherwise stated, we assume $\boldsymbol{s}$ is a stationary and ergodic sequence whose elements have zero-mean and variance P_s.

An intuitive outer bound for the rates of the channel in (5.1) would be the point-to-point channel capacity in the absence of the state $\boldsymbol{s}$, as follows

$$C \leq \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w} \boldsymbol{H}^H \boldsymbol{H}\big)\Big]. \tag{5.3}$$

Had the channel coefficients been known non-causally at the transmitter, the rate in (5.3) would have been achieved in a straightforward manner from Costa's result since the new state $\tilde{\boldsymbol{s}}_i \triangleq \boldsymbol{H}_i \boldsymbol{s}_i$ would be known at the transmitter [67, Chapter 9.5]. However, in the present model $\boldsymbol{H}$ is unknown at the transmitter, causing the problem to become more challenging. In the sequel, two different inner bounds are studied that approach the outer bound in (5.3).

5.2 Dirty Paper Coding Inner Bound

In this section we aim at deriving an achievable scheme for the channel in (5.1), that is as close as possible to capacity. The channel in (5.1) is a variation of Gel'fand and Pinsker's discrete memoryless channel with a state known non-causally at the transmitter [38], whose capacity can be expressed by

$$C = \max_{P_{V,X|S}} I(V; Y, H) - I(V; S), \tag{5.4}$$

where S represents the state and V is an auxiliary random variable. Y and H represent the receiver observation and the available channel state information at the receiver, respectively. Unfortunately the single-letter capacity optimization in (5.4) is not tractable. In [33], Costa studied the non-fading dirty paper channel with additive Gaussian noise, where he showed that

the point-to-point capacity can be achieved and the impact of $\boldsymbol{s}$ can be entirely eliminated. In the sequel we propose an achievable scheme for the fading MIMO dirty paper channel.

Theorem 10. *For the ergodic fading MIMO dirty paper channel in* (5.1)*, any rate satisfying*

$$R \leq \Big(\mathbb{E}\big[\log\det\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_{N_t}+\frac{P_x}{N_tP_w}\boldsymbol{H}^H\boldsymbol{H}\big)\big]\Big)^+, \tag{5.5}$$

is achievable, as long as the elements of $\boldsymbol{s}$ *are i.i.d. Gaussian.*

Proof. We first consider real-valued channels. We follow in the footsteps of the encoding and decoding schemes in [33, 64], where random binning at the encoder and typicality decoding were used. Details are as follows.

Encoding: The transmitted signal of length Mn is in the form $\boldsymbol{x}=\boldsymbol{v}-\boldsymbol{U}\boldsymbol{s}$, where $\boldsymbol{v}$ is drawn from a codebook consisting of $2^{n\tilde{R}}$ codewords for some $\tilde{R}>0$, each of length N_tn. The codewords are drawn from a Gaussian distribution with zero mean and covariance $\frac{P_x}{N_t}\boldsymbol{I}_{N_tn}+P_s\boldsymbol{U}\boldsymbol{U}^T$, where $\boldsymbol{U}\in\mathbb{R}^{N_tn\times N_tn}$ will be determined later. These codewords are randomly assigned to 2^{nR} bins for some $0<R<\tilde{R}$, so that each bin will contain approximately $2^{n(\tilde{R}-R)}$ codewords. As long as $2^{n(\tilde{R}-R)}>2^{nI(V;S)}$, typicality arguments guarantee the existence of a codeword $\boldsymbol{v}_0$ in each bin that is jointly typical with the state $\boldsymbol{s}_0$, i.e., $\boldsymbol{v}_o-\boldsymbol{U}\boldsymbol{s}_o$ is nearly orthogonal to $\boldsymbol{s}_o$ [33]. The bin index is chosen according to the message to be transmitted, and from that bin the appropriate codeword is transmitted that is jointly typical with the state. The transmitter emits $\boldsymbol{x}_o\triangleq\boldsymbol{v}_o-\boldsymbol{U}\boldsymbol{s}_o$, which satisfies the power constraint, $\mathbb{E}[||\boldsymbol{x}||^2]\leq nP_x$ (recall $\boldsymbol{x}_o,\boldsymbol{s}_o$ are orthogonal).

Decoding: Given the occurrence of state $\boldsymbol{s}_o$, the received signal is given by

$$\begin{aligned}\boldsymbol{y}=&\boldsymbol{H}_d\boldsymbol{x}_o+\boldsymbol{H}_d\boldsymbol{s}_o+\boldsymbol{w}_o\\ =&\boldsymbol{H}_d\boldsymbol{v}_o+(\boldsymbol{I}_{N_tn}-\boldsymbol{U})\boldsymbol{H}_d\boldsymbol{s}_o+\boldsymbol{w},\end{aligned} \tag{5.6}$$

where $\boldsymbol{H}_d\triangleq\mathrm{diag}\{\boldsymbol{H}_1,\ldots,\boldsymbol{H}_n\}$, and the receiver knows the codebook of $\boldsymbol{v}$. From standard typicality arguments, $\boldsymbol{v}_o$ can be decoded reliably as long as $2^{n\tilde{R}}<2^{nI(V;Y,H)}$ for sufficiently large n.

Rate analysis: Based on the encoding and decoding procedures, the number of distinguishable messages that can be transmitted is equal to the number of bins 2^{nR}. The achievable rate can then be analyzed as follows

$$\begin{aligned}
nR &< I(V;Y,H) - I(V;S) \\
&= h(V) + h(Y|H) - h(Y,V|H) - h(V) + h(V|S) \\
&= h(Y|H) + h(V - \alpha S|S) - h(Y,V|H) \\
&= h(Y|H) + h(X) - h(Y,V|H) \\
&= \frac{1}{2}\log\big((2\pi e)^{N_t n}\det(\boldsymbol{Q})\big) + \frac{1}{2}\log\big((2\pi e)^{N_t n}\det(\frac{P_x}{N_t}\boldsymbol{I}_{N_t n})\big) \\
&\quad - \frac{1}{2}\log\big((2\pi e)^{(N_t+N_r)n}\det(\boldsymbol{Q})\det\big(\frac{P_x}{N_t}\boldsymbol{I}_{N_t n} + P_s\boldsymbol{U}\boldsymbol{U}^T \\
&\quad\quad - (\frac{P_x}{N_t}\boldsymbol{I}_{N_t n} + P_s\boldsymbol{U})\boldsymbol{H}_d^T\boldsymbol{Q}^{-1}\boldsymbol{H}_d(\frac{P_x}{N_t}\boldsymbol{I}_{N_t n} + P_s\boldsymbol{U}^T)\big)\big),
\end{aligned} \tag{5.7}$$

where $\boldsymbol{Q} \triangleq \frac{P_x}{N_t}\boldsymbol{H}_d\boldsymbol{H}_d^T + P_s\boldsymbol{H}_d\boldsymbol{H}_d^T + P_w\boldsymbol{I}_{N_r n}$. On choosing $\boldsymbol{U} = \boldsymbol{I}_{N_t n}$,

$$R < \frac{1}{2n}\sum_{i=1}^{n}\log\det\big(\frac{P_x}{P_x + N_t P_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w}\boldsymbol{H}_i^T\boldsymbol{H}_i\big)\Big]. \tag{5.8}$$

From the law of large numbers, the rate bound in (5.8) converges to

$$R < \frac{1}{2}\mathbb{E}\Big[\log\det\big(\frac{P_x}{P_x + N_t P_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w}\boldsymbol{H}^T\boldsymbol{H}\big)\Big] - \epsilon \tag{5.9}$$

with probability 1, where ϵ vanishes as $n \to \infty$. Finally, the result can be extended to the complex-valued channel model through the following equivalent channel model

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{H}}_i\tilde{\boldsymbol{x}}_i + \tilde{\boldsymbol{H}}_i\tilde{\boldsymbol{s}}_i + \tilde{\boldsymbol{w}}_i, \tag{5.10}$$

where

$$\tilde{\boldsymbol{H}}_i \triangleq \begin{bmatrix} \boldsymbol{H}_i^{(R)} & -\boldsymbol{H}_i^{(I)} \\ \boldsymbol{H}_i^{(I)} & \boldsymbol{H}_i^{(R)} \end{bmatrix} \tag{5.11}$$

is a $2N_r \times 2N_t$ real-valued channel matrix and $\tilde{\boldsymbol{x}}_i \triangleq [\boldsymbol{x}_i^{(R)T}\boldsymbol{x}_i^{(I)T}]^T$ and similarly for $\tilde{\boldsymbol{s}}_i, \tilde{\boldsymbol{w}}_i$, and the rate can then be expressed by the form in (5.5). This concludes the proof of Theorem 10. □

In the following we bound the gap between the inner and outer bounds in (5.3),(5.5). We assume $N_t \leq N_r$.

Corollary 7. *The rate achieved in* (5.5) *is within* N_t *bits of the capacity.*

Proof. The gap between the capacity outer bound in (5.3) and (5.5) is bounded by

$$
\begin{aligned}
\Delta \triangleq & C - R \\
\leq & \mathbb{E}\big[\log\det\big(\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w}\boldsymbol{H}^H\boldsymbol{H}\big)\big] \; - \Big(\mathbb{E}\big[\log\det\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_tP_w}\boldsymbol{H}^H\boldsymbol{H}\big)\big]\Big)^+ \\
\leq & \mathbb{E}\big[\log\det\big(\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w}\boldsymbol{H}^H\boldsymbol{H}\big)\big] \; - \mathbb{E}\big[\log\det\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_tP_w}\boldsymbol{H}^H\boldsymbol{H}\big)\big] \\
= & \mathbb{E}\big[\sum_{j=1}^{N_t}\log(1+\frac{P_x}{N_tP_w}\sigma_j^2)\big] \; - \mathbb{E}\big[\sum_{j=1}^{N_t}\log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\big] \qquad (5.12) \\
= & \sum_{j=1}^{N_t}\mathbb{E}_{\sigma_j}\Big[\log(1+\frac{P_x}{N_tP_w}\sigma_j^2) - \log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\Big] \\
= & \sum_{j=1}^{N_t}\Bigg(\mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\big)\mathbb{E}\Big[\log(1+\frac{P_x}{N_tP_w}\sigma_j^2) - \log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\Big|\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\Big] \\
& + \mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2< 1\big)\mathbb{E}\Big[\log(1+\frac{P_x}{N_tP_w}\sigma_j^2) - \log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\Big|\frac{P_x}{N_tP_w}\sigma_j^2< 1\Big]\Bigg) \qquad (5.13) \\
< & \sum_{j=1}^{N_t}\Bigg(\mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\big)\mathbb{E}\Big[1+\log(\frac{P_x}{N_tP_w}\sigma_j^2) - \log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\Big|\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\Big] \\
& + \mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2< 1\big)\mathbb{E}\Big[1 - \log(\frac{P_x}{P_x+N_tP_s}+\frac{P_x}{N_tP_w}\sigma_j^2)\Big|\frac{P_x}{N_tP_w}\sigma_j^2< 1\Big]\Bigg) \\
< & \sum_{j=1}^{N_t}\Big(\mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\big)\mathbb{E}\big[1\,\big|\,\frac{P_x}{N_tP_w}\sigma_j^2\geq 1\big] + \mathbb{P}\big(\frac{P_x}{N_tP_w}\sigma_j^2< 1\big)\mathbb{E}\big[1\,\big|\,\frac{P_x}{N_tP_w}\sigma_j^2< 1\big]\Big) \\
= & N_t\,, \qquad (5.14)
\end{aligned}
$$

where σ_j^2 are the eigenvalues of $\boldsymbol{H}^H\boldsymbol{H}$ for $j = 1,\ldots,N_t$, as explained in Section 5.1, and hence (5.12) is an alternative representation of the expressions in (5.3),(5.5) in terms of the channel eigenvalues. (5.13) follows from the law of total expectation on each of the rate expressions. The gap to capacity is then shown to be bounded from above by N_t bits. □

Remark 3. *The rates achieved in [64, Section IV] were shown to approach capacity at high SNR, i.e., $\Delta \to 0$ as $P_x \to \infty$. Meanwhile, Corollary 1 bounds the gap to capacity within a constant number of bits, irrespective of the values of P_x, P_s as well as the fading distribution. This result does not contradict with that in [64], however. For instance, when $N_r = N_t = 1$ and $P_w = 1$, the gap to capacity would be as follows*

$$\begin{aligned}\Delta &\leq \mathbb{E}\big[\log(1+P_x|h|^2)\big] - \mathbb{E}\big[\log(\frac{P_x}{P_x+P_s}+P_x|h|^2)\big] \\ &= \mathbb{E}\big[\log(1+\frac{\frac{P_s}{P_x+P_s}}{\frac{P_x}{P_x+P_s}+|h|^2P_x})\big] \\ &< \mathbb{E}\big[\log(1+\frac{1}{|h|^2P_x})\big], \end{aligned} \tag{5.15}$$

which vanishes as $P_x \to \infty$, confirming the result in [64].

5.3 Lattice coding inner bound

Although the scheme proposed in Section 5.2 achieves rates that are close to capacity, it has large computational complexity at both the transmitter and receiver since it uses a Gaussian codebook. In this section a lattice coding and decoding scheme is proposed that transmits a dithered lattice codeword and at the receiver uses a single-tap equalizer and lattice decoding, similar to [5, 34]. In our scheme the use of CSIR is limited to the equalizer and the lattice decision regions do not depend on the instantaneous realizations of the fading channel.

Theorem 11. *For the ergodic fading MIMO dirty paper channel given in* (5.1)*, any rate*

$$R < \left(-\log\det\left(\mathbb{E}\big[\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_M + \frac{P_x}{N_tP_w}\boldsymbol{H}^H\boldsymbol{H}\big)^{-1}\big]\right)\right)^+ \tag{5.16}$$

is achievable using lattice coding and decoding.

Proof. We first consider real-valued channels, where $\boldsymbol{H}_i \in \mathbb{R}^{N_r\times N_t}$ and the elements of $\boldsymbol{w}$ are zero-mean i.i.d. Gaussian with variance P_w.

Encoding: Nested lattice codes are used where $\Lambda \subseteq \Lambda_1$. The transmitter emits a signal $\boldsymbol{x}$ as follows

$$\begin{aligned}\boldsymbol{x} &= [\boldsymbol{t} - \boldsymbol{B}\boldsymbol{s} - \boldsymbol{d}] \bmod \Lambda, \\ &= \boldsymbol{t} + \boldsymbol{\lambda} - \boldsymbol{B}\boldsymbol{s} - \boldsymbol{d},\end{aligned} \tag{5.17}$$

where $\boldsymbol{t} \in \mathcal{L}_1$ is a point drawn from a nested lattice code with $\Lambda \subseteq \Lambda_1$, dithered with $\boldsymbol{d}$ which is drawn uniformly over $\mathcal{V}$, and $\boldsymbol{\lambda} = -Q_{\mathcal{V}}(\boldsymbol{t} - \boldsymbol{B}\boldsymbol{s} - \boldsymbol{d}) \in \Lambda$ from (2.5). $\boldsymbol{B} \in \mathbb{R}^{N_t n \times N_t n}$ is a matrix to be chosen in the sequel.[3] Note from Lemma 1 that the dither guarantees the independence of $\boldsymbol{x}$ from both $\boldsymbol{t}$ and $\boldsymbol{s}$.

Decoding: The received signal in (5.1) is multiplied by a single-tap equalization matrix $\boldsymbol{U}_d$ and the dither is removed as follows

$$\begin{aligned}\boldsymbol{y}' =& \boldsymbol{U}_d^T \boldsymbol{y} + \boldsymbol{d} \\ =& \boldsymbol{U}_d^T \boldsymbol{H}_d \boldsymbol{x} + \boldsymbol{U}_d^T \boldsymbol{H}_d \boldsymbol{s} + \boldsymbol{U}_d^T \boldsymbol{w} + \boldsymbol{d} \\ =& \boldsymbol{x} + (\boldsymbol{U}_d^T \boldsymbol{H}_d - \boldsymbol{I}_{N_t n})\boldsymbol{x} + \boldsymbol{U}_d^T \boldsymbol{H}_d \boldsymbol{s} + \boldsymbol{U}_d^T \boldsymbol{w} + \boldsymbol{d} \\ =& \boldsymbol{t} + \boldsymbol{\lambda} + \boldsymbol{z},\end{aligned} \tag{5.18}$$

where

$$\boldsymbol{z} \triangleq (\boldsymbol{U}_d^T \boldsymbol{H}_d - \boldsymbol{I}_{N_t n})\boldsymbol{x} + (\boldsymbol{U}_d^T \boldsymbol{H}_d - \boldsymbol{B})\boldsymbol{s} + \boldsymbol{U}_d^T \boldsymbol{w} \tag{5.19}$$

is independent of $\boldsymbol{t}$ according to Lemma 1. For the special case $\boldsymbol{H}_d = \boldsymbol{I}_{N_t n}$, the problem reduces to the non-fading dirty paper channel, where $\boldsymbol{U}_d = \boldsymbol{B} = \frac{P_x}{P_x + P_w} \boldsymbol{I}_{N_t n}$ is optimal and the point-to-point channel capacity can be achieved via the lattice coding and decoding scheme in [34]. However, this scheme cannot be directly extended to ergodic fading, since the channel realizations are unknown at the transmitter. We choose $\boldsymbol{B} = \boldsymbol{I}_{N_t n}$ so that

$$\boldsymbol{z} \triangleq (\boldsymbol{U}_d^T \boldsymbol{H}_d - \boldsymbol{I}_{N_t n})(\boldsymbol{x} + \boldsymbol{s}) + \boldsymbol{U}_d^T \boldsymbol{w}. \tag{5.20}$$

[3]Note that $\boldsymbol{B}$ must be independent of the channel realizations.

The motivation is to align the dirt and self-interference terms in (5.19). The equalization matrix $\boldsymbol{U}_d$ is then chosen to minimize $\mathbb{E}\big[||\boldsymbol{z}||^2\big]$, which is a block-diagonal matrix whose diagonal block i, $\boldsymbol{U}_i$ is given by[4]

$$\boldsymbol{U}_i = (\frac{P_x}{N_t} + P_s)\big((\frac{P_x}{N_t} + P_s)\boldsymbol{H}_i\boldsymbol{H}_i^T + P_w\boldsymbol{I}_{N_r}\big)^{-1}\boldsymbol{H}_i. \tag{5.21}$$

From (5.20) and (5.21)

$$\begin{aligned}\boldsymbol{z}_i = &- \big(\boldsymbol{I}_{N_t} + \frac{1}{P_w}(\frac{P_x}{N_t} + P_s)\boldsymbol{H}_i^T\boldsymbol{H}_i\big)^{-1}(\boldsymbol{x}_i + \boldsymbol{s}_i) \\ &+ (\frac{P_x}{N_t} + P_s)\boldsymbol{H}_i^T((\frac{P_x}{N_t} + P_s)\boldsymbol{H}_i\boldsymbol{H}_i^T + P_w\boldsymbol{I}_{N_r})^{-1}\boldsymbol{w}_i.\end{aligned} \tag{5.22}$$

Naturally, the distribution of $\boldsymbol{z}$ conditioned on $\boldsymbol{H}_i$ (which is known at the receiver) varies across time. This variation produces complications, so in order to simplify the decoding process, we ignore the instantaneous channel knowledge at the decoder following the equalization step, i.e., after equalization the receiver considers $\boldsymbol{H}_i$ a random matrix. The same approach has been adopted in the point-to-point scheme in Section 3.1.1.

Lemma 9. *For any $\epsilon > 0$ and $\gamma > 0$, there exists $n_{\gamma,\epsilon}$ such that for all $n > n_{\gamma,\epsilon}$,*

$$\mathbb{P}\big(\boldsymbol{z} \notin \Omega\big) < \gamma, \tag{5.23}$$

where Ω is the following sphere

$$\Omega \triangleq \Big\{\boldsymbol{v} \in \mathbb{R}^{N_t n} \,:\, ||\boldsymbol{v}||^2 \leq (1+\epsilon)n\, tr\big(\mathbb{E}\big[\bar{\boldsymbol{\Sigma}}\big]\big)\Big\}, \tag{5.24}$$

and $\bar{\boldsymbol{\Sigma}} \triangleq \mathbb{E}\Big[\big(\frac{1}{\frac{P_x}{N_t}+P_s}\boldsymbol{I}_{N_t} + \frac{1}{P_w}\boldsymbol{H}^T\boldsymbol{H}\big)^{-1}\Big]$ is a scaled identity matrix, and ϵ is an arbitrary positive constant.

Proof. See Appendix C.1. □

[4]Unlike [5, 34], $\boldsymbol{U}_i$ is *not* the MMSE equalization matrix of the channel in (5.1).

Now, we apply an ambiguity decoder with a spherical decision region Ω in (5.24). The decoder decides $\hat{\boldsymbol{t}} \in \Lambda_1$ if and only if the received point falls exclusively within the decision region of the lattice point $\hat{\boldsymbol{t}}$, i.e., $\boldsymbol{y}' \in \hat{\boldsymbol{t}} + \Omega$.

Probability of error: On averaging over the set of all fine lattices $\mathcal{C}$ of rate R whose construction follows that in Section 2.1, the probability of error can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i \in \mathcal{C}} \mathbb{P}_e < \mathbb{P}(\boldsymbol{z} \notin \Omega) + (1+\delta)2^{nR}\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V})}, \tag{5.25}$$

for any $\delta > 0$. From Lemma 9, the first term in (5.25) vanishes with n. For convenience define $\boldsymbol{\Psi} = \frac{P_x}{N_t}\bar{\boldsymbol{\Sigma}}^{-1}$. The volume of Ω is given by

$$\mathrm{Vol}(\Omega) = (1+\epsilon)^{\frac{N_t n}{2}}\mathrm{Vol}\big(\mathcal{B}_{N_t n}(\sqrt{nP_x})\big)\ \det\left(\boldsymbol{\Psi}^{\frac{-1}{2}}\right). \tag{5.26}$$

The second term in (5.25) is bounded by

$$\begin{aligned}(1+\delta)2^{nR}(1+\epsilon)^{N_t n/2}&\frac{\mathrm{Vol}(\mathcal{B}_{N_t n}(\sqrt{nP_x}))}{\mathrm{Vol}(\mathcal{V})}\ \det\left(\boldsymbol{\Psi}^{\frac{-1}{2}}\right)\\ &= (1+\delta)2^{-N_t n\left(-\frac{1}{N_t n}\log\left(\frac{\mathrm{Vol}(\mathcal{B}_{N_t n}(\sqrt{nP_x}))}{\mathrm{Vol}(\mathcal{V})}\right)+\xi\right)},\end{aligned} \tag{5.27}$$

where

$$\begin{aligned}\xi &\triangleq \frac{-1}{2}\log(1+\epsilon) - \frac{1}{N_t}R - \frac{1}{2N_t n}\log\det(\boldsymbol{\Psi}^{-1})\\ &= \frac{-1}{2}\log(1+\epsilon) - \frac{1}{N_t}R - \frac{1}{2N_t}\log\det\left(\mathbb{E}\big[\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_tP_w}\boldsymbol{H}^T\boldsymbol{H}\big)^{-1}\big]\right).\end{aligned} \tag{5.28}$$

From (2.8), since the lattice Λ is good for covering, the first term of the exponent in (5.27) vanishes. From (5.27), $\mathbb{P}_e \to 0$ as $n \to \infty$ whenever $\xi > 0$. A sufficient condition for positive ξ is

$$R < \frac{-1}{2}\log\det\left(\mathbb{E}\big[\big(\frac{P_x}{P_x+N_tP_s}\boldsymbol{I}_{N_t} + \frac{P_x}{N_tP_w}\boldsymbol{H}^T\boldsymbol{H}\big)^{-1}\big]\right) - \frac{1}{2}\log(1+\epsilon) - \epsilon',$$

where ϵ, ϵ' are positive numbers that can be made arbitrarily small by increasing n. The existence of at least one lattice $\mathcal{C}_i \in \mathcal{C}$ that achieves (5.25) is straightforward. For the coarse lattice, any

covering-good lattice from $\mathcal{C}$ with second moment P_x can be picked, e.g., a pair of self-similar lattices can be used for the coarse and fine lattices.

In the event of successful decoding, from (5.18) the outcome of the decoding process would be $\hat{\boldsymbol{t}} = \boldsymbol{t} + \boldsymbol{\lambda}$. On applying the modulo-Λ operation on $\hat{\boldsymbol{t}}$,

$$[\hat{\boldsymbol{t}}] \operatorname{mod} \Lambda \;=\; [\boldsymbol{t} + \boldsymbol{\lambda}] \operatorname{mod} \Lambda \;=\; \boldsymbol{t}, \tag{5.29}$$

where the second equality follows from (2.6) since $\boldsymbol{\lambda} \in \Lambda$. This concludes the proof for real-valued channels.

For complex-valued channels, we follow in the footsteps of Theorem 2, therefore only a sketch of the proof is provided. With a slight abuse of notation, we denote the complex-valued elements by a superscript $\tilde{}$.

Encoding: Since the channel is complex-valued, *two* independent codewords are selected from the same nested lattice code $\tilde{\Lambda}_1 \supseteq \tilde{\Lambda}$ where $\tilde{\Lambda}$ has a second moment $P_x/2$. The transmitted signal is the combination of the two lattice codewords

$$\tilde{\boldsymbol{x}} = \big[\tilde{\boldsymbol{t}}^{(R)} - \tilde{\boldsymbol{s}}^{(R)} - \tilde{\boldsymbol{d}}^{(R)}\big] \operatorname{mod} \tilde{\Lambda} + j\big[\tilde{\boldsymbol{t}}^{(I)} - \tilde{\boldsymbol{s}}^{(I)} - \tilde{\boldsymbol{d}}^{(I)}\big] \operatorname{mod} \tilde{\Lambda}, \tag{5.30}$$

with $\mathbb{E}\big[||\tilde{\boldsymbol{x}}||^2\big] \leq nP_x$, and the dithers $\tilde{\boldsymbol{d}}^{(R)}, \tilde{\boldsymbol{d}}^{(I)}$ are independent.

Decoding: The equalization matrix at time i is given by

$$\tilde{\boldsymbol{U}}_i = \big(\frac{P_x}{N_t} + P_s\big)\big(\big(\frac{P_x}{N_t} + P_s\big)\tilde{\boldsymbol{H}}_i \tilde{\boldsymbol{H}}_i^H + P_w \boldsymbol{I}_{N_t n}\big)^{-1} \tilde{\boldsymbol{H}}_i. \tag{5.31}$$

Following MMSE equalization and dither removal similar to (5.18), the real and imaginary equivalent channels at the receiver are as follows

$$\tilde{\boldsymbol{y}}'^{(R)} = \tilde{\boldsymbol{t}}^{(R)} + \boldsymbol{\lambda}' + \tilde{\boldsymbol{z}}^{(R)} \quad , \quad \tilde{\boldsymbol{y}}'^{(I)} = \tilde{\boldsymbol{t}}^{(I)} + \boldsymbol{\lambda}'' + \tilde{\boldsymbol{z}}^{(I)}, \tag{5.32}$$

where $\boldsymbol{\lambda}'$,$\boldsymbol{\lambda}'' \in \tilde{\Lambda}$. $\tilde{\boldsymbol{z}}^{(R)}, \tilde{\boldsymbol{z}}^{(I)}$ are the equivalent noise components over the real and imaginary channels, respectively. Hence, the complex-valued channel is transformed to two parallel real-valued channels at the receiver, where the decoder recovers the lattice points $\tilde{\boldsymbol{t}}^{(R)}$ and $\tilde{\boldsymbol{t}}^{(I)}$

independently over the real and imaginary domains using the following decision regions

$$\Omega^{(R)} = \Omega^{(I)} \triangleq \left\{ \boldsymbol{v} \in \mathbb{R}^{N_t n} \,:\, ||\boldsymbol{v}||^2 \leq (1+\epsilon)\frac{n}{2} \operatorname{tr}\Big(\mathbb{E}\Big[\big(\frac{1}{\frac{P_x}{N_t} + P_s} \boldsymbol{I}_{N_t} + \frac{1}{P_w} \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}} \big)^{-1} \Big] \Big) \right\}, \quad (5.33)$$

where reliable rates can be achieved on each channel as long as

$$\tilde{R}^{(R)} = \tilde{R}^{(I)} < \left(\frac{-1}{2} \log\det \left(\mathbb{E}\big[\big(\frac{P_x}{P_x + N_t P_s} \boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_w} \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}} \big)^{-1} \big] \right) \right)^+, \quad (5.34)$$

and $\tilde{R} \triangleq \tilde{R}^{(R)} + \tilde{R}^{(I)}$ is the achievable rate for the complex-valued channel given in (5.16).

Finally, utilizing Lemma 5, it can be shown that a Euclidean lattice decoder achieves the same (or less) error probability under the same transmission rate. This concludes the proof of Theorem 11. □

We compute bounds on the gap between the outer and inner bounds of the capacity in (5.3) and (5.16). The gap results are computed under different antenna configurations and fading distributions. For convenience we define the SNR per transmit antenna $\rho \triangleq \frac{P_x}{N_t P_w}$.

Corollary 8. *For the ergodic fading dirty paper channel given in* (5.1)*, the gap between the lattice coding inner bound in* (5.16) *and the outer bound of the capacity in* (5.3) *for any* $P_s > 0$ *is upper bounded by*

- $N_r \geq N_t$ *and* $\rho \geq 1$*: For any channel for which all elements of* $\mathbb{E}\big[(\boldsymbol{H}^H \boldsymbol{H})^{-1}\big] < \infty$

$$\Delta < \log\det \Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\boldsymbol{H}^H \boldsymbol{H}] \big) \mathbb{E}\big[(\boldsymbol{H}^H \boldsymbol{H})^{-1} \big] \Big). \quad (5.35)$$

- $N_r > N_t$ *and* $\rho \geq 1$*: When the elements of* $\boldsymbol{H}$ *are i.i.d. complex Gaussian with zero mean and unit variance,*

$$\Delta < N_t \log\big(1 + \frac{N_t + 1}{N_r - N_t}\big). \quad (5.36)$$

- $N_r = N_t = 1$*: Under Nakagami-m fading with* $m > 1$ *and* $\mathbb{E}[|h|^2] = 1$*,*

$$\Delta < 1 + \log\big(1 + \frac{1}{m-1}\big). \quad (5.37)$$

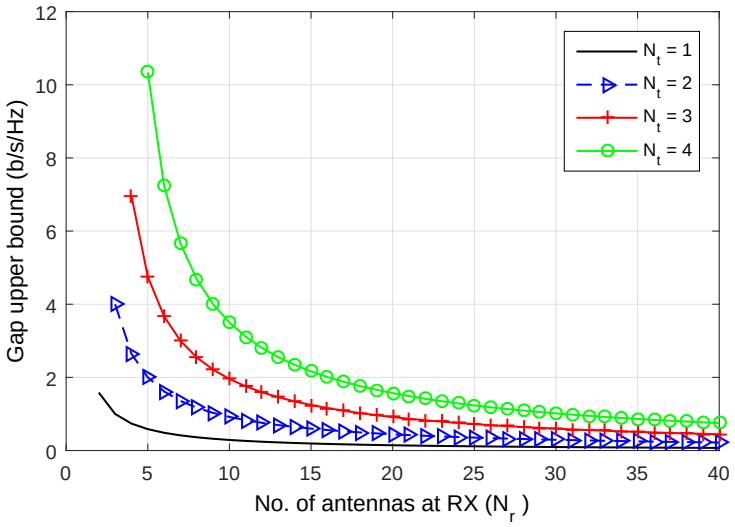

Figure 5.1. The upper bound on the gap to capacity of the MIMO Dirty paper channel vs. N_r, for $\rho \geq 1$.

- $N_r = N_t = 1$*: Under Rayleigh fading with* $\mathbb{E}[|h|^2] = 1$,

$$\Delta < 1.48 + \log\big(\log(1+\kappa)\big), \tag{5.38}$$

where $\kappa \triangleq \max\{\frac{P_x}{P_w}, \frac{P_s}{P_w}, 1\}$.

Proof. See Appendix C.2. □

Under Rayleigh fading with $M = N = 1$ the gap expression in (5.38) varies with P_x, P_s. However, it can be shown that $\lim_{P_x\to\infty} \frac{\Delta}{C} \to 0$ for any fixed ratio $\frac{P_x}{P_s}$. When $N_t < N_r$, Δ is independent of P_x, P_s, and also vanishes when $N_r \gg N_t$ even at finite P_x. This result implies that lattice coding and decoding- along with a channel independent decision rule- approach the capacity of the Rayleigh-fading MIMO channel with $N_r > N_t$.

In Figure 5.1, a bound on the gap to capacity is plotted for various antenna configurations, subject to $\rho \geq 1$. The gap bound diminishes as N_r grows beyond N_t. For the square MIMO channel with $N_r = N_t = 2$, the throughput of the proposed lattice scheme is plotted in Figure 5.2. The gap to capacity is also plotted, which saturates as the SNR increases. For single-antenna

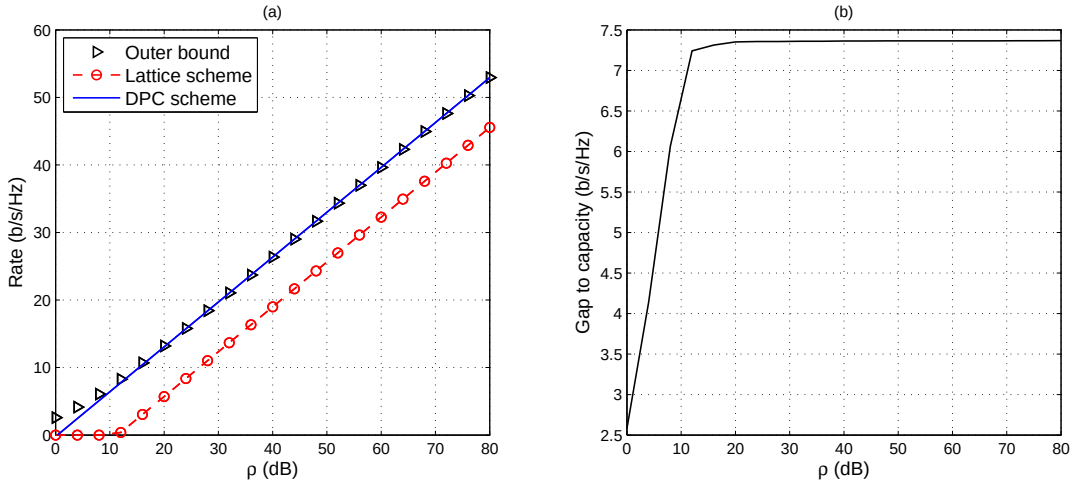

Figure 5.2. (a) Rates achieved by lattice codes vs. dirty paper coding vs. outer bound under Rayleigh fading with $N_r = N_t = 2$ and $P_s = 80\,\text{dB}$. (b) The gap to capacity of the lattice scheme.

nodes, the rates achieved using lattice codes are compared with dirty paper coding rates in (5.5). The capacity outer bound in (5.3) is plotted under Nakagami-m fading with $m = 2$ and Rayleigh fading in Figure 5.3 and Figure 5.4, respectively. While the differences between the rates achieved are notable under Rayleigh fading, they are insignificant under Nakagami-m fading, which again suggests the point-to-point capacity is almost realized under this scenario.

5.4 Fading Broadcast Channel with CSIR

We first consider a two-user broadcast channel where the channel coefficients of Receiver 1 are quasi-static, and that of Receiver 2 are stationary and ergodic. The transmitter and the two receivers have N_t, N_{r_1}, N_{r_2} antennas, respectively. The received signals are given by

$$\begin{aligned} \boldsymbol{y}_{1,i} =& \boldsymbol{G}\boldsymbol{x}_{1,i} + \boldsymbol{G}\boldsymbol{x}_{2,i} + \boldsymbol{w}_{1,i} \\ \boldsymbol{y}_{2,i} =& \boldsymbol{H}_i\boldsymbol{x}_{1,i} + \boldsymbol{H}_i\boldsymbol{x}_{2,i} + \boldsymbol{w}_{2,i}, \end{aligned} \tag{5.39}$$

Each receiver has its own CSIR, but not global CSIR. The transmitter power constraint for the two signals is $\mathbb{E}[||\boldsymbol{x}_1||^2] \leq n\alpha P_x$ and $\mathbb{E}[||\boldsymbol{x}_2||^2] \leq n(1-\alpha)P_x$, with $\alpha \in [0,1]$ and n is the time duration of each codeword. The noise terms $\boldsymbol{w}_1, \boldsymbol{w}_2$ are zero mean i.i.d. circularly-symmetric

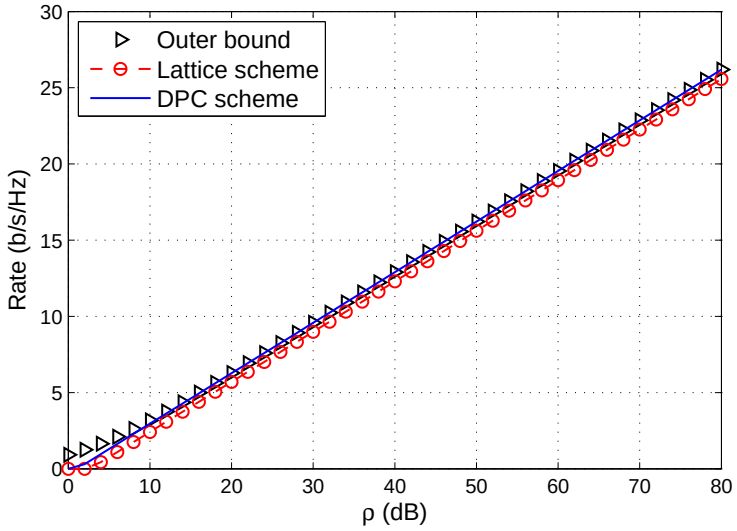

Figure 5.3. Rates achieved using lattice codes vs. dirty paper coding vs. outer bound under Nakagami-m fading with $m = 2$ and $P_s = 80\,\text{dB}$.

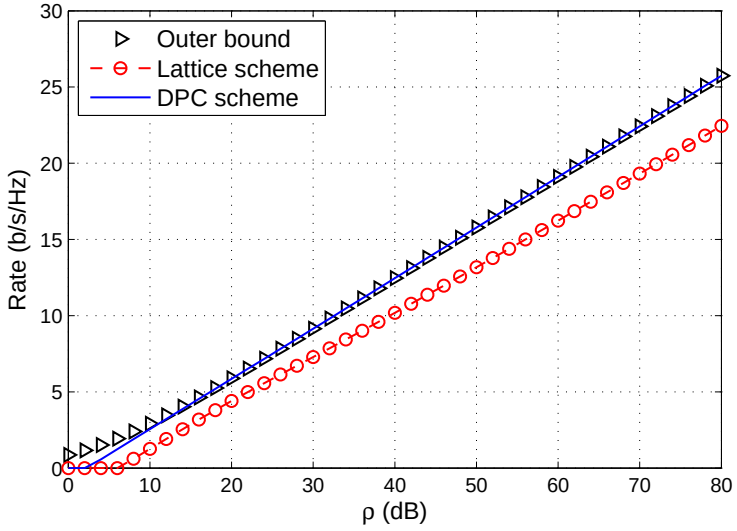

Figure 5.4. Rates achieved using lattice codes vs. dirty paper coding vs. outer bound under Rayleigh fading and $P_s = 80\,\text{dB}$.

complex Gaussian with variances P_{w_1}, P_{w_2}, respectively. A set of achievable rates for this channel under lattice coding and decoding are as follows

Theorem 12. *For the two-user broadcast channel given in* (5.39)*, lattice coding and decoding achieve*

$$R_1 < \log\det\Big(\boldsymbol{I}_{N_t} + \frac{\alpha P_x}{N_t}\boldsymbol{G}^H\boldsymbol{\Phi}^{-1}\boldsymbol{G}\Big) \tag{5.40}$$

$$R_2 < \Big(-\log\det\Big(\frac{1}{1-\alpha}\mathbb{E}\big[\big(\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_{w_2}}\boldsymbol{H}^H\boldsymbol{H}\big)^{-1}\big]\Big)\Big)^+, \tag{5.41}$$

where $\boldsymbol{\Phi} \triangleq \frac{(1-\alpha)P_x}{N_t}\boldsymbol{G}\boldsymbol{G}^H + P_{w_1}\boldsymbol{I}_{N_{r_1}}$.

Proof. Receiver 1*:* The transmitter emits a superposition of two codewords, i.e., $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, where Receiver 1 decodes $\boldsymbol{x}_1$ while treating $\boldsymbol{x}_2$ as noise. Hence, with respect to Receiver 1, the channel is a special case of the dirty paper channel with $P_s = 0$, and colored noise given by $\boldsymbol{G}\boldsymbol{x}_{2,i} + \boldsymbol{w}_{1,i}$. The equalization matrix is then time invariant, given by

$$\boldsymbol{U}_1 = \frac{\alpha P_x}{N_t}\big(\frac{P_x}{N_t}\boldsymbol{G}\boldsymbol{G}^H + \boldsymbol{I}_{N_{r_1}}\big)^{-1}\boldsymbol{G}\,. \tag{5.42}$$

Since the channel is fixed, the decision region becomes ellipsoidal, given by

$$\Omega \triangleq \Big\{\boldsymbol{v} \in \mathbb{R}^{N_t n} \,:\, \boldsymbol{v}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{v} \le (1+\epsilon)n\Big\}, \tag{5.43}$$

where $\boldsymbol{\Sigma}_1$ is an $N_t n \times N_t n$ block-diagonal matrix whose diagonal blocks are equal, and given by

$$\boldsymbol{\Sigma}_{1,i} \triangleq \frac{\alpha P_x}{N_t}\big(\boldsymbol{I}_{N_t} + \frac{\alpha P_x}{N_t}\boldsymbol{G}^H\boldsymbol{\Phi}^{-1}\boldsymbol{G}\big)^{-1}. \tag{5.44}$$

Following in the footsteps of the proof of Theorem 10, it can be shown that R_1 satisfies (5.40).

Receiver 2*:* Since $\boldsymbol{x}_1$ is known non-causally at the transmitter, the channel between the transmitter and Receiver 2 is equivalent to an ergodic fading dirty paper channel with $P_s = \alpha P_x$, where R_2 is given by (5.41) from Theorem 11. The remainder of the rate region is obtained by varying α. □

In the absence of channel state information at the transmitter, the capacity of the fading MIMO broadcast channel is not known in general. In Figure 5.5 we compare the rate region of the lattice coding scheme to the time-sharing inner bound as well as a version of Costa's dirty paper coding under non-causal channel state information at the transmitter and white-input covariance. In this non-causal scheme, Receiver 1 decodes its message while treating $\boldsymbol{x}_2$ as noise. Since $\{\boldsymbol{H}_i \boldsymbol{x}_{1,i}\}_{i=1}^n$ are known non-causally at the transmitter, dirty paper coding allows the rate for Receiver 2 to be unaffected by interference. The rate region is then given by[5]

$$\begin{aligned} \check{R}_1 < & \log\det\Big(\boldsymbol{I}_{N_t} + \frac{\alpha P_x}{N_t}\boldsymbol{G}^H \boldsymbol{\Phi}^{-1}\boldsymbol{G}\Big) \\ \check{R}_2 < & \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_t} + \frac{(1-\alpha)P_x}{N_t P_{w_2}}\boldsymbol{H}^H\boldsymbol{H}\big)\Big]. \end{aligned} \tag{5.45}$$

In Figure 5.5 we compute the rates through Monte-Carlo simulations when $N_t = N_{r_1} = 2, N_{r_2} = 4$, and the channel of Receiver 2 is Rayleigh faded.[6] For the special case of single-antenna nodes, the rates of the lattice coding scheme are plotted in Figure 5.6 and compared with the time sharing inner bound as well as the Tuninetti-Shamai rate region for the two-user fading broadcast channel [40].[7] The results are also compared to the white-input capacity when the channel state is available at all nodes [30]. For the single-antenna case we assume the channel of Receiver 1 has unit gain, i.e., $|g| = 1$. Note that unlike both [40, 30], the proposed lattice scheme presumes each receiver is oblivious to the codebook designed for the other receiver.

In addition, we study the two-user broadcast channel with CSIR, where the fading processes of the two users are stationary, ergodic and independent of each other, as follows

$$\begin{aligned} \boldsymbol{y}_{1,i} = & \boldsymbol{H}_{1,i}\boldsymbol{x}_{1,i} + \boldsymbol{H}_{1,i}\boldsymbol{x}_{2,i} + \boldsymbol{w}_{1,i} \\ \boldsymbol{y}_{2,i} = & \boldsymbol{H}_{2,i}\boldsymbol{x}_{1,i} + \boldsymbol{H}_{2,i}\boldsymbol{x}_{2,i} + \boldsymbol{w}_{2,i}, \end{aligned} \tag{5.46}$$

[5]It is unknown whether the rate region in (5.45) is an outer bound for the capacity region of the channel in (5.39).

[6]Jafar and Goldsmith [43] showed that for isotropic fading broadcast channels with single-antenna receivers, increasing N_t beyond one does not increase the capacity under CSIR only. Therefore, we focus in our simulations on cases where $N_t \leq \min\{N_{r_1}, N_{r_2}\}$.

[7]The authors of [40] conjecture that their inner bound is tight.

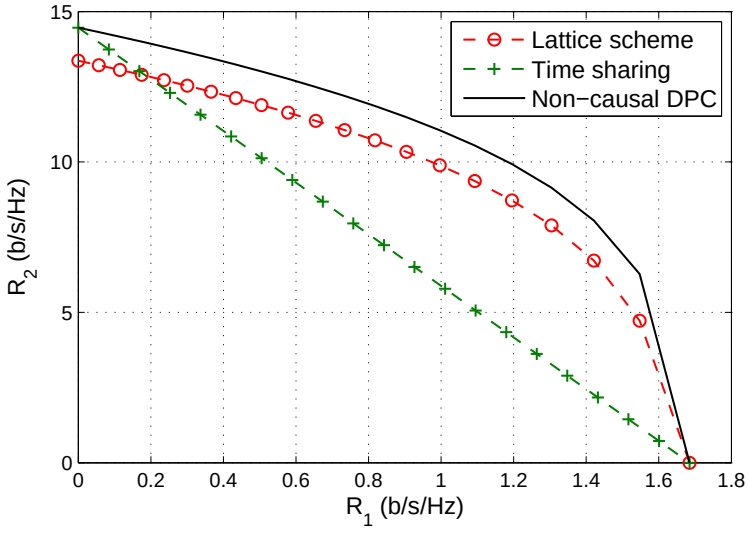

Figure 5.5. Rate regions for the broadcast channel with $N_t = N_{r_1} = 2$ and $N_{r_2} = 4$, where $\frac{P_x}{P_{w_1}} = 0\,\text{dB}$ and $\frac{P_x}{P_{w_2}} = 20\,\text{dB}$. Fading of second user is Rayleigh.

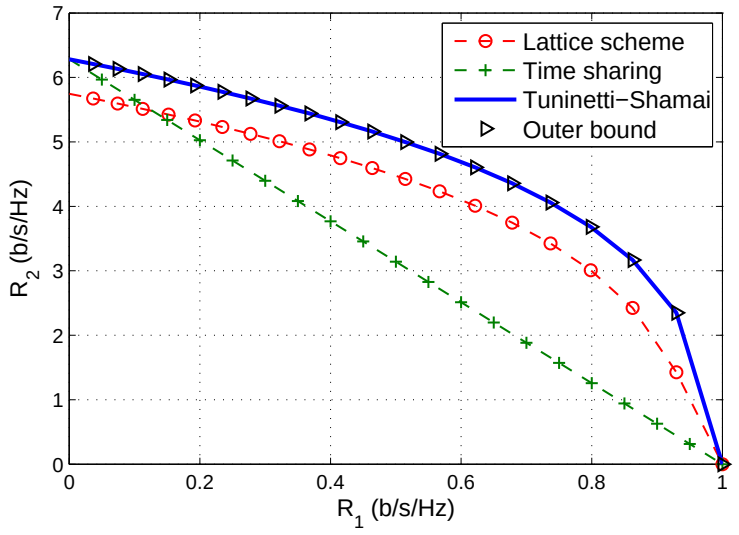

Figure 5.6. Rate regions for the SISO broadcast channel, where $\frac{P_x}{P_{w_1}} = 0\,\text{dB}$ and $\frac{P_x}{P_{w_2}} = 20\,\text{dB}$. Fading of second user is Nakagami-m with $m = 2$.

Theorem 13. *For the two-user broadcast channel given in* (5.46)*, lattice coding and decoding achieve*

$$R_1 < \Big(-\log\det\Big(\mathbb{E}\big[\big(\boldsymbol{I}_M + \frac{\alpha P_x}{N_t P_{w_1}}\boldsymbol{H}_1^H \boldsymbol{\Phi}^{-1}\boldsymbol{H}_1\big)^{-1}\big]\Big)\Big)^+, \tag{5.47}$$

$$R_2 < \Big(-\log\det\Big(\frac{1}{1-\alpha}\mathbb{E}\big[\big(\boldsymbol{I}_{N_t} + \frac{P_x}{N_t P_{w_2}}\boldsymbol{H}_2^H \boldsymbol{H}_2\big)^{-1}\big]\Big)\Big)^+, \tag{5.48}$$

where $\boldsymbol{\Phi} \triangleq \frac{(1-\alpha)P_x}{N_t}\boldsymbol{H}_1\boldsymbol{H}_1^H + P_{w_1}\boldsymbol{I}_{N_{r_1}}$.

Proof. The achievability proof of the rate of Receiver 2 in (5.48) is identical to that of Theorem 12. At Receiver 1, the received signal is multiplied by a time-varying equalization matrix, given by

$$\boldsymbol{U}_{1,i} = \frac{\alpha P_x}{N_t}\big(\frac{P_x}{N_t}\boldsymbol{H}_{1,i}\boldsymbol{H}_{1,i}^H + \boldsymbol{I}_{N_{r_1}}\big)^{-1}\boldsymbol{H}_{1,i}\,, \tag{5.49}$$

with spherical decision region as follows

$$\Omega \triangleq \Big\{\boldsymbol{v} \in \mathbb{R}^{N_t n} \,:\, ||\boldsymbol{v}||^2 \le (1+\epsilon)n\frac{\alpha P_x}{N_t}\,\mathrm{tr}\Big(\mathbb{E}\Big[\big(\boldsymbol{I}_{N_t} + \frac{\alpha P_x}{N_t}\boldsymbol{H}_1^H\boldsymbol{\Phi}^{-1}\boldsymbol{H}_1\big)^{-1}\Big]\Big)\Big\}. \tag{5.50}$$

The remainder of the analysis resembles that in the proof of Theorem 12, where it can be shown that (5.47) is achievable. □

The rate region of the lattice scheme is plotted in Figure 5.7 under Nakagami fading with $N_t = 1$ and $N_{r_1} = N_{r_2} = 2$, and compared with time sharing and dirty paper coding with non-causal CSIT.

5.5 Broadcast Channel with CSIT

The two-user SISO ergodic broadcast channel under lattice codes is studied, when channel-state information is available *causally* at all communication nodes. As a byproduct, we show that lattice codes achieve the capacity of the (non-fading) two-user AWGN broadcast channel. For simplicity, only the real-valued case is addressed.

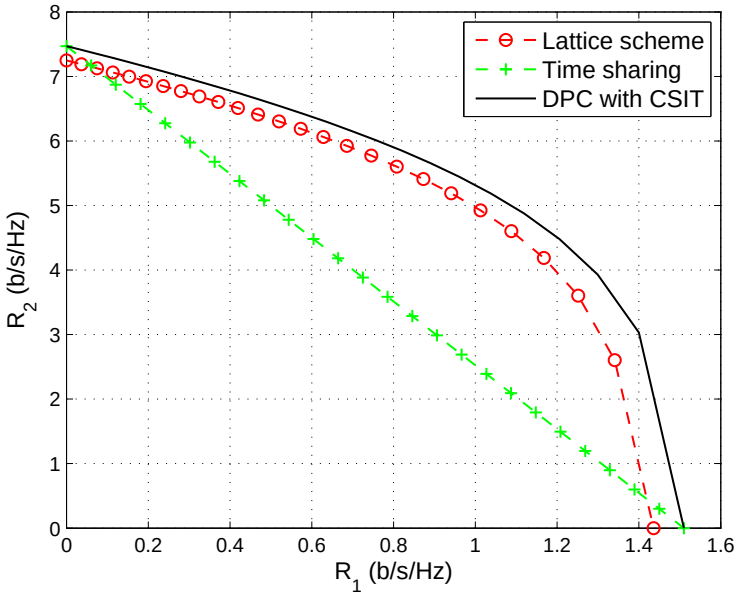

Figure 5.7. Rate regions for the MIMO broadcast channel with $N_t = 1$ and $N_{r_1} = N_{r_2} = 2$, where $\frac{P_x}{P_{w_1}} = 0\,\text{dB}$ and $\frac{P_x}{P_{w_2}} = 20\,\text{dB}$. Both users experience Nakagami-m fading with $m = 2$.

We consider the case where only a private message is intended for each receiver. The transmitter emits the sum of two lattice codewords, and hence the received signals at each receiver are given by

$$\begin{aligned} \boldsymbol{y}_1 =& \boldsymbol{H}_1\boldsymbol{x}_1 + \boldsymbol{H}_1\boldsymbol{x}_2 + \boldsymbol{w}_1 \\ \boldsymbol{y}_2 =& \boldsymbol{H}_2\boldsymbol{x}_1 + \boldsymbol{H}_2\boldsymbol{x}_2 + \boldsymbol{w}_2, \end{aligned} \tag{5.51}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ denote the signals intended at receivers 1 and 2, respectively. In order to fulfill the power constraint, we assume $\mathbb{E}[||\boldsymbol{x}_1||^2] \leq n\alpha\rho$ and $\mathbb{E}[||\boldsymbol{x}_2||^2] \leq n(1-\alpha)\rho$. $\boldsymbol{H}_1$, $\boldsymbol{H}_2$ are diagonal matrices representing the channel coefficients across time. The vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are zero-mean i.i.d. Gaussian noise vectors each with covariance $P_w\boldsymbol{I}_n$, and independent of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$.

5.5.1 AWGN Broadcast Channel

For the AWGN broadcast channel, the channel gains are time invariant, and hence the channel matrices simplify to $\boldsymbol{H}_1 = h_1\boldsymbol{I}_n$ and $\boldsymbol{H}_2 = h_2\boldsymbol{I}_n$. Assume without loss of generality that $|h_2| \leq |h_1|$.

The capacity region of the two-user Gaussian broadcast channel is given by [67]

$$\begin{aligned} R_1 <& \frac{1}{2}\log\left(1+\alpha\rho h_1^2\right) \\ R_2 <& \frac{1}{2}\log\left(1+\frac{(1-\alpha)\rho h_2^2}{\alpha\rho h_1^2+1}\right). \end{aligned} \tag{5.52}$$

Theorem 14. *Lattice codes achieve the capacity of the two-user Gaussian broadcast channel with full channel state information.*

Proof. Encoding: The transmitter emits the sum of two lattice codewords as follows

$$\boldsymbol{x} = [\boldsymbol{t}_1 - \boldsymbol{d}_1] \bmod \Lambda \; + \; [\boldsymbol{t}_2 - \boldsymbol{d}_2] \bmod \Lambda', \tag{5.53}$$

where $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are two lattice points drawn from $\Lambda_1 \supseteq \Lambda$ and $\Lambda_2 \supseteq \Lambda'$, respectively. The second moment of the lattices Λ and Λ' are $\alpha\rho$ and $(1-\alpha)\rho$, respectively. The dithers $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are independent and uniform over $\mathcal{V}$ and $\mathcal{V}'$. Similar to the MAC, Λ_1 and Λ_2 are not necessarily independent.

Decoding (User 2): User 2 uses MMSE scaling to decode its desired signal, while treating the other signal as noise. The (time-invariant) MMSE coefficient is given by

$$u_2 = \frac{(1-\alpha)\rho h_2}{\rho h_2^2+1}. \tag{5.54}$$

The scaled signal can thus be written as

$$\boldsymbol{y}_2' = x_2 + z_2,$$

where

$$\boldsymbol{z}_2 = -\frac{\alpha\rho h_2^2+1}{\rho h_2^2+1}\boldsymbol{x}_1 + \frac{(1-\alpha)\rho h_2^2}{\rho h_2^2+1}\boldsymbol{x}_2 + \frac{(1-\alpha)\rho h_2}{\rho h_2^2+1}\boldsymbol{w}_2. \tag{5.55}$$

We apply the ambiguity decoder defined by a spherical decision region Ω_2 as follows

$$\Omega_2 \triangleq \left\{ \boldsymbol{v} \in \mathbb{R}^n \ : \ \boldsymbol{v}^T \boldsymbol{v} \leq (1+\epsilon) n(1-\alpha)\rho \frac{\alpha\rho h_2^2 + 1}{\rho h_2^2 + 1} \right\}, \tag{5.56}$$

where ϵ is an arbitrary positive constant.

Probability of error (User 2): There exists a nested lattice code with rate R_2 whose probability of error can be bounded by

$$P_e < P(\boldsymbol{z} \notin \Omega_1) + (1+\delta) 2^{nR_2} \frac{\mathrm{Vol}(\Omega_1)}{\mathrm{Vol}(\mathcal{V}')}, \tag{5.57}$$

for any $\delta > 0$. Following in the footsteps of Lemma 7, $P(\boldsymbol{z} \notin \Omega_1) < \gamma$ for any $\gamma > 0$. Consequently, the error probability can be bounded by

$$\mathbb{P}_e < \gamma + (1+\delta) 2^{nR_2} \frac{\mathrm{Vol}(\Omega_2)}{\mathrm{Vol}(\mathcal{V}')}, \tag{5.58}$$

for any $\gamma, \delta > 0$. The volume of Ω_2 is given by

$$\mathrm{Vol}(\Omega_2) = (1+\epsilon)^{\frac{n}{2}} \big(\frac{\alpha\rho h_2^2 + 1}{h_2^2 \rho + 1} \big)^{\frac{n}{2}} \mathrm{Vol}\big(\mathcal{B}(\sqrt{n(1-\alpha)\rho}) \big). \tag{5.59}$$

The second term in (5.58) is bounded by

$$\begin{aligned} &(1+\delta) 2^{nR_2} (1+\epsilon)^{n/2} \Big(\frac{\alpha\rho h_2^2 + 1}{\rho h_2^2 + 1} \Big)^{\frac{n}{2}} \frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n(1-\alpha)\rho}))}{\mathrm{Vol}(\mathcal{V}')} \\ &= (1+\delta) 2^{-n \Big(-\frac{1}{n} \log \big(\frac{\mathrm{Vol}(\mathcal{B}(\sqrt{n(1-\alpha)\rho}))}{\mathrm{Vol}(\mathcal{V}')} \big) + \xi \Big)}, \end{aligned} \tag{5.60}$$

where

$$\xi \triangleq \frac{-1}{2} \log(1+\epsilon) - \frac{1}{2} \log \Big(\frac{\alpha\rho h_2^2 + 1}{\rho h_2^2 + 1} \Big) - R_2. \tag{5.61}$$

From (2.8), since the lattice Λ' is good for covering, the first term of the exponent in (5.60) vanishes. Then, whenever ξ is a positive constant we have $\lim_{n\to\infty} \mathbb{P}_e = 0$. Since ϵ is arbitrary, positive ξ can be achieved when

$$R_2 < \frac{1}{2} \log \big(1 + \frac{(1-\alpha)\rho h_2^2}{\rho h_2^2 + 1} \big). \tag{5.62}$$

Decoding (User 1): The condition $|h_2| \leq |h_1|$ guarantees that User 1 can successfully decode any signal that User 2 has decoded, since

$$\frac{1}{2}\log\big(1+\frac{(1-\alpha)\rho h_2^2}{\rho h_2^2+1}\big) \leq \frac{1}{2}\log\big(1+\frac{(1-\alpha)\rho h_1^2}{\rho h_1^2+1}\big). \tag{5.63}$$

This is always true because the expression in (5.62) is non-decreasing with $|h|$. Hence, User 1 can follow the same steps as User 2 to decode $\boldsymbol{x}_2$, and then subtract it. The equivalent channel is then $\boldsymbol{y}_1'' = h_1\boldsymbol{x}_1 + \boldsymbol{w}_1$. User 1 can then scale $\boldsymbol{y}_1''$ by an MMSE coefficient

$$u_1 = \frac{\alpha\rho h_1}{\alpha\rho h_1^2+1}, \tag{5.64}$$

and then decode using an ambiguity decoder with a spherical decision region

$$\Omega_1 \triangleq \left\{\boldsymbol{v}\in\mathbb{R}^n \;:\; \boldsymbol{v}^T\boldsymbol{v} \leq (1+\epsilon)n\alpha\rho\frac{1}{\alpha\rho h_1^2+1}\right\}, \tag{5.65}$$

where it can be shown that $\boldsymbol{x}_1$ can be decoded reliably as long as

$$R_1 < \frac{1}{2}\log\big(1+\alpha\rho h_1^2\big).$$

This concludes the proof of Theorem 14. □

Note that the aforementioned achievable scheme can be extended to the K-user AWGN broadcast channel. This is done through properly ordering the channel strengths of the users such that each receiver can decode and subtract all the signals intended for the users with weaker channel strengths.

5.5.2 Ergodic Broadcast Channel with CSIT

We study the two-user ergodic broadcast channel with full channel state information at all nodes. The problem is different from the AWGN case since the relative channel magnitudes of the two users alternate across different channel realizations. The capacity region of this class of channels was established by Li and Goldsmith in [30] as follows

$$\begin{aligned} R_1 <& \frac{1}{2}\mathbb{E}\Big[\log\big(1+\frac{\alpha\rho h_1^2}{(1-\alpha)\rho h_1^2 I(h_2\geq h_1)+1}\big)\Big] \\ R_2 <& \frac{1}{2}\mathbb{E}\Big[\log\big(1+\frac{(1-\alpha)\rho h_2^2}{\alpha\rho h_2^2 I(h_2< h_1)+1}\big)\Big], \end{aligned} \tag{5.66}$$

where $I(.)$ is an indicator function such that $I(\upsilon) = 1$ if the argument υ is true and zero otherwise.

Theorem 15. *The capacity region of the two-user ergodic broadcast channel with full channel state information can be achieved using lattice codes.*

Proof. The argument broadly follows the same lines as [30]. We briefly sketch the achievability proof for completeness, which is based on *separable coding*, i.e., coding independently across the different fading states. This is done through discretizing the range of the time-varying channel of each user into κ states, and hence we have $L \triangleq \kappa^2$ joint channel states (Generally κ^K states for K users). The states are denoted by $S_0, \ \ldots, \ S_{L-1}$. In each joint state, the channel is approximated by a non-fading broadcast channel, whose capacity can be achieved as shown in Section 5.5.1. Hence, allowing sufficiently many channel states (large κ) in addition to using the lattice scheme for the Gaussian broadcast channel over each joint state yields the rates given in (5.66). □

CHAPTER 6

ERGODIC STRONG INTERFERENCE CHANNEL

Recently, lattice codes have been proposed to replace Gaussian codes for interference channels. In [22] it has been shown that a naive extension of the Han-Kobayashi scheme with Gaussian codes is suboptimal for the Gaussian interference channel with more than two users, and that linear codes (more specifically lattice codes) outperform Gaussian codes in such case. Following this result, Jafar and Vishwanath [23] showed that the *generalized* Degrees-of-Freedom of the K-user symmetric Gaussian interference channel is achieved via lattice codes. Subsequently, Ordentlich et al. [68] derived an achievable region for the K-user symmetric Gaussian interference using nested lattice codes. In [47] Sankar et al. considered the two-user ergodic fading interference channel with channel state information available at all nodes, where Gaussian codes were shown to achieve the sum capacity under some scenarios, namely the *ergodic very strong* (a mix of strong and weak fading states satisfying specific strict fading average conditions) and *uniformly strong*, (every cross fading coefficient is larger in magnitude than the direct fading coefficient) interference regimes. Farsani [48] also studied a variant of the latter problem with partial channel state information at the transmitters.

In this chapter, the two-user ergodic strong interference channel is defined, and its capacity is computed using Gaussian codebooks. Then, an achievable rate region under lattice coding and decoding is derived for this channel and compared to capacity.

6.1 Capacity Region

Consider a two-user Gaussian interference channel with received signals at time instant i given by

$$\begin{aligned} y_{1,i} =& h_{11,i}x_{1,i} + h_{12,i}x_{2,i} + w_{1,i}, \\ y_{2,i} =& h_{21,i}x_{1,i} + h_{22,i}x_{2,i} + w_{2,i}, \end{aligned} \tag{6.1}$$

where in general $h_{kj,i} \in \mathbb{R}$ denotes the coefficient of the channel between transmitter j and receiver k at time i, which are stationary and ergodic with time-varying gain. Channel coefficients across different links are independent of each other, and are only known at the receivers. $x_{j,i} \in \mathbb{R}$ is the symbol sent from transmitter j at time i, where the codeword $\boldsymbol{x}_j \triangleq [x_{j,1}, \ldots, x_{j,n}]^T$ is transmitted throughout n channel uses and satisfies $\mathbb{E}[||\boldsymbol{x}_j||^2] \leq nP_j$. The noise $\boldsymbol{w}_j \in \mathbb{R}^n$ is zero-mean i.i.d. Gaussian with covariance $\boldsymbol{I}_n$ for $j \in \{1, 2\}$.

Definition 1. *A two-user* ergodic strong *interference channel is defined by*

$$\begin{aligned} I(X_1^n; Y_1^n | X_2^n, H) &\leq I(X_1^n; Y_2^n | X_2^n, H), \\ I(X_2^n; Y_2^n | X_1^n, H) &\leq I(X_2^n; Y_1^n | X_1^n, H). \end{aligned} \tag{6.2}$$

for all $\mathbb{P}(X_1^n)\mathbb{P}(X_2^n)$*, where* H *represents the set of all channel coefficients.*

Theorem 16. *The capacity region of the ergodic strong interference channel is given by*

$$\begin{aligned} R_1 &< \mathbb{E}\big[\log(1 + P_1|h_{11}|^2))\big], \\ R_2 &< \mathbb{E}\big[\log(1 + P_2|h_{22}|^2))\big], \\ R_1 + R_2 &< \min\big\{\mathbb{E}\big[\log(1 + P_1|h_{11}|^2 + P_2|h_{12}|^2))\big], \\ &\qquad \mathbb{E}\big[\log(1 + P_1|h_{21}|^2 + P_2|h_{22}|^2))\big]\big\}. \end{aligned} \tag{6.3}$$

Proof. The first two bounds in (6.3) are trivial. Following in the footsteps of [67, Section 6.3], the third bound can be derived as follows

$$\begin{aligned}
n(R_1+R_2) =& h(M_1)+h(M_2) \\
=& h(M_1|H)+h(M_2|H) && (6.4)\\
\leq& I(M_1;Y_1^n,H)+I(M_2;Y_2^n,H)+n\epsilon && (6.5)\\
=& I(X_1^n;Y_1^n,H)+I(X_2^n;Y_2^n,H)+n\epsilon \\
\leq& I(X_1^n;Y_1^n|H)+I(X_2^n;Y_2^n|X_1^n,H)+n\epsilon && (6.6)\\
\leq& I(X_1^n;Y_1^n|H)+I(X_2^n;Y_1^n|X_1^n,H)+n\epsilon && (6.7)\\
=& I(X_1^n,X_2^n;Y_1^n|H)+n\epsilon \\
=& \mathcal{H}(Y_1^n|H)-\mathcal{H}(Y_1^n|X_1^n,X_2^n,H)+n\epsilon \\
=& \mathcal{H}(Y_1^n|H)-\mathcal{H}(W_1^n)+n\epsilon \\
\leq& \sum_{i=1}^{n}\log(1+P_1|h_{11,i}|^2+P_2|h_{12,i}|^2)+n\epsilon, && (6.8)
\end{aligned}$$

where (6.4) follows since channel knowledge is not available at transmitters and (6.5) follows from *Fano's inequality*. (6.6) holds since $I(X_1^n;X_2^n|H)=0$, whereas (6.7) holds from Definition 1. At large n, (6.8) converges to

$$R_1+R_2\leq\mathbb{E}\big[\log(1+P_1|h_{11}|^2+P_2|h_{12}|^2))\big]+\epsilon+\epsilon',$$

where ϵ,ϵ' vanish at large n. The other bound on R_1+R_2 follows similarly by switching the indices in (6.6). This concludes the proof of Theorem 16. □

Obviously, the capacity region in (6.3) is the intersection of that of two multiple-access channels where both messages are required at both receivers. Hence (6.3) is achievable with Gaussian codes, similar to the Gaussian strong interference channel without fading [45]. Although the ergodic interference channel has been studied in [47, 48] under some channel knowledge at the transmitters, capacity regions were only fully characterized for the *uniformly strong* interference

channel, i.e., every fading state is strong, and the *ergodic very strong* interference channel, i.e., a mix of strong and weak fading states satisfying specific fading averaged conditions. The ergodic strong interference channel model studied in this section requires milder conditions than the aforementioned models, yet includes them both as special cases.

6.2 Lattice Coding Inner Bound

Although the capacity region of the the ergodic strong interference channel is achievable using Gaussian codes and typicality (or maximum likelihood) decoding, their significant encoding and decoding complexity hinder their use in practice. Recently, lattice coding and decoding have been shown to achieve the capacity of Gaussian channels [5], opening the door for significant reduction in complexity without compromising communication rates. In this section, a lattice coding and decoding scheme is proposed for the ergodic strong interference channel, and its rate region is computed. We limit our study to the case where the direct and cross fading coefficients are drawn from the same distribution, but are independent of each other, as follows

$$\begin{aligned} y_{1,i} =& \sqrt{\alpha_{11}} h_{11,i} x_{1,i} + \sqrt{\alpha_{12}} h_{12,i} x_{2,i} + w_{1,i}, \\ y_{2,i} =& \sqrt{\alpha} h_{21,i} x_{1,i} + \sqrt{\alpha_{22}} h_{22,i} x_{2,i} + w_{2,i}, \end{aligned} \tag{6.9}$$

where $h_{kj,i}$ are identically distributed but independent of each other for all j, k, and $\alpha_{21} \geq \alpha_{11}$, $\alpha_{12} \geq \alpha_{22}$. This models the case where all transmitters and receivers are in a similar terrain, with the direct links being more distant. For ease of exposition let $\alpha_{11} = \alpha_{22} = 1$ and $\alpha_{12} = \alpha_{21} \triangleq \alpha$. A *sufficient condition* for the channel to satisfy ergodic strong interference in Definition 1 is $\alpha \geq 1$.

Theorem 17. *For the ergodic strong interference channel with complex-valued channels, lattice coding and decoding achieve the following rate region*

$$\begin{aligned} R_1 &< \mathbb{L}(P_1 g_{11}) \\ R_2 &< \mathbb{L}(P_2 g_{22}) \\ \beta_{11} R_1 + \beta_{12} R_2 &< \omega_1 \\ \beta_{21} R_1 + \beta_{22} R_2 &< \omega_2 \,, \end{aligned} \tag{6.10}$$

where $\mathbb{L}(x) \triangleq -\log\left(\mathbb{E}[\frac{1}{1+x}]\right)$, $g_{jk} \triangleq |h_{jk}^2|$, *and*

$$\begin{aligned} \beta_{11} &\triangleq \mathbb{L}(\alpha P_2 g_{12}) - \mathbb{L}(\frac{\alpha P_2 g_{12}}{1+P_1 g_{11}}) \\ \beta_{12} &\triangleq \mathbb{L}(P_1 g_{11}) - \mathbb{L}(\frac{P_1 g_{11}}{1+\alpha P_2 g_{12}}) \\ \beta_{21} &\triangleq \mathbb{L}(P_2 g_{22}) - \mathbb{L}(\frac{P_2 g_{22}}{1+\alpha P_1 g_{21}}) \\ \beta_{22} &\triangleq \mathbb{L}(\alpha P_1 g_{21}) - \mathbb{L}(\frac{\alpha P_1 g_{21}}{1+P_2 g_{22}}) \\ \omega_1 &\triangleq \mathbb{L}(P_1 g_{11})\mathbb{L}(\alpha P_2 g_{12}) - \mathbb{L}(\frac{P_1 g_{11}}{1+\alpha P_2 g_{12}})\mathbb{L}(\frac{\alpha P_2 g_{12}}{1+P_1 g_{11}}) \\ \omega_2 &\triangleq \mathbb{L}(\alpha P_1 g_{21})\mathbb{L}(P_2 g_{22}) - \mathbb{L}(\frac{\alpha P_1 g_{21}}{1+P_2 g_{22}})\mathbb{L}(\frac{P_2 g_{22}}{1+\alpha P_1 g_{21}}) \,. \end{aligned}$$

Proof. Encoding: We first study real-valued channels. Nested lattice codewords are used where at transmitter j, $\Lambda^{(j)} \subseteq \Lambda_1^{(j)}$. Transmitter j emits a lattice point $\boldsymbol{t}_j \in \Lambda_1^{(j)}$ dithered uniformly with $\boldsymbol{d}_j$, as follows

$$\boldsymbol{x}_j = [\boldsymbol{t}_j - \boldsymbol{d}_j] \bmod \Lambda^{(j)} \;=\; \boldsymbol{t}_j - \boldsymbol{d}_j + \boldsymbol{\lambda}_j \,, \quad j = 1, 2, \tag{6.11}$$

where $\boldsymbol{\lambda} = -Q_{\mathcal{V}^{(j)}}(\boldsymbol{t}_j - \boldsymbol{d}_j) \in \Lambda^{(j)}$ from (2.5), and $\Lambda^{(j)} \in \mathbb{R}^n$ has second moment P_j. The dithers $\boldsymbol{d}_j$ are independent of each other and uniform over $\mathcal{V}^{(j)}$, leading to the independence of $\boldsymbol{x}_1, \boldsymbol{x}_2$, according to Lemma 1.

Decoding: First we discuss the received signal at receiver 1 which can be expressed in the form

$$\boldsymbol{y}_1 = \boldsymbol{H}_{11}\boldsymbol{x}_1 + \sqrt{\alpha}\boldsymbol{H}_{12}\boldsymbol{x}_2 + \boldsymbol{w}_1, \tag{6.12}$$

where $\boldsymbol{H}_{1j}$ is a diagonal matrix whose elements are $h_{1j,i}$. The receiver performs successive cancellation decoding, i.e., it decodes and subtracts $\boldsymbol{x}_2$ before decoding $\boldsymbol{x}_1$. At the receiver, $\boldsymbol{y}_1$ is multiplied by a single-tap equalization matrix $\boldsymbol{U}_1 \in \mathbb{R}^{n\times n}$ and the dither is removed as follows

$$\begin{aligned}\boldsymbol{y}_1' \triangleq& \boldsymbol{U}_1^T\boldsymbol{y}_1 + \boldsymbol{d}_2 \\ =& \boldsymbol{x}_2 + (\sqrt{\alpha}\boldsymbol{U}_1^T\boldsymbol{H}_{12} - \boldsymbol{I}_n)\boldsymbol{x}_2 + \boldsymbol{U}_1^T\boldsymbol{H}_{11}\boldsymbol{x}_1 + \boldsymbol{U}_1^T\boldsymbol{w}_1 + \boldsymbol{d}_2 \\ =& \boldsymbol{t}_2 + \boldsymbol{\lambda}_2 + \boldsymbol{z}_1,\end{aligned} \tag{6.13}$$

where

$$\boldsymbol{z}_1 \triangleq (\sqrt{\alpha}\boldsymbol{U}_1^T\boldsymbol{H}_{12} - \boldsymbol{I}_n)\boldsymbol{x}_2 + \boldsymbol{U}_1^T\boldsymbol{H}_{11}\boldsymbol{x}_1 + \boldsymbol{U}_1^T\boldsymbol{w}_1, \tag{6.14}$$

and $\boldsymbol{t}_2$ is independent of $\boldsymbol{z}_1$, according to Lemma 1. $\boldsymbol{U}_1$ is chosen to minimize $\mathbb{E}[||\boldsymbol{z}_1||^2]$, and is then a diagonal matrix whose elements $u_{1,i}$, are the time-varying MMSE coefficient, given by

$$u_{1,i} = \frac{\sqrt{\alpha}P_2 h_{12,i}}{1 + P_1 h_{11,i}^2 + \alpha P_2 h_{12,i}^2}, \tag{6.15}$$

From (6.14),(6.15), $z_{1,i}$ is expressed as

$$\begin{aligned}z_{1,i} =& \frac{-(1 + P_1 h_{11,i}^2)x_{2,i}}{1 + P_1 h_{11}^2 + \alpha P_2 h_{12}^2} + \frac{\sqrt{\alpha}P_2 h_{11,i}h_{12,i}x_{1,i}}{1 + P_1 h_{11,i}^2 + \alpha P_2 h_{12,i}^2} \\ &+ \frac{\sqrt{\alpha}P_2 h_{12,i}w_{1,i}}{1 + P_1 h_{11,i}^2 + \alpha P_2 h_{22,i}^2}.\end{aligned} \tag{6.16}$$

Obviously, $z_{1,i}$ depends on $h_{11,i}, h_{12,i}$, causing the noise distribution to be channel-dependent. In order to simplify the decoding process, we ignore the instantaneous channel knowledge subsequent to MMSE scaling. We apply a spherical ambiguity decoder as follows

$$\Omega \triangleq \Big\{\boldsymbol{v} \in \mathbb{R}^n \ : \ ||\boldsymbol{v}||^2 \leq (1+\epsilon)nP_2\mathbb{E}\big[\frac{1}{1 + \frac{\alpha P_2 h_{12}^2}{1+P_1 h_{11}^2}}\big]\Big\}, \tag{6.17}$$

where ϵ is an arbitrary positive constant.

Error Probability: On averaging over the set of all fine lattices $\mathcal{C}$ of rate R that belong to the class of lattices proposed in Section 2.1, the probability of error can be bounded by

$$\frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i\in\mathcal{C}}\mathbb{P}_e < \mathbb{P}(\boldsymbol{z}\notin\Omega)+(1+\delta)2^{nR}\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V}^{(2)})}, \tag{6.18}$$

for any $\delta>0$. Following in the footsteps of Lemma 4, it can be shown that $P(\boldsymbol{z}\notin\Omega)<\gamma$ for any $\gamma>0$ at sufficiently large n. Consequently, the error probability can be bounded by

$$\epsilon'' \triangleq \frac{1}{|\mathcal{C}|}\sum_{\mathcal{C}_i\in\mathcal{C}}\mathbb{P}_e < \epsilon'+(1+\delta)2^{nR}\frac{\mathrm{Vol}(\Omega)}{\mathrm{Vol}(\mathcal{V}^{(2)})}, \tag{6.19}$$

for any $\gamma,\delta>0$. The volume of Ω is given by

$$\mathrm{Vol}(\Omega)=(1+\epsilon)^{\frac{n}{2}}\mathrm{Vol}\big(\mathcal{B}(\sqrt{nP_2})\big)\left(\mathbb{E}\big[\frac{1}{1+\frac{\alpha P_2 h_{12}^2}{1+P_1h_{11}^2}}\big]\right)^{\frac{n}{2}}. \tag{6.20}$$

The second term in (6.19) is then bounded by

$$\begin{aligned}&(1+\delta)2^{nR}(1+\epsilon)^{\frac{n}{2}}\left(\mathbb{E}\big[\frac{1}{1+\frac{\alpha P_2 h_{12}^2}{1+P_1h_{11}^2}}\big]\right)^{\frac{n}{2}}\frac{\mathrm{Vol}(\mathcal{B}(\sqrt{nP_2}))}{\mathrm{Vol}(\mathcal{V}^{(2)})}\\ &=(1+\delta)2^{-n\left(-\frac{1}{n}\log\left(\frac{\mathrm{Vol}(\mathcal{B}(\sqrt{nP_2}))}{\mathrm{Vol}(\mathcal{V}^{(2)})}\right)+\xi\right)},\end{aligned} \tag{6.21}$$

where

$$\xi\triangleq\frac{-1}{2}\log(1+\epsilon)-\frac{1}{2}\log\left(\mathbb{E}\big[\frac{1}{1+\frac{\alpha P_2 h_{12}^2}{1+P_1h_{11}^2}}\big]\right)-R. \tag{6.22}$$

From (2.8), since the lattice $\Lambda^{(2)}$ is good for covering, the first term of the exponent in (6.21) vanishes. From (6.21), $\lim_{n\to\infty}\mathbb{P}_e=0$ whenever $\xi>0$. Hence,

$$R<-\frac{1}{2}\log\left(\mathbb{E}\big[\frac{1}{1+\frac{\alpha P_2 h_{12}^2}{1+P_1h_{11}^2}}\big]\right)-\frac{1}{2}\log(1+\epsilon)-\epsilon', \tag{6.23}$$

is achievable where ϵ, ϵ' can be made arbitrarily small by increasing n. The existence of at least one lattice $\mathcal{C}_i$ that achieves the error probability averaged over the set of lattices $\mathcal{C}$ is straightforward.

In the event of successful decoding, from (6.13) the outcome of the decoding process would be $\hat{\boldsymbol{t}} = \boldsymbol{t}_2 + \boldsymbol{\lambda}_2$. On applying the modulo-Λ operation on $\hat{\boldsymbol{t}}$,

$$[\hat{\boldsymbol{t}}] \,\mathrm{mod}\Lambda \;=\; [\boldsymbol{t}_2 + \boldsymbol{\lambda}_2] \,\mathrm{mod}\Lambda^{(2)} \;=\; \boldsymbol{t}_2, \tag{6.24}$$

where the second equality follows from (2.6) since $\boldsymbol{\lambda}_2 \in \Lambda^{(2)}$.

Following in the footsteps of [5, 7], the existence of a sequence of coarse lattices with second moment P_2 that are good for covering and quantization, and nested in $\Lambda_1^{(2)}$ can be shown. Since a sphere is define by the *Euclidean metric*, one can replace the spherical region in (6.17) with a *Euclidean lattice decoder*, given by

$$\hat{\boldsymbol{t}} = \Big[\arg \min_{\boldsymbol{t}' \in \Lambda_1^{(2)}} ||(\boldsymbol{y}_2' - \boldsymbol{t}_2')||^2\Big] \,\mathrm{mod}\Lambda^{(2)}, \tag{6.25}$$

which cannot increase the error probability. Let $R_2^{(1)} \triangleq R$ denote the rate of message 2 that is decodable at receiver 1. Following the first decoding stage, $\boldsymbol{x}_2$ is then subtracted and $\boldsymbol{x}_1$ is decoded interference-free. When following similar steps to that above, it can be shown that

$$R_1^{(1)} < -\frac{1}{2} \log \Big(\mathbb{E}\big[\frac{1}{1 + P_1 h_{11}^2}\big] \Big), \tag{6.26}$$

where the rate pair $\{R_1, R_2\}$ in (6.23),(6.26) represent one corner in the rate region. On reversing the decoding order, a new rate pair can be obtained, given by

$$\begin{aligned} \tilde{R}_1^{(1)} &< -\frac{1}{2} \log \Big(\mathbb{E}\big[\frac{1}{1 + \frac{P_1 h_{11}^2}{1 + \alpha P_2 h_{12}^2}}\big] \Big), \\ \tilde{R}_2^{\,(1)} &< -\frac{1}{2} \log \Big(\mathbb{E}\big[\frac{1}{1 + \alpha P_2 h_{12}^2}\big] \Big). \end{aligned} \tag{6.27}$$

Using time sharing between the two aforementioned strategies, the first sum rate constraint in (6.10) can be obtained.[1] A similar rate region can be obtained with respect to receiver 2, and the intersection of both regions yields the rate region in (6.10).[2] This concludes the proof of Theorem 17. □

[1] Unlike the sum capacity bounds in (6.3), the sum rate bounds in (6.10) does not have a unit slope in general.

[2] The extension to complex-valued channels is similar to that in Theorem 2.

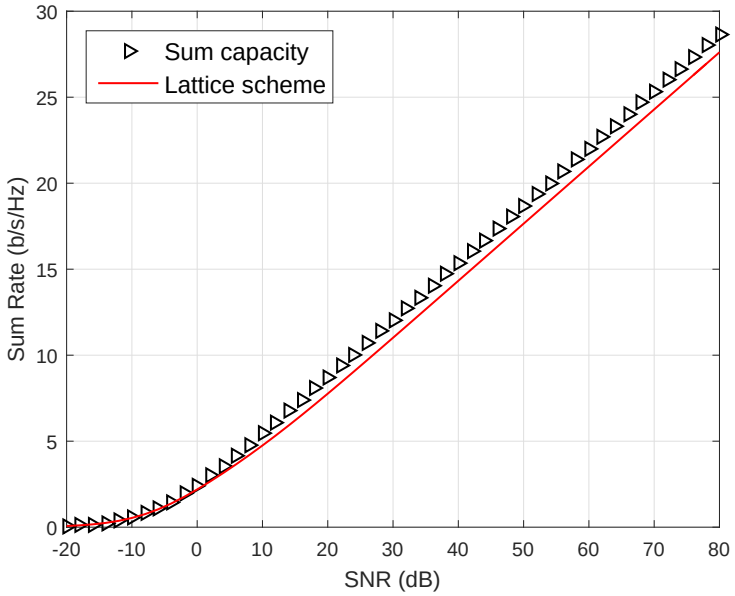

Figure 6.1. The sum rate of the ergodic strong interference channel with $\alpha = 4$ vs. sum capacity under Nakagami-m fading with $m = 2$, and $P_1 = P_2 \triangleq$ SNR.

In Figure 6.1, the sum rate in (6.10) is compared with the sum capacity in (6.3) versus SNR when all fading coefficients are Nakagami-m distributed with $m = 2$, $\alpha = 4$. The gap between the two rates is very small. Similar results are plotted in Figure 6.2 under Rayleigh fading.

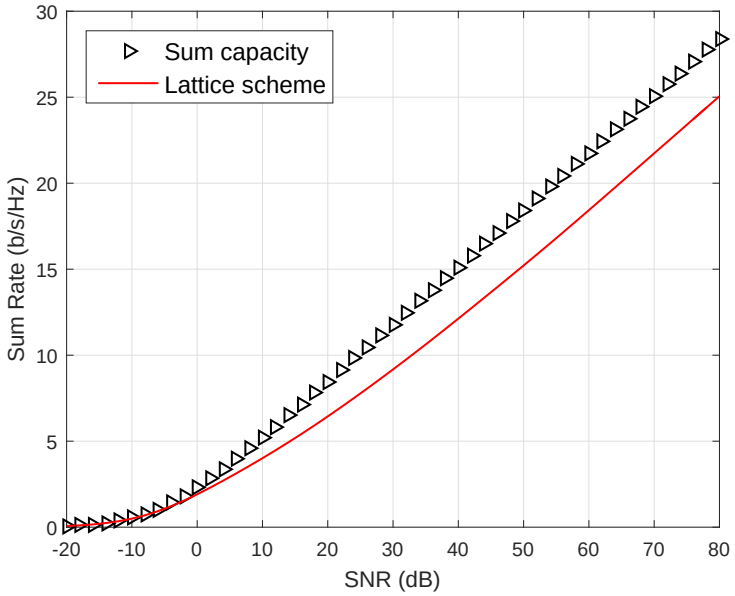

Figure 6.2. The sum rate of the ergodic strong interference channel with $\alpha = 4$ vs. sum capacity under Rayleigh fading, and $P_1 = P_2 \triangleq \text{SNR}$.

CHAPTER 7

CONCLUSION

This dissertation studies lattice coding and decoding in the context of several ergodic fading channels. Achievable lattice coding and decoding schemes are proposed and explicit achievable rates are computed. For most cases, the gap to respective capacities of these rates are computed. For the point-to-point channel and MAC, it is shown that the achievable rates are within a constant gap to capacity under a wide range of distributions. Under i.i.d. Rayleigh fading, the gap to capacity is a constant that vanishes as the number of receive antennas grows beyond that of transmit antennas, even at finite SNR. For the special case of single-antenna nodes, the gap to capacity is also shown to be a constant for even a wider range of distributions, including Nakagami-m fading. At low SNR, the gap to capacity is shown to be a diminishing fraction of the achievable rate. Simulation results are provided, confirming that lattice codes perform competitively with Gaussian codes under several channel and antenna configurations. An alternative decoding strategy is presented for block fading channels drawn from a discrete distribution, where channel-matching decision regions are proposed that achieve rates within a constant gap to capacity. The gap depends on both the coherence length of the channel as well as the fading distribution size, and is shown to vanish under some special cases of interest. Results are also extended to the imperfect CSIR case.

Also, the block fading MIMO point-to-point channel with channel state information at the transmitter is studied, where it is shown that a precoded lattice coding scheme with a fixed decoder and non-separable coding achieve the capacity.

The MIMO dirty paper channel with CSIR is also studied, where it is shown that a variant of Costa's dirty paper coding achieves rates within a constant gap to the capacity outer bound. A lattice coding and decoding scheme is proposed that achieve rates within a constant gap to capacity for a wide range of fading distributions. The gap to capacity diminishes as the number of receive antennas increases, even at finite SNR. The results imply that lattice coding and

decoding approach optimality for the fading dirty paper channel, and that under some antenna configurations the capacity of the fading dirty paper channel with CSIR is within a very small gap to the capacity of the point-to-point channel. The results are applied to various broadcast channel scenarios. Simulations show that the proposed lattice coding scheme has near-capacity performance.

Finally, a two-user ergodic fading interference channel is studied, where the interference is statistically strong. The capacity region of this channel is established. Additionally, a lattice coding and decoding scheme is proposed, whose achievable sum rate is close to the sum capacity for a variety of fading distributions.

Interesting future directions for research include:

- *Ergodic Interference Channel with arbitrary strength:* The results in this dissertation include performance characterization of the ergodic strong interference channel only. Other scenarios where the average interference power is lower than that of the direct link remain open.

- *Relay-aided Wireless Networks:* One plausible direction for future work includes studying relay channels. This would open the door for addressing the optimality of relaying techniques under non-Gaussian coding.

- *Closing the Gap to Capacity:* Although we have shown that lattice coding and decoding achieve rates within a small gap to capacity in several scenarios, it remains of interest to explore whether lattice codes can achieve capacity. This is an active area of research and some recent results have been able to answer the question in the affirmative for a point-to-point link [15, 69]. Extending the analysis to more general cases remains an open problem.

APPENDIX A

APPENDICES FOR CHAPTER 3

A.1 Proof of Lemma 4

The aim of Lemma 4 is showing that $\boldsymbol{z}$ lies with high probability within the sphere Ω. However, computing the distribution of $\boldsymbol{z}$ is challenging since it depends on that of $\boldsymbol{x}, \boldsymbol{H}$ as shown in (3.6), where the distribution of $\boldsymbol{x}$ is not known at arbitrary block length, and no fixed distribution is imposed for $\boldsymbol{H}$. The outline of the proof is as follows. First, we replace the original noise sequence with a *noisier* sequence whose statistics are known. Then, we use the weak law of large numbers to show that the noisier sequence is confined with high probability within Ω, which implies that the original noise $\boldsymbol{z}$ is also confined within Ω. We decompose the noise $\boldsymbol{z}$ in (3.8) in the form $\boldsymbol{z} = \boldsymbol{A}_s\boldsymbol{x} + \sqrt{\rho}\boldsymbol{B}_s\boldsymbol{w}$, where both $\boldsymbol{A}_s$, $\boldsymbol{B}_s$ are block-diagonal matrices with diagonal blocks $\boldsymbol{A}_i \triangleq -(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}_i^T\boldsymbol{H}_i)^{-1}$ and $\boldsymbol{B}_i \triangleq \sqrt{\rho}\boldsymbol{H}_i^T(\boldsymbol{I}_{N_r} + \rho\boldsymbol{H}_i\boldsymbol{H}_i^T)^{-1}$, respectively, such that

$$\boldsymbol{A}_s\boldsymbol{A}_s^T + \boldsymbol{B}_s\boldsymbol{B}_s^T = (\boldsymbol{I}_{N_tn} + \rho\boldsymbol{H}_s^T\boldsymbol{H}_s)^{-1}. \tag{A.1}$$

Since $\boldsymbol{H}_i$ is a stationary and ergodic process, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ can also be shown to be stationary and ergodic. Denote the eigenvalues of the random matrix $\boldsymbol{H}^T\boldsymbol{H}$ (arranged in ascending order) by $\sigma_{H,1}^2, \ldots, \sigma_{H,N_t}^2$. Then its eigenvalue decomposition is given by $\boldsymbol{H}^T\boldsymbol{H} \triangleq \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$, where $\boldsymbol{V}$ is a unitary matrix and $\boldsymbol{D}$ is a diagonal matrix whose *unordered* entries are $\sigma_{H,1}^2, \ldots, \sigma_{H,N_t}^2$. Owing to the isotropy of the distribution of $\boldsymbol{H}$, $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{V}(\boldsymbol{I}_{N_t}+\rho\boldsymbol{D})^{-2}\boldsymbol{V}^T$ can be shown to be *unitarily invariant*, i.e., $\mathbb{P}(\boldsymbol{A}\boldsymbol{A}^T) = \mathbb{P}(\check{\boldsymbol{V}}\boldsymbol{A}\boldsymbol{A}^T\check{\boldsymbol{V}}^T)$ for any unitary matrix $\check{\boldsymbol{V}}$ that is independent of $\boldsymbol{A}$. As a result of unitary invariance $\boldsymbol{V}$ is independent of $\boldsymbol{D}$ [70]. Hence,

$$\begin{aligned}
\mathbb{E}\big[\boldsymbol{A}\boldsymbol{A}^T\big] &= \mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^T\boldsymbol{H})^{-2}\big] \\
&= \mathbb{E}\big[\boldsymbol{V}(\boldsymbol{I}_{N_t} + \rho\boldsymbol{D})^{-2}\boldsymbol{V}^T\big] \\
&= \mathbb{E}_{\boldsymbol{V}|\boldsymbol{D}}\big[\boldsymbol{V}\mathbb{E}_{\boldsymbol{D}}[(\boldsymbol{I}_{N_t} + \rho\boldsymbol{D})^{-2}]\boldsymbol{V}^T\big] \\
&= \mathbb{E}_{\boldsymbol{V}|\boldsymbol{D}}\big[\boldsymbol{V}\sigma_A^2\boldsymbol{I}_{N_t}\boldsymbol{V}^T\big] \\
&= \sigma_A^2\boldsymbol{I}_{N_t},
\end{aligned} \tag{A.2}$$

where $\sigma_A^2 \triangleq \mathbb{E}_j\big[\mathbb{E}_{\sigma_{\boldsymbol{H},j}}[\frac{1}{(1+\rho\sigma^2_{\boldsymbol{H},j})^2}]\big]$. Similarly, it can be shown that $\mathbb{E}\big[\boldsymbol{B}\boldsymbol{B}^T\big] = \sigma_B^2 \boldsymbol{I}_{N_t}$, where

$$\sigma_B^2 \triangleq \mathbb{E}_j\bigg[\mathbb{E}_{\sigma_{\boldsymbol{H},j}}\big[\frac{\rho\sigma^2_{\boldsymbol{H},j}}{(1+\rho\sigma^2_{\boldsymbol{H},j})^2}\big]\bigg].$$

For convenience define $\sigma_z^2 \triangleq \sigma_A^2 + \sigma_B^2$. Next, we compute the autocorrelation of $\boldsymbol{z}$ as follows

$$\boldsymbol{\Sigma_z} \triangleq \mathbb{E}\big[\boldsymbol{z}\boldsymbol{z}^T\big] = \mathbb{E}\big[\boldsymbol{A}_s\boldsymbol{\Sigma_x}\boldsymbol{A}_s^T\big] + \rho\mathbb{E}\big[\boldsymbol{B}_s\boldsymbol{B}_s^T\big], \tag{A.3}$$

where $\boldsymbol{\Sigma_x} \triangleq \mathbb{E}\big[\boldsymbol{x}\boldsymbol{x}^T\big]$. Unfortunately, $\boldsymbol{\Sigma_x}$ is not known for all n, yet it approaches $\rho\boldsymbol{I}_{N_tn}$ for large n, according to Lemma 2. Hence one can rewrite

$$\boldsymbol{\Sigma_z} = \underbrace{\sigma_x^2\mathbb{E}\big[\boldsymbol{A}_s\boldsymbol{A}_s^T + \boldsymbol{B}_s\boldsymbol{B}_s^T\big]}_{\sigma_x^2\,\sigma_z^2\boldsymbol{I}_{N_tn}} + \underbrace{\mathbb{E}\big[\boldsymbol{A}_s(\boldsymbol{\Sigma_x} - \sigma_x^2\boldsymbol{I}_{N_tn})\boldsymbol{A}_s^T\big] + (\rho - \sigma_x^2)\mathbb{E}\big[\boldsymbol{B}_s\boldsymbol{B}_s^T\big]}_{\succ 0}, \tag{A.4}$$

where $\sigma_x^2 \triangleq \lambda_{\min}(\boldsymbol{\Sigma_x}) - \delta$, and $\lambda_{\min}(\boldsymbol{\Sigma_x})$ is the minimum eigenvalue of $\boldsymbol{\Sigma_x}$. Note that $\rho \geq \sigma_x^2$, from the definition in (2.2). As a result the second term in (A.4) is positive-definite, and $\boldsymbol{\Sigma_z} \succ \sigma_x^2\,\sigma_z^2\boldsymbol{I}_{N_tn}$. This implies that

$$\boldsymbol{\Sigma_z}^{-1} \prec \frac{1}{\sigma_x^2\sigma_z^2}\boldsymbol{I}_{N_tn}\,. \tag{A.5}$$

To make noise calculations more tractable, we introduce a related noise variable that modifies the second term of $\boldsymbol{z}$ as follows

$$\boldsymbol{z}^* = \boldsymbol{A}_s\boldsymbol{x} + \boldsymbol{B}_s\big(\sqrt{\rho}\boldsymbol{w} + \sqrt{\frac{1}{N_tn}R_c^2 - \rho}\,\boldsymbol{w}^*\big), \tag{A.6}$$

where $\boldsymbol{w}^*$ is i.i.d. Gaussian with zero mean and unit variance, and R_c is the covering radius of $\mathcal{V}$. We now wish to bound the probability that $\boldsymbol{z}^*$ is outside a sphere of radius $\sqrt{(1+\epsilon)N_tn\sigma_x^2\sigma_z^2}$. First, we rewrite

$$||\boldsymbol{z}^*||^2 = \boldsymbol{x}^T\boldsymbol{A}_s^T\boldsymbol{A}_s\boldsymbol{x} + \frac{1}{N_tn}R_c^2\boldsymbol{w}^T\boldsymbol{B}_s\boldsymbol{B}_s^T\boldsymbol{w} + 2\sqrt{\frac{1}{N_tn}R_c^2}\,\boldsymbol{x}^T\boldsymbol{A}_s^T\boldsymbol{B}_s\boldsymbol{w}. \tag{A.7}$$

Then, we bound each term separately using the weak law of large numbers. The third term satisfies [1]

$$\mathbb{P}\big(2\sqrt{\frac{1}{N_tn}R_c^2}\,\boldsymbol{x}^T\boldsymbol{A}_s^T\boldsymbol{B}_s\boldsymbol{w} > N_tn\epsilon_3\big) < \gamma_3. \tag{A.8}$$

[1]The third term in (A.7) is a sum of zero mean uncorrelated random variables to which the law of large numbers applies [71].

Addressing the second term in (A.7), [2]

$$\begin{aligned}
&\mathbb{P}\big(\frac{1}{N_t n} R_c^2\, \boldsymbol{w}^T \boldsymbol{B}_s^T \boldsymbol{B}_s \boldsymbol{w} > \sigma_B^2 R_c^2 + N_t n \epsilon_2\big) \\
&\quad = \mathbb{P}\big(\frac{1}{N_t n} R_c^2\, \mathrm{tr}(\boldsymbol{B}_s \boldsymbol{w}\boldsymbol{w}^T \boldsymbol{B}_s^T \boldsymbol{w}) > \sigma_B^2 R_c^2 + N_t n \epsilon_2\big) \\
&\quad < \gamma_2. \qquad \text{(A.9)}
\end{aligned}$$

Now, we bound the first term in (A.7). Given that $\boldsymbol{A}_s^T \boldsymbol{A}_s$ is a block-diagonal matrix with $\mathbb{E}\big[\boldsymbol{A}_s^T \boldsymbol{A}_s\big] = \sigma_A^2 \boldsymbol{I}_{N_t n}$, and that $\boldsymbol{\Sigma_x} \to \rho \boldsymbol{I}_{N_t n}$ as $n \to \infty$, it can be shown using [72, Theorem 1] that $\frac{1}{||\boldsymbol{x}||^2} \boldsymbol{x}^T \boldsymbol{A}_s^T \boldsymbol{A}_s \boldsymbol{x} \to \sigma_A^2$ as $n \to \infty$. More precisely,

$$\begin{aligned}
&\mathbb{P}\big(\boldsymbol{x}^T \boldsymbol{A}_s^T \boldsymbol{A}_s \boldsymbol{x} > \sigma_A^2 R_c^2 + N_t n \epsilon_1\big) \\
&\quad < \mathbb{P}\big(\boldsymbol{x}^T \boldsymbol{A}_s^T \boldsymbol{A}_s \boldsymbol{x} > \sigma_A^2 ||\boldsymbol{x}||^2 + N_t n \epsilon_1\big) \\
&\quad < \gamma_1, \qquad \text{(A.10)}
\end{aligned}$$

where $||\boldsymbol{x}||^2 < R_c^2$, and $\epsilon_1, \epsilon_2, \epsilon_3$ and $\gamma_1, \gamma_2, \gamma_3$ can be made arbitrarily small by increasing n. Using a union bound,

$$\mathbb{P}\big(||\boldsymbol{z}^*||^2 > (1+\epsilon_4) R_c^2 \sigma_z^2\big) < \gamma, \qquad \text{(A.11)}$$

where $\epsilon_4 \triangleq \frac{(\epsilon_1+\epsilon_2+\epsilon_3)}{R_c^2 \sigma_z^2}$ and $\gamma \triangleq \gamma_1 + \gamma_2 + \gamma_3$. For large n, $\frac{1}{N_t n} R_c^2 \leq (1+\epsilon_6)\rho$ for covering-good lattices and $\rho \leq (1+\epsilon_7)\sigma_x^2$ according to Lemma 2. Let $\epsilon_5 \triangleq (1+\epsilon_6)(1+\epsilon_7) - 1$, then for any ϵ such that $\epsilon \leq (1+\epsilon_4)(1+\epsilon_5) - 1$,

$$\begin{aligned}
&\mathbb{P}\big(\boldsymbol{z}^{*T} \boldsymbol{\Sigma_z} \boldsymbol{z}^* > (1+\epsilon) N_t n\big) < \mathbb{P}\big(||\boldsymbol{z}^*||^2 > (1+\epsilon) N_t n \rho \sigma_z^2\big) \qquad \text{(A.12)} \\
&\quad = \mathbb{P}\Big(\boldsymbol{z}^{*T} \big(\mathbb{E}\big[(\boldsymbol{I}_{N_t n} + \rho \boldsymbol{H}_s^T \boldsymbol{H}_s)^{-1}\big]\big)^{-1} \boldsymbol{z}^* > (1+\epsilon) N_t n \rho\Big) \\
&\quad < \gamma, \qquad \text{(A.13)}
\end{aligned}$$

where (A.12) holds from (A.5) and (A.13) holds since $\mathbb{E}\big[(\boldsymbol{I}_{N_t n} + \rho \boldsymbol{H}_s^T \boldsymbol{H}_s)^{-1}\big] = \sigma_z^2 \boldsymbol{I}_{N_t n}$, according to (A.1). The final step is to show that $||\boldsymbol{z}^* - \boldsymbol{z}|| \to 0$ as $n \to \infty$, where $\boldsymbol{z}^* - \boldsymbol{z} =$

[2] Note that $\mu_i \triangleq \boldsymbol{w}_i^T \boldsymbol{B}_i^T \boldsymbol{B}_i \boldsymbol{w}_i$ is also a stationary and ergodic process that obeys the law of large numbers.

$\sqrt{\frac{1}{N_t n}R_c^2 - \rho}\,\boldsymbol{B}_s\boldsymbol{w}^*$. From the structure of $\boldsymbol{B}_s$, the norm of each of its rows is less than N_t, and hence the variance of each of the elements of $\boldsymbol{B}_s\boldsymbol{w}^*$ is no more than N_t. Since $\lim_{n\to\infty}\frac{1}{N_t n}R_c^2 = \rho$ for a covering-good lattice, it can be shown using Chebyshev's inequality that the elements of $\sqrt{\frac{1}{N_t n}R_c^2 - \rho}\,\boldsymbol{B}_s\boldsymbol{w}^*$ vanish and $|\boldsymbol{z}_j^* - \boldsymbol{z}_j| \to 0$ as $n \to \infty$ for all $j \in \{1, \ldots, N_t n\}$. This concludes the proof of Lemma 4.

A.2 Proof of Lemma 5

Denote by $\mathcal{S}$ the event that the post-processed received point $\boldsymbol{y}'$ falls exclusively within one decision sphere, defined in (3.9), where the probability of occurrence of $\mathcal{S}$ is $\mathbb{P}_{\mathcal{S}} \triangleq 1 - \gamma_s$. Using the law of total probability, the probability of error (in general) is given by

$$\mathbb{P}_e = \mathbb{P}_{e|\mathcal{S}}\mathbb{P}_{\mathcal{S}} + \mathbb{P}_{e|\mathcal{S}^c}\mathbb{P}_{\mathcal{S}^c} \tag{A.14}$$

First we analyze the ambiguity decoder with spherical decision regions (denoted by superscript $^{(SD)}$). From the definition of ambiguity decoding, $\mathbb{P}_{e|\mathcal{S}^c}^{(SD)} = 1$. Hence,

$$\mathbb{P}_e^{(SD)} = \eta'(1 - \gamma_s) + \gamma_s, \tag{A.15}$$

where $\mathbb{P}_{e|\mathcal{S}}^{(SD)} \triangleq \eta'$. Now we analyze the Euclidean lattice decoder (denoted by superscript $^{(LD)}$). Since a sphere is defined by the Euclidean metric, the outcomes of the spherical decoder and the Euclidean lattice decoder *conditioned on the event* $\mathcal{S}$ are identical, and hence yield the same error probability, i.e., $\mathbb{P}_{e|\mathcal{S}}^{(LD)} = \mathbb{P}_{e|\mathcal{S}}^{(SD)} = \eta'$. However, from (3.17), the Euclidean lattice decoder declares a valid output even under the event $\mathcal{S}^c$. Hence, $\mathbb{P}_{e|\mathcal{S}^c}^{(LD)} \triangleq \eta'' \leq 1$. Thereby,

$$\mathbb{P}_e^{(LD)} = \eta'(1 - \gamma_s) + \eta''\gamma_s \leq \mathbb{P}_e^{(SD)}. \tag{A.16}$$

A.3 Proof of Corollary 1

Lemma 10. *For an i.i.d. complex Gaussian $K \times M$ matrix $\boldsymbol{G}$ whose elements have zero mean, unit variance and $K > M$, then $\mathbb{E}\big[(\boldsymbol{G}^H\boldsymbol{G})^{-1}\big] = \frac{1}{K-M}\,\boldsymbol{I}_M$.*

Proof. See [73, Section V]. □

A.3.1 Case 1: $N_r \geq N_t$ and the elements of $\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big] < \infty$

$$
\begin{aligned}
\Delta =& C - R \\
=& \mathbb{E}\big[\log\det(\boldsymbol{I}_{N_t} + \rho\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})\big] + \log\det\Big(\mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big) \\
\leq& \log\det\Big(\boldsymbol{I}_{N_t} + \rho\mathbb{E}[\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}]\Big) + \log\det\Big(\mathbb{E}\big[(\boldsymbol{I}_{N_t} + \rho\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big) \qquad \text{(A.17)} \\
<& \log\det\Big(\boldsymbol{I}_{N_t} + \rho\mathbb{E}[\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}]\Big) + \log\det\Big(\mathbb{E}\big[(\rho\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big) \qquad \text{(A.18)} \\
=& \log\det\Big(\big(\frac{1}{\rho}\boldsymbol{I}_{N_t} + \mathbb{E}[\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}]\big)\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big) \qquad \text{(A.19)} \\
\leq& \log\det\Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}]\big)\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big), \qquad \text{(A.20)}
\end{aligned}
$$

where (A.17),(A.18) follow since $\log\det(\boldsymbol{A})$ is a concave and non-decreasing function over the set of all positive definite matrices [74]. (A.20) follows since $\rho \geq 1$.

A.3.2 Case 2: $N_r > N_t$ and the elements of $\tilde{\boldsymbol{H}}$ are i.i.d. complex Gaussian

When the elements of $\tilde{\boldsymbol{H}}$ are i.i.d. complex Gaussian with zero mean and unit variance,

$$
\begin{aligned}
\Delta <& \log\det\Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}]\big)\mathbb{E}\big[(\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}})^{-1}\big]\Big) \qquad \text{(A.21)} \\
=& \log\det\big((1+N_r)\frac{1}{N_r - N_t}\boldsymbol{I}_{N_t}\big) \qquad \text{(A.22)} \\
=& N_t\log\big(1 + \frac{N_t+1}{N_r - N_t}\big),
\end{aligned}
$$

where (A.21),(A.22) follow from Case 1 and Lemma 10, respectively.

A.3.3 Case 3: $N_t = 1$ and $\rho < \frac{1}{\mathbb{E}[||\tilde{\boldsymbol{h}}||^2]}$

$$
\begin{aligned}
\Delta = \; & C - R \\
& = \mathbb{E}\big[\log\left(1+\rho||\tilde{\boldsymbol{h}}||^2\right)\big] + \log\left(\mathbb{E}[\frac{1}{1+\rho||\tilde{\boldsymbol{h}}||^2}]\right) \\
& \leq \; \log\left(1+\rho\mathbb{E}[||\tilde{\boldsymbol{h}}||^2]\right) + \log\left(\mathbb{E}\big[\frac{1}{1+\rho||\tilde{\boldsymbol{h}}||^2}\big]\right) \qquad \text{(A.23)} \\
& \leq \log e\, \mathbb{E}[||\tilde{\boldsymbol{h}}||^2]\rho + \log e\, \mathbb{E}\Big[\frac{-\rho||\tilde{\boldsymbol{h}}||^2}{1+\rho||\tilde{\boldsymbol{h}}||^2}\Big] \qquad \text{(A.24)} \\
& = \; \log e\, \mathbb{E}\Big[\rho||\tilde{\boldsymbol{h}}||^2 - \frac{\rho||\tilde{\boldsymbol{h}}||^2}{1+\rho||\tilde{\boldsymbol{h}}||^2}\Big] \\
& = \log e\, \mathbb{E}\Big[\frac{||\tilde{\boldsymbol{h}}||^4}{1+\rho||\tilde{\boldsymbol{h}}||^2}\Big]\,\rho^2 \\
& < \; 1.45\, \mathbb{E}\big[||\tilde{\boldsymbol{h}}||^4\big]\,\rho^2,
\end{aligned}
$$

where (A.23) is due to Jensen's inequality and (A.24) utilizes $\ln x \leq x - 1$.

A.4 Proof of Corollary 2

The results in Case 1 and Case 2 are straightforward from Corollary 1. The proofs are therefore omitted.

A.4.1 Case 3: $\rho \geq 1$, Nakagami-m fading with $m > 1$

The Nakagami-m distribution with $m > 1$ satisfies the condition $\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big] < \infty$. For a Nakagami-$m$ variable with unit power, i.e., $\mathbb{E}[|\tilde{h}|^2] = 1$, $\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big]$ is computed as follows

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big] &= \frac{2m^m}{\Gamma(m)} \int_0^\infty \frac{1}{x^2} x^{2m-1} e^{-mx^2} dx \\
&= \frac{2m^m}{\Gamma(m)} \frac{1}{2m^{m-1}} \int_0^\infty y^{m-2} e^{-y} dy \\
&= \frac{m\,\Gamma(m-1)}{\Gamma(m)} \\
&= \frac{m\,\Gamma(m-1)}{(m-1)\,\Gamma(m-1)} \\
&= 1 + \frac{1}{m-1}\,,
\end{aligned}$$

where $\Gamma(\cdot)$ denotes the *gamma function*. Substituting in (3.25), $\Delta < 1 + \log\big(1 + \frac{1}{m-1}\big)$.

A.4.2 Case 4: $\rho \geq 1$, Rayleigh fading

Lemma 11. *For any $z > 0$, the exponential integral function defined by $\bar{E}_1(z) = \int_z^\infty \frac{e^{-t}}{t} dt$ is upper bounded by*

$$\bar{E}_1(z) < \frac{1}{\log e}\, e^{-z}\, \log(1 + \frac{1}{z}).$$

Proof. See [75, Section 5.1]. □

Under Rayleigh fading, $|\tilde{h}|^2$ is exponentially distributed with unit power. Hence,

$$\begin{aligned}
\Delta =& \ \mathbb{E}[\log\left(1+\rho|\tilde{h}|^2\right)] + \log\left(\mathbb{E}[\frac{1}{1+\rho|\tilde{h}|^2}]\right) \\
\leq& \ \log\left(1+\rho\mathbb{E}[|\tilde{h}|^2]\right) + \log\left(\mathbb{E}[\frac{1}{1+\rho|\tilde{h}|^2}]\right) \qquad \text{(A.25)} \\
\leq& \ 1+\log\left(\mathbb{E}[\frac{1}{|\tilde{h}|^2+\frac{1}{\rho}}]\right) \qquad \text{(A.26)} \\
\leq& \ 1+\log\big(\int_0^\infty \frac{1}{x+\frac{1}{\rho}}e^{-x}dx\big) \\
=& \ 1+\log\big(e^{\frac{1}{\rho}}\int_{\frac{1}{\rho}}^\infty \frac{1}{y}e^{-y}dy\big) \\
=& \ 1+\log\big(e^{\frac{1}{\rho}}\bar{E}_1(\frac{1}{\rho})\big) \\
<& \ 1+\log\big(\frac{1}{\log e}\log(1+\rho)\big) \qquad \text{(A.27)} \\
<& \ 0.48+\log\big(\log(1+\rho)\big),
\end{aligned}$$

where (A.25) follows from Jensen's inequality. (A.26) holds from the condition $\rho \geq 1$ and (A.27) follows from Lemma 11.

A.5 Proof of Lemma 7

The proof resembles that in [5, 7]. Consider a noise vector $\boldsymbol{z}^* \in \mathbb{R}^{Mn}$ that is closely related to $\boldsymbol{z}$ as follows

$$\begin{aligned}
\boldsymbol{z}^* =& -\big(\boldsymbol{I}_{N_tn}+\rho\boldsymbol{H}_s^T\boldsymbol{H}_s\big)^{-1}\boldsymbol{g}+\sqrt{\rho}\boldsymbol{H}_s^T\big(\boldsymbol{I}_{N_tn}+\rho\boldsymbol{H}_s\boldsymbol{H}_s^T\big)^{-1}\big(\sqrt{\rho}\,\boldsymbol{w}+\sqrt{\sigma_{\mathcal{B}}^2-\rho}\,\boldsymbol{w}^*\big) \\
\triangleq& \ \boldsymbol{A}_s\boldsymbol{g}+\boldsymbol{B}_s\big(\sqrt{\rho}\,\boldsymbol{w}+\sqrt{\sigma_{\mathcal{B}}^2-\rho}\,\boldsymbol{w}^*\big), \qquad \text{(A.28)}
\end{aligned}$$

where $\boldsymbol{g}$ and $\boldsymbol{w}^*$ are i.i.d. Gaussian with zero mean and variances $\sigma_{\mathcal{B}}^2\boldsymbol{I}_{N_tn}$, $\boldsymbol{I}_{N_rn}$, respectively. $\sigma_{\mathcal{B}}^2$ is the second moment of the smallest sphere covering $\mathcal{V}$, where $\sigma_{\mathcal{B}}^2 > \rho$. It is then easy to show that the autocorrelation matrix of $\boldsymbol{z}^*$ is $\frac{\sigma_{\mathcal{B}}^2}{\rho}\boldsymbol{\Sigma}$, given in (3.38). The probability $\mathbb{P}\big(\boldsymbol{z}^* \notin \Omega\big)$ is then equivalent to $\mathbb{P}\big(||\boldsymbol{z}^{(w)}||^2 > (1+\gamma)\frac{\rho}{\sigma_{\mathcal{B}}^2}N_tn\big)$, where $\boldsymbol{z}^{(w)} \triangleq (\frac{\sigma_{\mathcal{B}}^2}{\rho}\boldsymbol{\Sigma}_s)^{\frac{-1}{2}}\boldsymbol{z}^*$ is i.i.d. Gaussian with

zero-mean and unit variance, and $||\boldsymbol{z}^{(w)}||^2$ is a *chi-squared* random variable with Mn degrees-of-freedom. Using the Chernoff bound,

$$\begin{aligned}
&\mathbb{P}\big(||\boldsymbol{z}^{(w)}||^2 > (1+\gamma)N_t n\big) \\
&\quad\leq \min_{t\geq 0}\left\{e^{-N_t n\left((1+\gamma)t+\log_e(1-t)\right)}\right\} \\
&\quad= (1+\gamma)^{\frac{N_t n}{2}} e^{-\frac{N_t n\gamma}{2}} \\
&\quad= e^{-\frac{N_t n}{2}\left(\gamma-\log_e(1+\gamma)\right)}. \qquad \text{(A.29)}
\end{aligned}$$

Now, we show that the distribution of $\boldsymbol{z}$, i.e., $f_{\boldsymbol{z}}(\boldsymbol{v})$, is upper-bounded (up to a constant) by $f_{\boldsymbol{z}^*}(\boldsymbol{v})$. It was shown in [5, Lemma 11], that $f_{\boldsymbol{x}}(\boldsymbol{v}) \leq e^{c_n} f_{\boldsymbol{g}}(\boldsymbol{v})$, where $c_n/n \to 0$ as $n \to \infty$. Hence, $f_{\boldsymbol{A}_s\boldsymbol{x}}(\boldsymbol{v}) \leq e^{c_n} f_{\boldsymbol{A}_s\boldsymbol{g}}(\boldsymbol{v})$ and $f_{\boldsymbol{z}}(\boldsymbol{v}) \leq e^{c_n} f_{\boldsymbol{z}^*}(\boldsymbol{v})$ follow as well, where the former inequality is obtained using transformation of random variables whereas the latter is obtained via convolution of both terms in (A.28).

$$\begin{aligned}
\mathbb{P}(z \notin \Omega) &= \int_{v\notin\Omega} f_{\boldsymbol{z}}(\boldsymbol{v})dv \\
&\leq e^{c_n}\int_{v\notin\Omega} f_{\boldsymbol{z}^*}(\boldsymbol{v})dv \\
&= e^{c_n}\,\mathbb{P}\big(||\boldsymbol{z}^{(w)}||^2 > (1+\gamma)\frac{\rho}{\sigma_{\mathcal{B}}^2}N_t n\big) \\
&= e^{c_n}\,\mathbb{P}\big(||\boldsymbol{z}^{(w)}||^2 > (1+\gamma')Mn\big) \qquad \text{(A.30)} \\
&\leq e^{-\frac{N_t n}{2}\left(\gamma'-\log_e(1+\gamma')-\frac{2c_n}{N_t n}\right)}. \qquad \text{(A.31)}
\end{aligned}$$

where (A.30) holds since $\rho \leq (1+\frac{2}{N_t n})\sigma_{\mathcal{B}}^2$ [5], and hence $\gamma' \leq (1+\gamma)(1+\frac{2}{N_t n})$. Since $\gamma > \log_e(1+\gamma')$ for all $\gamma' > 0$ and $c_n/n \to 0$ as $n \to \infty$, the exponent in (A.31) remains negative and it can be shown that there exists $n_{\epsilon'} \in \mathbb{Z}^+$ such that for all $n > n_{\epsilon'}$, $\mathbb{P}(\boldsymbol{z} \notin \Omega) < \epsilon'$.

The final step is to show that the elements of $\sqrt{\sigma_{\mathcal{B}}^2 - \rho}\,\boldsymbol{B}\boldsymbol{w}^* \to 0$ as $n \to \infty$. Since $|B_{ii}| < 1$, the variance of each of the elements of $\boldsymbol{B}\boldsymbol{w}^*$ is no more than 1. Since $\lim_{n\to\infty} \sigma_{\mathcal{B}}^2 = \rho$ for a covering-good lattice, it can be shown using *Chebyshev's inequality* that the elements of $\sqrt{\sigma_{\mathcal{B}}^2 - \rho}\,\boldsymbol{B}\boldsymbol{w}^*$

vanish with high probability, as follows

$$\mathbb{P}\big(\sqrt{\sigma_{\mathcal{B}}^2-\rho}\,\boldsymbol{B}\boldsymbol{w}^*\geq\gamma^*\,\kappa\big)\leq\frac{1}{\kappa^2},\qquad for\ all\ \kappa>0,$$
$$\mathbb{P}\big(\sqrt{\sigma_{\mathcal{B}}^2-\rho}\,\boldsymbol{B}\boldsymbol{w}^*\geq\sqrt{\gamma^*}\big)\leq\gamma^*, \tag{A.32}$$

where $\gamma^*\triangleq\sqrt{\sigma_{\mathcal{B}}^2-\rho}$ vanishes with n and (A.32) follows when $\kappa=\frac{1}{\sqrt{\gamma^*}}$. This concludes the proof of Lemma 7.

A.6 Designing the Permutation Matrix $\boldsymbol{V}$

As mentioned earlier in Section 3.2.1, the role of $\boldsymbol{V}$ is ordering the channel coefficients in the diagonal matrix $\boldsymbol{H}$ in ascending order of the magnitudes. Recall that the coefficients take on the values $\mathfrak{h}_1,\ldots,\mathfrak{h}_{|\mathcal{H}|}$, where $\mathfrak{h}_k$ appears exactly n_k times. Let us define a counter $\nu_k^{(i)}\leq n_k$, that counts the number of occurrences of the coefficient $\mathfrak{h}_k$ immediately after channel use i. Typically, when $h_i=\mathfrak{h}_k$, then $\boldsymbol{V}(i,:)$ is $\boldsymbol{\delta}_m^T$, where $m=\nu_k^{(i)}+\sum_{l=1}^{k-1}n_l$. Using basic linear algebraic results, it is straightforward that $\boldsymbol{V}^T\boldsymbol{H}\boldsymbol{V}=\boldsymbol{H}_{\boldsymbol{\pi}}$ for any $\boldsymbol{H}$ that belongs to the random location channel model.

A.7 Proof of Theorem 5

Before we proceed, we bound the total number of channel occurrences that deviate from $n\mathbb{P}_k$ in (2.10) as follows[3]

$$\sum_{k=1}^{|\mathcal{H}|}|n_k-n\mathbb{P}_k|\leq\delta n\mathbb{P}_k<\delta n\triangleq n_{out}. \tag{A.33}$$

From (2.11),(A.33) on varying δ, a tradeoff occurs between n_{out} and the total number of δ-typical sequences. However, for the choice $\delta\triangleq\delta' n^{\frac{-1}{2}(1-\gamma)}$, where $\delta'>0$ and $0<\gamma<1$, $n_{out}=\delta' n^{\frac{1}{2}(1+\gamma)}$ is a vanishing fraction of n, and the probability of atypical sequences would be upper-bounded by $2|\mathcal{H}|e^{-\frac{\mu n^\gamma}{3}}$. Hence, negligible n_{out} can be guaranteed for almost all sequences satisfying a distribution at large n.

[3]For simplicity, we assume $n\mathbb{P}_k$ and $n\delta$ are positive integers.

Now we are ready to present the capacity achieving scheme. The encoding, the choice of $\boldsymbol{D}$ as well as $\boldsymbol{U}$ are identical to that in Section 3.2.1, i.e., D_{ii}^2 are the normalized optimal power allocations for h_i and U_{ii} are the MMSE coefficients. However, designing the permutation matrix $\boldsymbol{V}$ is now more challenging, since the number of occurrences of the different channel values does not exactly fit the statistical distribution of the channel. We adopt a best-effort approach to choose $\boldsymbol{V}$, whose design is given in details in Table A.1, which can be phrased as follows. First, we make a rough assumption that coefficient α_k occurs exactly $n\mathbb{P}_k$ times. Designing the first $n - n_{out}$ rows of $\boldsymbol{V}$ is identical to that in Appendix A.6. For the remaining n_{out} rows, one or more slots dedicated for α_k may be exhausted. Hence, we utilize the unoccupied slots dedicated for α_j where $j < k$, whose channel magnitudes are smaller. If all are occupied, we utilize the last n_{out} time slots available. This implies that the n_{out} channel coefficients with the largest magnitudes have no dedicated slots. Following the permutation operation, we use an ellipsoidal decision region $\tilde{\Omega}_1$ as follows

$$\tilde{\Omega}_1 \triangleq \{\boldsymbol{s} \in \mathbb{R}^n \ : \ \boldsymbol{s}^T \tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{s} \leq (1+\epsilon)n\}, \tag{A.34}$$

where $\tilde{\boldsymbol{\Sigma}}$ is a diagonal matrix given by

$$\tilde{\Sigma}_{ii} = \begin{cases} \frac{\rho}{\rho h^2_{\boldsymbol{\pi}(i)}+1} & \text{for } i \in \{1, \ldots, n - n_{out}\} \\ \rho & \text{for } i \in \{n - n_{out} + 1, \ldots, n\}. \end{cases} \tag{A.35}$$

Owing to $\boldsymbol{V}$, $\Omega_1^{(p)} \subseteq \tilde{\Omega}$, where $\Omega^{(p)}$ is an ellipsoid parametrized by a perfectly ordered auto-correlation matrix $\boldsymbol{\Sigma^{(p)}}$ whose elements are perfectly ordered in descending order (recall the channel coefficients are in ascending order), and hence achieves capacity. $\Omega_1^{(p)} \subseteq \tilde{\Omega}$ follows since $\Sigma_{ii}^{(p)} \leq \tilde{\Sigma}_{ii}$ for any i, as guaranteed by the structure of $\boldsymbol{V}$. The rate achieved is then

$$R < \frac{1}{2n} \sum_{i=1}^{n} \log\left(1 + h_i^2 \rho^*(h_i)\right),$$

which converges to $\frac{1}{2}\mathbb{E}\Big[\log\left(1 + h^2\rho^*(h)\right)\Big]$, by the weak law of large numbers. The final step needed is showing that the suboptimal decision region $\tilde{\Omega}_1$ has negligible impact on the achievable

Design of the permutation matrix $\boldsymbol{V}$

Set $\nu_{out} = 0$, $\nu_k^{(0)} = 0$ for $k = 1, \ldots, |\mathcal{H}|$.
for $i = 1 : n$
  Set $flag = 0$.
  if $h_i == \alpha_k$
    Set $\nu_k^{(i)} = \nu_k^{(i-1)} + 1$ and $\nu_l^{(i)} = \nu_l^{(i-1)}$ for $l \neq k$.
    if $\nu_k^{(i)} \leq np_k$
      Set $\boldsymbol{V}(i,:)$ to $\boldsymbol{\delta}_m^T$, where $m = \nu_k^{(i)} + n\sum_{l=1}^{k-1} \mathbb{P}_l$.
      Set $flag = 1$.
    else
      for $j = k-1 : -1 : 1$
        if $\nu_j^{(i)} \leq n\mathbb{P}_j$
          Set $\nu_j^{(i)} = \nu_j^{(i-1)} + 1$ and $\nu_l^{(i)} = \nu_l^{(i-1)}$ for $l \neq j$.
          Set $\boldsymbol{V}(i,:)$ to $\boldsymbol{\delta}_m^T$, where $m = \nu_j^{(i)} + n\sum_{l=1}^{j-1} \mathbb{P}_l$.
          Set $flag = 1$.
          break;
        end
      end
      if *flag*==0
        Set $\nu_{out} = \nu_{out} + 1$.
        Set $\boldsymbol{V}(i,:)$ to $\boldsymbol{\delta}_m^T$, where $m = \nu_{out} + \lfloor n - n_{out} \rfloor$.
      end
    end
  end
end

Table A.1. Steps of designing $\boldsymbol{V}$ for a general fading channel.

rate. From (2.7) as well as the error analysis in Section 3.2.1, it can be shown that the gap $\Delta \triangleq C - R$ is given by

$$\begin{aligned}\Delta =& \frac{1}{n}\Big(\log\Big(\frac{\mathrm{Vol}(\tilde{\Omega}_1)}{\mathrm{Vol}(\Omega_1^{(p)})}\Big) + o(n)\Big) \\ =& \frac{n_{out}}{n}\,\frac{1}{n_{out}} \sum_{i=n-n_{out}+1}^{n} \log(1 + \rho^*_{\boldsymbol{\pi}(i)} h^2_{\boldsymbol{\pi}(i)}) + \frac{o(n)}{n} \\ <& \frac{n_{out}}{n}\, \log(1 + \rho^*_{\boldsymbol{\pi}(n)} h^2_{\boldsymbol{\pi}(n)}) + \frac{o(n)}{n}\,, \end{aligned} \tag{A.36}$$

where $o(n)$ satisfies $\lim_{n\to\infty} \frac{o(n)}{n} = 0$. Hence, Δ vanishes since $\frac{n_{out}}{n} \to 0$ as $n \to \infty$. This concludes the proof of Theorem 5.

A.8 Proof of Corollary 3

For ease of exposition let $\beta \triangleq \frac{P_x}{1+P_x\sigma_h^2}$.

A.8.1 Case 1: General fading distribution with $\mathbb{E}\big[\frac{1}{|\hat{h}|^2}\big] < \infty$

$$\begin{aligned}
\Delta =& \mathbb{E}\big[\log(1+\beta|\hat{h}|^2)\big] + \log\Big(\mathbb{E}\big[\frac{1}{1+\beta|\hat{h}|^2}\big]\Big) \\
\leq& \log\Big(\big(1+\frac{1}{\beta}\big)\,\mathbb{E}\big[\frac{1}{|\hat{h}|^2+\frac{1}{\beta}}\big]\Big) && \text{(A.37)} \\
\leq& 1+\log\Big(\mathbb{E}\big[\frac{1}{|\hat{h}|^2+\frac{1}{\beta}}\big]\Big) && \text{(A.38)} \\
<& 1+\log\Big(\mathbb{E}\big[\frac{1}{|\hat{h}|^2}\big]\Big),
\end{aligned}$$

and (A.37) follows from Jensen's inequality. (A.38) follows since $\sigma_h^2 \leq \frac{P_x-1}{P_x}$ implies that $\beta \geq 1$.

A.8.2 Case 2: Nakagami-m fading with $m > 1$

First, we compute $\mathbb{E}\big[\frac{1}{g^2}\big]$ when g is Nakagami-m distributed with $m > 1$ and $\mathbb{E}[g^2] = 1$.

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{g^2}\big] =& \frac{2m^m}{\Gamma(m)} \int_0^\infty \frac{1}{g^2}\, g^{2m-1} e^{-mg^2} dx \\
=& \frac{2m^m}{\Gamma(m)} \frac{1}{2m^{m-1}} \int_0^\infty y^{m-2} e^{-y} dy \\
=& m\,\frac{\Gamma(m-1)}{\Gamma(m)} \\
=& 1+\frac{1}{m-1}\,.
\end{aligned}$$

Since $|\hat{h}|$ is Nakagami distributed, $\Delta < 1+\log\big(1+\frac{1}{m-1}\big)$.

A.8.3 Case 3: Rayleigh fading

Under Rayleigh fading, $|\hat{h}|^2$ is exponentially distributed with unit rate parameter, i.e., $\mathbb{E}[|\hat{h}|^2] = 1$. Hence,

$$\begin{aligned}
\Delta \leq & 1 + \log\Big(\mathbb{E}\big[\frac{1}{|\hat{h}|^2 + \frac{1}{\beta}}\big]\Big) && \text{(A.39)}\\
= & 1 + \log\big(\int_0^\infty \frac{1}{x + \frac{1}{\beta}} e^{-x} dx\big)\\
= & \log(2 + \sigma_h^2) + \log\big(e^{\sigma_h^2} \int_{\sigma_h^2}^\infty \frac{1}{y} e^{-y} dy\big)\\
< & 0.48 + \log\big(\log(1+\beta)\big) && \text{(A.40)}\\
< & 0.48 + \log\Big(\log\big(1 + \frac{1}{\sigma_h^2}\big)\Big),
\end{aligned}$$

where (A.39) follows from (A.38) and (A.40) follows from Lemma 11.

A.9 Proof of Corollary 4

A.9.1 Case 1: Medard's scheme

$$\begin{aligned}
R_G = & \mathbb{E}\Big[\log\big(1 + \frac{P_x|\hat{h}|^2}{1 + P_x\sigma_h^2}\big)\Big]\\
= & \mathbb{E}\Big[\log\big(\frac{P_x|\hat{h}|^2 + P_x\sigma_h^2 + 1}{P_x\sigma_h^2 + 1}\big)\Big]\\
< & \mathbb{E}\Big[\log\big(\frac{P_x|\hat{h}|^2 + P_x\sigma_h^2}{P_x\sigma_h^2}\big)\Big] && \text{(A.41)}\\
< & 1 + \mathbb{E}\Big[\log\big(\frac{\mathrm{SNR}}{\mathrm{INR}}\big)\Big], && \text{(A.42)}
\end{aligned}$$

where (A.41) follows since $\frac{x+y+1}{x+1} < \frac{x+y}{x}$ for all $x, y > 0$, and (A.42) follows since $\log(1+z) < 1 + \log(z)$ for $z > 1$. Hence,

$$\frac{R_G}{\mathbb{E}[\log \mathrm{SNR}]} = \frac{1 + \mathbb{E}\Big[\log\big(\frac{\mathrm{SNR}}{\mathrm{INR}}\big)\Big]}{\mathbb{E}[\log \mathrm{SNR}]} = 1 - \alpha + \frac{1}{\mathbb{E}[\log \mathrm{SNR}]}.$$

Hence, $\mathcal{D}_R(\alpha) = 1 - \alpha$ is straightforward.

A.9.2 Case 2: Lattice scheme

$$\begin{aligned} R_L &= -\log\Big(\mathbb{E}\big[\frac{1}{1+\frac{P_x|\hat{h}|^2}{P_x\sigma_h^2+1}}\big]\Big) \\ &= -\log\Big(\mathbb{E}\big[\frac{P_x\sigma_h^2+1}{P_x|\hat{h}|^2+P_x\sigma_h^2+1}\big]\Big) \\ &> -\log\Big(\mathbb{E}\big[(P_x\sigma_h^2+1)\frac{1}{P_x|\hat{h}|^2+1}\big]\Big) \\ &> -\log\Big(\mathbb{E}\big[\frac{1}{P_x|\hat{h}|^2+1}\big]\Big) - 1 - \log \mathrm{INR}. \end{aligned}$$

On replacing β with P_x in Appendix A.4, it is easy to show that under Nakagami-m fading with $m > 1$,

$$\frac{-\log\big(\mathbb{E}[\frac{1}{1+\mathrm{SNR}}]\big)}{\mathbb{E}[\log(1+\mathrm{SNR})]} > 1 + \frac{1+\log(1+\frac{1}{m-1})}{\mathbb{E}[\log(1+\mathrm{SNR})]},$$

whereas for Rayleigh fading,

$$\frac{-\log\big(\mathbb{E}[\frac{1}{1+\mathrm{SNR}}]\big)}{\mathbb{E}[\log(1+\mathrm{SNR})]} > 1 + \frac{0.48+\log\big(\log(1+\mathrm{SNR})\big)}{\mathbb{E}[\log(1+\mathrm{SNR})]}\,.$$

Hence, it is obvious that $\mathcal{D}_L(\alpha) = 1-\alpha$ in both cases.

APPENDIX B

APPENDICES FOR CHAPTER 4

B.1 Proof of Corollary 5

Lemma 12. *For any two independent i.i.d. Gaussian matrices* $\boldsymbol{A} \in \mathbb{C}^{r\times m}$, $\boldsymbol{B} \in \mathbb{C}^{r\times q}$ *where* $r \geq q+1$ *whose elements have zero mean and unit variance,*

$$\boldsymbol{A}^H\big(c\boldsymbol{I}_r + \boldsymbol{B}\boldsymbol{B}^H\big)^{-1}\boldsymbol{A} \succ \frac{1}{c}\,\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}}, \tag{B.1}$$

where the elements of $\bar{\boldsymbol{A}} \in \mathbb{C}^{(r-q)\times m}$ *are i.i.d. Gaussian with zero-mean and unit variance, and* c *is a positive constant.*

Proof. Using the eigenvalue decomposition of $\big(c\boldsymbol{I}_r + \boldsymbol{B}\boldsymbol{B}^H\big)^{-1}$,

$$\boldsymbol{A}^H\big(c\boldsymbol{I}_r + \boldsymbol{B}\boldsymbol{B}^H\big)^{-1}\boldsymbol{A} = \boldsymbol{A}^H\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^H\boldsymbol{A} = \check{\boldsymbol{A}}^H\boldsymbol{D}\check{\boldsymbol{A}}, \tag{B.2}$$

where the columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{B}\boldsymbol{B}^H$. The corresponding eigenvalues of $\boldsymbol{B}\boldsymbol{B}^H$ are then in the form $\sigma_1^2,\ldots,\sigma_q^2,0,\ldots,0$. Hence, q of the diagonal entries of $\boldsymbol{D}$ are in the form $1/(c+\sigma_j^2)$, whereas the remaining $r-q$ entries are $1/c$. Since $\boldsymbol{V}$ is unitary, then $\check{\boldsymbol{A}} \triangleq \boldsymbol{V}^H\boldsymbol{A}$ is i.i.d. Gaussian, similar to $\boldsymbol{A}$ [70]. One can rewrite (B.2) as follows

$$\begin{aligned}
\check{\boldsymbol{A}}^H\boldsymbol{D}\check{\boldsymbol{A}} &= \sum_{j=1}^{r-q}\frac{1}{c}\,\check{\boldsymbol{a}}_j\check{\boldsymbol{a}}_j^H + \sum_{j=r-q+1}^{r}\frac{1}{c+\sigma_j^2}\,\check{\boldsymbol{a}}_j\check{\boldsymbol{a}}_j^H \\
&\succ \sum_{j=1}^{r-q}\frac{1}{c}\,\check{\boldsymbol{a}}_j\check{\boldsymbol{a}}_j^H \\
&= \frac{1}{c}\,\bar{\boldsymbol{A}}^H\bar{\boldsymbol{A}},
\end{aligned} \tag{B.3}$$

where $\check{\boldsymbol{a}}_j$ is the conjugate transposition of row j in $\check{\boldsymbol{A}}$, and the columns of the matrix $\bar{\boldsymbol{A}}$ are $\check{\boldsymbol{a}}_j$ for $j \in \{1,\ldots,r-q\}$. The generalized inequality in (B.3) follows since $\boldsymbol{X}+\boldsymbol{Y} \succeq \boldsymbol{X}$ for any two positive semidefinite matrices $\boldsymbol{X},\boldsymbol{Y}$. □

Let $\tilde{\boldsymbol{F}}_{\pi(k)} \triangleq \boldsymbol{I}_{N_r} + \rho \sum_{l=k+1}^{K} \tilde{\boldsymbol{H}}_{\pi(l)} \tilde{\boldsymbol{H}}_{\pi(l)}^H$, where $\pi(\cdot)$ is an arbitrary permutation as described in Section 4.1. We first bound the sum capacity in (4.18) (from above) as follows

$$
\begin{aligned}
C_{\text{sum}} &\triangleq \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_r} + \sum_{k=1}^{K} \rho \tilde{\boldsymbol{H}}_k \tilde{\boldsymbol{H}}_k^H\big)\Big] \\
&= \sum_{k=1}^{K} \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}_{\pi(k)}^H \tilde{\boldsymbol{F}}_{\pi(k)}^{-1} \tilde{\boldsymbol{H}}_{\pi(k)}\big)\Big] \\
&\leq \sum_{k=1}^{K} \mathbb{E}\Big[\log\det\big(\boldsymbol{I}_{N_t} + \rho \tilde{\boldsymbol{H}}_{\pi(k)}^H \tilde{\boldsymbol{H}}_{\pi(k)}\big)\Big] \qquad \text{(B.4)} \\
&\leq \sum_{k=1}^{K} \log\det\Big(\boldsymbol{I}_{N_t} + \rho\, \mathbb{E}\big[\tilde{\boldsymbol{H}}_{\pi(k)}^H \tilde{\boldsymbol{H}}_{\pi(k)}\big]\Big) \\
&= K\, \log\det\big((1+\rho N_r)\boldsymbol{I}_{N_t}\big) \\
&= N_t K \log\left(1+\rho N_r\right) \\
&\leq N_t K\big(\log\rho + \log\left(1+N_r\right)\big), \qquad \text{(B.5)}
\end{aligned}
$$

where (B.4) follows since interference cannot increase capacity, and (B.5) follows since $\rho \geq 1$.

Now, we bound (from below) R_{sum}. Since the sum of the rate expressions in both (4.16) and (4.17) are equal, we bound each of the $N_t K$ terms in (4.16), where the power is allocated uniformly over each virtual user, given by ρ as follows

$$
\begin{aligned}
R_{\pi(\ell)} &= -\log\Big(\mathbb{E}\big[\frac{1}{1+\rho\tilde{\boldsymbol{h}}_{\pi(\ell)}^H(\boldsymbol{I}_{N_r} + \rho\sum_{j=\ell+1}^{L}\tilde{\boldsymbol{h}}_{\pi(j)}\tilde{\boldsymbol{h}}_{\pi(j)}^H)^{-1}\tilde{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big) \\
&= -\log\Big(\mathbb{E}\big[\frac{1}{1+\rho\tilde{\boldsymbol{h}}_{\pi(\ell)}^H(\frac{1}{\rho}\boldsymbol{I}_{N_r} + \tilde{\boldsymbol{G}}_{\pi(\ell)}\tilde{\boldsymbol{G}}_{\pi(\ell)}^H)^{-1}\tilde{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big) \\
&\geq -\log\Big(\mathbb{E}\big[\frac{1}{1+\rho\check{\boldsymbol{h}}_{\pi(\ell)}^H\check{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big) \qquad \text{(B.6)} \\
&> -\log\Big(\mathbb{E}\big[\frac{1}{\rho\check{\boldsymbol{h}}_{\pi(\ell)}^H\check{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big) \\
&= \log\rho - \log\Big(\mathbb{E}\big[\frac{1}{\check{\boldsymbol{h}}_{\pi(\ell)}^H\check{\boldsymbol{h}}_{\pi(\ell)}}\big]\Big) \\
&= \log\rho + \log\big(N_r - (K-\ell+1)\big), \qquad \text{(B.7)}
\end{aligned}
$$

where $\tilde{\boldsymbol{G}}_{\pi(\ell)} \triangleq [\tilde{\boldsymbol{h}}_{\pi(\ell+1)}, \ldots, \tilde{\boldsymbol{h}}_{\pi(L)}]$. (B.6) follows from Lemma 12 where $\check{\boldsymbol{h}}_{\pi(\ell)} \in \mathbb{C}^{N_r - L + \ell}$ is an i.i.d. Gaussian distributed vector whose elements have unit variance, and (B.7) follows from Lemma 10.

Hence, from (B.5),(B.7) the gap is bounded as follows

$$\begin{aligned}
\Delta &\triangleq C_{\text{sum}} - R_{\text{sum}} \\
&< \sum_{\ell=1}^{N_t K} \Big(\log\left(1 + N_r\right) - \log\left(N_r - (N_t K - \ell + 1)\right) \Big) \\
&= \sum_{\ell=1}^{N_t K} \log\big(\frac{1 + N_r}{N_r - (N_t K - \ell + 1)}\big) \\
&= \sum_{\ell=1}^{N_t K} \log\big(\frac{1 + N_r}{N_r - \ell}\big) \\
&= \sum_{\ell=1}^{N_t K} \log\big(1 + \frac{\ell + 1}{N_r - \ell}\big). \qquad \text{(B.8)}
\end{aligned}$$

B.2 Proof of Corollary 6

B.2.1 Case 1: $\rho < \frac{1}{2}$

$$\begin{aligned}
\Delta =& \mathbb{E}\big[\log\left(1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2\right)\big] + \log\Big(\mathbb{E}\big[\frac{1 + \rho|\tilde{h}_1|^2}{1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2}\big]\mathbb{E}\big[\frac{1}{1 + \rho|\tilde{h}_1|^2}\big]\Big) \\
\leq& \log\left(1 + \rho\mathbb{E}[|\tilde{h}_1|^2] + \rho\mathbb{E}[|\tilde{h}_2|^2]\right) + \log\Big(\mathbb{E}\big[\frac{1}{1 + \rho|\tilde{h}_1|^2}\big]\Big) + \log\Big(\mathbb{E}\big[\frac{1 + \rho|\tilde{h}_1|^2}{1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2}\big]\Big) \\
<& \log e\Big(\rho\mathbb{E}[|\tilde{h}_1|^2] + \rho\mathbb{E}[|\tilde{h}_2|^2] + \mathbb{E}\big[\frac{-\rho|\tilde{h}_2|^2}{1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2}\big] + \mathbb{E}\big[\frac{-\rho|\tilde{h}_1|^2}{1 + \rho|\tilde{h}_1|^2}\big]\Big) \\
<& \log e\Big(\mathbb{E}\big[\rho|\tilde{h}_2|^2 - \frac{\rho|\tilde{h}_2|^2}{1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2}\big] + \mathbb{E}\big[\rho|\tilde{h}_1|^2 - \frac{\rho|\tilde{h}_1|^2}{1 + \rho|\tilde{h}_1|^2}\big]\Big) \\
=& \log e\Big(\mathbb{E}\big[\frac{\rho^2|\tilde{h}_1|^2|\tilde{h}_2|^2 + \rho^2|\tilde{h}_2|^4}{1 + \rho|\tilde{h}_1|^2 + \rho|\tilde{h}_2|^2}\big] + \mathbb{E}\big[\frac{\rho^2|\tilde{h}_1|^4}{1 + \rho|\tilde{h}_1|^2}\big]\Big) \\
\leq& \log e\left(\mathbb{E}[\rho^2|\tilde{h}_1|^2|\tilde{h}_2|^2 + \rho^2|\tilde{h}_2|^4] + \mathbb{E}[\rho^2|\tilde{h}_1|^4]\right) \\
=& \log e\left(1 + \mathbb{E}[|\tilde{h}_1|^4] + \mathbb{E}[|\tilde{h}_2|^4]\right)\rho^2 \\
<& 1.45\left(1 + 2\mathbb{E}[|\tilde{h}_1|^4]\right)\rho^2.
\end{aligned}$$

B.2.2 Case 2: $\rho \geq \frac{1}{2}$ and $\mathbb{E}\big[\frac{1}{|\tilde{h}|^2}\big] < \infty$

$$\begin{aligned}
\Delta =& \mathbb{E}\big[\log\left(1+\rho|\tilde{h}_1|^2+\rho|\tilde{h}_2|^2\right)\big] + \log\Big(\mathbb{E}\big[\frac{1+\rho|\tilde{h}_1|^2}{1+\rho|\tilde{h}_1|^2+\rho|\tilde{h}_2|^2}\big]\mathbb{E}\big[\frac{1}{1+\rho|\tilde{h}_1|^2}\big]\Big) \\
\leq& \log\left(1+\rho\mathbb{E}[|\tilde{h}_1|^2]+\rho\mathbb{E}[|\tilde{h}_2|^2]\right)] + \log\Big(\mathbb{E}\big[\frac{1+\rho|\tilde{h}_1|^2}{1+\rho|\tilde{h}_1|^2+\rho|\tilde{h}_2|^2}\big]\mathbb{E}\big[\frac{1}{1+\rho|\tilde{h}_1|^2}\big]\Big) \qquad \text{(B.9)} \\
<& \log\Big((1+2\rho)\mathbb{E}\big[\frac{1}{1+\rho|\tilde{h}_1|^2}\big]\Big) \\
\leq& 2+\log\big(\mathbb{E}[\frac{1}{|\tilde{h}_1|^2+\frac{1}{\rho}}]\big) \qquad \text{(B.10)} \\
<& 2+\log\Big(\mathbb{E}\big[\frac{1}{|\tilde{h}_1|^2}\big]\Big),
\end{aligned}$$

where (B.9) follows from Jensen's inequality and (B.10) follows since $\rho \geq \frac{1}{2}$.

B.2.3 Case 3: $\rho \geq \frac{1}{2}$, Nakagami-m fading with $m > 1$

Since the Nakagami-m distribution with $m > 1$ belongs to the class of distributions in Case 2, then

$$\begin{aligned}
\Delta <& \ 2+\log\Big(\mathbb{E}\big[\frac{1}{|\tilde{h}_1|^2}\big]\Big) \\
=& \ 2+\log\big(1+\frac{1}{m-1}\big), \qquad \text{(B.11)}
\end{aligned}$$

where (B.11) follows from the proof of Case 3 in Appendix A.4.

B.2.4 Case 4: $\rho \geq \frac{1}{2}$, Rayleigh fading

$$\begin{aligned}
\Delta \leq& \ 2+\log\big(\mathbb{E}[\frac{1}{|\tilde{h}_1|^2+\frac{1}{\rho}}]\big) \qquad \text{(B.12)} \\
<& \ 1.48+\log\big(\log(1+\rho)\big), \qquad \text{(B.13)}
\end{aligned}$$

where (B.12) follows from Case 2 and (B.13) follows from Case 4 in Appendix A.4.

APPENDIX C

APPENDICES FOR CHAPTER 5

C.1 Proof of Lemma 9

We rewrite the noise expression $\boldsymbol{z}$ in (5.22) in the form $\boldsymbol{z} = \boldsymbol{A}_d\boldsymbol{x} + \boldsymbol{A}_d\boldsymbol{s} + \sqrt{\frac{\mu}{P_w}}\boldsymbol{B}_d\boldsymbol{w}$, where $\mu \triangleq \frac{P_x}{N_t} + P_s$, and both $\boldsymbol{A}_d$, $\boldsymbol{B}_d$ are block-diagonal matrices with diagonal blocks $\boldsymbol{A}_i, \boldsymbol{B}_i$, as follows

$$\boldsymbol{A}_i \triangleq -\big(\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{H}_i^T\boldsymbol{H}_i\big)^{-1}, \tag{C.1}$$

$$\boldsymbol{B}_i \triangleq \sqrt{P_w}\mu\boldsymbol{H}_i^T(\mu\boldsymbol{H}_i\boldsymbol{H}_i^T + P_w\boldsymbol{I}_{N_r})^{-1}. \tag{C.2}$$

Note that

$$\boldsymbol{A}_d\boldsymbol{A}_d^T + \boldsymbol{B}_d\boldsymbol{B}_d^T = (\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{H}_d^T\boldsymbol{H}_d)^{-1}. \tag{C.3}$$

Since $\boldsymbol{H}_i$ is a stationary and ergodic process, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are stationary and ergodic processes as well. In the proceeding we omit the time index i whenever it is clear from the context. Denote the ordered eigenvalues of the random matrix $\boldsymbol{H}^T\boldsymbol{H}$ by $\sigma_{H,1}^2, \ldots, \sigma_{H,M}^2$ (non-decreasing). Then the eigenvalue decomposition of $\boldsymbol{H}^T\boldsymbol{H}$ is $\boldsymbol{H}^T\boldsymbol{H} \triangleq \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$, where $\boldsymbol{V}$ is a unitary matrix and $\boldsymbol{D}$ is a diagonal matrix whose unordered entries are $\sigma_{H,1}^2, \ldots, \sigma_{H,N_t}^2$. Owing to the isotropy of the distribution of $\boldsymbol{H}$, $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{V}(\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{D})^{-2}\boldsymbol{V}^T$ is unitarily invariant, i.e., $\mathbb{P}(\boldsymbol{A}\boldsymbol{A}^T) = \mathbb{P}(\check{\boldsymbol{V}}\boldsymbol{A}\boldsymbol{A}^T\check{\boldsymbol{V}}^T)$ for any unitary matrix $\check{\boldsymbol{V}}$ independent of $\boldsymbol{A}$. As a result $\boldsymbol{V}$ is independent of $\boldsymbol{D}$ [70]. Hence,

$$\begin{aligned}
\mathbb{E}\big[\boldsymbol{A}\boldsymbol{A}^T\big] &= \mathbb{E}\big[(\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{H}^T\boldsymbol{H})^{-2}\big] \\
&= \mathbb{E}\big[\boldsymbol{V}(\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{D})^{-2}\boldsymbol{V}^T\big] \\
&= \mathbb{E}_{\boldsymbol{V}|\boldsymbol{D}}\big[\boldsymbol{V}\mathbb{E}_{\boldsymbol{D}}[(\boldsymbol{I}_{N_t} + \frac{\mu}{P_w}\boldsymbol{D})^{-2}]\boldsymbol{V}^T\big] \\
&= \mathbb{E}_{\boldsymbol{V}|\boldsymbol{D}}\big[\boldsymbol{V}\sigma_A^2\boldsymbol{I}_{N_t}\boldsymbol{V}^T\big] \\
&= \sigma_A^2\boldsymbol{I}_{N_t},
\end{aligned} \tag{C.4}$$

where $\sigma_A^2 \triangleq \mathbb{E}_j\big[\mathbb{E}_{\sigma_{\boldsymbol{H},j}}[\frac{1}{(1+\frac{\mu}{P_w}\sigma^2_{\boldsymbol{H},j})^2}]\big]$. Similarly, it can be shown that $\mathbb{E}\big[\boldsymbol{B}\boldsymbol{B}^T\big] = \sigma_B^2 \boldsymbol{I}_{N_t}$, where

$$\sigma_B^2 \triangleq \mathbb{E}_j\Big[\mathbb{E}_{\sigma_{\boldsymbol{H},j}}\big[\frac{\frac{\mu}{P_w}\sigma^2_{\boldsymbol{H},j}}{(1+\frac{\mu}{P_w}\sigma^2_{\boldsymbol{H},j})^2}\big]\Big].$$

For convenience define $\sigma_z^2 \triangleq \sigma_A^2 + \sigma_B^2$. Next, we compute the autocorrelation of $\boldsymbol{z}$ as follows

$$\boldsymbol{\Sigma_z} \triangleq \mathbb{E}\big[\boldsymbol{z}\boldsymbol{z}^T\big] = \mathbb{E}\big[\boldsymbol{A}_d(\boldsymbol{\Sigma_x} + P_s\boldsymbol{I}_{N_tn})\boldsymbol{A}_d^T\big] + \mu\mathbb{E}\big[\boldsymbol{B}_d\boldsymbol{B}_d^T\big], \tag{C.5}$$

where $\boldsymbol{\Sigma_x} \triangleq \mathbb{E}\big[\boldsymbol{x}\boldsymbol{x}^T\big]$. Unfortunately, $\boldsymbol{\Sigma_x}$ is not known for all n, yet it approaches $\frac{P_x}{N_t}\boldsymbol{I}_{N_tn}$ for large n, according to Lemma 2. Hence one can rewrite

$$\boldsymbol{\Sigma_z} = \underbrace{(\sigma_x^2 + P_s)\mathbb{E}\big[\boldsymbol{A}_d\boldsymbol{A}_d^T + \boldsymbol{B}_d\boldsymbol{B}_d^T\big]}_{(\sigma_x^2+P_s)\,\sigma_z^2\boldsymbol{I}_{N_tn}} + \underbrace{\mathbb{E}\big[\boldsymbol{A}_d(\boldsymbol{\Sigma_x} - \sigma_x^2\boldsymbol{I}_{N_tn})\boldsymbol{A}_d^T\big] + \big(\frac{P_x}{N_t} - \sigma_x^2\big)\mathbb{E}\big[\boldsymbol{B}_d\boldsymbol{B}_d^T\big]}_{\succ\boldsymbol{0}}, \tag{C.6}$$

where $\sigma_x^2 \triangleq \lambda_{\min}(\boldsymbol{\Sigma_x}) - \delta$. It follows that $\boldsymbol{\Sigma_z} \succ (\sigma_x^2 + P_s)\,\sigma_z^2\boldsymbol{I}_{N_tn}$, therefore

$$\boldsymbol{\Sigma_z}^{-1} \prec \frac{1}{(\sigma_x^2 + P_s)\sigma_z^2}\boldsymbol{I}_{N_tn}. \tag{C.7}$$

To make noise calculations more tractable, we introduce a related noise variable that modifies the second term of $\boldsymbol{z}$ as follows

$$\boldsymbol{z}^* = \boldsymbol{A}_d\boldsymbol{x} + \boldsymbol{A}_d\boldsymbol{s} + \sqrt{\frac{\mu}{P_w}}\boldsymbol{B}_d\boldsymbol{w} + \sqrt{\frac{\frac{1}{N_tn}R_c^2 + P_s}{P_w} - \frac{\mu}{P_w}}\,\boldsymbol{B}_d\boldsymbol{w}^*, \tag{C.8}$$

where $\boldsymbol{w}^*$ is i.i.d. Gaussian with zero mean and unit variance, and R_c is the covering radius of $\mathcal{V}$, and hence $\frac{1}{n}R_c^2 > P_x$. We now wish to bound the probability that $\boldsymbol{z}^*$ is outside a sphere of radius $\sqrt{(1+\epsilon)N_tn(\sigma_x^2 + P_s)\sigma_z^2}$ for some ϵ that vanishes with n. First, we rewrite

$$\begin{aligned}||\boldsymbol{z}^*||^2 = {}& \frac{\frac{1}{N_tn}R_c^2 + P_s}{P_w}\,\boldsymbol{w}^T\boldsymbol{B}_d^T\boldsymbol{B}_d\boldsymbol{w} + \boldsymbol{s}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{s} + 2\sqrt{\frac{\frac{1}{N_tn}R_c^2 + P_s}{P_w}}\,\boldsymbol{s}^T\boldsymbol{A}_d^T\boldsymbol{B}_d\boldsymbol{w} + \boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{x} \\ & + 2\sqrt{\frac{\frac{1}{N_tn}R_c^2 + P_s}{P_w}}\,\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{B}_d\boldsymbol{w} + 2\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{s}.\end{aligned} \tag{C.9}$$

We now bound the probability of deviation each of the terms on the right hand side of (C.9) from its mean using the law of large numbers. To begin with, the first term in (C.9) is the

sum of n terms of an ergodic sequence, where $\mathbb{E}[\boldsymbol{w}^T\boldsymbol{B}_d^T\boldsymbol{B}_d\boldsymbol{w}] = \mathrm{tr}\big(\mathbb{E}[\boldsymbol{B}_d\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{B}_d^T]\big) = N_t n P_w \sigma_B^2$. Hence for any $\epsilon, \gamma \in (0,1)$ there exists sufficiently large n so that:

$$\mathbb{P}\Big(\frac{\frac{1}{N_t n}R_c^2 + P_s}{P_w}\,\boldsymbol{w}^T\boldsymbol{B}_d^T\boldsymbol{B}_d\boldsymbol{w} > (R_c^2 + N_t n P_s)\sigma_B^2 + N_t n\epsilon\Big) < \gamma, \tag{C.10}$$

Similarly:

$$\mathbb{P}\big(\boldsymbol{s}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{s} > N_t n P_s \sigma_A^2 + N_t n\epsilon\big) < \gamma, \tag{C.11}$$

and

$$\mathbb{P}\Big(2\sqrt{\frac{\frac{1}{N_t n}R_c^2 + P_s}{P_w}}\,\boldsymbol{s}^T\boldsymbol{A}_d^T\boldsymbol{B}_d\boldsymbol{w} > N_t n\epsilon\Big) < \gamma. \tag{C.12}$$

The next term in (C.9) involves $\boldsymbol{A}_d^T\boldsymbol{A}_d$, a block-diagonal matrix with $\mathbb{E}\big[\boldsymbol{A}_d^T\boldsymbol{A}_d\big] = \sigma_A^2\boldsymbol{I}_{N_t n}$. Considering $\boldsymbol{\Sigma}_{\boldsymbol{x}} \to \rho\boldsymbol{I}_{N_t n}$ as $n \to \infty$, it can be shown that $\frac{1}{||\boldsymbol{x}||^2}\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{x} \to \sigma_A^2$, using [72, Theorem 1]. More precisely,

$$\mathbb{P}\big(\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{x} > \sigma_A^2||\boldsymbol{x}||^2 + N_t n\epsilon\big) < \gamma. \tag{C.13}$$

Since $||\boldsymbol{x}||^2 < R_c^2$, (C.13) implies

$$\mathbb{P}\big(\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{x} > \sigma_A^2 R_c^2 + N_t n\epsilon\big) < \gamma. \tag{C.14}$$

The penultimate term in (C.9) can be bounded as follows:[1]

$$\mathbb{P}\Big(2\sqrt{\frac{\frac{1}{N_t n}R_c^2 + P_s}{P_w}}\,\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{B}_d\boldsymbol{w} > N_t n\epsilon\Big) < \gamma, \tag{C.15}$$

where the elements of $\boldsymbol{x}$ are also bounded. Similarly for the final term in (C.9),

$$\mathbb{P}\big(2\boldsymbol{x}^T\boldsymbol{A}_d^T\boldsymbol{A}_d\boldsymbol{s} > N_t n\epsilon\big) < \gamma, \tag{C.16}$$

In (C.10) through (C.16), ϵ, γ can be made arbitrarily small by increasing n. Moreover, for any fixed ϵ, γ there is a sufficiently large n so that simultaneously all the above bounds are satisfied, because their number is finite and for each one a sufficiently large n exists.

[1]This term can be expressed as the sum of zero-mean and uncorrelated random variables to which the law of large numbers apply [71].

We now produce a union bound on all the terms above:

$$\mathbb{P}\big(||\boldsymbol{z}^*||^2 > (1+\epsilon')(R_c^2 + N_t n P_s)\sigma_z^2\big) < 6\gamma, \tag{C.17}$$

where $\epsilon' \triangleq \frac{6\epsilon}{(R_c^2+N_t n P_s)\sigma_z^2}$. For sufficiently large n, we can find $\frac{1}{N_t n}R_c^2 \leq (1+\epsilon)\frac{P_x}{N_t}$ for covering-good lattices and $\frac{P_x}{N_t} \leq (1+\epsilon)\sigma_x^2$ according to Lemma 2. Then, take $\epsilon'' \triangleq (1+\epsilon)^2 - 1$, and any $\epsilon''' \leq (1+\epsilon')(1+\epsilon'') - 1$, we have

$$\begin{aligned}
&\mathbb{P}\big(\boldsymbol{z}^{*T}{\boldsymbol{\Sigma}_{\boldsymbol{z}}}^{-1}\boldsymbol{z}^* > (1+\epsilon''')N_t n\big) \\
<&\mathbb{P}\big(||\boldsymbol{z}^*||^2 > (1+\epsilon''')n(N_t\sigma_x^2 + N_t P_s)\sigma_z^2\big) \qquad \text{(C.18)} \\
<&\mathbb{P}\big(||\boldsymbol{z}^*||^2 > (1+\epsilon'''')n(P_x + N_t P_s)\sigma_z^2\big) \\
=&\mathbb{P}\Big(||\boldsymbol{z}^*||^2 > (1+\epsilon'''')\,\mathrm{tr}\big(\mathbb{E}\big[\big(\frac{1}{\frac{P_x}{N_t}+P_s}\boldsymbol{I}_{N_t n} + \frac{1}{P_w}\boldsymbol{H}_d^T\boldsymbol{H}_d\big)^{-1}\big]\big)\Big) \qquad \text{(C.19)} \\
=&\mathbb{P}\Big(||\boldsymbol{z}^*||^2 > (1+\epsilon'''')n\,\mathrm{tr}\big(\mathbb{E}\big[\big(\frac{1}{\frac{P_x}{N_t}+P_s}\boldsymbol{I}_{N_t} + \frac{1}{P_w}\boldsymbol{H}^T\boldsymbol{H}\big)^{-1}\big]\big)\Big) < 6\gamma,
\end{aligned}$$

where (C.18) holds from (C.7) and (C.19) holds since $\mathbb{E}\big[(\boldsymbol{I}_{N_t n} + \rho\boldsymbol{H}_d^T\boldsymbol{H}_d)^{-1}\big] = \sigma_z^2\boldsymbol{I}_{N_t n}$, according to (C.3). The final step is to show that $||\boldsymbol{z}^*|| \to ||\boldsymbol{z}||$ as $n \to \infty$, where $\boldsymbol{z}^* - \boldsymbol{z} = \sqrt{\frac{\frac{1}{N_t n}R_c^2+P_s}{P_w} - \frac{\mu}{P_w}}\,\boldsymbol{B}_d\boldsymbol{w}^*$. From the structure of $\boldsymbol{B}_d$, the norm of each of its rows is less than N_t, and hence the variance of each of the elements of $\boldsymbol{B}_d\boldsymbol{w}^*$ is no more than N_t. Since $\lim_{n\to\infty}\frac{1}{N_t n}R_c^2 = \frac{P_x}{N_t}$ for a covering-good lattice, it can be shown using Chebyshev's inequality [76] that the elements of $\sqrt{\frac{\frac{1}{N_t n}R_c^2+P_s}{P_w} - \frac{\mu}{P_w}}\,\boldsymbol{B}_d\boldsymbol{w}^*$ vanish and $|z_i^* - z_i| \to 0$ as $n \to \infty$ for all $i \in \{1,\ldots,n\}$, as follows

$$\begin{aligned}
\mathbb{P}\big(|z_i^* - z_i| \geq \gamma^*\kappa\big) &\leq \frac{1}{\kappa^2}, \qquad \textit{for all } \kappa > 0, \\
\mathbb{P}\big(|z_i^* - z_i| \geq \sqrt{\gamma^*}\big) &\leq \gamma^*, \qquad \text{(C.20)}
\end{aligned}$$

where $\gamma^* \triangleq \sqrt{N_t\big(\frac{\frac{1}{N_t n}R_c^2+P_s}{P_w} - \frac{\mu}{P_w}\big)}$ vanishes with n and (C.20) follows when $\kappa = \frac{1}{\sqrt{\gamma^*}}$. This concludes the proof of Lemma 9.

C.2 Proof of Corollary 8

C.2.1 $N_r \geq N_t$ and $\mathbb{E}\big[(\boldsymbol{H}^H\boldsymbol{H})^{-1}\big] < \infty$

$$
\begin{aligned}
\Delta =& C - R \\
=& \mathbb{E}\big[\log\det(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^H\boldsymbol{H})\big] - \Big(- \log\det\Big(\mathbb{E}\big[(\frac{P_x}{P_x + MP_s}\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big)\Big)^+ \\
\leq& \mathbb{E}\big[\log\det(\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^H\boldsymbol{H})\big] + \log\det\Big(\mathbb{E}\big[(\frac{P_x}{P_x + N_tP_s}\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big) \\
\leq& \log\det\Big(\boldsymbol{I}_{N_t} + \rho\mathbb{E}[\boldsymbol{H}^H\boldsymbol{H}]\Big) + \log\det\Big(\mathbb{E}\big[(\frac{P_x}{P_x + N_tP_s}\boldsymbol{I}_{N_t} + \rho\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big) \qquad \text{(C.21)} \\
<& \log\det\Big(\boldsymbol{I}_{N_t} + \rho\mathbb{E}[\boldsymbol{H}^H\boldsymbol{H}]\Big) + \log\det\Big(\mathbb{E}\big[(\rho\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big) \qquad \text{(C.22)} \\
=& \log\det\Big((\frac{1}{\rho}\boldsymbol{I}_{N_t} + \mathbb{E}[\boldsymbol{H}^H\boldsymbol{H}])\mathbb{E}\big[(\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big) \\
\leq& \log\det\Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\boldsymbol{H}^H\boldsymbol{H}]\big)\mathbb{E}\big[(\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big), \qquad \text{(C.23)}
\end{aligned}
$$

where (C.21),(C.22) follow since $\log\det(\boldsymbol{A})$ is a concave and non-decreasing function over the set of all positive definite matrices [74]. (C.23) follows since $\rho \geq 1$.

C.2.2 $N_r > N_t$ and $\boldsymbol{H}$ is Gaussian

$$
\begin{aligned}
\Delta <& \log\det\Big(\big(\boldsymbol{I}_{N_t} + \mathbb{E}[\boldsymbol{H}^H\boldsymbol{H}]\big)\mathbb{E}\big[(\boldsymbol{H}^H\boldsymbol{H})^{-1}\big]\Big) \qquad \text{(C.24)} \\
=& \log\det\big((1 + N_r)\,\frac{1}{N_r - N_t}\boldsymbol{I}_{N_t}\big) \qquad \text{(C.25)} \\
=& \; N_t\log\big(1 + \frac{N_t + 1}{N_r - N_t}\big),
\end{aligned}
$$

where (C.24) and (C.25) follow from (C.23) and Lemma 10, respectively.

C.2.3 $N_r = N_t = 1$ and $|h|$ is Nakagami-m with $m > 1$

The Nakagami-m distribution with $m > 1$ satisfies the condition $\mathbb{E}\big[\frac{1}{|h|^2}\big] < \infty$, and hence the gap is in the form (C.23). When $\mathbb{E}[|h|^2] = 1$ and $\rho \geq 1$, then

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{|h|^2}\big] &= \frac{2m^m}{\Gamma(m)} \int_0^\infty \frac{1}{x^2}\, x^{2m-1} e^{-mx^2} dx \\
&= \frac{2m^m}{\Gamma(m)} \frac{1}{2m^{m-1}} \int_0^\infty y^{m-2} e^{-y} dy \\
&= m \frac{\Gamma(m-1)}{\Gamma(m)} \\
&= 1 + \frac{1}{m-1}\,.
\end{aligned}$$

Hence, $\Delta < \log\big((1 + \mathbb{E}[|h|^2])\mathbb{E}\big[\frac{1}{|h|^2}\big]\big) = 1 + \log\big(1 + \frac{1}{m-1}\big)$.

The case where $\rho < 1$ is trivial, since [2]

$$\Delta < C = \mathbb{E}\big[\log(1 + \rho|h|^2)\big] \leq \log\big(1 + \rho\mathbb{E}[|h|^2]\big) < 1, \tag{C.26}$$

and hence $\Delta < 1 + \log\big(1 + \frac{1}{m-1}\big)$ is universal for all ρ.

[2] The result in (C.26) holds for any fading distribution with $\mathbb{E}\big[|h|^2\big] = 1$.

C.2.4 $N_r = N_t = 1$ and $|h|$ is Rayleigh

When $|h|$ is a Rayleigh random variable, $|h|^2$ is exponentially distributed with $\mathbb{E}[|h|^2] = 1$. For the case where $\rho \geq 1$,

$$\begin{aligned}
\Delta \leq & \log\Big(\big(1+\rho\mathbb{E}[|h|^2]\big)\mathbb{E}\big[\frac{1}{\rho|h|^2+\frac{\rho}{\frac{P_x}{P_w}+\frac{P_s}{P_w}}}\big]\Big) && \text{(C.27)}\\
= & \log\Big((\frac{1}{\rho}+1)\,\mathbb{E}\big[\frac{1}{|h|^2+\frac{1}{\frac{P_x}{P_w}+\frac{P_s}{P_w}}}\big]\Big)\\
\leq & 1+\log\Big(\mathbb{E}\big[\frac{1}{|h|^2+\frac{1}{\frac{P_x}{P_w}+\frac{P_s}{P_w}}}\big]\Big) && \text{(C.28)}\\
\leq & 1+\log\Big(\mathbb{E}\big[\frac{1}{|h|^2+\frac{1}{2\kappa}}\big]\Big)\\
= & 1+\log\big(\int_0^\infty \frac{e^{-x}}{x+\frac{1}{2\kappa}}dx\big)\\
= & 1+\log\big(e^{\frac{1}{2\kappa}}\int_{\frac{1}{2\kappa}}^\infty \frac{e^{-y}}{y}dy\big)\\
< & 0.48+\log\big(\log(1+2\kappa)\big) && \text{(C.29)}\\
< & 1.48+\log\big(\log(1+\kappa)\big), && \text{(C.30)}
\end{aligned}$$

where (C.27) follows from (C.21), (C.28) follows since $\rho \geq 1$ and (C.29) follows from Lemma 11. Recall $\kappa \triangleq \max\{\frac{P_x}{P_w}, \frac{P_s}{P_w}, 1\}$. When $\rho < 1$, the gap to capacity is within one bit, and hence (C.30) is also an upper bound for the gap in this regime.

REFERENCES

[1] R. de Buda, "Some optimal codes have structure," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 893–899, Aug. 1989.

[2] T. Linder, C. Schlegal, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1735–1737, Sep. 1993.

[3] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

[4] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, Jan. 1998.

[5] U. Erez and R. Zamir, "Achieving 1/2log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[6] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[7] H. El-Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, Jun. 2004.

[8] N. Prasad and M. Varanasi, "Analysis and optimization of diagonally layered lattice schemes for MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1162–1185, Mar. 2008.

[9] P. Dayal and M. Varanasi, "An algebraic family of complex lattices for fading channels with application to space-time codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4184–4202, Dec. 2005.

[10] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[11] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 323–340, Jan. 2015.

[12] S. Vituri, "Dispersion analysis of infinite constellations in ergodic fading channels," *CoRR*, vol. abs/1309.4638, 2015. [Online]. Available: http://arxiv.org/abs/1309.4638v2

[13] L. Luzzi and R. Vehkalahti, "Almost universal codes achieving ergodic MIMO capacity within a constant gap," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3224–3241, May 2017.

[14] L. Liu and C. Ling, "Polar codes and polar lattices for independent fading channels," *CoRR*, vol. abs/1601.04967, 2016. [Online]. Available: http://arxiv.org/abs/1601.04967

[15] A. Campello, C. Ling, and J. Belfiore, "Algebraic lattice codes achieving the capacity of the ergodic fading channel," in *IEEE Information Theory Workshop (ITW)*, Sep. 2016.

[16] Y. Song and N. Devroye, "Lattice codes for the Gaussian relay channel: Decode-and-forward and compress-and-forward," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4927–4948, Aug. 2013.

[17] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[18] A. Özgür and S. Diggavi, "Approximately achieving gaussian relay network capacity with lattice-based qmf codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8275–8294, Dec. 2013.

[19] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3450–3482, Jun. 2014.

[20] S. C. Lin, P. H. Lin, C. P. Lee, and H. J. Su, "Filter and nested lattice code design for mimo fading channels with side-information," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1489–1494, Jun. 2011.

[21] I. Bergel, D. Yellin, and S. Shamai, "Dirty paper coding with partial channel state information," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2014, pp. 334–338.

[22] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566–4592, Sep. 2010.

[23] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.

[24] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[25] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[26] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[27] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[28] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[29] S. Shamai and A. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels. I," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1877–1894, Nov. 1997.

[30] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels. I. ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[31] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[32] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[33] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[34] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[35] S. Rini and S. Shamai, "On the dirty paper channel with fast fading dirt," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, Jun. 2015, pp. 2286–2290.

[36] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[37] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, 1958.

[38] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[39] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

[40] D. Tuninetti and S. Shamai, "On two-user fading Gaussian broadcast channels with perfect channel state information at the receivers," in *Information Theory Proceedings (ISIT), 2003 IEEE International Symposium on*, Jun. 2003, pp. 345–345.

[41] D. Tse and R. Yates, "Fading broadcast channels with state information at the receivers," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3453–3471, Jun. 2012.

[42] A. Jafarian and S. Vishwanath, "The two-user Gaussian fading broadcast channel," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, Jul. 2011, pp. 2964–2968.

[43] S. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels:the scalar upper bound and loss in degrees of freedom," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[44] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[45] H. Sato, "The capacity of the Gaussian interference channel under strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. 27, no. 6, pp. 786–788, Nov. 1981.

[46] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[47] L. Sankar, X. Shang, E. Erkip, and H. V. Poor, "Ergodic fading interference channels: Sum-capacity and separability," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2605–2626, May 2011.

[48] R. K. Farsani, "The capacity region of the wireless ergodic fading interference channel with partial CSIT to within one bit," in *2013 IEEE Int. Symp. on Inf. Theory*, Jul. 2013, pp. 759–763.

[49] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[50] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[51] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, NY, USA, 2014.

[52] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[53] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, NY, USA, 2005.

[54] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[55] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[56] G. D. Forney, "Coset codes. I. Introduction and geometrical classification," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1123–1151, Sep. 1988.

[57] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[58] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 206–218, Feb. 1998.

[59] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[60] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, Jul. 2011, pp. 194–198.

[61] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[62] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[63] K. N. Pappi, G. K. Karagiannidis, and R. Schober, "How sensitive is compute-and-forward to channel estimation errors?" in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, Jul. 2013, pp. 3110–3114.

[64] C. S. Vaze and M. K. Varanasi, "Dirty paper coding for fading channels with partial transmitter side information," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Oct. 2008, pp. 341–345.

[65] A. Bennatan and D. Burshtein, "On the fading-paper achievable region of the fading MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 100–115, Jan. 2008.

[66] W. Zhang, S. Kotagiri, and J. N. Laneman, "Writing on dirty paper with resizing and its application to quasi-static fading broadcast channels," in *2007 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2007, pp. 381–385.

[67] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, NY, USA, 2011.

[68] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3450–3482, Jun. 2014.

[69] A. Campello, C. Ling, and J. C. Belfiore, "Algebraic lattice codes achieve the capacity of the compound block-fading channel," in *2016 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 910–914.

[70] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.

[71] R. Lyons, "Strong laws of large numbers for weakly correlated random variables," *Michigan Math. J.*, vol. 35, no. 3, pp. 353–359, 1988.

[72] N. Etemadi, "Convergence of weighted averages of random variables revisited," in *Proc. of the American Mathematical Society*, vol. 134, no. 9, Sep. 2006, pp. 2739–2744.

[73] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEEE Proc. Radar, Sonar and Navig.*, vol. 147, no. 4, pp. 162–168, Aug. 2000.

[74] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, NY, USA, 2004.

[75] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*. Dover Publications, 1972.

[76] H. Kobayashi, B. Mark, and W. Turin, *Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance*. Cambridge University Press, 2011.

BIOGRAPHICAL SKETCH

Ahmed Monier Hindy was born in Alexandria, Egypt. He received his B.Sc. degree with honors in Electronics and Communications Engineering from Alexandria University, Egypt, in 2010 and his M.Sc. degree in Communication and Information Technology from Nile University, Egypt, in 2012. He is a PhD candidate in Electrical Engineering at The University of Texas at Dallas, Richardson, TX, USA.

He has been with InterDigital Communications Inc., Melville, NY, USA, since January 2017. He was a research/teaching assistant at The University of Texas at Dallas, Richardson, TX, USA, from 2012 to 2016, and a research assistant at Nile University, Cairo, Egypt, from 2010 to 2012. He was a co-op wireless engineer at Vodafone, Cairo, Egypt, from 2009 to 2010. His research interests include information theory, signal processing and wireless cellular systems, such as LTE-A and beyond. He received the Ericsson Graduate Fellowship in 2015 and the Erik Jonsson Graduate Fellowship in 2012.

CURRICULUM VITAE

Ahmed M. Hindy

June 30, 2017

Contact Information:

Department of Electrical Engineering
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

Voice: (972) 883-6433
Email: ahmed.hindy@utdallas.edu

Educational History:

B.Sc., Electrical Engineering, Alexandria University, Egypt, 2010
M.Sc., Electrical Engineering, Nile University, Egypt, 2012

Approaching the Ergodic Capacity of Wireless Networks with Lattice Codes
Ph.D. Dissertation
Electrical Engineering, The University of Texas at Dallas
Advisor: Prof. Aria Nosratinia

Employment History:

PhD Research Fellow, InterDigital Communications Inc., January 2017 – present
Research Assistant, The University of Texas at Dallas, May 2015 – December 2016
Teaching Assistant, The University of Texas at Dallas, August 2012 – May 2015
Research Assistant, Nile University, Egypt, October 2010 – August 2012
Co-op Wireless Engineer, Vodafone, Egypt, November 2009 – July 2010
RF Engineering Intern, Huawei Technologies, Egypt, June 2009 – September 2009

Professional Recognitions and Honors:

Advanced Graduate Teaching Certificate, 2016
Ericsson Graduate Fellowship, 2015
Erik Jonsson Graduate Fellowship, UTD, 2012
Certificate of Appreciation, Vodafonne, 2010
Award of Excellence, Alexandria University, 2009
Egyptian Ministry of Education Award (Top 100 Students Nationwide), 2005

Professional Memberships:

Institute of Electrical and Electronics Engineers (IEEE), 2007–present
IEEE Communications Society (IEEE ComSoc), 2014–present

Journal Publications:

1. A. Hindy and A. Nosratinia, "Lattice Coding and Decoding for Multiple-Antenna Ergodic Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 1873 - 1885, May 2017.

2. A. Hindy and A. Nosratinia, "Ergodic Fading MIMO Dirty Paper and Broadcast Channels: Capacity Bounds and Lattice Strategies," *IEEE Transactions on Wireless Communications, to appear*, 2017.

3. A. Hindy and A. Nosratinia, "On the Separability of Ergodic Fading MIMO Channels: A Lattice Coding Approach," *submitted to IEEE Transactions on Communications*, 2017.

4. A. Hindy and A. Nosratinia, "On the Ergodic Interference Channel under Lattice Coding and Decoding," *IEEE Transactions on Information Theory, under preparation*, 2017.

Conference Publications:

1. A. Hindy and A. Nosratinia, "On the Universality of Lattice Codes for a Class of Ergodic Fading Channels," *in IEEE International Symposium on Information Theory (ISIT), 2017.*

2. A. Hindy and A. Nosratinia, "Approaching the Ergodic Capacity of the MIMO Channel with Lattice Codes," *in IEEE Global Communications Conference (GLOBECOM), 2016.*

3. A. Hindy and A. Nosratinia, "On the Fading MIMO Dirty Paper Channel with Lattice Coding and Decoding," *in IEEE Global Communications Conference (GLOBECOM), 2016.*

4. A. Hindy and A. Nosratinia, "The capacity of fast fading channels using lattice codes: Is separability necessary?," *in International Symposium on Information Theory and its Applications (ISITA), 2016.*

5. A. Hindy and A. Nosratinia, "Lattice codes under imperfect channel state information," *in International Symposium on Information Theory and its Applications (ISITA), 2016.*

6. A. Hindy and A. Nosratinia, "Lattice strategies for the ergodic fading dirty paper channel," *in IEEE International Symposium on Information Theory (ISIT), 2016.*

7. A. Hindy and A. Nosratinia, "Achieving the ergodic capacity with lattice codes," *in IEEE International Symposium on Information Theory (ISIT), 2015.*

8. A. Hindy and A. Nosratinia, "Approaching the ergodic capacity with lattice coding," *in IEEE Global Communications Conference (GLOBECOM), 2014.*

9. A. Hindy, A. El-Keyi, M. Nafie and A. Tulino, "A proper throughput-leakage balance for downlink cellular networks," *in IEEE Wireless Communications and Networking Conference (WCNC), 2014.*

10. A. Hindy, A. El-Keyi, M. Nafie and A. Tulino, "Maximizing the signal to leakage ratio in downlink cellular networks," *in International Conference on Communications and Information Technology (ICCIT), 2013*.

11. M. Fadel, A. Hindy, A. El-Keyi, M. Nafie, O. Koyluoglu and A. Tulino, "Resource allocation for throughput enhancement in cellular shared relay networks," *in IEEE Sarnoff Symposium, 2012*.