Mathematical Engineering

Victor Beresnevich
Alister Burr
Bobak Nazer
Sanju Velani *Editors*

Number Theory Meets Wireless Communications

Springer

Mathematical Engineering

Today, the development of high-tech systems is unthinkable without mathematical modeling and analysis of system behavior. As such, many fields in the modern engineering sciences (e.g. control engineering, communications engineering, mechanical engineering, and robotics) call for sophisticated mathematical methods in order to solve the tasks at hand.

The series Mathematical Engineering presents new or heretofore little-known methods to support engineers in finding suitable answers to their questions, presenting those methods in such manner as to make them ideally comprehensible and applicable in practice.

Therefore, the primary focus is—without neglecting mathematical accuracy—on comprehensibility and real-world applicability.

To submit a proposal or request further information, please use the PDF Proposal Form or contact directly: Dr. Thomas Ditzinger (thomas.ditzinger@springer.com)

Indexed by SCOPUS, zbMATH, SCImago.

More information about this series at http://www.springer.com/series/8445

Victor Beresnevich • Alister Burr • Bobak Nazer • Sanju Velani
Editors

Number Theory Meets Wireless Communications

Springer

Editors
Victor Beresnevich
Department of Mathematics
University of York
York, UK

Alister Burr
Department of Electronic Engineering
University of York
York, UK

Bobak Nazer
Department of Electrical and Computer Engineering
Boston University
Boston, MA, USA

Sanju Velani
Department of Mathematics
University of York
York, UK

ISSN 2192-4732 ISSN 2192-4740 (electronic)
Mathematical Engineering
ISBN 978-3-030-61302-0 ISBN 978-3-030-61303-7 (eBook)
https://doi.org/10.1007/978-3-030-61303-7

Mathematics Subject Classification: 11K60, 11H06, 11J83, 11H71, 94A15, 94A40, 11J25

Preface

Powerful techniques from various areas of Number Theory have played important roles in breakthrough developments in areas of Wireless Communications. These include the impact of geometry of numbers, Diophantine approximation and algebraic number theory on lattice coding and interference alignment. This book introduces and describes some of these developments as well as the techniques that have made them possible. It lays particular emphasis on those that are at the forefront of current research. The chapters are all written by leading researchers in both areas. They present the state-of-the-art research, which illustrates the deep interaction between number theory and wireless communications. Together, they show that there is currently great scope to develop the mutual understanding of methods and problems.

The book has been developed from lectures given at the international meeting "Workshop on Interactions Between Number Theory and Wireless Communication" held at the University of York in July 2016. Details, including list of participants, programme and slides of talks, can be found at:

https://www.york.ac.uk/maths/events/2016/workshop-interactions-between-number-theory-wirele/

The primary goal of the workshop was to inspire both early career and established researchers to consolidate and build new and exciting bridges between Number Theory and Wireless Communications. Naturally, this is also the overarching goal of this book. With this in mind, we encouraged the speakers to develop their written contributions in an expository way and to provide an overview of current tools and developments. Each chapter thus foregrounds the main concepts behind the topic under consideration while keeping technicalities to a bare minimum. The chapters thus offer direct and accessible information about highly exciting current research developments to researchers in both Number Theory and Communication Theory. Breaking down the superficial "language barrier" between the two disciplines is key to understanding the respective central problems and is the first step towards fruitful collaboration and progress.

To the best of our knowledge, this book is the first volume jointly edited by individuals working in Number Theory and Communication Theory. We hope it provides a unique insight into key concepts, cutting-edge results, and modern techniques that play an essential role in contemporary research. Great effort has been made to present the material in a manner that is accessible to new researchers, including Ph.D. students. The book will also be useful for established researchers working in Number Theory or Wireless Communications who wish to broaden their outlook and contribute towards the deep interplay between the two.

Many communication techniques involve choosing discrete sets of points that represent information being sent over a communication channel. These discrete sets represent a key element of codes and are often conveniently chosen to have a linear structure. The presence of linear structure allows for efficient and low-complexity decoding. Furthermore, it provides powerful properties that underpin state-of-the-art techniques for managing interference and other challenges in wireless networks. In number theory, discrete linear structures are natural objects of study within the geometry of numbers and Diophantine approximation. Roughly speaking, the geometry of numbers characterizes geometric properties (such as packing and covering radii and Minkowski minima) of linear structures known as lattices. On the contrary, Diophantine approximation studies the properties of linear maps on such linear structures. The basis for the theory of Diophantine approximation is Dirichlet's classical theorem based on the Pigeonhole principle. The opening chapter discusses the role it plays in the world of wireless communications. The authors present an informal discussion of aspects of wireless communications via a series of basic examples. These allow them to introduce a variety of concepts (such as badly approximable, singular and well approximable points) and aspects (such as probabilistic and manifolds theories) from Diophantine approximation while explaining their role in wireless communications. In particular, they introduce a new concept of jointly non-singular points and use it to improve a well-known result of Motahari et al. regarding the Degrees of Freedom (DoF) of a two-user X-channel. An overarching goal of Chap. 1 is to provide an answer to the question: *What is the role of number theory in the world of wireless communications?*

Consider multiple transmitters and receivers that communicate with each other across a shared wireless channel. The two main challenges to establishing reliable communication between users are the noise introduced by the channel and the interference between simultaneously transmitted signals. Over the past few decades, experts in network information theory have strived to determine the fundamental limits of reliable communication over multi-user channels. At the same time, they attempted to realize network architectures that, in practice, could approach these limits. In Chap. 2, the authors discuss the recent developments in network information theory based on the use of lattice codebooks (i.e. codebooks that are a subset of a lattice over $\mathbb{R}^n$). The inherent linearity of lattice codebooks can be effectively used as a building block for communication strategies that operate beyond the performance available for classical coding schemes. In general, the performance of these lattice-based strategies is determined by how closely the channel coefficients can be approximated by integer coefficients. In other words, the

performance is intertwined with the theory of Diophantine approximation. Overall, this chapter provides a unified view of recent results that connect the performance of the compute-and-forward strategy of recovering an integer-linear combination of codewords to Diophantine approximation bounds. The chapter concludes by highlighting scenarios in which novel applications of Diophantine approximation, such as non-asymptotic approximation bounds over manifolds, have the potential to yield new exciting results in network information theory.

With the roll out of the fifth-generation (5G) wireless systems, constructing efficient space–time codes offering complexity reduction is crucial for many applications including massive multiple-input multiple-output (MIMO) systems. Traditionally, space–time codes have been developed in the context of point-to-point MIMO communications. However, today's wireless networks need to accommodate numerous types of applications and devices. In view of this, the so-called distributed space–time codes have become a prominent area of research. In practice, such codes often exhibit a high decoding complexity. Algebraic number theory and lattice theory provide a framework for overcoming this issue. In Chap. 3, the authors give an overview on the topic of fast decodable algebraic space–time codes. More precisely, the chapter provides a basic introduction to space–time coding and brings to the forefront the powerful algebraic tools needed for the design and construction of such codes. In particular, it describes the algebraic techniques used for reducing the decoding complexity of both single-user and multiuser space–time codes. The key lies in utilizing the carefully chosen underlying algebraic structure. The necessary background to both the lattice theory and algebraic number theory is provided. The chapter concludes by describing explicit construction methods for fast-decodable algebraic space–time codes. These are crucial for practical implementation.

The problem of finding the densest arrangement of spheres in an n-dimensional Euclidean space has been extensively pursued in mathematics. It is a classical and central problem in the geometry of numbers. The celebrated Minkowski–Hlawka theorem (dating back to 1943) states that there is a lattice in $\mathbb{R}^n$, such that the corresponding best packing of spheres with centres at the lattice points has density greater than $\zeta(n)\,2^{-(n-1)}$—a constant dependent on the dimension. This is an existence statement, and to date, excluding a handful of dimensions, no explicit lattice construction achieving the Minkowski–Hlawka lower bound for the sphere packing density is known. The proof uses probabilistic methods to analyse a random ensemble of lattices rather than individual instances. Recent improvements to the Minkowski-Hlawka lower bound exploit lattices with inherited algebraic structures. For example, Venkatesh has successfully used the structure of cyclotomic number field lattices to obtain a super-linear improvement. The sphere packing problem has well-established and deep connections to coding theory. Indeed, building upon the Shannon's seminal work from 1948, it was shown in the nineties that the same random ensembles used to produce lattices that give rise to the Minkowski–Hlawka lower bound can be used to construct optimal lattice codes for the basic additive white Gaussian noise (AWGN) communication channels. In other words, for such channels the probabalistic strategy of Minkowski and Hlawka leads to the existence

of capacity-achieving codes. Although the AWGN channel is a good model for deep-space or satellite channels, which operate over a line of sight, modern wireless communications call for more general models, which take into account the sending of information propagated over different media (e.g. fading channels) via multiple transmit and receive antennas (e.g. MIMO channels) and to various users (e.g. relay networks). Such channels cannot be abstracted into a simple AWGN model and require a different strategy. With this in mind, it is fruitful to consider lattices with additional algebraic (multiplicative) structure, often inherited by the properties of number fields. Indeed, the cyclotomic lattices exploited by Venkatesh play a crucial role in some recent channel constructions. In Chap. 4, the authors start by providing a self-contained exposition of random lattices and the sphere packing problem. This includes both the classical aspects and the recent developments. Regarding the latter, the use of algebraic number theory to utilize the structure of algebraic lattices is brought to the forefront. A general construction that naturally incorporates a number of important families of algebraic lattices (such as cyclotomic, Lipschitz and Hurwitz lattices) is described. The emphasis then switches to describing how such lattices can be applied to build effective, reliable and secure transmission schemes for wireless communications. The main focus on the application side is threefold: (i) to block fading, (ii) to certain forms of MIMO channels and (iii) to improving information security.

As alluded to in the discussion above, one of the classical problems in information and coding theory is that of designing codes that can approach the capacity of the AWGN channel. One promising approach is to draw codewords from a lattice and draw upon deep results from the geometry of numbers to establish performance bounds. However, the AWGN channel model is not sufficient for modelling the phenomena observed in wireless communication scenarios. Recent efforts have thus shifted towards designing codes that can approach the capacity of fading channels, which model the wireless medium via a random matrix multiplication of the channel input followed by the addition of Gaussian noise. Here, it is also of interest to design good lattice codebooks that can operate near the capacity. In Chap. 5, the authors begin by reviewing the Hermite invariant approach to the design of lattice codes for classical AWGN channels. They then propose a reduced Hermite invariant criterion for the design of lattice codes for fading channels. Using this criterion, they are able to translate the problem of operating within a constant gap of the fading capacity to the problem of finding totally complex number fields with the smallest determinant. Drawing upon powerful results from this area, they demonstrate the existence of lattice codebooks that can operate within a constant gap of the fading capacity. They then discuss the limitations of this approach and outline a promising direction based on the construction of lattice codes from ideals. They conclude the chapter with a discussion on the connection between the reduced Hermite invariant and homogeneous forms.

A key property of the wireless medium is that a receiver's observation can be written as a linear superposition of all transmitted signals and Gaussian noise. By employing codebooks based on nested lattices at each transmitter, this property can be leveraged in order to allow the receiver to directly recover a function of

the codewords. The compute-and-forward framework was proposed to characterize the achievable rates for recovering integer-linear combinations, and lattice network coding was subsequently proposed to connect this framework to module theory from abstract algebra, which in turns allows for a much richer set of lattice codebook constructions. In Chap. 6, the authors provide a comprehensive overview of these frameworks, beginning with a review of necessary concepts from abstract algebra, lattice codes and classical construction. Furthermore, methods for obtaining performance bounds for compute-and-forward are also discussed. They then propose multilevel lattice codes as a powerful method for reducing the decoding complexity while maintaining the performance advantages of lattice codes. They introduce detailed procedures for constructing such multilevel lattices, including a novel elementary divisor construction, which captures prior methods as special cases. From here, they generalize compute-and-forward and lattice network coding to utilize multilevel lattices and demonstrate that this approach can yield a more efficient method for decoding multiple messages. They conclude by proposing an iterative decoding procedure for multilevel lattice codes and demonstrate its advantages via numerical simulations.

Shannon's beautiful theorem concerning the existence of capacity-achieving codes for an AWGN channel makes fundamental use of a random coding argument. In short, independent identically distributed (i.i.d.) random ensembles according to some codeword distribution are exploited to prove the existence of "optimal" codes. In Chap. 7, the final chapter, the authors begin by reviewing the main steps in Shannon's proof, in particular the use of the i.i.d. random ensembles in the achievability part. They then revisit the achievability part from the viewpoint of exploiting random structured ensembles such as random linear codes and random lattice codes. For certain scenarios (e.g. those involving relay networks or physical layer secrecy), random structured codes achieve better "rates" than random i.i.d. codes. Furthermore, random linear codes allow for computationally efficient encoding (since the encoding operation essentially involves simple matrix multiplication), and random lattice codes allow for lattice decoding, which, for example, enjoy lower complexity than maximum likelihood (ML) decoding. The main goal of the chapter is to provide an accessible account of recent developments and simplifications in the use of random structured codes in achievability proofs. The focus is on addressing the two questions: *Can random linear codes achieve the discrete memoryless channel (DMC) capacity?* and, *Can random lattice codes achieve the additive white Gaussian noise (AWGN) channel capacity?* These two questions are discussed separately but in a parallel manner. Indeed, the introduced framework unifies the approaches for DMC and AWGN channels into a streamlined analysis.

York, UK — Victor Beresnevich
York, UK — Alister Burr
Boston, MA, USA — Bobak Nazer
York, UK — Sanju Velani
June 2020

Acknowledgements

As part of the organizing committee of the 2016 workshop, we would like to thank all the people who helped to make this meeting highly enjoyable and successful. We are also extremely grateful to everybody who made this volume possible. In particular, we would like to acknowledge the help of:

- Dr. Rémi Lodh at Springer for all his encouragement, patience, good humour and fantastic support in producing this book—we have truly enjoyed working with you;
- the distinguished speakers for their inspiring talks;
- all the contributors to this volume, for their hard work in producing high-quality chapters and for having faith in us during its 3-year gestation period;
- all the reviewers of the chapters, for their work in producing informative reports, which have helped improve the accuracy and clarity of the material presented;
- all the participants of the workshop for creating an informal and inspiring atmosphere.

Finally, we would like to stress that the workshop was fully funded from the EPSRC Programme Grant: EP/J018260/1. The workshop and in turn this book would not have been possible without this generous support.

Contents

Contributors

Amaro Barreal Department of Mathematics and Systems Analysis, Aalto University, Aalto, Finland

Victor Beresnevich Department of Mathematics, University of York, York, UK

Alister G. Burr Department of Electronic Engineering, University of York, York, UK

Antonio Campello Wellcome Trust, London, UK

Chen Feng The School of Engineering, University of British Columbia (Okanagan Campus), Kelowna, BC, Canada

Camilla Hollanti Department of Mathematics and Systems Analysis, Aalto University, Aalto, Finland

Yu-Chih (Jerry) Huang Institute of Communications Engineering, National Chiao Tung University, Hsinchu City, Taiwan

Cong Ling Department of Electrical and Electronic Engineering, Imperial College London, London, UK

Laura Luzzi Laboratoire ETIS (UMR 8051, CY Université, ENSEA, CNRS) Cergy-Pontoise, France

Krishna R. Narayanan Department of Electrical & Computer Engendering, Texas A&M University, College Station, TX, USA

Bobak Nazer Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA

Or Ordentlich The Rachel and Selim Benin School of Computer Science, Engineering Hebrew University of Jerusalem, Jerusalem, Israel

Renming Qi The School of Engineering, University of British Columbia (Okanagan Campus), Kelowna, BC, Canada

Roope Vehkalahti Department of Communications and Networking, Aalto University, Helsinki, Finland

Sanju Velani Department of Mathematics, University of York, York, UK

Yi Wang Department of Electronic Engineering, University of York, York, UK

Chapter 1
Number Theory Meets Wireless Communications: An Introduction for Dummies Like Us

Victor Beresnevich and Sanju Velani

Abstract In this chapter we introduce the theory of Diophantine approximation via a series of basic examples from information theory relevant to wireless communications. In particular, we discuss Dirichlet's theorem, badly approximable points, Dirichlet improvable and singular points, the metric (probabilistic) theory of Diophantine approximation including the Khintchine-Groshev theorem and the theory of Diophantine approximation on manifolds. We explore various number theoretic approaches used in the analysis of communication characteristics such as Degrees of Freedom (DoF). In particular, we improve the result of Motahari et al. regarding the DoF of a two-user X-channel. In essence, we show that the total DoF can be achieved for all (rather than almost all) choices of channel coefficients with the exception of a subset of strictly smaller dimension than the ambient space. The improvement utilises the concept of jointly non-singular points that we introduce and a general result of Kadyrov et al. on the δ-escape of mass in the space of lattices. We also discuss follow-up open problems that incorporate a breakthrough of Cheung and more generally Das et al. on the dimension of the set of singular points.

1.1 Basic Examples and Fundamentals of Diophantine Approximation

Let us start by addressing a natural question that a number theorist or more generally a mathematician who has picked up this book may well ask: *what is the role of number theory in the world of wireless communications?* We will come clean straightaway and say that by number theory we essentially mean areas such as Diophantine approximation and the geometry of numbers, and by wireless communication we essentially mean the design and analysis of lattice/linear codes for wireless communications which thus falls in the realm of information theory. To

V. Beresnevich (✉) · S. Velani
Department of Mathematics, University of York, York, UK
e-mail: victor.beresnevich@york.ac.uk; sanju.velani@york.ac.uk

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*, Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_1

begin with, with this confession in mind, let us start by describing the role of one-dimensional Diophantine approximation. Recall, that at the heart of Diophantine approximation is the classical theorem of Dirichlet on rational approximations to real numbers.

Theorem 1.1 (Dirichlet, 1842) *For any $\xi \in \mathbb{R}$ and any $Q \in \mathbb{N}$ there exist $p, q \in \mathbb{Z}$ such that*

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{qQ} \qquad \text{and} \qquad 1 \leq q \leq Q\,. \tag{1.1}$$

The proof can be found in many elementary number theory books and makes use of the wonderfully simple yet powerful Pigeonhole Principle: if n objects are placed in m boxes and $n > m$, then some box will contain at least two objects. See, for example, [16, §1.1] for details. An easy consequence of the above theorem is the following statement.

Corollary 1.1 *Let $\xi \in \mathbb{R} \setminus \mathbb{Q}$, that is ξ is a real irrational number. Then there exist infinitely many reduced rational fractions p/q $(p, q \in \mathbb{Z})$ such that*

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^2}. \tag{1.2}$$

The following exposition illustrates one of the many aspects of the role of Diophantine approximation in wireless communication. In particular, within this section we consider a basic example of a communication channel which brings into play the theory of Diophantine approximation. In Sect. 1.2 we consider a slightly more sophisticated example which also brings into play the theory of Diophantine approximation in higher dimensions. This naturally feeds into Sect. 1.3 in which the role of the theory of Diophantine approximation of dependent variables is discussed. The latter is also referred to as Diophantine approximation on manifolds since the parameters of interest are confined by some functional relations. To begin with, we consider a 'baby' example of a communication channel intended to remove the language barrier for mathematicians and explicitly expose an aspect of communications that invites the use of Diophantine approximation.

1.1.1 A 'baby' Example

Suppose there are two *users* S_1 and S_2 wishing to send (*transmit*) their *messages* u_1 and u_2 respectively along a shared (radio/wireless) communication channel to a *receiver* R. For obvious reasons, users are often also referred to as transmitters. Suppose for simplicity that $u_1, u_2 \in \{0, 1\}$. Typically, prior to transmission, every message is encoded with what is called a *codeword*. Suppose that $x_1 = x_1(u_1)$ and $x_2 = x_2(u_2)$ are the codewords that correspond to u_1 and u_2. In general, x_1

and x_2 could be any functions on the set of messages. In principle, one can take $x_1 = u_1$ and $x_2 = u_2$. When the codewords x_1 and x_2 are being transmitted along a wireless communication channel, there is normally a certain degree of fading of the transmitted signals. This for instance could be dependent on the distance of the transmitters from the receiver and the reflection caused by obstacles such as buildings in the path of the signal. Let h_1 and h_2 denote the fading factors (often referred to as *channel gains* or *channel coefficients* or *paths loss*) associated with the transmission of signals from S_1 and S_2 to R respectively. These are strictly positive numbers and for simplicity we will assume that their sum is one: $h_1 + h_2 = 1$. Mathematically, the meaning of the channel coefficients is as follows: if S_i transmits signal x_i, the receiver R observes $h_i x_i$. However, due to fundamental physical properties of wireless medium, when S_1 and S_2 simultaneously use the same wireless communication channel, R will receive the superposition of $h_1 x_1$ and $h_2 x_2$, that is

$$y = h_1 x_1 + h_2 x_2 \,. \tag{1.3}$$

For instance, assuming that $x_1 = u_1$ and $x_2 = u_2$, the outcomes of y are

$$y = \begin{cases} 0 & \text{if} \quad u_1 = u_2 = 0 \\ h_1 & \text{if} \quad u_1 = 0 \text{ and } u_2 = 1 \,, \\ h_2 & \text{if} \quad u_1 = 1 \text{ and } u_2 = 0 \,, \\ 1 = h_1 + h_2 & \text{if} \quad u_1 = u_2 = 1 \,. \end{cases} \tag{1.4}$$

A pictorial description of the above setup is given below in Fig. 1.1.

The ultimate goal is for the receiver R to identify (*decode*) the messages u_1 and u_2 from the observation of y. For example, with reference to (1.4), assuming the channel coefficients h_1 and h_2 are known at the receiver and are different, that is $h_1 \neq h_2$, the receiver is obviously able to do so. However, in real life there is

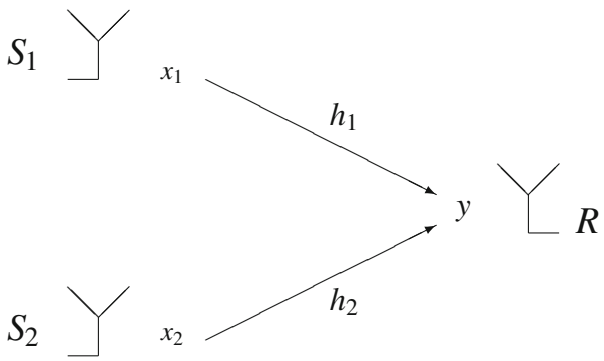


Fig. 1.1 Two user multiple access channel (no noise)

Fig. 1.2 Separation of intervals of radius $|z|$ around each possible outcome of y which contain the values of y'

always a degree of error in the transmission process, predominantly caused by the received signal y being corrupted by (additive) *noise*. The noise can result from a combinations of various factors including the interference of other users and natural electromagnetic radiation. In short, if z denotes the noise, then instead of (1.3), R receives the signal

$$y' = y + z = h_1x_1 + h_2x_2 + z\,. \tag{1.5}$$

Equation (1.5) represents one the simplest models of what is known as an *Additive White Gaussian Noise Multiple Access Channel* (AWGN-MAC), see Chap. 2 for a formal definition. As before, the goal for the receiver R remains to decode the messages u_1 and u_2, but now from the observation of $y' = y + z$. Let $d_{\min}$ denote the *minimum distance* between the four outcomes of y. Then as long as the absolute value $|z|$ of the noise is strictly less than $d_{\min}/2$, the receiver is able to recover y and consequently the messages u_1 and u_2 from the value of y'. This is simply due to the fact that the intervals of radius $d_{\min}/2$ centered at the four outcomes of y are disjoint and y' will lie in exactly one of these intervals, see Fig. 1.2. In other words, R is able to identify y by rounding y' to the closest possible outcome of y.
For example, it is easy to see that the maximum separation between the four outcomes given by (1.4) is attained when $h_1 = 1/3$ and $h_2 = 2/3$. In this case $d_{\min} = 1/3$, and we are able to recover the messages u_1 and u_2 assuming that $|z| < 1/6$. The upshot of the above discussion is the following simple but fundamental conclusion.

Conclusion The greater the mutual separation $d_{\min}$ of the outcomes of y, the better the tolerance for noise we have during the transmission of the signal.

In information theory achieving good separation between received signals translates into obtaining good lower bounds on the fundamental parameters of communication channels such as Rates-of-Communications, Channel Capacity and Degrees-of-Freedom, see Chap. 2 for formal definitions of these notions. Within this chapter we will concentrate on the role of Diophantine approximation in answering the following natural and important question:

How can a good separation of received signals be achieved and how often?

Indeed, to some extent, answering this and related questions using the tools of Diophantine approximation, algebraic number theory and the geometry of numbers is a reoccurring theme throughout the whole book. We will solely use *linear encoding* to achieve 'good' separation. In particular, within the above 'baby'

example, one is able to achieve the optimal separation ($d_{\min} = 1/3$) at the receiver regardless of the values of h_1 and h_2 by applying the following simple linear encoding of the messages u_1 and u_2:

$$x_1 = \frac{1}{3}h_1^{-1}u_1 \qquad \text{and} \qquad x_2 = \frac{2}{3}h_2^{-1}u_2\,.$$

Indeed, before taking noise into consideration, under the above encoding the received signals become

$$y = h_1x_1 + h_2x_2 = \tfrac{1}{3}u_1 + \tfrac{2}{3}u_2 = \begin{cases} 0 & \text{if} \quad u_1 = u_2 = 0\,, \\ 1/3 & \text{if} \quad u_1 = 0 \text{ and } u_2 = 1\,, \\ 2/3 & \text{if} \quad u_1 = 1 \text{ and } u_2 = 0\,, \\ 1 & \text{if} \quad u_1 = u_2 = 1\,. \end{cases} \tag{1.6}$$

To summarise, the above discussion brings to the forefront the importance of maximizing the minimal distance/separation $d_{\min}$ of the received (noise-free) signals and at the same time indicates how a linear encoding allows us to achieve this. Nevertheless, the assumption that the messages u_1 and u_2 being sent by the transmitters S_1 and S_2 are binary in nature makes the discussion over simplistic—especially in terms of the use of number theory to analyse the outcomes. We now modify the 'baby' example to a more general situation in which S_1 and S_2 wish to send messages u_1 and u_2 from the set of integers $\{0, \ldots, Q\}$ to a single receiver R.

1.1.2 Example 1 (Modified 'baby' Example)

Unless stated otherwise, here and throughout, $Q \geq 2$ is a fixed integer. As we shall see, this slightly more complex setup, in which $u_1, u_2 \in \{0, \ldots, Q\}$, naturally bring into play the rich theory of Diophantine approximation. So with this in mind, let us assume that the codewords x_1 and x_2 that are being transmitted by S_1 and S_2 are simply obtained by the linear encoding of the messages u_1 and u_2 as follows

$$x_1 = \alpha u_1 \quad \text{and} \quad x_2 = \beta u_2 \qquad (0 \leq u_1, u_2 \leq Q)\,, \tag{1.7}$$

where α and β are some positive real numbers. We emphasise that the parameters α and β are at our disposal and this fact will be utilized later. As in the 'baby' example let h_1 and h_2 denote the channel coefficients associated with S_1 and S_2 respectively. Then, before taking noise into account, R will receive the signal

$$y = h_1x_1 + h_2x_2 = h_1\alpha u_1 + h_2\beta u_2\,. \tag{1.8}$$

Clearly, y takes the values

$$h_1\alpha u_1 + h_2\beta u_2 : 0 \le u_1, u_2 \le Q\,. \tag{1.9}$$

Thus, there are potentially $(Q+1)^2$ distinct outcomes of y and they lie in the interval $[0, (h_1\alpha + h_2\beta)Q]$. It is easily verified that if they were equally separated then their mutual separation would be precisely

$$\frac{h_1\alpha + h_2\beta}{Q+2}\,. \tag{1.10}$$

However, this is essentially never the case. Indeed, let $d_{\min}$ denote the minimal distance between the points y given by (1.9). Without loss of generality, suppose for the sake of simplicity that

$$0 < h_1\alpha < h_2\beta$$

and define the real number

$$\xi := \frac{h_1\alpha}{h_2\beta}\,, \tag{1.11}$$

which in view of the above assumption is between 0 and 1; *i.e.* $0 < \xi < 1$. Then, by Dirichlet's theorem, we have that

$$\left|\xi - \frac{p}{q}\right| \le \frac{1}{qQ} \tag{1.12}$$

for an integer pair $(p, q) \in \mathbb{Z}^2$ satisfying $1 \le q \le Q$. Since $0 < \xi < 1$ and $1 \le q \le Q$, we also have that $0 \le p \le q \le Q$. On multiplying (1.12) by $h_2\beta q$, we find that

$$|h_1\alpha q - h_2\beta p| \le \frac{C_1}{Q} \qquad (C_1 := h_2\beta\,), \tag{1.13}$$

for some integer pair $(p, q) \in \mathbb{Z}^2$ satisfying $1 \le q \le Q$ and $0 \le p \le q$. Now observe that the quantity $|h_1\alpha q - h_2\beta p|$ on the left hand side of (1.13) is exactly the distance between the two specific values of y within (1.9) corresponding to $u_1 = q, u_2 = 0$ and $u_1 = 0, u_2 = p$. Since $q \neq 0$, this demonstrates that the minimal distance $d_{\min}$ between the values of y given by (1.9) is always bounded above by C_1/Q; i.e.

$$d_{\min} \le \frac{C_1}{Q}\,. \tag{1.14}$$

For all intents and purposes, this bound on the minimal distance is smaller than the hypothetical 'perfect' separation given by (1.10). In general, we have that

$$d_{\min} \leq \min\left\{\frac{C_1}{Q}\,,\,\frac{h_1\alpha + h_2\beta}{Q+2}\right\}.$$

It is easily seen that we can remove the assumption that $0 < h_1\alpha < h_2\beta$ if we put $C_1 = \max\{h_1\alpha, h_2\beta\}$.

Remark 1.1 On looking at (1.14), the reader may be concerned (rightly) that the minimal distance $d_{\min}$ vanishes as Q grows. Luckily, this can be easily rectified by introducing a scaling factor $\lambda \geq 1$ into the linear encoding of the messages u_1 and u_2. The point of doing this is that the codeword x_1 (resp. x_2) given by (1.7) becomes $\lambda\alpha u_1$ (resp. $\lambda\beta u_2$) and this has no effect on the point of interest ξ given by (1.11) but it scales up by λ the constant C_1 appearing in (1.13). Thus, by choosing λ appropriately (namely, proportional to Q) we can avoid the right hand side of (1.14) from vanishing as Q grows. In subsequent more 'sophisticated' examples, the scaling factor will be relevant to the discussion and will appear at the point of linear encoding the messages.

Now let us bring noise into the above setup. As in the 'baby' example, if z denotes the (additive) noise, then instead of (1.8), R receives the signal

$$y' = y + z = h_1\alpha u_1 + h_2\beta u_2 + z\,. \tag{1.15}$$

Note that as long as the absolute value $|z|$ of the noise is strictly less than $d_{\min}/2$, the receiver R is able to recover y and consequently u_1 and u_2 from the value of y'. Commonly, the nature of noise is such that z is a random variable having normal distribution. Without loss of generality we will assume that $z \sim \mathcal{N}(0, 1)$, that is the mean value of noise is 0 and its variance is 1. Therefore, when taking the randomness of noise into account, the problem of whether or not the receiver is able to recover messages sent by the transmitters becomes probabilistic in nature. Loosely speaking, we are interested in the probability that $|z| < d_{\min}/2$—the larger the probability the more likely the receiver is able to recover messages by rounding y' to the closest possible outcome of y. Of course, if it happens that $|z| \geq d_{\min}/2$, then we will have an error in the recovery of y and thus the messages u_1 and u_2. When $z \sim \mathcal{N}(0, 1)$, the probability of this error can be computed using the Gauss error function and is explicitly equal to

$$1 - \sqrt{2/\pi}\int_0^{d_{\min}/2} e^{-\theta^2/2}d\theta\,.$$

This gets smaller as $d_{\min}$ gets larger. Clearly, in view of the theoretic upper bound on $d_{\min}$ given by (1.14) the probability of error is bounded above by the probability that $|z| < C_1/2Q$. Thus, the closer $d_{\min}$ is to the theoretic upper bound, the closer we are to minimizing the probability of the error and in turn the higher the threshold for tolerating noise. With this in mind, we now demonstrate that on appropriately

choosing the parameters α and β associated with the encoding procedure it is possible to get within a constant factor of the theoretic upper bound.

1.1.3 Badly Approximable Numbers

The key is to make use of the existence of badly approximable numbers—a fundamental class of real numbers in the theory of Diophantine approximation.

Definition 1.1 (Badly Approximable Numbers) A real number ξ is said to be *badly approximable* if there exists a constant $\kappa = \kappa(\xi) > 0$ such that for all $q \in \mathbb{N},\ p \in \mathbb{Z}$

$$\left|\xi - \frac{p}{q}\right| \geq \frac{\kappa}{q^2}\,. \tag{1.16}$$

Note that by definition, badly approximable numbers are precisely those real numbers for which the right hand side of inequality (1.2) associated with Dirichlet's corollary (Corollary 1.1) cannot be 'improved' by an arbitrary constant factor. By Hurwitz's theorem [16], if ξ is badly approximable then for the associated badly approximable constant $\kappa(\xi)$ we have that

$$0 < \kappa(\xi) < 1/\sqrt{5}\,.$$

It is well known that the set of badly approximable numbers can be characterized as those real numbers whose continued fraction expansions have bounded partial quotients. Moreover, an irrational number has a periodic continued fraction expansion if and only if it is a quadratic irrational and thus every quadratic irrational is badly approximable. In particular, it is easily verified that for any given $\varepsilon > 0$, the golden ratio

$$\gamma := (\sqrt{5}+1)/2$$

satisfies inequality (1.16) with $\kappa = 1/(\sqrt{5}+\varepsilon)$ for all $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ with $q^2 \geq 1/(\sqrt{5}\varepsilon)$. This is obtained using the standard argument that involves substituting p/q into the minimal polynomial f of γ over $\mathbb{Z}$ and using the obvious fact that $1 \leq q^2|f(p/q)| \leq q^2|\gamma - p/q|\cdot|\bar{\gamma} - p/q|$, where $\bar{\gamma} = (\sqrt{5}-1)/2$ is the conjugate of γ. We leave further computational details to the reader. Observe that on taking $\varepsilon = 1/\sqrt{5}$, we find that γ is badly approximable with $\kappa(\gamma) \geq \sqrt{5}/6$.

The reason for us bringing into play the notion of badly approximable numbers is very easy to explain. By definition, on choosing the parameters α and β so that $\xi := h_1\alpha/h_2\beta$ is badly approximable guarantees the existence of a constant $\kappa(\xi) > 0$ such that

$$|h_1\alpha q - h_2\beta p| \ \geq\ \kappa(\xi)\frac{C_1}{q} \qquad \forall\, q \in \mathbb{N},\ p \in \mathbb{Z}\,.$$

Thus, it follows that the separation between the points given by (1.9) is at least $\kappa(\xi)C_1/Q$. In other words, the minimal distance $d_{\min}$ is within a constant factor of the theoretic upper bound C_1/Q given by (1.14). Indeed, if we choose α and β so that $h_1\alpha/h_2\beta$ is the golden ratio γ we obtain that

$$\kappa(\gamma)\frac{C_1}{Q} \leq d_{\min} \leq \frac{C_1}{Q}\,. \tag{1.17}$$

The upshot is that equation (1.17) gives an explicit 'safe' threshold for the level of noise that can be tolerated. Namely, the probability that $|z| < d_{\min}/2$ is at least the probability that $|z| < \kappa(\gamma)C_1/Q$. In principle, one can manipulate the values of $Q \in \mathbb{N}$ and $\varepsilon > 0$ within the above argument to improve the lower bound in (1.17). However, any such manipulation will not enable us to surpass the hard lower bound limit of $C_1/(\sqrt{5}Q)$ imposed by the aforementioned consequence of Hurwitz's theorem. Therefore, we now explore a different approach in an attempt to make improvements to (1.17) beyond this hard limit. Ideally, we would like to replace $1/\sqrt{5}$ by a constant arbitrarily close to one. We would also like to move away from insisting that ξ is badly approximable since this is a rare event. Indeed, although the set of badly approximable number is of full Hausdorff dimension (a result of Jarník from the 1920s), it is a set of Lebesgue measure zero (a result of Borel from 1908). In other words, the (uniform) probability that a real number in the unit interval is badly approximable is zero. We will return to this in Sects. 1.2.2 and 1.2.7 below.

1.1.4 Probabilistic Aspects

The approach we now pursue is motivated by the following probabilistic problem: *Given* $0 < \kappa' < 1$ *and* $Q \in \mathbb{N}$, *what is the probability that a given real number* $\xi \in \mathbb{I} := (0, 1)$ *satisfies*

$$\left|\xi - \frac{p}{q}\right| \geq \frac{\kappa'}{qQ} \tag{1.18}$$

for all integers p *and* $1 \leq q \leq Q$*?* Note that these are the real numbers for which the right hand side of inequality (1.1) associated with Dirichlet's theorem cannot be improved by the factor of κ' (Q is fixed here). It is worth mentioning at this point, in order to avoid confusion later, that these real numbers are not the same as Dirichlet non-improvable numbers which will be introduced below in Sect. 1.1.5. To estimate the probability in question, we consider the complementary inequality

$$\left|\xi - \frac{p}{q}\right| < \frac{\kappa'}{qQ}\,. \tag{1.19}$$

Let $1 \leq q \leq Q$. Then for a fixed q, the probability that a given $\xi \in \mathbb{I} := (0,1)$ satisfies (1.19) for some $p \in \mathbb{Z}$ is exactly $2\kappa'/Q$—it corresponds to the measure of the set

$$E_q := \bigcup_{p \in \mathbb{Z}} \left(\tfrac{p}{q} - \tfrac{\kappa'}{qQ}, \tfrac{p}{q} + \tfrac{\kappa'}{qQ} \right) \cap \mathbb{I} .$$

On summing up these probabilities over q, we conclude that the probability that a given $\xi \in \mathbb{I}$ satisfies (1.19) for some integers p and $1 \leq q \leq Q$ is trivially bounded above by $2\kappa'$. This in turn implies that for any $\kappa' < 1/2$ and any $Q \in \mathbb{N}$ the probability that (1.18) holds for all integers p, q with $1 \leq q \leq Q$ is at least

$$1 - 2\kappa' .$$

The following result shows that with a little more extra work it is possible to improve this trivial bound.

Lemma 1.1 *For any $0 < \kappa' < 1$ and any $Q \in \mathbb{N}$ the probability that (1.18) holds for all integers p, q with $1 \leq q \leq Q$ is at least*

$$1 - \frac{12\kappa'}{\pi^2} \approx 1 - 1.216\kappa' . \tag{1.20}$$

Remark 1.2 Observe that when

$$\kappa' < \pi^2/12 \approx 0.822 ,$$

the quantity $12\kappa'/\pi^2$ is strictly less than 1 and therefore the probability given by (1.20) is greater than zero. Hence for any $Q \in \mathbb{N}$, there exist real numbers ξ satisfying (1.18) for all integers p and $1 \leq q \leq Q$.

Remark 1.3 Within Lemma 1.1 the word 'probability' refers to the uniform probability over [0, 1]. However, in real world applications the parameter ξ appearing in (1.18) may not necessarily be a uniformly distributed random variable. For instance, the channel coefficients could be subject to Rayleigh distribution and this will have an obvious effect on the distribution of ξ via (1.11). Nevertheless, as long as the distribution of ξ is absolutely continuous, a version of Lemma 1.1 can be established, albeit the constant that accompanies κ' will be different. For further details we refer the reader to [1].

Proof The proof of Lemma 1.1 relies on 'removing' the overlaps between the different sets E_q as q varies. Indeed, it is easily seen that

$$E := \bigcup_{q=1}^{Q} E_q = \bigcup_{q=1}^{Q} \bigcup_{\substack{0 \leq p \leq q \\ \gcd(p,q)=1}} \left(\tfrac{p}{q} - \tfrac{\kappa'}{qQ}, \tfrac{p}{q} + \tfrac{\kappa'}{qQ} \right) \cap \mathbb{I} .$$

Therefore,

$$\mathbf{Prob}(E) \le \sum_{q=1}^{Q} \sum_{\substack{1\le p\le q \\ \gcd(p,q)=1}} \frac{2\kappa'}{qQ} = \sum_{q=1}^{Q} \frac{2\kappa'\varphi(q)}{qQ} = \frac{2\kappa'}{Q} \sum_{q=1}^{Q} \frac{\varphi(q)}{q}, \tag{1.21}$$

where φ is the Euler function. To estimate the above sum, it is convenient to use the Möbius inversion formula, which gives that

$$\frac{\varphi(q)}{q} = \sum_{d|q} \frac{\mu(d)}{d}$$

where μ is the *Möbius function*. Recall that

$$\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

Then

$$\begin{aligned}
\sum_{q=1}^{Q} \frac{\varphi(q)}{q} &= \sum_{q=1}^{Q} \sum_{d|q} \frac{\mu(d)}{d} = \sum_{q=1}^{Q} \sum_{dd'=q} \frac{\mu(d)}{d} \\
&= \sum_{dd'\le Q} \frac{\mu(d)}{d} = \sum_{1\le d\le Q} \frac{\mu(d)}{d} \sum_{d'\le Q/d} 1 \\
&= \sum_{1\le d\le Q} \frac{\mu(d)}{d} [Q/d] \le Q \sum_{1\le d\le Q} \frac{\mu(d)}{d^2} \\
&\le \frac{6Q}{\pi^2}.
\end{aligned}$$

Combining this with (1.21) gives the required estimate, that is a lower bound on $1 - \mathbf{Prob}(E)$, the probability of the complement to E. □

Let $0 < \kappa' < \pi^2/12$ and $Q \in \mathbb{N}$ be given. The upshot of the above discussion is that there exist parameters α and β so that with probability greater than $1 - 12\kappa'/\pi^2 > 0$, the real number $\xi := h_1\alpha/h_2\beta$ satisfies (1.18) for all integers p and $1 \le q \le Q$. It follows that for such ξ (or equivalently parameters α and β) the separation between the associated points given by (1.9) is at least $\kappa' C_1/Q$ and so the minimal distance $d_{\min}$ satisfies

$$\kappa' \frac{C_1}{Q} \le d_{\min} \le \frac{C_1}{Q}. \tag{1.22}$$

In particular, we can choose κ' so that $\kappa(\gamma) < \kappa'$ in which case the lower bound in (1.22) is better than that in (1.17) obtained by making use of badly approximable numbers. That is to say, that the lower bound involving κ' is closer to the theoretic upper bound C_1/Q. Moreover, the set of badly approximable numbers is a set of measure zero whereas the set of real numbers satisfying (1.18) for all integers p and $1 \le q \le Q$ has Lebesgue measure at least $1 - 12\kappa'/\pi^2$. This is an important advantage of the probabilistic approach since in reality it is often the case that the channel coefficients h_1 and h_2 are random in nature. For example, when dealing with mobile networks one has to take into consideration the obvious fact that the transmitters are not fixed. The upshot is that in such a scenario, we do not have the luxury of specifying a particular choice of the parameters α and β that leads to the corresponding points given by (1.9) being well separated as in the sense of (1.17). The probabilistic approach provides a way out. In short, it enables us to ensure that the minimal distance $d_{\min}$ between the points given by (1.9) satisfies (1.22) with good (explicitly computable) probability. See [54, Section VI.B] for a concrete example where the above probabilistic approach is used for the analysis of the capacity of symmetric Gaussian multi-user interference channels.

Up to this point, Q has been a fixed integer greater than or equal to 2 and reflects the size of the set of messages. We end our discussion revolving around Example 1 by considering the scenario in which we have complete freedom in choosing Q. In particular, one is often interested in the effect of allowing Q to tend to infinity on the model under consideration. This is relevant to understanding the so-called Degrees of Freedom (DoF) of communication channels, see Sect. 1.2.4.

1.1.5 Dirichlet Improvable and Non-improvable Numbers

We now show that there are special values of Q for which the minimal distance $d_{\min}$ satisfies (1.22) with κ' as close to one as desired. The key is to exploit the (abundant) existence of numbers for which Dirichlet's theorem cannot be improved. Note that in the argument leading to (1.17) we made use of the existence of badly approximable numbers; that is numbers for with Dirichlet's corollary cannot be improved.

Definition 1.2 (Dirichlet Improvable and Non-improvable Numbers) Let $0 < \kappa' < 1$. A real number ξ is said to be κ'-*Dirichlet improvable* if for all sufficiently large $Q \in \mathbb{N}$ there are integers p and $1 \le q \le Q$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{\kappa'}{qQ}\,.$$

A real number ξ is said to be *Dirichlet non-improvable* if for any $\kappa' < 1$ it is not κ'-Dirichlet improvable. In other words, a real number ξ is *Dirichlet non-improvable* if for any $0 < \kappa' < 1$ there exists arbitrarily large $Q \in \mathbb{N}$ such that for all integers p

and $1 \leq q \leq Q$

$$\left|\xi - \frac{p}{q}\right| \geq \frac{\kappa'}{qQ}.$$

A well know result of Davenport & Schmidt [28] states that:

a real number is Dirichlet non − improvable

$$\Updownarrow$$

it is not badly approximable.

Consequently, a randomly picked real number is Dirichlet non-improvable with probability one. The upshot of this is the following remarkable consequence: for any random choice of channel coefficients h_1, h_2 and parameters α, β, **with probability one** for any $\varepsilon > 0$ there exist arbitrarily large integers Q such that the minimal distance $d_{\min}$ between the associated points given by (1.9) satisfies

$$(1-\varepsilon)\frac{C_1}{Q} \leq d_{\min} \leq \frac{C_1}{Q}.$$

Clearly, this is the best possible outcome for the basic wireless communication model considered in Example 1. We now consider a slightly more sophisticated model which demonstrates the role of higher dimensional Diophantine approximation in wireless communication.

1.2 A 'toddler' Example and Diophantine Approximation in Higher Dimensions

The discussion in this section is centred on analysing the model arising from adding another receiver within the setup of the modified 'baby' example.

1.2.1 Example 2

Suppose there are two users S_1 and S_2 as in Example 1 but this time there are also two receivers R_1 and R_2. Let $Q \geq 1$ be an integer and suppose S_1 wishes to simultaneously transmit independent messages $u_1, v_1 \in \{0, \ldots, Q\}$, where u_1 is intended for R_1 and v_1 for R_2. Similarly, suppose S_2 wishes to simultaneously transmit independent messages $u_2, v_2 \in \{0, \ldots, Q\}$, where u_2 is intended for

R_1 and v_2 for R_2. After (linear) encoding, S_1 transmits $x_1 := x_1(u_1, v_1)$ and S_2 transmits $x_2 := x_2(u_2, v_2)$; that is to say

$$x_1 = \alpha_1 u_1 + \beta_1 v_1 \quad \text{and} \quad x_2 = \alpha_2 u_2 + \beta_2 v_2 \tag{1.23}$$

where $\alpha_1, \alpha_2, \beta_1$ and β_2 are some positive real numbers. Next, for $i, j = 1, 2$, let h_{ij} denote the channel coefficients associated with the transmission of signals from S_j to R_i. Also, let y_i denote the signal received by R_i before noise is taken into account. Thus,

$$y_1 = h_{11}x_1 + h_{12}x_2\,, \tag{1.24}$$

$$y_2 = h_{21}x_1 + h_{22}x_2\,. \tag{1.25}$$

A pictorial description of the above setup is given in Fig. 1.3 below.
Substituting (1.23) into (1.24) and (1.25) gives that

$$y_1 = h_{11}\alpha_1 u_1 + h_{11}\beta_1 v_1 + h_{12}\alpha_2 u_2 + h_{12}\beta_2 v_2\,, \tag{1.26}$$

$$y_2 = h_{21}\alpha_1 u_1 + h_{21}\beta_1 v_1 + h_{22}\alpha_2 u_2 + h_{22}\beta_2 v_2\,. \tag{1.27}$$

Note that there are potentially $(Q+1)^4$ distinct outcomes of y_i and they lie in the interval $[0, (h_{i1}\alpha_1 + h_{i1}\beta_1 + h_{i2}\alpha_2 + h_{i2}\beta_2)Q]$.

Now let us bring noise into the setup. If z_i denotes the (additive) noise at receiver R_i $(i = 1, 2)$, then instead of (1.26) and (1.27), R_1 and R_2 receive the signals

$$y_1' = y_1 + z_1 \qquad \text{and} \qquad y_2' = y_2 + z_2 \tag{1.28}$$

respectively. Equations (1.23)–(1.28) represent one of the simplest models of what is known as a *two-user X-channel*. The ultimate goal is for the receiver R_1 to decode

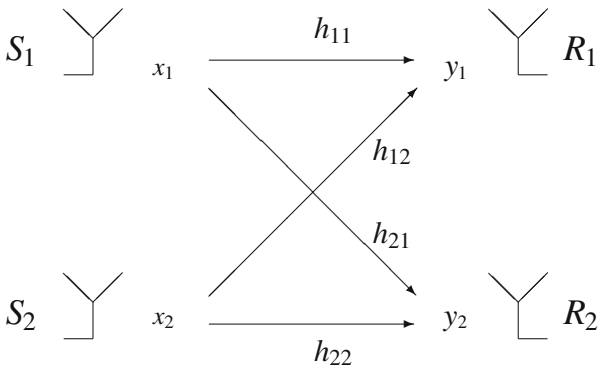


Fig. 1.3 Two-user X-channel

the messages u_1 and u_2 from the observation of y_1' and for the receiver R_2 to decode the messages v_1 and v_2 from the observation of y_2'. Clearly, this goal is attainable if $2|z_1|$ and $2|z_2|$ are smaller than the minimal distance between the outcomes of y_1 given by (1.26) and the minimal distance between the outcomes of y_2 given by (1.27) respectively.

Assume for the moment that $u_1, u_2, v_1, v_2 \in \{0, 1\}$ and for the ease of discussion, let us just concentrate on the signal y_1' received at R_1. Then there are generally up to 16 different outcomes for y_1. Now there is one aspect of the above setup that we have not yet exploited: the receiver R_1 is not interested in the signals v_1 and v_2. So if these 'unwanted' signals could be deliberately aligned (at the transmitters) via encoding into a single component $v_1 + v_2$, then there would be fewer possible outcomes for y_1. This is merely down to the simple fact that there are 4 different pairs (v_1, v_2) as opposed to 3 different sums $v_1 + v_2$ when v_1 and v_2 take on binary values. With this in mind, suppose that

$$x_1 = \lambda(h_{22}u_1 + h_{12}v_1) \quad \text{and} \quad x_2 = \lambda(h_{21}u_2 + h_{11}v_2) \tag{1.29}$$

respectively. Here $\lambda \geq 1$ is simply some scaling factor. Thus, with reference to (1.23), we have that

$$\alpha_1 = \lambda h_{22},\ \beta_1 = \lambda h_{12},\ \alpha_2 = \lambda h_{21},\ \beta_2 = \lambda h_{11}\,, \tag{1.30}$$

and so (1.24) and (1.25) become

$$y_1 = \lambda\Big((h_{11}h_{22})u_1 + (h_{21}h_{12})u_2 + (h_{11}h_{12})(v_1 + v_2)\Big) \tag{1.31}$$

$$y_2 = \lambda\Big((h_{21}h_{12})v_1 + (h_{11}h_{22})v_2 + (h_{21}h_{22})(u_1 + u_2)\Big)\,. \tag{1.32}$$

Clearly, there are now only 12 outcomes for either y_1 or y_2 rather than 16. The above discussion is a simplified version of that appearing in [52, §III: Example 3] and constitutes the basis for *real interference alignment*—a concept introduced and developed in [48, 51, 52] and subsequent publications.

Remark 1.4 The original idea of interference alignment exploits the availability of 'physical' dimensions of wireless systems such as the frequency of the signal or the presence of multiple antennae. In short, an antenna is a device (such as an old fashioned radio or television ariel) that is used to transmit or receive signals. In any case, by using several antennae it is possible for a user to simultaneously transmit several messages and these can naturally be thought of as the coordinates of a point in a vector space, say $\mathbb{R}^n$. Thus, when analysing such wireless systems the transmitted signals can be treated as vectors in $\mathbb{R}^n$. The art of interference alignment is to attempt to introduce an encoding at the transmitters (users) which result in unwanted (interfering) signals at the receivers being forced to lie in a subspace of $\mathbb{R}^n$ of smaller (ideally single) dimension. Such alignment is achieved

by exploiting elementary methods from linear algebra, see for instance [37, Section 2.1] for concrete examples and a detailed overview of the process. The novel idea of Motahari et al. involves exploiting instead the abundance of rationally independent points in the real line $\mathbb{R}$. For instance, with reference to Example 2 above and the transmitted signals given by (1.29), assuming that h_{22}/h_{12} is irrational, the signal x_1 transmitted by S_1 lies in the 2-dimensional vector subspace of $\mathbb{R}$ over $\mathbb{Q}$ given by

$$V_1 = \lambda h_{22}\mathbb{Q} + \lambda h_{12}\mathbb{Q}\,.$$

Similarly, assuming that h_{21}/h_{11} is irrational, the signal x_2 transmitted by S_2 lies in the 2-dimensional vector subspace of $\mathbb{R}$ over $\mathbb{Q}$ given by

$$V_2 = \lambda h_{21}\mathbb{Q} + \lambda h_{11}\mathbb{Q}\,.$$

In view of the alignment, the unwanted messages v_1 and v_2 at receiver R_1 are forced to lie in a subspace of $\mathbb{R}$ over $\mathbb{Q}$ of dimension one; namely $W_1 = \lambda h_{11}h_{12}\mathbb{Q}$. Similarly, the unwanted messages u_1 and u_2 at receiver R_2 lie in the one-dimensional $\mathbb{Q}$-subspace $W_2 = \lambda h_{21}h_{22}\mathbb{Q}$.

As with the 'baby' example, we can easily modify the above 'binary' consideration to the more general situation when the messages u_1, u_2, v_1, v_2 are integers lying in $\{0, \dots, Q\}$; i.e., the setup of Example 2. It is easily seen that in this more general situation the savings coming from interference alignment are even more stark: there are $(2Q+1)(Q+1)^2 \sim 2Q^3$ outcomes for either y_1 or y_2 after alignment as opposed to $(Q+1)^4 \sim Q^4$ outcomes before alignment. Consequently, based on the outcomes for y_1 and y_2 after alignment being equally spaced, we have the following trivial estimates for the associated minimal distances:

$$d_{\min,1} \le \frac{\lambda\Big(h_{11}h_{22} + h_{21}h_{12} + 2h_{11}h_{12}\Big)Q}{(2Q+1)(Q+1)^2} \tag{1.33}$$

and

$$d_{\min,2} \le \frac{\lambda\Big(h_{21}h_{12} + h_{11}h_{22} + 2h_{21}h_{22}\Big)Q}{(2Q+1)(Q+1)^2}\,. \tag{1.34}$$

We stress that $d_{\min,1}$ is the minimal distance between the outcomes of y_1 given by (1.31) and $d_{\min,2}$ is the minimal distance between the outcomes of y_2 given by (1.32). As in Example 1, 'perfect' separation is essentially never the case and to demonstrate this we need to bring into play the appropriate higher dimensional version of Dirichlet's theorem.

Theorem 1.2 (Minkowski's Theorem for Systems of Linear Forms) *Let* $\beta_{i,j} \in \mathbb{R}$, *where* $1 \le i, j \le k$, *and let* $\lambda_1, \ldots, \lambda_k > 0$. *If*

$$|\det(\beta_{i,j})_{1\le i,j\le k}| \le \prod_{i=1}^{k} \lambda_i, \tag{1.35}$$

then there exists a non-zero integer point $\mathbf{a} = (a_1, \ldots, a_k)$ *such that*

$$\begin{cases} |a_1\beta_{i,1} + \cdots + a_k\beta_{i,k}| < \lambda_i\,, \quad (1 \le i \le k-1) \\ |a_1\beta_{k,1} + \cdots + a_k\beta_{k,k}| \le \lambda_k\,. \end{cases} \tag{1.36}$$

The simplest proof of the theorem makes use of Minkowski's fundamental convex body theorem from the geometry of numbers; see, for instance [16, §1.4.1] or, indeed, Chap. 2 of this book.

We now show how the minimal distance $d_{\min,1}$ (and similarly, $d_{\min,2}$) can be estimated from above using Minkoswki's theorem. For simplicity, consider the case when

$$\max\{h_{11}h_{22},\ h_{21}h_{12}\,,\, h_{11}h_{12}\} = h_{11}h_{12}\,; \tag{1.37}$$

that is, $h_{11} \ge h_{21}$ and $h_{12} \ge h_{22}$. Then, on applying Theorem 1.2 with $k = 3$, $\lambda_1 = (h_{11}h_{12})Q^{-2}$, $\lambda_2 = \lambda_3 = Q$ and

$$(\beta_{i,j})_{1\le i,j\le k} = \begin{pmatrix} h_{11}h_{22} & h_{21}h_{12} & h_{11}h_{12} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

we deduce the existence of integers a_1, a_2 and a_3, not all zero, such that

$$\begin{cases} |(h_{11}h_{22})a_1 + (h_{21}h_{12})a_2 + (h_{11}h_{12})a_3| < (h_{11}h_{12})Q^{-2}\,, \\ |a_1| < Q, \\ |a_2| \le Q\,. \end{cases} \tag{1.38}$$

Remark 1.5 It is worth pointing out that the argument just given above can be appropriately adapted to establish the following generalisation of Dirichlet's theorem. For the details see for instance [16, Corollary 1.4.7]. Here and throughout, given a point $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we let $|\mathbf{x}| := \max\{|x_1|, \ldots, |x_n|\}$.

Theorem 1.3 *For any* $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$ *and any* $Q \in \mathbb{N}$ *there exists* $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ *such that*

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| < \frac{1}{Q^n} \qquad \textit{and} \qquad 1 \le |\mathbf{q}| \le Q. \tag{1.39}$$

We now return to determining an upper bound for $d_{\min,1}$. A consequence of (1.38) is that for any given $Q \geq 1$ there exist integers a_1, a_2, a_3, not all zero, such that

$$\left| \frac{h_{11}h_{22}}{h_{11}h_{12}} a_1 + \frac{h_{21}h_{12}}{h_{11}h_{12}} a_2 + a_3 \right| < Q^{-2} \leq 1 \, .$$

This together with the triangle inequality implies that

$$|a_3| < \left| \frac{h_{11}h_{22}}{h_{11}h_{12}} a_1 \right| + \left| \frac{h_{21}h_{12}}{h_{11}h_{12}} a_2 \right| + 1,$$

and so in view of our 'maximal' assumption (1.37), it follows that

$$|a_3| < |a_1| + |a_2| + 1 \leq Q + (Q-1) + 1 = 2Q \, .$$

Now observe that the quantity

$$\lambda \times \left| (h_{11}h_{22})a_1 + (h_{21}h_{12})a_2 + (h_{11}h_{12})a_3 \right|$$

is precisely the distance between the two specific outcomes of y_1 associated with (1.31) given by the following choices:

$$\begin{aligned} &\text{Choice 1: } u_1 = \max\{0, a_1\}, \quad u_2 = \max\{0, a_2\}, \quad v_1 + v_2 = \max\{0, a_3\}, \\ &\text{Choice 2: } u_1 = \max\{0, -a_1\}, \; u_2 = \max\{0, -a_2\}, \; v_1 + v_2 = \max\{0, -a_3\}. \end{aligned}$$

We have just observed that Theorem 1.2 guarantees that $|a_1| \leq Q$, $|a_2| \leq Q$ and $|a_3| \leq 2Q$ and so u_1, u_2, v_1, v_2 are integers lying in $\{0, \ldots, Q\}$. Hence, in view of (1.38) it follows (under the assumption (1.37)) that

$$d_{\min,1} \leq \frac{\lambda \, h_{11}h_{12}}{Q^2} = \frac{C_2}{Q^2} \, , \qquad \text{where } \; C_2 := \lambda \, h_{11}h_{12} \, . \tag{1.40}$$

For all intents and purposes, this bound on the minimal distance is smaller than the 'perfect' separation estimate given by (1.33). A similar analysis can be carried out when the maximum in (1.37) is attained on another term, and for estimating $d_{\min,2}$. Obviously the parameter C_2 would reflect the situation under consideration.

As mentioned earlier, the receivers R_1 and R_2 can decode the respective messages provided that the respective minimal distances $d_{\min,1}$ and $d_{\min,2}$ are at least two times larger than the noise at each receiver. Given that the nature of noise is often a random variable with normal distribution, the overarching goal is to ensure the probability that $|z_1| < \frac{1}{2} d_{\min,1}$ and $|z_2| < \frac{1}{2} d_{\min,2}$ is large. Indeed, as in Example 1, the larger the probability the more likely the receivers R_i $(i = 1, 2)$ are able to recover messages by rounding y_i' (given by (1.28)) to the closest possible outcome of y_i (given by (1.31) if $i = 1$ and (1.32) if $i = 2$). It is therefore imperative to understand how close $d_{\min,1}$ and $d_{\min,2}$ can be to their theoretical upper bounds.

With this in mind we now describe various tools and notions from Diophantine approximation that can be used for this purpose. In short, they allow us to get within a constant factor of the theoretical upper bounds. As in Example 1, we start by attempting to manipulate the encoding process so as to exploit the existence of badly approximable points in $\mathbb{R}^n$. Before we embark on this discussion we make a remark concerning the scaling factor λ that first appears in (1.29).

Remark 1.6 Observe that estimating $d_{\min,1}$ and $d_{\min,2}$ from below is essentially the same as estimating from below the size of the linear forms

$$(h_{11}h_{22})u_1 + (h_{21}h_{12})u_2 + (h_{11}h_{12})(v_1 + v_2)\,, \tag{1.41}$$

$$(h_{21}h_{12})v_1 + (h_{11}h_{22})v_2 + (h_{21}h_{22})(u_1 + u_2)\,. \tag{1.42}$$

The factor λ appearing in (1.31) and (1.32) only determines the scaling of $d_{\min,1}$ and $d_{\min,2}$ and can be used to 'adjust' these quantities, namely, to prevent them from vanishing as Q grows, see Remark 1.1 for a similar consideration within Example 1. Indeed, the effect of multiplication by λ can be simply understood as increasing the separation in the constellation of messages; i.e. the messages u_1, v_1, u_2, v_2 could be associated with $\{0, \lambda, 2\lambda, 3\lambda, \ldots, Q\lambda\}$ instead of $\{0, 1, 2, 3, \ldots, Q\}$.

1.2.2 Badly Approximable Points

We start by stating the following simple consequence of Theorem 1.3. It is the higher dimensional analogue of Corollary 1.1.

Corollary 1.2 *For any point $\boldsymbol{\xi} \in \mathbb{R}^n$ there exists infinitely many $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n \backslash \{\mathbf{0}\}$ such that*

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| < \frac{1}{|\mathbf{q}|^n}\,. \tag{1.43}$$

Note that in the corollary we have not imposed the condition that $\boldsymbol{\xi}$ is not a point on a rational hyperplane. This is since we do not impose, as in the one-dimensional statement, the requirement that $(p, \mathbf{q})$ is *primitive*; that is, without a non-trivial common divisor. Naturally, badly approximable points in $\mathbb{R}^n$ are defined by requiring that the right hand side of (1.43) cannot be 'improved' by an arbitrary constant factor. This we now formally state.

Definition 1.3 (Badly Approximable Points) A point $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be *badly approximable* if there exists a constant $\kappa = \kappa(\boldsymbol{\xi}) > 0$ such that for all $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n \backslash \{\mathbf{0}\}$

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \frac{\kappa}{|\mathbf{q}|^n}\,. \tag{1.44}$$

The set of badly approximable points in $\mathbb{R}^n$ will be denoted by $\mathbf{Bad}(n)$. It is relatively simple to verify that for any real algebraic number ξ of degree $n+1$ the point $(\xi, \xi^2, \ldots, \xi^n) \in \mathbb{R}^n$ is badly approximable. Indeed, consider the norm of the algebraic number

$$\alpha_1 = q_1\xi + q_2\xi^2 + \cdots + q_n\xi^n + p \in \mathbb{Q}(\xi)$$

which (up to sign) is the product of α_1 and its other conjugates, say $\alpha_2, \ldots, \alpha_{n+1}$. For simplicity one can assume that ξ is an algebraic integer. Furthermore, we can assume that the right hand side of (1.44) is less than one and so without loss of generality we have that $|p| \ll |\mathbf{q}|$. Then, it is easily seen that $|\alpha_j| \ll |\mathbf{q}|$ for all j, while the norm of α_1 is bounded below by 1. Here and elsewhere $\gg$ (respectively, $\ll$) is the Vinogradov symbol meaning $\geq$ (respectively $\leq$) up to a multiplicative constant factor. The upshot is that

$$|q_1\xi + q_2\xi^2 + \cdots + q_n\xi^n + p| = |\alpha_1| \gg \prod_{j=2}^{n+1} |\alpha_j|^{-1} \gg |\mathbf{q}|^{-n}\,,$$

whence the claim that $(\xi, \xi^2, \ldots, \xi^n) \in \mathbf{Bad}(n)$ follows. This argument can be made explicit to obtain a specific lower bound for the badly approximable constant $\kappa(\xi, \ldots, \xi^n)$. Examples of badly approximable algebraic points of this ilk were first given by Perron [55].

The reason for us bringing into play the notion of badly approximable numbers is similar to that in Example 1. If the channel coefficients happen to be such that

$$\boldsymbol{\xi} = (\xi_1, \xi_2) := \left(\frac{h_{11}h_{22}}{h_{11}h_{12}}, \frac{h_{21}h_{12}}{h_{11}h_{12}}\right) = \left(\frac{h_{22}}{h_{12}}, \frac{h_{21}}{h_{11}}\right) \tag{1.45}$$

is a badly approximable point in $\mathbb{R}^2$, then we are guaranteed the existence of a constant $\kappa(\boldsymbol{\xi}) > 0$ such that

$$\left|\frac{h_{11}h_{22}}{h_{11}h_{12}}q_1 + \frac{h_{21}h_{12}}{h_{11}h_{12}}q_2 + p\right| \geq \frac{\kappa(\boldsymbol{\xi})}{|\mathbf{q}|^2}$$

for all non-zero integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^2 \backslash \{\mathbf{0}\}$. Thus, it follows that for every $Q \in \mathbb{N}$:

$$|h_{11}h_{22}q_1 + h_{21}h_{12}q_2 + h_{11}h_{12}p| \geq \frac{\kappa(\boldsymbol{\xi})h_{11}h_{12}}{Q^2}$$

for all $(q_1, q_2, p) \in \mathbb{Z}^3$ with $1 \leq |\mathbf{q}| \leq Q$, and so the separations between any two points given by (1.31) is at least $\frac{\kappa(\boldsymbol{\xi})\lambda h_{11}h_{12}}{Q^2}$. In other worlds,

$$d_{\min,1} \geq \frac{\kappa(\boldsymbol{\xi})C_2}{Q^2} \tag{1.46}$$

which complements the upper bound (1.40). Note that instead of (1.45) one can equivalently consider $\boldsymbol{\xi} = (\xi_1, \xi_2)$ to be either of the points

$$\left(\frac{h_{21}h_{12}}{h_{11}h_{22}}, \frac{h_{11}h_{12}}{h_{11}h_{22}}\right), \quad \left(\frac{h_{11}h_{22}}{h_{21}h_{12}}, \frac{h_{11}h_{12}}{h_{21}h_{12}}\right), \tag{1.47}$$

which will also be badly approximable if (1.45) is badly approximable. Thus, we can in fact show that (1.46) with appropriately adjusted constant $\kappa(\boldsymbol{\xi})$ holds with C_2 redefined as

$$C_2 := \max\{h_{11}h_{22},\ h_{21}h_{12}\,, h_{11}h_{12}\}\,. \tag{1.48}$$

A similar lower bound to (1.46) can be established for $d_{\min,2}$ if

$$\left(\frac{h_{21}h_{12}}{h_{21}h_{22}}, \frac{h_{11}h_{22}}{h_{21}h_{22}}\right) = \left(\frac{h_{12}}{h_{22}}, \frac{h_{11}}{h_{21}}\right) \tag{1.49}$$

or equivalently

$$\left(\frac{h_{11}h_{22}}{h_{21}h_{12}}, \frac{h_{21}h_{22}}{h_{21}h_{12}}\right) \text{ or } \left(\frac{h_{21}h_{12}}{h_{11}h_{22}}, \frac{h_{21}h_{22}}{h_{11}h_{22}}\right) \tag{1.50}$$

is a badly approximable point in $\mathbb{R}^2$.

Remark 1.7 We end this subsection with a short discussion that brings to the forefront the significant difference between Examples 1 & 2, in attempting to exploit the existence of badly approximable points. In short, the encoding process (1.30) leading to the alignment of the unwanted signals in (1.31) and (1.32) comes at a cost. Up to a scaling factor, it fixes the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ in terms of the given channel coefficients. This in turn, means that our analysis of the linear forms (1.41) and (1.42) gives rise to the points (1.45) and (1.49) in $\mathbb{R}^2$ that are dependent purely on the channel coefficients. Now either these points are in **Bad**(2) or not. In other words, there is no flexibility left in the encoding procedure (after alignment) to force (1.45) or (1.49) to be badly approximable in $\mathbb{R}^n$. This is very different to the situation in Example 1. There we had total freedom to choose the parameters α and β in order to force the point (1.11) to be a badly approximable number. The upshot is that in Example 2, there is no such flexibility and this exacerbates the fact that the probability of (1.45) or (1.49) being badly approximable is already zero. The fact that **Bad**(n) has measure zero can be easily deduced from Khintchine's theorem, which will be discussed below in Sect. 1.2.4—however see Sect. 1.2.7 for the actual derivation. Although of measure zero, for the sake of completeness, it is worth mentioning that **Bad**(n) is of full Hausdorff dimension, the same as the whole of $\mathbb{R}^n$. This was established by Schmidt [57, 58] as an application of his remarkably powerful theory of (α, β)-games. In fact, he proved the full dimension statement for badly approximable sets associated with systems of linear forms (see Sect. 1.2.7).

Remark 1.8 We note that if $\boldsymbol{\xi}$ is any of the points (1.45) or (1.47) and $\boldsymbol{\xi}'$ is any of the points (1.49) or (1.50), then in order to simultaneously guarantee (1.46) and its analogue for $d_{\min,2}$ both $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ need to be badly approximable. This adds more constraints to an already unlikely (in probabilistic terms) event, since the points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ are dependent. Indeed, concerning the latter, it is easily seen that

$$\boldsymbol{\xi}' = \boldsymbol{f}(\boldsymbol{\xi}) \tag{1.51}$$

for one of the following choices of $\boldsymbol{f} : \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{f}(x,y) = \left(\frac{1}{x},\frac{1}{y}\right),\ \left(x,\frac{x}{y}\right),\ \left(\frac{x}{y},x\right),\ \left(y,\frac{y}{x}\right),\ \text{or}\ \left(\frac{y}{x},y\right). \tag{1.52}$$

Clearly, the set of pairs $(\boldsymbol{\xi},\boldsymbol{\xi}')$ of badly approximable points confined by (1.51) is a subset of the already measure zero set $\mathbf{Bad}(2)\times\mathbf{Bad}(2)$. Nevertheless, they do exist, as was proved by Davenport [26], and are in ample supply in the following sense: the set of pairs $(\boldsymbol{\xi},\boldsymbol{\xi}')$ of badly approximable points subject to (1.51) has full Hausdorff dimension, which is two. In other words, the dimension of $\mathbf{Bad}(2)\cap\boldsymbol{f}(\mathbf{Bad}(2))$ is equal to the dimension of $\mathbf{Bad}(2)$. This follows from the results of [19].

1.2.3 Probabilistic Aspects

In this section, we consider within the higher dimensional context of Example 2, the probabilistic approach set out in Sect. 1.1.4. Given $0<\kappa'<1$ and $Q\in\mathbb{N}$, let $\mathcal{B}_n(Q,\kappa')$ be the set of $\boldsymbol{\xi}\in\mathbb{I}^n:=(0,1)^n$ such that

$$|q_1\xi_1+\cdots+q_n\xi_n+p|\geq\frac{\kappa'}{Q^n} \tag{1.53}$$

for all integer points $(p,\mathbf{q})\in\mathbb{Z}\times\mathbb{Z}^n$ such that $1\leq|\mathbf{q}|\leq Q$. Note that $\boldsymbol{\xi}\in\mathcal{B}_n(Q,\kappa')$ are precisely the points in $\mathbb{I}^n$ for which the right hand side of inequality (1.39) appearing in Dirichlet's n-dimensional theorem, cannot be improved by the factor of κ' (Q is fixed here). To estimate the probability of $\mathcal{B}_n(Q,\kappa')$, we consider the complementary inequality

$$|q_1\xi_1+\cdots+q_n\xi_n+p|<\frac{\kappa'}{Q^n}\,. \tag{1.54}$$

Let $1\leq|\mathbf{q}|\leq Q$. Then for a fixed $\mathbf{q}$, it can be verified that the probability that a given $\boldsymbol{\xi}\in\mathbb{I}^n$ satisfies (1.54) for some $p\in\mathbb{Z}$ is exactly $2\kappa' Q^{-n}$—this is a relatively straightforward calculation the details of which can be found in [63, Lemma 8]. On summing up these probabilities over $\mathbf{q}$ with $q_1\geq 0$ (this can be assumed without

loss of generality), we conclude that the probability that a given $\boldsymbol{\xi} \in \mathbb{I}^n$ satisfies (1.54) for some integers p and $1 \leq |\mathbf{q}| \leq Q$, is bounded above by

$$2\kappa' Q^{-n}(2Q+1)^{n-1}(Q+1) \sim 2^n \kappa' \qquad (\text{as } Q \to \infty).$$

This in turn implies the following statement.

Lemma 1.2 *For any* $0 < \kappa' < 1$ *and any* $Q \in \mathbb{N}$

$$\mathbf{Prob}(\mathcal{B}_n(Q, \kappa')) \geq 1 - 2^n \kappa' \left(1 + \frac{1}{2Q}\right)^{n-1} \left(1 + \frac{1}{Q}\right). \tag{1.55}$$

Similarly to the one-dimensional case (cf. Sect. 1.1.4), the above trivial estimate can be improved, however, we leave this task to the energetic reader. We also note that the probability in Lemma 1.2 is assumed to be uniform but it is possible to obtain a version of Lemma 1.2 for other (absolutely continuous) distributions as mentioned in Remark 1.3. In any case, the upshot of the above discussion is that for sufficiently small $\kappa' > 0$ the probability that the point $\boldsymbol{\xi}$ given by (1.45) modulo 1 belongs to $\mathcal{B}_n(Q, \kappa')$ is positive. Hence, it follows that for any $\rho \in (0, 1)$ there exists an explicitly computable constant $\kappa' > 0$ with the following property: with probability greater than ρ, for a random choice of the four channel coefficients h_{ij} $(i, j = 1, 2)$, the separation between the associated points y_1 given by (1.31) is at least $\kappa' C_2 / Q^2$, and so the minimal distance $d_{\min,1}$ satisfies

$$d_{\min,1} \geq \frac{\kappa' C_2}{Q^2}. \tag{1.56}$$

Moreover, the probability ρ can be made arbitrarily close to one. However, the cost is that the constant κ' becomes arbitrarily small. The above analysis holds equally well at receiver R_2 and we obtain an analogous probabilistic bound for the minimal distance $d_{\min,2}$ associated with the points y_2 given by (1.32).

Remark 1.9 Obviously (1.56) is a better lower bound for $d_{\min,1}$ than (1.46) whenever κ' is greater than the badly approximable constant $\kappa(\boldsymbol{\xi})$ appearing in (1.46). However, this really is not the point—both approaches yield lower bounds for the minimal distance that lie within a constant factor of the theoretic upper bound (1.40). The main point is that the badly approximable approach has zero probability of actually delivering (1.46) whereas the probabilistic approach yields (1.46) with positive probability (whenever $\kappa(\boldsymbol{\xi})$ is sufficiently small so that the right hand side of (1.55) with $\kappa' = \kappa(\boldsymbol{\xi})$ is positive).

Remark 1.10 In the same vein as Remark 1.8, we first observe that in order to simultaneously guarantee (1.56) and its analogue for $d_{\min,2}$, both the points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ modulo one, where $\boldsymbol{\xi}$ is given by (1.45) or (1.47) and $\boldsymbol{\xi}'$ is given by (1.49) or (1.50), need to simultaneously lie in $\mathcal{B}_n(Q, \kappa')$. Thus to obtain the desired (simultaneous) probabilistic statement, we need to show the probability of both $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ modulo one

belonging to $\mathcal{B}_n(Q,\kappa')$ is positive; say $1-\kappa'$ in line with (1.55). This would be an easy task if the points under consideration were independent. However, the points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ are confined by (1.51) and therefore the events $\boldsymbol{\xi}(\mathrm{mod}1) \in \mathcal{B}_n(Q,\kappa')$ and $\boldsymbol{\xi}'(\mathrm{mod}1) \in \mathcal{B}_n(Q,\kappa')$ are dependent. Nevertheless, it can be shown that the probability of these two events holding simultaneously is at least $1-\sigma\times\kappa'$, where σ is an explicitly computable positive constant. We leave the details to the extremely energetic reader.

Remark 1.11 For another specific (and powerful) application of the probabilistic approach outlined in this section we refer the reader to [53]. In short, in [53] the probabilistic approach is used to estimate the capacity of the two-user X channel from below and above with only a constant gap between the bounds.

Notice that the fundamental set $\mathcal{B}_n(Q,\kappa')$ that underpins the probabilistic approach is dependent on Q. Thus, as Q varies, so does the random choice of channel coefficients that achieve (1.56). As we shall see in the next section, this can be problematic.

1.2.4 The Khintchine-Groshev Theorem and Degrees of Freedom

The probabilistic approach of Sect. 1.2.3, relies on the point $\boldsymbol{\xi}$ associated with the channel coefficients via (1.45) being in the set $\mathcal{B}_n(Q,\kappa')$. Now, however large the probability of the latter (a lower bound is given by (1.55)), it can be verified that

$$\mathbf{Prob}(\mathcal{B}_n(Q,\kappa')) \le 1-\omega\kappa' \,, \tag{1.57}$$

where $\omega > 0$ is a constant depending only on n. The proof of this can be obtained by utilizing the notion of ubiquity; in particular, exploiting the ideas used in establishing Proposition 4 in [12, Section 12.1]. Moreover, for any $\kappa' > 0$ and any infinite subset $\mathcal{Q} \subset \mathbb{N}$ the probability that $\boldsymbol{\xi}$ lies in $\mathcal{B}_n(Q,\kappa')$ for all sufficiently large $Q \in \mathcal{Q}$ (let alone all sufficiently large Q in $\mathbb{N}$) is zero. This is a fairly straightforward consequence of Theorem 1.3 and [9, Lemma 4]. This is an unfortunate downside of the probabilistic approach, especially when it comes to estimating the so called *Degrees of Freedom* (DoF) of communication channels. Indeed, when estimating the DoF it is desirable to achieve, with probability one, close to optimal bounds on the minimal distances ($d_{\min,1}$ and $d_{\min,2}$ within the context of Example 2) for all sufficiently large Q. Of course, the badly approximable approach described in Sect. 1.2.2 does this in the sense that it yields (1.56) for all large Q whenever $\boldsymbol{\xi} \in \mathbf{Bad}(2)$. However, as already discussed in Remark 1.9, the downside of the badly approximable approach is that the probability of hitting $\mathbf{Bad}(2)$ is zero. In this section we describe another approach which overcomes the inadequacies of both the probabilistic and badly approximable approaches. It gives an 'ε-weaker'

estimate for the minimal distance but as we shall soon see it is more than adequate for estimating the DoF. The key is to make use of the fundamental Khintchine-Groshev theorem in metric Diophantine approximation and this is what we first describe.

Given a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers, let

$$\mathcal{W}_n(\psi) := \left\{ \boldsymbol{\xi} \in \mathbb{I}^n : \begin{array}{l} |q_1\xi_1 + \cdots + q_n\xi_n + p| < \psi(|\mathbf{q}|) \\ \text{for i.m. } (p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n \backslash \{\mathbf{0}\} \end{array} \right\}. \tag{1.58}$$

Here and elsewhere, 'i.m.' is short for 'infinitely many' and given a subset X in $\mathbb{R}^n$, we will write $|X|_n$ for its n–dimensional Lebesgue measure. For obvious reasons, points in $\mathcal{W}_n(\psi)$ are referred to as ψ-approximable. When $n = 1$, it is easily seen that $\mathcal{W}(\psi) := \mathcal{W}_1(\psi)$ is the set of $\xi = \xi_1 \in \mathbb{I}$ such that

$$\left| \xi - \frac{p}{q} \right| < \frac{\psi(q)}{q}$$

has infinitely many solutions $(p, q) \in \mathbb{Z} \times \mathbb{N}$. Investigating the measure theoretic properties of $\mathcal{W}(\psi)$ was the subject of the pioneering work of Khintchine [40] almost a century ago. The following generalisation of Khintchine's theorem is a special case of a result of Groshev [36] concerning systems of linear form (see Theorem 1.12 in Sect. 1.2.7). In the one-dimensional case, it provides a quantitative analysis of the density of the rationals in the reals.

Theorem 1.4 (Khintchine-Groshev for One Linear Form) *Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonic function. Then*

$$|\mathcal{W}_n(\psi)|_n = \begin{cases} 0 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q) < \infty\,, \\ 1 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q) = \infty\,. \end{cases}$$

Remark 1.12 The convergence case of Theorem 1.4 is a relatively simple application of the Borel–Cantelli Lemma from probability theory and it holds for arbitrary functions ψ. In the divergence case, the theorem was first obtained by Groshev under the stronger assumption that $q^n\psi(q)$ is monotonic. In fact, the monotonicity assumption can be completely removed from the statement of theorem if $n \geq 2$. This is a consequence of Schmidt's paper [56, Theorem 2] from the swinging sixties if $n \geq 3$ and the relatively recent paper [10] covers the $n = 2$ case. In 1941, Duffin & Schaeffer [29] constructed a non-monotonic approximating function ψ for which the sum $\sum_q \psi(q)$ diverges but $|W(\psi)| = 0$. Thus, the monotonicity assumption cannot be removed in dimension one. For completeness, we mention that in the same paper Duffin & Schaeffer formulated an alternative statement for arbitrary functions. This soon became known as the notorious Duffin-Schaeffer Conjecture

and it remained unsolved for almost eighty years until the breakthrough work of Koukoulopoulos & Maynard [47].

An immediate consequence of the convergence case of Theorem 1.4 is the following statement.

Corollary 1.3 *Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a function such that*

$$\sum_{q=1}^{\infty} q^{n-1}\psi(q) < \infty\,. \tag{1.59}$$

Then, for almost all $\boldsymbol{\xi} \in \mathbb{I}^n$ there exists a constant $\kappa(\boldsymbol{\xi}) > 0$ such that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \;>\; \kappa(\boldsymbol{\xi})\,\psi(|\mathbf{q}|) \qquad \forall\ (p,\mathbf{q}) \in \mathbb{Z}\times\mathbb{Z}^n\backslash\{\mathbf{0}\}\,. \tag{1.60}$$

Now consider the special case when $\psi : q \to q^{-n-\varepsilon}$ for some $\varepsilon > 0$. Then Corollary 1.3 implies that for almost all $\boldsymbol{\xi} \in \mathbb{I}^n$ there exists a constant $\kappa(\boldsymbol{\xi}) > 0$ such that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \frac{\kappa(\boldsymbol{\xi})}{|\mathbf{q}|^{n+\varepsilon}}$$

for all $(p,\mathbf{q}) \in \mathbb{Z}\times\mathbb{Z}^n\backslash\{\mathbf{0}\}$. In particular, for almost all $\boldsymbol{\xi} \in \mathbb{I}^n$ and every $Q \in \mathbb{N}$ we have that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \frac{\kappa(\boldsymbol{\xi})}{Q^{n+\varepsilon}} \tag{1.61}$$

for all $(p,\mathbf{q}) \in \mathbb{Z}\times\mathbb{Z}^n\backslash\{\mathbf{0}\}$ with $1 \leq |\mathbf{q}| \leq Q$. Now in the same way if $\boldsymbol{\xi}$ given by (1.45) is badly approximable leads to the minimal distance estimate (1.46), the upshot of (1.61) is the following statement: with probability one, for every $Q \in \mathbb{N}$ and a random choice of channel coefficients h_{ij} $(i,j = 1,2)$, the separation between the associated points y_1 given by (1.31) is at least $\kappa(\boldsymbol{\xi})C_2/Q^{2+\varepsilon}$ and so

$$d_{\min,1} \geq \frac{\kappa(\boldsymbol{\xi})C_2}{Q^{2+\varepsilon}}\,. \tag{1.62}$$

Just to clarify, that $\boldsymbol{\xi}$ in the above corresponds to the point given by (1.45) associated with the choice of the channel coefficients. Note that instead of (1.45), one can equivalently consider $\boldsymbol{\xi}$ to be either of the points given by (1.47) and this would lead to (1.62) with C_2 defined by (1.48). A similar lower bound statement holds for the minimal distance $d_{\min,2}$ associated with the points y_2 given by (1.32). Of course, in this case $\boldsymbol{\xi}$ need to be replaced by $\boldsymbol{\xi}'$ given by (1.49) or equivalently (1.50).

Remark 1.13 Recall that $\boldsymbol{\xi}$ is given by (1.45) or (1.47) and $\boldsymbol{\xi}'$ is given by (1.49) or (1.50) and they are dependent via (1.51) and (1.52). Note that any of the maps

in (1.52) is a diffeomorphism on a sufficiently small neighborhood of almost every point in $\mathbb{R}^2$. Therefore, if $\boldsymbol{\xi}$ avoids a subset of $\mathbb{R}^2$ of measure zero, then so does $\boldsymbol{\xi}'$. Thus, (1.62) and an analogous bound for $d_{\min,2}$ are simultaneously valid for almost all choices of the channel coefficients.

Remark 1.14 Note that in the above analysis, if we had worked with the function $\psi : q \to q^{-n}(\log q)^{-1-\varepsilon}$ for some $\varepsilon > 0$, we would have obtained the stronger estimate

$$d_{\min,1} \geq \frac{\kappa(\boldsymbol{\xi})C_2}{Q^2(\log Q)^{1+\varepsilon}}\,.$$

It will be soon be clear that (1.62) is all we need for estimating the DoF within the context of Example 2.

A natural question arising from the above discussion is: *can the constant* $\kappa(\boldsymbol{\xi})$ *within Corollary 1.3 and thus* (1.62) *be made independent of* $\boldsymbol{\xi}$*?* Unfortunately, this is impossible to guarantee with probability one; that is, for almost all $\boldsymbol{\xi} \in \mathbb{I}^n$. To see this, consider the set

$$\mathcal{B}_n(\psi,\kappa) := \left\{ \boldsymbol{\xi} \in \mathbb{I}^n : \begin{array}{c} |q_1\xi_1 + \cdots + q_n\xi_n + p| > \kappa\psi(|\mathbf{q}|) \\ \forall\ (p,\mathbf{q}) \in \mathbb{Z}\times\mathbb{Z}^n\backslash\{\mathbf{0}\} \end{array} \right\}. \tag{1.63}$$

Then for any κ and ψ, observe that $\mathcal{B}_n(\psi,\kappa)$ will not contain the region

$$[-\kappa\psi(|\mathbf{q}|), \kappa\psi(|\mathbf{q}|)] \times \mathbb{R}^{n-1}$$

when $\mathbf{q} = (1,0,\ldots,0) \in \mathbb{Z}^n$. This region has positive probability; namely $2\kappa\psi(1)$), and so the complement (which contains $\mathcal{B}_n(\psi,\kappa)$) cannot have probability one. Nevertheless, the following result provides not only an explicit dependence on the probability of $\mathcal{B}_n(\psi,\kappa)$ on κ, but shows that it can be made arbitrarily close to one upon taking κ sufficiently small.

Theorem 1.5 (Effective Convergence Khintchine-Groshev for One Linear Form) *Let* $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ *be a function such that*

$$\sum_{q=1}^{\infty} q^{n-1}\psi(q) < \infty\,.$$

Then, for any $\kappa > 0$

$$\mathbf{Prob}(\mathcal{B}_n(\psi,\kappa)) \geq 1 - 4n\kappa\sum_{q=1}^{\infty}(2q+1)^{n-1}\psi(q)\,.$$

Proof Note that

$$\mathcal{B}_n(\psi,\kappa) = \mathbb{I}^n \setminus \bigcup_{\mathbf{q}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} E_{\mathbf{q}}(\psi)\,,$$

where

$$E_{\mathbf{q}} := \left\{\boldsymbol{\xi} \in \mathbb{I}^n : |q_1\xi_1 + \cdots + q_n\xi_n + p| \ \le \ \kappa\psi(|\mathbf{q}|) \text{ for some } p \in \mathbb{Z}\right\}.$$

Now, it is not difficult to verify that $|E_{\mathbf{q}}|_n = 2\kappa\psi(|\mathbf{q}|)$ - see [63, Lemma 8] for details. Thus, it follows that

$$\begin{aligned}
\mathbf{Prob}(\mathcal{B}_n(\psi,\kappa)) := |\mathcal{B}_n(\psi,\kappa)|_n \ &\ge \ 1 - \sum_{\mathbf{q}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} |E_{\mathbf{q}}|_n \\
&= \ 1 - \sum_{\mathbf{q}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} 2\kappa\psi(|\mathbf{q}|) \\
&= \ 1 - \sum_{q=1}^{\infty} \sum_{\substack{\mathbf{q}\in\mathbb{Z}^n \\ |\mathbf{q}|=q}} 2\kappa\psi(|\mathbf{q}|) \\
&= \ 1 - 2\kappa \sum_{q=1}^{\infty} \psi(q) \sum_{\substack{\mathbf{q}\in\mathbb{Z}^n \\ |\mathbf{q}|=q}} 1 \\
&\ge \ 1 - 2\kappa \sum_{q=1}^{\infty} \psi(q) 2n(2q+1)^{n-1}\,,
\end{aligned}$$

as desired. □

Having set up the necessary mathematical theory, we now turn our attention to calculating the DoF for the two-user X-channel considered in Example 2. The advantage of utilising the Khintchine-Groshev approach rather than the badly approximable approach, is that the value we obtain is not only sharp but it is valid for almost every realisation of the four channel coefficients h_{ij} $(i, j = 1, 2)$. Here, almost every is naturally with respect to 4-dimensional Lebesgue measure. At this point, a mathematician with little or no background in communication theory (like us) may rightly be crying out for an explanation of what is meant by the Degrees of Freedom of communication channels. We will attempt to provide a basic and in part a heuristic explanation within the context of Example 2. For a more in depth and general discussion we refer the reader to Chap. 2.

The simplest example of a communication channel is one involving just one transmitter and one receiver. For obvious reasons, such a setup is referred to as a *point to point channel*. The DoF of any other communication channel model is in

essence a measure of its efficiency compared with using multiple point to point channels. In making any comparison, it is paramount to compare like with like. Thus, given that the noise z_i $(i = 1, 2)$ at both receivers R_i within Example 2 is assumed to have normal distribution $\mathcal{N}(0, 1)$, we assume that the noise within the benchmark point to point channel has normal distribution $\mathcal{N}(0, 1)$. In the same vein, we assume that the messages the users transmit within both models are integers lying in $\{0, \ldots, Q\}$; that is to say that Q is the same in Example 2 and the point to point channel model. The parameter $Q \in \mathbb{N}$ is obviously a bound on the message size and it provides a bound on the number of binary digits (*bits*) that can be transmitted instantaneously as a single bundle. Indeed, sending the integer Q requires transmitting a bundle of $\lfloor \log Q \rfloor + 1 \approx \log Q$ bits, where the logarithm is to the base 2. Loosely speaking, the larger the message to be sent the larger the "power" required to transmit the message (transmitting instantaneously more bits requires more energy). Thus a bound on the message size Q corresponds to imposing a *power constraint* P on the channel model under consideration. For physical reasons, that are not particularly relevant to the discussion here, the power is comparable to the square of the message size. The upshot is that a power constraint P on the channel model places a bound on the maximal number of bits that can be reliably transmitted as a single bundle. With this in mind, the (total) DoF of the channel characterises the number (possibly fractional) of simple point-to-point channels, needed to reliably transmit the same maximal number of bits as the power constraint P tends to infinity. We now calculate the total DoF for the concrete setup of Example 2. The exposition given below is a simplified version of that presented in [52].

In relation to Example 2, the power constraint P means that

$$|x_1|^2 \leq P \text{ and } |x_2|^2 \leq P\,, \tag{1.64}$$

where x_1 and x_2 are the codewords transmitted by S_1 and S_2 as given by (1.29). Now notice that since the messages u_1, u_2, v_1, v_2 are integers lying in $\{0, \ldots, Q\}$, it follows that P is comparable to $(\lambda Q)^2$—the channel coefficients h_{ij} are fixed. Recall, that $\lambda \geq 1$ is a scaling factor which is at our disposal and this will be utilized shortly. It is shown in [52], that the probability of error in transmission within Example 2 is bounded above by

$$\exp\left(-\frac{d_{\min}^2}{8}\right), \tag{1.65}$$

where

$$d_{\min} = \min\{d_{\min,1}, d_{\min,2}\}.$$

It is a standard requirement that this probability should tend to zero as $P \to \infty$. In essence, this is what it means for the transmission to be reliable. Then, on assuming (1.62)—which holds for almost every realisation of the channel coefficients—it

follows that

$$d_{\min} \gg \frac{\lambda}{Q^{2+\varepsilon}}\,, \tag{1.66}$$

and so the quantity (1.65) will tend to zero as $Q \to \infty$ if we set

$$\lambda = Q^{2+2\varepsilon}\,.$$

The upshot of this is that we will achieve reliable transmission under the power constraint (1.64) if we set P to be comparable to $Q^{6+4\varepsilon}$; that is

$$Q^{6+4\varepsilon} \ll P \ll Q^{6+4\varepsilon}\,.$$

Now in Example 2, we simultaneously transmit 4 messages, namely u_1, u_2, v_1, v_2, which independently take values between 0 and Q. Therefore, in total we transmit approximately $4 \times \log Q$ bits, which with our choice of P is an achievable total rate of reliable transmission; however, it may not be maximal. We now turn our attention to the simple point to point channel in which the noise has normal distribution $\mathcal{N}(0,1)$. In his pioneering work during the forties, Shannon [62] showed that such a channel subject to the power constraint P achieves the maximal rate of reliable transmission $\frac{1}{2}\log(1+P)$—for further details see Sect. 2.1 of Chap. 2. On comparing the above rates of reliable transmission for the two models under the same power constraint, we get that the total DoF of the two-user X-channel described in Example 2 is at least

$$\lim_{P\to\infty} \frac{4\log Q}{\frac{1}{2}\log(1+P)} = \lim_{Q\to\infty} \frac{4\log Q}{\frac{1}{2}\log(1+Q^{6+4\varepsilon})} = \frac{4}{3+2\varepsilon}. \tag{1.67}$$

Given that $\varepsilon > 0$ is arbitrary, it follows that for almost every realisation of the channel coefficients

$$\mathrm{DoF} \geq \frac{4}{3}\,.$$

Now it was shown in [38] that the total DoF of a two-user X-channel is upper bounded by $4/3$ for all choices of the channel coefficients, and so it follows that for almost every realisation of the channel coefficients

$$\mathrm{DoF} = \frac{4}{3}\,. \tag{1.68}$$

For ease of reference we formally state these findings, the full details of which can be found in [52], as a theorem.

Theorem 1.6 *For almost every realisation of the four channel coefficients* h_{ij} $(i, j = 1, 2)$*, the total DoF of the two-user X-channel is* $\frac{4}{3}$.

Remark 1.15 We reiterate that by utilising the Khintchine-Groshev approach rather than the badly approximable approach (i.e. exploiting the lower bound (1.62) instead of (1.46) or equivalently (1.56) for the minimal distance), we obtain (1.68) for the DoF that is valid for almost every realisation of the four channel coefficients h_{ij} $(i, j = 1, 2)$ rather than on a set of 4-dimensional Lebesgue measure zero. In Sect. 1.2.6, we shall go further and show that any exceptional set of channel coefficients for which (1.68) fails is a subset arising from the notion of jointly singular points. This subset is then shown (see Theorem 1.9) not only to have measure zero but to have dimension strictly less than 4—the dimension of the space occupied by the channel coefficients. In short, our improvement of Theorem 1.6 is given by Theorem 1.10.

1.2.5 Dirichlet Improvable and Non-Improvable Points: Achieving Optimal Separation

We now show that there are special values of Q for which the minimal distance $d_{\min,1}$ satisfies (1.56) with κ' as close to one as desired. Recall, the larger the minimal distance the more tolerance we have for noise. The key is to exploit the (abundant) existence of points for which Dirichlet's theorem cannot be improved.

Definition 1.4 (Dirichlet Improvable and Non-Improvable Points) Let $0 < \kappa' < 1$. A point $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be κ'*-Dirichlet improvable* if for all sufficiently large $Q \in \mathbb{N}$ there are integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ with $1 \leq |\mathbf{q}| \leq Q$ such that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| < \kappa' Q^{-n}\,. \tag{1.69}$$

A point $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be *Dirichlet non-improvable* if for any $\kappa' < 1$ it is not κ'-Dirichlet improvable. Thus, explicitly, $\boldsymbol{\xi} \in \mathbb{R}^n$ is *Dirichlet non-improvable* if for any $0 < \kappa' < 1$ there exists arbitrarily large $Q \in \mathbb{N}$ such that for all integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ with $1 \leq |\mathbf{q}| \leq Q$

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \kappa' Q^{-n}\,. \tag{1.70}$$

Remark 1.16 Note that Dirichlet non-improvable points are not the same as those considered in the probabilistic approach of Sect. 1.2.3. There the emphasis is on both κ' and Q being uniform.

In a follow-up paper [27] to their one-dimensional work cited in Sect. 1.1.5, Davenport & Schmidt showed that Dirichlet improvable points in $\mathbb{R}^n$ form a set $\mathbf{DI}(n)$ of n-dimensional Lebesgue measure zero. Hence, a randomly picked point in $\mathbb{R}^n$ is Dirichlet non-improvable with probability one. The upshot of this is

the following consequence: for almost every random choice of the four channel coefficients h_{ij} $(i, j = 1, 2)$ and for any $\varepsilon > 0$ there exist arbitrarily large integers Q such that the minimal distance $d_{\min,1}$ between the associated points given by (1.31) satisfies

$$d_{\min,1} \geq \frac{(1-\varepsilon)\lambda h_{11}h_{12}}{Q^2} = (1-\varepsilon)\,\frac{C_2}{Q^2}\,. \tag{1.71}$$

To conclude, the Dirichlet non-improvable approach allows us to almost surely achieve the best possible separation, within the factor $(1-\varepsilon)$ of the theoretic upper bound (1.40), for an infinite choice of integer parameters $Q \in \mathcal{Q}_1$.

Remark 1.17 Obviously, we can obtain an analogous lower bound statement for $d_{\min,2}$ for an infinite choice of integer parameters $Q \in \mathcal{Q}_2$. However, it is not guaranteed that the integer sets $\mathcal{Q}_1$ and $\mathcal{Q}_2$ overlap and thus the problem of optimising $d_{\min,1}$ and $d_{\min,2}$ simultaneously remains open.

1.2.6 Singular and Non-Singular Points: The DoF of X-Channel Revisited

With reference to Example 2, the Khintchine-Groshev and the Dirichlet non-improvable approaches allows us to achieve good separation for the minimal distances (i.e., lower bounds for $d_{\min,1}$ and $d_{\min,2}$ that are at most 'ε-weaker' than the theoretic upper bounds) for almost all choices of the four channel coefficients h_{ij} $(i, j = 1, 2)$. We now turn to the question of *whether good separation can be achieved for a larger class of channel coefficients? For example, is it possible that the set of exceptions not only has measure zero (as is the case with the aforementioned approaches) but has dimension strictly less than four (the dimension of the space occupied by the channel coefficients)?* In short the answer is yes. The key is to make use of the following weaker notion than that of Dirichlet non-improvable points (cf. Definition 1.4).

Definition 1.5 (Singular and Non-Singular Points) A point $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be *singular* if it is κ'-Dirichlet improvable for any $\kappa' > 0$. A point $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be *non-singular* (or *regular*) if it is not singular. Thus, explicitly, $\boldsymbol{\xi} \in \mathbb{R}^n$ is *non-singular* if there exists a constant $\kappa' = \kappa'(\boldsymbol{\xi}) > 0$ such that there exist arbitrarily large integers $Q \in \mathbb{N}$ so that for all integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ with $1 \leq |\mathbf{q}| \leq Q$

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \kappa' Q^{-n}\,. \tag{1.72}$$

By definition, any singular point is trivially Dirichlet improvable. Equivalently, any Dirichlet non-improvable point is trivially non-singular.

We let $\mathbf{Sing}(n)$ denote the set of singular points in $\mathbb{R}^n$. It is easily verified that $\mathbf{Sing}(n)$ contains every rational hyperplane in $\mathbb{R}^n$. Therefore,

$$n - 1 \leq \dim \mathbf{Sing}(n) \leq n \,.$$

Here and throughout, $\dim X$ will denote the Hausdorff dimension of a subset X of $\mathbb{R}^n$. For the sake of completeness, we provide the definition.

Definition 1.6 (Hausdorff Dimension) Let $X \subset \mathbb{R}^n$. Then the Hausdorff dimension $\dim X$ of X is defined to be the infimum of $s > 0$ such that for any $\rho > 0$ and any $\varepsilon > 0$ there exists a cover of X by a countable family B_i of balls of radius $r(B_i) < \rho$ such that

$$\sum_{i=1}^{\infty} r(B_i)^s < \varepsilon \,.$$

Remark 1.18 For most sets upper bounds for the Hausdorrf dimension can be obtained using natural covering by small balls. Indeed, let $X \subset \mathbb{R}^n$ and $\rho > 0$ and suppose X can be covered by $N_\rho(X)$ balls of radius at most ρ. Then, it immediately follows for the above definition that

$$\dim X \leq \limsup_{\rho \to 0} \frac{\log N_\rho(X)}{-\log \rho} \,.$$

Note that the Hausdorff dimension of planes and more generally smooth submanifolds of $\mathbb{R}^n$ is the same as their usual 'geometric' dimension. The middle third Cantor set $\mathcal{K}$ is the standard classical example of a set with fractal dimension. Recall, $\mathcal{K}$ consists of all real numbers in the unit interval whose base 3 expansion does not contain the 'digit' 1; that is

$$\mathcal{K} := \{\xi \in [0, 1] : \xi = \textstyle\sum_{i=1}^{\infty} a_i 3^{-i} \quad \text{with } a_i = 0 \text{ or } 2\} \,.$$

It is well known that

$$\dim \mathcal{K} = \frac{\log 2}{\log 3} \,.$$

For a proof of this and a lovely introduction to the mathematical world of fractals, see the bible [30].

Now returning to singular points, in the case $n = 1$, a nifty argument due to Khintchine [40] dating back to the twenties shows that a real number is singular if and only if it is rational; that is

$$\mathbf{Sing}(1) = \mathbb{Q} \,. \tag{1.73}$$

Recently, Cheung & Chevallier [22], building on the spectacular $n = 2$ work of Cheung [21], have proved the following dimension statement for $\mathbf{Sing}(n)$.

Theorem 1.7 (Cheung & Chevallier) *Let $n \geq 2$. Then*

$$\dim \mathbf{Sing}(n) = \frac{n^2}{n+1}\,.$$

Thus,

$$\operatorname{codim} \mathbf{Sing}(n) = \frac{n}{n+1}\,.$$

Remark 1.19 Note that since $\frac{n^2}{n+1} > n-1$, the theorem immediately implies that in higher dimensions $\mathbf{Sing}(n)$ does not simply correspond to rationally dependent $\boldsymbol{\xi} \in \mathbb{R}^n$ as in the one-dimensional case—the theory is much richer. Also observe, that since $\frac{n^2}{n+1} < n$, the set $\mathbf{Sing}(n)$ is strictly smaller than $\mathbb{R}^n$ in terms of its Hausdorff dimension. How much smaller is measured by its codimension; *i.e.* $n - \dim \mathbf{Sing}(n)$.

Now if the four channel coefficients h_{ij} $(i, j = 1, 2)$ happen to be such that the corresponding point $\boldsymbol{\xi} \in \mathbb{R}^2$ given by (1.45) is non-singular, then there exist arbitrarily large integers Q such that the minimal distance $d_{\min,1}$ between the associated points given by (1.31) satisfies

$$d_{\min,1} \geq \frac{\kappa'(\boldsymbol{\xi})\lambda h_{11}h_{12}}{Q^2} = \frac{\kappa'(\boldsymbol{\xi})\, C_2}{Q^2}\,. \tag{1.74}$$

This of course is similar to the statement in which the point $\boldsymbol{\xi}$ is Dirichlet non-improvable with the downside that we cannot replace the constant $\kappa'(\boldsymbol{\xi})$ by $(1-\varepsilon)$ as in (1.71). However, the advantage is that it is valid for a much larger set of channel coefficients; namely, the exceptional set of channel coefficients $(h_{11}, h_{12}, h_{21}, h_{22}) \in \mathbb{R}^4_+$ for which (1.74) is not valid has dimension $\frac{10}{3}$, which is strictly smaller than 4—the dimension of the ambient space occupied by $(h_{11}, h_{12}, h_{21}, h_{22})$. This result seems to be new and we state it formally.

Proposition 1.1 *For all choices of channel coefficients $(h_{11}, h_{12}, h_{21}, h_{22}) \in \mathbb{R}^4_+$, except on a subset of codimension $\frac{2}{3}$, there exist arbitrarily large integers Q such that the minimal distance $d_{\min,1}$ between the associated points given by (1.31) satisfies (1.74).*

The proof of the proposition will make use of the following two well known results from fractal geometry [50].

Lemma 1.3 (Marstrand's Slicing Lemma) *For any $X \subset \mathbb{R}^k$ and $l \in \mathbb{N}$, we have that*

$$\dim(X \times \mathbb{R}^\ell) = \dim X + \ell\,.$$

Lemma 1.4 *Let $X \subset \mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}^k$ be a locally bi-Lipschitz map. Then*

$$\dim\big(g(X)\big) = \dim X\,.$$

Proof *(Proof of Proposition 1.1)* Consider the following map on the channel coefficients

$$g : \mathbb{R}^4_+ \to \mathbb{R}^4_+ \quad \text{such that} \quad g(h_{11}, h_{12}, h_{21}, h_{22}) = \left(h_{11}, h_{12}, \frac{h_{22}}{h_{12}}, \frac{h_{21}}{h_{11}}\right).$$

As we have already discussed, for any $\boldsymbol{\xi}$ given by (1.45) such that $\boldsymbol{\xi} \in \mathbb{R}^2_+ \setminus \mathbf{Sing}(2)$ we have that (1.74) holds. Hence, (1.74) holds for any choice of channel coefficients such that

$$(h_{11}, h_{12}, h_{21}, h_{22}) \notin g^{-1}\Big(\mathbb{R}^2_+ \times \big(\mathbb{R}^2_+ \cap \mathbf{Sing}(2)\big)\Big)\,. \tag{1.75}$$

By Lemma 1.3 and Theorem 1.7, it follows that

$$\operatorname{codim}\Big(\mathbb{R}^2_+ \times \big(\mathbb{R}^2_+ \cap \mathbf{Sing}(2)\big)\Big) = \frac{2}{3}\,.$$

Finally, note that locally at every point of $\mathbb{R}^4_+$ the map g is a C^1 diffeomorphism and hence is bi-Lipschitz. Therefore, by Lemma 1.4 it follows that g^{-1} preserves dimension and thus the codimension of the right hand side of (1.75) is $\frac{2}{3}$. This completes the proof. □

Remark 1.20 Just to clarify, that $\boldsymbol{\xi}$ appearing in (1.74) corresponds to the point given by (1.45) associated with the choice of the channel coefficients h_{ij} $(i, j = 1, 2)$ and $\kappa'(\boldsymbol{\xi}) > 0$ is a constant dependent on $\boldsymbol{\xi}$. Note that instead of (1.45), one can equivalently consider $\boldsymbol{\xi}$ to be either of the points given by (1.47) and this would lead to (1.74) with C_2 defined by (1.48).

Naturally, the analogue of Proposition 1.1 holds for the minimal distance $d_{\min,2}$ between the associated points given by (1.34). However, as in the Dirichlet non-improvable setup (cf. Remark 1.17), we cannot guarantee that the arbitrary large integers Q on which the lower bounds for the minimal distances are attained, overlap. If we could guarantee infinitely many overlaps, it would enable us to strengthen Theorem 1.6 concerning the Degrees of Freedoms (DoF) of the two-user X-channel described in Example 2. With this goal in mind, it is appropriate to introduce the following notion of jointly singular points.

Definition 1.7 (Jointly Singular and Non-Singular Points) The pair of points $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in \mathbb{R}^n \times \mathbb{R}^n$ is said to be *jointly singular* if for any $\varepsilon > 0$ for all sufficiently large $Q \in \mathbb{N}$ there exists an integer point $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ with $1 \le |\mathbf{q}| \le Q$ satisfying

$$\min_{1 \le j \le 2} |q_1\xi_{j,1} + \cdots + q_n\xi_{j,n} + p| < \varepsilon Q^{-n}\,,$$

where $\boldsymbol{\xi}_j = (\xi_{j,1}, \dots, \xi_{j,n})$, $j = 1, 2$. The pair $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in \mathbb{R}^n \times \mathbb{R}^n$ will be called *jointly non-singular* if it is not jointly singular, that is if there exists a constant $\kappa' = \kappa'(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) > 0$ such that there exist arbitrarily large $Q \in \mathbb{N}$ so that for all integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n$ with $1 \le |\mathbf{q}| \le Q$

$$\min_{1\le j\le 2} \left|q_1\xi_{j,1} + \dots + q_n\xi_{j,n} + p\right| \ge \kappa' Q^{-n} \,. \tag{1.76}$$

The set of jointly singular pairs in $\mathbb{R}^n \times \mathbb{R}^n$ will be denoted by $\mathbf{Sing}^2(n)$. This set is not and should not be confused with the standard simultaneous singular set corresponding to two linear forms in n variables (see Sect. 1.2.7).

The above notion of jointly non-singular pairs enables us to prove the following DoF statement.

Proposition 1.2 *Let $(h_{11}, h_{12}, h_{21}, h_{22}) \in \mathbb{R}^4_+$ be given and let $\boldsymbol{\xi}$ be any of the points (1.45) or (1.47), let $\boldsymbol{\xi}'$ be any of the points (1.49) or (1.50). Suppose that*

$$(\boldsymbol{\xi}, \boldsymbol{\xi}') \notin \mathbf{Sing}^2(2) \,. \tag{1.77}$$

Then (1.68) holds, that is the total DoF of the two-user X-channel with h_{ij} $(i, j = 1, 2)$ as its channel coefficients is $\frac{4}{3}$.

Proof To start with, simply observe that condition (1.77) means that there exist $\kappa' > 0$ and an infinite subset $\mathcal{Q} \subset \mathbb{N}$ such that for every $Q \in \mathcal{Q}$ and all integer points $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^2$ with $1 \le |\mathbf{q}| \le Q$

$$|q_1\xi_1 + q_2\xi_2 + p| \ge \kappa' Q^{-2} \quad \text{and} \quad \left|q_1\xi'_1 + q_2\xi'_2 + p\right| \ge \kappa' Q^{-2} \,. \tag{1.78}$$

Consequently, for every $Q \in \mathcal{Q}$ we can guarantee that (1.74) and its analogue for $d_{\min,2}$ are simultaneously valid. This in turn implies (1.66) for every $Q \in \mathcal{Q}$. From this point onwards, the rest of the argument given in Sect. 1.2.4 leading to (1.68) remains unchanged apart from the fact that the limit in (1.67) is now along $Q \in \mathcal{Q}$ rather than the natural numbers. □

Proposition 1.2 provides a natural pathway for strengthening Theorem 1.6. This we now describe. It is reasonable to expect that the set of $(\boldsymbol{\xi}, \boldsymbol{\xi}')$ not satisfying (1.77) is of dimension strictly smaller than four—the dimension of the ambient space. Indeed, this is something that we are able to prove.

Theorem 1.8 *Let $n \ge 2$. Then*

$$\dim \mathbf{Sing}^2(n) = 2n - \frac{n}{(n+1)} \,. \tag{1.79}$$

The theorem will easily follow from a more general statement concerning systems of linear forms proved in Sect. 1.2.7 below; namely, Theorem 1.14. Note that Theorem 1.8 is not enough for improving Theorem 1.6. Within Proposition 1.2,

the point $\boldsymbol{\xi}$ is given by (1.45) or (1.47) and $\boldsymbol{\xi}'$ is given by (1.49) or (1.50), and are therefore dependent via (1.51) and (1.52). The above theorem does not take into consideration this dependency. This is rectified by the following result.

Theorem 1.9 *Let $\boldsymbol{f} : U \to \mathbb{R}^n$ be a locally bi-Lipschitz map defined on an open subset $U \subset \mathbb{R}^n$ and let*

$$\mathbf{Sing}^2_{\boldsymbol{f}}(n) := \{\boldsymbol{\xi} \in U : (\boldsymbol{\xi}, \boldsymbol{f}(\boldsymbol{\xi})) \in \mathbf{Sing}^2(n)\}\,.$$

Then

$$\dim \mathbf{Sing}^2_{\boldsymbol{f}}(n) \leq n - \frac{n}{2(n+1)} < n. \tag{1.80}$$

As with Theorem 1.8, we defer the proof of the above theorem till Sect. 1.2.7. Combining the $n = 2$ case of Theorem 1.9 with Proposition 1.2 gives the following strengthening of the result of Motahari et al. on the DoF of a two-user X-channel (Theorem 1.6).

Theorem 1.10 *The total DoF of the two-user X-channel given by (1.68) can be achieved for all realisations of the channel coefficients h_{ij} $(i, j = 1, 2)$ except on a subset of Hausdorff dimension $\leq 4 - \frac{1}{3}$; that is, of codimension $\geq \frac{1}{3}$.*

Clearly, $\mathbf{Sing}(n)$ is a subset $\mathbf{Sing}^2_{\boldsymbol{f}}(n)$. Therefore, it follows that

$$\dim \mathbf{Sing}^2_{\boldsymbol{f}}(n) \geq \dim \mathbf{Sing}(n)$$

which together with Theorem 1.7 implies that for $n \geq 2$

$$\dim \mathbf{Sing}^2_{\boldsymbol{f}}(n) \geq \frac{n^2}{n+1} = n - \frac{n}{n+1}\,.$$

The gap between this lower bound and the upper bound of Theorem 1.9 leaves open the natural problem of determining $\dim \mathbf{Sing}^2_{\boldsymbol{f}}(n)$ precisely. We suspect that the lower bound is sharp.

Problem 1.1 Let $n \geq 2$ and $\boldsymbol{f} : U \to \mathbb{R}^n$ be a locally bi-Lipschitz map defined on an open subset $U \subset \mathbb{R}^n$. Verify if

$$\dim \mathbf{Sing}^2_{\boldsymbol{f}}(n) = \frac{n^2}{n+1}\,.$$

Note that to improve Theorem 1.10 we are only interested in the case $n = 2$ of Problem 1.1 with $\boldsymbol{f}$ given by (1.52).

1.2.7 Systems of Linear Forms

To date, we have in one form or another exploited the theory of Diophantine approximation of a single linear form in n real variables. In fact, Example 1 only really requires the notions and results with $n = 1$ while Example 2 requires them with $n = 2$. It is easily seen, that in either of these examples, if we increase the number of users (transmitters) S then we increase the numbers of variables appearing in the linear form(s) associated with the received message(s) y. Indeed, within the setup of Example 2 (resp. Example 1) we would need to use the general n (resp. $n-1$) variable theory if we had n transmitters.

The majority of the Diophantine approximation theory for a single linear form is a special case of a general theory addressing systems of m linear forms in n real variables. For the sake of completeness, it is appropriate to provide a brief taster of the general Diophantine approximation theory with an emphasis on those aspects used in analysing communication channel models. It should not come as a surprise that the natural starting point is Dircihlet's theorem for systems of linear forms. Throughout, let $n, m \geq 1$ be integers and $\mathbb{M}_{n,m}$ denote the set of $n \times m$ matrices $\boldsymbol{\Xi} = (\xi_{i,j})$ with entries from $\mathbb{R}$. Clearly, such a matrix represents the coordinates of a point in $\mathbb{R}^{nm}$. Also, given $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n$ let

$$|\mathbf{q}\boldsymbol{\Xi} + \mathbf{p}| := \max_{1 \leq j \leq m} |\mathbf{q}.\boldsymbol{\xi}_j + p_j| \,,$$

where $\boldsymbol{\xi}_j := (\xi_{1,j}, \ldots, \xi_{n,j})^t \in \mathbb{R}^n$ is the j'th column vector of $\boldsymbol{\Xi}$ and $\mathbf{q}.\boldsymbol{\xi}_j := q_1\xi_{1,j} + \ldots + q_n\xi_{n,j}$ is the standard dot product.

Theorem 1.11 (Dirichlet's Theorem for Systems of Linear Forms) *For any $\boldsymbol{\Xi} \in \mathbb{M}_{n,m}$ and any $Q \in \mathbb{N}$ there exists $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n$ such that*

$$|\mathbf{q}\boldsymbol{\Xi} + \mathbf{p}| < Q^{-\frac{n}{m}} \qquad \text{and} \qquad 1 \leq |\mathbf{q}| \leq Q \,.$$

The theorem is a relatively straightforward consequence of Minkowski's theorem for systems of linear forms; namely Theorem 1.2 in Sect. 1.2.1. For the details of the deduction see for example [60, Chapter 2]. In turn, a straightforward consequence of the above theorem is the following natural extension of Corollary 1.1 to systems of linear form.

Corollary 1.4 *For any $\boldsymbol{\Xi} \in \mathbb{M}_{n,m}$ there exists infinitely many $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n \setminus \{\mathbf{0}\}$ such that*

$$|\mathbf{q}\boldsymbol{\Xi} + \mathbf{p}| < |\mathbf{q}|^{-\frac{n}{m}} \,.$$

Armed with Theorem 1.11 and its corollary, it does not require much imagination to extend the single linear form notions of badly approximable (cf. Definition 1.3)

and Dirchlet improvable (cf. Definition 1.4) to systems of linear forms. Indeed, concerning the former we arrive at the set

$$\mathbf{Bad}(n,m) := \left\{ \boldsymbol{\Xi} \in \mathbb{M}_{n,m} \ : \ \liminf_{\substack{\mathbf{q}\in\mathbb{Z}^n: \\ |\mathbf{q}|\to\infty}} |\mathbf{q}|^{\frac{n}{m}} |\mathbf{q}\boldsymbol{\Xi} - \mathbf{p}| > 0 \right\} .$$

This clearly coincides with $\mathbf{Bad}(n)$ when $m = 1$. As we shall soon see, $\mathbf{Bad}(n,m)$ it is a set of zero nm-dimensional Lebesgue measure. Even still, Schmidt [57, 58] showed that it is a large set in the sense that it is of maximal dimension; i.e. $\dim \mathbf{Bad}(n,m) = nm$. Moving swiftly on, given a function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ let

$$\mathcal{W}_{n,m}(\psi) := \left\{ \boldsymbol{\Xi} \in \mathbb{M}_{n,m}(\mathbb{I}) : \begin{array}{l} |\mathbf{q}\boldsymbol{\Xi} - \mathbf{p}| < \psi(|\mathbf{q}|) \quad \text{for} \\ \text{i.m. } (\mathbf{p},\mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n \backslash \{\mathbf{0}\} \end{array} \right\} .$$

Here and below, $\mathbb{M}_{n,m}(\mathbb{I}) \subset \mathbb{M}_{n,m}$ denotes the set of $n \times m$ matrices with entries from $\mathbb{I} = (0,1)$. The following provides an elegant criterion for the size of the set $\mathcal{W}_{n,m}(\psi)$ expressed in terms of nm-dimensional Lebesgue measure. When $m = 1$, it coincides with Theorem 1.4 appearing in Sect. 1.2.4.

Theorem 1.12 (The Khintchine-Groshev Theorem) *Given any monotonic function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$, we have that*

$$|\mathcal{W}_{n,m}(\psi)|_{nm} = \begin{cases} 0 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q)^m < \infty , \\ 1 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q)^m = \infty . \end{cases}$$

Consider for the moment the function $\psi(r) = r^{-\frac{n}{m}} (\log r)^{-m}$ and observe that

$$\mathbf{Bad}(n,m) \cap \mathbb{M}_{n,m}(\mathbb{I}) \ \subset \ \mathbb{M}_{n,m}(\mathbb{I}) \setminus \mathcal{W}_{n,m}(\psi) .$$

By Theorem 1.12, $|\mathcal{W}_{n,m}(\psi)|_{nm} = 1$. Thus $|\mathbb{M}_{n,m}(\mathbb{I}) \setminus \mathcal{W}_{n,m}(\psi)|_{nm} = 0$ and on using the fact that set $\mathbf{Bad}(n,m)$ is invariant under translation by integer $n \times m$ matrices, it follows that

$$|\mathbf{Bad}(n,m)|_{nm} = 0 .$$

Another immediate consequence of the Khintchine-Groshev Theorem is the following statement (cf. Corollary 1.3).

Corollary 1.5 *Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be any function such that*

$$\sum_{q=1}^{\infty} q^{n-1}\psi(q)^m < \infty .$$

Then, for almost all $\boldsymbol{\Xi} \in \mathbb{M}_{n,m}$ *there exists a constant* $\kappa(\boldsymbol{\Xi}) > 0$ *such that*

$$|\mathbf{q}\boldsymbol{\Xi} + \mathbf{p}| \ > \ \kappa(\boldsymbol{\Xi})\,\psi(|\mathbf{q}|) \qquad \forall \ (\mathbf{p},\mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n \backslash \{\mathbf{0}\}\,.$$

The following is the natural generalisation of the set given by (1.63) to systems of linear forms and the subsequent statement is the natural generalisation of Theorem 1.5. Let

$$\mathcal{B}_{n,m}(\psi,\kappa) := \left\{ \boldsymbol{\Xi} \in \mathbb{M}_{n,m}(\mathbb{I}) : \begin{array}{c} |\mathbf{q}\boldsymbol{\Xi} + \mathbf{p}| > \kappa\psi(|\mathbf{q}|) \\ \forall \ (\mathbf{p},\mathbf{q}) \in \mathbb{Z}^m \times \mathbb{Z}^n \backslash \{\mathbf{0}\} \end{array} \right\}. \tag{1.81}$$

Theorem 1.13 (Effective Convergence Khintchine-Groshev Theorem) *Suppose that*

$$\sum_{q=1}^{\infty} q^{n-1}\psi(q)^m \ < \infty\,.$$

Then, for any $\kappa > 0$

$$\mathbf{Prob}(\mathcal{B}_{n,m}(\psi,\kappa)) \geq 1 - 2^m n\kappa^m \sum_{q=1}^{\infty}(2q+1)^{n-1}\psi(q)^m\,.$$

We highlight the fact that the probability in Theorem 1.13 is assumed to be uniform but it is possible to obtain a version for absolutely continuous distributions as already mentioned in Remark 1.3. Recall, that the Khintchine-Groshev theorem (with $m = 1$ and $n = 2$) underpinned the approach taken in Sect. 1.2.4 for calculating the Degrees of Freedom of the two-user X-channel (cf. Theorem 1.6).

We bring our selective overview of the general Diophantine approximation theory to a close by describing singular and jointly singular sets for systems of linear forms. In the process we shall prove Theorems 1.8 and 1.9. Recall, that the latter allows us to improve Theorem 1.6. For ease of comparison, it is convenient to define the sets of interest as follows:

$$\mathbf{Sing}(n,m) := \left\{ \boldsymbol{\Xi} \in \mathbb{M}_{n,m} : \begin{array}{c} \min\limits_{\substack{(\mathbf{p},\mathbf{q})\in\mathbb{Z}^m\times\mathbb{Z}^n: \\ 0<|\mathbf{q}|\leq Q}} \ \max\limits_{1\leq j\leq m} Q^{\frac{n}{m}}|\mathbf{q}.\boldsymbol{\xi}_j + p_j| \to 0 \\ \text{as} \ \ Q \to \infty \end{array} \right\}$$

and

$$\mathbf{Sing}^m(n) := \left\{ \boldsymbol{\Xi} \in \mathbb{M}_{n,m} : \begin{array}{c} \min\limits_{\substack{(\mathbf{p},\mathbf{q})\in\mathbb{Z}^m\times\mathbb{Z}^n: \\ 0<|\mathbf{q}|\leq Q}} \ \min\limits_{1\leq j\leq m} Q^{n}|\mathbf{q}.\boldsymbol{\xi}_j + p_j| \to 0 \\ \text{as} \ \ Q \to \infty \end{array} \right\}. \tag{1.82}$$

Clearly, when $m = 1$ the above two sets are equal and the elements coincide with the single linear form notion of singular points (cf. Definition 1.5). In recent groundbreaking work [25], Das, Fishman, Simmons & Urbański proved the following dimension statement (cf. Theorem 1.7) for the set of singular $n \times m$ matrices: *for all* $(n, m) \neq (1, 1)$*, we have that*

$$\dim \mathbf{Sing}(n, m) = mn\left(1 - \frac{1}{m+n}\right).$$

This resolved a conjecture of Kadyrov, Kleinbock, Lindenstrauss & Margulis [39]. In short, they showed that $\dim \mathbf{Sing}(n, m) \leq mn(1 - 1/(m+n))$ and conjectured that their upper bound is in fact sharp.

Regarding the set of jointly singular $n \times m$ matrices, it is clear that when $m = 2$ its elements coincide with the single linear form notion of jointly singular points (cf. Definition 1.7). Furthermore, it follows from the definition that for any integers $m_1, m_2 \geq 1$

$$\mathbf{Sing}^{m_1}(n) \times \mathbb{R}^{n \times m_2} \subset \mathbf{Sing}^{m_1+m_2}(n).$$

This together with Marstrand's Slicing Lemma and the fact $\mathbf{Sing}^1(n) = \mathbf{Sing}(n)$, implies that

$$\dim \mathbf{Sing}^m(n) \geq (m-1)n + \dim \mathbf{Sing}(n). \tag{1.83}$$

In turn, this together with Theorem 1.7, implies that for $n \geq 2$

$$\dim \mathbf{Sing}^m(n) \geq nm - \frac{n}{(n+1)}. \tag{1.84}$$

The following statement showing that we have equality in (1.84) is a natural generalisation of Theorem 1.8 to systems of linear forms.

Theorem 1.14 *Let* $m \geq 1$, $n \geq 2$. *Then*

$$\dim \mathbf{Sing}^m(n) = nm - \frac{n}{(n+1)}. \tag{1.85}$$

Clearly, when $m = 2$ the theorem coincides with Theorem 1.8. In view of (1.84), the key to establishing Theorem 1.14 (and thus Theorem 1.8) is the following upper bound statement.

Theorem 1.15 *Let* $m, n \geq 1$. *Then*

$$\dim \mathbf{Sing}^m(n) \leq nm - \frac{n}{(n+1)}. \tag{1.86}$$

Note that this upper bound estimate is valid for $n = 1$. Clearly, in this case it is not sharp when $m = 1$ since $\mathbf{Sing}^1(1) = \mathbf{Sing}(1) = \mathbb{Q}$ and so $\dim \mathbf{Sing}^1(1) = 0$. Also, note that the lower bound given by (1.83) does not match the upper bound given by (1.86). Nevertheless, we suspect that (1.86) is sharp when $m \geq 2$.

Problem 1.2 Let $m \geq 2$. Verify if $\dim \mathbf{Sing}^m(1) = m - \frac{1}{2}$.

Clearly, if true then we can replace the conditions on m and n in Theorem 1.14 by $mn > 1$. Although, not explicitly stated or even discussed, it is worth mentioning that Problem 1.1 concerning the set $\mathbf{Sing}_f^2(n)$ also has a natural generalisation to systems of linear form.

The proof of Theorem 1.15 (and indeed Theorem 1.9) makes use of the powerful connection between problems in Diophantine approximation an homogeneous dynamics. This we now briefly explain. The various Diophantine notions discussed in this chapter correspond to certain types of orbits of unimodular lattices under the action by diagonal matrices. For instance, as was famously discovered by Dani [23], a point $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$ is badly approximable if and only if the orbit

$$\left\{ g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1} : t > 0 \right\}$$

is bounded in the homogeneous space $X_{n+1} = \mathrm{SL}_{n+1}(\mathbb{R})/\mathrm{SL}_{n+1}(\mathbb{Z})$ of unimodular lattices in $\mathbb{R}^{n+1}$. Here and throughout,

$$g_t := \begin{pmatrix} e^{nt} & & & \\ & e^{-t} & & \\ & & \ddots & \\ & & & e^{-t} \end{pmatrix} \qquad \text{for} \quad t \in \mathbb{R}_+$$

and

$$u_{\boldsymbol{\xi}} := \begin{pmatrix} 1 & \xi_1 & \ldots & \xi_n \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} \qquad \text{for} \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n .$$

Today this beautiful and powerful equivalence between badly approximable points and the behaviour of orbits in X_{n+1} is simply refereed to as Dani's correspondence. For background and further details see for instance [24, 46].

Recall that the homogeneous space X_{n+1} is non-compact and, by Mahler's criterion, every bounded subset of X_{n+1} is contained in

$$K_\varepsilon := \left\{ \Lambda \in X_{n+1} : \inf_{\mathbf{v} \in \Lambda,\, \mathbf{v} \neq \mathbf{0}} \|\mathbf{v}\| \geq \varepsilon \right\}$$

for some $\varepsilon > 0$, where $\|\cdot\|$ is any norm on $\mathbb{R}^{n+1}$. With this in mind, in the same paper [23], Dani went on to show that $\boldsymbol{\xi} \in \mathbb{R}^n$ is singular if and only if the orbit $g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1}$ diverges as $t \to \infty$; that is, for any $\varepsilon > 0$ there exists a constant $t_{\varepsilon,\boldsymbol{\xi}} > 0$ such that

$$\forall\, t \geq t_{\varepsilon,\boldsymbol{\xi}} \qquad g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1} \notin K_\varepsilon\,.$$

This means that the orbit $g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1}$ leaves any bounded set 'forever' from some 'time' point $t_{\varepsilon,\boldsymbol{\xi}}$. In the same vein, it can be verifed that the matrix $\boldsymbol{\Xi} \in \mathbb{M}_{n,m}$ composed of the columns $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m \in \mathbb{R}^n$ is jointly singular if and only if for any $\varepsilon > 0$ there exists a constant $t_{\varepsilon,\boldsymbol{\Xi}} > 0$ such that

$$\forall\, t \geq t_{\varepsilon,\boldsymbol{\Xi}} \quad \exists\, j \in \{1, \ldots, m\} \qquad g_t u_{\boldsymbol{\xi}_j} \mathbb{Z}^{n+1} \notin K_\varepsilon\,. \tag{1.87}$$

Unlike for singular points, for every $j \in \{1, \ldots, m\}$ the individual orbit $g_t u_{\boldsymbol{\xi}_j} \mathbb{Z}^{n+1}$ need not be divergent and could in fact for some $\varepsilon > 0$ return to the bounded set K_ε arbitrarily often.

The proof of Theorem 1.15 and indeed Theorem 1.9 rely on the following powerful statement adapted for our application in mind due to Kadyrov, Kleinbock, Lindenstrauss & Margulis [39, Theorem 1.5]. Given $\boldsymbol{\xi} \in \mathbb{R}^n$, $N > 1$, $s > 0$ and $\varepsilon > 0$, let

$$S_{\boldsymbol{\xi}}(N, s, \varepsilon) := \{\ell \in \{1, \ldots, N\} : g_{s\ell} u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1} \notin K_\varepsilon\}\,.$$

Thus, $S_{\boldsymbol{\xi}}(N, s, \varepsilon)$ corresponds to those times $t = sl$ $(1 \leq \ell \leq N)$ for which the orbit $g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1}$ does not lie in K_ε. In what follows, given a set X we let $\#X$ denote its cardinality.

Theorem 1.16 (Kadyrov, Kleinbock, Lindenstrauss & Margulis) *Let B_1^n be the unit ball in $\mathbb{R}^n$ centred at the origin. Then there exist $s_0 > 1$ and $C > 0$ such that for any $s > s_0$, there exists $\varepsilon > 0$ such that for any $N \in \mathbb{N}$ and $\delta \in [0, 1)$, the set*

$$Z(\varepsilon, N, s, \delta) := \left\{\boldsymbol{\xi} \in B_1^n : \frac{\#S_{\boldsymbol{\xi}}(N, s, \varepsilon)}{N} \geq \delta\right\}$$

can be covered with $Cs^{3N} e^{(n+1-\delta)nsN}$ balls of radius $e^{-(n+1)sN}$.

Note that $\boldsymbol{\xi} \in Z(\varepsilon, N, s, \delta)$ if and only if the proportion of times $t = sl \leq sN$ $(1 \leq \ell \leq N)$ for which the orbit $g_t u_{\boldsymbol{\xi}} \mathbb{Z}^{n+1}$ avoids K_ε is at least δ. To be absolutely precise, the case when $\delta = 0$ is not covered by [39, Theorem 1.5]. However, it is trivially true since then $Z(\varepsilon, N, s, \delta) = B_1^n$ and the unit ball can easily be seen to be covered with $Ce^{(n+1-\delta)nsN}$ balls of radius $e^{-(n+1)sN}$. The next statement relates the jointly singular sets of interest to those appearing in Theorem 1.16.

Proposition 1.3 *Let $\varepsilon > 0$ and $s \geq 1$. Then*

$$\mathbf{Sing}^m(n) \cap \big(B_1^n\big)^m \subset \bigcup_{\boldsymbol{\delta} \in \Delta_s} \bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} Z_m(\varepsilon, N, s, \boldsymbol{\delta})\,, \tag{1.88}$$

where

$$\Delta_s := \Big\{ \boldsymbol{\delta} = (\delta_1, \ldots, \delta_m) \in \tfrac{1}{s}\mathbb{Z}^m \cap [0,1)^m : \delta_1 + \cdots + \delta_m \geq 1 - \tfrac{m+1}{s} \Big\}$$

and

$$Z_m(\varepsilon, N, s, \boldsymbol{\delta}) := Z(\varepsilon, N, s, \delta_1) \times \cdots \times Z(\varepsilon, N, s, \delta_m)\,.$$

Proof Recall, that given any $\Xi \in \mathbb{M}_{n,m}$ its column vectors are denoted by $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m \in \mathbb{R}^n$. Now, suppose that $\Xi \in \mathbf{Sing}^m(n) \cap \big(B_1^n\big)^m$. Then, by (1.87), for any $\varepsilon > 0$ and all $N > s^{-1} t_{\varepsilon,\Xi}$ we have that

$$\{\ell \in \mathbb{N} : s^{-1} t_{\varepsilon,\Xi} \leq \ell \leq N\} \subset \bigcup_{j=1}^{m} S_{\boldsymbol{\xi}_j}(N, s, \varepsilon)\,.$$

It follows that

$$\sum_{j=1}^{m} \#S_{\boldsymbol{\xi}_j}(N, s, \varepsilon) \;\geq\; N - s^{-1} t_{\varepsilon,\Xi}\,.$$

This implies that

$$\sum_{j=1}^{m} \frac{\#S_{\boldsymbol{\xi}_j}(N, s, \varepsilon)}{N} \geq 1 - \frac{t_{\varepsilon,\Xi}}{sN}\,. \tag{1.89}$$

For each $j \in \{1, \ldots, m\}$, let $\delta_j \in \frac{1}{s}\mathbb{Z}$ be the largest number such that

$$\frac{\#S_{\boldsymbol{\xi}_j}(N, s, \varepsilon)}{N} \geq \delta_j\,.$$

Then, with $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ we have that

$$\Xi \in Z_m(\varepsilon, N, s, \boldsymbol{\delta})\,. \tag{1.90}$$

We now show that $\boldsymbol{\delta} \in \Delta_s$. Since $\#S_{\boldsymbol{\xi}_j}(N, s, \varepsilon) \leq N$, we have that $0 \leq \delta_j \leq 1$. By the maximality of δ_j we have that

$$\delta_j + \frac{1}{s} \geq \frac{\#S_{\boldsymbol{\xi}_j}(N, s, \varepsilon)}{N} \geq \delta_j \,.$$

By (1.89), it follow that for N sufficiently large

$$\sum_{j=1}^{m} \delta_j \;\geq\; 1 - \frac{m}{s} - \frac{t_{\varepsilon,\boldsymbol{\Xi}}}{sN} \;\geq\; 1 - \frac{m+1}{s} \,. \tag{1.91}$$

Therefore, $\boldsymbol{\delta} \in \Delta_s$. Since Δ_s is finite, the latter condition together with (1.90) implies (1.88) and thereby completes the proof of the proposition. □

As we shall now see, armed with Theorem 1.16 and Proposition 1.3, it is relatively straightforward to establish Theorem 1.15 and indeed Theorem 1.9.

Proof *(Proof of Theorem 1.15)* Without loss of generality, it suffices to show (1.86) for the set $\mathbf{Sing}^m(n) \cap \left(B_1^n\right)^m$ instead of $\mathbf{Sing}^m(n)$. In short, this makes use of the fact that $\mathbf{Sing}^m(n)$ is contained in a countable union of translates of $\mathbf{Sing}^m(n) \cap \left(B_1^n\right)^m$. By Theorem 1.16, for $s > s_0$ and each $\boldsymbol{\delta} \in \Delta_s$, there exists a cover of $Z_m(\varepsilon, N, s, \boldsymbol{\delta})$ by

$$\prod_{j=1}^{m} Cs^{3N} e^{(n+1-\delta_j)nsN} \ll s^{3mN} e^{(n+1)nmsN-(1-\frac{m+1}{s})nsN}$$

balls of the same radius

$$r = e^{-(n+1)sN} \,. \tag{1.92}$$

Thus, in view of Proposition 1.3 and the trivial fact that

$$\#\Delta_s \leq (s+1)^m \,,$$

it follows that we have a cover of $\mathbf{Sing}^m(n) \cap \left(B_1^n\right)^m$ by

$$\ll (s+1)^m s^{3mN} e^{(n+1)nmsN-(1-\frac{m+1}{s})nsN}$$

balls of the same radius satisfying (1.92). Therefore, by the definition of Hausdorff dimension (see Definition 1.6 and Remark 1.18 immediately following it), for every $s > s_0$ we have that

$$\dim\left(\mathbf{Sing}^m(n) \cap \left(B_1^n\right)^m\right) \;\leq$$

$$\leq \limsup_{N\to\infty} \frac{\log\left((s+1)^m s^{3mN} e^{(n+1)nmsN-(1-\frac{m+1}{s})nsN}\right)}{-\log(e^{-(n+1)sN})}$$

$$= \limsup_{N\to\infty} \frac{3mN\log s + \left((n+1)nmsN - (1-\frac{m+1}{s})nsN\right)}{(n+1)sN}$$

$$= \frac{3m\log s + (n+1)nms - (1-\frac{m+1}{s})ns}{(n+1)s}\,.$$

Letting $s \to \infty$ gives

$$\dim\left(\mathbf{Sing}^m(n)\cap\left(B_1^n\right)^m\right) \;\le\; \frac{(n+1)nm-n}{n+1} \;=\; mn - \frac{n}{n+1}\,,$$

and thereby completes the proof of Theorem 1.15. □

Proof *(Proof of Theorem 1.9)* Given $\boldsymbol{f} : U \to \mathbb{R}^n$ as in the statement of the theorem, let

$$\mathcal{M}_{\boldsymbol{f}} := \{\boldsymbol{\Xi} \in \mathbb{M}_{n,2} : \boldsymbol{\xi}_2 = \boldsymbol{f}(\boldsymbol{\xi}_1)\}\,.$$

Since f is bi-Lipschitz,

$$\dim\left(\mathbf{Sing}^2_{\boldsymbol{f}}(n)\right) = \dim\left(\mathbf{Sing}^2(n)\cap\mathcal{M}_{\boldsymbol{f}}\right).$$

Therefore, (1.80) is equivalent to

$$\dim\left(\mathbf{Sing}^2(n)\cap\mathcal{M}_{\boldsymbol{f}}\right) \le n - \frac{n}{2(n+1)}. \tag{1.93}$$

As in the previous proof, it suffices to show (1.93) for $\mathbf{Sing}^2(n)\cap\mathcal{M}_{\boldsymbol{f}}\cap\left(B_1^n\right)^2$ instead of $\mathbf{Sing}^2(n)\cap\mathcal{M}_{\boldsymbol{f}}$. With this in mind, by Proposition 1.3, for any $\varepsilon > 0$ and any $s \ge 1$ we have that

$$\mathbf{Sing}^2(n)\cap\mathcal{M}_{\boldsymbol{f}}\cap\left(B_1^n\right)^2 \subset \bigcup_{\boldsymbol{\delta}\in\Delta_s}\bigcup_{N_0=1}^{\infty}\bigcap_{N=N_0}^{\infty} Z_2(\varepsilon,N,s,\boldsymbol{\delta})\cap\mathcal{M}_{\boldsymbol{f}}\,. \tag{1.94}$$

Observe that

$$\max\{\delta_1,\delta_2\} \ge \tfrac{1}{2} - \tfrac{3}{2s}\,,$$

and so by Theorem 1.16, for $s > s_0$ and each $\boldsymbol{\delta} \in \Delta_s$, we have a cover of $Z_2(\varepsilon,N,s,\boldsymbol{\delta})\cap\mathcal{M}_{\boldsymbol{f}}$ by

$$\min_{1\le j\le 2} Cs^{3N}e^{(n+1-\delta_j)nsN} \le Cs^{3N}e^{\left(n+1-\frac{1}{2}+\frac{3}{2s}\right)nsN}$$

balls of the same radius

$$r = e^{-(n+1)sN}. \tag{1.95}$$

Thus, in view of Proposition 1.3 and the trivial fact that

$$\#\Delta_s \leq (s+1)^2,$$

it follows that we have a cover of $\mathbf{Sing}^2(n) \cap \mathcal{M}_f \cap \left(B_1^n\right)^2$ by

$$\ll (s+1)^2 s^{3N} e^{\left(n+1-\frac{1}{2}+\frac{3}{2s}\right)nsN}$$

balls of the same radius r as given by (1.95). Therefore, for every $s > s_0$ we have that

$$\begin{aligned}
\dim\left(\mathbf{Sing}^2(n) \cap \mathcal{M}_f \cap \left(B_1^n\right)^2\right) \leq \\
&\leq \limsup_{N\to\infty} \frac{\log\left((s+1)^2 s^{3N} e^{\left(n+1-\frac{1}{2}+\frac{3}{2s}\right)nsN}\right)}{-\log(e^{-(n+1)sN})} \\
&= \limsup_{N\to\infty} \frac{3N\log s + \left(n+1-\frac{1}{2}+\frac{3}{2s}\right)nsN}{(n+1)sN} \\
&= \frac{3\log s + \left(n+1-\frac{1}{2}+\frac{3}{2s}\right)ns}{(n+1)s}.
\end{aligned}$$

On letting $s \to \infty$, gives

$$\dim\left(\mathbf{Sing}^2(n) \cap \mathcal{M}_f \cap \left(B_1^n\right)^2\right) \leq \frac{\left(n+1-\frac{1}{2}\right)n}{n+1} = n - \frac{n}{2(n+1)},$$

and thereby completes the proof of Theorem 1.9. □

As mentioned at the start of this subsection, even if we increased the number of users in the basic setup of Examples 1 & 2 we would still only need to call upon the general Diophantine approximation theory described above for a singular linear form (i.e., $m = 1$). A natural question that a reader may well be asking at this point is, whether or not there is a model of a communication channel that in its analysis requires us to genuinely exploit the general systems of linear forms theory with $m > 1$? The answer to this is emphatically yes. The simplest setup

that demonstrates this involves n users and one receiver equipped with m antennae. Recall, an antenna is a device (such as an old fashioned radio or television ariel) that is used to transmit or receive signals. Within Examples 1 & 2, each transmitter and receiver are implicitly understood to have a single antenna. This convention is pretty standard whenever the number of antennae at a transmitter or receiver is not specified. For a single receiver to be equipped with m antennae is in essence equivalent to m receivers (each with a single antenna) in cahoots with one another. The overall effect of sharing information is an increase in the probability that the receivers will be able to decode the transmitted messages. We now briefly explain how the setup alluded to above naturally brings into play the general Diophantine approximation theory for systems of linear forms.

Example 2A (Multi-Antennae Receivers) Suppose there are n users $S_1, \dots, S_n$ and two receivers R_1 and R_2 which 'cooperate' with one another. Furthermore, assume that $n \geq 3$. Let $Q \geq 1$ be an integer and suppose S_j wishes to transmit the message $u_j \in \{0, \dots, Q\}$ simultaneously to R_1 and R_2. Next, as in Example 2, for $i = 1, 2$ and $j = 1, \dots, n$, let h_{ij} denote the channel coefficients associated with the transmission of signals from S_j to R_i. Also, let y_i denote the signal received by R_i after (linear) encoding but before noise z_i is taken into account. Thus,

$$y_1 = \lambda \sum_{j=1}^{n} h_{1j}\alpha_j u_j \,, \tag{1.96}$$

$$y_2 = \lambda \sum_{j=1}^{n} h_{2j}\alpha_j u_j \,. \tag{1.97}$$

where $\lambda, \alpha_1, \dots, \alpha_n$ are some positive real numbers. Now let $d_{\min,i}$ the minimal distance between the $(Q+1)^n$ potential outcomes of y_i. Now, the larger the minimal distance $d_{\min,i}$ $(i = 1, 2)$ the greater the tolerance for noise and thus the more likely the receivers R_i are able to recover the messages $u_1, \dots, u_n$ by rounding $y_i' = y_i + z_i$ to the closest possible outcome of y_i (given by (1.97)). Thus, it is imperative to understand how $d_{\min,i}$ can be bounded below. Since R_1 and R_2 are sharing information (in fact it is better than that, they are actually the same person but they are not aware of it!), it is only necessary that at least one of $d_{\min,1}$ or $d_{\min,2}$ is relatively large compared to the noise. In other words, we need that the points $(y_1, y_2) \in \mathbb{R}^2$ are sufficiently separated. In order to analysis this, we first apply the inverse to the linear transformation

$$L := \begin{pmatrix} h_{11}\alpha_1 & h_{12}\alpha_2 \\ h_{21}\alpha_1 & h_{22}\alpha_2 \end{pmatrix}$$

to $(y_1, y_2)^t$. Without loss of generality, we can assume that the matrix norm of L and its inverse L^{-1} are bounded above. Therefore, the separation between the points

$(y_1, y_2) \in \mathbb{R}^2$ is comparable to the separation between the points $(\tilde{y}_1, \tilde{y}_2) \in \mathbb{R}^2$, where

$$(\tilde{y}_1, \tilde{y}_2)^t := L(y_1, y_2)^t .$$

Let $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \in \mathbb{R}^{n-2} \times \mathbb{R}^{n-2}$ be the pair corresponding to the two columns vectors of the matrix

$$\boldsymbol{\Xi} := \left(L^{-1} \begin{pmatrix} h_{13}\alpha_3 & \dots & h_{1n}\alpha_n \\ h_{23}\alpha_3 & \dots & h_{2n}\alpha_n \end{pmatrix} \right)^t .$$

The upshot, after a little manipulation, is that analysing the separation of the points $(y_1, y_2) \in \mathbb{R}^2$ equates to understanding the quantity

$$\max\{|\mathbf{q}\boldsymbol{\xi}_1 + p_1|, \mathbf{q}\boldsymbol{\xi}_2 + p_2|\}$$

for $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^2 \times \mathbb{Z}^{n-2}$ with $1 \leq |\mathbf{q}| \leq Q$. In particular, asking for good separation equates to obtaining good lower bounds on the quantity in question. In turn, this naturally brings into play the general Diophantine approximation theory for systems of 2 linear forms in $n - 2$ real variables. Note that assuming the number n of users is strictly greater than two (the number of cooperating receivers) simply avoids the degenerate case. For further details of the setup just described and its more sophisticated variants, we refer the reader to [49, Example 1] and [37, Section 3.2] and references within.

1.3 A 'child' Example and Diophantine Approximation on Manifolds

The theory of *Diophantine approximation on manifolds* (as coined by Bernik & Dodson in their Cambridge Tract [17]) or *Diophantine approximation of dependent quantities* (as coined by Sprindžuk in his monograph [63]) refers to the study of Diophantine properties of points in $\mathbb{R}^n$ whose coordinates are confined by functional relations or equivalently are restricted to a submanifold $\mathcal{M}$ of $\mathbb{R}^n$. In this section we consider an example of a communication channel which brings to the forefront the role of the theory of Diophantine approximation on manifolds in wireless communication.

Remark 1.21 The reader may well argue that in our analysis of the wireless communication model considered in Example 2, we have already touched upon the theory of Diophantine approximation on manifolds. Indeed, as pointed out on several occasions (see in particular Remarks 1.8 and 1.13), the points of interest $\boldsymbol{\xi} = (\xi_1, \xi_2)$ and $\boldsymbol{\xi}' = (\xi_1', \xi_2')$ associated with the example are functionally dependent. The explicit dependency is given by (1.51) and (1.52). However, it is

important to stress that the actual coordinates of each of these points are not subject to any dependency and so are not restricted to a sub-manifold of $\mathbb{R}^2$. The upshot of this is that we can analyse the points independently using the standard single linear form theory of Diophantine approximation in $\mathbb{R}^n$. In other words, the analysis within Example 2 does not require us to exploit the theory of Diophantine approximation on manifolds.

1.3.1 Example 3

In this example we will consider a model that involves several "transmitter-receiver" pairs who simultaneously communicate using shared communication channels. For the sake of simplicity we will concentrate on the case of three transmitter-receiver pairs; that is, we suppose that there are three users S_1, S_2 and S_3 and there are also three receivers R_1, R_2 and R_3. Let $Q \geq 1$ be an integer and suppose for each $j = 1, 2, 3$ the user S_j wishes to send a message $u_j \in \{0, \dots, Q\}$ to receiver R_j. After (linear) encoding, S_j transmits

$$x_j := \lambda \alpha_j u_j \tag{1.98}$$

where α_j is a positive real number and $\lambda \geq 1$ is a scaling factor. Note that apart form the obvious extra user S_3 and receiver R_3, the current setup is significantly different to that of Example 2 in that S_j does not wish to send independent messages to the receivers R_i $(i \neq j)$. In other words, we are not considering a three-user X-channel and thus, unlike Example 2, the codeword of user S_j does not have any component intended for any other receiver but R_j. Nevertheless, since the communication channel is being shared, as in Example 2, the signal x_j transmitted by S_j is being received by every receiver R_i with appropriate channel coefficients and thereby causing interference. Formally, for $i, j = 1, 2, 3$ let h_{ij} denote the channel coefficients associated with the transmission of signals from S_j to R_i. Also, let y_i denote the signal received by R_i before noise is taken into account. Thus,

$$y_i = \sum_{j=1}^{3} h_{ij} x_j \overset{(1.98)}{=} \lambda \sum_{j=1}^{3} h_{ij} \alpha_j u_j \,. \tag{1.99}$$

Now as usual, let us bring noise into the setup. If z_i denotes the (additive) noise at receiver R_i $(i = 1, 2, 3)$, then instead of (1.99), R_i receives the signal

$$y_i' = y_i + z_i \,. \tag{1.100}$$

Equations (1.99) and (1.100) represent one the simplest models of what is known as a *Gaussian Interference Channel* (GIC). The ultimate goal is for the receivers R_i $(i = 1, 2, 3)$ to decode the messages u_i from the observation of y_i'. This is

attainable if $2|z_i|$ is smaller than the minimal distance between the outcomes of y_i given by (1.99), which will be denoted by $d_{\min,i}$. As before, given that the nature of noise is often a random variable with normal distribution, the overarching goal is to ensure the probability that $|z_i| < \frac{1}{2}d_{\min,i}$ is large. Indeed, as in Examples 1 & 2, the larger the probability the more likely the receivers R_i $(i = 1, 2, 3)$ are able to recover messages by rounding y_i' (given by (1.100)) to the closest possible outcome of y_i (given by (1.99)). Thus, as in previous examples it is imperative to understand how $d_{\min,i}$ can be bounded below. Note that there are potentially $(Q+1)^3$ distinct outcomes of y_i and that

$$0 \leq y_i \ll \lambda Q \qquad (1 \leq i \leq 3), \tag{1.101}$$

where the implicit implied constants depend on the maximum of the channel coefficients h_{ij} and the encoding coefficients α_j. It is then easily verified, based on the outcomes of y_i given by (1.99) being equally spaced, that the minimal distance satisfies the following inequality

$$d_{\min,i} \ll \frac{\lambda}{Q^2} \qquad (1 \leq i \leq 3)\,. \tag{1.102}$$

Ideally, we would like to obtain lower bounds for $d_{\min,i}$ that are both "close" to this "theoretic" upper bound and are valid for a large class of possible choices of channel coefficients. Before we embark on the discussion of tools from Diophantine approximation that can be used for this purpose, we discuss how the idea of interference alignment introduced in the context of Example 2 extends to the setup of Example 3. This will naturally bring the theory of Diophantine approximation on manifolds into play.

Assume for the moment that $u_j \in \{0, 1\}$ and for the ease of discussion, let us just concentrate on the signal y_1 received at R_1. Then there are generally up to $2^3 = 8$ different outcomes for y_1. However, receiver R_1 is not interested in the signals u_2 and u_3. So if these signals could be deliberately aligned (at the transmitters) via encoding into a single component, then there would be fewer possible outcomes for y_1. Clearly, such an alignment would require that the ratio $h_{12}\alpha_2/h_{13}\alpha_3$ is a rational number. For example, if this ratio is equal to one, that is $h_{12}\alpha_2 = h_{13}\alpha_3$, then

$$y_1 = \lambda\Big(h_{11}u_1 + h_{12}\alpha_2(u_2 + u_3)\Big)\,.$$

Clearly, in this case the number of distinct outcomes of y_1 is reduced from 8 to 6, since there are 4 different pairs (u_2, u_3) as opposed to 3 different sums $u_2 + u_3$ when u_j take on binary values. Let us call the scenario described above *a perfect alignment*. For the received signals to be perfectly aligned at each receiver would require imposing highly restrictive constraints on the channel coefficients, which in practice would never be realised. Indeed, an encoding realising perfect alignment

simultaneously at each receiver would necessarily have that the following three ratios

$$\frac{h_{12}\alpha_2}{h_{13}\alpha_3}\,, \qquad \frac{h_{21}\alpha_1}{h_{23}\alpha_3}\,, \qquad \frac{h_{31}\alpha_1}{h_{32}\alpha_2}$$

are all rational numbers. For example, if all these ratios are equal to one then we have that

$$\det \begin{pmatrix} 0 & h_{12} & -h_{13} \\ h_{21} & 0 & -h_{23} \\ h_{31} & -h_{32} & 0 \end{pmatrix} = 0\,,$$

or equivalently, that

$$h_{12}h_{23}h_{31} = h_{32}h_{21}h_{13}\,.$$

In reality, for the channel coefficients to satisfy this equality would be so extraordinary that it is not worth considering. The upshot is that perfect alignment is simply not feasible.

Motahari et al. [52] proposed a scheme based on the method introduced by Cadambe et al. [20], which simultaneously at each receiver realises a *partial alignment* that is effectively arbitrarily close to perfect alignment. The basic idea is to split the messages u_j into 'blocks' and apply different linear encodings to each 'block'. As it happens, there is a choice of encodings that allows for all but a few of the received 'blocks' to be appropriately aligned as each receiver. On increasing the number of blocks one can approach perfect alignment with arbitrary accuracy. We now provide the details of the alluded scheme within the context of Example 3. Recall, the user S_j ($j = 1, 2, 3$) wishes to send a message $u_j \in \{0, \dots, Q\}$ to receiver R_j. In the first instance, given an integer $B \geq 2$ we let

$$u_{j,\mathbf{s}} \in \{0, \dots, B-1\}$$

be a collection of 'blocks' that determine (up to order) the coefficients in the base B expansion of u_j. Here and throughout, for $m, k \in \mathbb{N}$

$$\mathbf{s} = (s_1, \dots, s_m) \in \mathcal{S}_k := \{0, \dots, k-1\}^m$$

is a multi-index which is used to enumerate the blocks—in a moment we will take $m = 6$. Clearly, the number of different blocks (i.e. digits available to us when considering the base B expansion of a number) is equal to

$$M := k^m$$

and so the size of the message u_j that S_j can send to R_j is bounded above by $B^M - 1$. Without loss of generality, we can assume that

$$Q = B^M - 1\,. \tag{1.103}$$

Now, instead of transmitting (1.98), after encoding S_j transmits the message

$$x_j = \lambda \sum_{\mathbf{s}\in\mathcal{S}_k} \mathbf{T}^{\mathbf{s}} u_{j,\mathbf{s}}\,. \tag{1.104}$$

Here and throughout, for $\mathbf{s} \in \mathcal{S}_k$

$$\mathbf{T}^{\mathbf{s}} := T_1^{s_1} \cdots T_m^{s_m} \tag{1.105}$$

are real parameters called *transmit directions* obtained from a fixed finite set

$$\mathbf{T} := \{T_1, \ldots, T_m\}$$

of positive real numbers, called *generators*. As we shall soon see, the generators will be determined by the channel coefficients. In short, they play the role the positive real numbers α_j appearing in the encoding leading to (1.98). It is worth highlighting that the (linear) encoding leading to (1.104) varies from block to block. It follows that with this more sophisticated 'block' setup, instead of (1.99), the signal received by R_i before noise is taken into account is given by

$$\begin{aligned} y_i = \sum_{j=1}^{3} h_{ij} x_j \overset{(1.104)}{=} \lambda \sum_{j=1}^{3} h_{ij} \underbrace{\sum_{\mathbf{s}\in\mathcal{S}_k} \mathbf{T}^{\mathbf{s}} u_{j,\mathbf{s}}}_{x_j} \\ = \lambda \left(\underbrace{\sum_{\mathbf{s}\in\mathcal{S}_k} h_{ii} \mathbf{T}^{\mathbf{s}} u_{i,\mathbf{s}}}_{\text{wanted at } R_i} + \underbrace{\sum_{\substack{j=1 \\ j\neq i}}^{3} \sum_{\mathbf{s}\in\mathcal{S}_k} h_{ij} \mathbf{T}^{\mathbf{s}} u_{j,\mathbf{s}}}_{\text{unwanted at } R_i} \right). \end{aligned} \tag{1.106}$$

Thus, the unwanted message blocks $u_{j,\mathbf{s}}$ from S_j ($j \neq i$) arrive at R_i with the transmit directions $\mathbf{T}^{\mathbf{s}}$ multiplied by two possible channel coefficients h_{ij}. It follows that the unwanted blocks appearing in (1.106) constitute a linear form with $2M = 2k^m$ terms. We now choose the generators in such a way so as to align some of these unwanted blocks with the net effect of reducing the number of terms in the linear form. With this in mind, define the set of generators to be the collection of all channel coefficient with $i \neq j$; namely

$$\mathbf{T} = \{h_{12}, h_{13}, h_{21}, h_{23}, h_{31}, h_{32}\}\,. \tag{1.107}$$

Thus, $m = 6$ with respect to the general description above. With this choice of generators, it follows that the unwanted part within (1.106) can now be written as

$$\sum_{\mathbf{s}\in\mathcal{S}_{k+1}} \mathbf{T}^{\mathbf{s}} v_{i,\mathbf{s}} \tag{1.108}$$

where the terms

$$v_{i,\mathbf{s}} \in \{0, \dots, 2B-2\}$$

are integers formed as sums of up to two blocks $u_{j,\mathbf{s}}$. Note that the coefficients of $v_{i,\mathbf{s}}$ are monomials in the generators given by (1.107). Due to the multiplication by h_{ij} in (1.106) the exponents in the monomials appearing in (1.108) are up to k rather than just $k-1$. This explains why the summation in (1.108) is taken over $\mathcal{S}_{k+1}$ rather than just $\mathcal{S}_k$. The upshot of choosing $\mathbf{T}$ as in (1.107) is that the 'unwanted' linear form of $2M = 2k^6$ terms appearing in (1.106) is replaced by a linear form given by (1.108) of $(k+1)^6 = M(1+1/k)^6$ terms. In other words, asymptotically (as k increases) we have halved the number of terms associated with unwanted message blocks. On substituting (1.108) into (1.106) we get that

$$y_i = \lambda \left(\underbrace{\sum_{\mathbf{s}\in\mathcal{S}_k} h_{ii}\mathbf{T}^{\mathbf{s}} u_{i,\mathbf{s}}}_{\text{wanted at } R_i} + \underbrace{\sum_{\mathbf{s}\in\mathcal{S}_{k+1}} \mathbf{T}^{\mathbf{s}} v_{i,\mathbf{s}}}_{\text{unwanted at } R_i} \right). \tag{1.109}$$

Thus, y_i is a linear form of

$$M' := k^6 + (k+1)^6$$

terms[1] . Up to the factor λ, the coefficients of the integers $u_{i,\mathbf{s}}$ and $v_{i,\mathbf{s}}$ in (1.109) are monomials in the six generators of $\mathbf{T}$ and are all different. It is convenient to represent these coefficients as a 'coefficient' vector

$$\mathbf{G}_i := (G_{i,0}, G_{i,1}, \dots, G_{i,n}) \quad \text{where} \quad n := M' - 1\,. \tag{1.110}$$

To reiterate, the components $G_{i,0}, G_{i,1}, \dots, G_{i,n}$ are the real numbers

$$\mathbf{T}^{\mathbf{s}} \text{ with } \mathbf{s} \in \mathcal{S}_{k+1} \quad \text{and} \quad h_{ii}\mathbf{T}^{\mathbf{s}} \text{ with } \mathbf{s} \in \mathcal{S}_k \tag{1.111}$$

[1]Observe that essentially half of the terms in (1.109) are wanted at R_i compared to only a third (before alignment) in (1.106) or indeed in (1.101).

written in any fixed order. It is easily verified that for any $\epsilon > 0$, for k sufficiently large

$$2M < n < 2M + \epsilon\,. \tag{1.112}$$

Now let

$$\boldsymbol{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,n}) := \Big(\frac{G_{i,1}}{G_{i,0}}, \ldots, \frac{G_{i,n}}{G_{i,0}}\Big) \qquad (1 \le i \le 3)\,. \tag{1.113}$$

Returning to (1.109), it is easily seen that there are potentially $B^{M'}$ distinct outcomes of y_i and as before (cf. (1.101))

$$0 \le y_i \ll 2\lambda B \qquad (1 \le i \le 3), \tag{1.114}$$

where the implicit implied constants depend on the maximum of the channel coefficients h_{ij} and the integer k. Now let $d_{\min,i}$ denote the minimal distance between the outcomes of y_i given by (1.109). It is then easily verified, based on these outcomes being equally spaced, that the minimal distance satisfies the following inequality (cf. (1.102))

$$d_{\min,i} \ll \frac{\lambda B}{B^{M'}} = \frac{\lambda}{B^n} \le \frac{\lambda}{Q^2} \qquad (1 \le i \le 3)\,. \tag{1.115}$$

The last inequality makes use of (1.103) and (1.112). Recall, that our goal is the same as in all previous examples. We wish to obtain lower bounds for $d_{\min,i}$ that are both "close" to this "theoretic" upper bound and at the same time are valid for a large class of possible choices of channel coefficients. As we have seen in Examples 1 & 2, the goal is intimately related to the Diophantine properties of certain points defined via the channel coefficients. Within the context of Example 3, the points of interest are precisely those corresponding to $\boldsymbol{\xi}_i \in \mathbb{R}^n$ as given by (1.113). In Sect. 1.3.2, we will demonstrate that this is indeed the case by calculating the DoF of the three-user Gaussian Interference Channel (GIC). First we make an important observation: *the coordinates of each point $\boldsymbol{\xi}_i$ ($i = 1, 2, 3$) are functions of seven variables and are therefore dependent*. The latter follows since $k \ge 1$ and so by definition $n \ge 2^6 > 7$. The fact that the point $\boldsymbol{\xi}_i$ of interest is of dependent variables implies that $\boldsymbol{\xi}_i$ lies on a submanifold $\mathcal{M}$ of $\mathbb{R}^n$ of dimension strictly smaller that n. Trivially, since the dimension of $\mathcal{M}$ is strictly less than n, we have that the n-dimension Lebesgue measure of $\mathcal{M}$ is zero. The upshot of the dependency is that all the measure theoretic Diophantine approximation results (such as those concerning badly approximable, ψ-approximable, Dirichlet improvable, singular, etc etc) that we have exploited so far in our analysis of Examples 1 & 2 are pretty much redundant. We need a theory which takes into account that the points of interest lie on a submanifold $\mathcal{M}$ of $\mathbb{R}^n$. Luckily, today the metric theory of Diophantine

approximation on manifolds is in reasonable shape. Indeed, for a large class of so called non-degenerate manifolds there exists

(i) a rich badly approximable theory concerning $\mathbf{Bad}(n) \cap \mathcal{M}$—see for example [3, 5, 7, 14, 15, 64] and references within,
(ii) a rich ψ-approximable theory concerning $\mathcal{W}_n(\psi) \cap \mathcal{M}$—see for example [1, 6, 11, 18, 31–35, 41, 42] and references within, and
(iii) a rich Dirichlet improvable theory concerning $\mathbf{DI}(n) \cap \mathcal{M}$—see for example [43, 44, 61] and references within.

For a general overview of the manifold theory we refer the reader to [16, Section 6]. In short, the recent state of the art results for the sets just listed suffice to implement the approaches taken in Sects. 1.2.2 to 1.2.5 within the context of Example 3. As already mentioned, we will shortly provide the details of how the 'Khintchine-Groshev' approach of Sect. 1.2.4 translates to the current setup.

Observe that in above list of Diophantine sets restricted to $\mathcal{M}$ there is a notable exception. We have not mentioned singular (resp. jointly singular) sets $\mathbf{Sing}(n)$ (resp. $\mathbf{Sing}^2(n)$) and in turn we have avoided mentioning the approach taken in Sect. 1.2.6 that enables us to improve the result of Motahari et al. on the DoF of a two-user X-channel. The reason for this is simple—our current knowledge of $\mathbf{Sing}(n) \cap \mathcal{M}$ is not sufficient. We will come back to this in Sect. 1.3.3.

1.3.2 The Khintchine-Groshev Theorem for Manifolds and DoF

The goal of this section is twofold. The first is to introduce the analogue of the Khintchine-Groshev Theorem for one linear form (i.e. Theorem 1.4 in Sect. 1.2.4) in which the points of interest are restricted to a submanifold of $\mathbb{R}^n$. The second is to exploit this so called Khintchine-Groshev theorem for manifolds to calculate the DoF of the three-user GIC considered in Example 3.

Let $\mathcal{M}$ be a submanifold of $\mathbb{R}^n$ and let $\mathcal{W}_n(\psi)$ be the set of ψ-approximable points in $\mathbb{R}^n$ defined by (1.58). In short, if the manifold is "sufficiently" curved the Khintchine-Groshev theorem for manifolds provides a 'zero-one' criterion for the Lebesgue measure of the set

$$\mathcal{W}_n(\psi) \cap \mathcal{M}\,.$$

Observe that if the dimension of the manifold is strictly less than n, then with respect to n-dimensional Lebesgue measure we trivially have that $|\mathcal{W}_n(\psi) \cap \mathcal{M}|_n = 0$ irrespective of the approximating function ψ. Thus, when referring to the Lebesgue measure of the set $\mathcal{W}_n(\psi) \cap \mathcal{M}$ it is always with reference to the induced Lebesgue measure on $\mathcal{M}$. More generally, given a subset S of $\mathcal{M}$ we shall write $|S|_{\mathcal{M}}$ for the measure of S with respect to the induced Lebesgue measure on $\mathcal{M}$. Without loss of

generality, we will assume that

$$|\mathcal{M}|_{\mathcal{M}} = 1$$

since otherwise the induced measure can be re-normalized accordingly. It is not particularly difficult to show that in order to obtain an analogue of Theorem 1.4 (both the convergence and divergence aspects) for $\mathcal{W}_n(\psi) \cap \mathcal{M}$ we need to avoid hyperplanes—see [16, Section 4.5]. To overcome such natural counterexamples, we insist that $\mathcal{M}$ is a *non–degenerate* manifold.

Non-Degenerate Manifolds Essentially, these are smooth submanifolds of $\mathbb{R}^n$ which are sufficiently curved so as to deviate from any hyperplane. Formally, a manifold $\mathcal{M}$ of dimension d embedded in $\mathbb{R}^n$ is said to be *non-degenerate* if it arises from a non-degenerate map $\mathbf{f} : U \to \mathbb{R}^n$ where U is an open subset of $\mathbb{R}^d$ and $\mathcal{M} := \mathbf{f}(U)$. The map $\mathbf{f} : U \to \mathbb{R}^n, \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}))$ is said to be *l-non-degenerate at* $\mathbf{x} \in U$, where $l \in \mathbb{N}$, if $\mathbf{f}$ is l times continuously differentiable on some sufficiently small ball centred at $\mathbf{x}$ and the partial derivatives of $\mathbf{f}$ at $\mathbf{x}$ of orders up to l span $\mathbb{R}^n$. The map $\mathbf{f}$ is *non-degenerate* at $\mathbf{x}$ if it is l-non-degenerate at $\mathbf{x}$ for some $l \in \mathbb{N}$. The map $\mathbf{f}$ is *non-degenerate* if it is non-degenerate at almost every (in terms of d-dimensional Lebesgue measure) point $\mathbf{x}$ in U; in turn the manifold $\mathcal{M} = \mathbf{f}(U)$ is also said to be non-degenerate. It is well known, that any real connected analytic manifold not contained in any hyperplane of $\mathbb{R}^n$ is non-degenerate at every point [42]. In the case the manifold $\mathcal{M}$ is a planar curve C, a point on C is non-degenerate if the curvature at that point is non-zero. Moreover, it is not difficult to show that the set of points on a planar curve at which the curvature vanishes but the curve is non-degenerate is at most countable, see [8, Lemmas 2 & 3]. In view of this, the curvature completely describes the non-degeneracy of planar curves. Clearly, a straight line is degenerate everywhere.

The convergence part of the following statement was independently established in [6] and [18], while the divergence part was established in [11].

Theorem 1.17 (Khintchine-Groshev for Manifolds) *Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonic function and let $\mathcal{M}$ be a non-degenerate submanifold of $\mathbb{R}^n$. Then*

$$|\mathcal{W}_n(\psi) \cap \mathcal{M}|_{\mathcal{M}} = \begin{cases} 0 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q) < \infty\,, \\ 1 \text{ if } \sum_{q=1}^{\infty} q^{n-1}\psi(q) = \infty\,. \end{cases}$$

Remark 1.22 In view of Corollary 1.2 in Sect. 1.2.2, it follows that

$$\mathcal{W}_n(\psi) \cap \mathcal{M} = \mathcal{M} \quad \text{if} \quad \psi : q \mapsto q^{-n}\,.$$

Now, given $\varepsilon > 0$ consider the function $\psi_\varepsilon : q \mapsto q^{-n-\varepsilon}$. A submanifold $\mathcal{M}$ of $\mathbb{R}^n$ is called *extremal* if

$$|\mathcal{W}_n(\psi_\varepsilon) \cap \mathcal{M}|_{\mathcal{M}} = 0\,.$$

Sprindžuk (1980) conjectured that any analytic non-degenerate submanifold is extremal. In their pioneering work [42], Kleinbock & Margulis proved that any non-degenerate submanifold $\mathcal{M}$ of $\mathbb{R}^n$ is extremal and thus established Sprindžuk's conjecture. It is easy to see that this implies the convergence case of Theorem 1.17 for functions of the shape ψ_ε.

Remark 1.23 For the sake of completeness, it is worth mentioning that the externality theorem for non-degenerate submanifolds of $\mathbb{R}^n$ has been extended in recent years to submanifolds of $n \times m$ matrices, see [2, 13, 45].

An immediate consequence of the convergence case of Theorem 1.17 is the following statement (cf. Corollary 1.3).

Corollary 1.6 *Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a function such that*

$$\sum_{q=1}^{\infty} q^{n-1}\psi(q) < \infty\,. \tag{1.116}$$

Suppose that $\mathcal{M}$ is as in Theorem 1.17. Then, for almost all $\boldsymbol{\xi} \in \mathcal{M}$ there exists a constant $\kappa(\boldsymbol{\xi}) > 0$ such that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \;>\; \kappa(\boldsymbol{\xi})\,\psi(|\mathbf{q}|) \qquad \forall\ (p,\mathbf{q}) \in \mathbb{Z}\times\mathbb{Z}^n\backslash\{\mathbf{0}\}\,. \tag{1.117}$$

In line with the discussion in Sect. 1.2.4 preceding the statement of the effective convergence Khintchine-Groshev theorem (i.e. Theorem 1.5), a natural question to consider is: *can the constant $\kappa(\boldsymbol{\xi})$ within Corollary 1.6 be made independent of $\boldsymbol{\xi}$?* The argument involving the set $\mathcal{B}_n(\psi,\kappa)$ given by (1.63) can be modified to show that this is impossible to guarantee with probability one; that is, for almost all $\boldsymbol{\xi} \in \mathcal{M}$. Nevertheless, the following result provides an effective solution to the above question. It is a special case of [1, Theorem 3].

Theorem 1.18 (Effective Convergence Khintchine-Groshev for Manifolds) *Let $l \in \mathbb{N}$ and let $\mathcal{M}$ be a compact d-dimensional C^{l+1} submanifold of $\mathbb{R}^n$ that is l-non-degenerate at every point. Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonically decreasing function such that*

$$\Sigma_\psi := \sum_{q=1} q^{n-1}\psi(q) < \infty\,. \tag{1.118}$$

Then there exist positive constants κ_0, C_1 depending on ψ and $\mathcal{M}$ only and C_0 depending on the dimension of $\mathcal{M}$ only such that for any $0 < \delta < 1$, the inequality

$$|\mathcal{B}_n(\psi,\kappa) \cap \mathcal{M}|_{\mathcal{M}} \geq 1 - \delta \tag{1.119}$$

holds with

$$\kappa := \min\left\{\kappa_0,\ \frac{C_0\delta}{\Sigma_\psi},\ C_1\delta^{d(n+1)(2l-1)}\right\}. \tag{1.120}$$

Remark 1.24 The constants appearing in (1.120) are explicitly computable, see [1, Theorem 6] for such a statement. In [31] Theorem 1.18 was also extended to a natural class of affine subspaces, which by definition are degenerate.

We now move onto our second goal: to exploit the Khintchine-Groshev theorem for manifolds to calculate the DoF of the three-user GIC considered in Example 3. The overall approach is similar to that used in Sect. 1.2.4 to calculate the DoF of the two-user X-channel considered in Example 2. In view of this we will keep the following exposition rather brief and refer the reader to Sect. 1.2.4 for both the motivation and the details. With this in mind, let $\mathcal{M}$ denote the 7-dimensional submanifold of $\mathbb{R}^n$ arising from the implicit dependency within (1.113). In other words, a point $\boldsymbol{\xi}_i \in \mathcal{M}$ if and only if it is of the form (1.113). That $\mathcal{M}$ is of dimension 7 follows from the fact that the monomials $G_{i,0}, G_{i,1}, \ldots, G_{i,n}$ depend on h_{ii} and the other 6 channel coefficients that form the set $\mathbf{T}$ of generators. It is also not difficult to see that these monomials are all different and therefore linearly independent over $\mathbb{R}$. Consequently, $1, \xi_{i,1}, \ldots, \xi_{i,n}$ are linearly independent over $\mathbb{R}$ as functions of the corresponding channel coefficients. Hence $\mathcal{M}$ cannot be contained in any hyperplane of $\mathbb{R}^n$. Also note that $\mathcal{M}$ is connected and analytic, and therefore, it is non-degenerate.

Now suppose that

$$\boldsymbol{\xi} \notin \mathcal{W}_n(\psi) \tag{1.121}$$

where $\psi : q \to q^{-n-\varepsilon}$ for some $\varepsilon > 0$. Then, Corollary 1.6 implies that for almost all $\boldsymbol{\xi} \in \mathcal{M}$ there exists a constant $\kappa(\boldsymbol{\xi}) > 0$ such that

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \frac{\kappa(\boldsymbol{\xi})}{|\mathbf{q}|^{n+\varepsilon}}$$

for all $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n \backslash \{\mathbf{0}\}$. Here and throughout the rest of this section, almost all is with respect to 7-dimensional Lebesgue measure induced on $\mathcal{M}$. In particular, it follows that for almost all $\boldsymbol{\xi} \in \mathcal{M}$ and every $B \in \mathbb{N}$ we have that (cf. (1.61))

$$|q_1\xi_1 + \cdots + q_n\xi_n + p| \geq \frac{\kappa(\boldsymbol{\xi})}{B^{n+\varepsilon}} \tag{1.122}$$

for all $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^n \backslash \{\mathbf{0}\}$ with $1 \leq |\mathbf{q}| \leq B$. Then, the analysis as in Sect. 1.2.4 that leads to (1.62), enables us to make the following analogous statement: with probability one, for every $B \geq 2$ and a random choice of channel coefficients h_{ij}

$(i, j = 1, 2, 3)$, the minimum separation between the associated points y_i given by (1.109) satisfies

$$d_{\min,i} \gg \frac{\lambda\, \kappa(\boldsymbol{\xi}_i)}{B^{n+\varepsilon}} \qquad (1 \le i \le 3)\,. \tag{1.123}$$

We stress, that $\boldsymbol{\xi}_i$ corresponds to the point given by (1.113) associated with the choice of channel coefficients. Recall, that the latter determine the set of generators (1.107) which in turn determine the coefficient vector $\mathbf{G}_i$ and therefore the point $\boldsymbol{\xi}_i$. Note that apart from the extra ε term in the power, the lower bound (1.123) coincides (up to constants) with the upper bound (1.115).

Now, in relation to Example 3, the power constraint P on the channel model means that

$$|x_j|^2 \le P \qquad (j = 1, 2, 3)\,, \tag{1.124}$$

where x_j is the codeword transmitted by S_j as given by (1.104). Now notice that since the blocks $u_{j,\mathbf{s}}$ ($\mathbf{s} \in \mathcal{S}_k$) are integers lying in $\{0, \dots, B-1\}$, it follows that

$$|x_j| \ll \lambda B\,,$$

where the implied implicit constant is independent from B and λ. Hence, we conclude that P is comparable to $(\lambda B)^2$. It is shown in [52, §5], that the probability of error in transmission within Example 3 is bounded above by (1.65) with

$$d_{\min} = \min\{d_{\min,1}, d_{\min,2}, d_{\min,3}\}.$$

Recall, in order to achieve reliable transmission one requires that this probability tends to zero as $P \to \infty$. Then, on assuming (1.123)—which holds for almost every $\boldsymbol{\xi}_i \in \mathcal{M}$—it follows that

$$d_{\min} \gg \frac{\lambda}{B^{n+\varepsilon}}\,, \tag{1.125}$$

and so the quantity (1.65) will tend to zero as $B \to \infty$ if we set

$$\lambda = B^{n+2\varepsilon}\,.$$

The upshot of this is that we will achieve a reliable transmission rate under the power constraint (1.124) if we set P to be comparable to $B^{2n+2+4\varepsilon}$; that is

$$B^{2n+2+4\varepsilon} \ll P \ll B^{2n+2+4\varepsilon}\,.$$

Next, recall that the largest message u_j that user S_j can send to R_j is given by (1.103). Thus, it follows that the number of bits (binary digits) that user S_j transmits is approximately

$$\log B^M = M \log B \,.$$

Therefore, in total the three users S_j $(j = 1, 2, 3)$ transmit approximately $3M \times \log B$ bits, which with our choice of P is an achievable total rate of reliable transmission; however, it may not be maximal. On comparing this to the rate of reliable transmission for the simple point to point channel under the same power constraint, we get that the total DoF of the three-user GIC is at least

$$\lim_{P\to\infty} \frac{3M \log B}{\frac{1}{2}\log(1+P)} = \lim_{B\to\infty} \frac{3M \log B}{\frac{1}{2}\log(1+B^{2n+2+4\varepsilon})} = \frac{3M}{n+1+2\varepsilon} \,. \tag{1.126}$$

Given that $\varepsilon > 0$ is arbitrary, it follows that for almost every (with respect to the 7-dimensional Lebesgue measure) realisation of the channel coefficients

$$\text{DoF} \geq \frac{3M}{n+1} \,.$$

Now recall that $n + 1 = M' = (k+1)^6 + k^6$ and $M = k^6$. On substituting these values into the above lower bound, we obtain that

$$\text{DoF} \geq \frac{3k^6}{(k+1)^6 + k^6} \,.$$

Given that k is arbitrary, it follows (on letting $k \to \infty$) that for almost every realisation of the channel coefficients

$$\text{DoF} \geq \frac{3}{2} \,.$$

Now it was shown in [20] that the DoF of a three-user GIC is upper bounded by $3/2$ for all choices of the channel coefficients, and so it follows that for almost every realisation of the channel coefficients

$$\text{DoF} = \frac{3}{2} \,. \tag{1.127}$$

1.3.3 Singular and Non-Singular Points on Manifolds

With reference to Example 3, we have seen in the previous section that the Khintchine-Groshev theorem for non-degenerate manifolds allows us to achieve

good separation between the received signals y_i given by (1.123). More precisely, for almost all choices of the channel coefficients h_{ij} $(i, j = 1, 2, 3)$ we obtain the lower bounds (1.115) for the minimal distances $d_{\min,i}$ that are only 'ε-weaker' than the 'theoretic' upper bounds as given by (1.123). As in the discussion at the start of Sect. 1.2.6, this motivates the question of *whether good separation and indeed if the total DoF of 3/2 for the three-user GIC can be achieved for a larger class of channel coefficients?* Concerning the latter, what we have in mind is a statement along the lines of Theorem 1.10 that improves the Motahari et al. result (Theorem 1.6) for the total DoF of the two-user X-channel. Beyond this, but still in a similar vein, one can ask if the more general DoF results of Motahari et al. [52] for communications channels involving more users and receivers can be improved? Clearly, the approach taken in Sects. 1.2.6 and 1.2.7 based on the Diophantine approximation theory of non-singular and jointly non-singular points can be utilized to make the desired improvements. However there is a snag—we would require the existence of such a theory in which the points of interest are restricted to non-degenerate manifolds. Unfortunately, the analogues of Theorems 1.7, 1.8, 1.9, 1.14 and 1.15 for manifolds are not currently available. In short, obtaining any such statement represents a significant open problem in the theory of Diophantine approximation on manifolds. Indeed, even partial statements such as the following currently seem out of reach. As we shall see, it has non-trivial implications for both number theory and wireless communication.

Problem 1.3 Let $n \geq 2$ and $\mathcal{M}$ be any analytic non-degenerate submanifold of $\mathbb{R}^n$ of dimension d. Verify if

$$\dim \big(\mathbf{Sing}(n) \cap \mathcal{M}\big) < d := \dim \big(\mathcal{M}\big)\,. \tag{1.128}$$

Recall, that $\mathbf{Sing}(n)$ is the set of singular points in $\mathbb{R}^n$—see Definition 1.5 in Sect. 1.2.6.

Remark 1.25 Determining the actual value for the Hausdorff dimension of the set $\mathbf{Sing}(n) \cap \mathcal{M}$ for special classes of submanifolds $\mathcal{M}$ (such as polynomial curves—see below) would be most desirable. It is not difficult to see that the intersection of $\mathcal{M}$ with any rational hyperplane is contained in $\mathbf{Sing}(n)$. Therefore,

$$\dim \big(\mathbf{Sing}(n) \cap \mathcal{M}\big) \geq d - 1\,.$$

When $d > 1$, this gives a non-trivial lower bound. Obviously, when $d = 1$ the lower bound is trivial.

From a purely number theoretic point of view, Problem 1.3 is of particular interest when the manifold is a curve ($d = 1$). It has a well-known connection to the famous and notorious problem posed by Wirsing (1961) and later restated in a stronger form by Schmidt [59, pg. 258]. This we now briefly describe. The Wirsing-Schmidt conjecture is concerned with the approximation of real numbers by algebraic numbers of bounded degree. The proximity of the approximation is

measured in terms of the height of the algebraic numbers. Recall, that given a polynomial P with integer coefficients, the height $H(P)$ of P is defined to be the maximum of the absolute values of the coefficients of P. In turn the height $H(\alpha)$ of an algebraic number α is the height of the minimal defining polynomial P of α over $\mathbb{Z}$.

Conjecture 1 (Wirsing-Schmidt) Let $n \geq 2$ and ξ be any real number that is not algebraic of degree $\leq n$. Then there exists a constant $C = C(n, \xi)$ and infinitely many algebraic numbers α of degree $\leq n$, such that

$$|\xi - \alpha| < C\, H(\alpha)^{-n-1}\,. \tag{1.129}$$

Note that when $n = 1$ the conjecture is trivially true since it coincides with the classical corollary to Dirichlet's theorem—the first theorem stated in this chapter. For $n = 2$ the conjecture was proved by Davenport & Schmidt (1967). For $n \geq 3$ there are only partial results. For recent progress and an overview of previous results we refer the reader to [4] and references within.

The connection between the Wirsing-Schmidt conjecture and Problem 1.3 comes about via the well know fact that the former is intimately related to singular points on the Veronese curves $\mathcal{V}_n := \{(\xi, \xi^2, \ldots, \xi^n) : \xi \in \mathbb{R}\}$.

Lemma 1.5 *Let $n \geq 2$ and $\xi \in \mathbb{R}$. If $(\xi, \xi^2, \ldots, \xi^n) \notin \mathbf{Sing}(n)$, then the Wirsing-Schmidt conjecture holds for ξ.*

The proof of the lemma is pretty standard. For example, it easily follows by adapting the argument appearing in [7, Appendix B] in an obvious manner. A straightforward consequence of the lemma is that any upper bound for $\dim\big(\mathbf{Sing}(n) \cap \mathcal{V}\big)$ gives an upper bound on the dimension of the set of potential counterexamples to the Wirsing-Schmidt conjecture. When $n \geq 3$, currently we do not even know that the set of potential counterexamples has dimension strictly less than one—the trivial bound. Clearly, progress on Problem 1.3 with $\mathcal{M} = \mathcal{V}_n$ would rectify this gaping hole in our knowledge.

We now turn our attention to the question raised at the start of this subsection; namely, whether good separation and the total DoF of 3/2 within the setup of Example 3 can be achieved for a larger class of channel coefficients? To start with we recall that the 7-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^n$ arising from the implicit dependency within (1.113) is both analytic and non-degenerate. Thus it falls under the umbrella of Problem 1.3. In turn, on naturally adapting the argument used to establish Proposition 1.1, a consequence of the upper bound (1.128) is the following statement: for all choice of channel coefficients $\{h_{ii}, h_{12}, h_{13}, h_{21}, h_{23}, h_{31}, h_{32}\}$ $(i = 1, 2, 3)$ except on a subset of strictly positive codimension, the minimum separation $d_{\min,i}$ between the associated points y_i given by (1.109) satisfies (1.123). The upshot is that if true, Problem 1.3 enables us to obtain good separation for a larger class of channel coefficients than the (unconditional) Khintchine-Groshev approach outlined in Sect. 1.3.2.

As we have seen within the setup of Example 2, in order to improve the 'almost all' DoF result (Theorem 1.6) of Motahari et al. we need to work with the jointly singular set $\mathbf{Sing}^2_{\boldsymbol{f}}(n)$ appearing in Theorem 1.9. This theorem provides a non-trivial upper bound for the Hausdorff dimension of such sets and is the key to establishing the stronger DoF statement Theorem 1.10. With this in mind, we suspect that progress on the following problem is at the heart of improving the 'almost all' DoF result for the three-user GIC (see (1.127)) obtained via the Khintchine-Groshev approach. In any case, we believe that the problem is of interest in its own right. Recall, that $\mathbf{Sing}^m(n)$ is given by (1.82) and is the jointly singular set for systems of linear forms.

Problem 1.4 Let $k, \ell, m, d \in \mathbb{N}$, $n = k+\ell$, $U \subset \mathbb{R}^d$ and $V \subset \mathbb{R}^m$ be open subsets. Suppose that $\boldsymbol{f} : U \to \mathbb{R}^k$ and $\boldsymbol{g} : U \to \mathbb{R}^\ell$ are polynomial non-degenerate maps. For each $\mathbf{u} \in U$ and $\mathbf{v} \in V$ let $\Xi(\mathbf{u}, \mathbf{v})$ be the matrix with columns $(v_i \boldsymbol{f}(\mathbf{u}), \boldsymbol{g}(\mathbf{u}))^t$ and let

$$\mathbf{Sing}^m_{\boldsymbol{f},\boldsymbol{g}}(n) := \Big\{ (\mathbf{u}, \mathbf{v}) \in U \times V : \Xi(\mathbf{u}, \mathbf{v}) \in \mathbf{Sing}^m(n) \Big\} .$$

Verify if

$$\dim \Big(\mathbf{Sing}^m_{\boldsymbol{f},\boldsymbol{g}}(n) \Big) < d + m .$$

Of course, it would be natural to generalise the problem by replacing 'polynomial' with 'analytic' and by widening the scope of the $n \times m$ matrices under consideration. On another front, staying within the setup of Problem 1.4, it would be highly desirable to determine the actual value for the Hausdorff dimension of the set $\mathbf{Sing}^m_{\boldsymbol{f},\boldsymbol{g}}(n)$. This represents a major challenge.

Acknowledgments The authors are grateful to Anish Ghosh and Cong Ling for their valuable comments on an earlier version of this chapter. We would also like to thank Mohammad Ali Maddah-Ali for bringing [53] to our attention (see Remark 1.11).

References

1. Adiceam, F., Beresnevich, V., Levesley, J., Velani, S., Zorin, E.: Diophantine approximation and applications in interference alignment. Adv. Math. **302**, 231–279 (2016)
2. Aka, M., Breuillard, E., Rosenzweig, L., de Saxcé, N.: Diophantine approximation on matrices and Lie groups. Geom. Funct. Anal. **28**(1), 1–57 (2018)
3. An, J., Beresnevich, V., Velani, S.: Badly approximable points on planar curves and winning. Adv. Math. **324**, 148–202 (2018)
4. Badziahin, D., Schleischitz, J.: An improved bound in Wirsing's problem. https://arxiv.org/abs/1912.09013 (2019)
5. Badziahin, D., Velani, S.: Badly approximable points on planar curves and a problem of Davenport. Math. Ann. **359**(3–4), 969–1023 (2014)

6. Beresnevich, V.: A Groshev type theorem for convergence on manifolds. Acta Math. Hungar. **94**(1–2), 99–130 (2002)
7. Beresnevich, V.: Badly approximable points on manifolds. Invent. Math. **202**(3), 1199–1240 (2015)
8. Beresnevich, V., Bernik, V.: On a metrical theorem of W. Schmidt. Acta Arith. **75**(3), 219–233 (1996)
9. Beresnevich, V., Velani, S.: A note on zero-one laws in metrical Diophantine approximation. Acta Arith. **133**(4), 363–374 (2008)
10. Beresnevich, V., Velani, S.: Classical metric Diophantine approximation revisited: the Khintchine-Groshev theorem. Int. Math. Res. Not. IMRN **2010**(1), 69–86 (2010)
11. Beresnevich, V.V., Bernik, V.I., Kleinbock, D.Y., Margulis, G.A.: Metric Diophantine approximation: the Khintchine-Groshev theorem for nondegenerate manifolds. Mosc. Math. J. **2**(2), 203–225 (2002). Dedicated to Yuri I. Manin on the occasion of his 65th birthday
12. Beresnevich, V., Dickinson, D., Velani, S.: Measure theoretic laws for lim sup sets. Mem. Am. Math. Soc. **179**(846), x+91 (2006)
13. Beresnevich, V., Kleinbock, D., Margulis, G.: Non-planarity and metric Diophantine approximation for systems of linear forms. J. Théor. Nombres Bordeaux **27**(1), 1–31 (2015)
14. Beresnevich, V., Nesharim, E., Velani, S., Yang, L.: Schmidt's conjecture and badly approximable matrices. In preparation
15. Beresnevich, V., Nesharim, E., Yang, L.: Winning property of badly approximable points on curves. https://arxiv.org/abs/2005.02128 (2020)
16. Beresnevich, V., Ramírez, F., Velani, S.: Metric Diophantine approximation: aspects of recent work. In: Dynamics and Analytic Number Theory, vol. 437. London Math. Soc. Lecture Note Ser., pp. 1–95. Cambridge Univ. Press, Cambridge (2016)
17. Bernik, V.I., Dodson, M.M.: Metric Diophantine Approximation on Manifolds, vol. 137. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1999)
18. Bernik, V., Kleinbock, D., Margulis, G.A.: Khintchine-type theorems on manifolds: the convergence case for standard and multiplicative versions. Int. Math. Res. Notices **2001**(9), 453–486 (2001)
19. Broderick, R., Fishman, L., Kleinbock, D., Reich, A., Weiss, B.: The set of badly approximable vectors is strongly C^1 incompressible. Math. Proc. Camb. Philos. Soc. **153**(2), 319–339 (2012)
20. Cadambe, V.R., Jafar, S.A.: Interference alignment and degrees of freedom of the K-user interference channel. IEEE Trans. Inform. Theory **54**(8), 3425–3441 (2008)
21. Cheung, Y.: Hausdorff dimension of the set of singular pairs. Ann. Math. (2) **173**(1), 127–167 (2011)
22. Cheung, Y., Chevallier, N.: Hausdorff dimension of singular vectors. Duke Math. J. **165**(12), 2273–2329 (2016)
23. Dani, S.G.: Divergent trajectories of flows on homogeneous spaces and diophantine approximation. Journal für die reine und angewandte Mathematik **1985**(359), 55–89 (1985)
24. Dani, S.G.: On badly approximable numbers, Schmidt games and bounded orbits of flows. In: Number Theory and Dynamical Systems (York, 1987), vol. 134 London Math. Soc. Lecture Note Ser., pp. 69–86. Cambridge Univ. Press, Cambridge (1989)
25. Das, T., Fishman, L., Simmons, D., Urbański, M.: A variational principle in the parametric geometry of numbers (2019). arXiv:1901.06602
26. Davenport, H.: A note on Diophantine approximation. II. Mathematika **11**, 50–58 (1964)
27. Davenport, H., Schmidt, W.M.: Dirichlet's theorem on diophantine approximation. II. Acta Arith. **16**, 413–424 (1969/70)
28. Davenport, H., Schmidt, W.M.: Dirichlet's theorem on diophantine approximation. In: Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), pp. 113–132. Academic Press, London (1970)
29. Duffin, R.J., Schaeffer, A.C.: Khintchine's problem in metric Diophantine approximation. Duke Math. J. **8**, 243–255 (1941)
30. Falconer, K.: Fractal Geometry. Wiley, Chichester (1990). Mathematical foundations and applications

31. Ganguly, A., Ghosh, A.: Quantitative Diophantine approximation on affine subspaces. Math. Z. **292**(3–4), 923–935 (2019)
32. Ghosh, A.: A Khintchine-type theorem for hyperplanes. J. Lond. Math. Soc. (2) **72**(2), 293–304 (2005)
33. Ghosh, A.: Diophantine exponents and the Khintchine Groshev theorem. Monatsh. Math. **163**(3), 281–299 (2011)
34. Ghosh, A.: A Khintchine-Groshev theorem for affine hyperplanes. Int. J. Number Theory **7**(4), 1045–1064 (2011)
35. Ghosh, A., Royals, R.: An extension of the Khinchin-Groshev theorem. Acta Arith. **167**(1), 1–17 (2015)
36. Groshev, A.: A theorem on a system of linear forms. Dokl. Akad. Nauk SSSR **19**, 151–152 (1938)
37. Jafar, S.A.: Interference Alignment—A New Look at Signal Dimensions in a Communication Network. Now Publishers (2011)
38. Jafar, S.A., Shamai, S.: Degrees of freedom region of the MIMO X channel. IEEE Trans. Inform. Theory **54**(1), 151–170 (2008)
39. Kadyrov, S., Kleinbock, D., Lindenstrauss, E., Margulis, G.A.: Singular systems of linear forms and non-escape of mass in the space of lattices. J. Anal. Math. **133**, 253–277 (2017)
40. Khintchine, A.: Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen. Math. Ann. **92**(1–2), 115–125 (1924)
41. Kleinbock, D.: Extremal subspaces and their submanifolds. Geom. Funct. Anal. **13**(2), 437–466 (2003)
42. Kleinbock, D.Y., Margulis, G.A.: Flows on homogeneous spaces and Diophantine approximation on manifolds. Ann. Math. (2) **148**(1), 339–360 (1998)
43. Kleinbock, D., Wadleigh, N.: An inhomogeneous Dirichlet theorem via shrinking targets. Compos. Math. **155**(7), 1402–1423 (2019)
44. Kleinbock, D., Weiss, B.: Dirichlet's theorem on Diophantine approximation and homogeneous flows. J. Mod. Dyn. **2**(1), 43–62 (2008)
45. Kleinbock, D.Y., Margulis, G.A., Wang, J.: Metric Diophantine approximation for systems of linear forms via dynamics. Int. J. Number Theory **6**(5), 1139–1168 (2010)
46. Kleinbock, D., Shah, N., Starkov, A.: Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory. In: Handbook of Dynamical Systems, Vol. 1A, pp. 813–930. North-Holland, Amsterdam (2002)
47. Koukoulopoulos, D., Maynard, J.: On the Duffin-Schaeffer conjecture. Ann. Math. (2) (2019, to appear). arXiv:1907.04593
48. Maddah-Ali, M.A., Motahari, A.S., Khandani, A.K.: Communication over MIMO X channels: interference alignment, decomposition, and performance analysis. IEEE Trans. Inform. Theory **54**(8), 3457–3470 (2008)
49. Mahboubi, S.H., Motahari, A.S., Khandani, A.K.: Layered interference alignment: achieving the total DOF of MIMO X-channels. In: 2010 IEEE International Symposium on Information Theory, pp. 355–359. IEEE (2010)
50. Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces, vol. 44. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1995). Fractals and rectifiability
51. Motahari, A.S., Gharan, S.O., Khandani, A.K.: Real Interference Alignment with Real Numbers (2009). arXiv:0908.1208
52. Motahari, A.S., Oveis-Gharan, S., Maddah-Ali, M.-A., Khandani, A.K.: Real interference alignment: exploiting the potential of single antenna systems. IEEE Trans. Inform. Theory **60**(8), 4799–4810 (2014)
53. Niesen, U., Maddah-Ali, M.A.: Interference alignment: From degrees of freedom to constant-gap capacity approximations. IEEE Trans. Inf. Theory **59**(8), 4855–4888 (2013)
54. Ordentlich, O., Erez, U., Nazer, B.: The approximate sum capacity of the symmetric Gaussian K-user interference channel. IEEE Trans. Inform. Theory **60**(6), 3450–3482 (2014)
55. Perron, O.: Über diophantische approximationen. Math. Ann. **83**(1–2), 77–84 (1921)

56. Schmidt, W.: A metrical theorem in diophantine approximation. Can. J. Math. **12**, 619–631 (1960)
57. Schmidt, W.M.: On badly approximable numbers and certain games. Trans. Am. Math. Soc. **123**, 178–199 (1966)
58. Schmidt, W.M.: Badly approximable systems of linear forms. J. Number Theory **1**, 139–154 (1969)
59. Schmidt, W.M.: Diophantine Approximation. Springer, Berlin and New York (1980)
60. Schmidt, W.M.: Diophantine Approximation. Springer Science & Business Media (1996)
61. Shah, N.A.: Equidistribution of expanding translates of curves and Dirichlet's theorem on Diophantine approximation. Invent. Math. **177**(3), 509–532 (2009)
62. Shannon, C.E.: Communication in the presence of noise. Proc. I.R.E. **37**, 10–21 (1949)
63. Sprindžuk, V.G.: Metric Theory of Diophantine Approximations. V. H. Winston & Sons, Washington, D.C.; A Halsted Press Book, Wiley, New York, Toronto, London (1979). Translated from the Russian and edited by Richard A. Silverman, With a foreword by Donald J. Newman, Scripta Series in Mathematics
64. Yang, L.: Badly approximable points on manifolds and unipotent orbits in homogeneous spaces. Geom. Funct. Anal. **29**(4), 1194–1234 (2019)

Chapter 2
Characterizing the Performance of Wireless Communication Architectures via Basic Diophantine Approximation Bounds

Bobak Nazer and Or Ordentlich

Abstract Consider a wireless network where several users are transmitting simultaneously. Each receiver observes a linear combination of the transmitted signals, corrupted by random noise, and attempts to recover the codewords sent by one or more of the users. Within the context of network information theory, it is of interest to determine the maximum possible data rates as well as efficient strategies that operate at these rates. One promising recent direction has shown that if the users utilize a lattice-based strategy, then a receiver can recover an integer-linear combination of the codewords at a rate that depends on how well the real-valued channel gains can be approximated by integers. In other words, the performance of this lattice-based strategy is closely linked to a basic question in Diophantine approximation. This chapter provides an overview of the key findings in this emerging area, starting from first principles, and expanding towards state-of-the art results and open questions, so that it is accessible to researchers with either an information theory or Diophantine approximation background.

2.1 Introduction

Consider multiple transmitters and receivers that communicate with each other across a shared wireless channel. The two main challenges to establishing reliable communication between users are the noise introduced by the channel and the interference between simultaneously transmitted signals. Over the past few decades, the field of network information theory has striven to determine the fundamental

B. Nazer
Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA
e-mail: bobak@bu.edu

O. Ordentlich (✉)
The Rachel and Selim Benin School of Computer Science and Engineering,
Hebrew University of Jerusalem, Jerusalem, Israel
e-mail: or.ordentlich@mail.huji.ac.il

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*,
Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_2

limits of reliable communication over multi-user channels as well as architectures that can approach these limits in practice [11, 14, 34].

In this chapter, we discuss recent developments in network information theory based on the use of lattice codebooks, i.e., codebooks that are a subset of a lattice over $\mathbb{R}^n$ [36]. The inherent linearity of these codebooks is appealing for two reasons. First, linearity lends itself to more efficient encoding and decoding algorithms. Second, since lattices are closed under integer-linear combinations, it is possible for a receiver to directly decode an integer-linear combination of transmitted codewords (without first recovering the individual codewords) [26]. This phenomenon can be used as a building block for communication strategies that operate beyond the performance available for classical coding schemes.

In general, the performance of these lattice-based strategies is determined by how closely the channel coefficients can be *approximated by* integer coefficients. For any particular choice of channel coefficients, we can identify the optimal integer coefficients, and the resulting performance. However, it is often of interest to have universal bounds that do not depend on the specific realization of the channel. As we will demonstrate, classical and modern results from Diophantine approximation can be used to establish such bounds.

Overall, this chapter attempts to provide a unified view of recent results that connect the performance of the "compute-and-forward" strategy of recovering an integer-linear combination to Diophantine approximation bounds. We also highlight scenarios where novel applications of Diophantine approximation techniques may lead to new results in network information theory.

2.1.1 Single-User Gaussian Channels

Consider the following channel model for time $t \in \{1, 2, \ldots, T\}$:

$$y[t] = x[t] + z[t] \tag{2.1}$$

where

- $y[t] \in \mathbb{R}$ represents the channel output at the receiver at time t,
- $x[t] \in \mathbb{R}$ is the channel input of the transmitter at time t,
- and $z[t] \in \mathbb{R}$ is the noise at time t, which is assumed to be Gaussian, $z[t] \sim \mathcal{N}(0, 1)$, and generated independently for each time t.

Our goal is for the transmitter to reliably send information to the receiver at the highest possible data rates. To this end, the channel may be used during T time slots, which is often referred to as the *blocklength* of the communication scheme. The *communication rate* $R \geq 0$ is defined as the average number of bits that it transmits per time slot. One practical consideration is that the transmitter has a maximum power level that it can sustain during its transmission. This is modeled in

the definition below via the *power constraint* $P \geq 0$. Let $\|\cdot\|$ denote the Euclidean norm.

Definition 2.1 (Code) A $(2^{TR}, T, P)$ code for the channel (2.1) consists of

- a message set $\{1, 2, \ldots, 2^{TR}\}$,
- an encoder that assigns a T-dimensional vector $\mathbf{x}(m) \in \mathbb{R}^T$ to each message $m \in \{1, 2, \ldots, 2^{TR}\}$. The encoder is subject to a power constraint $P > 0$, which dictates that $\|\mathbf{x}(m)\|^2 \leq TP$ for all $m \in \{1, 2, \ldots, 2^{TR}\}$,
- and a decoder that assigns an estimate $\hat{m}$ of the transmitted message to each possible received sequence $[y[1]\ y[2]\ \cdots\ y[T]]$.

The message M is assumed to be uniformly distributed over $\{1, 2, \ldots, 2^{TR}\}$. The average error probability of a code is defined as

$$p_{\text{error}} = \mathscr{P}\big(\hat{M} \neq M\big). \tag{2.2}$$

Definition 2.2 (Achievable Rate) A rate R is said to be achievable over the channel (2.1) with power constraint P if, for any $\epsilon > 0$ and T large enough, there exists a $(2^{TR}, T, P)$ code with $p_{\text{error}} < \epsilon$.

Definition 2.3 (Capacity) The capacity of the channel (2.1) with power constraint P is the supremum of the set of all achievable rates.

The capacity of the Gaussian channel is due to Shannon [33].

Theorem 2.1 (Gaussian Capacity) *The capacity of the channel* (2.1) *with power constraint* P *is*

$$C = \frac{1}{2}\log(1 + P)\ . \tag{2.3}$$

The proof of Theorem 2.1 consists of two parts: a *converse* part where it is shown that if a $(2^{TR}, T, P)$ code with small error probability exists, then the rate R must satisfy $R \leq \frac{1}{2}\log(1 + P)$, and a *direct* part, where it is shown that there exists a sequence of codes $(2^{TR}, T, P)$, with growing T and vanishing error probability so long as $R < \frac{1}{2}\log(1 + P)$.

The main observation leading to the direct part is that, in high dimensions, the noise sequence lives inside a ball of radius $\sqrt{T(1+\delta)}$ for $\delta > 0$ with high probability. Thus, the coding task reduces to placing the centers of 2^{TR} balls of this radius inside a larger ball of radius $\sqrt{TP}$, with some small overlap between balls that corresponds to the small allowable error probability. Shannon's insight was that the existence of such a packing can be shown via the probabilistic method, i.e., by drawing the centers of the balls independently and uniformly within a ball of radius $\sqrt{TP}$. In this manner, the codewords are ensured to not violate the power constraint. Alternatively, we can draw the codewords i.i.d. according to a $\mathcal{N}(\mathbf{0}, P\mathbf{I})$ distribution.

A typical member of the i.i.d. code ensemble lacks structure, and thus the encoding and decoding operations require exponential complexity in T (i.e.,

they essentially correspond to lookup tables for all 2^{TR} codewords) and are not practically realizable. The field of coding theory has striven to develop families of codes with low encoding and decoding complexity and performance close to the capacity limit.

The art of coding for the AWGN channel is by now well-developed and low-complexity coding schemes operating near capacity, e.g., low-density parity-check (LDPC) codes [9, 16, 32], turbo codes [6], polar codes [3], etc., are known and implemented in various communication standards. A lot of these coding schemes are based on mapping a binary linear code, i.e., a subspace in $\mathbb{F}_2^T$, (or more generally, a p-ary linear code) to the Euclidean space. Consequently, the resulting code often has some linear structure, and can be thought of as a lattice code, as we define below.

A lattice Λ is a discrete subgroup of $\mathbb{R}^T$ that is closed under reflection and real addition. Formally, for any $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda$, we have that $-\boldsymbol{\lambda}_1, -\boldsymbol{\lambda}_2 \in \Lambda$ and $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 \in \Lambda$. Note that, by definition, the zero vector $\mathbf{0}$ is always a member of the lattice. Any lattice Λ in $\mathbb{R}^T$ is spanned by some $T \times T$ matrix $\mathbf{G}$ such that

$$\Lambda = \{\boldsymbol{\lambda} = \mathbf{G}\mathbf{q} : \mathbf{q} \in \mathbb{Z}^T\}.$$

We say that a lattice is full-rank if its spanning matrix $\mathbf{G}$ is full-rank.

Let $\mathcal{B}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^T : \|\mathbf{x}\| \le r\}$ be the T-dimensional, zero-centered, closed ball of radius $r > 0$. A lattice code is constructed by intersecting a base lattice Λ, with some shaping region $\mathcal{V} \subset \mathcal{B}(\mathbf{0}, \sqrt{TP})$, whose role is to enforce the power constraint. The rate of the lattice code $\mathcal{L} = \Lambda \cap \mathcal{V}$ is therefore $R = \frac{1}{T} \log |\Lambda \cap \mathcal{V}|$.

The main motivation for using lattice codes for the AWGN channel is to exploit the linear structure of Λ for simplified encoding and decoding algorithms. In particular, for the AWGN channel, the optimal decoder corresponds to finding the codeword with the smallest Euclidean distance from the channel output. When a lattice code $\mathcal{L} = \Lambda \cap \mathcal{V}$ is used, this can be approximated by applying the nearest neighbor lattice quantizer defined as

$$Q_\Lambda(\mathbf{y}) = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \|\mathbf{y} - \boldsymbol{\lambda}\|, \tag{2.4}$$

to the channel output, and returning the corresponding message if $Q_\Lambda(\mathbf{y}) \in \mathcal{V}$, or declaring an error otherwise.

The choice of the shaping region $\mathcal{V}$ should on the one hand result in a high rate, and on the other hand maintain much of the structure of the base lattice, such that there is a "convenient" mapping between the message set $\{1, 2, \ldots, 2^{RT}\}$ and the points in $\mathcal{L}$, and that a lattice decoder, which essentially ignores the shaping region, would still perform well. Erez and Zamir [15] showed that for any $P > 0$, there exists a base lattice Λ and a shaping region $\mathcal{V}$ (more precisely, a sequence in T of $\Lambda^{(T)}, \mathcal{V}^{(T)}$), such that the lattice code $\mathcal{L} = \Lambda \cap \mathcal{V}$ achieves the AWGN channel capacity under (a slight modification of) lattice decoding. In particular, they took $\mathcal{V}$ as the Voronoi region of a coarse lattice $\Lambda_c \subset \Lambda$.

For the point-to-point AWGN channel, the interest in lattice codes is motivated by the need to lower the complexity of encoding and decoding operations so as to render them practically feasible. For networks with multiple transmitters or receivers, lattice codes can also be used to approach the performance suggested by i.i.d. random codes. Interestingly, as we will explore below, lattice codes can also be used to derive lower bounds on multi-user capacity that cannot be established via i.i.d. ensembles.

2.2 Gaussian Multiple-Access Channel Model

We will focus on bounds for the Gaussian multiple-access channel (MAC), which is a canonical model for a wireless network where multiple transmitters simultaneously communicate with a single receiver. We assume that there are K users, each equipped with a single antenna, that wish to communicate with an N-antenna receiver for time $t \in \{1, 2, \ldots, T\}$, leading to the following model:

$$\mathbf{y}[t] = \sum_{k=1}^{K} \mathbf{h}_k x_k[t] + \mathbf{z}[t] \tag{2.5}$$

where

- $\mathbf{y}[t] \in \mathbb{R}^N$ represents the channel output at the receiver at time t,
- $x_k[t] \in \mathbb{R}$ is the channel input of the kth user at time t,
- $\mathbf{h}_k \in \mathbb{R}^N$ is the vector of channel gains from the kth user to the N antennas of the receiver,
- and $\mathbf{z}[t] \in \mathbb{R}^N$ is the noise vector at time t, which is assumed to be Gaussian, $\mathbf{z}[t] \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and generated independently for each time t.

It will be useful to express all of the channel gains together in matrix notation,

$$\mathbf{y}[t] = \mathbf{H}\mathbf{x}[t] + \mathbf{z}[t] \tag{2.6}$$

$$\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \cdots\ \mathbf{h}_K] \tag{2.7}$$

$$\mathbf{x}[t] = [x_1[t]\ x_2[t]\ \cdots\ x_K[t]]^{\mathsf{T}} \tag{2.8}$$

where the (n, k)th entry $h_{n,k}$ of $\mathbf{H}$ represents the channel gain from the kth user to the nth antenna.

Definition 2.4 A $(2^{TR_1}, \ldots, 2^{TR_K}, T, P)$ code for the channel (2.6) consists of

- K message sets $\{1, 2, \ldots, 2^{TR_k}\}$, $k = 1, \ldots, K$,
- K encoders, where encoder k assigns a T-dimensional vector $\mathbf{x}_k(m_k) \in \mathbb{R}^T$ to each message $m_k \in \{1, 2, \ldots, 2^{TR_k}\}$. All encoders are subject to a power

constraint $P > 0$, which dictates that $\|\mathbf{x}_k(m_k)\|^2 \leq TP$ for all $k = 1, \ldots, K$ and $m_k \in \{1, 2, \ldots, 2^{TR_k}\}$,
- and a decoder that assigns an estimate $(\hat{m}_1, \ldots, \hat{m}_K)$ of the transmitted messages to each possible received sequence $\mathbf{Y} = [\mathbf{y}[1]\, \mathbf{y}[2]\, \cdots\, \mathbf{y}[T]] \in \mathbb{R}^{N \times T}$.

In the sequel, it will be useful to compactly represent all time slots t together:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \tag{2.9}$$

where

$$\mathbf{Y} = [\mathbf{y}[1]\, \cdots\, \mathbf{y}[T]] \in \mathbb{R}^{N \times T} \tag{2.10}$$

$$\mathbf{X} = [\mathbf{x}[1]\, \cdots\, \mathbf{x}[T]] = [\mathbf{x}_1(m_1)\, \cdots\, \mathbf{x}_K(m_K)]^{\mathsf{T}} \in \mathbb{R}^{K \times T} \tag{2.11}$$

$$\mathbf{Z} = [\mathbf{z}[1]\, \cdots\, \mathbf{z}[T]] \in \mathbb{R}^{N \times T}. \tag{2.12}$$

The message M_k of the kth user is assumed to be uniformly distributed over $\{1, 2, \ldots, 2^{TR_k}\}$, and $M_1, \ldots, M_K$ are assumed to be mutually independent. The average error probability of a code is defined as

$$p_{\text{error}} = \mathscr{P}\left((\hat{M}_1, \ldots, \hat{M}_K) \neq (M_1, \ldots, M_K)\right). \tag{2.13}$$

Definition 2.5 (Achievable Rates) A rate tuple $(R_1, \ldots, R_K)$ is said to be achievable over the channel (2.6) with power constraint P if, for any $\epsilon > 0$ and T large enough, there exists a $(2^{TR_1}, \ldots, 2^{TR_k}, T, P)$ code with $p_{\text{error}} < \epsilon$.

Definition 2.6 (Capacity Region) The capacity region of the channel (2.6) with power constraint P is the closure of the set of all achievable rate tuples.

The Gaussian MAC (2.6) with power constraint P, is a special case of the family of discrete memoryless MACs, for which the capacity region is known, and can be expressed in closed form [2, 11, 22].

Theorem 2.2 (MAC Capacity Region) *The capacity region of the Gaussian MAC* (2.6) *with power constraint P is the set of all rates satisfying*

$$\sum_{k \in \mathcal{S}} R_k \leq \frac{1}{2} \log \det\left(\mathbf{I} + P\mathbf{H}_{\mathcal{S}}^{\mathsf{T}}\mathbf{H}_{\mathcal{S}}\right), \tag{2.14}$$

for all $\mathcal{S} = \{i_1, \ldots, i_{|\mathcal{S}|}\} \subset [K]$, where $\mathbf{H}_{\mathcal{S}} = [\mathbf{h}_{i_1}\, \cdots\, \mathbf{h}_{i_{|\mathcal{S}|}}]$.

As in the point-to-point AWGN case, the direct (achievability) part of Theorem 2.2 is established by drawing each user's codebook independently at random from an i.i.d. ensemble [14, §9.2.1]. Consequently, the proof does not lead to practical communication schemes for this channel.

Note also that unlike the point-to-point AWGN model, here the channel is characterized by a channel matrix $\mathbf{H}$. Thus, in general, different codes are needed for different channel matrices, even if $R_1, \ldots, R_K, T$ and P are fixed. In practical scenarios, $\mathbf{H}$ is seldom known in advance, and typically it is changing with time. Thus, a more natural approach is to design the encoders independently of $\mathbf{H}$, and to only adapt the decoder w.r.t. the actual channel matrix $\mathbf{H}$. Moreover, since capacity-approaching codes with low-complexity for the point-to-point AWGN channel exist, a very appealing approach is to manipulate the MAC output $\mathbf{Y}$ using signal processing, in order to induce parallel point-to-point channels from it.

The most natural, and widely used, example of such an approach is based on linear estimation. In particular, in order to decode $\mathbf{x}_k = \mathbf{x}_k(M_k)$, we can first set $\tilde{\mathbf{y}}_k^{\mathsf{T}} = \mathbf{b}_k^{\mathsf{T}}\mathbf{Y}$, where the vector $\mathbf{b}_k \in \mathbb{R}^N$ is selected to minimize $\sigma_k^2 = \mathbb{E}\|\mathbf{x}_k - \tilde{\mathbf{y}}_k\|^2$. Now, the channel from $\mathbf{x}_k$ to $\tilde{\mathbf{y}}_k$ can be thought of as a point-to-point AWGN channel with noise variance σ_k^2. Thus, if $\mathbf{x}_k$ is encoded via a "good" code for the AWGN channel, we can apply the corresponding decoder, and decode $\mathbf{x}_k$ from $\tilde{\mathbf{y}}_k$ with small error probability, if $R_k < \frac{1}{2}\log\left(\frac{P}{\sigma_k^2}\right)$.[1] We refer to the above communication scheme as a *linear equalization* scheme, since roughly speaking, the vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_K\}$ attempt to equalize the channel matrix $\mathbf{H} \in \mathbb{R}^{N\times K}$ to $\mathbf{I}_K$, the identity matrix of size K. The achievable rates for linear equalization are characterized in the following theorem (see, e.g., [17]).

Theorem 2.3 (Performance of Linear Equalization) *Let* $\mathbf{\Sigma} = \left(P^{-1}\mathbf{I}_K + \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}$ *and let* $\sigma_k^2 = \mathbf{\Sigma}_{kk}$. *Then, any rate tuple* $(R_1, \ldots, R_K)$ *that satisfies*

$$R_k < \frac{1}{2}\log\left(\frac{P}{\sigma_k^2}\right) \tag{2.15}$$

is achievable over the Gaussian MAC (2.6) *with power constraint* P, *under linear equalization.*

2.3 Exploiting Linear Structure

As discussed above, many of the coding strategies employed in practice can be viewed as lattice codes. It turns out that the linear structure of these lattice code ensembles opens up a new equalization possibility: rather than decoding each codeword individually, we can directly decode any integer-linear combination of codewords. Specifically, since the lattice is closed under addition, any integer-linear

[1] The $1+$ term from the capacity expression $C = \frac{1}{2}\log(1+P)$ is lost to compensate for the dependence between $\mathbf{x}_k$ and $\mathbf{e}_k = \mathbf{x}_k - \tilde{\mathbf{y}}_k$. However, if we set $\mathbf{H} = 1$ to model a point-to-point AWGN channel, we find that (2.15) is equal to the AWGN capacity $1/2\log(1+P)$ as desired. See [36], [10, Lemma 2] for more details.

combination of lattice points is itself a lattice point, and thus afforded the same protection against noise as the original codewords.

To illustrate the potential gains of this approach, consider the following example from [37].

Example 2.1 There are $K = 2$ users and $N = 2$ receive antennas. The channel matrix is integer-valued

$$\mathbf{H} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.16}$$

From (2.6), the receiver observes

$$\mathbf{Y} = \begin{bmatrix} 2\mathbf{x}_1^{\mathsf{T}} + \mathbf{x}_2^{\mathsf{T}} \\ \mathbf{x}_1^{\mathsf{T}} + \mathbf{x}_2^{\mathsf{T}} \end{bmatrix} + \mathbf{Z} \,. \tag{2.17}$$

For large P, the linear equalizer roughly reduces to inverting the matrix $\mathbf{H}$, i.e., $\mathbf{b}_1^{\mathsf{T}} = [1 \quad -1]$ and $\mathbf{b}_2^{\mathsf{T}} = [-1 \;\; 2]$, which yields the effective channel outputs

$$\tilde{\mathbf{y}}_1 = \mathbf{x}_1 + \mathbf{b}_1^{\mathsf{T}}\mathbf{Z} \tag{2.18}$$

$$\tilde{\mathbf{y}}_2 = \mathbf{x}_2 + \mathbf{b}_2^{\mathsf{T}}\mathbf{Z} \,, \tag{2.19}$$

and rates

$$R_1 = \frac{1}{2}\log\left(1 + \frac{P}{2}\right) \approx \frac{1}{2}\log\left(\frac{P}{2}\right) \tag{2.20}$$

$$R_2 = \frac{1}{2}\log\left(1 + \frac{P}{5}\right) \approx \frac{1}{2}\log\left(\frac{P}{5}\right) \tag{2.21}$$

where the approximations become tight as P increases. On the other hand, if both encoders employ the same lattice code, then the integer-linear combinations $2\mathbf{x}_1 + \mathbf{x}_2$ and $\mathbf{x}_1 + \mathbf{x}_2$ are themselves codewords and can be decoded at rates

$$R_1 = \frac{1}{2}\log\left(\frac{1}{5} + P\right) \approx \frac{1}{2}\log(P) \tag{2.22}$$

$$R_2 = \frac{1}{2}\log\left(\frac{1}{2} + P\right) \approx \frac{1}{2}\log(P) \,. \tag{2.23}$$

as will be shown by Theorem 2.4. After removing the noise, we can solve for the desired codewords $\mathbf{x}_1$ and $\mathbf{x}_2$. The high-level intuition is that this strategy offers an advantage since it does not enhance the noise during the linear equalization step.

The example above demonstrates that there can be performance advantages to recovering integer-linear combinations as an intermediate step towards decoding

the transmitted messages. We now turn to the general case where the channel coefficients are not necessarily integer-valued. As we will see, it is still possible to decode integer-linear combinations of codewords, and the performance is determined by how closely the integer coefficients approximate the real-valued channel gains. First, we need to be a bit more precise about what we mean by recovering linear combinations.

Definition 2.7 A $(2^{TR_1}, \ldots, 2^{TR_K}, T, P)$ computation code for the channel (2.6) consists of

- K message sets $\{1, 2, \ldots, 2^{TR_k}\},\ k = 1, \ldots, K$,
- K encoders, where encoder k assigns a *unique* T-dimensional vector $\mathbf{x}_k(m_k) \in \mathbb{R}^T$ to each message $m_k \in \{1, 2, \ldots, 2^{TR_k}\}$. All encoders are subject to a power constraint $P > 0$, which dictates that $\|\mathbf{x}_k(m_k)\|^2 \leq TP$ for all $k = 1, \ldots, K$ and $m_k \in \{1, 2, \ldots, 2^{TR_k}\}$,
- and, for a chosen integer vector $\mathbf{a} = [a_1 \ \cdots \ a_K]^\mathsf{T} \in \mathbb{Z}^K$, a decoder that assigns an estimate $\hat{\mathbf{v}}$ of the integer-linear combination of the codewords $\mathbf{v} = \sum_{k=1}^K a_k \mathbf{x}_k(m_k)$ to each possible received sequence $\mathbf{Y} \in \mathbb{R}^{N \times T}$.

For a given channel matrix $\mathbf{H} \in \mathbb{R}^{N \times K}$ and integer vector $\mathbf{a} \in \mathbb{Z}^K$, the average error probability of a computation code is defined as

$$p_{\text{error}} = \mathscr{P}(\hat{\mathbf{v}} \neq \mathbf{v}) \ . \tag{2.24}$$

The rates at which it is possible to recover an integer-combination depends on both the vector of integer coefficients $\mathbf{a} \in \mathbb{Z}^K$ and the channel matrix $\mathbf{H} \in \mathbb{R}^{N \times K}$ as well as the power P. The definition below is useful for concisely describing the computation rate.

Definition 2.8 The computation rate function $R(\mathbf{H}, \mathbf{a}, P)$ is achievable over the channel (2.6) if, for any $\epsilon > 0$ and T large enough, there exists a $(2^{TR_1}, \ldots, 2^{TR_K}, T, P)$ computation code such that, for any $\mathbf{H} \in \mathbb{R}^{N \times K}$ and $\mathbf{a} \in \mathbb{Z}^K$, we have that $p_{\text{error}} < \epsilon$ if

$$R_k < R(\mathbf{H}, \mathbf{a}, P) \quad \forall k \ . \tag{2.25}$$

According to the definition above, the receiver is free to recover *any* integer-linear combination of codewords for which (2.25) is satisfied. That is, the transmitters are completely agnostic as to the choice of the integer coefficients as well as the channel matrix $\mathbf{H}$, i.e., a codeword depends only on the selected message.

Remark 2.1 For the sake of conciseness, we have focused on the symmetric case $R_1 = \cdots = R_K$. Specifically, for a given $\mathbf{H}$ and $\mathbf{a}$, all rates $R_1, \ldots, R_K$ must be below the scalar rate threshold given by $R(\mathbf{H}, \mathbf{a}, P)$, which can be thought of as setting all rates equal to one another. More generally, we might expect to describe the attainable performance by a region. See [27] for relevant definitions and theorems.

Example 2.2 We can interpret a capacity-achieving multiple-access code as a computation code in the following sense. A multiple-access code allows the receiver to decode all of the transmitted messages, from which it can reconstruct the transmitted codewords, and then any integer-linear combination of interest. It follows from Theorem 2.2 that the computation rate described by the function

$$R(\mathbf{H}, \mathbf{a}, P) = \min_{\mathcal{S} \subset [K]} \frac{1}{2|\mathcal{S}|} \log \det \left(\mathbf{I} + P \mathbf{H}_{\mathcal{S}}^{\mathsf{T}} \mathbf{H}_{\mathcal{S}} \right), \tag{2.26}$$

which has no dependence on the integer vector $\mathbf{a}$, is achievable.

Intuitively, we expect that, for a more interesting computation code, $R(\mathbf{H}, \mathbf{a}, P)$ should depend on $\mathbf{a}$ and should be larger than (2.26) whenever $\mathbf{H}$ and $\mathbf{a}$ are "close." Our approach is for each encoder to employ the same lattice codebook $\mathcal{L} = \Lambda \cap \mathcal{V}$. Since all codewords can be viewed as elements of the lattice, $\mathbf{x}_k(m_k) \in \Lambda$, then we have that integer-linear combinations are elements of the lattice as well $\sum_{k=1}^{K} a_k \mathbf{x}_k(m_k) \in \Lambda$. The key idea is that, if the lattice codebook is designed to tolerate noise up to a certain variance, then we can recover any integer-linear combinations for which the effective noise variance is below this level. Overall, the job of the each encoder is simple: it maps its message m_k into the corresponding lattice codeword $\mathbf{x}_k(m_k)$, and transmits it, paying no attention to nature's choice of the channel matrix $\mathbf{H}$ or the receiver's choice of the integer vector $\mathbf{a}$.

At the receiver, our goal is to recover $\mathbf{v} = \sum_{k=1}^{K} a_k \mathbf{x}_k(m_k) = \mathbf{a}^{\mathsf{T}} \mathbf{X}$ from $\mathbf{Y}$. We are free to select the integer vector $\mathbf{a}$ based on our knowledge of $\mathbf{H}$. As a first step, we use an equalization vector $\mathbf{b} \in \mathbb{R}^N$ to create the effective channel

$$\tilde{\mathbf{y}}^{\mathsf{T}} = \mathbf{b}^{\mathsf{T}} \mathbf{Y} \tag{2.27}$$

$$= \mathbf{b}^{\mathsf{T}} \mathbf{H} \mathbf{X} + \mathbf{b}^{\mathsf{T}} \mathbf{Z} \tag{2.28}$$

$$= \mathbf{a}^{\mathsf{T}} \mathbf{X} + \mathbf{z}_{\text{eff}}^{\mathsf{T}} \tag{2.29}$$

where

$$\mathbf{z}_{\text{eff}}^{\mathsf{T}} = (\mathbf{b}^{\mathsf{T}} \mathbf{H} - \mathbf{a}^{\mathsf{T}}) \mathbf{X} + \mathbf{b}^{\mathsf{T}} \mathbf{Z} \,. \tag{2.30}$$

It can be shown that the effective noise variance is

$$\frac{1}{n} \mathbb{E} \|\mathbf{z}_{\text{eff}}\|^2 = \|\mathbf{b}\|^2 + P \|\mathbf{H}^{\mathsf{T}} \mathbf{b} - \mathbf{a}\|^2 \,. \tag{2.31}$$

This variance is minimized by taking $\tilde{\mathbf{y}}^{\mathsf{T}}$ to be the linear least-squares error (LLSE) estimator of the integer-linear combination $\mathbf{a}^{\mathsf{T}} \mathbf{X}$ from the channel output $\mathbf{Y}$, which corresponds to setting the equalization vector to

$$\mathbf{b} = P \mathbf{a}^{\mathsf{T}} \mathbf{H}^{\mathsf{T}} \left(\mathbf{I} + P \mathbf{H} \mathbf{H}^{\mathsf{T}} \right)^{-1} . \tag{2.32}$$

We define the resulting effective noise variance to be

$$\sigma^2_{\text{eff}}(\mathbf{H}, \mathbf{a}, P) = \mathbf{a}^{\mathsf{T}}\left(P^{-1}\mathbf{I} + \mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}\mathbf{a}\,. \tag{2.33}$$

After this equalization step, the receiver uses a lattice quantizer to obtain an estimate of the integer-linear combination $\hat{\mathbf{v}} = Q_\Lambda(\tilde{\mathbf{y}})$. For a good lattice code, the receiver can successfully decode if $R < \log(P/\sigma^2_{\text{eff}}(\mathbf{H}, \mathbf{a}, P))$. Overall, this strategy leads to the following theorem [26, 27, 37].

Theorem 2.4 (Computation Rate Region) *The computation rate region described by the function*

$$R(\mathbf{H}, \mathbf{a}, P) = \frac{1}{2}\log\left(\frac{P}{\sigma^2_{\text{eff}}(\mathbf{H}, \mathbf{a}, P)}\right) \tag{2.34}$$

$$= -\frac{1}{2}\log\left(\mathbf{a}^{\mathsf{T}}\left(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}\mathbf{a}\right) \tag{2.35}$$

is achievable over the channel (2.6) *with power constraint* P.

Note that the matrix $\left(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}$ is symmetric and positive definite, and therefore admits a Cholesky decomposition

$$\left(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1} = \mathbf{L}\mathbf{L}^{\mathsf{T}}, \tag{2.36}$$

where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries. With this notation, we can express the computation rate function as

$$R(\mathbf{H}, \mathbf{a}, P) = -\frac{1}{2}\log\|\mathbf{L}^{\mathsf{T}}\mathbf{a}\|^2. \tag{2.37}$$

In many cases, the receiver is interested in decoding L linearly independent linear combinations, but does not care about the particular coefficients. Therefore, we can use the L linearly independent integer vectors $\mathbf{a}_1, \ldots, \mathbf{a}_L$ that yield the highest computation rates $R(\mathbf{H}, \mathbf{a}_1, P) \geq \cdots \geq R(\mathbf{H}, \mathbf{a}_L, P)$. Accordingly, we define the kth computation rate $R_{\text{comp},k}(\mathbf{H}, P) \triangleq R(\mathbf{H}, \mathbf{a}_k, P)$ to be the rate associated with decoding the kth best integer coefficient vector $\mathbf{a}_k$ that is linearly independent of $\{\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}\}$.

In some applications, it suffices to recover $L < K$ linear combinations at a single receiver. For instance, K receivers could each decode one (linearly independent) integer-linear combination and forward it to a single node that solves for the transmitted codewords. In other cases, it will be of interest to recover K (linearly independent) integer-linear combinations at a single receiver. Overall, if we wish to recover L linear combinations, then the rate of the lattice codebook must be smaller than $R_{\text{comp},L}(\mathbf{H}, P)$.

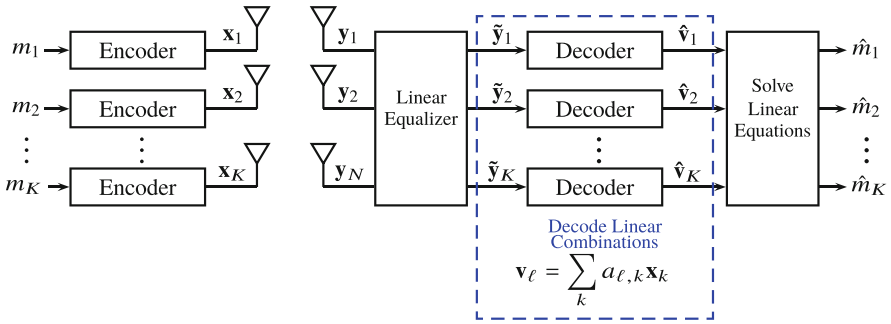


Fig. 2.1 The integer-forcing receiver architecture. The receiver employs linear equalization followed by parallel decoding to recover K linear combinations of the transmitted codewords. It can then solve for the individual codewords (and thus the original messages)

As a concrete example, consider the *integer-forcing* architecture for a Gaussian MAC as illustrated in Fig. 2.1. Each of the K users employs the same lattice codebook. Similarly to the strategy used to establish Theorem 2.3, the receiver applies a linear equalizer $\mathbf{B}$ to its observation $\mathbf{Y}$ to obtain the effective channel output $\tilde{\mathbf{Y}} = \mathbf{BY}$. In Theorem 2.3, this equalization step is used to induce an effective channel that is close to the identity matrix, which facilitates the parallel decoding of the K transmitter codewords. For the integer-forcing receiver, the equalization is instead used to create any effective integer-valued, full-rank channel matrix $\mathbf{A}$. Parallel decoding can then be used to reliably decode the integer-linear combinations $\mathbf{AX}$, which can then be solved for the desired individual messages.

2.4 Universal Bounds via Successive Minima

In this section, we derive bounds on the computation rates $\{R_{\mathrm{comp},k}(\mathbf{H}, P)\}_{k=1}^{K}$ using known results about the successive minima of a lattice. These bounds can be used to approximate computation rates without first finding the optimal integer coefficients.

Definition 2.9 (Successive Minima) Let $\Lambda(\mathbf{G})$ be the lattice spanned by the full-rank matrix $\mathbf{G} \in \mathbb{R}^{K\times K}$. For $k = 1, \ldots, K$, we define the kth successive minimum as

$$\lambda_k(\mathbf{G}) \triangleq \inf\left\{r \ : \ \dim\left(\mathrm{span}\left(\Lambda(\mathbf{G}) \bigcap \mathcal{B}(\mathbf{0}, r)\right)\right) \geq k\right\}.$$

In words, the kth successive minimum of a lattice is the minimal radius of a ball centered around $\mathbf{0}$ that contains k linearly independent lattice points.

Let $\mathbf{L}$ be the matrix defined in (2.36), $\Lambda(\mathbf{L}^{\mathsf{T}})$ be the lattice generated by $\mathbf{L}^{\mathsf{T}}$, and $\lambda_k(\mathbf{L}^{\mathsf{T}})$ its kth successive minimum. By (2.37) and the definition of $R_{\mathrm{comp},k}(\mathbf{H}, P)$,

we have that

$$R_{\text{comp},k}(\mathbf{H}, P) = -\log \lambda_k(\mathbf{L}^{\mathsf{T}}). \tag{2.38}$$

It follows that any upper bound on $\lambda_k(\mathbf{L}^{\mathsf{T}})$ immediately translates to a lower bound on $R_{\text{comp},k}(\mathbf{H}, P)$. For $k = 1$, such bounds are given by Minkowski's first theorem. Let $V_K = \text{Vol}(\mathcal{B}(\mathbf{0}, 1))$ be the volume of the K-dimensional unit ball. While an explicit expression

$$V_K = \frac{\pi^{K/2}}{\Gamma(K/2+1)},$$

exists, we will be content with the estimate $V_K \geq 2^K K^{-K/2}$, which is obtained by noting that $\mathcal{B}(\mathbf{0}, 1)$ contains a cube with side $2/\sqrt{K}$ [24].

Theorem 2.5 (Minkowski's First Theorem) *For any full-rank* $\mathbf{G}$*,*

$$\lambda_1(\mathbf{G}) \leq 2\left(\frac{\left|\det(\mathbf{G})\right|}{V_K}\right)^{\frac{1}{K}} \leq \sqrt{K}\left|\det(\mathbf{G})\right|^{\frac{1}{K}}. \tag{2.39}$$

From Minkowki's first theorem we immediately obtain a lower bound on $R_{\text{comp},1}(\mathbf{H}, P)$, given as a simple function of $\mathbf{H}$, K, and P.

Theorem 2.6

$$R_{\text{comp},1}(\mathbf{H}, P) \geq \frac{1}{2K}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}) - \frac{1}{2}\log K. \tag{2.40}$$

Proof From (2.38) and Theorem 2.5 we have that

$$\begin{aligned} R_{\text{comp},1}(\mathbf{H}, P) &\geq -\frac{1}{K}\log\left|\det(\mathbf{L}^T)\right| - \frac{1}{2}\log K \\ &= -\frac{1}{2K}\log\left|\det(\mathbf{L}\mathbf{L}^T)\right| - \frac{1}{2}\log K \\ &= \frac{1}{2K}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}) - \frac{1}{2}\log K, \end{aligned}$$

where the last equality follows from (2.36). □

Theorem 2.2 implies that for any rate-tuple $(R_1, \ldots, R_K)$ that is achievable over the channel (2.6) with power constraint P, we must have

$$\sum_{k=1}^{K} R_k \leq \frac{1}{2}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}). \tag{2.41}$$

The expression on the right hand side of (2.41) is referred to as the *sum-capacity* of the channel.[2] Consequently, if the symmetric rate-tuple $(R, \ldots, R)$ is achievable, then we must have that

$$R \leq \frac{1}{2K} \log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}), \tag{2.42}$$

where the expression in the right hand side of (2.42) is an upper bound on the *symmetric capacity* of the channel. In light of this, the interpretation of Theorem 2.6 is that $R_{\mathrm{comp},1}(\mathbf{H}, P)$ cannot be much smaller than the symmetric capacity, for all $\mathbf{H}$ and P.

Next, we turn to estimating $\sum_{k=1}^{K} R_{\mathrm{comp},k}(\mathbf{H}, P)$. By (2.38), we have

$$\begin{aligned}\sum_{k=1}^{K} R_{\mathrm{comp},k}(\mathbf{H}, P) &= -\sum_{k=1}^{K} \log \lambda_k(\mathbf{L}^{\mathsf{T}}) \\ &= -\log\left(\prod_{k=1}^{K} \lambda_k(\mathbf{L}^{\mathsf{T}})\right).\end{aligned} \tag{2.43}$$

Our goal is therefore to estimate the product of successive minima. Let $\mathbf{a}_1, \ldots, \mathbf{a}_K \in \mathbb{Z}^K$ be linearly independent vectors such that $\lambda_k(\mathbf{L}^{\mathsf{T}}) = \|\mathbf{L}^{\mathsf{T}}\mathbf{a}_k\|$, and let $\mathbf{A} = [\mathbf{a}_1|\cdots|\mathbf{a}_K] \in \mathbb{Z}^{K\times K}$. Since $|\det(\mathbf{A})| \geq 1$, we have

$$\left|\det(\mathbf{L}^{T})\right| \leq \left|\det(\mathbf{L}^{T})\right| \cdot \left|\det(\mathbf{A})\right| = \left|\det(\mathbf{L}^{T}\mathbf{A})\right| \leq \prod_{k=1}^{K} \|\mathbf{L}^{\mathsf{T}}\mathbf{a}_k\| = \prod_{k=1}^{K} \lambda_k(\mathbf{L}^{\mathsf{T}}). \tag{2.44}$$

An upper bound on the product of the successive minima is given by Minkowski's second theorem.

Theorem 2.7 (Minkowski's Second Theorem) *For any full-rank* $\mathbf{G}$*,*

$$\prod_{k=1}^{K} \lambda_k(\mathbf{G}) \leq 2^K \left(\frac{\left|\det(\mathbf{G})\right|}{V_K}\right) \leq K^{K/2}\left|\det(\mathbf{G})\right|. \tag{2.45}$$

With (2.43), (2.44) and Theorem 2.7, we can establish the following.

[2] Specifically, it can be shown that there is a choice of rates $R_1, \ldots, R_K$ satisfying $\sum_k R_k = \frac{1}{2}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H})$ that satisfies the capacity region constraints from Theorem 2.2 and any choice of rates with a higher sum rate $\sum_k R_k > \frac{1}{2}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H})$ will violate these capacity constraints.

Theorem 2.8 ([31, Theorem 3])

$$\frac{1}{2}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})-\frac{K}{2}\log K\le\sum_{k=1}^{K}R_{\mathrm{comp},k}(\mathbf{H},P)\le\frac{1}{2}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H}). \tag{2.46}$$

Proof By the definition of $\mathbf{L}$ in (2.36), we have

$$\log\left|\det(\mathbf{L}^{\mathsf{T}})\right|=\frac{1}{2}\log\left|\det(\mathbf{L}\mathbf{L}^{\mathsf{T}})\right|=\frac{1}{2}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H}). \tag{2.47}$$

The lower bound now follows from (2.43), (2.45), and (2.47), whereas the upper bound follows from (2.43), (2.44), and (2.47). □

Theorem 2.8 asserts that the sum of the computation rates is never too far from the sum capacity of the channel (2.6) with power constraint P. An operational meaning for $\sum_{k=1}^{K}R_{\mathrm{comp},k}(\mathbf{H},P)$ is given in [31], where a low-complexity coding scheme based on compute-and-forward for the Gaussian MAC (2.6) that achieves this sum-rate is proposed. The remarkable conclusion from Theorem 2.8, is that while the individual computation rates $\{R_{\mathrm{comp},k}(\mathbf{H},P)\}$ may be very sensitive to the entries of $\mathbf{H}$, their sum is, to the first order, only influenced by the corresponding sum-capacity. This phenomenon is illustrated in Fig. 2.2.

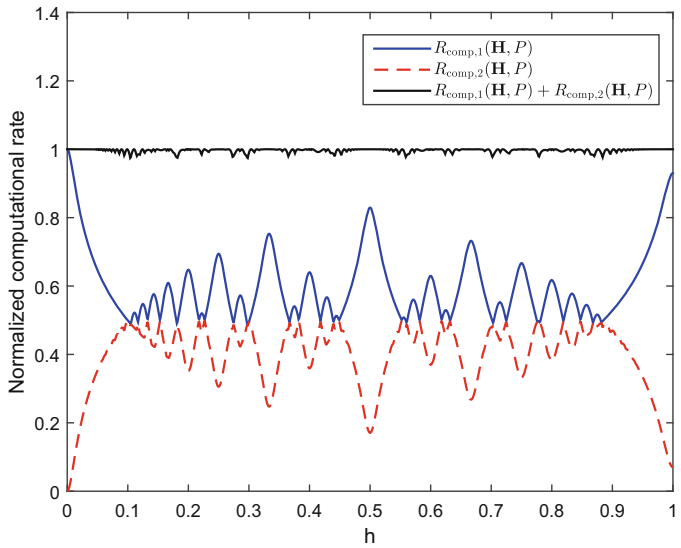


Fig. 2.2 $R_{\mathrm{comp},1}(\mathbf{H},P)$ and $R_{\mathrm{comp},2}(\mathbf{H},P)$ as a function of h for the channel $\mathbf{y}=\mathbf{x}_1+h\mathbf{x}_2+\mathbf{z}$ at $P=40\,\mathrm{dB}$. The sum of these computation rates is nearly equal to the multiple-access sum capacity. All rates are normalized by this sum capacity $1/2\log(1+(1+h^2)P)$

We are often particularly interested in estimating the value of $R_{\mathrm{comp},K}(\mathbf{H}, P)$, as this is the quantity that dictates the symmetric communication rate over the MAC channel (2.6) with power constraint P, when decoding is done via first recovering K integer linear combinations. However, directly estimating this quantity may be challenging, as it requires to first find $K-1$ linearly independent shortest lattice vectors. Estimating $R_{\mathrm{comp},1}(\mathbf{H}, P)$, on the other hand, is a much simpler task, as it only involves one shortest lattice vector. It is thus desirable to estimate $R_{\mathrm{comp},K}(\mathbf{H}, P)$ as a function of $R_{\mathrm{comp},1}(\mathbf{H}, P)$. Using the monotonicity of $R_{\mathrm{comp},k}(\mathbf{H}, P)$ in k and Theorem 2.8, yields the following simple estimate, which shows that if $R_{\mathrm{comp},1}(\mathbf{H}, P)$ is close to $\frac{1}{2K}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})$, then so is $R_{\mathrm{comp},K}(\mathbf{H}, P)$.

Proposition 2.1

$$\begin{aligned}&\left[\frac{1}{2K}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})-\frac{K}{2}\log K\right.\\&\left.-(K-1)\left(R_{\mathrm{comp},1}(\mathbf{H}, P)-\frac{1}{2K}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})\right)\right]^{+}\\&\le R_{\mathrm{comp},K}(\mathbf{H}, P)\le\frac{1}{2K}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H}),\end{aligned}\tag{2.48}$$

where $[x]^{+}=\max\{0, x\}$.

Proof By definition, we have that $R_{\mathrm{comp},1}(\mathbf{H}, P)\ge\cdots\ge R_{\mathrm{comp},K}(\mathbf{H}, P)$, which implies that

$$\sum_{k=1}^{L}R_{\mathrm{comp},k}(\mathbf{H}, P)\le L\cdot R_{\mathrm{comp},1}(\mathbf{H}, P)\tag{2.49}$$

$$\sum_{k=1}^{L}R_{\mathrm{comp},k}(\mathbf{H}, P)\ge L\cdot R_{\mathrm{comp},L}(\mathbf{H}, P).\tag{2.50}$$

The upper bound in (2.48) follows from (2.50) with $L=K$, combined with the upper bound from (2.43). To establish the lower bound in (2.48) we can write

$$\begin{aligned}R_{\mathrm{comp},K}(\mathbf{H}, P)&=\sum_{k=1}^{K}R_{\mathrm{comp},k}(\mathbf{H}, P)-\sum_{k=1}^{K-1}R_{\mathrm{comp},k}(\mathbf{H}, P)\\&\ge\frac{1}{2}\log\det(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})-\frac{K}{2}\log K-(K-1)R_{\mathrm{comp},1}(\mathbf{H}, P),\end{aligned}$$

where we have used the lower bound from (2.43), and (2.49) applied with $L = K-1$ in the last inequality. To arrive at the left hand side of (2.48), we write

$$R_{\text{comp},1}(\mathbf{H}, P) = \frac{1}{2K}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}) + \left(R_{\text{comp},1}(\mathbf{H}, P) - \frac{1}{2K}\log\det(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H})\right). \tag{2.51}$$

An alternative route for estimating $R_{\text{comp},K}(\mathbf{H}, P)$ involves studying the *dual lattice* of $\Lambda(\mathbf{L}^T)$.

Definition 2.10 (Dual Lattice) For a lattice $\Lambda(\mathbf{G})$ with a full-rank generator matrix $\mathbf{G} \in \mathbb{R}^{K\times K}$, the dual lattice is defined by

$$\Lambda^*(\mathbf{G}) \triangleq \Lambda\left((\mathbf{G}^{\mathsf{T}})^{-1}\right). \tag{2.52}$$

By definition, we have that if $\mathbf{x} \in \Lambda(\mathbf{G})$ and $\mathbf{x}^* \in \Lambda^*(\mathbf{G})$, then $\mathbf{x}^{\mathsf{T}}\mathbf{x}^* \in \mathbb{Z}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \Lambda(\mathbf{G})$ be linearly independent vectors such that $\|\mathbf{x}_k\| = \lambda_k(\mathbf{G})$ for $k = 1, \ldots, K$ and let $\mathbf{x}^* \in \Lambda^*(\mathbf{G})$ be such that $\|\mathbf{x}^*\| = \lambda_1\left((\mathbf{G}^{\mathsf{T}})^{-1}\right)$. Since $\{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ form a basis for $\mathbb{R}^K$, we must have that $\mathbf{x}_k^T\mathbf{x}^* \neq 0$ for some $k \in \{1, \ldots, K\}$. Thus, for this k, we must have that

$$\lambda_k(\mathbf{G})\,\lambda_1\left((\mathbf{G}^{\mathsf{T}})^{-1}\right) = \|\mathbf{x}_k\| \cdot \|\mathbf{x}^*\| \geq |\mathbf{x}_k^T\mathbf{x}^*| \geq 1, \tag{2.53}$$

where we have used the Cauchy–Schwartz inequality and the fact that $\mathbf{x}_k^T\mathbf{x}^* \in \mathbb{Z}$. Since $\lambda_k(\mathbf{G})$ is monotone in k and $k \leq K$, we conclude that

$$\lambda_K(\mathbf{G})\,\lambda_1\left((\mathbf{G}^{\mathsf{T}})^{-1}\right) \geq 1. \tag{2.54}$$

It turns out that the product of successive minima of a lattice and its dual can also be upper bounded.

Theorem 2.9 (Banaszczyk [4, Theorem 2.1]) *Let $\Lambda(\mathbf{G})$ be a lattice with a full-rank generating matrix $\mathbf{G} \in \mathbb{R}^{K\times K}$ and let $\Lambda^*(\mathbf{G}) = \Lambda\left((\mathbf{G}^{\mathsf{T}})^{-1}\right)$ be its dual lattice. The successive minima of $\Lambda(\mathbf{G})$ and $\Lambda^*(\mathbf{G})$ satisfy the following inequality*

$$\lambda_k(\mathbf{G})\,\lambda_{K-k+1}\left((\mathbf{G}^{\mathsf{T}})^{-1}\right) \leq K, \quad \forall k = 1, 2, \ldots, K.$$

Banaszczyk's theorem and (2.54) yield the following estimate on $R_{\text{comp},K}(\mathbf{H}, P)$.

Theorem 2.10 ([29])

$$\frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\left(\|\mathbf{a}\|^2+P\|\mathbf{H}\mathbf{a}\|^2\right)\right)-\log K \le R_{\mathrm{comp},K}(\mathbf{H},P) \le \frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\left(\|\mathbf{a}\|^2+P\|\mathbf{H}\mathbf{a}\|^2\right)\right) \tag{2.55}$$

Proof By (2.38), Theorem 2.9, applied with $k = K$, and (2.54) we have that

$$\log\lambda_1(\mathbf{L}^{-1})-\log K \le R_{\mathrm{comp},K}(\mathbf{H},P) \le \log\lambda_1(\mathbf{L}^{-1}). \tag{2.56}$$

By definition of successive minima,

$$\begin{aligned}\lambda_1^2(\mathbf{L}^{-1}) &= \min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\|\mathbf{L}^{-1}\mathbf{a}\|^2\\ &= \min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\mathbf{a}^{\mathsf{T}}(\mathbf{L}\mathbf{L}^{\mathsf{T}})^{-1}\mathbf{a}\\ &= \min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\mathbf{a}^{\mathsf{T}}(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H})\mathbf{a},\end{aligned} \tag{2.57}$$

where we have used the definition of $\mathbf{L}$ from (2.36) in the last equality. The theorem now follows by substituting (2.57) in (2.56). □

2.5 Asymptotic Bounds

For the single-user AWGN channel (2.1) with power constraint P, the capacity is $C(P) = \frac{1}{2}\log(1+P)$ bits/channel use, by Theorem 2.1. The MAC channel model (2.6) with power constraint P is richer than the AWGN model (unless $N = K = 1$), but we would nevertheless like to compare it to a simple AWGN channel. In our context, the notion of degrees-of-freedom (DoF) is a first-order approximation that measures how many AWGN channels (or fractions thereof) are needed to attain the same rate as the MAC sum capacity. To be precise, let $C(\mathbf{H}, P)$ be the sum-capacity of the channel (2.6), i.e.,

$$C(\mathbf{H},P) \triangleq \frac{1}{2}\log\det\left(\mathbf{I}+P\mathbf{H}^{\mathsf{T}}\mathbf{H}\right). \tag{2.58}$$

Then, the DoF offered by the MAC channel (2.6) with channel matrix $\mathbf{H}$ is defined as

$$\mathsf{DoF}(\mathbf{H}) \triangleq \lim_{P\to\infty} \frac{C(\mathbf{H}, P)}{C(P)} = \lim_{P\to\infty} \frac{\log\det\left(\mathbf{I} + P\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)}{\log(1+P)}. \tag{2.59}$$

It is well known that $\mathsf{DoF}(\mathbf{H}) = \mathrm{rank}(\mathbf{H})$. In particular, for almost all $\mathbf{H} \in \mathbb{R}^{N\times K}$ (w.r.t. Lebesgue measure) we have that $\mathsf{DoF}(\mathbf{H}) = \min(K, N)$.

In order to characterize the asymptotic behavior of communication schemes based on decoding integer-linear combinations, we define the DoF associated with decoding the best ℓ equations as

$$d_{\mathrm{comp},\ell}(\mathbf{H}) = \lim_{P\to\infty} \frac{R_{\mathrm{comp},\ell}(\mathbf{H}, P)}{\frac{1}{2}\log(1+P)}. \tag{2.60}$$

By Theorem 2.8, we have that

$$\frac{C(\mathbf{H}, P) - \frac{K}{2}\log(K)}{C(P)} \le \sum_{k=1}^{K} \frac{R_{\mathrm{comp},k}(\mathbf{H}, P)}{\frac{1}{2}\log(1+P)} \le \frac{C(\mathbf{H}, P)}{C(P)}. \tag{2.61}$$

Since the upper and lower bounds coincide in the limit of $P \to \infty$, we see that

$$\sum_{k=1}^{K} d_{\mathrm{comp},k}(\mathbf{H}) = \mathsf{DoF}(\mathbf{H}) = \mathrm{rank}(\mathbf{H}) \le \min\{K, N\}. \tag{2.62}$$

The main purpose of this section is to show that for almost all $\mathbf{H} \in \mathbb{R}^{N\times K}$ (w.r.t. the Lebesgue measure) we have that $d_{\mathrm{comp},1}(\mathbf{H}) = \cdots = d_{\mathrm{comp},K}(\mathbf{H}) = \frac{\min\{K,N\}}{K}$. By (2.62) and the monotonicity of $d_{\mathrm{comp},k}(\mathbf{H})$, it suffices to show that for almost every $\mathbf{H}$ we have $d_{\mathrm{comp},K}(\mathbf{H}) \ge \frac{\min\{K,N\}}{K}$. Our focus will therefore be on establishing lower bounds for $d_{\mathrm{comp},K}(\mathbf{H})$.

Our starting point is Theorem 2.10. Denoting the Kth singular value of $\mathbf{H}$ by $\sigma_K(\mathbf{H})$, this theorem gives

$$\begin{aligned} R_{\mathrm{comp},K}(\mathbf{H}, P) &\ge \frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P\|\mathbf{H}\mathbf{a}\|^2\right)\right) - \log K \\ &\ge \frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P\sigma_K^2(\mathbf{H})\|\mathbf{a}\|^2\right)\right) - \log K \\ &\ge \frac{1}{2}\log\left(1 + P\sigma_K^2(\mathbf{H})\right) - \log K. \end{aligned} \tag{2.63}$$

Since $\sigma_K^2(\mathbf{H})$ is strictly above 0 whenever $\text{rank}(\mathbf{H}) = K$, we conclude that if $\text{rank}(\mathbf{H}) = K$ then $d_{\text{comp},K}(\mathbf{H}) = 1$. For $K \leq N$, this is indeed the case for almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$. Thus, we have established that if $K \leq N$ then $d_{\text{comp},K}(\mathbf{H}) \geq \frac{\min\{K,N\}}{K}$ for almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$. The interesting case is therefore $N < K$, which we assume in the derivation.

Instead of bounding (2.63) in terms of $\sigma_K(\mathbf{H})$, we can resort to the tradeoff between the allowed length of $\mathbf{a}$ and the smallest attainable $\|\mathbf{Ha}\|$ [29]

$$\begin{aligned} R_{\text{comp},K}(\mathbf{H}, P) &\geq \frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\left(\|\mathbf{a}\|^2 + P\|\mathbf{Ha}\|^2\right)\right) - \log K \\ &\geq \frac{1}{2}\log\left(\min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\left(\|\mathbf{a}\|_\infty^2 + P\|\mathbf{Ha}\|_\infty^2\right)\right) - \log K, \end{aligned} \tag{2.64}$$

where for $\mathbf{x} \in \mathbb{R}^m$ we define $\|\mathbf{x}\|_\infty = \max\{|x_1|, \ldots, |x_m|\}$. For $0 < \epsilon < 1$, define $\kappa_\epsilon(\mathbf{H}) \geq 0$ as

$$\kappa_\epsilon(\mathbf{H}) \triangleq \inf_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}} \frac{\|\mathbf{Ha}\|_\infty}{\|\mathbf{a}\|_\infty^{1-\frac{K}{N}\frac{1}{1-\epsilon}}}. \tag{2.65}$$

We have that

$$\begin{aligned} \min_{\mathbf{a}\in\mathbb{Z}^K\setminus\{\mathbf{0}\}}\left(\|\mathbf{a}\|_\infty^2 + P\|\mathbf{Ha}\|_\infty^2\right) &\geq \min_{\ell=1,2,\ldots}\left(\ell^2 + P\kappa_\epsilon^2(\mathbf{H})\ell^{2\left(1-\frac{K}{N}\frac{1}{1-\epsilon}\right)}\right) \\ &\geq \min_{t>0}\left(t + P\kappa_\epsilon^2(\mathbf{H})t^{1-\frac{K}{N}\frac{1}{1-\epsilon}}\right) \\ &= \frac{1}{1-\frac{N}{K}(1-\epsilon)}\cdot\left(\frac{K}{N}\frac{1}{1-\epsilon}-1\right)^{\frac{N}{K}(1-\epsilon)}\cdot\left(\kappa_\epsilon^2(\mathbf{H})P\right)^{\frac{N}{K}(1-\epsilon)}, \end{aligned} \tag{2.66}$$

where the last equality is obtained by straightforward differentiation. Substituting (2.66) in (2.64) and recalling the definition of $d_{\text{comp},K}(\mathbf{H})$, we have established that for any $0 < \epsilon < 1$, the following holds

$$d_{\text{comp},K}(\mathbf{H}) \geq \frac{N}{K}(1-\epsilon)\left(1 + 2\lim_{P\to\infty}\frac{\log\kappa_\epsilon(\mathbf{H})}{\log P}\right). \tag{2.67}$$

It now remains to show that $\kappa_\epsilon(\mathbf{H}) > 0$ for every $0 < \epsilon < 1$, and almost every $\mathbf{H}$. To this end, we resort to the literature on *systems of small linear forms*. Several results in this field can be used, depending on whether the entries of $\mathbf{H}$ are independent or dependent (i.e., they can be characterized by fewer than NK parameters). Below, we state the most general available result, which was recently obtained by Beresnevich, Bernik, and Budarina [5].

2.5.1 Small Linear Forms

We will need several definitions before we can state (an adaptation of) the main result from [5].

For $j = 1, \ldots, N$, let $U_j \subset \mathbb{R}^{d_j}$ be an open ball, and $\mathbf{f}_j = (f_{j1}, \ldots, f_{jK}) : U_j \mapsto \mathbb{R}^K$ be functions. For $(\mathbf{x}_1, \ldots, \mathbf{x}_N) \in U_1 \times \cdots \times U_N$, we define

$$\mathbf{F} = \mathbf{F}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \triangleq \begin{bmatrix} \mathbf{f}_1(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}_N(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} f_{11}(\mathbf{x}_1) & \ldots & f_{1K}(\mathbf{x}_1) \\ \vdots & \vdots & \vdots \\ f_{N1}(\mathbf{x}_N) & \ldots & f_{NK}(\mathbf{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times K}. \tag{2.68}$$

For $\rho > 0$, define the set

$$\mathcal{W}(\mathbf{F}, \rho) \triangleq \Big\{ (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in U_1 \times \cdots \times U_N \ : \ \|\mathbf{F}(\mathbf{x}_1, \ldots, \mathbf{x}_N)\mathbf{a}\|_\infty < (\|\mathbf{a}\|_\infty)^{-\rho}) \\ \text{for infinitely many } \mathbf{a} \in \mathbb{Z}^K \setminus \{\mathbf{0}\} \Big\}. \tag{2.69}$$

Theorem 2.11 ([5, Theorem 2]) *Let $K > N \geq 1$ be integers, and let $U_1, \ldots, U_N$, $\mathbf{f}_1, \ldots, \mathbf{f}_N$, $\mathbf{F}$ and $\mathcal{W}(\mathbf{F}, \rho)$ be as above. Suppose that for each $j = 1, \ldots, N$ the coordinate functions $f_{j1}, \ldots, f_{jK}$ of the map $\mathbf{f}_j$ are analytic and linearly independent over $\mathbb{R}$. Then,*

$$\mu\left(\mathcal{W}(\mathbf{F}, \rho)\right) = \begin{cases} 0 & \text{if } \rho > \frac{K}{N} - 1, \\ \prod_{j=1}^{N} \mu(U_j) & \text{if } \rho \leq \frac{K}{N} - 1 \end{cases} \tag{2.70}$$

where $\mu(B)$ denotes the Lebesgue measure of a set $B \subset \mathbb{R}^d$.

An immediate corollary of Theorem 2.11 is the following.

Corollary 2.1 *Let $K > N \geq 1$ be integers, and let $U_1, \ldots, U_N$, $\mathbf{f}_1, \ldots, \mathbf{f}_N$, $\mathbf{F}$ and $\mathcal{W}(\mathbf{F}, \rho)$ be as above. Suppose that, for each $j = 1, \ldots, N$, the coordinate functions $f_{j1}, \ldots, f_{jK}$ of the map $\mathbf{f}_j$ are analytic and linearly independent over $\mathbb{R}$. Then, for any $0 < \epsilon < 1$ and almost every $(\mathbf{x}_1, \ldots, \mathbf{x}_N) \in U_1 \times \cdots \times U_N$, we have that $\kappa_\epsilon(\mathbf{F}(\mathbf{x}_1, \ldots, \mathbf{x}_N)) > 0$.*

We can now combine Corollary 2.1 and (2.67) for several cases of particular interest.

2.5.2 Independent Channel Gains

A common assumption in wireless communication is that the entries h_{ij} of the channel matrix $\mathbf{H} \in \mathbb{R}^{N\times K}$ are independent. In the context of Theorem 2.11, this corresponds to taking $U_j = [-\tau, \tau]^K$ for all $j = 1, \ldots, N$, where $\tau \in \mathbb{R}^+$ is some large number, and $\mathbf{f}_j(\mathbf{x}_j) = (x_{j1}, \ldots, x_{jK})$. These functions certainly satisfy the conditions of Corollary 2.1, and we can therefore deduce the following.

Corollary 2.2 *Let $K > N \geq 1$. For almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$ we have that $\kappa_\epsilon(\mathbf{H}) > 0$.*

Now, combining the corollary above and (2.67) we see that for $K > N \geq 1$ we have that $d_{\mathrm{comp},K}(\mathbf{H}) \geq \frac{N}{K}$ for almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$. Recalling that for $1 \leq K \leq N$ we have that $d_{\mathrm{comp},K}(\mathbf{H}) \geq 1$ for almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$, we recover the following lemma from [29].[3]

Lemma 2.1 ([29, Lemma 3]) *For almost every $\mathbf{H} \in \mathbb{R}^{N\times K}$,*

$$K \cdot d_{comp,K}(\mathbf{H}) = \min\{K, N\}. \tag{2.71}$$

Roughly speaking, this allows us to conclude that, in the limit of large P, the integer-forcing strategy does as well as the optimal sum-capacity-achieving scheme.

2.5.3 Dependent Channel Gains

In many applications of interest, the channel model $\mathbf{Y} = \mathbf{HX} + \mathbf{Z}$ represents an effective channel induced by certain signal processing operations performed at the transmitters and the receivers. Often, these operations create dependencies between the entries of $\mathbf{H}$, which requires replacing the Lebesgue measure in the DoF analysis with a measure on a suitable manifold.

As a canonical example, we will consider the symmetric two-user X-channel [20, 21, 23, 25, 28]. This channel consists of two transmitters emitting the signals $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, each in $\mathbb{R}^{1\times T}$ and satisfying the power constraint $\|\mathbf{x}_k\|^2 \leq TP$, and two receivers observing the signals

$$\mathbf{y}_1 = \mathbf{x}_1 + g\mathbf{x}_2 + \mathbf{z}_1$$
$$\mathbf{y}_2 = g\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{z}_2,$$

[3] The proof of Lemma 3 from [29] relied on [19, Corollary 2], which can be obtained as a special case of [5, Theorem 2].

respectively, where $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{1\times T}$ are two statistically independent i.i.d. $\mathcal{N}(0, 1)$ noises. Each transmitter has two messages, one for the first receiver and one for the second receiver, and we assume all four messages are of the same rate R. We now describe one particular transmission scheme for this channel. We use one lattice codebook of rate R and power P, such that the message from user k to receiver j is encoded to a lattice codeword $\tilde{\mathbf{x}}_{jk}$. The users then transmit

$$\mathbf{x}_1 = \frac{1}{\sqrt{1+g^2}}\left(\tilde{\mathbf{x}}_{11} + g\tilde{\mathbf{x}}_{21}\right)$$
$$\mathbf{x}_2 = \frac{1}{\sqrt{1+g^2}}\left(\tilde{\mathbf{x}}_{22} + g\tilde{\mathbf{x}}_{12}\right).$$

Consequently, the receivers observe

$$\mathbf{y}_1 = \frac{1}{\sqrt{1+g^2}}\left(\tilde{\mathbf{x}}_{11} + g(\tilde{\mathbf{x}}_{21} + \tilde{\mathbf{x}}_{22}) + g^2\tilde{\mathbf{x}}_{12}\right) + \mathbf{z}_1$$
$$\mathbf{y}_2 = \frac{1}{\sqrt{1+g^2}}\left(\tilde{\mathbf{x}}_{22} + g(\tilde{\mathbf{x}}_{12} + \tilde{\mathbf{x}}_{11}) + g^2\tilde{\mathbf{x}}_{21}\right) + \mathbf{z}_1.$$

Since the channel output is symmetric across receivers, it suffices to analyze the rates that allow the first receiver to decode its two desired codewords $\tilde{\mathbf{x}}_{11}$ and $\tilde{\mathbf{x}}_{12}$. Noting that $\tilde{\mathbf{x}}_2 \triangleq \tilde{\mathbf{x}}_{12} + \tilde{\mathbf{x}}_{11}$ is a lattice codeword itself, we can write

$$\mathbf{y}_1 = \mathbf{h}^{\mathsf{T}}\mathbf{X}_1 + \mathbf{z}_1, \tag{2.72}$$

where

$$\mathbf{h} = \mathbf{h}(g) = \frac{1}{\sqrt{1+g^2}}[1\ g^2\ g],\ \mathbf{X}_1 = \left[\tilde{\mathbf{x}}_{11}^{\mathsf{T}}\ \tilde{\mathbf{x}}_{12}^{\mathsf{T}}\ \tilde{\mathbf{x}}_{2}^{\mathsf{T}}\right]^{\mathsf{T}}. \tag{2.73}$$

Thus, the effective channel (2.72) induced by our transmission scheme falls within our generic model introduced in the first section. We can decode the two desired codewords $\tilde{\mathbf{x}}_{11}$ and $\tilde{\mathbf{x}}_{12}$, as well as the nuisance codeword $\tilde{\mathbf{x}}_2$, by decoding three integer-linear combinations and then inverting them. Thus, the asymptotic performance of our scheme depends on $d_{\mathrm{comp},3}(\mathbf{h})$.

We would like to apply Corollary 2.1 in order to show that $\kappa_\epsilon(\mathbf{h}(g)) > 0$ for almost every $g \in \mathbb{R}$. To this end, we take $U_1 = [-\tau, \tau]$ for some large $\tau \in \mathbb{R}^+$ and set

$$\mathbf{f}_1(x) = \left(\frac{1}{\sqrt{1+x^2}}, \frac{x^2}{\sqrt{1+x^2}}, \frac{x}{\sqrt{1+x^2}}\right), \tag{2.74}$$

such that $\mathbf{h}(g) = \mathbf{f}_1(g) \in \mathbb{R}^{1\times 3}$. Certainly, $\mathbf{f}_1$ satisfies the conditions of Corollary 2.1, and we therefore obtain the following.

Corollary 2.3 *For almost every $g \in \mathbb{R}$ we have that $\kappa_\epsilon(\mathbf{h}(g)) > 0$.*

Combining the corollary above with (2.67), we have established that for almost every $g \in \mathbb{R}$, the proposed communication scheme attains $d_{\text{comp},3}(\mathbf{h}(g)) = 1/3$.

The operational implication of this result, is that using the lattice-based communication scheme proposed above, each user can send both of its messages reliably, each with a rate that scales like $\frac{1}{3} \cdot \frac{1}{2}\log(P)$ with P. To appreciate this, note that the naïve scheme, which avoids interference by transmitting each of the 4 messages over different $T/4$ channel uses, can only achieve reliable communication with rates below $\frac{1}{4} \cdot \frac{1}{2}\log(1 + 4P)$.

2.6 Non-Asymptotic Bounds

For communication applications, it is often of interest to understand performance for finite P, as in practice the allowed transmission power is limited, and usually quite moderate.

As a canonical example, consider the symmetric K-user Gaussian interference channel, depicted in Fig. 2.3. In this channel model, there are K users, each transmitting a signal $\mathbf{x}_k \in \mathbb{R}^T$, $k = 1, \ldots, K$, subject to the power constraint $\|\mathbf{x}_k\|^2 \le TP$. There are also K receivers with observations

$$\mathbf{y}_k = \mathbf{x}_k + g \sum_{j \neq k} \mathbf{x}_j + \mathbf{z}_k, \quad k = 1, \ldots, K \tag{2.75}$$

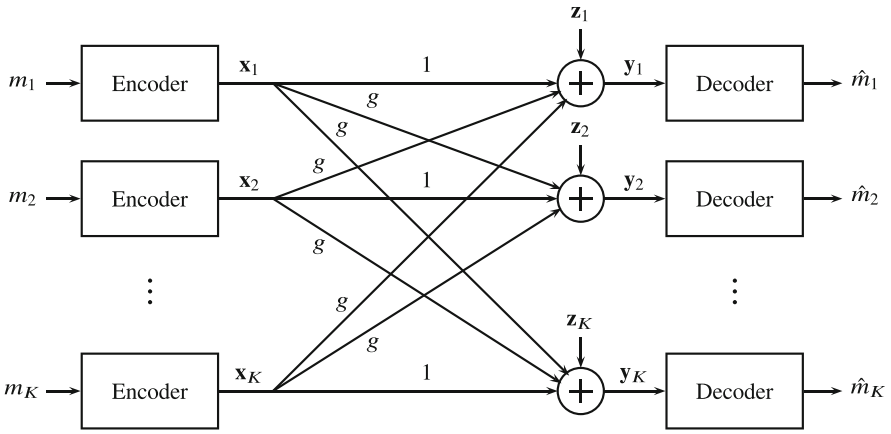


Fig. 2.3 Block diagram of a symmetric Gaussian K-user interference channel

where $g \in \mathbb{R}$ is the (symmetric) interference gain, and $\mathbf{z}_k$ is i.i.d. Gaussian noise with zero mean and unit variance. The goal of the kth receiver is to decode only the codeword $\mathbf{x}_k$, whereas all other codewords are interference. In the proceeding discussion, we will assume that $1 < g < \sqrt{P}$.

The naïve approach for dealing with interference is to avoid it entirely. This corresponds to splitting the channel uses into T/K different slots, and letting only one user transmit within each slot. In this scheme, when the kth user transmits, the kth receiver observes its signal without any interference, and the resulting achievable rate is $\frac{1}{K} \cdot \frac{1}{2} \log(1 + KP)$.

A different, and sometimes more efficient, approach, is *interference alignment*. For the symmetric interference channel, this approach boils down to having all users encode their messages using the same lattice codebook. Consider the sum of interfering codewords at receiver k, $\mathbf{x}_{\text{interference},k} = \sum_{j \neq k} \mathbf{x}_j$. Owing to the fact that the lattice is closed under integer-linear combinations, $\mathbf{x}_{\text{interference},k}$ is itself a lattice codeword. Consequently, the effective two-user channel seen by receiver k is

$$\mathbf{y}_k = \mathbf{x}_k + g\,\mathbf{x}_{\text{interference},k} + \mathbf{z}_k. \tag{2.76}$$

Now, it is possible to recover $\mathbf{x}_k$ by decoding two linearly independent integer-linear combinations of $\mathbf{x}_k$ and $\mathbf{x}_{\text{interference},k}$.

The achievable rate of this interference alignment scheme is therefore the second computation rate[4] for the channel $\mathbf{H} = [1\ g]$. By Theorem 2.10, we can lower bound the second computation rate by

$$R_{\text{comp},2}(\mathbf{H}, P) \geq \frac{1}{2} \log \left(\min_{\mathbf{a} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P \|\mathbf{H}\mathbf{a}\|^2 \right) \right) - 1. \tag{2.77}$$

Setting $\mathbf{H} = [1\ g]$, $\mathbf{a} = [-p\ q]$, and assuming without loss of generality that $q \geq 0$, we can write

$$\min_{\mathbf{a} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P \|\mathbf{H}\mathbf{a}\|^2 \right) = \min_{p \in \mathbb{Z}, q \in \mathbb{N}, (p,q) \neq (0,0)} \left(p^2 + q^2 + P|qg - p|^2 \right). \tag{2.78}$$

Defining $\tilde{p} = p - q\lfloor g \rfloor$ and $\tilde{g} = g - \lfloor g \rfloor$, we can rewrite this as

$$\min_{\tilde{p} \in \mathbb{Z}, q \in \mathbb{N}, (\tilde{p},q) \neq (0,0)} \left((q\lfloor g \rfloor + \tilde{p})^2 + q^2 + P\left|q\lfloor g \rfloor + q\tilde{g} - q\lfloor g \rfloor - \tilde{p}\right|^2 \right) \tag{2.79}$$

$$= \min_{\tilde{p} \in \mathbb{Z}, q \in \mathbb{N}, (\tilde{p},q) \neq (0,0)} \left((q\lfloor g \rfloor + \tilde{p})^2 + q^2 + P\left|q\tilde{g} - \tilde{p}\right|^2 \right). \tag{2.80}$$

[4] Up to a small correction term, due to the fact that the effective user $\mathbf{x}_{\text{interference},k}$ has power $(K-1)P$ instead of P. See [31] for more details.

Since $\tilde{g} \geq 0$ by definition, we see that for $\tilde{p} < 0$ the expression above is lower bounded by P. We can therefore write

$$\begin{aligned}\min_{\mathbf{a}\in\mathbb{Z}^2\setminus\{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P\|\mathbf{H}\mathbf{a}\|^2\right) &\geq \min_{\tilde{p}\in\mathbb{Z},q\in\mathbb{N},(\tilde{p},q)\neq(0,0)} \left((q\lfloor g\rfloor + \tilde{p})^2 + q^2 + P|q\tilde{g} - \tilde{p}|^2\right) \\ &\geq \min\left\{P, \min_{(\tilde{p},q)\in\mathbb{N}^2\setminus\{\mathbf{0}\}} \left((q\lfloor g\rfloor + \tilde{p})^2 + q^2 + P|q\tilde{g} - \tilde{p}|^2\right)\right\} \\ &\geq \min\left\{P, \min_{(\tilde{p},q)\in\mathbb{N}^2\setminus\{\mathbf{0}\}} \max(q^2\lfloor g\rfloor^2,\ P|q\tilde{g} - \tilde{p}|^2)\right\}. \end{aligned} \tag{2.81}$$

Next, we will study the behavior of the last term in (2.81). In particular, for an integer $1 \leq b \leq \sqrt{P}$ and $0 < \delta < 1$, we will study the Lebesgue measure of the "outage set"

$$\begin{aligned}\mathcal{W}_{b,\delta} &= \left\{g \in [b, b+1) \ : \min_{(\tilde{p},q)\in\mathbb{N}^2\setminus\{\mathbf{0}\}} \max(q^2\lfloor g\rfloor^2,\ P|q\tilde{g} - \tilde{p}|^2) < \frac{\sqrt{g}}{2}P^{\frac{1}{2}(1-\delta)}\right\} \\ &\subset b + \left\{x \in [0,1) : |qx - \tilde{p}| < \sqrt{b}P^{-\frac{1}{4}(1+\delta)} \text{ for some } q \leq \frac{P^{\frac{1}{4}(1-\delta)}}{\sqrt{b}},\ \tilde{p} \in \mathbb{N}\right\}. \end{aligned} \tag{2.82}$$

Note that for all $g \in [b, b+1) \setminus \mathcal{W}_{b,\delta}$, we have that

$$\min_{\mathbf{a}\in\mathbb{Z}^2\setminus\{\mathbf{0}\}} \left(\|\mathbf{a}\|^2 + P\|\mathbf{H}\mathbf{a}\|^2\right) \geq \frac{\sqrt{g}}{2}P^{\frac{1}{2}(1-\delta)}, \tag{2.83}$$

which implies, by (2.77), that

$$R \geq \frac{1}{4}\log(g^2 P) - \frac{\delta}{4}\log P - \frac{3}{2} \tag{2.84}$$

for all $g \in [b, b+1) \setminus \mathcal{W}_{b,\delta},\ 1 \leq b \leq \sqrt{P}$.

In order to upper bound $\mu(\mathcal{W}_{b,\delta})$, for any $q \in \mathbb{Z}^+$, we define the set

$$\mathcal{T}_{b,\delta}(q) \triangleq \left[\left\{0, \frac{1}{q}, \ldots, \frac{q-1}{q}\right\} + \frac{\Phi_{b,\delta}}{q}\mathcal{I}\right] \bmod [0,1), \tag{2.85}$$

where $\mathcal{I} \triangleq [-1, 1]$ and $\Phi_{b,\delta} \triangleq \sqrt{b}P^{-\frac{1}{4}(1+\delta)}$. It is easy to see that

$$\mathcal{W}_{b,\delta} \subset b + \bigcup_{q=1}^{q_{\max}(b,\delta)} \mathcal{T}_{b,\delta}(q), \tag{2.86}$$

where $q_{\max}(b,\delta) \triangleq \left\lfloor \frac{P^{\frac{1}{4}(1-\delta)}}{\sqrt{b}} \right\rfloor$. Therefore,

$$\begin{aligned}
\mu(\mathcal{W}_{b,\delta}) &\le \mu\left(\bigcup_{q=1}^{q_{\max}(b,\delta)} \mathcal{T}_{b,\delta}(q)\right) \\
&\le \sum_{q=1}^{q_{\max}(b,\delta)} \mu\left(\mathcal{T}_{b,\delta}(q)\right) \\
&\le \sum_{q=1}^{q_{\max}(b,\delta)} 2\Phi_{b,\delta} \\
&= 2q_{\max}(b,\delta)\Phi_{b,\delta} \\
&\le 2P^{-\frac{\delta}{2}}. \qquad (2.87)
\end{aligned}$$

Now, setting $\delta = 2(\gamma+1)/\log(P)$, (2.84) and (2.87) imply that we can achieve a rate satisfying

$$\begin{aligned}
R &\ge \frac{1}{4}\log(g^2P) - \frac{\gamma+1}{2} - \frac{3}{2} \\
&= \frac{1}{4}\log(g^2P) - \frac{\gamma}{2} - 2 \qquad (2.88)
\end{aligned}$$

for all $g \in [b, b+1) \setminus \mathcal{W}$, where $\mathcal{W} = \mathcal{W}_{b,2(\gamma+1)/\log(P)}$ has Lebesgue measure at most $2^{-\gamma}$.

To appreciate this result, it should be contrasted with the rate attained by interference avoidance. The interference alignment rate scales with P as $\frac{1}{4}\log(g^2P)$ whereas that of interference avoidance only scales as $\frac{1}{2K}\log(P)$. For $K \ge 3$ and large P, the improvement is very significant. It can also be shown that the symmetric capacity of the symmetric K-user Gaussian interference channel is upper bounded by $\frac{1}{4}\log(g^2P) + 1$. Thus, we have the following theorem.

Theorem 2.12 ([31]) *The lattice interference alignment scheme described above attains the symmetric capacity of the symmetric K-user Gaussian interference channel to within $3 + \gamma/2$ bits for all $g \in [1, \sqrt{P}) \setminus \{\mathcal{W}\}$, where the set $\mathcal{W} \subset [1, \sqrt{P})$ has Lebesgue measure at most $(\sqrt{P} - 1)2^{-\gamma}$.*

2.7 Conclusions and Open Problems

In this chapter, we demonstrated that classical and modern results from the theory of Diophantine approximation are extremely useful for obtaining upper and lower bounds for the performance of lattice-based communication strategies. In particular,

the compute-and-forward strategy makes it possible for a receiver to obtain integer-linear combinations of codewords, with the rate determined by how well the real-valued channel coefficients are approximated by the chosen integer coefficients. Though not discussed in this survey, similar ideas have been found useful for distributed data compression, where the compression rates are determined by how well the source covariance matrix can be approximated by a matrix with integer coefficients [12, 18, 30]. While explicitly identifying these integer coefficients is a challenging optimization problem, we can obtain universal bounds on the achievable communication rates via Diophantine approximation.

A major focus of this chapter was on degrees-of-freedom characterizations, i.e., the first-order term in the rate expression as the power P tends to infinity. For this regime, Diophantine approximation results allow us to obtain tight bounds up to a set of channel matrices with Lebesgue measure zero, even when dependencies exist between the channel gains, as in interference alignment. Going further, one can follow a similar approach to determine the degrees-of-freedom of essentially any interference network (see, for instance, [7, 8, 25, 35] for more details).

We also considered non-asymptotic bounds that hold for any choice of P. Specifically, we examined the symmetric K-user Gaussian interference, and derived a lower bound on the capacity whose gap to the upper bound depends on the measure of the excluded channel gains. Similar results are available for the two-user X channel [28]. For larger networks, we need to rely on more sophisticated interference alignment schemes, and more research is needed to develop non-asymptotic bounds that can handle the resulting dependencies. Specifically, alignment schemes for K-user interference channels (with arbitrary channel gains) utilize many signaling directions based on monomials constructed from the channel gains [7, 25]. This corresponds to a codeword emitted per signaling direction with a rate penalty for each additional codeword layer. In the limit as P tends to infinity, these rate penalties can be safely ignored to approach the optimal degrees-of-freedom of $1/2$ per user. However, for finite P, we must carefully tradeoff the number of codeword layers with the measure of excluded channel gains to attain the best performance. This in turn requires non-asymptotic Diophantine approximation bounds over manifolds. See [1, 13] for recent progress in this direction.

References

1. Adiceam, F., Beresnevich, V., Levesley, J., Velani, S., Zorin, E.: Diophantine approximation and applications in interference alignment. Adv. Math. **302**, 231–279 (2016)
2. Ahlswede, R.: Multi-way communication channels. In: Proceedings of the 2nd International Symposium on Information Theory, Prague, pp. 23–52. Publishing House of the Hungarian Academy of Sciences, Hungarian (1971)
3. Arikan, E.: Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **55**(7), 3051–3073 (2009)

4. Banaszczyk, W.: New bounds in some transference theorems in the geometry of numbers. Math. Ann. **296**(1), 625–635 (1993)
5. Beresnevich, V., Bernik, V., Budarina, N.: Systems of small linear forms and Diophantine approximation on manifolds. ArXiv e-prints (2017). http://arxiv.org/abs/1707.00371
6. Berrou, C., Glavieux, A.: Near optimum error correcting coding and decoding: Turbo-codes. IEEE Trans. Commun. **44**(10), 1261–1271 (1996)
7. Cadambe, V.R., Jafar, S.A.: Interference alignment and the degrees of freedom for the K-user interference channel. IEEE Trans. Inf. Theory **54**(8), 3425–3441 (2008)
8. Cadambe, V.R., Jafar, S.A.: Interference alignment and the degrees of freedom of wireless X networks. IEEE Trans. Inf. Theory **55**(5), 2334–2344 (2009)
9. Chung, S.Y., Forney, G.D., Richardson, T.J., Urbanke, R.: On the design of low-density parity-check codes within 0.0045 db of the Shannon limit. IEEE Commun. Lett. **5**(2), 58–60 (2001)
10. Cioffi, J.M., Dudevoir, G.P., Eyuboglu, M.V., Forney, G.D.: MMSE decision-feedback equalizers and coding. I. Equalization results. IEEE Trans. Commun. **43**(10), 2582–2594 (1995)
11. Cover, T., Thomas, J.: Elements of Information Theory, 2nd edn. Wiley-Interscience, Hoboken, (2006)
12. Domanovitz, E., Erez, U.: Outage probability bounds for integer-forcing source coding. In: Proceedings of the IEEE Information Theory Workshop (ITW 2017). Kaohsiung, Taiwan (2017)
13. Domanovitz, E., Erez, U.: Outage behavior of integer forcing with random unitary pre-processing. IEEE Trans. Inf. Theory **64**(4), 2774–2790 (2018)
14. El Gamal, A., Kim, Y.H.: Network Information Theory. Cambridge University, Cambridge (2011)
15. Erez, U., Zamir, R.: Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. IEEE Trans. Inf. Theory **50**(10), 2293–2314 (2004)
16. Gallager, R.: Low-density parity-check codes. IRE Trans. Inf. Theory **8**(1), 21–28 (1962)
17. Guess, T., Varanasi, M.K.: An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels. IEEE Trans. Inf. Theory **51**(1), 173–187 (2005)
18. He, W., Nazer, B.: Integer-forcing source coding: Successive cancellation and source-channel duality. In: Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), pp. 155–159 (2016)
19. Hussain, M., Levesley, J.: The metrical theory of simultaneously small linear forms. Functiones et Approximatio Commentarii Mathematici **48**(2), 167–181 (2013)
20. Jafar, S.A.: Interference alignment—a new look at signal dimensions in a communication network. In: Foundations and Trends in Communications and Information Theory, vol. 7. NOW Publishers, Boston (2011)
21. Jafar, S.A., Shamai (Shitz), S.: Degrees of freedom region for the MIMO X channel. IEEE Trans. Inf. Theory **54**(1), 151–170 (2008)
22. Liao, H.: Multiple access channels. Ph.D. thesis, University of Hawaii, Honolulu (1972)
23. Maddah-Ali, M.A., Motahari, A.S., Khandani, A.K.: Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis. IEEE Trans. Inf. Theory **54**(8), 3457–3470 (2008)
24. Micciancio, D., Goldwasser, S.: Complexity of Lattice Problems: A Cryptographic Perspective. The Kluwer International International Series in Engineering and Computer Science, vol. 671. Kluwer Academic Publishers, Cambridge, (2002)
25. Motahari, A.S., Oveis-Gharan, S., Maddah-Ali, M.A., Khandani, A.K.: Real interference alignment: Exploiting the potential of single antenna systems. IEEE Trans. Inf. Theory **60**(8), 4799–4810 (2014)
26. Nazer, B., Gastpar, M.: Compute-and-forward: harnessing interference through structured codes. IEEE Trans. Inf. Theory **57**(10), 6463–6486 (2011)
27. Nazer, B., Cadambe, V.R., Ntranos, V., Caire, G.: Expanding the compute-and-forward framework: unequal powers, signal levels, and multiple linear combinations. IEEE Trans. Inf. Theory **62**(9), 4879–4909 (2016)

28. Niesen, U., Maddah-Ali, M.A.: Interference alignment: from degrees-of-freedom to constant-gap capacity approximations. IEEE Trans. Inf. Theory **59**(8), 4855–4888 (2013)
29. Ordentlich, O., Erez, U.: Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap. IEEE Trans. Inf. Theory **61**(1), 323–340 (2015)
30. Ordentlich, O., Erez, U.: Integer-forcing source coding. IEEE Trans. Inf. Theory **63**(2), 1253–1269 (2017)
31. Ordentlich, O., Erez, U., Nazer, B.: The approximate sum capacity of the symmetric K-user Gaussian interference channel. IEEE Trans. Inf. Theory **60**(6), 3450–3482 (2014)
32. Richardson, T., Urbanke, R.: Modern Coding Theory. Cambridge University, Cambridge (2008)
33. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
34. Tse, D., Viswanath, P.: Fundamentals of Wireless Communication. Cambridge University Press, Cambridge (2005)
35. Wu, Y., Shamai (Shitz), S., Verdú, S.: Information dimension and the degrees of freedom of the interference channel. IEEE Trans. Inf. Theory **61**(1), 256–279 (2015)
36. Zamir, R.: Lattice Coding for Signals and Networks. Cambridge University, Cambridge (2014)
37. Zhan, J., Nazer, B., Erez, U., Gastpar, M.: Integer-forcing linear receivers. IEEE Trans. Inf. Theory **60**(12), 7661–7685 (2014)

Chapter 3
On Fast-Decodable Algebraic Space–Time Codes

Amaro Barreal and Camilla Hollanti

Abstract In the near future, the 5th generation (5G) of wireless systems will be well established. They will consist of an integration of different techniques, including distributed antenna systems and massive multiple-input multiple-output (MIMO) systems, and the overall performance will highly depend on the channel coding techniques employed. Due to the nature of future wireless networks, space–time codes are no longer merely an object of choice, but will often appear naturally in the communications setting. However, as the involved communication devices often exhibit a modest computational power, the complexity of the codes to be utilised should be reasonably low for possible practical implementation. Fast-decodable codes enjoy reduced complexity of maximum-likelihood (ML) decoding due to a smart inner structure allowing for parallelisation in the ML search. The complexity reductions considered in this chapter are entirely owing to the algebraic structure of the considered codes, and could be further improved by employing non-ML decoding methods, however yielding suboptimal performance. The aim of this chapter is twofold. First, we provide a tutorial introduction to space–time coding and study powerful algebraic tools for their design and construction. Secondly, we revisit algebraic techniques used for reducing the worst-case decoding complexity of both single-user and multiuser space-time codes, alongside with general code families and illustrative examples.

3.1 Introduction

Let us start this chapter by introducing, very briefly, the reader to the field of algebraic space–time coding. While there are various design criteria to be considered as well as a plethora of code constructions for a variety of different channel models and

A. Barreal · C. Hollanti (✉)
Department of Mathematics and Systems Analysis, Aalto University, Aalto, Finland
e-mail: amaro.barreal@aalto.fi; camilla.hollanti@aalto.fi

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*,
Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_3

communications settings, we will here only review the developments most relevant to the rest of this chapter.

The first space–time code, the Alamouti code [1], was introduced in 1998 and gave rise to a massive amount of research in the attempt to construct well-performing codes for various multi-antenna wireless communications settings. It was discovered that the code matrices constituting this particular code actually depict an algebraic structure known as the Hamiltonian quaternions, and by restriction to Lipschitz (i.e., integral) quaternions, the (unconstrained) code becomes a lattice. As Hamiltonian quaternions are the most popular example of a division algebra, this finding prompted the study of general division algebra space–time lattice codes [4, 34].

Division algebras are related to achieving full diversity by maximising the rank of the code matrices [38]. Soon it was noticed that by choosing the related field extensions carefully, one can achieve non-vanishing determinants (NVD) [4] for the codewords, implying a non-vanishing coding gain [38]. As the coding gain is inversely proportional to the decoding error probability, this in turn prevents the error probability from blowing up. A related notion, the diversity–multiplexing gain [43] captures the tradeoff between the decay speed of the decoding error probability and available degrees of freedom. It is known that for symmetric systems, that is, with an equal number of transmit and receive antennas, full-rate space-time codes with the NVD property achieve the optimal tradeoff of the channel.

Several explicit constructions of space–time codes based on cyclic division algebras exist in the literature. For instance, Perfect space-time codes and their generalisations [5, 12, 30] provide orthogonal lattices for any number of antennas, whereas maximal order codes [14, 15, 39] optimise the coding gain, while giving up on the orthogonality of the underlying lattice.

In the multiuser settings considered in this chapter, multiple users are communicating to a joint destination, with or without cooperating with each other. When cooperation is allowed, it is possible to take advantage of intermediate distributed relays which aid the active transmitter in the communication process. Various protocols exist for enabling this type of diversity—the one considered here is the non-orthogonal half-duplex amplify-and-forward protocol, see [42]. The non-cooperative case is referred to as the multiple access channel (MAC), where users transmit signals independently of each other. Some algebraic MAC codes are presented in [22, 23], among others.

One of the biggest obstacles in utilising space-time lattice codes and realising the theoretical promise of performance gains is their decoding complexity. Namely, maximum-likelihood (ML) decoding boils down to closest lattice point search, the complexity of which grows exponentially in the lattice dimension. More efficient methods exist, most prominently sphere decoding [41], which limits the search to a hypersphere of a given radius. However, the complexity remains prohibitive for higher dimensional lattices. To this end, several attempts have been made to reduce the ML decoding complexity. In principle, there are two ways to do this: either one can resort to reduced-complexity decoders yielding suboptimal performance, or try to build the code lattice in such a way that its structure naturally

allows for parallelisation of the decoding process, hence yielding reduction in the dimensionality of the search. In this chapter, we are interested in the latter: we will show how to design codes that inherently yield reduced complexity thanks to a carefully chosen underlying algebraic structure.

On our way to this goal, we will introduce the reader to the basics of lattices and algebraic number theory, to the extent that is relevant to this chapter. We will also lay out the typical channel models for the considered communications settings. Whenever we cannot explain everything in full detail in the interest of space, suitable references will be given for completeness. We assume the reader is familiar with basic abstract algebra and possesses some mathematical maturity, while assuming no extensive knowledge on wireless communications.

The rest of the chapter is organised as follows. We begin in Sect. 3.2 by familiarising the reader with the important notion of lattices and recall related results. Following a section introducing concepts and results from algebraic number theory, we study a particular class of central simple algebras, specifically cyclic division algebras, and their orders. We then move on to providing a background in wireless communications in Sect. 3.3, introducing the well-known multiple-input multiple-output fading channel model and related performance parameters. As a coding technique employed in this multiple-antenna communications setup, we then introduce the main object of this chapter, space–time codes. We recall code design criteria, and furthermore show how codes can be constructed from cyclic division algebras. In Sect. 3.4, maximum-likelihood decoding is introduced, and we discuss a possible decoding complexity reduction by algebraic means, defining the concept of fast-decodable space–time codes. The definition of fast decodability is then further refined, which allows us to consider more specific families of space–time codes with reduced decoding complexity. We further recall a useful iterative method for code construction. Finally, in Sect. 3.5 we discuss two specific communication scenarios as well as explicit methods to construct fast-decodable space–time codes.

3.2 Algebraic Tools for Space–Time Coding

Although space-time codes are primarily a tool for data transmission, they are of a highly mathematical nature. Indeed, design criteria derived for minimising the probability of incorrect decoding, which we will revisit in Sect. 3.3.2.1, can be met by ensuring certain algebraic properties of the underlying structure used for code construction. For this reason, we first devote a chapter to the mathematical notions needed for space–time code analysis and design.

We start with basic concepts and results about lattices, objects which are of particular interest as almost all space–time codes with good performance arise from lattice structures. This is both to ensure a linear structure—a lattice is simply a free $\mathbb{Z}$-module, thus an abelian group—as well as to avoid accumulation points at the receiver, to which end the discreteness property of a lattice is useful. Our main references for all lattice related results are [10, 11].

In a successive section, we then introduce relevant tools and objects from algebraic number theory, such as number fields, their rings of integers, and prime ideal factorisation. These tools will play a crucial role in the construction of space–time codes. As references, we have [28, 29].

Most importantly, we finally introduce central simple algebras and their orders, the main objects that will determine the performance of the constructed codes. Over number fields, every central simple algebra is cyclic, and we study these in detail. We refer to [6, 27, 31] for good general references.

3.2.1 Lattices

We begin with the simplest definition of a lattice in the ambient space $\mathbb{R}^n$.

Definition 3.1 A *lattice* $\Lambda \subset \mathbb{R}^n$ is the $\mathbb{Z}$-span of a set of vectors of $\mathbb{R}^n$ that are linearly independent over $\mathbb{R}$.

Note that we do not require that the number of vectors spanning Λ equals the dimension n. Indeed, any lattice is isomorphic to $\mathbb{Z}^t$ as *groups* for $t \leq n$. A lattice is thus a free abelian group of rank $\mathrm{rk}(\Lambda) = t$, and is called *full-rank* or shortly *full*, if the rank and dimension coincide, i.e., $t = n$. We give an alternative and useful group theoretic definition.

Definition 3.2 A *lattice* $\Lambda \subset \mathbb{R}^n$ is a discrete[1] subgroup of $\mathbb{R}^n$.

A lattice $\Lambda \subseteq \mathbb{R}^n$ can hence be expressed as a set

$$\Lambda = \left\{ \mathbf{x} = \sum_{i=1}^{t} z_i \mathbf{b}_i \,\middle|\, z_i \in \mathbb{Z} \right\},$$

with $\mathbf{b}_i \in \mathbb{R}^n$ (and the z_i uniquely determined by x). We say that $\{\mathbf{b}_1, \ldots, \mathbf{b}_t\}$ forms a $\mathbb{Z}$-basis of Λ.

We can conveniently define a *generator matrix* and the corresponding *Gram matrix* for Λ

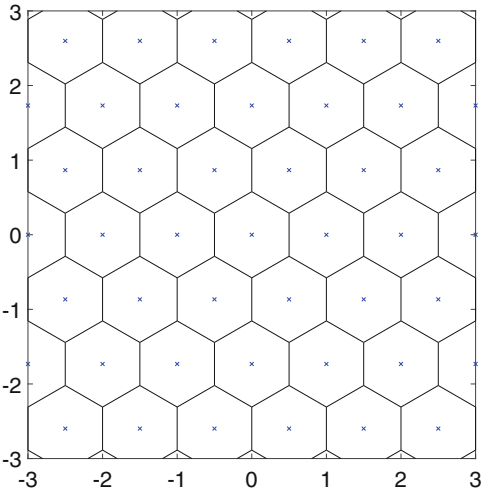
$$M_\Lambda = \left[\mathbf{b}_1 \cdots \mathbf{b}_n\right]; \quad G_\Lambda = M_\Lambda^t M_\Lambda,$$

so that every element of Λ can be expressed as $\mathbf{x} = M_\Lambda \mathbf{z}$ for some $\mathbf{z} \in \mathbb{Z}^n$.

Example 3.1 The simplest lattice is the integer lattice $\mathbb{Z}^n$ in arbitrary dimension $n \geq 1$. A generator and Gram matrix for $\mathbb{Z}^n$ is simply the $n \times n$ identity matrix.

A more interesting example in dimension $n = 2$ is the *hexagonal lattice* A_2. A $\mathbb{Z}$-basis for this lattice can be taken to be $\mathbf{b}_1 = (1, 0)^t$ and $\mathbf{b}_2 = (-1/2, \sqrt{3}/2)^t$. A

[1] By discrete, we mean that the metric on $\mathbb{R}^n$ defines the discrete topology on Λ, i.e., any bounded region of $\mathbb{R}^n$ contains only finitely many points of the subgroup.



$$M_{A_2} = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix};$$

$$G_{A_2} = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

Fig. 3.1 The Voronoi regions of the hexagonal lattice A_2

graphical representation of the lattice, as well as a generator and Gram matrix with respect to this basis are presented in Fig. 3.1.

To each lattice Λ, we can associate its *fundamental parallelotope*, defined as $\mathcal{P}_\Lambda := \{M_\Lambda \mathbf{y} | \, \mathbf{y} \in [0, 1)^n\}$. Note that we can recover $\mathbb{R}^n$ as a disjoint union of the sets $\mathbf{x} + \mathcal{P}_\Lambda$ for all $\mathbf{x} \in \Lambda$. Since M_Λ contains a $\mathbb{Z}$-basis of Λ, any change of basis is obtained via an integer matrix with determinant ± 1. Hence, the Lebesgue measure of $\mathcal{P}_\Lambda$ is invariant under change of basis. Thus, we define the *volume* of a lattice $\Lambda \subset \mathbb{R}^n$ as the Lebesgue measure of its fundamental parallelotope,

$$\mathrm{vol}\,(\Lambda) := \mathrm{vol}\,(\mathcal{P}_\Lambda) = \sqrt{\det(G_\Lambda)}.$$

We have defined a lattice to be a discrete subgroup of $\mathbb{R}^n$ and they are, by definition, free $\mathbb{Z}$-modules. It is however possible and often desirable to extend the definition to other rings and ambient spaces, such as the ring of integers of a number field, or an order in a cyclic division algebra. In this more general context, we define a lattice Λ to be a discrete and finitely generated abelian subgroup of a real or complex ambient space V. In the previous derivations, we have set $V = \mathbb{R}^n$. Of interest for purposes of space–time coding is to consider lattices in $V = \mathrm{Mat}(n, \mathbb{C})$. In this case, we can also identify a full lattice in V with a full lattice in $\mathbb{R}^{2n^2}$ via the $\mathbb{R}$-linear isometry

$$\begin{aligned} &\iota : \mathrm{Mat}(n, \mathbb{C}) \to \mathbb{R}^{2n^2}, \\ &[\mathbf{u}_1, \ldots, \mathbf{u}_n] \mapsto (\mathrm{Re}(u_{11}), \mathrm{Im}(u_{11}), \ldots, \mathrm{Im}(u_{1n}), \ldots, \mathrm{Re}(u_{nn}), \mathrm{Im}(u_{nn}))^t . \end{aligned} \tag{3.1}$$

We have $\|U\|_F = \|\iota(U)\|$, where $\|\cdot\|$ (resp. $\|\cdot\|_F$) denotes the Euclidean (resp. Frobenius) norm, and ι maps full lattices in V to full lattices in the target Euclidean space. This map will be crucial for decoding considerations in later sections.

Let $\Lambda \subset \mathrm{Mat}(n, \mathbb{C})$ be a full lattice with $\mathbb{Z}$-basis $\{B_1, \ldots, B_n\}$, $B_i \in \mathrm{Mat}(n, \mathbb{C})$. A generator matrix and the corresponding Gram matrix for Λ can be given as

$$M_\Lambda = \left[\iota(B_1) \cdots \iota(B_n)\right]; \quad G_\Lambda = M_\Lambda^t M_\Lambda = \left(\mathrm{Re}(\mathrm{Tr}(B_i^\dagger B_j))\right)_{i,j}.$$

The volume of Λ is the volume of the corresponding lattice $\iota(\Lambda)$ in $\mathbb{R}^{2n^2}$, i.e., $\mathrm{vol}(\Lambda) = \sqrt{\det(G_\Lambda)}$.

Example 3.2 We exemplify the notion of a lattice in $\mathrm{Mat}(n, \mathbb{C})$ and corresponding vectorisation on the famous Alamouti code [1]. As we shall see later, the Alamouti code is constructed from a lattice in $\mathrm{Mat}(2, \mathbb{C})$ corresponding to Hamiltonian (or more precisely Lipschitz) quaternions. More concretely, it is a finite subset

$$\mathcal{X}_A \subset \left\{ \begin{bmatrix} x_1 + ix_2 & -(x_3 - ix_4) \\ x_3 + ix_4 & x_1 - ix_2 \end{bmatrix} \middle| (x_1, \ldots, x_4) \in \mathbb{Z}^4 \right\}.$$

A basis of the underlying lattice Λ_A consists of the matrices

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad B_2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}; \quad B_3 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}; \quad B_4 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

Using the defined isometry ι, we can identify Λ_A with a lattice in $\mathbb{R}^8$, which we again denote by Λ_A, with generator and Gram matrix

$$M_{\Lambda_A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}; \qquad G_{\Lambda_A} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

The volume of this lattice is $\mathrm{vol}(\Lambda_A) = \sqrt{\det(G_{\Lambda_A})} = 4$.

3.2.2 Algebraic Number Theory

In this section, we recall fundamental notions from algebraic number theory which are indispensable for space–time code constructions. We assume that the reader is familiar with basic Galois theory.

Let L/K be an arbitrary field extension. If we view L as a vector space over K, we can define the degree of the field extension as the vector space dimension, that is, $[L : K] := \dim_K(L)$. If the degree is finite, we call the extension finite. An element $\alpha \in L$ is called *algebraic* over K if there exists a non-zero polynomial $f(x) \in K[x]$ such that $f(\alpha) = 0$, and the field extension L/K is called *algebraic* if all elements of L are algebraic over K. Consider the homomorphism $\phi : K[x] \to L,\ f(x) \mapsto f(\alpha)$. Since α is algebraic, $\ker(\phi) \neq 0$, and can be generated by a single polynomial $m_{K,\alpha}(x)$, chosen to be monic of smallest degree admitting α as a root. We call this unique polynomial the *minimal polynomial* of α over K. When $K = \mathbb{Q}$ or when the field is clear from context, we may shortly denote $m_\alpha(x)$.

Definition 3.3 An *algebraic number field* is a finite extension of $\mathbb{Q}$.

Example 3.3 The simplest example of a field extension over $\mathbb{Q}$ is the Gaussian field $\mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$, where $i = \sqrt{-1}$ is the imaginary unit. The minimal polynomial of $i \in \mathbb{C}$ over $\mathbb{Q}$ is given by $m_i(x) = x^2 + 1$.

We will henceforth consider L/K to be an extension of algebraic number fields. In the above example, we constructed the field $\mathbb{Q}(i)$ by *adjoining* an algebraic element $i \in \mathbb{C}$ to $\mathbb{Q}$. By the notation $\mathbb{Q}(i)$ we hence mean the smallest field that contains both $\mathbb{Q}$ and i. This is a more general phenomenon.

Theorem 3.1 (Primitive Element Theorem) *Let L/K be an extension of number fields. Then, there exists an element $\alpha \in L$ such that $L = K(\alpha)$.*

We see that we can construct the field L by adjoining the algebraic element $\alpha \in L$ to K and, since $m_{K,\alpha}(x)$ is irreducible, we have the isomorphism

$$L \cong K[x]/\langle m_{K,\alpha}(x)\rangle.$$

It now becomes apparent that the degree of the field extension equals the degree of the minimal polynomial of the adjoined element, $[L : K] = \deg(m_{K,\alpha}(x))$.

Example 3.4 Consider the number field $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. We claim that $K = \mathbb{Q}(\sqrt{2} + \sqrt{3})$ and is hence generated by a single element. The inclusion $\mathbb{Q}(\sqrt{2} + \sqrt{3}) \subseteq K$ is trivial, as $\sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$. For the reverse inclusion, it suffices to express $\sqrt{2}$ and $\sqrt{3}$ in terms of elements of $\mathbb{Q}(\sqrt{2} + \sqrt{3})$. Note that as $(\sqrt{2} + \sqrt{3})^2 = 5 + 2\sqrt{6}$ it follows that $\sqrt{6} \in \mathbb{Q}(\sqrt{2} + \sqrt{3})$, and we have

$$\sqrt{2} = \frac{2 + \sqrt{6}}{\sqrt{2} + \sqrt{3}}; \quad \sqrt{3} = \frac{3 + \sqrt{6}}{\sqrt{2} + \sqrt{3}}.$$

This shows that $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. The minimal polynomial of $\alpha := \sqrt{2} + \sqrt{3}$ is $m_\alpha(x) = x^4 - 10x^2 + 1$, and we see that $\mathbb{Q}(\alpha)$ is an extension of degree 4.

We now define a very important ring associated with a number field K.

Definition 3.4 Let K be a number field. The integral closure of $\mathbb{Z}$ in K consists of all the elements $\alpha \in K$ for which $m_\alpha(x) \in \mathbb{Z}[x]$. The integral closure is a ring, called the *ring of integers* O_K of K. We call any element $\alpha \in O_K$ an *algebraic integer*.

Example 3.5 Consider the field extension $\mathbb{Q}(i)/\mathbb{Q}$. The ring of integers of $\mathbb{Q}(i)$ is precisely $\mathbb{Z}[i]$. It is however not always true that $O_{K(\alpha)} = \mathbb{Z}[\alpha]$. Consider for example $\mathbb{Q}(\sqrt{5})/\mathbb{Q}$. We have that $\mathbb{Z}[\sqrt{5}]$ is composed of algebraic integers, but $\mathbb{Z}[\sqrt{5}] \neq O_K$. For example, the element $\frac{1+\sqrt{5}}{2}$ is a root of the polynomial $x^2 - x - 1$, but $\frac{1+\sqrt{5}}{2} \notin \mathbb{Z}[\sqrt{5}]$. In fact, it turns out that $O_K = \mathbb{Z}\left[\frac{1+\sqrt{5}}{2}\right]$.

As we have seen, $\alpha \in K$ is an algebraic integer if and only if $m_\alpha(x) \in \mathbb{Z}[x]$. Further, the field of fractions of O_K is precisely K. In the above examples, the ring of integers $O_K = \mathbb{Z}[\theta]$ admits a $\mathbb{Z}$-basis $\{1, \theta\}$. In fact, we have the following result.

Theorem 3.2 *Let K be a number field of degree n. The ring of integers O_K of K is a free $\mathbb{Z}$-module of rank n.*

As a consequence, the ring of integers O_K admits an integral basis over $\mathbb{Z}$, that is, a basis as a $\mathbb{Z}$-module. Given an extension L/K of number fields, it is however not true in general that the ring of integers O_L is a free O_K-module. This holds for instance if O_K is a principal ideal domain (PID). We will be considering extensions of $\mathbb{Q}$ and $\mathbb{Q}(i)$, hence circumventing this problem.[2]

Consider a number field K of degree n over $\mathbb{Q}$. We fix compatible embeddings of K into $\mathbb{C}$, and identify the field with its image under these embeddings. More precisely, there exist exactly n pairwise distinct embeddings (i.e., injective ring homomorphisms) $\sigma_i : K \to \mathbb{C}$, forming the set $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C}) = \{\sigma_1, \ldots, \sigma_n\}$.

We split the embeddings into those whose image is real or complex, respectively. More concretely, let $\sigma_1, \ldots, \sigma_r : K \to \mathbb{R}$, and $\sigma_{r+1}, \ldots, \sigma_n : K \to \mathbb{C}$. Note that the embeddings with complex image come in conjugate pairs, of which there are exactly $s := \frac{n-r}{2}$. We call the tuple (r, s) the *signature* of the number field K.

We can use the embeddings to define two important functions, namely the norm and trace of elements in K. For each $\alpha \in K$, consider the induced $\mathbb{Q}$-linear homomorphism $\varphi_\alpha : K \to K$, where for all $\beta \in K$, we have $\varphi_\alpha(\beta) = \alpha\beta$. By fixing a basis of K over $\mathbb{Q}$, φ_α can be represented by a matrix $A_\alpha \in \mathrm{Mat}(n, \mathbb{Q})$. This is referred to as the *left regular representation*.

Definition 3.5 Let K be a number field of degree n, and let $\alpha \in K$. The *norm* and *trace* of α, respectively, are defined as

$$\mathrm{Nm}_K(\alpha) = \det(A_\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha); \qquad \mathrm{Tr}_K(\alpha) = \mathrm{Tr}(A_\alpha) = \sum_{i=1}^{n} \sigma_i(\alpha).$$

These definitions are independent of the choice of a basis for A_α.

[2] The practical reason behind this choice is that the popular modulation alphabets, referred to as pulse amplitude modulation (PAM) and quadrature amplitude modulation (QAM), correspond to the rings of integers of these fields.

We note that the norm and trace are generally rational elements. When $\alpha \in O_K$, however, we have $\mathrm{Nm}_K(\alpha), \mathrm{Tr}_K(\alpha) \in \mathbb{Z}$.

Definition 3.6 Let K be a number field of degree n, with ring of integers O_K, and let $\{b_1, \ldots, b_n\}$ be an integral basis of O_K. The *discriminant* of K is the well-defined integer

$$d_K = \det\left(\begin{bmatrix} \mathrm{Tr}_K(b_1 b_1) & \cdots & \mathrm{Tr}_K(b_1 b_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}_K(b_n b_1) & \cdots & \mathrm{Tr}_K(b_n b_n) \end{bmatrix}\right)$$
$$= \det\left(\begin{bmatrix} \sigma_1(b_1) & \cdots & \sigma_1(b_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(b_1) & \cdots & \sigma_n(b_n) \end{bmatrix}\right)^2.$$

The determinants above can indeed be shown to be equal. The discriminant d_K is independent of the choice of basis, and hence an invariant of the number field.

Example 3.6 Consider the number field $K = \mathbb{Q}(\sqrt{-5})$, with ring of integers $O_K = \mathbb{Z}[\sqrt{-5}]$. As K is a degree-2 extension of $\mathbb{Q}$, and generated by a complex element, we have that its signature is $(r, s) = (0, 1)$. A representative of the pair of complex embeddings is given by $\sigma_1 : \sqrt{-5} \mapsto -\sqrt{-5}$, and the complex conjugate σ_2 is simply the identity.

Given an element $\alpha = x_0 + \sqrt{-5}x_1 \in K$, the norm and trace of α can be computed to be

$$\mathrm{Nm}_K(\alpha) = \sigma_1(\alpha)\sigma_2(\alpha) = x_0^2 + 5x_1^2; \quad \mathrm{Tr}_K(\alpha) = \sigma_1(\alpha) + \sigma_2(\alpha) = 2x_0.$$

Moreover, we can compute the discriminant of K by choosing a basis $\left\{1, \sqrt{-5}\right\}$ of O_K and computing the determinant

$$d_K = \det\left(\begin{bmatrix} 1 & -\sqrt{-5} \\ 1 & \sqrt{-5} \end{bmatrix}\right)^2 = -20.$$

The motivation for studying number fields has its origins in the factorisation of integers into primes. In the ring $\mathbb{Z}$, prime and irreducible elements coincide, and every natural number factors uniquely into prime numbers. By generalising the ring $\mathbb{Z}$ to the ring of integers O_K of a number field, unique factorisation into prime elements is no longer guaranteed. However, the underlying structure of the ring O_K allows for a generalisation of unique factorisation by making use of ideals, instead of elements.

Let K be a number field of degree n, and $\mathfrak{a} \subset \mathcal{O}_K$ a non-zero ideal. Then $\mathfrak{a}$ factors into a product of prime ideals, unique up to permutation,

$$\mathfrak{a} = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i},$$

where $e_i > 0$. We define the *norm* of the ideal $\mathfrak{a}$ as the cardinality of the finite ring $\mathrm{N}(\mathfrak{a}) := |\mathcal{O}_K/\mathfrak{a}|$. The ideal norm extends multiplicatively, and moreover $\mathrm{N}(\mathfrak{a}) \in \mathfrak{a}$. Consequently, if $\mathrm{N}(\mathfrak{a})$ is prime, then $\mathfrak{a}$ is a prime ideal. More importantly, if $\mathrm{N}(\mathfrak{a}) = p_1^{e_1} \cdots p_k^{e_k}$ is the prime factorisation, then (as we can show that $\mathfrak{a}$ divides $\mathrm{N}(\mathfrak{a})\mathcal{O}_K$) it is clear that every prime divisor of $\mathfrak{a}$ is a prime divisor of $p_i\mathcal{O}_K$ for some i.

Remark 3.1 If all prime divisors of $p\mathcal{O}_K$ are known for all primes $p \in \mathbb{Z}$, then all ideals of $\mathcal{O}_K$ are known.

Let $\mathfrak{p} \subset \mathcal{O}_K$ be a prime ideal. Then $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$, p prime. We can hence write

$$p\mathbb{Z} = \mathfrak{p}^e \mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_k^{e_k}$$

for $\mathfrak{p}_i$ distinct prime ideals of $\mathcal{O}_K$. The number $e = e(\mathfrak{p}/p\mathbb{Z})$ is referred to as the *ramification index* of $p\mathbb{Z}$ at $\mathfrak{p}$. We further define the *residue class degree* of $\mathfrak{p}/p\mathbb{Z}$ as the integer $f \geq 1$ which satisfies $\mathrm{N}(\mathfrak{p}) = p^f$.

Example 3.7 Consider $K = \mathbb{Q}(i)$, and let $p > 2$ be a rational prime. We want to study the factorisation of p in $\mathcal{O}_K = \mathbb{Z}[i]$. We have the following isomorphisms:

$$\mathbb{Z}[i]/\langle p \rangle \cong \mathbb{Z}[x]/\langle p, x^2 + 1 \rangle \cong \mathbb{F}_p[x]/\langle x^2 + 1 \rangle$$

By norm considerations, as $\mathrm{N}(p\mathbb{Z}[i]) = |\mathbb{Z}[i]/\langle p \rangle| = |\mathbb{F}_p[x]/\langle x^2 + 1 \rangle| = p^2$, we have that p can either remain prime in $\mathbb{Z}[i]$, or be the product of two prime ideals. On the other hand, we know that $p\mathbb{Z}[i]$ is prime if and only if $\mathbb{Z}[i]/\langle p \rangle$ is a field. In fact,

$$\mathbb{Z}[i]/\langle p \rangle \cong \mathbb{Z}[x]/\langle p, x^2 + 1 \rangle \cong \mathbb{F}_p[x]/\langle x^2 + 1 \rangle,$$

so that the residue class degree is $f = 2$. This quotient is a field precisely when $x^2 + 1$ is irreducible. This is the case for $p \not\equiv 1 \bmod 4$.

For the case $p \equiv 1 \bmod 4$, we can factor $x^2 + 1 = (x - a)(x - b)$, and we get a factorisation $p\mathbb{Z}[i] = (i - a)(i - b)$.

3.2.3 Central Simple Algebras

Let K be a field, and $\mathcal{A} \supseteq K$ a finite-dimensional associative K-algebra, i.e., a finite-dimensional K-vector space and a ring together with a K-bilinear product.

The algebra is *simple*, if it contains no non-trivial two-sided ideals, and moreover *central* if its centre is precisely K. The algebra is a *division algebra* if all of its non-zero elements are invertible. We have the following important theorem, which is a simplified version of a more general result.

Theorem 3.3 (Wedderburn) *Every central simple K-algebra is isomorphic to* $\mathrm{Mat}(n, D)$ *for some uniquely determined n and some division K-algebra D, unique up to isomorphism.*

If $\mathcal{A}$ is a central simple K-algebra and D is the division algebra from the above theorem, we denote by $\mathrm{ind}(\mathcal{A}) = \sqrt{[D : K]}$ the *index*, and by $\deg(\mathcal{A}) = \sqrt{[\mathcal{A} : K]}$ the *degree* of the algebra. $\mathcal{A}$ is a division algebra if and only if $\mathrm{ind}(\mathcal{A}) = \deg(\mathcal{A})$.

If $\mathcal{A}$ is a finite-dimensional central simple algebra over a field K, then $\mathcal{A}$ is said to be cyclic if it contains a strictly maximal subfield L such that L/K is a cyclic field extension, i.e., the Galois group is a cyclic group. If K is a number field, every K-central simple algebra is *cyclic*, and vice versa. This family of central simple algebras has been widely used for space–time coding since the work [34]. We start with the special case of cyclic algebras of degree 2, also known as *quaternion algebras*.

Definition 3.7 Let K be a field, and $a, \gamma \in K^\times$ not necessarily distinct. A *quaternion algebra* $(a, \gamma)_K$ is a K-central algebra defined as

$$(a, \gamma)_K := \{x = x_0 + ix_1 + jx_2 + kx_3 | \, x_i \in K\},$$

where the basis elements satisfy the rules

$$i^2 = a, \quad j^2 = \gamma, \quad ij = -ji = k.$$

Example 3.8 The most famous example is the set of *Hamiltonian quaternions*, which can be defined as $\mathbb{H} = (-1, -1)_{\mathbb{R}}$. An element $x \in \mathbb{H}$ is of the form $x = x_0 + ix_1 + jx_2 + kx_3$ with $(x_0, x_1, x_2, x_3) \in \mathbb{R}^4$, $i^2 = j^2 = -1$ and $ij = -ji = k$.

For quaternion algebras, we have the following deep and important classification result.

Theorem 3.4 *Let $(a, \gamma)_K$ be a quaternion algebra. We have two possibilities.*

(a) $(a, \gamma)_K$ *is a division algebra.*
(b) $(a, \gamma)_K \cong \mathrm{Mat}(2, K)$.

We can determine which of the cases apply by means of a simple quaternary quadratic form. To be more precise, consider an element $x = x_0 + ix_1 + jx_2 + kx_3 \in (a, \gamma)_K$, and define the *norm* of x as

$$\mathrm{Nm}(x) = xx^* = x_0^2 - ax_1^2 - \gamma x_2^2 + a\gamma x_3^2,$$

where $x^* = x_0 - ix_1 - jx_2 - kx_3$ is the conjugate of x. Then, the quaternion algebra $(a, \gamma)_K$ is division if and only if $\mathrm{Nm}(x) = 0$ implies $x = 0$.

Example 3.9 Recall the set of Hamiltonian quaternions $\mathbb{H}$. The norm of an element $x = x_0 + ix_1 + jx_2 + kx_3 \in \mathbb{H}$ is $\mathrm{Nm}(x) = x_0^2 + x_1^2 + x_2^2 + x_3^2 \geq 0$. As $x_i \in \mathbb{R}$, we have equality if and only if $x = 0$. Hence, $\mathbb{H}$ is a division algebra.

A quaternion algebra is a degree-4 vector space over the centre K. They are a special case of the more general cyclic algebras, a family of central simple algebras which we study in the sequel.

Definition 3.8 Let L/K be a degree-n cyclic Galois extension of number fields, and denote by $\langle\sigma\rangle = \mathrm{Gal}\,(L/K)$ its Galois group. A *cyclic algebra* is a tuple

$$C = (L/K, \sigma, \gamma) := \bigoplus_{i=0}^{n-1} e^i L,$$

where $e^n = \gamma \in K^\times$ and multiplication satisfies $le = e\sigma(l)$ for all $l \in L$.

The algebra C is K-central simple, and is called a *cyclic division algebra* if it is division.

The usefulness of cyclic division algebras for purposes of space–time coding starts with the existence of a matrix representation of elements of the algebra. To be more precise, each element $x = \sum_{i=0}^{n-1} e^i x_i \in C$ induces for all $y \in C$ a right L-linear map $\rho : y \mapsto xy$, which is referred to as the *left-regular representation* of the algebra, and describes left multiplication with x. We can define a matrix associated with ρ, given by

$$x \mapsto \rho(x) := \begin{bmatrix} x_0 & \gamma\sigma(x_{n-1}) & \gamma\sigma^2(x_{n-2}) & \cdots & \gamma\sigma^{n-1}(x_1) \\ x_1 & \sigma(x_0) & \gamma\sigma^2(x_{n-1}) & & \gamma\sigma^{n-1}(x_2) \\ \vdots & & \vdots & & \vdots \\ x_{n-2} & \sigma(x_{n-3}) & \sigma^2(x_{n-4}) & & \gamma\sigma^{n-1}(x_{n-1}) \\ x_{n-1} & \sigma(x_{n-2}) & \sigma^2(x_{n-3}) & \cdots & \sigma^{n-1}(x_0) \end{bmatrix}.$$

Example 3.10 Let us consider again the Hamiltonian quaternions. Using the above notation, we write $e = j$ and

$$\mathbb{H} = (\mathbb{C}/\mathbb{R}, \sigma = *, \gamma = -1) = \mathbb{C} \oplus j\mathbb{C},$$

with $cj = jc^*$ for all $c \in \mathbb{C}$ and $j^2 = \gamma = -1$. Note that we have intentionally chosen to represent $\mathbb{H}$ as a right vector space in order to be compatible with the left regular representation.

Let now $x = x_0 + jx_1$ with $x_0, x_1 \in \mathbb{C}$. If we multiply the basis elements $\{1, j\}_{\mathbb{C}}$ from the left by x, we get

$$\begin{aligned} x \cdot 1 &= x_0 + jx_1\,, \\ x \cdot j &= (x_0 + jx_1)j = x_0 j + jx_1 j = jx_0^* + j^2 x_1^* = -x_1^* + jx_0^*\,. \end{aligned}$$

In a matrix form, we have

$$x \mapsto \rho(x) = \begin{bmatrix} x_0 & -x_1^* \\ x_1 & x_0^* \end{bmatrix}.$$

Note that this matrix corresponds to the Alamouti code.

Example 3.11 Let L/K be a number field extension of degree $n = 3$. Then, we can pick a basis $\left\{1, e, e^2\right\}$ of a cyclic algebra C over its maximal subfield L, where $e^3 = \gamma \in K^\times$. Let $x = x_0 + ex_1 + e^2 x_2$, and consider left multiplication. Similarly as above,

$$\begin{aligned} x \cdot 1 &= x_0 + ex_1 + e^2 x_2\,, \\ x \cdot e &= (x_0 + ex_1 + e^2 x_2)e = e\sigma(x_0) + e^2\sigma(x_1) + e^3\sigma(x_2) \\ &= \gamma\sigma(x_2) + e\sigma(x_0) + e^2\sigma(x_1)\,, \\ x \cdot e^2 &= (x_0 + ex_1 + e^2 x_2)e^2 = e^2\sigma(x_0) + e^3\sigma(x_1) + e^4\sigma(x_2) \\ &= \gamma\sigma(x_1) + \gamma e\sigma(x_2) + e^2\sigma(x_0)\,. \end{aligned}$$

We see that in this basis, left multiplication by x can be represented by the matrix

$$\rho(x) = \begin{bmatrix} x_0 & \gamma\sigma(x_2) & \gamma\sigma^2(x_1) \\ x_1 & \sigma(x_0) & \gamma\sigma^2(x_2) \\ x_2 & \sigma(x_1) & \sigma^2(x_0) \end{bmatrix}$$

We close this section by recalling how to ensure that a cyclic algebra $(L/K, \sigma, \gamma)$ is division by means of the element $\gamma \in K^\times$. The result is a simple corollary to a result due to A. Albert.

Theorem 3.5 *Let $C = (L/K, \sigma, \gamma)$ be a cyclic algebra. If $\gamma^{n/p}$ is not a norm of some element of L for all prime divisors p of n, then C is division.*

3.2.3.1 Orders

Given a number field K, the collection of integral elements form the ring of integers O_K of K. This ring is the unique *maximal order* of K, a concept which we will now recall in a more general context.

Definition 3.9 Let $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra. An O_K*-order* Γ in C is a subring of C sharing the same identity as C and such that Γ is a finitely generated O_K-module which generates C as a linear space over K.

An order is *maximal* if it is not properly contained in another order of C.

Every order of a cyclic division algebra is contained in a maximal order. Within a number field K, the ring of integers O_K is integrally closed and the unique maximal order of K. In general, a maximal order Γ of C is not integrally closed, and a division algebra C may contain multiple maximal orders. In contrast, the following special order is often of interest due to its simple structure. It is in fact the initial source for space–time codes with non-vanishing determinants.

Definition 3.10 Let $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra. The *natural order* of C is the O_K-module

$$\Gamma_{\text{nat}} := \bigoplus_{i=0}^{n-1} e^i O_L.$$

Note that Γ_{nat} is not closed under multiplication unless $\gamma \in O_K$.

Remark 3.2 Given a cyclic division algebra $C = (L/K, \sigma, \gamma)$ and an element $c \in \Gamma$, where $\Gamma \subset C$ is an order, we can define concepts like the *reduced norm* $\text{nm}(c) = \det(\rho(c))$ and *reduced trace* $\text{tr}(c) = \text{Tr}(\rho(c))$. These are elements of the ring of integers of the centre, i.e., $\text{nm}(c), \text{tr}(c) \in O_K$. Consequently, for $K = \mathbb{Q}$ or K imaginary quadratic, we have $|\text{nm}(c)| \geq 1$ for any non-zero $c \in \Gamma$, an observation which is crucial for achieving the *non-vanishing determinant* property (cf. Sect. 3.3.2.1).

3.3 Physical Layer Communications

In this section, we study the characteristics and properties of a wireless channel, discussing various methods for combating the effects of fading and noise.

3.3.1 Rayleigh Fading MIMO Channel

In a wireless environment, in contrast to wired channels, the signal traverses several different paths between a transmitter and receiver. Consequently, different versions

of the signal distorted by (independent) environmental effects will come together at the receiver, causing a superimposed channel output. Together with dissipation effects caused by urban scatterers as well as interference, the signal experiences *fading*, and various statistical models exist to describe these phenomena. Here, we consider the widely used *Rayleigh fading* channel model. In addition, thermal noise at the receiver further distorts the channel output.

To be more precise, assume a single source, equipped with $n_t \geq 1$ transmit antennas, and a single destination, with $n_r \geq 1$ receive antennas. If $n_t, n_r \geq 2$ we refer to the setup as the *multiple-input multiple-output* (MIMO) model, while the case $(n_t, n_r) = (1, 1)$ is termed the *single-input single-output* (SISO) channel model. The mixed cases $(n_t = 1, n_r > 1)$ and $(n_t > 1, n_r = 1)$ are the SIMO and MISO channel setups, respectively.

Consider a channel between n_t transmit antennas and n_r receive antennas. The wireless channel is modelled by a random matrix

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n_t} \\ h_{21} & h_{22} & & h_{2n_t} \\ \vdots & & \ddots & \vdots \\ h_{n_r 1} & h_{n_r 2} & \cdots & h_{n_r n_t} \end{bmatrix} \in \mathrm{Mat}(n_r \times n_t, \mathbb{C}),$$

We assume that the channel remains static for $T \geq n_t$ time slots and then changes independently of its previous state, and refer to T as the *channel delay* or *channel coherence time*. Each of the entries h_{ij} of H denotes the path gain from transmit antenna j to receive antenna i. They are modelled as complex variables with i.i.d. normal distributed real and imaginary parts,

$$\mathrm{Re}(h_{ij}), \mathrm{Im}(h_{ij}) \sim \mathcal{N}(0, \sigma_h^2),$$

yielding a Rayleigh distributed envelope

$$|h_{ij}| = \sqrt{\mathrm{Re}(h_{ij})^2 + \mathrm{Im}(h_{ij})^2} \sim \mathrm{Ray}(\sigma_h)$$

with scale parameter σ_h, which gives this fading model its name.

The additive noise is modelled by a matrix $N \in \mathrm{Mat}(n_r \times T, \mathbb{C})$ with i.i.d. complex Gaussian entries with finite variance σ_n^2. To combat the destructive effects of fading, the transmitter encodes its data into a codeword matrix $X \in \mathrm{Mat}(n_t \times T, \mathbb{C})$. Each column $\mathbf{x}_i$ of X corresponds to the signal vector transmitted in the ith time slot, across the available transmit antennas. If we denote each column of the noise matrix N by $\mathbf{n}_i$, the received signal at each time slot $1 \leq i \leq T$ is given by the channel equation

$$\mathbf{y}_i = H\mathbf{x}_i + \mathbf{n}_i.$$

We assume that the destination waits for the T subsequent transmissions before starting any decoding process. As usual, we assume perfect channel state information at the receiver, while the transmitter only has statistical channel information. The channel is supposed to remain fixed during the entire transmission process, and hence we can summarise the overall channel equation in a compact form to read

$$Y = HX + N.$$

Thus, by allowing the use of multiple antennas at the transmitter and/or receiver, we have created *spatial diversity*. By ensuring a separation of the antennas by at least half the used wavelength, the multiple signals will fade independently of each other. On the other hand, the transmission over multiple time slots enables *temporal diversity*, providing copies of the signal at the receiver.

The physical conditions in an actual wireless channel are rapidly changing. Consequently, the comparison in performance of two different codes needs to be considered with respect to a standardised quantity. We define the *signal-to-noise ratio* (SNR) at the receiver as the ratio of the received signal power to noise power, that is,

$$\mathrm{SNR} = \frac{\mathbb{E}\left[\|HX\|^2\right]}{\mathbb{E}\left[\|N\|^2\right]}.$$

3.3.1.1 Performance Parameters of a Wireless Channel

Consider a MIMO channel with n_t transmit antennas and n_r receive antennas. The first quantity that we need to mention is the *capacity* of the channel.

Definition 3.11 Assume that the receiver knows the realisation of the channel matrix H. For a fixed power constraint on the channel input, the *capacity* of a MIMO channel is the upper bound on the mutual information between the channel input and output, given the channel realisation.

As the capacity depends on the channel matrix, it needs to be viewed as a random variable. The ergodic capacity of a MIMO channel is given by

$$C_H = \mathbb{E}_H\left[\log\det\left(I_{n_r} + \frac{\mathrm{SNR}}{n_t}H^{\dagger}H\right)\right].$$

Recently, the authors in [24] gave criteria for algebraic space–time codes from division algebras to achieve the channel capacity up to a constant gap.

Equivalently we can interpret the capacity of the channel as the upper bound on the amount of information that can be transmitted, so that the probability of error can be maintained arbitrarily low. At high SNR, the capacity of the channel scales with the number of antennas. More specifically, an SNR increase of 3 dB results in an increase in capacity by $\min\{n_t, n_r\}$.

We now define two quantities which allow us to compare different coding strategies for the MIMO channel.

Definition 3.12 Consider a MIMO channel.

(i) The *diversity gain* of a coding strategy is the asymptotic slope of the corresponding error probability curve with respect to the SNR in a log − log scale.
(ii) The *coding gain* measures the difference in SNR required for two different full-diversity coding strategies to achieve the same error probability.

3.3.2 Space–Time Codes

This section introduces the main object of the survey: space–time codes. These codes are tailor-made for MIMO communications. We start with basic definitions and relate the enabled spatial and temporal diversity to the matrix structure of space–time codewords.

In the first subsection, the basic code design criteria for minimising the probability of incorrect decoding are derived. While the design criteria are independent of the actual code construction method and hold for any matrix codebook, various results are then introduced exposing how the criteria can be met by purely algebraic means. Hence, it becomes clear which properties the underlying structures should exhibit in order to construct well-performing codes.

After this, we utilise the algebraic tools introduced in Sect. 3.2 in order to construct space–time codes meeting the derived criteria.

3.3.2.1 Design Criteria for Space–Time Codes

Recall the Rayleigh fading $n_t \times n_r$ MIMO channel model with channel coherence time T. We have seen that the codewords X need to be taken from some collection of matrices $\mathcal{X} \subset \mathrm{Mat}(n_t \times T, \mathbb{C})$. Very naively, and this is our first definition, we simply define a code to be a finite collection of such matrices.

Definition 3.13 Let $\mathfrak{C} \subset \mathbb{R}^\times$ be a finite subset and $k \in \mathbb{Z}_+$. A *space–time code* is the image of an injective map $\phi : \mathfrak{C}^k \to \mathrm{Mat}(n_t \times T, \mathbb{C})$.

Having no structure, however, may lead to accumulation of the received signals. To circumvent this problem, forcing a discrete and linear structure on the code is helpful, e.g., a lattice structure. We give the more specialised definition of *linear* space–time codes, which we will consider henceforth.

Definition 3.14 Let $\{B_i\}_{i=1}^k$ be an $\mathbb{R}$-linearly independent set of matrices in $\mathrm{Mat}(n_t \times T, \mathbb{C})$. A *linear space–time block code* of rank k is a set of the form

$$\mathcal{X} = \left\{ \sum_{i=1}^{k} B_i s_i \,\middle|\, s_i \in S \right\},$$

where $S \subset \mathbb{Z}$ is a finite *signalling alphabet*. In relation to the previous definition, we have $\mathcal{X} = \phi(\mathfrak{C}^k)$, where $\mathfrak{C} = S$.

As the matrices $\{B_i\}_{i=1}^k$ form a basis of a *lattice* $\Lambda \subset \mathrm{Mat}(n_t \times T, \mathbb{C})$, $\mathcal{X}$ is called a *space–time lattice code* of rank $k = \mathrm{rk}\,(\Lambda) \leq 2n_tT$, the upper bound being imposed by the $\mathbb{R}$-dimension of $\mathrm{Mat}(n_t \times T, \mathbb{C})$.

We henceforth refer to such a code $\mathcal{X}$ simply as a space–time code. As the transmit power consumption is directly related to the Frobenius norm of the transmitted codeword, the finite codebook is usually carved out to consist of a desired number of lattice elements with smallest possible Frobenius norms.[3]

The *code rate* of $\mathcal{X}$ is defined as $R = k/T$ real symbols per channel use.[4] In the literature, a code is often said to be *full rate* if all available degrees of freedom from the transmitter's point of view are utilised, i.e., $k = 2n_tT$ and $R = 2n_tT/n_t = 2T$. This is a consequence of mainly having considered symmetric square systems, that is, the case $n_t = n_r = T$. Here, we do not restrict to symmetric systems and define full rate as the maximum rate that still maintains the discrete structure at the receiver and allows for linear detection methods such as sphere-decoding [41]. More precisely, for n_r receive antennas we define full rate as $2n_r$. Hence, in order to achieve full rate as defined in this chapter (avoiding accumulation points at the receiver's space), for n_r receive antennas we should choose a lattice of rank $2n_rT$ (instead of $2n_tT$).

Consider a space–time code $\mathcal{X}$, and let $X \in \mathcal{X}$ be the transmitted codeword. A receiver observes its channel output Y and, as it is assumed to know the channel H and the noise is zero-mean, decodes a maximum-likelihood estimate of the transmitted codeword by computing

$$\hat{X} = \arg\min_{X \in \mathcal{X}} \|Y - HX\|_F^2. \tag{3.2}$$

The probability $\mathscr{P}(X \to X')$ that a codeword $X' \neq X$ is decoded when X was transmitted is asymptotically upper bounded with increasing SNR as

$$\mathscr{P}(X \to X') \leq \left(\det\left((X - X')(X - X')^\dagger\right)\mathrm{SNR}^{n_t}\right)^{-n_r}.$$

From this upper bound, two design criteria can be derived [38]. The *diversity gain* of a code as defined above relates to the minimum rank of $(X - X')$ over all

[3]The smallest Frobenius norms correspond to the shortest Euclidean norms of the vectorised matrices. Directly, this would mean spherical constellation shaping. However, it is often more practical to consider a slightly more relaxed cubic shaping. This is the case in particular when the lattice in question is orthogonal, as then the so-called Gray-mapping [13] will give an optimal bit labelling of the lattice points.

[4]In the literature, the code rate is often defined in complex symbols per channel use. We prefer using real symbols, as not every code admits a Gaussian basis, while every lattice has a $\mathbb{Z}$-basis.

pairs of distinct code matrices $(X, X') \in \mathcal{X}^2$. Thus, the minimum rank of $\mathcal{X}$ should satisfy

$$\min_{X \neq X'} \operatorname{rk}\left(X - X'\right) = \min\{n_t, T\} = n_t.$$

A code satisfying this criterion is called a *full-diversity* code.

On the other hand, the *coding gain* can be shown to be proportional to the determinant $\det\left((X - X')(X - X')^\dagger\right)$. As a consequence, the minimum taken over all pairs of distinct codewords,

$$\min_{X \neq X'} \det\left((X - X')(X - X')^\dagger\right),$$

should be as large as possible. For the infinite code

$$\mathcal{X}_\infty = \left\{ \sum_{i=1}^{k} s_i B_i \,\middle|\, s_i \in \mathbb{Z} \right\},$$

we define the *minimum determinant* as the infimum

$$\Delta_{\min}(\mathcal{X}_\infty) := \inf_{X \neq X'} \det\left((X - X')(X - X')^\dagger\right).$$

If $\Delta_{\min}(\mathcal{X}_\infty) > 0$, i.e., the determinants do not vanish as the code size increases, the code is said to have the *non-vanishing determinant* property.

We assume henceforth that the number of transmit antennas and channel delay coincide, $n_t = T =: n$. Given a lattice $\Lambda \subset \mathrm{Mat}(n, \mathbb{C})$, we have by linearity

$$\Delta_{\min}(\Lambda) = \inf_{0 \neq X \in \Lambda} |\det(X)|^2.$$

This implies that any lattice Λ with non-vanishing determinants can be scaled so that $\Delta_{\min}(\Lambda)$ achieves any wanted nonzero value. Consequently, the comparison of two different lattices requires some sort of normalisation. Let Λ be a full lattice with volume $\mathrm{vol}\,(\Lambda)$. The *normalised minimum determinant* and *normalised density* of Λ are the normalised quantities

$$\delta(\Lambda) = \frac{\Delta_{\min}(\Lambda)}{\mathrm{vol}\,(\Lambda)^{\frac{1}{2n}}}; \quad \eta(\Lambda) = \frac{\Delta_{\min}(\Lambda)^{2n}}{\mathrm{vol}\,(\Lambda)},$$

and satisfy the relation $\delta(\Lambda)^2 = \eta(\Lambda)^{\frac{1}{n}}$. Thus, for a fixed minimum determinant, the coding gain can be increased by maximising the density of the code lattice. Or, the other way around, for a fixed volume, the coding gain can be increased by maximising the minimum determinant of the lattice.

3.3.2.2 Constructions from Cyclic Division Algebras

We move on to illustrate how space–time codes satisfying the two introduced criteria can be designed. We begin by ensuring full diversity, to which end the following result is helpful.

Theorem 3.6 ([34, Prop. 1]) *Let K be a field and $\mathcal{D}$ an index-n division K-algebra with a maximal subfield L. Any finite subset $\mathcal{X}$ of the image of a ring homomorphism $\phi : \mathcal{D} \mapsto \mathrm{Mat}(n, L)$ satisfies $\mathrm{rk}\left(X - X'\right) = n$ for any distinct $X, X' \in \mathcal{X}$.*

This leads to a straightforward approach for constructing full-diversity codes, namely by choosing the underlying structure to be a division algebra. In the same article, cyclic division algebras were proposed for code construction as a particular example of division algebras. The ring homomorphism ϕ is the link between the division algebra and a full-diversity space–time code, as we illustrate in the following.

Let $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra of degree n. The left-regular representation $\rho : C \to \mathrm{Mat}(n, \mathbb{C})$ is an injective ring homomorphism (cf. Definition 3.8 and the discussion beneath). We identify elements in C with elements in $\mathrm{Mat}(n, \mathbb{C})$ via ρ. This leads to the following definition.

Definition 3.15 Let C be an index-n cyclic division algebra with left-regular representation $\rho : C \to \mathrm{Mat}(n, \mathbb{C})$. A *space–time code* constructed from C is a finite subset

$$\mathcal{X} \subset \rho(C).$$

To be consistent with Definition 3.14, let $\{B_i\}_{i=1}^{k} \subset \mathrm{Mat}(n, \mathbb{C})$ with $k \leq 2n^2$ be a set of $\mathbb{Q}$-linearly independent matrices in $\rho(C)$. For a fixed signalling alphabet $S \subset \mathbb{Z}$, symmetric around the origin, the space–time code $\mathcal{X}$ is of the form

$$\mathcal{X} = \left\{ \sum_{i=1}^{k} s_i B_i \,\middle|\, s_i \in S \right\}.$$

If C admits a basis over $\mathbb{Z}[i]$, we may also consider the lattice with respect to its $\mathbb{Z}[i]$-basis, and the signalling alphabet will then be a subset in $\mathbb{Z}[i]$.

Note that, given an element $X = \rho(x)$, where $x \in C$, we have that $\det(X) = \det(\rho(x)) \in K$. We can however restrict to certain subrings of the cyclic division algebra, for instance an order Γ. For any $x \in \Gamma$, we have $\det(\rho(x)) \in O_K$. This yields $|\det(\rho(x))| \geq 1$ for $K = \mathbb{Q}$ or K an imaginary quadratic number field. Then, we can consider finite subsets of $\rho(\Gamma)$ as space–time lattice codes guaranteeing non-vanishing determinants (cf. Remark 3.2).

Example 3.12 Consider a MIMO system with $n = n_t = T = 2$, and consider the index-2 number field extension $L/K = \mathbb{Q}(i, \sqrt{5})/\mathbb{Q}(i)$. The ring of integers of L is $O_L = \mathbb{Z}[i, \theta]$ with $\theta = \frac{1+\sqrt{5}}{2}$, and we pick the relative integral basis $\{1, \theta\}$ of

$\mathcal{O}_L$ over $\mathcal{O}_K = \mathbb{Z}[i]$. The *Golden code* [5] is constructed from the cyclic division algebra

$$\mathcal{G} = (L/K, \sigma, \gamma) \cong (5, \gamma)_{\mathbb{Q}(i)}$$

with $\sigma : \sqrt{5} \mapsto -\sqrt{5}$ and $\gamma \in \mathbb{Q}(i)$ non-zero and such that $\gamma \neq \mathrm{Nm}_{L/K}(l)$ for any $l \in L$. We pick $\gamma = i$, leading to a (left regular) matrix representation of $\mathcal{G}$ of the form

$$\begin{aligned} X = \rho(x) &= \begin{bmatrix} x_0 + \theta x_1 & i(x_2 + \sigma(\theta)x_3) \\ x_2 + \theta x_3 & x_0 + \sigma(\theta)x_1 \end{bmatrix} \\ &= x_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x_1 \begin{bmatrix} \theta & 0 \\ 0 & \sigma(\theta) \end{bmatrix} + x_2 \begin{bmatrix} 0 & i \\ 1 & 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 & i\sigma(\theta) \\ \theta & 0 \end{bmatrix}, \end{aligned}$$

where $x_i \in K$.

The algebra $\mathcal{G}$ is a division algebra by Theorem 3.5, so that the Golden code is indeed a full-diversity space–time code. Moreover, by restricting the codewords to the natural order Γ_{nat} by choosing $x_i \in \mathbb{Z}[i]$ guarantees the non-vanishing determinant property (cf. Remark 3.2).

The actual Golden code lattice is a twisted version of $\rho(\Gamma_{nat})$ in order to get an orthogonal lattice. The twisting does not affect the normalised minimum determinant.

3.4 Codes with Reduced ML Decoding Complexity

Using multiple antennas for increased diversity—and additionally enabling temporal diversity—comes at the cost of a higher complexity in decoding. The worst-case complexity of maximum-likelihood (ML) decoding is upper bounded by that of exhaustive search, and is often computationally too expensive for practical use for higher-dimensional code lattices. A fast-decodable space–time code is, in colloquial terms, simply a space–time code whose worst-case ML decoding complexity is lower than that of exhaustive search.

Yet, independently of the actual decoder used, the ML decoding complexity of a space–time code can sometimes be reduced by algebraic means, allowing for parallelisation in the ML search. If the underlying code lattice is of rank k, this requires in principle joint decoding of k information symbols. One way to achieve fast-decodability (this is also how we define fast decodability more formally below) is then to reduce the dimensionality of the (e.g., sphere) decoder, that is, to enable parallelisation where each parallel set contains less than k symbols to be jointly decoded.

In this section we introduce the technique of ML decoding and revise criteria for a space–time code to be fast-decodable. We further specify different families of fast-decodable codes and study their potential decoding complexity reduction.

3.4.1 Maximum-Likelihood Decoding

In the previous sections, we have seen what properties a space–time code should exhibit to potentially ensure a reasonable performance, at least in terms of reliability. There are however more aspects of the communication process which need to be taken into consideration. Orthogonal lattices allow for efficient encoding of the information symbols and bit-labelling of the codewords, while not necessarily yielding the best possible error performance. On the other hand, a too complicated lattice structure makes it more complex to encode a signal in the first place, and may require brute force bit labelling of the codewords.

On the receiver's side, the structure of the code lattice determines the complexity of the decoding process. Indeed, as already mentioned, the major bottleneck in effective implementation of algebraic space–time codes has been their decoding complexity. The concept of fast-decodability was introduced in [9] in order to address the possibility for reducing the dimensionality of the ML decoding problem (cf. (3.2)) without having to resort to suboptimal decoding methods.

Given a finite signalling alphabet $S \subset \mathbb{Z}$, the ML decoding complexity of a rank-k space–time code $\mathcal{X}$ is defined as the minimum number of values that have to be computed for finding the solution to (3.2). The upper bound is the worst-case decoding complexity that we denote by $\mathfrak{D}(S)$, which for its part is upper bounded by the exhaustive search complexity, $\mathfrak{D}(S) \leq |S|^k$. The following definition is hence straightforward.

Definition 3.16 A space–time code $\mathcal{X}$ is said to be *fast-decodable* if its ML decoding complexity is upper bounded by

$$\mathfrak{D}(S) = c|S|^{k'},$$

where $k' < k$ is the number of symbols to be jointly decoded and $c \leq k$ is a constant describing the number of parallel symbol groups to be decoded. If $c = k$, this means that we can decode symbol-wise ($k' = 1$) with linear complexity. We refer to k' as the *complexity order*.

We will mostly drop the constant c in the rest of the chapter and concentrate only on the order k', and also by abuse of notation write $\mathfrak{D}(S) = |S|^{k'}$ without the constant.

Now let us proceed to investigate how to determine the complexity order of a space–time code $\mathcal{X}$. Let $\{B_i\}_{i=1}^k$ be a basis of $\mathcal{X}$ over $\mathbb{Z}$, and $X \in \mathcal{X}$ the transmitted signal. Recall the isometry (3.1), which allows us to identify the space–time code lattice with a lattice in Euclidean space. In addition, for $c \in \mathbb{C}$ let

$$\tilde{c} = \begin{bmatrix} \mathrm{Re}(c) & -\mathrm{Im}(c) \\ \mathrm{Im}(c) & \mathrm{Re}(c) \end{bmatrix}.$$

From the channel output $Y = HX + N$, define the matrices

$$B = \left[\iota(B_1) \cdots \iota(B_k)\right] \in \mathrm{Mat}(2n_t T \times k, \mathbb{R}),$$
$$B_H = \left[\iota(HB_1) \cdots \iota(HB_k)\right] \in \mathrm{Mat}(2n_r T \times k, \mathbb{R}).$$

The equivalent received codeword under the isometry can be expressed as $\iota(HX) = B_H\mathbf{s}$ for a coefficient vector $\mathbf{s}^t = (s_1, \ldots, s_k) \in S^k$, and we get an equivalent vectorized channel equation

$$\begin{aligned}\iota(Y) &= B_H\mathbf{s} + \iota(N)\\ &= (I_T \otimes \tilde{H})B\mathbf{s} + \iota(N),\end{aligned}$$

where $\tilde{H} = (\tilde{h}_{ij})_{i,j}$ and $\otimes$ denotes the Kronecker product.

We go on to perform QR-decomposition on B_H, or equivalently on $(I_T \otimes \tilde{H})B$. We write $B_H = QR$ with $Q \in \mathrm{Mat}(2n_r T \times k, \mathbb{R})$ unitary and $R \in \mathrm{Mat}(k, \mathbb{R})$ upper triangular. More precisely, if we write

$$B_H = \left[\mathbf{b}_1 \cdots \mathbf{b}_k\right], \qquad Q = \left[\mathbf{q}_1 \cdots \mathbf{q}_k\right],$$

the matrix R is precisely given by

$$R = \begin{bmatrix} \|\mathbf{r}_1\| & \langle \mathbf{q}_1, \mathbf{b}_2\rangle & \langle \mathbf{q}_1, \mathbf{b}_3\rangle & \cdots & \langle \mathbf{q}_1, \mathbf{b}_k\rangle \\ 0 & \|\mathbf{r}_2\| & \langle \mathbf{q}_2, \mathbf{b}_3\rangle & \cdots & \langle \mathbf{q}_2, \mathbf{b}_k\rangle \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & \|\mathbf{r}_k\| \end{bmatrix},$$

where

$$\mathbf{r}_1 = \mathbf{b}_1; \quad \mathbf{r}_i = \mathbf{b}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{q}_j, \mathbf{b}_i\rangle}{\langle \mathbf{q}_j, \mathbf{q}_j\rangle}\mathbf{q}_j; \quad \mathbf{q}_i = \frac{\mathbf{b}_i}{\|\mathbf{b}_u\|}.$$

Since the receiver has channel state information, and as the noise is zero-mean, the decoding process, as we have already seen, requires to solve the minimisation problem

$$\hat{X} = \arg\min_{X \in \mathbf{X}} \|Y - HX\|_F^2.$$

Using the QR decomposition, we can solve the equivalent problem

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s} \in S^k} \|\iota(Y) - B_H\mathbf{s}\|^2 = \arg\min_{\mathbf{s} \in S^k} \|Q^\dagger \iota(Y) - R\mathbf{s}\|^2,$$

a problem which can be solved using a real sphere-decoder [41]. It is now clear that the structure of the matrix R determines the complexity of decoding. With

zero entries at specific places, the involved variables can be decoded independently of each other, allowing for parallelisation in the decoding process, and potentially reducing the decoding complexity.

Moreover, different orderings of the weight matrices B_i, or equivalently of the symbols s_i, result in different zero patterns in the matrix R. An algorithm for the optimal ordering of the weight matrices resulting in the minimum possible decoding complexity is given in [20], and was implemented in [18]. We use the implementation found in the latter article for the explicit computation of the decoding complexity reduction of the example codes exposed in the remaining of this section.

Before we move on to define more specialized families of fast-decodable codes, we present the usual approach to give sufficient conditions for a code to be fast-decodable. This so-called *Hurwitz-Radon quadratic form* approach is discussed in [19, 20, 36], among others. The idea behind the Hurwitz-Radon approach on which the quadratic form is based is to give a criterion for when two variables of the considered code can be decoded independently. More specifically, the variables s_i, s_j can be decoded independently if their corresponding weight matrices B_i, B_j are *mutually orthogonal*, i.e.,

$$B_i B_j^\dagger + B_j B_i^\dagger = 0.$$

To be more precise, we give the following result

Proposition 3.1 ([36, Thm. 2][8, Thm. 1]) *Let $\mathcal{X}$ be a space–time code of rank k with weight matrices $\{B_i\}_{i=1}^k$. The matrices B_i and B_j are mutually orthogonal, if and only if the columns $\mathbf{b}_i$ and $\mathbf{b}_j$ of B_H are orthogonal.*

In particular, the entry (i, j) of the associated matrix R is zero. Relating to this condition, the Hurwitz-Radon quadratic form is a tool which allows to deduce the actual ML decoding complexity of a space–time code based on the mutually orthogonality of the weight matrices. In particular, the criterion based on the quadratic form shows that fast decodability can be achieved solely by designing the weight matrices cleverly, and is independent of the channel and number of antennas. We give the following definition.

Definition 3.17 Let $\mathcal{X}$ be a space–time code of rank k, and let $X \in \mathcal{X}$. The *Hurwitz-Radon quadratic form* is the map

$$Q : \mathcal{X} \to \mathbb{R},$$

$$X = \sum_{i=1}^{k} B_i s_i \mapsto \sum_{1 \le i \le j \le k} s_i s_j m_{ij},$$

where $m_{ij} := \| B_i B_j^\dagger + B_j B_i^\dagger \|_F^2$.

Note that B_i, B_j are mutually orthogonal if and only if $m_{ij} = 0$.

3.4.1.1 Multi-Group Decodable Codes

We begin with the family of multi-group decodable codes.

Definition 3.18 Consider a space–time code $\mathcal{X}$ defined by the weight matrices $\{B_i\}_{i=1}^{k}$.

(i) The code is *g-group decodable* if there exists a partition of $\{1, \ldots, k\}$ into g non-empty subsets $\Gamma_1, \ldots, \Gamma_k$ such that for $i \in \Gamma_u$, $j \in \Gamma_v$ with $u \neq v$, the matrices B_i and B_j are mutually orthogonal.

(ii) The code is *conditionally g-group decodable* if there exists a partition of $\{1, \ldots, k\}$ into $g+1$ non-empty subsets $\Gamma_1, \ldots, \Gamma_g, \Gamma$ such that for $i \in \Gamma_u$, $j \in \Gamma_v$ with $1 \leq u < v \leq g$, the matrices B_i and B_j are mutually orthogonal.

The family of codes which we refer to as conditionally g-group decodable codes are in the literature also known as *fast ML decodable codes*. We use the terminology of conditionally g-group decodable so as to not confuse the general notion of fast decodability with this specific family of fast-decodable codes.

In the following, we consider a space–time code $\mathcal{X}$ with weight matrices $\{B_i\}_{i=1}^{k}$ and corresponding real information symbols $s_1, \ldots, s_k \in S$. For $\mathcal{X}$ g-group decodable or conditionally g-group decodable, we may without loss of generality order the variables according to the g groups $\Gamma_1, \ldots, \Gamma_g$, i.e.,

$$\begin{aligned} &\left\{s_1, \ldots, s_{|\Gamma_1|}\right\} \in \Gamma_1, \\ &\left\{s_{|\Gamma_1|+1}, \ldots, s_{|\Gamma_1|+|\Gamma_2|}\right\} \in \Gamma_2, \\ &\qquad\vdots \\ &\left\{ s_{\sum_{i=1}^{g-1}|\Gamma_i|+1}, \ldots, s_{\sum_{i=1}^{g-1}|\Gamma_i|+|\Gamma_g|} \right\} \in \Gamma_g. \end{aligned} \tag{3.3}$$

We have the following result, which will be helpful in determining the decoding complexity of a code (cf. Theorem 3.7).

Proposition 3.2 ([19, Lemma 1]) *Let $\mathcal{X}$ be a g-group decodable space–time code, and let $M = (m_{ij})_{i,j}$ be the Hurwitz-Radon quadratic form matrix (cf. Definition 3.17) and $R = (r_{ij})_{i,j}$ the R-matrix from the QR decomposition of B_H. Then, $m_{ij} = r_{ij} = 0$ for $i < j$ whenever $s_i \in \Gamma_u$ and $s_j \in \Gamma_v$ with $u \neq v$. In particular, the R-matrix takes the form*

$$R = \begin{bmatrix} D_1 & & \\ & \ddots & \\ & & D_g \end{bmatrix},$$

where $D_i \in \mathrm{Mat}(|\Gamma_i|, \mathbb{R})$ is upper triangular, $1 \leq i \leq g$, and the empty spaces are filled with zeros.

Example 3.13 The first example we give is the complexity order of the Alamouti code $\mathcal{X}_A$ (cf. Sect. 3.2). We recall that this code consists of codewords

$$X = \begin{bmatrix} x_1 + ix_2 & -(x_3 - ix_4) \\ x_3 + ix_4 & x_1 - ix_2 \end{bmatrix},$$

where $(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4$ are usually taken to be integers to guarantee non-vanishing determinants.

The R-matrix associated with this code is in fact a diagonal 4×4 matrix with equal diagonal entries. Hence, $\mathcal{X}_A$ is 4-group decodable, and exhibits a complexity order $k' = 1$. In other words, it is single-symbol decodable.

Example 3.14 We recall the code constructed for multiple-access channels in [3, Ex. 6]. Consider the cyclic division algebra

$$C = \left(F(\sqrt{-3}, i)/F(i), \sigma, -\frac{2}{\sqrt{5}} \right),$$

where $F = \mathbb{Q}(\sqrt{5})$ and $\sigma : \sqrt{-3} \mapsto -\sqrt{-3}$ but fixes $F(i)$. Let τ be a generator of the cyclic Galois group $\mathrm{Gal}(F(i)/F)$, i.e., $\tau(i) = -i$. Let us extend the action of τ from $F(i)$ to $F(i, \sqrt{-3}, \sqrt{-\gamma})$ by letting it act as identity on both $\sqrt{-3}$ and $\sqrt{-\gamma}$, as justified by the isomorphism extension theorem. Consider codewords of the form

$$X = \begin{bmatrix} X_1 & \tau(X_1) \\ X_2 & \tau(X_2) \end{bmatrix},$$

where τ acts element-wise, and for $\theta = \frac{1+\sqrt{-3}}{2}$ and $k = 1, 2$ we have

$$X_k = \begin{bmatrix} x_{k,1} + x_{k,2}\theta & -\sqrt{-\gamma}(x_{k,3} + x_{k,4}\sigma(\theta)) \\ \sqrt{-\gamma}(x_{k,3} + x_{k,4}\theta) & x_{k,1} + x_{k,2}\sigma(\theta) \end{bmatrix}$$

with $x_{k,j} \in O_{F(i)}$. Hence, each X_k corresponds to the left-regular representation of an element in the natural order $\Gamma_{nat} \subseteq C$, after balancing the effect of γ by spreading it on the diagonal.[5]

The complexity of exhaustive search for a signalling alphabet S is $|S|^{32}$. The above code, however, is 2-group decodable. In fact, the associated R-matrix is of the form

$$R = \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix}$$

with $D_i \in \mathrm{Mat}(16, \mathbb{R})$ upper triangular. The code hence exhibits a complexity order $k' = 16$, resulting in a reduction of 50%.

[5] This is a usual trick to balance the average energies of the codeword entries more evenly. See [3, Ex. 1] for more details.

In the case of conditionally g-group decodable codes, i.e., where we have a further non-empty group Γ, the R matrix is not entirely block-diagonal. Instead, we have the following result.

Proposition 3.3 ([7, Lem. 2]) *Let $\mathcal{X}$ be a conditionally g-group decodable code, and let $M = (m_{ij})_{i,j}$ be the Hurwitz-Radon quadratic form matrix and $R = (r_{ij})_{i,j}$ the R-matrix from the QR decomposition. Then, $m_{ij} = r_{ij} = 0$ for $i < j$ whenever $s_i \in \Gamma_u$ and $s_j \in \Gamma_v$ with $1 \leq u < v \leq g$. In particular, the R-matrix takes the form*

$$R = \begin{bmatrix} D_1 & & N_1 \\ & \ddots & \vdots \\ & D_g & N_g \\ & & N \end{bmatrix},$$

with $D_i \in \mathrm{Mat}(|\Gamma_i|, \mathbb{R})$ and $N \in \mathrm{Mat}(|\Gamma|, \mathbb{R})$ are upper triangular, and $N_i \in \mathrm{Mat}(|\Gamma_i| \times |\Gamma|, \mathbb{R})$.

Example 3.15 As an example of a conditionally g-group space–time code we recall the famous *Silver code* [16, 32]. The code is contained as a subset in the cyclic division algebra

$$C = (\mathbb{Q}(i, \sqrt{-7})/\mathbb{Q}(\sqrt{-7}), \sigma, \gamma),$$

Note that σ is not just complex conjugation, as $\sigma(i) = -i$ and $\sigma(\sqrt{7}) = -\sqrt{7}$. With $\gamma = -1$, the algebra is division, and the resulting code is fully diverse and has non-vanishing determinants. The Silver code is however not directly constructed as a subset of $\rho(\Gamma)$ for Γ an order of C. Instead, it is defined as

$$\mathcal{X}_S = \{ X = X_A(x_1, x_2) + T X_B(x_3, x_4) \,|\, x_1, \ldots, x_4 \in \mathbb{Z}[i] \},$$

where $x_1, \ldots, x_4 \in \mathbb{Z}[i]$ and

$$T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \qquad X_A(x_1, x_2) = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix};$$

$$X_B(x_3, x_4) = \frac{1}{\sqrt{7}} \begin{bmatrix} (1+i)x_3 + (-2+2i)x_4 & -(1-2i)x_3^* - (1+i)x_4^* \\ (1+2i)x_3 + (1-i)x_4 & (1-i)x_3^* + (-1-2i)x_4^* \end{bmatrix}.$$

In particular, a codeword is of the form

$$X = \tfrac{1}{\sqrt{7}} \begin{bmatrix} x_1\sqrt{7} + (1+i)x_3 + (-1+2i)x_4 & -x_2^*\sqrt{7} - (1-2i)x_3^* - (1+i)x_4^* \\ x_2\sqrt{7} - (1+2i)x_3 - (1-i)x_4 & x_1^*\sqrt{7} - (1-i)x_3^* - (-1-2i)x_4^* \end{bmatrix}.$$

Using the optimal ordering of the weight matrices, we find that the complexity order of the Silver code is $k' = 5$, resulting in a complexity reduction of 37.5%.

Example 3.16 As a second example, we recall the Srinath-Rajan code, originally proposed in [36] for a 4×2-MIMO channel. To the best of the authors' knowledge, this is the best performing code known for a 4×2 system among codes with the same complexity order. We recall the construction illustrated in [37], where the underlying algebraic structure was discovered.

Let L/F be a cyclic Galois extension with cyclic Galois group $\mathrm{Gal}(L/F) = \langle \tau \rangle$, and consider a cyclic division algebra $C' = (L/F, \tau, \gamma')$. Moreover, let $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra of degree n, where $K \neq F$ and $\tau\sigma = \sigma\tau$. We require $\gamma \in K \cap F$ and $\gamma' \in F \backslash K$.

For the 4×2 Srinath-Rajan code, we make the choices

(i) $L = \mathbb{Q}(i, \sqrt{5}),\ K = \mathbb{Q}(\sqrt{5}),\ F = \mathbb{Q}(i)$.
(ii) $C' = (L/F, \tau, \gamma')$ with $\gamma' = i \notin K$ and $\tau : \sqrt{5} \mapsto -\sqrt{5}$. This cyclic division algebra gives rise to the Golden code.
(iii) $C = (L/K, \sigma, \gamma)$ with $\gamma = -1$ and $\sigma : i \mapsto -i$.

Fix the F-basis $\{\theta_1, \theta_2\}$ of L, with $\theta_1 = 1 + i(1 - \theta), \theta_2 = \theta_1\theta \in O_L$, where $\theta = \frac{1+\sqrt{5}}{2}$. Codewords are of the form

$$X = \begin{bmatrix} x_0 & -\sigma(x_1) & i\tau(x_2) & -i\tau\sigma(x_3) \\ x_1 & \sigma(x_0) & i\tau(x_3) & i\tau\sigma(x_2) \\ x_2 & -\sigma(x_3) & \tau(x_0) & -\tau\sigma(x_1) \\ x_3 & \sigma(x_2) & \tau(x_1) & \tau\sigma(x_0) \end{bmatrix},$$

where $x_i = x_{i1}\theta_1 + x_{i2}\theta_2$ with $x_{ij} \in \mathbb{Z}[i]$.

This code is conditionally 4-group decodable, where 8 real variables need to be conditioned, and the remaining 8 variables can be grouped in 4 groups of 2. This can be seen from the structure of the R-matrix, which for this code takes the form

$$R = \begin{bmatrix} D_1 & & & & N_1 \\ & D_2 & & & N_2 \\ & & D_3 & & N_3 \\ & & & D_4 & N_4 \\ & & & & N \end{bmatrix},$$

where D_i are 2×2 upper triangular matrices, N_i are 2×8 matrices, and N is an 8×8 upper triangular matrix. This yields a decoding complexity order $k' = 10$. This is a reduction in complexity of 37.5%.

To summarize, we observe that the R matrix allows to directly read the decoding complexity of a g-group decodable and conditionally g-group decodable code. After conditioning the last $|\Gamma|$ variables, the variables in each group Γ_i can be decoded independently of the other groups. This is summarized in the following result.

Theorem 3.7 *The decoding complexity order of a (conditionally) g-group decodable code $\mathcal{X}$ with possibly empty subset Γ is given by*

$$k' = |\Gamma| + \max_{1 \le i \le g} |\Gamma_i|.$$

Unfortunately, there is a trade-off between the maximum rate and maximum decoding complexity reduction of space–time codes. The recent work [8] treats both these questions for multi-group decodable codes by analysing the mutually orthogonality of matrices in central simple subalgebras of $\mathrm{Mat}(n, \mathbb{C})$ over number fields. The authors show on one hand that there is a lower bound for the decoding complexity of full-rate $n \times n$ space–time codes. They furthermore derive an upper bound on the number of groups of a multi-group decodable code. We summarise the results relevant to our chapter in the following theorem. For a more general setting, see Theorems 7–8 and Corollary 16 in [8].

Theorem 3.8 ([8]) *Let $\mathcal{X}$ be an $n \times n$ space–time code defined by the weight matrices $\{B_i\}_{i=1}^{2k^2}$, and let S denote the employed real signalling alphabet.*

(i) If $\mathcal{X}$ is full-rate, then the decoding complexity order is not better than $n^2 + 1$.
(ii) If $\mathcal{X}$ is multi-group decodable and the weight matrices are chosen from a K-central division algebra with K a number field, we have $g \le 4$.

3.4.1.2 Fast-Group Decodable Codes

Fast-group decodable codes combine the structure of the block-diagonal R-matrix with further parallelisation within each of the independent groups. We start with the formal definition.

Definition 3.19 Consider a space–time code $\mathcal{X}$ defined by the weight matrices $\{B_i\}_{i=1}^{k}$. The code is *fast-group decodable* if

(a) There is a partition of $\{1, \ldots, k\}$ into g non-empty subsets $\Gamma_1, \ldots, \Gamma_g$ such that whenever $i \in \Gamma_u$, $j \in \Gamma_v$ with $u \neq v$, the matrices B_i and B_j are mutually orthogonal.
(b) In addition, for at least one group Γ_i, we have $\langle \mathbf{q}_{l_1}, \mathbf{b}_{l_2} \rangle = 0$, where $l_1 = 1, \ldots L_i - 1$ and $l_2 = l_1 + 1, \ldots, L_i$ with $L_i \le |\Gamma_i|$.

Consider a fast-group decodable space–time code $\mathcal{X}$, and denote by $\Gamma_1, \ldots, \Gamma_g$ the groups in which the corresponding symbols can be jointly decoded. Assume that the variables $s_1, \ldots, s_k$ are without loss of generality ordered according to their groups, as described above (3.3).

Proposition 3.4 ([19, Lem. 3]) *Let $\mathcal{X}$ be a g fast-group decodable space–time code, and let $M = (m_{ij})_{i,j}$ be the Hurwitz-Radon quadratic form matrix and $R = (r_{ij})_{i,j}$ the R-matrix from the QR decomposition. Then, $m_{ij} = r_{ij} = 0$ for $i < j$ whenever $s_i \in \Gamma_u$, $s_j \in \Gamma_v$ with $u \neq v$. Furthermore, each group*

Γ_i admits to remove L_i levels from the sphere-decoder tree if $m_{i_{l_1} i_{l_2}} = 0$, where $l_1 = 1, \ldots, L_i - 1$ and $l_2 = l_1 + 1, \ldots, L_i$, with $L_i \leq |\Gamma_i|$. In particular, the R-matrix takes the form

$$R = \begin{bmatrix} R_1 & & \\ & \ddots & \\ & & R_g \end{bmatrix},$$

where the empty spaces are filled with zeros. Each of the matrices $R_i \in \mathrm{Mat}(|\Gamma_i|, \mathbb{R})$ is of the form

$$R_i = \begin{bmatrix} D_i & B_{i_1} \\ & B_{i_2} \end{bmatrix},$$

with $D_i \in \mathrm{Mat}(L_i, \mathbb{R})$ is diagonal, B_{i_2} is a square upper triangular matrix and B_{i_1} is a rectangular matrix.

Theorem 3.9 *The decoding complexity of a g fast-group decodable space–time code $\mathcal{X}$ with real signalling alphabet S is given by*

$$\mathfrak{D}(S) = |S|^{\max\limits_{1 \leq i \leq g} \{|\Gamma_i| - L_i + 1\}}.$$

Example 3.17 The authors in [33] construct a 4×4 fast-group decodable code based on an orthogonal space–time code. Codewords are of the form

$$X = \begin{bmatrix} x_1 + ix_2 + ix_{15} + ix_{16} + ix_{17} & x_7 + ix_8 + x_{13} + ix_{14} & x_3 + ix_4 + x_{11} + ix_{12} & -x_5 - ix_6 + x_9 + ix_{10} \\ -x_7 + ix_8 - x_{13} + ix_{14} & x_1 + ix_2 + ix_{15} - ix_{16} - ix_{17} & x_5 - ix_6 + x_9 - ix_{10} & x_3 - ix_4 - x_{11} + ix_{12} \\ -x_3 + ix_4 - x_{11} + ix_{12} & -x_5 - ix_6 - x_9 - ix_{10} & x_1 - ix_2 + ix_{15} - ix_{16} + ix_{17} & x_7 - ix_8 - x_{13} + ix_{14} \\ x_5 - ix_6 - x_9 + ix_{10} & -x_3 - ix_4 + x_{11} + ix_{12} & -x_7 - ix_8 + x_{13} + ix_{14} & x_1 - ix_2 + ix_{15} + ix_{16} - ix_{17} \end{bmatrix},$$

where x_i are real symbols. We refer to the original paper for more details on the explicit construction. The algebraic structure of this code allows to remove 5 levels from the sphere decoding tree. In particular, the decoding complexity order is $k' = 12$, resulting in a reduction in decoding complexity of $\sim$30%.

3.4.1.3 Block Orthogonal Codes

The last family of fast-decodable codes that we treat are *block orthogonal codes*. We define this family by means of the structure of the associated R-matrix.

Definition 3.20 Let $\mathcal{X}$ be a space–time code. The code is said to be *g-block orthogonal* if the associated R-matrix has the structure

$$R = \begin{bmatrix} R_1 & B_{12} & \cdots & B_{1g} \\ & R_2 & \cdots & B_{2g} \\ & & \ddots & \vdots \\ & & & R_g \end{bmatrix},$$

where the empty spaces are filled with zeros and the matrices B_{ij} are non-zero rectangular matrices. Further, the matrices R_i are block diagonal matrices of the form

$$R_i = \begin{bmatrix} U_{i,1} & & \\ & \ddots & \\ & & U_{i,k_i} \end{bmatrix},$$

with each of the blocks $U_{i,j}$ is a square upper triangular matrix.

Assuming that each of the matrices R_i has the same number of blocks k, we can determine a block orthogonal code by the three parameters (g, k, p), where g is the number of matrices R_i, k denotes the number of block matrices which compose each matrix R_i and p is the number of diagonal entries in the block matrices $U_{i,j}$.

Example 3.18 The aforementioned Golden code is a $(2, 2, 2)$ block orthogonal code. However, as its decoding complexity order is $k' = 6 < 8 = k$, it is not fast-decodable by the requirement of a strict inequality as per Definition 3.16.

As an example of a fast-decodable block orthogonal code, we consider the $(2, 4, 2)$ block orthogonal code from [21]. For a signalling vector $(s_1, \ldots, s_{16})$, a codeword is of the form

$$X = X'(s_1, \ldots, s_8) + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} X'(s_9, \ldots, s_{16}),$$

where

$$X'(s_1, \ldots, s_8) = \begin{bmatrix} (s_1 - s_2) + i(s_3 - s_4) & 0 & (s_7 - s_8) + i(s_5 - s_6) & 0 \\ 0 & (s_1 - s_2) + i(s_4 - s_3) & 0 & (s_8 - s_7) + i(s_6 - s_5) \\ -(s_7 + s_8) + i(s_5 + s_6) & 0 & (s_1 + s_2) - i(s_3 + s_4) & 0 \\ 0 & (s_7 + s_8) - i(s_5 + s_6) & 0 & (s_1 + s_2) + i(s_3 + s_4) \end{bmatrix},$$

Remark 3.3 Recall that the property of fast decodability relates to the reduction in decoding complexity without resorting to suboptimal decoding methods. By modifying the decoding algorithm used, the decoding complexity of certain codes

can be lowered. For example, the main algorithm of [35] reduces the complexity order of the Golden code from $k = 6$, corresponding to the complexity of ML-decoding, to $k' = 4$, while maintaining nearly-ML performance. The algorithm is specific to the Golden code, but has been generalized to the 3×3 and 4×4 perfect codes in, respectively, [2, 17].

In contrast to the previously introduced families, the approach via the Hurwitz-Radon quadratic form does not capture the complexity reduction for block orthogonal codes. This was recently addressed in [26], where relaxed conditions are derived for classifying codes into the here treated families of fast-decodable codes. More precisely, for block orthogonal codes we do not have an analogue of Proposition 3.3 or 3.4 relating the matrix M of the quadratic form to the R-matrix in the QR decomposition of B_H.

3.4.2 Inheriting Fast Decodability

Crucial for space–time codes to exhibit desirable properties is the underlying algebraic framework. Constructing codes for larger number of antennas means dealing with higher degree field extensions and algebras, which are harder to handle. We briefly recall an iterative space–time code construction proposed in [25] which, starting with an $n \times n$ space–time code, results in a new $2n \times 2n$ space–time code with the same code rate and double (lattice) rank. The advantage of this construction is that when applied carefully, the resulting codes inherit good properties from the original space–time codes.

As the general setup, consider the tower of extensions depicted in Fig. 3.2.

The cyclic Galois group of L/K is generated by σ, i.e., $\mathrm{Gal}(L/K) = \langle\sigma\rangle$, and we denote the left-regular representation by $\rho : C \to \mathrm{Mat}(n, L)$. Let $\tau \in \mathrm{Aut}(L)$ be an automorphism of L, and make the following assumptions:

$$\tau(\gamma) = \gamma; \quad \tau\sigma = \sigma\tau. \tag{3.4}$$

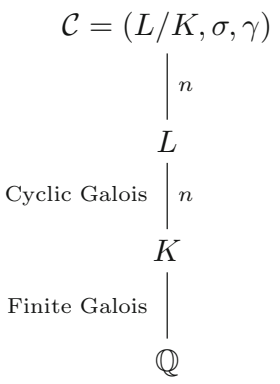


Fig. 3.2 Tower of extensions for the MIMO example code

By the above assumptions we have $\tau\rho = \rho\tau$. Moreover, τ can be extended to an automorphism of C and $\rho(C)$, respectively, by

$$\tau\left(\sum_{i=0}^{n-1} e^i x_i\right) = \sum_{i=0}^{n-1} e^i \tau(x_i); \qquad \tau\left((a_{ij})_{i,j}\right) = \left(\tau(a_{ij})\right)_{i,j}.$$

We can now fix an element $\theta \in C$, as well as a $\mathbb{Q}$-automorphism of L, $\tau \in \mathrm{Aut}_{\mathbb{Q}}(L)$, and have the following important definition.

Definition 3.21 Let K be a finite Galois extension of $\mathbb{Q}$ and $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra of degree n. Fix $\theta \in C$ and $\tau \in \mathrm{Aut}_{\mathbb{Q}}(L)$ as above.

(a) Define the function

$$\alpha_{\tau,\theta} : \mathrm{Mat}(n, L) \times \mathrm{Mat}(n, L) \to \mathrm{Mat}(2n, L)$$
$$(X, Y) \mapsto \begin{bmatrix} X & \theta\tau(Y) \\ Y & \tau(X) \end{bmatrix}.$$

(b) If $\theta = \zeta\theta'$ is totally real or totally imaginary, $\theta' > 0$ and $\zeta \in \{\pm 1, \pm i\}$, define the alike function

$$\tilde{\alpha}_{\tau,\theta} : \mathrm{Mat}(n, L) \times \mathrm{Mat}(n, L) \to \mathrm{Mat}(2n, L)$$
$$(X, Y) \mapsto \begin{bmatrix} X & \zeta\sqrt{\theta'}\tau(Y) \\ \sqrt{\theta'}Y & \tau(X) \end{bmatrix}.$$

The defined maps restrict to $C \times C \to \mathrm{Mat}(2, C)$ by identifying $x, y \in C$ with their representation $X = \rho(x)$, $Y = \rho(y)$.

Suppose that the algebra C gives rise to a rank-k space–time code $\mathcal{X}$ defined via matrices $\{B_i\}_{i=1}^k$. Then, the matrices $\left\{\alpha_{\tau,\theta}(B_i, 0), \alpha_{\tau,\theta}(0, B_i)\right\}_{i=1}^k$ (or applying $\tilde{\alpha}_{\tau,\theta}(\cdot,\cdot)$, respectively) define a rank-$2k$ code

$$\mathcal{X}_{\mathrm{it}} = \left\{ \sum_{i=1}^{k} \left[s_i \alpha_{\tau,\theta}(B_i, 0) + s_{k+i} \alpha_{\tau,\theta}(0, B_i) \right] \middle| \, s_i \in S \right\}.$$

We summarise the main results of [25] in the following proposition.

Proposition 3.5 ([25, Thm. 1, Thm. 2]) *Let $C = (L/K, \sigma, \gamma)$ be a cyclic division algebra giving rise to a rank-k space–time code $\mathcal{X}$ defined by the matrices $\{B_i\}_{i=1}^k$. Assume that $\tau \in \mathrm{Aut}_{\mathbb{Q}}(L)$ commutes with σ and complex conjugation, and further $\tau(\gamma) = \gamma$, $\tau^2 = \mathrm{id}$. Fix $\theta \in K^{\langle\tau\rangle}$, where $K^{\langle\tau\rangle}$ is the subfield of K fixed by τ. Identifying an element of C with its left-regular representation ρ, we have:*

(i) The image $\mathcal{I} = \alpha_{\tau,\theta}(C, C)$ is an algebra and is division if and only if $\theta \neq z\tau(z)$ for all $z \in C$. Moreover, for any $\alpha_{\tau,\theta}(x, y) \in \mathcal{I}$, we have $\det(\alpha_{\tau,\theta}(x, y)) \in K^{\langle\tau\rangle}$.

(ii) If in addition $\theta = \zeta\theta'$ is totally real or totally imaginary, the image $\mathcal{I} = \tilde{\alpha}_\theta(C, C)$ retains both the full-diversity and non-vanishing determinant property. If for some i, j, $B_i B_j^\dagger + B_j B_i^\dagger = 0$, we have

$$\tilde{\alpha}_{\tau,\theta}(B_i, 0)\tilde{\alpha}_{\tau,\theta}(B_j, 0)^\dagger + \tilde{\alpha}_{\tau,\theta}(B_j, 0)\tilde{\alpha}_{\tau,\theta}(B_i, 0)^\dagger = 0\,,$$

$$\tilde{\alpha}_{\tau,\theta}(0, B_i)\tilde{\alpha}_{\tau,\theta}(0, B_j)^\dagger + \tilde{\alpha}_{\tau,\theta}(0, B_j)\tilde{\alpha}_{\tau,\theta}(0, B_i)^\dagger = 0.$$

The second part of Proposition 3.5, in particular, states that under appropriate conditions, fast decodability is inherited from the rank-k space–time code $\mathcal{X}$ to the iterated code $\mathcal{X}_{\text{it}}$.

3.5 Explicit Constructions

All the notions and concepts introduced in the previous sections lead to this last part. To conclude the chapter, we focus on explicit construction methods for fast-decodable space–time codes.

Throughout this chapter, we have provided multiple examples of space–time codes with reduced ML decoding complexity. Such examples can sometimes be found by chance, but most often a clever design gives rise to infinite families of codes with reduced decoding complexity. In the following, we turn our attention to communication setups for which such general results are known. To the best of the authors' knowledge, the constructions presented here are the only general fast-decodable algebraic constructions found in literature.

3.5.1 Asymmetric Space–Time Codes

Above we have exemplified the 4×2 Srinath-Rajan code, the best performing code for this channel among codes with the same complexity order. Here, we discuss a methodology for constructing well-performing fast-decodable space–time codes for the 4×2 MIMO channel, offering a reduction in decoding complexity of up to 37.5%.

The motivation behind the following construction is the structure of the Alamouti code (cf. Example 3.13). As we have seen, the decoding complexity of the Alamouti code equals the size of the employed real signaling alphabet, $\mathfrak{D}(S) = |S|$ (or more precisely $\mathfrak{D}(S) = 4|S|$ as we are decoding each of the 4 real symbols in parallel). Motivated by this observation, it is of interest to study space–time codes which are subsets of the rings $\mathrm{Mat}(k, \mathbb{H})$. This motivates the next result.

Theorem 3.10 ([40]) *Let C be a cyclic division algebra of degree n, with center K of signature (r, s), $r + 2s = m$. There exists an injection*

$$\psi : C \hookrightarrow \operatorname{diag}\left(\operatorname{Mat}(n/2, \mathbb{H})^w \times \operatorname{Mat}(n, \mathbb{R})^{r-w} \times \operatorname{Mat}(n, \mathbb{C})^s\right),$$

where each $n \times n$ block is mapped to the corresponding diagonal block of a matrix in $\operatorname{Mat}(mn, \mathbb{C})$. Here, w is the number of places which ramify in C.

In particular, C can be embedded into $\operatorname{Mat}(n/2, \mathbb{H})$ if

(i) The center K is totally real, i.e., $r = m$.
(ii) The infinite places of K are ramified in C.

The ramification assumptions of places in the considered algebra are rather technical, and the interested reader is referred to [40] for further details.

While the above result guarantees the existence of an injection into $\operatorname{Mat}(n/2, \mathbb{H})$ when the conditions are satisfied, it does not make the embedding explicit. This is achieved in the following result.

Theorem 3.11 ([40, Prop. 11.1]) *Let $C = (L/\mathbb{Q}, \sigma, \gamma)$ be a cyclic division algebra satisfying the requirements from Theorem 3.10. Given for $x \in C$ an element $X = \rho(x) \in \mathcal{X}$, where $\mathcal{X}$ is a space–time code arising from the algebra C, we have an explicit map*

$$\begin{aligned} \psi : C &\to \operatorname{Mat}(n_t/2, \mathbb{H}) \\ X &\mapsto BPX(BP)^{-1}, \end{aligned}$$

where $P = (p_{ij})_{i,j}$ is a permutation matrix with entries

$$p_{ij} = \begin{cases} 1 & \text{if } 2 \nmid i \text{ and } j = \frac{i+1}{2}, \\ 1 & \text{if } 2 \mid i \text{ and } j = \frac{i+n_t}{2}, \\ 0 & \text{otherwise}, \end{cases}$$

and $B = \operatorname{diag}(\sqrt{|\gamma|}, |\gamma|, \ldots, \sqrt{|\gamma|}, |\gamma|)$.

We now turn our attention to the 4×2 MIMO channel. Given the results inroduced above, we recall a construction method for fast-decodable space–time codes for this channel.

Theorem 3.12 ([40]) *Let $C = (K/\mathbb{Q}, \sigma, \gamma)$ be a division algebra of index 4, where K is a totally complex field containing a totally real field of index 2. Assume that*

(i) $[K : \mathbb{Q}] = 4$,
(ii) $\gamma, \gamma^2 \notin \operatorname{Nm}_{K/\mathbb{Q}}\left(K^\times\right)$,
(iii) $\operatorname{Gal}(K/\mathbb{Q}) = \langle\sigma\rangle$ with σ^2 complex conjugation,
(iv) $\gamma < 0$.

Let $O_K = \mathbb{Z}w_1 + \mathbb{Z}w_2 + \mathbb{Z}w_3 + \mathbb{Z}w_4$ be the ring of integers of K, and consider the left regular representation ρ of $x \in C$, which under the above assumptions can be written as

$$\rho : x \mapsto \begin{bmatrix} x_1 & \gamma\sigma(x_4) & \gamma x_3^* & \gamma\sigma(x_2)^* \\ x_2 & \sigma(x_1) & \gamma x_4^* & \gamma\sigma(x_3)^* \\ x_3 & \sigma(x_2) & x_1^* & \gamma\sigma(x_4)^* \\ x_4 & \sigma(x_3) & x_2^* & \sigma(x_1)^* \end{bmatrix}$$

Here, $x_i = g_{4i-3}w_1 + g_{4i-2}w_2 + g_{4i-1}w_3 + g_{4i}w_4$ for $i = 1, \ldots, 4$ with $g_j \in \mathbb{Q}$ for all j, and $*$ denotes complex conjugation.

For ψ the explicit map given in Theorem 3.11, $\psi(\Gamma)$ is a lattice of dimension 16 in $\mathrm{Mat}(4, \mathbb{C})$ *with the non-vanishing determinant property. For a signaling alphabet S, codes arising from this construction have a decoding complexity order of* $10 \leq k' \leq 16$*, that is, enjoy a reduction in decoding complexity of up to* 37.5%.

Example 3.19 The MIDO_{A_4} code is a space–time code constructed in [40]. It is in fact a $(2, 2, 4)$ block orthogonal code, constructed from an algebra over the fifth cyclotomic field $\mathbb{Q}(\zeta_5)$. Consider the cyclic division algebra

$$C = \left(\mathbb{Q}(\zeta_5)/\mathbb{Q}, \sigma, -\frac{8}{9}\right),$$

where $\sigma : \zeta_5 \mapsto \zeta_5^3$.

Fix the $\mathbb{Z}$-basis $\left\{1 - \zeta_5, \zeta_5 - \zeta_5^2, \zeta_5^2 - \zeta_5^3, \zeta_5^3 - \zeta_5^4\right\}$ of O_K. Consider a maximal order Γ of C, and ψ the conjugation given in Theorem 3.11. Under this conjugation, codewords are of the form

$$X(x_1, \ldots, x_4) = \begin{bmatrix} x_1 & -r^2x_1^* & -r^3\sigma(x_4) & -r\sigma(x_3)^* \\ r^2x_2 & x_1^* & r\sigma(x_3) & -r^3\sigma(x_4)^* \\ rx_3 & -r^3x_3^* & \sigma(x_1) & -r^2\sigma(x_2)^* \\ r^3x_3 & rx_2^* & r^2\sigma(x_1) & \sigma(x_1)^* \end{bmatrix},$$

where $r = \left(\frac{8}{9}\right)^{1/4}$ and

$$x_i = g_{4i-3}(1 - \zeta_5) + g_{4i-2}(\zeta_5 - \zeta_5^2) + g_{4i-1}(\zeta_5^2 - \zeta_5^3) + g_{4i}(\zeta_5^3 - \zeta_5^4),$$
$$\sigma(x_i) = g_{4i-3}(1 - \zeta_5^3) + g_{4i-2}(\zeta_5^3 - \zeta_5) + g_{4i-1}(\zeta_5 - \zeta_5^4) + g_{4i}(\zeta_5^4 - \zeta_5^2).$$

The decoding complexity order of this code is $k' = 12$, resulting in a reduction in decoding complexity of 25%.

By choosing the basis $\left\{1, \frac{\zeta_5 + \zeta_5^{-1}}{2}, \frac{\zeta_5 - \zeta_5^{-1}}{2}, \frac{\zeta_5^2 - \zeta_5^{-2}}{4}\right\}$ of O_K instead, the decoding complexity can be further reduced. However, this is no longer an integral basis, and the price to pay is a smaller minimum determinant, yielding a slightly worse performance.

3.5.2 *Distributed Space–Time Codes*

The second setting we consider is a cooperative communications scenario. More concretely, we consider the communication of $(M+1)$ users with a single destination, where every user as well as the destination can be equipped with either a single antenna or multiple antennas. In this scenario, enabling cooperation and dividing the allocated transmission time allows for the M inactive users to aid the active source in communicating with the destination by acting as intermediate *relays*. For more details on the transmission model we refer to [3, 42]. While this is a more involved transmission scheme, from the destinations point of view it can be modeled as a virtual MIMO channel. Assume that the destination is equipped with n_r receive antennas. Setting $T = n := 2Mn_t$, where n_t is the number of transmit antennas available at each transmitter, we get the familiar channel equation $Y = HX + N$, where $X \in \mathrm{Mat}(n, \mathbb{C})$ and $Y \in \mathrm{Mat}(n_r \times n, \mathbb{C})$ are the (overall) transmitted and received signals, and the structure of the channel matrix $H \in \mathrm{Mat}(n_r \times n, \mathbb{C})$ is determined by the different relay paths.[6]

Furthermore, it is discussed in [42] that for this channel model, block-diagonal space–time codes, that is, where each $X \in \mathcal{X}$ takes the form

$$X = \mathrm{diag}\,(X_m)_m = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_M \end{bmatrix}$$

with $X_m \in \mathrm{Mat}(2n_t, \mathbb{C})$ are good choices for this channel if they additionally respect the usual design criteria such as non-vanishing determinants. To achieve this block structure, the following function is crucial.

Definition 3.22 Consider an M-relay channel as discussed above. Given a space–time code $\mathcal{X} \subset \mathrm{Mat}(2n_t, \mathbb{C})$ and a suitable function η of order M (i.e., $\eta^M(X) = X$), define the function

$$\Psi_{\eta,M} : \mathcal{X} \to \mathrm{Mat}(2n_t M, \mathbb{C})$$

$$X \mapsto \mathrm{diag}\left\{\eta^i(X)\right\}_{i=0}^{M-1} = \begin{bmatrix} X & & \\ & \ddots & \\ & & \eta^{M-1}(X) \end{bmatrix}.$$

We begin with the case where $n_t = 1$ and $n_r \geq 2$. Consider the tower of extensions depicted in Fig. 3.3, where ξ is taken to be totally real, $m \in \mathbb{Z}_{\geq 1}$ and $a \in \mathbb{Z} \setminus \{0\}$ are square-free.

[6] As remarked in Sect. 3.4.1, the property of fast decodability is independent of the channel. Hence, we omit details on the structure of the effective channel.

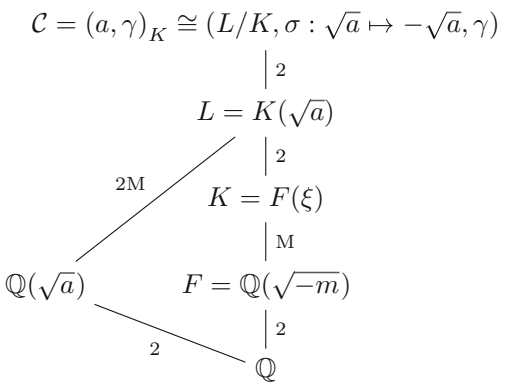


Fig. 3.3 Tower of extensions for the SIMO code construction

Assume that $\mathcal{C}$ is division. Let σ be the generator of $\mathrm{Gal}(L/K)$, and fix a generator η of $\mathrm{Gal}(K/F)$.

To have balanced energy and good decodability, it is necessary to slightly modify the matrix representation of the elements in $\mathcal{C}$, thus for $\Gamma \subset \mathcal{C}$ an order, instead of representing $x = x_0 + \sqrt{\gamma}x_1 \in \Gamma$ by its left-regular representation $\rho(x)$, we define the following similar and commonly used function that maintains the original determinant,

$$\tilde{\rho} : x \mapsto \begin{bmatrix} x_0 & -\sqrt{-\gamma}\sigma(x_1) \\ \sqrt{-\gamma}x_1 & \sigma(x_0) \end{bmatrix}. \tag{3.5}$$

Theorem 3.13 ([3, Thm. 1]) *Arising from the algebraic setup in Fig. 3.3 with $a < 0$, $\gamma < 0$, define the set*

$$\mathcal{X} = \left\{\Psi_{\eta,M}(X)\right\}_{X\in\tilde{\rho}(\Gamma)} = \left\{ \mathrm{diag}\left(\eta^i(X)\right)_{i=0}^{M-1} \middle| \, X \in \tilde{\rho}(\Gamma) \right\}.$$

The code $\mathcal{X}$ is of rank $8M$, rate $R = 4$ real symbols per channel use and has the non-vanishing determinant property. It is full-rate if $n_r = 2$. Moreover, $\mathcal{X}$ is conditionally 4-group decodable, and its decoding complexity order can be reduced from $k = 8M$ to $k' = 5M$, resulting in a complexity reduction of 37.5%.

Example 3.20 For $M = 2$ relays and $\xi = \sqrt{5}$, consider the tower of extensions in Fig. 3.4. The algebra $\mathcal{C}$ is division [3, Exp. 1].

Let $x = x_0 + \sqrt{-\gamma}x_1$ with $x_0, x_1 \in \mathcal{O}_L$ and $X = \tilde{\rho}(x)$. For $\langle\eta\rangle = \Gamma(K/F)$, define the 2-relay code

$$\mathcal{X} = \left\{\Psi_{\eta,2}(X)\right\}_{X\in\tilde{\rho}(\mathcal{O}_L)} = \left\{ \mathrm{diag}\left(\eta^i(X)\right)_{i=0}^{1} = \begin{bmatrix} X & \\ & \eta(X) \end{bmatrix} \middle| \, X \in \tilde{\rho}(\mathcal{O}_L) \right\}.$$

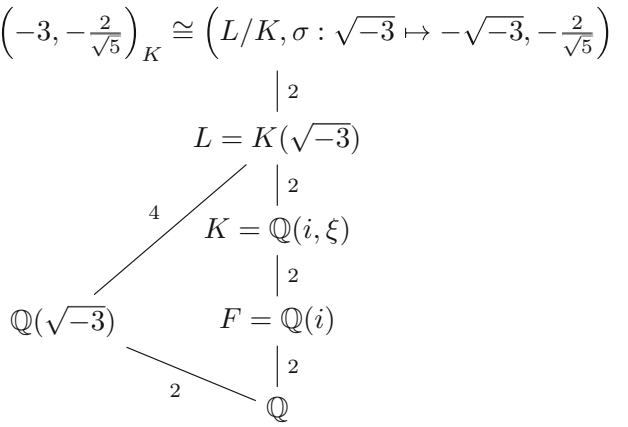


Fig. 3.4 Tower of extensions for the SIMO example code

The resulting code is a fully diverse code of rank 16 with non-vanishing determinants, which is conditionally 4-group decodable having decoding complexity order $k' = 10$ in contrast to $k = 16$.

We move on to the case where the transmitter and each relay is now equipped with $n_t \geq 1$ antennas. We require that the number of relays is expressible as $M = (p-1)/2$, with $p \geq 5$ prime. Let henceforth $n_t = 2$. Assume further a single destination with $n_r \geq 1$ antennas, and consider the tower of extensions in Fig. 3.5, where $K = \mathbb{Q}(\xi) = \mathbb{Q}^+(\zeta_p) \subset \mathbb{Q}(\zeta_p)$ is the maximal real subfield of the pth cyclotomic field, that is, $\xi = \zeta_p + \zeta_p^{-1}$, and $a \in \mathbb{Z} \setminus \{0\}$ is square-free. Let $\langle \sigma \rangle = \mathrm{Gal}(L/K)$ and $\langle \eta \rangle = \mathrm{Gal}(L/F)$.

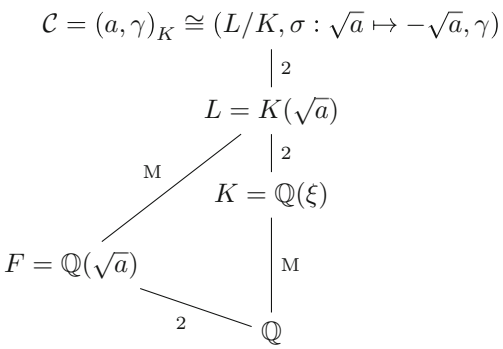


Fig. 3.5 Tower of extensions for the MIMO code construction

Theorem 3.14 ([3, Thm. 2]) *In the setup as in Fig. 3.5, choose $a \in \mathbb{Z}_{<0}$ such that $\mathfrak{p} = a\mathcal{O}_K$ is a prime ideal. Fix further $\gamma < 0$ and $\theta \in \mathcal{O}_K \cap \mathbb{R}^\times = \mathbb{Z}[\xi] \cap \mathbb{R}^\times$ such that*

- *γ and θ are both non-square* mod $\mathfrak{p}$,
- *the quadratic form $\langle \gamma, -\theta \rangle_L := l_1\gamma - l_2\theta$ with $l_1, l_2 \in L$ is anisotropic, i.e., evaluates to zero if and only if $\gamma = \theta = 0$,*

and further let $\tau = \sigma$. Then, if $\Gamma \subset \mathcal{C}$ is an order, the distributed space–time code

$$\mathcal{X} = \left\{ \Psi_{\eta,M}(\tilde{\alpha}_{\tau,\theta}(X,Y)) = \operatorname{diag}\left(\eta^i(\tilde{\alpha}_{\tau,\theta}(X,Y))\right)_{i=0}^{M-1} \middle| \, X, Y \in \tilde{\rho}(\Gamma) \right\}$$

is a full-diversity space–time code of rank $8M$, rate $R = 2$ real symbols per channel use (hence full-rate for $n_r = 1$), exhibits the non-vanishing determinant property and is g-group decodable, with $g \in \{2, 4\}$. Its decoding complexity order is

$$k' = \begin{cases} 4M & \text{if } a \equiv 1 \bmod 4, \\ 2M & \text{if } a \not\equiv 1 \bmod 4, \end{cases}$$

resulting in a reduction in complexity of 50% and 75%, respectively.

Example 3.21 We construct a 4-group decodable code for $M = 3$ relays, arising from the tower of extensions depicted in Fig. 3.6, where $\xi = \zeta_7 + \zeta_7^{-1}$ and $\gamma = -\frac{2}{1+\xi}$.

In the following, let $\tau = \sigma$ and $\langle \eta : \xi \mapsto \xi^2 - 2 \rangle = \Gamma(L/F)$. Choose further $\theta = 3(1-\xi) = \zeta\theta'$, with $\zeta = -1$ and $\theta' \in \mathbb{R}_{>0}$, and let $p_{\min}(x, \xi)$ be the minimal polynomial of ξ. With these choice of elements, the conditions from Theorem 3.14 are satisfied.

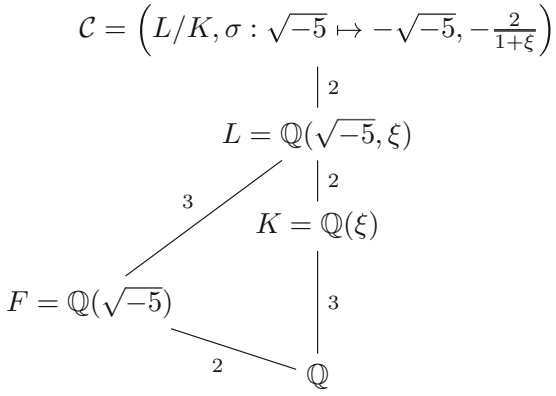


Fig. 3.6 Tower of extensions for the MIMO example code

Let $x \in \Gamma \subset C$, and set $\omega = \sqrt{-5}$. We define a space–time code $\mathcal{X}_0$ consisting of codewords of the form

$$X = \tilde{\rho}(x) = \begin{bmatrix} x_1 + x_2\omega & -\sqrt{-\gamma}(x_3 + x_4\sigma(\omega)) \\ \sqrt{-\gamma}(x_3 + x_4\omega) & x_1 + x_2\sigma(\omega) \end{bmatrix},$$

where $x_i \in O_K$, $1 \leq i \leq 4$. Next, we iterate $\mathcal{X}_0$ with the help of $\tilde{\alpha}(\cdot, \cdot)$ to obtain the set

$$\mathcal{X}_0^{\text{it}} = \left\{ \tilde{\alpha}_{\tau,\theta}(X, Y) = \begin{bmatrix} X & \zeta\sqrt{\theta'}\tau(Y) \\ \sqrt{\theta'}Y & \tau(X) \end{bmatrix} \middle| \, X, Y \in \tilde{\rho}(\Gamma) \right\}$$

and finally adapt the iterated code to the 3-relay channel by applying the map η, resulting in distributed space–time code

$$\mathcal{X} = \left\{ \Psi_{\eta,3}(\tilde{\alpha}_{\tau,\theta}(X, Y)) = \text{diag}\left(\eta^j(\tilde{\alpha}_{\tau,\theta}(X, Y))\right)_{j=0}^{2} \middle| \, X, Y \in \tilde{\rho}(\Gamma) \right\}$$

The resulting relay code is fully diverse, exhibit the non-vanishing determinant property and are fast-decodable. More concretely, $\mathcal{X}$ is 4-group decodable with decoding complexity order $k' = 6$ in contrast to $k = 24$, resulting in a complexity reduction of 75%.

3.6 Conclusions

In this chapter, we have given an overview on the topic of fast decodability of algebraic space–time codes. Traditionally, space–time codes have been developed in the context of point-to-point MIMO communications. However, with the development of new communication protocols in order to accommodate different types of applications and devices in modern wireless networks, so-called distributed space–time codes have recently become a popular subject of research. Due to the nature of the underlying communication protocols, such codes often exhibit a too high decoding complexity for practical use. Following the ideas of fast-decodability of more traditional space–time codes, this chapter aimed at giving an overview on the subdivision of space–time codes into different families of so-called fast-decodable codes. Moreover, we were particularly interested in the specific reduction in decoding complexity offered by these codes.

While crucial for practical implementation, only few explicit construction methods of fast-decodable space–time codes can be found in literature. In this chapter, we further recalled explicit constructions of asymmetric and distributed space–time codes with reduced decoding complexity, accompanied by example codes to illustrate the presented methods.

With the upcoming fifth generation (5G) wireless systems in mind, the development of new constructions of suitable well-performing space–time codes offering

complexity reduction is crucial for many applications, and opens up an interdisciplinary and rich research direction for future work.

References

1. Alamouti, S.: A simple transmitter diversity scheme for wireless communications. IEEE J. Sel. Areas Commun. **16**(8), 1451–1458 (1998)
2. Amani, E., Djouani, K., Kurien, A.: Low complexity decoding of the 4×4 perfect space–time block code. In: International Conference on Ambient Spaces, Networks and Technologies (2014)
3. Barreal, A., Hollanti, C., Markin, N.: Fast-decodable space–time codes for the n-relay and multiple-access MIMO channel. IEEE Trans. Wirel. Commun. **15**(3), 1754–1767 (2016)
4. Belfiore, J.-C., Rekaya, G.: Quaternionic lattices for space–time coding. In: Proceedings of the IEEE Information Theory Workshop (2003)
5. Belfiore, J-C., Rekaya, G., Viterbo, E.: The Golden code: a 2×2 full-rate space–time code with non-vanishing determinants. IEEE Trans. Inf. Theory **51**(4), 1432–1436 (2005)
6. Berhuy, G., Oggier, F.: An introduction to central simple algebras and their applications to wireless communication. In: Mathematical Surveys and Monographs. American Mathematical Society, New York (2013)
7. Berhuy, G., Markin, N., Sethuraman, B.A.: Fast lattice decodability of space–time block codes. In: Proceedings of the IEEE International Symposium on Information Theory (2014)
8. Berhuy, G., Markin, N., Sethuraman, B.A.: Bounds on fast decodability of space–time block codes, skew-Hermitian matrices, and Azumaya algebras. IEEE Trans. Inf. Theory **61**(4), 1959–1970 (2015)
9. Biglieri, E., Hong, Y., Viterbo, E.: On fast-decodable space–time block codes. IEEE Trans. Inf. Theory **55**(2), 524–530 (2009)
10. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices and Groups, 3rd edn.. Springer, Berlin (1999)
11. Ebeling, W.: Lattices and Codes, 3rd edn. Spektrum, Heidelberg (2013)
12. Elia, P., Sethuraman, B.A., Kumar, P.V.: Perfect space–time codes for any number of antennas. IEEE Trans. Inf. Theory **53**(11), 3853–3868 (2007)
13. Gray, F.: Pulse Code Communication (1953). US Patent 2,632,058
14. Hollanti, C., Lahtonen, J.: A new tool: Constructing STBCs from maximal orders in central simple algebras. In: Proceedings of the IEEE Information Theory Workshop (2006)
15. Hollanti, C., Lahtonen, J., Lu, H.f.: Maximal orders in the design of dense space–time lattice codes. IEEE Trans. Inf. Theory **54**(10), 4493–4510 (2008)
16. Hollanti, C., Lahtonen, J., Ranto, K., Vehkalahti, R.: On the algebraic structure of the silver code: a 2×2 perfect space–time block code. In: Proceedings of the IEEE Information Theory Workshop (2008)
17. Howard, S.D., Sirianunpiboon, S., Calderbank, A.R.: Low complexity essentially maximum likelihood decoding of perfect space-time block codes. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (2009)
18. Jäämeri, E.: Tila-aikakoodien nopea pallodekoodaus (2016). Bachelor's thesis
19. Jithamithra, G.R., Rajan, B.S.: A quadratic form approach to ML decoding complexity of STBCs. arXiv:1004.2844v2 (2010)
20. Jithamithra, G.R., Rajan, B.S.: Minimizing the complexity of fast sphere decoding of STBCs. IEEE Trans. Wirel. Commun. **12**(12), 6142–6153 (2013)
21. Jithamithra, G.R., Rajan, B.S.: Construction of block orthogonal STBCs and reducing their sphere decoding complexity. IEEE Trans. Wirel. Commun. **13**(5), 2906–2919 (2014)

22. Lu, H.f., Vehkalahti, R., Hollanti, C., Lahtonen, J., Hong, Y., Viterbo, E.: New space–time code constructions for two-user multiple access channels. IEEE J. Sel. Top. Sign. Proces. **3**(6), 939–957 (2009)
23. Lu, H.f., Hollanti, C., Vehkalahti, R., Lahtonen, J.: DMT optimal codes constructions for multiple-access MIMO channel. IEEE Trans. Inf. Theory **57**(6), 3594–3617 (2011)
24. Luzzi, L., Vehkalahti, R.: Almost universal codes achieving ergodic MIMO capacity within a constant gap. IEEE Trans. Inf. Theory **63**(5) (2017)
25. Markin, N., Oggier, F.: Iterated space–time code constructions from cyclic algebras. IEEE Trans. Inf. Theory **59**(9), 5966–5979 (2013)
26. Mejri, A., Khsiba, M.-A., Rekaya, G.: Reduced-complexity ML decodable STBCs: Revisited design criteria. In: Proceedings of the International Symposium on Wireless Communication Systems (2015)
27. Milne, J.: Class Field Theory (2013). Graduate course notes, v4.02
28. Milne, J.: Algebraic Number Theory (2014). Graduate course notes, v2.0
29. Neukirch, J.: Algebraic Number Theory. Springer, Berlin (2010)
30. Oggier, F., Rekaya, G., Belfiore, J.-C., Viterbo, E.: Perfect space–time block codes. IEEE Trans. Inf. Theory **52**(9), 3885–3902 (2006)
31. Oggier, F., Viterbo, E., Belfiore, J.-C.: Cyclic Division Algebras: A Tool for Space-Time Coding, vol. 4(1). Foundations and Trends in Communications and Information Theory (2007)
32. Paredes, J.M., Gershman, A.B., Gharavi-Alkhansari, M.: A new full-rate full-diversity space–time block code with nonvanishing determinants and simplified maximum-likelihood decoding. IEEE Trans. Signal Process. **56**(6) (2008)
33. Ren, T.P., Guan, Y.L., Yuen, C., Shen, R.J.: Fast-group-decodable space–time block code. In: Proceedings of the IEEE Information Theory Workshop (2010)
34. Sethuraman, B.A., Rajan, B.S., Shashidhar, V.: Full-diversity, high-rate space–time block codes from division algebras. IEEE Trans. Inf. Theory **49**(10), 2596–2616 (2003)
35. Sirinaunpiboon, S., Calderbank, A.R., Howard, S.D.: Fast essentially maximum likelihood decoding of the Golden code. IEEE Trans. Inf. Theory **57**(6), 3537–3541 (2011)
36. Srinath, K.P., Rajan, B.S.: Low ML-decoding complexity, large coding gain, full-rate, full-diversity STBCs for 2×2 and 4×2 MIMO systems. IEEE J. Sel. Top. Sign. Proces. **3**(6), 916–927 (2009)
37. Srinath, K.P., Rajan, B.S.: Fast-decodable MIDO codes with large coding gain. IEEE Trans. Inf. Theory **60**(2), 992–1007 (2014)
38. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space–time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans. Inf. Theory **44**(2), 744–765 (1998)
39. Vehkalahti, R., Hollanti, C., Lahtonen, J., Ranto, K.: On the densest MIMO lattices from cyclic division algebras. IEEE Trans. Inf. Theory **55**(8), 3751–3780 (2009)
40. Vehkalahti, R., Hollanti, C., Oggier, F.: Fast-decodable asymmetric space–time codes from division algebras. IEEE Trans. Inf. Theory **58**(4), 2362–2385 (2012)
41. Viterbo, E., Boutros, J.: A universal lattice code decoder for fading channels. IEEE Trans. Inf. Theory **45**(7), 1639–1642 (1999)
42. Yang, S., Belfiore, J.-C.: Optimal space–time codes for the MIMO amplify-and-forward cooperative channel. IEEE Trans. Inf. Theory **53**(2), 647–663 (2007)
43. Zheng, L., Tse, D.: Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. IEEE Trans. Inf. Theory **49**(5), 1073–1096 (2003)

Chapter 4
Random Algebraic Lattices and Codes for Wireless Communications

Antonio Campello and Cong Ling

Abstract In this chapter we will review classical and recent advances on "probabilistic" constructions for Euclidean lattices. We will then show recent refinements of these techniques using algebraic number theory. The interest in algebraic lattices is twofold: on the one hand, they are key elements for the construction of sphere packings with the best known asymptotic density; on the other hand, they provide effective solutions to a number of wireless communication problems. We will focus on applications to fading channels, multiple-input-multiple-output (MIMO) channels and to information-theoretic security.

4.1 Introduction

The problem of finding the densest arrangement of spheres in $\mathbb{R}^n$ is a central subject in the Geometry of Numbers, with a variety of well-established connections to Coding Theory. Let Δ_n denote the best possible sphere packing density achievable by a Euclidean lattice of rank n. The celebrated Minkowski-Hlawka theorem (or lower bound), e.g. [7, 10] asserts the existence of lattices with density

$$\Delta_n \geq \zeta(n)/2^{n-1} \tag{4.1}$$

for all $n \geq 2$, where $\zeta(n) = 1 + 1/2^n + 1/3^n + \ldots$ is the Riemann zeta function. Up to very modest asymptotic improvements, to date this is the best known lower bound for Δ_n in high dimensions.

No explicit construction of lattices achieving the lower bound is known. Typical methods for establishing the result rely on random ensembles of lattices and on mean-value arguments [11, 23]. Rush [20] and Loeliger [14] obtained the lower bound from integer lattices constructed from linear codes in $\mathbb{F}_p^n$, in the limit when

A. Campello · C. Ling (✉)
Department of Electrical and Electronic Engineering, Imperial College London, London, UK
e-mail: a.campello@wellcome.ac.uk; c.ling@imperial.ac.uk

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*,
Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_4

$p \to \infty$. A method due to Rogers [19] shows how to obtain packings which provide an improvement to (4.1) by a linear factor. More recent improvements entail the construction of lattices with additional (algebraic) structure. For instance, Vance [25] showed that the best quaternionic lattice improves linearly on (4.1), and Venkatesh [26] resorted to cyclotomic number fields lattices in order to obtain a super-linear improvement on the bound.

In the context of communications, the companion problem to sphere-packings is the one of reliably transmitting information over a Gaussian (AWGN) channel. The related "unconstrained" lattice problem can be stated as follows: Given a normal random vector $\mathbf{z}$ with entries distributed according to $\mathcal{N}(0, \sigma^2)$, what is the unit volume lattice that minimizes the "probability of error", i.e., the probability that $\mathbf{z}$ leaves the Voronoi cell of Λ? See Fig. 4.1b for an illustration. When σ is small, both problems coincide and dense lattices in low dimensions are also good in terms of probability of error. In high dimensions, the connection is also understood to some extent: the same random ensembles used to produce lattices satisfying the

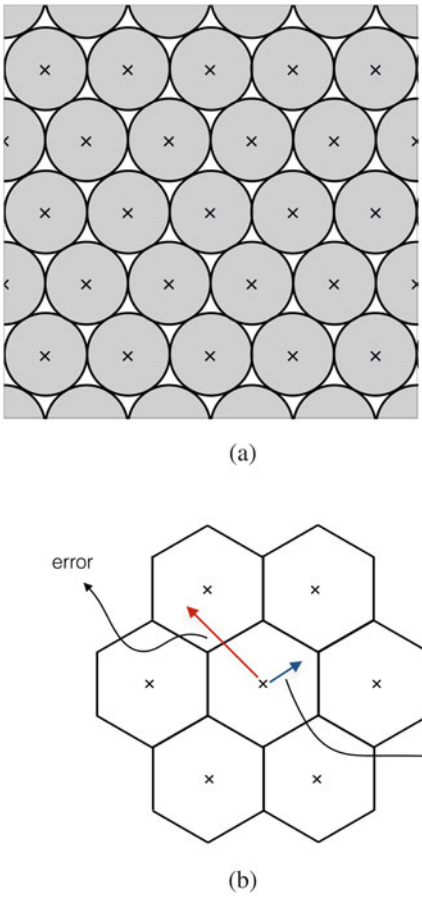


Fig. 4.1 Sphere packing in two dimensions (**a**) and an illustration of error and "no-error" events (**b**)

Minkoswki-Hlawka lower bound can be used to construct optimal lattice codes with vanishing probability of error.

Modern communications, on the other hand, go beyond the Gaussian channel and entails sending information propagated over different media (e.g., fading channels), using multiple antennas (e.g., MIMO channels) and to various users (e.g., relay networks). For such applications, it is desirable to enrich the lattices with some algebraic (multiplicative) structure, often inherited by the properties of number fields. Interestingly, lattices with precisely the same structure as the ones in the works of Vance [25] and Venkatesh [26] play a crucial role in some of these applications. There has been a recent increase in the literature on this relation and the applications of high dimensional algebraic lattices to various problems. In this text we provide a glimpse of some of these relations, showing how algebraic constructions can be advantageous from a mathematical and applications point of view.

4.1.1 Structure

The objective of this chapter is twofold. In the first part we provide a self-contained exposition of random lattices and their packing density. In the second part, we show how such lattices can be applied to building effective reliable and secure transmission schemes for wireless communications. We will focus on reliable and secure communications over block-fading and MIMO channels.

The content of Sect. 4.2 is fairly classical. We exhibit the original "analytical" method of finding dense lattices due to Davenport and Rogers, and Rogers' argument for obtaining the linear improvements depicted in Table 4.1. Further information on these and related results can be found individually in excellent references in the literature, e.g. [7, 19] or [10]. For a more "modern" and "information-theoretic" treatment on the analysis of random ensembles from codes, one can consult [28, Chapter 7].

The improvements in Sect. 4.3 are more recent and are not present in textbooks. We have thus attempted to include them in a general framework so as to provide a self-contained description. A suitable framework is based on "generalized reductions", as recently defined in [4]. In particular, this allows us to recover

Table 4.1 Best improvements on the lower bound

Dimension	$r(n)$	Reference
$n \in \mathbb{N}$	2	[11]
$n \in \mathbb{N}$	$\log(2(e-e^{-1})^{-1}n)$	[19]
$n = 4m$	$\log(24e^{-1}m)$	[25]
$n = 2\phi(m)$	$\log(2m)$	[26]
n large	$65{,}963.8n$	[26]

"coding-theoretic" versions of the results in Vance [25] for Hurwitz lattices and Venkatesh [26] for cyclotomic lattices.

Sections 4.4 and 4.5 are mainly based on [5] and [13]. The objective of these sections is to provide a glimpse of how random algebraic lattices can be used in applications to information security of wireless communications.

4.1.2 Summary of Results

Since the work of Hlawka that established (4.4), there have been many attempts to refine the lower bound, by establishing that

$$\log_2 \Delta_n \geq -n + r(n), \tag{4.2}$$

where $r(n) \geq 0$ is an "improvement" term (see Table 4.1). For all known improvements $r(n) = O(\log n)$, with the bound in [26] being slightly better than previous results (it can produce an extra $O(\log\log\log n)$ term in $r(n)$). These improvements are very modest in comparison to (4.1). On the other hand, there is no known upper bound of the form $r(n) = o(n)$, as one could expect. The best available bound on the literature is $r(n) \leq 0.41n$ for large n. Furthermore, if one considers more general (non-lattice) packings, there is experimental evidence that linear ($\Theta(n)$) improvements in $\log \Delta_n$ (i.e. exponential improvements on the density) are possible.

For the communications problem, the asymptotic behavior of lattices (and even non-lattices) as $n \to \infty$ is much better understood. A result due to Poltyrev [18] shows that any sequence of lattices $\Lambda_1, \Lambda_2, \ldots$ of increasing dimension and vanishing probability of error must satisfy

$$\lim_{n\to\infty} \sup \log V(\Lambda)^{1/n} > \log(2\pi e\sigma^2), \tag{4.3}$$

and the bound is achieved by random lattices [14]. Explicit constructions that achieve the bound have been recently found [27]. It is perhaps surprising that, in terms of packing density, these construction are very far from the Minkowski-Hlawka lower bound.

4.1.3 Notation

The Euclidean norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is denoted by $\|\mathbf{x}\| = (x_1^2 + \ldots + x_n^2)^{1/2}$. The ball of radius r in $\mathbb{R}^n$ is denoted by $B_r^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq r\}$. A *lattice* Λ is a discrete additive subgroup of $\mathbb{R}^n$. Denote by span Λ the minimal subspace of $\mathbb{R}^n$ that contains Λ. The *rank* of Λ is defined to be the dimension of span Λ. The

quotient (span Λ)/Λ is compact, and its volume, denoted by $V(\Lambda)$, is said to be the *volume of* Λ. It is also the volume of the Voronoi cell

$$\mathcal{V}_\Lambda = \{\mathbf{x} \in \text{span } \Lambda : \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{y} \in \Lambda\}.$$

The *first minimum* $\lambda_1(\Lambda)$ of Λ is the shortest Euclidean norm of non-zero vectors in Λ. In general, the i-th minima of Λ are defined as

$$\lambda_i(\Lambda) = \min\left\{r : \dim\left(\text{span}\left\{B_r^n \cap \Lambda\right\}\right) = i\right\}.$$

The *packing density* of a rank m lattice Λ is defined as

$$\Delta(\Lambda) = \frac{\text{vol } B_{\lambda_1/2}^m}{V(\Lambda)}.$$

In general, if Λ can pack a measurable set S (i.e., the translates of S by vectors of Λ are disjoint), then its packing density or efficiency can be defined as the ratio

$$\frac{\text{vol } S}{V(\Lambda)}.$$

4.2 Classical Methods

According to a translation due to Gruber [10], Hlawka described his proof of the lower bound as follows: "(...) consider the problem of catching fish of given length from a pond. Making one haul, one may catch such a fish only by chance. For this reason it makes sense to catch many fish, hoping that a fish of the desired length is among them. In probability theory this is called a random sample."

This description refers to the *probabilistic method* of analysing a random ensemble of lattices rather than individual instances. The method is nowadays well-established and widely used for a variety of discrete problems, including sphere-packing. Four years after the publication of Hlawka's proof, Shannon [22] used the probabilistic method, or "random coding" in the information theory jargon, to establish the existence of *capacity-achieving* codes. The resemblance is not incidental: as observed later by several authors [14, 18], and as we will see later, the Minkowski-Hlawka random argument essentially implies the existence of capacity-achieving codes for the Gaussian channel.

4.2.1 Random Lattices

4.2.1.1 Overview

In a nutshell, the analysis of the packing density is done indirectly by analyzing the lattice point enumerator. Suppose S is a convex set (for example, a ball), and let

$$\mathcal{N}_S(\Lambda) = \#(\Lambda \setminus \{0\} \cap S)$$

be the number of non-zero lattice points in S. If $\mathcal{N}_S(\Lambda) = 0$, then the translations of the set $S/2 = \{x/2 : x \in S\}$ by points of Λ are disjoint or, in other words, Λ can pack $S/2$. This packing has density

$$\frac{\text{vol } S/2}{V(\Lambda)} = \frac{\text{vol } S}{2^n V(\Lambda)}.$$

However, finding a lattice with small (or zero) $\mathcal{N}_S(\Lambda)$ is usually hard. Therefore, we opt to analyse the *average* behaviour of $\mathcal{N}_S(\Lambda)$ over a sufficiently large family of lattices, say $\mathbb{L}$, and guarantee that $\mathbb{E}_{\mathbb{L}}[\mathcal{N}_S(\Lambda)]$ is small.

These are the essential ideas for the establishment of the bound (4.1). The main differences between the various proofs in the literature is in the way of constructing $\mathbb{L}$. We present below a method due to Davenport-Rogers, which can be found in classical textbooks like [7]. An ensemble proposed by Loeliger [14] is particularly popular in applications, due to its relation to classical error-correcting codes. We will describe it in a generalized way in Sect. 4.3.4.

In what follows, we adopt the "information-theoretic" terminology in [28, Ch. 7] for random lattices. We say that a collection of lattices $\mathbb{L}$ of the same volume $V > 0$ is a Minkowski-Hlawka-Siegel (MHS) ensemble if:

$$\mathbb{E}_{\mathbb{L}}\left[\mathcal{N}_S(\Lambda)\right] = \frac{\text{vol } S}{V}, \tag{4.4}$$

for any measurable set S (in the sense of Jordan). Such an average implies the lower bound $1/2^{n-1}$ in the following way. From (4.4), it follows that it must exist at least one $\Lambda \in \mathbb{L}$ such that $\mathcal{N}_\Lambda(r) \leq (\zeta(m)V)^{-1}\text{vol } B_r$, where B_r denotes a ball of radius r. Now if we force the right-hand side to be equal to $2(1-\varepsilon)$, for some small $\varepsilon > 0$, then, since a lattice has at least two minimum vectors, we must have $\mathcal{N}_\Lambda(r) = 0$. Therefore Λ can pack balls of radius $r/2$; rearranging the terms gives us density:

$$\Delta = \frac{2(1-\varepsilon)}{2^{n-1}}.$$

which is, up to ε, the Minkowski-Hlawka bound. If Λ is a lattice with a guaranteed number of minimum vectors (say, L) we can, by similar arguments, achieve density $L(1-\varepsilon)/2^{nt}$.

Rogers-Davenport Proof Before stating the proof, we recall some facts about measurable sets. Notice that if $S \subset \mathbb{R}^n$ is a measurable set (in the sense of Jordan), then its volume can be calculated by discretizing it with fine scalings of the $\mathbb{Z}^n$ lattice, which gives us the formula:

$$\text{vol } S = \lim_{\beta \to \infty} \left(\mathcal{N}_S(\beta^{-1}\mathbb{Z}^n)\beta^{-n} \right). \tag{4.5}$$

Also notice that we can calculate the volume of S by slicing it into parallel hyperplanes and making the distance between successive hyperplanes tend to zero. For instance, consider

$$H = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : x_n = 0 \right\},$$

with normal vector $\mathbf{e}_n = (0, \ldots, 0, 1)$. We have:

$$\text{vol } S = \lim_{\rho \to \infty} \sum_{z \in \rho^{-1}\mathbb{Z} \setminus \{0\}} \rho^{-1} \text{vol}_{\,n-1}(S \cap (H + z\mathbf{e}_n)), \tag{4.6}$$

where $\text{vol}_{\,n-1}(S \cap (H + z\mathbf{e}_n)$ denotes the volume of the $(n-1)$-dimensional sets $S \cap (H + z\mathbf{e}_n)$ (i.e., its $(n-1)$-dimensional Jordan measure in the space $H + z\mathbf{e}_n$).

With the above facts in mind, Rogers-Davenport construct the random ensembles $\mathbb{L}_\rho$, $\rho > 0$, as follows. For a number $\rho > 0$ and a vector $\mathbf{u} \in \mathbb{R}^{n-1}$ in the cube $[0, \rho^{1/n-1})^{n-1}$, define the lattice

$$\Lambda(\mathbf{u}, \rho) = V^{-1/n} \left\{ \rho^{1/n}(\mathbf{x}, 0) + l(\mathbf{u}, \rho^{-1}) : (\mathbf{x}, l) \in \mathbb{Z}^{n-1} \right\}.$$

In other words, $\Lambda(\mathbf{u}, \rho)$ is the lattice generated by the columns of

$$V^{-1/n} \begin{pmatrix} \rho^{\frac{1}{n-1}} & 0 & 0 & \ldots & 0 & 0 \\ 0 & \rho^{\frac{1}{n-1}} & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \rho^{\frac{1}{n-1}} & 0 \\ u_1 & u_2 & u_3 & \ldots & u_{n-1} & \rho^{-1} \end{pmatrix}.$$

We then define the ensemble

$$\mathbb{L}_\rho = \left\{ \Lambda(\mathbf{u}, \rho) : u_i \in [0, \rho^{1/n}) \text{ for } i = 1, \ldots, n-1 \right\}. \tag{4.7}$$

Suppose a lattice in $\mathbb{L}_p$ is chosen by picking a point $\mathbf{u}$ in the cube uniformly at random. We have the following:

Theorem 4.1 *The ensemble defined in* (4.7) *satisfies*

$$\lim_{\rho\to\infty} \mathbb{E}_{\mathbb{L}_\rho}[\mathcal{N}_S(\Lambda)] = \frac{vol\ S}{V}. \tag{4.8}$$

Proof By scaling Λ appropriately we can suppose without loss of generality that $V = 1$. Let $C_\rho = [0, \rho^{1/(n-1)})^{n-1}$ be the cube with side-length $\rho^{1/(n-1)}$ and volume ρ. For ρ sufficiently large, since S is bounded, there is no point in S of the form $(\rho^{1/(n-1)}\mathbf{x}, z)$, with $\mathbf{x} \in \mathbb{Z}^{n-1}\setminus\{0\}$. The average of $\mathcal{N}_S$ over the ensemble is thus given by:

$$\begin{aligned}
\frac{1}{\rho}\int_{C_\rho} \mathcal{N}_S(\Lambda(\mathbf{u},\rho))d\mathbf{u} &= \sum_{z\in\mathbb{Z}\setminus\{0\}} \frac{1}{\rho}\int_{C_\rho} \mathcal{N}_{S\cap(H+\rho z\mathbf{e}_n)}(\Lambda(\mathbf{u},\rho))d\mathbf{u} \\
&= \sum_{z\in\mathbb{Z}\setminus\{0\}} \frac{1}{\rho}\int_{C_\rho} \mathcal{N}_{S\cap(H+\rho z\mathbf{e}_n)-\rho z\mathbf{e}_n}(\Lambda(\mathbf{u},\rho))d\mathbf{u} \\
&= \sum_{z\in\mathbb{Z}\setminus\{0\}} \rho^{-1}\mathrm{vol}_{\,n-1}((S\cap(H+\rho z\mathbf{e}_n)) - \rho z\mathbf{e}_n) \\
&= \sum_{z\in\mathbb{Z}\setminus\{0\}} \rho^{-1}\mathrm{vol}_{\,n-1}((S\cap(H+\rho z\mathbf{e}_n))).
\end{aligned}$$

From (4.6), the last equation tends to S as $\rho\to\infty$, finishing the proof. □

The limit in Theorem 4.1 is slightly weaker than Eq. (4.4), and it allows us to recover the Minkowski-Hlawka lower bound (4.1) up to a factor (say) $(1-\varepsilon)$ in the numerator, for any small (but positive) ε. To obtain the bound with $\varepsilon = 0$, we can resort to a compactness argument due to Mahler [7]. We omit the details.

Note that the lattice point enumerator

$$N_S(\Lambda) = \sum_{x\in\Lambda\setminus\{0\}} \mathbb{1}_S(x) \tag{4.9}$$

can be replaced by the sum of any integrable function (in the sense of Riemann) that vanishes outside a compact set, with essentially the same proof. In this case we obtain

$$\mathbb{E}_{\mathbb{L}_\rho}[\sum_{\mathbf{x}\in\Lambda\setminus\{0\}} f(\mathbf{x})] = V^{-1}\int_{\mathbb{R}^n} f(\mathbf{x})d\mathbf{x}. \tag{4.10}$$

As observed by Siegel [23, Rmk. 1, p. 346], the theorem can be further generalized to integrable functions that decay sufficiently fast with the norm of

$\mathbf{x}$. A sufficient condition is that $f(x)$ decays faster than the harmonic series, i.e. $f(x) \leq c/(1+\|\mathbf{x}\|)^{n+\delta}$, for constants $c, \delta > 0$. A function of particular interest satisfying this condition is $f(x) = e^{-\tau\|x\|^2}$ for $\tau > 0$, for which the sum over a lattice is known as its *theta series*

$$\Theta_\Lambda(\tau) = \sum_{x\in\Lambda} e^{-\tau\|x\|^2}.$$

The average behavior of the theta series becomes

$$E_{\mathbb{L}}[\Theta_\Lambda(\tau)] = V^{-1}\left(\frac{\pi}{\tau}\right)^{n/2} + 1. \tag{4.11}$$

As the dimension increases, for a fixed volume V, the point $\tau = \pi$ corresponds to a phase transition. If $\tau > \pi$ the theta series vanishes when the dimension increases, whereas for $\tau < \pi$ grows unboundedly.

4.2.2 Primitive Points

The extra term $\zeta(n)$ in the enumerator of (4.1) is obtained by considering *primitive lattice points*. A lattice point $\mathbf{x} \in \Lambda$ is said to be primitive if the equivalent conditions hold:

1. $\mathbf{x}$ is part of a basis for Λ
2. $\mathbf{x}$ is *visible* from the origin, i.e. the line segment $\{\lambda\mathbf{x} : \lambda \in (0,1)\}$ contains no point of Λ.
3. The greatest common divisor of the coefficients of $\mathbf{x}$ when written as a linear combination of a basis for Λ is equal to one.

An illustration of Condition 2 for the Eisenstein lattice (cf. Sect. 4.3.1) is given in Fig. 4.2. It is perhaps surprising that the "fraction" of primitive vectors of a lattice is very close to 1, even for small dimensions (the precise number is $1/\zeta(n)$, which tends to 1 very quickly as $n \to \infty$). Loosely speaking, in high dimensions almost any lattice point can be extended to a basis for Λ.

Statements on the sum of non-zero lattice points can usually be translated into statements for primitive points by means of the so-called Möbius inversion, namely:

$$\sum_{\mathbf{x}\in\Lambda'} f(\mathbf{x}) = \sum_{r=1}^{\infty} \frac{\mu(r)}{r^n} \sum_{\mathbf{x}\in\Lambda\setminus\{0\}} r^{-n} f(r\mathbf{x}),$$

where μ is the *Möbius function* (cf. [7]) that satisfies the identity

$$\sum_{r=1}^{\infty} \frac{\mu(r)}{r^n} = \frac{1}{\zeta(n)}.$$

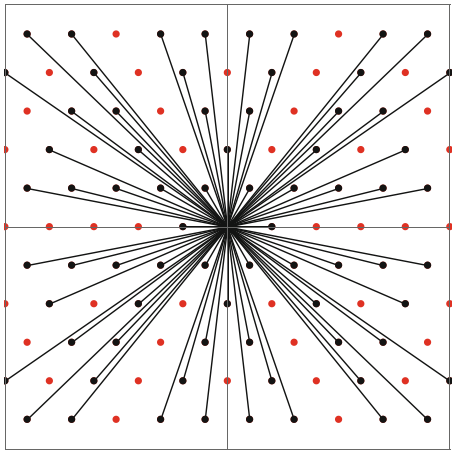


Fig. 4.2 Primitive (black) points in the Eisenstein lattice. Non-primitive points are depicted in red. Black points are visible from the origin, whereas red points have at least one other lattice point blocking the view

For simplicity of the statements, and in order to highlight the main ideas of the theorems, we omit the analysis of primitive points.

4.2.3 Linear Improvement

Rogers was the first author to obtain a linear improvement on the Minkowski-Hlawka theorem [19]. Instead of only looking at the packing radius directly, a new insight in Rogers' method is that the Minkowski-Hlawka theorem actually tells us more information about the successive minima of a lattice.

Let us define the successive densities of a lattice as

$$\Delta_i(\Lambda) = \frac{\text{vol } B^n_{\lambda_i(\Lambda)/2}}{V(\Lambda)}, i = 1, \ldots, n, \tag{4.12}$$

where $\lambda_i(\Lambda)$ are its successive minima. Notice that $\Delta_i(\Lambda)$ is not strictly speaking a "density" and it may in principle be greater than 1 for $i \geq 2$. However the following lemma (which we refer to as Minkowski's lemma), states that from a lattice with good average i-th densities we can construct a lattice with good density. We present a "matrix-based" proof below.

Lemma 4.1 (Minkowski) *Let Λ be a rank n-lattice with*

$$\left(\prod_{i=1}^{n} \Delta_i(\Lambda)\right)^{1/n} = \delta. \tag{4.13}$$

Then there exists another rank-n lattice $\tilde{\Lambda}$ with packing density $\Delta(\tilde{\Lambda}) = \Delta_1(\tilde{\Lambda}) = \delta$.

Proof Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \Lambda$ be linearly independent vectors that achieve the successive minima ($\|\mathbf{v}_i\| = \lambda_i(\Lambda)$). Let Q be the matrix whose columns are the Gram-Schmidt orthogonalization of the vectors $\mathbf{v}_i$ and $D = \text{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$. We claim that

$$\tilde{\Lambda} = \left\{ \mathbf{y} = Q^t D Q \mathbf{x} : \mathbf{x} \in \Lambda \right\} \tag{4.14}$$

satisfies the conditions of the theorem. To calculate the first minimum of $\tilde{\Lambda}$, we first write

$$\|\mathbf{y}\|^2 = \mathbf{x}^t Q^t D^2 Q \mathbf{x} = \mathbf{w}^t D^2 \mathbf{w} \geq \|\mathbf{w}\|^2 \lambda_k^{-2},$$

where $\mathbf{w} = Q\mathbf{x}$ and k is the smallest index such that $w_k \neq 0$ and $w_j = 0$ for $j > k$. Notice that by construction $\mathbf{x}$ is linearly independent of $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$ and thus $\|\mathbf{x}\| \geq \lambda_k$. Therefore $\lambda_1(\tilde{\Lambda}) \geq 1$, with equality achieved by $\mathbf{y} = Q^t D Q \mathbf{v}_1 \in \Lambda$. The density of $\tilde{\Lambda}$ is thus:

$$\Delta(\tilde{\Lambda}) = \frac{\text{vol } B_{\lambda_1(\tilde{\Lambda})/2}}{V(\tilde{\Lambda})} = (\lambda_1(\Lambda) \ldots \lambda_n(\Lambda)) \frac{\text{vol } B_{1/2}}{V(\Lambda)} = \delta.$$

□

Theorem 4.2 *For any $\varepsilon > 0$, there exists a lattice Λ with*

$$\left(\prod_{i=1}^{n} \Delta_i(\Lambda) \right)^{1/n} \geq \frac{2n\zeta(n)(1-\varepsilon)}{e(1-e^{-n})}. \tag{4.15}$$

Proof Consider the radial function $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \|\mathbf{x}\| \leq re^{(1-n)/n} \\ \frac{1}{n} - \log\left(\frac{\|\mathbf{x}\|}{r}\right) & \text{if } \|\mathbf{x}\| \in (re^{(1-n)/n}, re^{1/n}\,] \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors achieving the successive minima. Notice that if we guarantee that $\mathbf{v}_i$ are in the spherical-shell given by Condition 2 in the definition of f, then

$$\sum_{i=1}^{n} f(\mathbf{v}_i) = 1 - \log\left(\frac{\prod_{i=1}^{n} \lambda_i}{r^n} \right),$$

therefore upper-bounding f is the same as lower bounding the product of the minima (and consequently the successive densities). That is precisely the proof strategy we will follow.

Integrating f using hyperspherical coordinates, we get

$$\int_{\mathbb{R}^n} f(\mathbf{x})d\mathbf{x} = r^n V_n \frac{e(1-e^{-n})}{n}. \tag{4.16}$$

From Theorem 4.2, there exists Λ such that

$$\sum_{i=1}^{n} f(\mathbf{v}_i) = 2\sum_{i=1}^{n} f(\mathbf{v}_i) < \sum_{\mathbf{x}\in\Lambda\setminus\{0\}} f(\mathbf{x}) \leq \int_{\mathbb{R}^n} f(\mathbf{x})d(\mathbf{x}). \tag{4.17}$$

Now if we choose r such that the right-hand side of (4.16) is lesser or equal than $2(1-\varepsilon)$, we conclude that

$$\sum_{i=1}^{n} f(\mathbf{v}_i) = 1 - \log\left(\frac{\prod_{i=1}^{n}\lambda_i}{r^n}\right) \leq 1,$$

therefore $\lambda_1 \ldots \lambda_n > r^n$. Rearranging the terms and using the definition of successive minima gives us the bound. □

Again, we can remove the factor $(1-\epsilon)$ in the above Theorem by a compacity result due to Mahler. In the proof of Rogers notice that Eq. (4.17) uses the fact that Λ is closed under reflexion in the origin, i.e., under the action of the multiplicative group $\{1,-1\} \sim C_2$. This raises the natural question whether lattices with a larger symmetry group could improve the bound. This is precisely the nature of the improvements of [25] given in the next section.

4.3 Random Algebraic Lattices

In what follows we describe three examples of important families of lattices and their sublattices.

4.3.1 Eisenstein Integers $\mathbb{Z}[\omega]$

As an illustrative example, consider the ring of Eisenstein integers

$$\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}[\omega]\},$$

where $\omega = (-1 + i\sqrt{3})/2$. It has the largest unit group $\mathcal{U}$ among all complex quadratic ring of integers. Indeed, $\mathcal{U}$ is the cyclic group of order 6 given by

$$\mathcal{U} = \left\{\pm 1, \pm\omega, \pm\omega^2\right\},$$

with primitive elements $-\omega$ and $-\omega^2 = \omega + 1$. We have $\mathbb{Z}[\omega]\mathbb{Q} = \mathbb{Q}(\sqrt{-3}) \subset \mathbb{C}$. A $\mathbb{Z}[\omega]$-lattice is a free $\mathbb{Z}[\omega]$ sub-module of $\mathbb{C}^n$ with a free basis $\{b_1, \ldots, b_n\}$ which generates $\mathbb{C}^n$ as a vector space over $\mathbb{C}$. A $\mathbb{Z}[\omega]$-lattice is closed under multiplication by Eisenstein integers and, in particular by elements of $\mathcal{U}$. Indeed, the group $\mathcal{U}$ acts freely on the set of elements of the same norm, which means that each "layer" of a $\mathbb{Z}[\omega]$-lattice contains at least 6 vectors.

The trace form $\mathrm{tr}(x, y) = x\overline{y}$ induces a Hermitian inner product in the complex space $\mathbb{C}^n$, where $\overline{y}$ stands for the conjugate of y. We define the k-th complex minimum of an Eisenstein lattice as

$$\lambda_k^{\mathbb{C}}(\Lambda) = \min\left\{r : \dim_{\mathbb{C}} \mathrm{span}\left\{\mathbf{x} \in \Lambda : \sqrt{\mathrm{tr}(\langle \mathbf{x}, \mathbf{x}\rangle)} \leq r\right\} = k\right\}.$$

In other words, $\lambda_k^{\mathbb{C}}$ is the minimum value such that the ball of radius r contains k linearly independent vectors over $\mathbb{C}$.

Example 4.1 The exceptional lattice E_6 [8, p. 126] produces the densest packing in $\mathbb{R}^6$ [3]. It is generated by the vectors $(\theta, 0, 0)$, $(0, \theta, 0)$ and $(1, 1, 1)$ in $\mathbb{C}^3$, where $\theta = i\sqrt{3} = \omega - \bar{\omega}$. Its three complex minima are equal to $\sqrt{3}$ and achieved by the basis vectors.

Sub-lattices Ideals $\mathfrak{p} \subset \mathbb{Z}[\omega]$ produce complex sub-lattices of $\mathbb{Z}[\omega]$. For instance if $a \in \mathbb{Z}[\omega]$ with $N(a) = a\overline{a} = p$ a rational prime, then the quotient $\mathbb{Z}[\omega]/\langle a\rangle \sim \mathbb{F}_p$.

4.3.2 Cyclotomic Lattices

Let $K = Q(\zeta)$ be the cyclotomic field of degree $n = \phi(m)$, where ζ is a n-th root of unity. A cyclotomic lattice in K^t is a $\mathbb{Z}[\zeta]$-module. It can be embedded in $\mathbb{R}^n$ from the cyclotomic embeddings $\sigma_i(\zeta) = \zeta^i$. Multiplication by elements of $\mathbb{Z}[\zeta]$ translates into multiplication by a diagonal matrix in $\mathbb{R}^n$. Indeed, if $\overline{\Lambda}$ is a cyclotomic lattice, $\mathbf{u} \in \overline{\Lambda}$ and $a \in \mathbb{Z}[\zeta]$, then

$$\sigma(a\mathbf{u}) = (D_a \otimes I_t)\sigma(\mathbf{u}) = \begin{pmatrix} D_a & & & \\ & D_a & & \\ & & \ddots & \\ & & & D_a \end{pmatrix} \begin{pmatrix} \sigma(u_1) \\ \sigma(u_2) \\ \vdots \\ \sigma(u_t) \end{pmatrix},$$

where D_a is the diagonal matrix with elements $\sigma_1(a), \ldots, \sigma_n(a)$.

The following simple observation is key to the results of Venkatesh [26].

Proposition 4.1 *A cyclotomic lattice of degree $n = \phi(m)$ has at least m distinct shortest vectors.*

Proof Let $\mathbf{u} \in \overline{\Lambda}$ be a vector of Euclidean norm r, where $\overline{\Lambda}$ is a $\mathbb{Z}[\zeta]$-lattice. The vectors $\mathbf{u}_i = \zeta^i \mathbf{u}$, $i = 0, \ldots, m-1$ belong to $\overline{\Lambda}$ and have Euclidean norm:

$$\begin{aligned}\|\sigma(\mathbf{u}_i)\|^2 &= \sum_{j=1}^{n} \left\|\sigma_j(\zeta^i \mathbf{u})\right\|^2 = \sum_{j=1}^{n} |\sigma_j(\zeta^i)|^2 \left\|\sigma_j(\mathbf{u})\right\|^2 \\ &= \sum_{j=1}^{n} \left\|\sigma_j(\mathbf{u})\right\|^2 = \|\sigma(\mathbf{u})\|^2 = r^2.\end{aligned}$$

This shows that the group of cyclotomic units acts freely on the "layers" of $\overline{\Lambda}$ and proves the assertion of the proposition. □

Indeed, $\phi(m)$ of such vectors (corresponding to the distinct primitive roots of unity) are linearly independent over $\mathbb{R}$, i.e., the set of vectors of minimum norm of a cyclotomic lattice generate $\mathbb{R}^{tn}$. This property is referred in the literature to as *well-roundness*.

Example 4.2 If ζ is a p-th root of unity, and $\mathfrak{b}_i = (1-\zeta)^i \mathbb{Z}[\zeta]$ is the ideal generated by $(1-\zeta)^i$, the lattice obtained by embedding $\mathfrak{b}_i$ in $\mathbb{R}^{p-1}$ is called *Craig's lattice*, denoted by $A_{p-1}^{(i)}$. If i is choosen to be $l = \lfloor n/2 \log(n+1) \rceil$, a Craig's lattice has ratio:[1]

$$\frac{\lambda_1(\Lambda)}{V(\Lambda)^{1/(p+1)}} \geq \sqrt{\frac{2\pi}{\log n}} \left(\sqrt{\frac{n}{2\pi e}} + o(1) \right).$$

From this we obtain

$$\log_2 \Delta \gtrapprox -(1/2) \log\log n.$$

This is considerably weaker than (4.2) for high dimensions, but impressive for such a simple construction.

Sublattices Similarly to the Eisenstein lattices, one can construct ideal lattices with quotient equivalent to $\mathbb{F}_p$ as follows. Let $p \equiv 1 \pmod{\phi(m)}$. The ideal $p\mathbb{Z}[\zeta]$ can be factorized into $\phi(m)$ distinct prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_{\phi(m)}$. For instance, a prime can be factorized from the factorization of the corresponding cyclotomic polynomial modulo p. For any of these ideals we have $\mathbb{Z}[\zeta]/\mathfrak{p}_i \sim \mathbb{F}_p$.

[1] The square of this number is known as the *Hermite constant*.

Remark 4.1 This construction can be generalized to any (totally real or CM) field K, by considering its ring of integers O_K and a prime ideal $\mathfrak{p}$ above a prime that splits.

4.3.3 Lipschitz and Hurwitz Lattices

The quaternion skew-field $\mathbb{H}$ is given by

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\},$$

with the usual relations $k = ij$, $i^2 = j^2 = -1$ and $ij = -ji$. It has an Hermitian structure by considering the inner product $\langle x, y\rangle = x\overline{y}$, where $\overline{a + bi + (c + di)j} = a - bi - cj - dk$. The skew-field of quaternions can be identified with $\mathbb{R}^4$ under the natural mapping

$$\sigma(a + bi + cj + dk) = (a, b, c, d) \in \mathbb{R}^4, \tag{4.18}$$

with usual inner product. Lattices in $\mathbb{H}^n$ can be constructed from orders in $\mathbb{H}$. An order in $\mathbb{H}$ is a $\mathbb{Z}$-lattice which is also a subring of $\mathbb{H}$. A "natural way" of constructing an order in $\mathbb{H}$ is by taking

$$\mathcal{L} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{Z}\}.$$

This corresponds to the set of *Lipschitz integers*. This order is, however, not maximal, i.e., it is strictly contained in a bigger order. The *Hurwitz* order $\mathcal{H}$ is the maximal quaternionic order defined by:

$$\mathcal{H} = \{a + bi + cj + d(-1 + i + j + ij)/2 : a, b, c, d \in \mathbb{Z}\}.$$

By considering the mapping (4.18), the orders $\mathcal{L}$ and $\mathcal{H}$ correspond to $\mathbb{Z}$-lattices in $\mathbb{R}^4$. Indeed, $\sigma(\mathcal{L}) = \mathbb{Z}^4$, whereas $\sigma(\mathcal{H}) = D_4$ is the checkerboard lattice in dimension four [8, Sec. 7.2]. The Lipschitz order $\mathcal{L}$ has index 2 over the Hurwitz order $\mathcal{H}$.

A lattice in $\mathbb{H}^m$ is called a *Hurwitz* (resp. Lipschitz) lattice if it is a $\mathcal{H}$ left-module (resp. $\mathcal{L}$-module). Hurwitz lattices are, in particular, invariant under multiplication by elements of the Hurwitz unit group

$$\mathcal{H}^\times = \{\pm 1, \pm i, \pm j, \pm k, (\pm 1 \pm i \pm j \pm k)/2\},$$

which has order 24. Similarly to cyclotomic lattices, the units $\mathcal{H}^\times$ acts freely in the set of vectors of the same norm of a Hurwitz lattice $\Lambda \in \mathcal{H}$, meaning that each shell

has at least 24 vectors. The matrix-representation of a quaternion $x + yj$, $x, y \in \mathbb{C}$ is given by

$$\begin{pmatrix} x & -\overline{y} \\ y & \overline{x} \end{pmatrix}.$$

The *reduced norm* nrd of a quaternion is the discriminant of the corresponding matrix. We have

$$\text{nrd}(a + bi + cj + dk) = a^2 + b^2 + c^2 + d^2.$$

Sublattices One can consider (left) ideals in $\mathcal{H}$ to construct sub-lattices. For instance the sublattice corresponding to $p\mathcal{H}$, where p is a prime, has index p^4 in $\mathcal{H}$ and

$$\mathcal{H}/p\mathcal{H} \sim M_2(\mathbb{F}_p),$$

where $M_2(\mathbb{F}_p)$ is the ring of matrices with entries in $\mathbb{F}_p$. A possible ring isomorphism is obtained by setting

$$\phi(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \phi(i) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } \phi(j) = \begin{pmatrix} a & b \\ beta & -a \end{pmatrix},$$

where a and b are two integers such that $a^2 + b^2 \equiv -1 \pmod{p}$. Notice that such an isomorphism preserves the residue class of the reduced norm, i.e. $nrd(x) = \det\phi(x) \pmod{p}$, for any $x \in \mathcal{H}$.

4.3.4 A General Construction

It is possible to define a general construction that may be specialized in various ways and subsumes a number of constructions in the literature. We first recall the definition of a linear code.

Definition 4.1 Let $\mathbb{F}_p$ be a field with p elements, where p is a prime or a prime power. A k-dimensional vector subspace $C \subset \mathbb{F}_p^n$ is called a (linear) *code* with parameters (n, k, p) (or simply an (n, k, p)-code).

Let Λ be a rank m lattice and let $n \leq m$ be an integer. We define a reduction as follows:

Definition 4.2 Let $\phi_p : \Lambda \to \mathbb{F}_p^n$ be a surjective homomorphism. Given a linear code C, its associated lattice via ϕ_p is defined as

$$\Lambda_p(C) \triangleq \phi_p^{-1}(C).$$

It is not hard to see from the above definition that $\Lambda_p(C)$ is indeed a rank-m lattice. The reduction nests $\Lambda_p(C)$ between the base (fine) lattice Λ and the kernel (coarse) lattice $\Lambda_p = \ker \phi_p$. This definition generalizes the idea of lifting a code to the Euclidean space by shifting the codewords through vectors of $p\mathbb{Z}^n$, the so-called Construction A.

Example 4.3 By considering $\Lambda = \mathbb{Z}^n$, $\Lambda_p = p\mathbb{Z}^n$ and ϕ_p the componentwise reduction modulo p we recover the so-called Construction A, which was used by Rush [20] and Loeliger [14] to construct dense packings. Various important lattices can be built via Construction A. Most of the works on applications of lattices to communications hugely rely on Construction A [28]. See Fig. 4.3a for an illustration of ϕ_3.

Example 4.4 By replacing the ring $\mathbb{Z}$ by the ring of integers of a number field, we enable a number of different constructions. As an illustration, let $\mathbb{Q}[\sqrt{13}]$ be the quadratic field with ring of integers $\mathbb{Z}[\mu]$, where $\mu = \frac{1+\sqrt{13}}{2}$. The rational prime $3 = -\mu\bar{\mu}$ splits and the ideal $\mathfrak{p} = \mu\mathbb{Z}[\mu]$ is such that $\mathbb{Z}[\mu]/\mathfrak{p} \sim \mathbb{F}_3$. With a slight abuse of notation, define

$$\phi_p(x) = (x \pmod{\mu}, \overline{x} \pmod{\mu}),$$

where the modulo-μ operation identifies a point with its representative in $\mathbb{Z}[\mu]/\mathfrak{p}$, which can be chosen to be in $\{-1, 0, 1\}$.

We have $\ker \phi_p = 3\mathbb{Z}[\mu]$. Let $\sigma(x) = (x, \overline{x})$ be the embedding of x in $\mathbb{R}^2$. One possible full set of representatives for the quotient $\mathbb{Z}[\mu]/3\mathbb{Z}[\mu] \sim \mathbb{F}_3^2$ is

$$0 \xrightarrow{\sigma} (0,0) \xrightarrow{\phi_p} (0,0), 1 \xrightarrow{\sigma} (1,1) \xrightarrow{\phi_p} (1,1), -1 \xrightarrow{\sigma} -(1,1) \xrightarrow{\phi_p} -(1,1),$$

$$\mu \xrightarrow{\sigma} (\mu, \overline{\mu}) \xrightarrow{\phi_p} (0,1), (\mu+1) \xrightarrow{\sigma} (\mu+1, \overline{\mu}+1) \xrightarrow{\phi_p} (1,-1),$$

$$\bar{\mu} \xrightarrow{\sigma} (\overline{\mu}, \mu) \xrightarrow{\phi_p} (1,0), (\overline{\mu}-1) \xrightarrow{\sigma} (\overline{\mu}-1, \mu-1) \xrightarrow{\phi_p} (0,-1),$$

$$(\mu-1) \xrightarrow{\sigma} (\mu-1, \overline{\mu}-1) \xrightarrow{\phi_p} (-1,0), (-\mu-1) \xrightarrow{\sigma} (\mu-1, \overline{\mu}-1) \xrightarrow{\phi_p} (-1,1).$$

The pre-image by ϕ_3 of a code spreads its corresponding representatives along $\mathbb{R}^2$ (Fig. 4.3b).

Example 4.5 (Natural Reductions) In general, from any starting lattice Λ we can find a "natural" reduction to $\mathbb{F}_p^n$ as follows. Given a basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for Λ, take ϕ_p to be the linear mapping defined by $\phi_p(\mathbf{x}_i) = \boldsymbol{e}_i \in \mathbb{F}_p^n$, where $\mathbf{e}_i$ is the i-th canonical vector $(0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{F}_p^n$. It is clear that ϕ is surjective and $\ker \phi_p = p\Lambda$, therefore the associated sequence of reductions is non-degenerate. This provides a systematic way of constructing good sublattices of a given lattice.

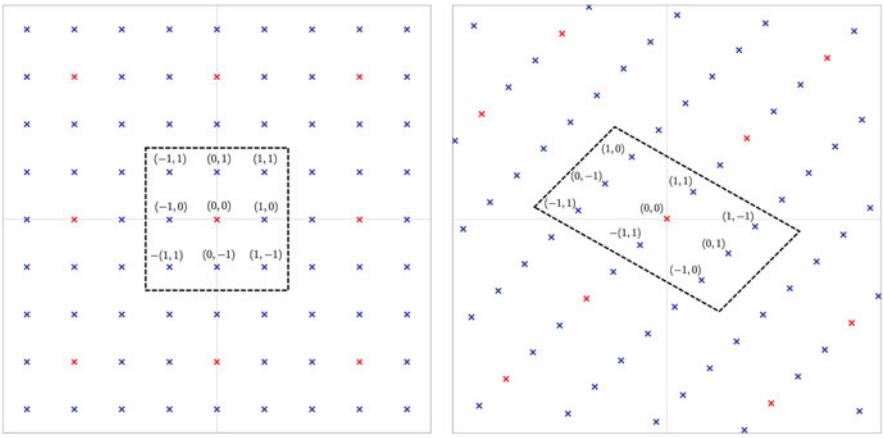


Fig. 4.3 Constructions from different base lattices and ϕ_3. Blue dots represent the fine lattice Λ and red dots represent the coarse lattice $\ker \phi_3$. (**a**) $\mathbb{Z}$-lattices (Example 4.3). (**b**) $\mathbb{Z}[\mu]$-lattices (Example 4.4)

The advantage of a reduction is that we can exploit properties of linear codes over $\mathbb{F}_p^n$ and, as it turns out, random codes over $\mathbb{F}_p$ are well behaved. This was recognized by Rush [20] and Loeliger [14], that showed from two different methods that the existence of certain codes over $\mathbb{F}_p$ implies the existence of good lattices.

Let $g : \mathbb{F}_p^n \to \mathbb{R}^+$ be a function and define $g^*(S)$ as the sum of g over all non-zero points of S. Let C be a code chosen at random from all codes of dimension k. It was proven in [14], from a simple counting argument, that the average of g assumes the simple form

$$\mathbb{E}[g^*(C)] = (p^k - 1)/(p^n - 1)g^*(\mathbb{F}_p^n). \tag{4.19}$$

Now let

$$\mathbb{L}_p = \left\{\beta \Lambda_p(C) : C \text{ is an } (n, k, p) - \text{code}\right\} \tag{4.20}$$

be the ensemble of all lattices associated to codes of dimension k, normalized to volume V (i.e., $\beta = V^{1/m}/(p^{n-k}V(\Lambda)^{1/m})$. The uniform distribution on the set of codes induces a uniform distribution on $\mathbb{L}_p$. In order to push Eq. (4.20) from codes to lattices, we need to impose some restrictions on the reductions.

Definition 4.3 A sequence of reductions $(\phi_{p_j})_{j=1}^{\infty}$, $\phi_j : \Lambda \to \mathbb{F}_{p_j}^n$, with increasing primes $p_1 < p_2 < \cdots$ is said to be *non-degenerate* if

$$\lambda(\Lambda_{p_j}) \geq c p_j^{\frac{n}{m}},$$

for all j some constants $c, \alpha > 0$. Similarly, the sequence of associated ensembles (Eq. (4.20)) is said to be non-degenerate.

Non degeneracy is indeed very mild and can be satisfied by the "natural" reduction in the previous examples. For the following theorem, let $\mathcal{P}$ be an infinite subset of the prime numbers.

Theorem 4.3 *If $(\phi_p)_{p\in\mathcal{P}}$ is a non-degenerate sequence of reductions and $\mathbb{L}_p$ are the corresponding ensembles, then*

$$\lim_{p\to\infty} \mathbb{E}_{\mathbb{L}_p}[\mathcal{N}_{B_r}(\Lambda)] = V^{-1} vol\ B_r.$$

Proof

$$\mathbb{E}\left[\mathcal{N}_{B_r}(\beta\Lambda)\right] = \mathbb{E}\left[\mathcal{N}_{B_r}(\beta\Lambda_p)\right] + \mathbb{E}\left[\mathcal{N}_{B_r}(\beta\Lambda_p(C)\backslash\beta\Lambda_p)\right].$$

Under the hypothesis, the first term tends to zero as $p \to \infty$. The second term, using Eq. (4.19) for $g(\mathbf{x}) = \mathbb{1}_{B_r}(\phi_p^{-1}(\{\mathbf{x}\}))$, is equal to

$$\mathbb{E}\left[\mathcal{N}_{B_r}(\Lambda_p(C)\backslash\Lambda_p)\right] = (p^k - 1)/(p^n - 1)\mathcal{N}_{B_r}(\beta\Lambda) \to \frac{\text{vol } B_r}{V},$$

where the last limit is obtained by applying a linear transformation to Eq. (4.5). □

From the above theorem and standard arguments, viz. Section 4.2, one can establish the existence of lattices constructed from any reduction ϕ_p approaching the lower bound.

Matrix Rings For the case of Hurwitz lattices (and more general orders over Division Algebras), as seen in Sect. 4.3, the "natural" underlying alphabet in the reduction is the ring $\mathcal{M}_n(\mathbb{F}_p)$ of $n \times n$ matrices with entries in $\mathbb{F}_p$. In these cases, a version of Theorem 4.3 over matrix rings is preferred. Such a theorem can be derived from the following Lemma 4.2 on random codes over matrix rings. Let $\mathcal{R}$ be a finite ring and $\mathcal{R}^*$ its units. Denote by $(\mathcal{R}^n)^*$ the set of vectors in $\mathcal{R}^n$ such that at least one coordinate is a unit. A linear code in $C \subset \mathbb{R}^n$ is a *free* $\mathcal{R}$-submodule of $\mathcal{R}^n$ (with the natural scalar multiplication). Let $g : \mathcal{R}^n \to \mathbb{R}^+$ be non-negative function. For a code C, we define $g^*(C) = \sum_{\mathbf{c}\in C\cap(\mathcal{R}^n)^*} g(c)$.

Lemma 4.2 ([4]) *If C_b is the set of all codes of rank k, and a code is chosen in C uniformly at random, then*

$$\mathbb{E}\left[g^*(C)\right] \leq \frac{|\mathcal{R}|^k}{\left|(\mathcal{R}^n)^*\right|} g^*(\mathcal{R}^n).$$

From this construction, and from a Rogers-like argument [25], it is possible to show the existence of Hurwitz lattices with real dimension $4m$ and density $\Delta_{4m} \geq 24m/e2^{4m}$, improving the results in Sect. 4.2.3.

4.4 A Glance at Applications to Wireless Communications

Up to now, we have only considered the problem of finding dense sphere-packings. As briefly described in the introduction, this problem is related to the one of finding good codes for the transmission of information over wireless media. This relation is well-established and can be interpreted via the Minkowski-Hlawka theorem.

In this second part of the chapter, we will show how the techniques considered previously, and the lattices constructed in Sects. 4.2 and 4.3 can be used in applications to wireless communications. In particular, we will show how algebraic lattices can be used to achieve the capacity of several communication channels.

Readers are referred to [24] for background of wireless communications. Practical design and applications of the proposed codes require more research in the future. See [6] for a code design based on a combination of algebraic lattices and polar codes.

4.4.1 Infinite Constellations

4.4.1.1 Classic AWGN Channel

The Gaussian channel problem can be described as follows. A signal, represented by a vector $\mathbf{x} \in \mathbb{C}^T$ is to be sent to a receiver through a noisy channel. One of the most fundamental way of modeling the noise is by supposing that it is additive and each entry is independent distributed according to a Gaussian distribution. The observed vector by a receiver after T slots of transmission (or "channel uses") is given by:

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

where $\mathbf{w}$ is a noise component, whose entries are iid, circularly symmetric Gaussian with variance σ_w^2 per complex dimensions. The objective of the receiver is to recover $\mathbf{x}$ with high probability, given the observation $\mathbf{y}$.

Infinite Constellation Problem First assume that the possible transmitted signals can be any point in a lattice[2] $\Lambda \subset \mathbb{C}^T$. One possible strategy is to find the closest lattice point to $\mathbf{y}$ and declare it as our estimate. This decoding strategy is usually

[2]Notice that in this part we consider *complex* lattices, since $\mathbb{C}^T$ is the typical ambient space in applications to wireless. The results in the previous sections can be "adapted" to complex lattices

known as *lattice decoding*. The probability of error of lattice decoding is given by

$$P_e(\Lambda) = P(\mathbf{w} \notin \mathcal{V}_\Lambda), \tag{4.21}$$

where $\mathcal{V}_\Lambda$ is the Voronoi cell of Λ (we recall Fig. 4.5a in the introduction for an illustration). If the volume of $\mathcal{V}_\Lambda$ is sufficiently large and its points are sufficiently separated apart (in comparison to the noise variance σ_w^2) we can clearly distinguish between each signal. However this strategy wastes too much volume. A more significant problem is the following: Given a target probability of error P_e what is the lattice that achieves P_e with the minimum possible volume? Conversely, one can ask, for a fixed a volume, what is the lattice that minimizes $P_e(\Lambda)$?

Another possible decoding strategy is to only decode points which are uniquely contained in a sphere of radius r (a threshold radius to be determined later). This strategy unveils the relation between the probability of error of a "random" lattice and the Minkowski-Hlawka theorem 4.1.

Proposition 4.2 (Lemma 7.7.1, [28]) *Let S be a sphere of radius r. The probability of error of a lattice is upper bounded by*

$$P_e(\Lambda) \leq P(\mathbf{w} \notin S) + \mathbb{E}_{\mathbf{w}}\left[\mathcal{N}_{S+\mathbf{w}}(\Lambda)\right],$$

where $\mathcal{N}_{S+\mathbf{w}}(\Lambda)$ is the lattice point enumerator (4.9).

Therefore, a bound on the probability of error can by obtained by bounding the lattice point enumerator. But the average behavior of $\mathcal{N}_{S+\mathbf{w}}(\Lambda)$ over an ensemble of lattices $\mathbb{L}$ is well-known from Theorems 4.1 and 4.3. The radius r can be choosen to be slightly greater than $\sqrt{T\sigma^2}$, which guarantees that the probability that $\mathbf{w} \notin S$ vanishes as the dimension increases.

We are particularly interested in high-dimensional signals, i.e., when $T \to \infty$ is large. In this case, from the above proposition we can deduce a result firstly proved by Poltyrev [18] that establishes that a vanishing probability of error is possible for a sequence of lattices $\Lambda_1, \Lambda_2, \ldots$ with

$$\lim_{T\to\infty} \sup \log V(\Lambda_T)^{1/T} \geq \log(\pi e \sigma^2).$$

Notice the slight difference between the above equation and (4.3), due the use of complex dimensions.

We will say that $\Lambda_1, \Lambda_2, \ldots,$ is *AWGN-good* if its probability of error vanishes with $\log V(\Lambda)^{1/T} \to \log(\pi e \sigma^2)$. The quantity

$$\gamma_\Lambda(\sigma) = V(\Lambda)^{1/T}/\sigma^2 \tag{4.22}$$

in a natural way. For instance, a complex full-rank lattice in $\mathbb{C}^T$ can be naturally identified with a lattice in $\mathbb{R}^n$, for $n = 2T$.

is usually called the *volume-to-noise ratio* (VNR) of a lattice, with respect to noise σ. Poltyrev's result can then be re-written as $\log \gamma_\Lambda(\sigma) \to \log \pi e$.

4.4.1.2 Compound Channel Model

Another, more general communication scenario arises when transmitting information using multiple antennas, through an unknown channel. After one channel use, the vector observed by a receiver can be modeled as follows

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{4.23}$$

where $\mathbf{H}$ is a fixed matrix with dimensions $n \times m$, and $\mathbf{w}$ is again a circularly symmetric Gaussian noise. In typical applications, the channel matrix $\mathbf{H}$ is known to the receiver but not known to the transmitter. After T channel uses, the channel equation can be written as:

$$\underbrace{\mathbf{Y}}_{n\times T} = \underbrace{\mathbf{H}}_{n\times m}\,\underbrace{\mathbf{X}}_{m\times T} + \underbrace{\mathbf{W}}_{n\times T} \tag{4.24}$$

or in vectorized form

$$\underbrace{\mathbf{y}}_{n\times 1} = \underbrace{\mathcal{H}}_{n\times mT}\,\underbrace{\mathbf{x}}_{mT\times 1} + \underbrace{\mathbf{w}}_{nT\times 1}, \tag{4.25}$$

where

$$\mathcal{H} = \mathbf{I}_T \otimes \mathbf{H} = \begin{pmatrix} \mathbf{H} & & & \\ & \mathbf{H} & & \\ & & \ddots & \\ & & & \mathbf{H} \end{pmatrix}.$$

is a block-diagonal matrix.

Infinite Compound Channel Model Again let us suppose that the transmitted signal $\mathbf{x}$ is in a lattice $\Lambda \subset \mathbb{C}^{nT}$. The probability of error is denoted by $P_e(\Lambda, \mathbf{H})$, and corresponds to the probability that $\mathbf{w}$ leaves the Voronoi cell of the transformed lattice $\mathcal{H}\Lambda$.

If $\mathbf{H}$ were known by both the transmitter and receiver, it would be possible to design the lattice $\mathbf{H}^{-1}\Lambda$, that completely ignores the effect of Λ. However multiplication by $\mathbf{H}^{-1}$ changes the volume of Λ to

$$\text{vol}\, \mathbf{H}^{-1}\Lambda = \mathcal{D} \times \text{vol}\, \Lambda,$$

where $\mathcal{D} = \sqrt{\det \mathbf{H}^{\dagger}\mathbf{H}}$ and $\mathbf{H}^{\dagger}$ is the Hermitian transpose of $\mathbf{H}$. Therefore it follows that with this strategy, according to the Poltyrev limit, the smallest possible volume for a lattice to have vanishing probability of error can be calculated as:

$$\begin{aligned} \log V(\mathbf{H}^{-1}\Lambda)^{1/nT} &= \log \det \mathcal{D}^{1/nT} + \log V(\Lambda)^{1/nT} \\ &\geq \log(\pi e \sigma^2) - \frac{1}{nT} \log \mathcal{D}. \end{aligned} \tag{4.26}$$

However, the assumption that $\mathbf{H}$ is known by a transmitter is very strong. It is perhaps surprising that even *without* this assumption, it is possible to design a sequence of lattices achieve the bound (4.26). In what follows, we will explain a construction with vanishing probability of error *for any matrix* H with fixed determinant $\mathcal{D} = \sqrt{\det \mathbf{H}^{\dagger}\mathbf{H}}$.

4.4.1.3 Block Fading Channel

In the block-fading channel, the channel matrix $\mathbf{H}$ in Eq. (4.24) is diagonal, with dimension $n \times n$. Note that here "block fading" means n parallel channels, all of which are fixed during a time interval of length T. Our objective is to design a lattice $\Lambda \subset \mathbb{C}^{nT}$ so that the probability of error $P_e(\Lambda, \mathbf{H})$ vanishes *simultaneously* for all $\mathbf{H}$ of the same determinant. For the following definition we recall that, from the definition of the volume-to-noise ratio (VNR) we have

$$\gamma_{\mathcal{H}\Lambda}(\sigma_w) = \frac{|\det \mathbf{H}|^{1/n} V(\Lambda)^{1/nT}}{\sigma_w^2}.$$

Definition 4.4 (Fading-Good Lattices [5]) We say that a sequence of lattices Λ of increasing dimension nT is universally good for the block-fading channel if $P_e(\Lambda, \mathbf{H}) \to 0$ as $T \to \infty$ for any VNR

$$\gamma_{(\mathbf{I}_T \otimes \mathbf{H})\Lambda}(\sigma_w) > \pi e$$

and all $\mathbf{H}$ with $|\det \mathbf{H}| = D$

In order to build lattices for the block fading channel, we resort to the generalized Construction A over O_K, as in Sect. 4.3.2 (see also the remark at the end of the section). In this case, we choose a totally complex (CM) number field $K/\mathbb{Q}(i)$ (or any other quadratic base field), of degree equal to n matching the number of rows/columns of matrix $\mathbf{H}$.

Let $\mathfrak{p} \subset O_K$ be a prime ideal above p with norm p^{ℓ}. Then $O_K/\mathfrak{p} \simeq \mathbb{F}_{p^{\ell}}$. The O_K-lattice Λ associated to a linear code $C \subset \mathbb{F}_{p^{\ell}}^{T}$ can be described using a convenient notation as:

$$\Lambda^{K}(C) = C + \mathfrak{p}^{T}. \tag{4.27}$$

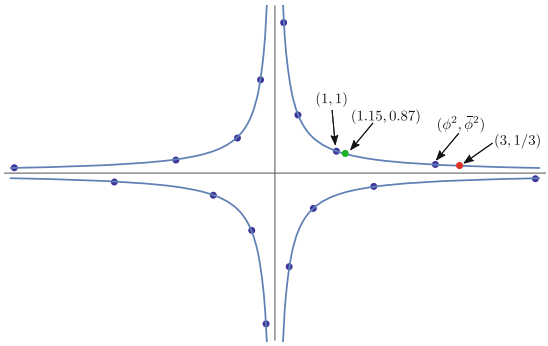


Fig. 4.4 Handling an ill-conditioned channel realization by the quantization of the channel space

The associated real lattice $\Lambda(C)$ will be then obtained by applying the embeddings of K in $\Lambda^K(C)$. Notice that this construction suits the generalised reductions as in Sect. 4.3.4, and therefore can generate lattices which satisfy the Minkowski-Hlawka theorem. One advantage of the following construction is that it can "compactify" the set of possible matrices with fixed discriminant $\mathbf{H}$ due the group of units of K and a theorem of Dirichlet. In other, words, for any matrix $\mathbf{H}$ with discriminant D it is possible to find $\mathbf{E}$ and $\mathbf{U}$ such that $\mathbf{EU} = \mathbf{H}$ and $\Lambda(C)$ is invariant under multiplication by $\mathbf{U}$, i.e., $\mathbf{U}\Lambda(C)$. An illustration of this process is in Fig. 4.4.

Using this property, and the Minkowski-Hlawka theorem, the existence of a universal lattice for the block-fading channel can be proven by averaging the aforementioned construction over random codes C (with $p \to \infty$) [5], as $T \to \infty$.

4.4.2 Power-Constrained General Model

4.4.2.1 Shaping

For practical applications, it is not possible to suppose that all lattice points are available for transmission. Due to physical limitations of the transmission devices, the power of the signal is usually constrained, and one can only send signals that satisfy

$$\frac{1}{T}\mathbb{E}\left[|x_1|^2 + |x_2|^2 + \ldots + |x_T|^2\right] \leq P,$$

where $P > 0$ is a given power parameter. One way of satisfying the power constraint, is to choose lattice points inside a sphere of radius $\sqrt{TP}$. We will explain next another technique, called *probabilistic shaping*, where the entire lattice is used, but the points are not picked uniformly, but chosen according to a *discrete Gaussian distribution*. In the power constrained case the normalized entropy of the signals

$$\frac{1}{n}\mathbb{H}(\mathbf{x}) = -\frac{1}{n}\sum_{\mathbf{x}\in\Lambda} P(\mathbf{x}) \log P(\mathbf{x}),$$

measures the communication rate of a scheme, where $P(\mathbf{x})$ denotes the probability (mass) of the sent point $\mathbf{x}$.

4.4.3 Lattice Gaussian Distribution

In order to deal with power constraints, we have to shape the infinite lattice constellation Λ. In this subsection, we define the lattice Gaussian distribution for $\mathbb{Z}[i]$-lattices. The definitions here are formally the same as for its real counterpart explained in [13], and the difference is a factor 2 in most cases.

Recall that an n-dimensional $\mathbb{Z}[i]$-lattice Λ in the Euclidean space $\mathbb{C}^n$ is defined as

$$\Lambda = \mathcal{L}(\mathbf{B}) = \left\{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}[i]^n\right\}$$

where $\mathbf{B} \in \mathbb{C}^{n\times n}$ is the generator matrix. The dual lattice Λ^* of a lattice Λ is defined as the set of vectors $\mathbf{v} \in \mathbb{C}^n$ such that $\langle \mathbf{v}, \boldsymbol{\lambda}\rangle = \mathbf{v}^\dagger\boldsymbol{\lambda} \in \mathbb{Z}[i]$, for all $\boldsymbol{\lambda} \in \Lambda$. The volume of Λ is defined as that of its real equivalent: $V(\Lambda) = |\det\mathbf{B}|^2$.

For $\sigma > 0$ and $\mathbf{c} \in \mathbb{C}^n$, the continuous Gaussian distribution of covariance matrix Σ centered at $\mathbf{c}$ is given by

$$f_{\sqrt{\Sigma},\mathbf{c}}(\mathbf{x}) = \frac{1}{\pi^n \det(\Sigma)} e^{-(\mathbf{x}-\mathbf{c})^\dagger \Sigma^{-1}(\mathbf{x}-\mathbf{c})},$$

for $\mathbf{x} \in \mathbb{C}^n$. For convenience, we write $f_{\sqrt{\Sigma}}(\mathbf{x}) = f_{\sqrt{\Sigma},\mathbf{0}}(\mathbf{x})$.

Consider the Λ-periodic function (see Fig. 4.5a)

$$f_{\sqrt{\Sigma},\Lambda}(\mathbf{x}) = \sum_{\boldsymbol{\lambda}\in\Lambda} f_{\sqrt{\Sigma},\boldsymbol{\lambda}}(\mathbf{x}) = \frac{1}{\pi^n \det(\Sigma)} \sum_{\boldsymbol{\lambda}\in\Lambda} e^{-(\mathbf{x}-\mathbf{c})^\dagger \Sigma^{-1}(\mathbf{x}-\mathbf{c})}, \tag{4.28}$$

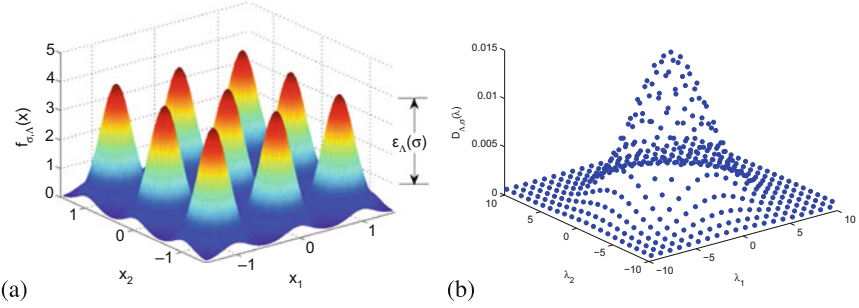


Fig. 4.5 Lattice Gaussian distributions. (**a**) Continuous periodic distribution $f_{\sigma,\Lambda}(\mathbf{x})$. (**b**) Discrete Gaussian distribution $D_{\Lambda,\sigma}(\boldsymbol{\lambda})$

for all $\mathbf{x} \in \mathbb{C}^n$. Observe that $f_{\sigma,\Lambda}$ restricted to a fundamental region $\mathcal{R}(\Lambda)$ is a probability density. We define the *discrete Gaussian distribution* over Λ centered at $\mathbf{c} \in \mathbb{C}^n$ as the following discrete distribution taking values in $\boldsymbol{\lambda} \in \Lambda$:

$$D_{\Lambda,\sqrt{\Sigma},\mathbf{c}}(\boldsymbol{\lambda}) = \frac{f_{\sqrt{\Sigma},\mathbf{c}}(\boldsymbol{\lambda})}{f_{\sqrt{\Sigma},\mathbf{c}}(\Lambda)}, \quad \forall \boldsymbol{\lambda} \in \Lambda,$$

where $f_{\sqrt{\Sigma},\mathbf{c}}(\Lambda) \triangleq \sum_{\boldsymbol{\lambda}\in\Lambda} f_{\sqrt{\Sigma},\mathbf{c}}(\boldsymbol{\lambda}) = f_{\sqrt{\Sigma},\Lambda}(\mathbf{c})$. Again for convenience, we write $D_{\Lambda,\sqrt{\Sigma}} = D_{\Lambda,\sqrt{\Sigma},\mathbf{0}}$. Figure 4.5b illustrates the discrete Gaussian distribution. As can be seen, it resembles a continuous Gaussian distribution, but is only defined over a lattice.

The flatness factor of a lattice Λ quantifies the maximum variation of $f_{\sqrt{\Sigma},\Lambda}(\mathbf{x})$ for $\mathbf{x} \in \mathbb{C}^n$.

Definition 4.5 (Flatness Factor) For a lattice Λ and for covariance matrix $\sqrt{\Sigma}$, the flatness factor is defined by:

$$\epsilon_\Lambda(\sqrt{\Sigma}) \triangleq \max_{\mathbf{x}\in\mathcal{R}(\Lambda)} \left| V(\Lambda) f_{\sqrt{\Sigma},\Lambda}(\mathbf{x}) - 1 \right|.$$

In words, $\frac{f_{\sqrt{\Sigma},\Lambda}(\mathbf{x})}{1/V(\Lambda)}$, the ratio between $f_{\sqrt{\Sigma},\Lambda}(\mathbf{x})$ and the uniform distribution over $\mathcal{R}(\Lambda)$, is within the range $[1 - \epsilon_\Lambda(\sqrt{\Sigma}), 1 + \epsilon_\Lambda(\sqrt{\Sigma})]$.

Proposition 4.3 (Expression of $\epsilon_\Lambda(\sqrt{\Sigma})$) *We have:*

$$\begin{aligned}\epsilon_\Lambda(\sqrt{\Sigma}) &= \frac{V(\Lambda)}{\pi^n \det(\Sigma)} \sum_{\boldsymbol{\lambda}\in\Lambda} e^{-\boldsymbol{\lambda}^\dagger \Sigma^{-1} \boldsymbol{\lambda}} \\ &= \sum_{\boldsymbol{\lambda}^*\in\Lambda^*} e^{-\pi^2 \boldsymbol{\lambda}^\dagger \Sigma^{-1} \boldsymbol{\lambda}} - 1\end{aligned}$$

In particular, if $\Sigma = \sigma^2 \mathbf{I}$*, then*

$$\begin{aligned}\epsilon_\Lambda(\sigma) &= \left(\frac{\gamma_\Lambda(\sigma)}{\pi}\right)^n \Theta_\Lambda\left(\frac{1}{\pi\sigma^2}\right) - 1 \\ &= \Theta_{\Lambda^*}\left(\pi\sigma^2\right) - 1\end{aligned}$$

where $\gamma_\Lambda(\sigma) = \frac{V(\Lambda)^{1/n}}{\sigma^2}$ *is the volume-to-noise ratio (VNR), and* $\Theta_\Lambda(\tau) = \sum_{\boldsymbol{\lambda}\in\Lambda} e^{-\pi\tau\|\boldsymbol{\lambda}\|^2}$ *is the theta series.*

A consequence of the Minkowski-Hlawka theorem of Sect. 4.2 applied to the theta series (see also the remark before Sect. 4.2.2) is the existence of sequences of lattices with vanishing flatness factor.

Theorem 4.4 (Minkowski-Hlawka) *$\forall\sigma > 0$ and $\forall\delta > 0$, there exists a sequence of lattices $\Lambda^{(n)}$ such that*

$$\epsilon_{\Lambda^{(n)}}(\sigma) \leq (1+\delta)\cdot\left(\frac{\gamma_{\Lambda^{(n)}}(\sigma)}{\pi}\right)^n, \tag{4.29}$$

i.e., the flatness factor can go to zero exponentially for any fixed VNR $\gamma_{\Lambda^{(n)}}(\sigma) < \pi$. More generally, $\epsilon_\Lambda(\sqrt{\Sigma}) \to 0$ if the generalized VNR $\gamma_{\Lambda^{(n)}}(\sqrt{\Sigma}) = \frac{V(\Lambda)^{1/n}}{\det(\Sigma)^{1/n}} < \pi$.

The significance of a small flatness factor is twofold. Firstly, it ensures that the "folded" distribution $f_{\sqrt{\Sigma},\Lambda}(\mathbf{x})$ is flat; secondly, it implies the discrete Gaussian distribution $D_{\Lambda,\sqrt{\Sigma},\mathbf{c}}$ is "smooth". We refer the reader to [12, 13] for more details.

The following lemma is particularly useful for communications and security [16].

Lemma 4.3 *Given $\mathbf{x}_1$ sampled from discrete Gaussian distribution $D_{\Lambda+\mathbf{c},\sqrt{\Sigma_1}}$ and $\mathbf{x}_2$ sampled from continuous Gaussian distribution $f_{\sqrt{\Sigma_2}}$. Let $\Sigma_0 = \Sigma_1 + \Sigma_2$ and let $\Sigma_3^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}$. If $\epsilon_\Lambda(\sqrt{\Sigma_3}) \leq \varepsilon \leq \frac{1}{2}$, then the distribution g of $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is close to $f_{\sqrt{\Sigma_0}}$:*

$$g(\mathbf{x}) \in f_{\sqrt{\Sigma_0}}(\mathbf{x})\,[1-4\varepsilon, 1+4\varepsilon]\,.$$

This lemma has profound implications. On one hand, it implies capacity, i.e., the discrete Gaussian distribution over a lattice is almost capacity-achieving if the flatness factor is small [12]. On the other hand, it implies security, i.e., Alice's signal received by Eve is indistinguishable from a continuous Gaussian distribution.

4.5 Achieving Channel Capacity

4.5.1 AWGN Channel

Consider the classic AWGN channel

$$\mathbf{y} = \mathbf{x} + \mathbf{w}$$

where the vectors have dimension T, the codeword length.

In [12], a new coding scheme based on the lattice Gaussian distribution was proposed. Let Λ be an AWGN-good lattice in $\mathbb{C}^T$ of dimension T, whose error probability vanishes if the VNR $\frac{V(\Lambda)^{1/T}}{\sigma_w^2} > \pi e$. The encoder maps the information bits to points in Λ, which obey the lattice Gaussian distribution (cf. Fig. 4.5b)

$$\mathbf{x} \sim D_{\Lambda,\sigma_s}.$$

Since the continuous Gaussian distribution is capacity-achieving, we want the lattice Gaussian distribution to behave like the continuous Gaussian distribution (in particular $P \approx \sigma_s^2$). This can be assured by a small flatness factor. Thus, while we are concerned with the discrete distribution D_{Λ,σ_s}, we in fact require the associated periodic distribution $f_{\sigma_s,\Lambda}$ to be flat.

Since the lattice points are not equally probable a priori in the lattice Gaussian coding, we will use maximum-a-posteriori (MAP) decoding. In [13], it was shown that MAP decoding is equivalent to Euclidean lattice decoding of Λ using a scaling coefficient $\alpha = \frac{\sigma_s^2}{\sigma_s^2+\sigma_w^2}$, which is asymptotically equal to the MMSE coefficient $\frac{P}{P+\sigma_w^2}$. In fact, the error probability of the proposed scheme under MMSE lattice decoding admits almost the same expression as that of Poltyrev [18], with σ_w replaced by $\tilde{\sigma}_w = \frac{\sigma_s\sigma_w}{\sqrt{\sigma_s^2+\sigma_w^2}}$. To satisfy the sphere bound, we choose the fundamental volume $V(\Lambda)$ such that

$$V(\Lambda)^{1/T} > \pi e \tilde{\sigma}_w^2. \tag{4.30}$$

Meanwhile, the rate of the scheme is given by the entropy of the lattice Gaussian distribution:

$$\begin{aligned}
\frac{1}{n}\mathbb{H}(\mathbf{x}) = R &\to \log(\pi e \sigma_s^2) - \frac{1}{T}\log V(\Lambda) \\
&< \log(\pi e \sigma_s^2) - \log\left(\pi e \frac{\sigma_s^2\sigma_w^2}{\sigma_s^2+\sigma_w^2}\right) \\
&= \log\left(1+\frac{\sigma_s^2}{\sigma_w^2}\right) \\
&\to \log(1+\mathrm{SNR}).
\end{aligned}$$

Combining these results, we arrive at the following theorem.

Theorem 4.5 (Coding Theorem) *Consider a lattice code whose codewords are drawn from the discrete Gaussian distribution D_{Λ,σ_s} for an AWGN-good lattice Λ. Any rate up to the channel capacity* $\log(1+\mathrm{SNR})$ *is achievable, while the error probability of MMSE lattice decoding vanishes exponentially fast.*

4.5.2 Compound Block Fading Channel

In the general form, our framework is able to tackle the compound MIMO channel; specializing this model we will obtain the block fading channel and the AWGN channel. More precisely, we consider an $n \times n$ MIMO channel described by the equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{4.31}$$

where $\mathbf{H} \in \mathbb{C}^{n\times n}$ is the channel matrix, and $\mathbf{x} \in \mathbb{C}^n$ is the input subject to the power constraint $E[\mathbf{x}^\dagger \mathbf{x}] \leq nP$. The noise entries of $\mathbf{w}$ are circularly symmetric complex Gaussian with zero-mean and variance σ_w^2. The signal-to-noise ratio (SNR) per receive antenna is defined by $\text{SNR} = nP/\sigma_w^2$. Assume that the receiver has complete knowledge of $\mathbf{H}$ (but the transmitter does not have CSIT), which is fixed during a whole transmission block. The (white input) achievable rate of this channel is

$$C = \log\det\left(\mathbf{I} + \text{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right). \tag{4.32}$$

Consider the set $\mathbb{H}$ of all channel matrices with fixed white-input capacity C:

$$\mathbb{H} = \left\{\mathbf{H} \in \mathbb{C}^{n\times n} : \log\det\left(\mathbf{I} + \text{SNR}\,\mathbf{H}^\dagger\mathbf{H}\right) = C\right\}. \tag{4.33}$$

This can be viewed as a compound channel with capacity C. The compound channel model (4.33) arises in several important scenarios in communications, such as the outage formulation in the open-loop mode and broadcast [17].

The compound channel demands a universal code that achieves the capacity for all members $\mathbf{H} \in \mathbb{H}$. This represents one of the most difficult problems in coding theory. Note that (4.33) reduces to a compound block fading channel if $\mathbf{H}$ is diagonal (here n denotes the number of blocks), and to the AWGN channel if $\mathbf{H} = \mathbf{I}$.

Applying the "unconstrained" construction described in Sect. 4.4.1.3 along with Gaussian shaping, it can be shown that the average error probability $\mathbb{E}_\Lambda[P_e(\Lambda)]$ vanishes as long as the VNR $> \pi e$ (as $T \to \infty$):

$$\frac{|\mathbf{I} + \text{SNR}\,\mathbf{H}^\dagger\mathbf{H}|^{\frac{1}{n}} V(\Lambda)^{\frac{1}{nT}}}{\sigma_s^2} > \pi e. \tag{4.34}$$

Thus, any rate

$$R \to n\log(\pi e\sigma_s^2) - \frac{1}{T}\log(V(\Lambda)) < \log|\mathbf{I} + \text{SNR}\,\mathbf{H}^T\mathbf{H}| = C$$

is achievable. Note that the achievable rate only depends on $\mathbf{H}$ through determinant $|\mathbf{I} + \text{SNR}\,\mathbf{H}^\dagger\mathbf{H}|$. Therefore, there exists a lattice Λ achieving capacity C of the compound channel.

4.5.3 MIMO Fading Channel

The case of MIMO channels is more technical due to non-commutativity of the underlying algebra. Let $\mathcal{O}$ be the natural order of cyclic division algebra $\mathcal{A}$. Take a two-sided ideal $\mathcal{J}$ of $\mathcal{O}$ and consider the quotient ring $\mathcal{O}/\mathcal{J}$. Define a reduction

$\beta : \mathcal{O} \to \mathcal{O}/\mathcal{J}$. For a linear code C over $\mathcal{O}/\mathcal{J}$, $\beta^{-1}(C)$ is a lattice Λ (in $\mathbb{C}^{n^2T}$). However, the quotient ring $\mathcal{O}/\mathcal{J}$ is non-commutative in general, e.g., a matrix ring, skew polynomial ring etc. Nevertheless, as we have seen in Sect. 4.3.4 it is still possible to prove the Minkowski-Hlawka theorem using codes over rings. Thus, there exists a sequence of lattices universally good for MIMO fading, hence achieving the capacity of compound MIMO channels. Note that recently [17] and [15] have achieved a constant gap to the capacity of compound MIMO channels.

4.5.4 Approaching Secrecy Capacity

4.5.4.1 Gaussian Wiretap Channel

Now consider the Gaussian wiretap channel where Alice and Bob are the legitimate users, while Eve is an eavesdropper. The outputs $\mathbf{y}$ and $\mathbf{z}$ at Bob and Eve's ends are respectively given by

$$\begin{cases} \mathbf{y} = \mathbf{x} + \mathbf{w}_b, \\ \mathbf{z} = \mathbf{x} + \mathbf{w}_e, \end{cases} \tag{4.35}$$

where $\mathbf{w}_b$, $\mathbf{w}_e$ are T-dimensional Gaussian noise vectors with zero mean and variance σ_b^2, σ_e^2 respectively.

For secrecy rate R_s, we use coset coding induced by a lattice partition $\Lambda_e \subset \Lambda_b$ such that

$$\frac{1}{T} \log |\Lambda_b/\Lambda_e| = R_s.$$

The fine lattice Λ_b is the usual coding lattice for Bob, i.e., it is an AWGN-good lattice. The coarse lattice Λ_e is new, and turns out to be a secrecy-good lattice. To encode, Alice uses the secret bits to select one coset of Λ_e and transmits a random point inside this coset.

Let us discuss intuitively why this scheme is secure. Informally, given message m, Alice samples a lattice point uniformly at random from a coset $\Lambda_e + \boldsymbol{\lambda}_m$ (this corresponds to Poltyrev's setting of infinite lattice coding [18]). Due to the channel noise, Eve observes the periodic distribution

$$\frac{1}{(\pi\sigma_e^2)^T} \sum_{\boldsymbol{\lambda} \in \Lambda + \boldsymbol{\lambda}_m} e^{-\frac{\|\mathbf{z}-\boldsymbol{\lambda}\|^2}{\sigma_e^2}}.$$

If the flatness factor $\epsilon_{\Lambda_e}(\sigma_e)$ is small, it will be close to a uniform distribution, regardless of message m. Then Eve would not be able to distinguish which message Alice has sent. With a careful design of Λ_e, this is possible, because Eve's channel

is noisier. Of course, the technical difficulty here is that one cannot really sample a lattice point uniformly from a lattice or its coset.

Now we describe the wiretap coding scheme more formally. Consider a message set $\mathcal{M} = \{1, \ldots, e^{TR}\}$, and a one-to-one function $\phi : \mathcal{M} \to \Lambda_b/\Lambda_e$ which associates each message $m \in \mathcal{M}$ to a coset $\widetilde{\boldsymbol{\lambda}}_m \in \Lambda_b/\Lambda_e$. One could choose the coset representative $\boldsymbol{\lambda}_m \in \Lambda_b \cap \mathcal{R}(\Lambda_e)$ for any fundamental region $\mathcal{R}(\Lambda_e)$. In order to encode the message $m \in \mathcal{M}$, Alice actually samples $\mathbf{x}_m$ from lattice Gaussian distribution

$$\mathbf{x}_m \sim D_{\Lambda_e + \boldsymbol{\lambda}_m, \sigma_s}.$$

equivalently, Alice transmits $\boldsymbol{\lambda} + \boldsymbol{\lambda}_m$ where $\boldsymbol{\lambda} \sim D_{\Lambda_e, \sigma_s, -\boldsymbol{\lambda}_m}$. Let $\tilde{\sigma}_e = \frac{\sigma_s \sigma_e}{\sqrt{\sigma_s^2 + \sigma_e^2}}$ and $\sigma_s' = \sqrt{\sigma_s^2 + \sigma_e^2}$. Regev's Lemma (cf. Lemma 4.3) implies that if $\epsilon_{\Lambda_e}(\tilde{\sigma}_e) < \frac{1}{2}$, then:

$$\mathbb{V}\left(p_{\mathsf{Z}|\mathsf{M}}(\cdot|m),\, f_{\sigma_s'}\right) \leq 4\epsilon_{\Lambda_e}(\tilde{\sigma}_e).$$

We see that the received signals converge to the same Gaussian distribution $f_{\sigma_s'}$. This already gives *distinguishing security*, which means that, asymptotically, the channel outputs are indistinguishable for different input messages.

An upper bound on the amount of leaked information then follows.

Theorem 4.6 (Information Leakage [13]) *Suppose that the wiretap coding scheme described above is employed on the Gaussian wiretap channel (4.35), and let $\varepsilon_T = \epsilon_{\Lambda_e}(\tilde{\sigma}_e)$. Assume that $\varepsilon_T < \frac{1}{2}$ for all T. Then the mutual information between the confidential message and the eavesdropper's signal is bounded as follows:*

$$\mathrm{i}(\mathsf{M}; \mathsf{Z}) \leq 8\varepsilon_T TR - 8\varepsilon_T \log 8\varepsilon_T. \tag{4.36}$$

A wiretap coding scheme is secure in the sense of *strong secrecy* if $\lim_{T\to\infty} \mathrm{i}(\mathsf{M}; \mathsf{Z}) = 0$. From (4.36), a flatness factor $\varepsilon_T = o(\frac{1}{T})$ would be enough. In practice, an exponential decay of the information leakage is desired, and this motivates the notion of secrecy-good lattices:

Definition 4.6 (Secrecy-Good Lattices) A sequence of lattices $\Lambda^{(T)}$ is *secrecy-good* if

$$\epsilon_{\Lambda^{(n)}}(\sigma) = e^{-\Omega(T)}, \quad \forall \gamma_{\Lambda^{(T)}}(\sigma) < \pi. \tag{4.37}$$

In the notion of strong secrecy, plaintext messages are often assumed to be random and uniformly distributed in $\mathcal{M}$. This assumption is deemed problematic from the cryptographic perspective, since in many setups plaintext messages are not random. This issue can be resolved by using the standard notion of *semantic security* [9] which means that, asymptotically, it is impossible to estimate any

function of the message better than to guess it without considering Z at all. The relation between strong secrecy and semantic security was recently revealed in [2, 13], namely, achieving strong secrecy for all distributions of the plaintext messages is equivalent to achieving semantic security. Since in our scheme we make no a priori assumption on the distribution of m, it achieves semantic security.

It can be shown that, under mild conditions (similar to those in [13]), the secrecy rate

$$R < \log(1 + \mathrm{SNR}_b) - \log(1 + \mathrm{SNR}_e) - 1 \tag{4.38}$$

is achievable, which is within 1 nat from the secrecy capacity. It is worth mentioning that this small gap may be fictitious, due to our proof technique.

4.5.4.2 Fading Wiretap Channel

The channels for Bob and for Eve are given by

$$\mathbf{y} = \mathbf{H}_b\mathbf{x} + \mathbf{w}_b, \quad \mathbf{z} = \mathbf{H}_e\mathbf{x} + \mathbf{w}_e,$$

respectively. We fix the capacity C_e of Eve's compound channel with white inputs

$$\mathbb{H}_e = \left\{\mathbf{H}_e \in \mathbb{C}^{n\times n} : \log\det\left(\mathbf{I} + \mathrm{SNR}\,\mathbf{H}_e^{\dagger}\mathbf{H}_e\right) = C_e\right\}. \tag{4.39}$$

as well as the capacity C_b of Bob's compound channel. The secrecy capacity of compound MIMO wiretap channels with white inputs is given by Schaefer and Loyka[21]:

$$C_s = C_b - C_e. \tag{4.40}$$

Similarly to lattice coding over the Gaussian wiretap channel, we use a pair of nested lattices $\Lambda_b \subset \Lambda_e$. These lattices are built in the same manner as above:

$$\Lambda_b = C_b + \mathfrak{p}^T \tag{4.41}$$

$$\Lambda_e = C_e + \mathfrak{p}^T \tag{4.42}$$

where the codes satisfy $C_e \subseteq C_b$.

In order to encode the message $m \in \mathcal{M}$, Alice samples $\mathbf{x}_m$ from distribution $D_{\Lambda_e+\boldsymbol{\lambda}_m,\sigma_s}$. Similarly to (4.25), let $\mathcal{H}_e = \mathbf{I}_T \otimes \mathbf{H}_e$ of size nT. Eve observes a discrete Gaussian distribution $D_{\mathcal{H}_e(\Lambda_e+\boldsymbol{\lambda}_m),\mathcal{H}_e\sigma_s}$, contaminated by i.i.d. Gaussian noise of standard deviation σ_e. We would like this to be indistinguishable from a continuous

Gaussian distribution of covariance matrix $\Sigma_0 = \sigma_s^2 \mathcal{H}_e \mathcal{H}_e^\dagger + \sigma_e^2 \mathbf{I}$, regardless of m. By Lemma 4.3, we need

$$\epsilon_{\sqrt{\Sigma_3}}(\mathcal{H}_e \Lambda_e) \to 0$$

where $\Sigma_3^{-1} = \sigma_s^{-2}(\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} + \sigma_e^{-2}\mathbf{I}$. In other words, we want the flatness factor $\epsilon_{\mathcal{H}_e \Lambda_e}(\sqrt{\Sigma_3}) = \varepsilon_T$ to vanish with T.

We derive the expression

$$\begin{aligned}
\epsilon_{\mathcal{H}_e \Lambda_e}(\sqrt{\Sigma_3}) &= \frac{V(\mathcal{H}_e \Lambda_e)}{\pi^{nT} \det(\Sigma_3)} \sum_{\boldsymbol{\lambda} \in \mathcal{H}_e \Lambda_e} e^{-\boldsymbol{\lambda}^T \Sigma_3^{-1} \boldsymbol{\lambda}} - 1 \\
&= \frac{V(\mathcal{H}_e \Lambda_e)}{\pi^{nT} \det(\Sigma_3)} \sum_{\boldsymbol{\lambda} \in \Lambda_e} e^{-\frac{\boldsymbol{\lambda}^T \mathcal{H}_e^\dagger \Sigma_3^{-1} \mathcal{H}_e \boldsymbol{\lambda}}{2}} - 1 \\
&= \frac{V(\Lambda_e)}{\pi^{nT}} \det(\sigma_s^{-2}\mathbf{I} + \sigma_e^{-2} \mathcal{H}_e^\dagger \mathcal{H}_e) \times \\
&\quad \sum_{\boldsymbol{\lambda} \in \Lambda_e} e^{-\frac{\boldsymbol{\lambda}^T \left(\sigma_s^{-2}\mathbf{I} + \sigma_e^{-2} \mathcal{H}_e^\dagger \mathcal{H}_e\right) \boldsymbol{\lambda}}{2}} - 1.
\end{aligned}$$

It is worth mentioning that this expression shares the same form of Eve's correct decoding probability given in [1, (13)] except the MMSE correction term $\sigma_s^{-2}\mathbf{I}$.

Applying Minkowski-Hlawka, we obtain

$$\begin{aligned}
&\mathbb{E}_{\Lambda_e}[\epsilon_{\mathcal{H}_e \Lambda_e}(\sqrt{\Sigma_3})] \\
&= \frac{V(\Lambda_e)}{\pi^{nT}} \det(\sigma_s^{-2}\mathbf{I} + \sigma_e^{-2} \mathcal{H}_e^\dagger \mathcal{H}_e) \\
&= \frac{V(\Lambda_e)}{(\pi \sigma_s^2)^{nT}} \det(\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e)^T.
\end{aligned}$$

Now we calculate the information leakage to Eve. If we slightly reduce the VNR of Λ_e, $\mathbb{E}_{\Lambda_e}[\epsilon_{\mathcal{H}_e \Lambda_e}(\sqrt{\Sigma_3})]$ in (4.43) will vanish exponentially with T. Similar to the Gaussian wiretap channel (4.36), the mutual information between Alice and Eve is bounded for any $\mathbf{H}_b$, $\mathbf{H}_e$ as

$$\mathsf{i}(\mathsf{M}; \mathsf{Z}) \leq 8\varepsilon_T T R_s - 8\varepsilon_T \log(8\varepsilon_T). \tag{4.43}$$

Again, it is tricky to exhibit the existence of a universal code for all $\mathbf{H}_b$, $\mathbf{H}_e$. Fortunately, thanks to the unit groups, this can be resolved by quantizing the channels in the same manner as for capacity [21].

For a vanishing flatness factor, we need the condition

$$\frac{\det(\mathbf{I}+\mathrm{SNR}_e\,\mathbf{H}_e^\dagger\mathbf{H}_e)^{1/n}\,V(\Lambda_e)^{\frac{1}{nT}}}{\sigma_s^2}<\pi. \tag{4.44}$$

From (4.34) and (4.44), we obtain the secrecy rate

$$R_s<\log\frac{\left|\mathbf{I}+\mathrm{SNR}_b\,\mathbf{H}_b^\dagger\mathbf{H}_b\right|}{\left|\mathbf{I}+\mathrm{SNR}_e\,\mathbf{H}_e^\dagger\mathbf{H}_e\right|}-n=C_b-C_e-n,$$

which is the secrecy capacity to within a constant gap of n nats. Again, this gap may well be fictitious.

Then one may claim the existence of a universal lattice code which achieves the secrecy capacity to within n nats, under semantic security. Extensions to the MIMO wiretap channel are also possible, using cyclic division algebras. The security proof is very much the same, except that $\mathbf{H}_b$ and $\mathbf{H}_e$ are full matrices.

References

1. Belfiore, J., Oggier, F.: An error probability approach to MIMO wiretap channels **61**(8), 3396–3403 (2013). https://doi.org/10.1109/TCOMM.2013.061913.120278
2. Bellare, M., Tessaro, S., Vardy, A.: Semantic security for the wiretap channel. In: Proceedings of CRYPTO 2012. Lecture Notes in Computer Science, vol. 7417, pp. 294–311. Springer, Berlin (2012)
3. Blichfeldt, H.: The minimum values of positive quadratic forms in six, seven and eight variables. Math. Zeitsch. **39**, 1–15 (1935). http://eudml.org/doc/168534
4. Campello, A.: Random ensembles of lattices from generalized reductions. IEEE Trans. Inf. Theory. **64**(7), 5231–5239 (2018)
5. Campello, A., Ling, C., Belfiore, J.C.: Algebraic lattice codes achieve the capacity of the compound block-fading channel. In: 2016 IEEE International Symposium on Information Theory (ISIT), pp. 910–914 (2016). https://doi.org/10.1109/ISIT.2016.7541431
6. Campello, A., Liu, L., Ling, C.: Multilevel code construction for compound fading channels. In: 2017 IEEE International Symposium on Information Theory (ISIT), pp. 1008–1012 (2017). https://doi.org/10.1109/ISIT.2017.8006680
7. Cassels, J.W.S.: An Introduction to the Geometry of Numbers. Springer, Berlin (1997)
8. Conway, J.H., Sloane, N.J.A.: Sphere-Packings, Lattices, and Groups. Springer, New York (1998)
9. Goldwasser, S., Micali, S.: Probabilistic encryption. J. Comput. Syst. Sci. **28**(2), 270–299 (1984)
10. Gruber, P.: Convex and Discrete Geometry. Springer, Berlin (2007)
11. Hlawka, E.: Zur geometrie der zahlen. Math. Zeitsch. **49**, 285–312 (1943). http://eudml.org/doc/169025
12. Ling, C., Belfiore, J.C.: Achieving AWGN channel capacity with lattice gaussian coding. IEEE Trans. Inf. Theory **60**(10), 5918–5929 (2014). https://doi.org/10.1109/TIT.2014.2332343

13. Ling, C., Luzzi, L., Belfiore, J.C., Stehle, D.: Semantically secure lattice codes for the Gaussian wiretap channel. IEEE Trans. Inf. Theory **60**(10), 6399–6416 (2014). https://doi.org/10.1109/TIT.2014.2343226
14. Loeliger, H.A.: Averaging bounds for lattices and linear codes. IEEE Trans. Inf. Theory **43**(6), 1767–1773 (1997). https://doi.org/10.1109/18.641543
15. Luzzi L., Vehkalahti, R.: Almost universal codes achieving ergodic MIMO capacity within a constant gap. IEEE Trans. Inf. Theory **63**(5), 3224–3241 (2017)
16. Luzzi, L., Vehkalahti, R., Ling, C.: Almost universal codes for MIMO wiretap channels. IEEE Trans. Inf. Theory **64**(11), 7218–7241 (2018)
17. Ordentlich, O., Erez, U.: Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap. IEEE Trans. Inf. Theory **61**(1), 323–340 (2015). https://doi.org/10.1109/TIT.2014.2370047
18. Poltyrev, G.: On coding without restrictions for the AWGN channel. IEEE Trans. Inf. Theory **40**, 409–417 (1994)
19. Rogers, C.A.: Packing and Covering. Cambridge University Press, Cambridge (1964)
20. Rush, J.A.: A lower bound on packing density. Invent. Math. **98**(3), 499–509 (1989). https://doi.org/10.1007/BF01393834
21. Schaefer, R.F., Loyka, S.: The secrecy capacity of compound Gaussian MIMO wiretap channels. IEEE Trans. Inf. Theory **61**(10), 5535–5552 (2015). https://doi.org/10.1109/TIT.2015.2458856
22. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
23. Siegel, C.L.: A mean value theorem in geometry of numbers. Ann. Math. **46**(2), 340–347 (1945). http://www.jstor.org/stable/1969027
24. Tse, D., Viswanath, P.: Fundamentals of Wireless Communication. Cambridge University Press, Cambridge (2005)
25. Vance, S.: Improved sphere packing lower bounds from Hurwitz lattices. Adv. Math. **227**(5), 2144–2156 (2011). https://doi.org/10.1016/j.aim.2011.04.016. http://www.sciencedirect.com/science/article/pii/S000187081100140X
26. Venkatesh, A.: A note on sphere packings in high dimension. Int. Math. Res. Not. (2012). https://doi.org/10.1093/imrn/rns096. http://imrn.oxfordjournals.org/content/early/2012/03/07/imrn.rns096.abstract
27. Liu, L., Yan, Y., Ling, C., Wu, X.: Construction of capacity-achieving lattice codes: Polar lattices. IEEE Trans. Commun. **67**(2), 915–928 (2019)
28. Zamir, R.: Lattice Coding for Signals and Networks. Cambridge University Press, Cambridge (2014)

Chapter 5
Algebraic Lattice Codes for Linear Fading Channels

Roope Vehkalahti and Laura Luzzi

Abstract There exists an old and well established connection between lattice code design for the additive white Gaussian noise (AWGN) channel and the mathematical theory of lattices. Code design principles can be translated into the language of geometry of numbers and are related to the most central problems in classical lattice theory. These connections appear both in the practical design of short lattice codes, and also in the asymptotic regime when designing codes that perform well from the capacity point of view. However, when considering modern wireless channels, one must take into account new features such as time or frequency selective fading and multiple antennas. Such channels can not be abstracted into a simple AWGN model, and require a different coding strategy. While in recent years plenty of research has been done on code design for fading channels, few works have focused on the problem of approaching capacity. In this survey, we review and generalize our recent works and show how it is possible to perform code design for a large class of different fading channels from a unified perspective and how this approach can be used to build very robust lattice codes that perform within a constant gap from the corresponding capacity. Our approach can be seen as a generalization to fading channels of the classical connection between sphere packing problems and design of capacity approaching lattice codes.

5.1 Introduction

In the decades following Shannon's seminal work, the quest to design codes for the additive white Gaussian noise (AWGN) channel led to the development of a rich theory, revealing a number of beautiful connections between information theory and

R. Vehkalahti (✉)
Department of Communications and Networking, Aalto University, Espoo, Finland
e-mail: roope.vehkalahti@aalto.fi

L. Luzzi
Laboratoire ETIS, CY Université, ENSEA, CNRS (UMR 8051), Cergy-Pontoise, France
e-mail: laura.luzzi@ensea.fr

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*,
Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_5

geometry of numbers. One of the most striking examples is the connection between classical lattice sphere packing and the capacity of the AWGN channel. The main result states that any family of lattice codes with linearly growing *Hermite invariant* achieves a constant gap to capacity. These classical results and many more can be found in the comprehensive book by Conway and Sloane [5].

The early sphere packing results suggested that lattice codes could achieve the capacity of the AWGN channel and led to a series of works trying to prove this, beginning with [6] and finally completed in [7]. Thus, while there are still plenty of interesting questions to consider, the theory of lattice codes for the single user AWGN channel is now well-established.

However, although the AWGN channel is a good model for deep-space or satellite links, modern wireless communications call for more general channel models which include time or frequency varying fading and possibly multiple transmit and receive antennas. Therefore, in the last 20 years, coding theorists have focused on the design of lattice codes for multiple and single antenna fading channels [3, 23].

Yet the question of whether lattice codes can achieve capacity in fading channels has only been addressed recently. The first work that we are aware of is due to S. Vituri [25, Section 4.5], and gives a proof of existence of lattice codes achieving a constant gap to capacity for i.i.d SISO channels. It seems that with minor modifications this proof is enough to guarantee the existence of capacity achieving lattices. In the single antenna i.i.d fading channels, this problem was considered also in [10] and in our paper [24].

In [13], it was shown that polar lattices achieve capacity in single antenna i.i.d fading channels. This is not only an existence result, it also gives an explicit code construction. In [4], the authors prove existence of lattice codes achieving capacity for the compound MIMO channel, where the fading is random during the first s time units, but then gets repeated in blocks of length s. This work is most closely related to [20], which was considering a similar question.

In [14, 15], we proved that lattice codes derived from number fields and division algebras do achieve a constant gap to capacity over single and multiple antenna fading channels. As far as we know, this was the first result achieving constant gap to MIMO capacity with lattice codes. In [11], the authors corrected and generalized [10] and improved on our gap in the case of Rayleigh fading MIMO channels.

However, while in our work [15], the gap to capacity is relatively large, our code construction is *almost universal* in the sense that a *single* code achieves a constant gap to capacity for *all* stationary ergodic fading channel models satisfying a certain condition for fading (5.9). With some limitations, this gap is also uniform for all such channels (see Remark 5.4).

In this work, we are revisiting some of the results in [15] and presenting them from a slightly different and more general perspective. Our approach is based on generalizing the classical sphere packing approach to fading channels. In [15], we introduced the concept of *reduced Hermite invariant* of a lattice with respect to a linear group of block fading matrices. As a generalization of the classical result for AWGN channels, we proved that if a family of lattices has linearly growing

reduced Hermite invariant, it achieves a constant gap to capacity in the block fading MIMO channel. In this work, we extend this result and show that given any linearly fading channel model we can define a corresponding notion of reduced Hermite invariant. We also prove that in some cases the reduced Hermite invariant of a lattice is actually a *homogeneous minimum* with respect to a *homogeneous form* (which depends on the fading model). From this perspective, the classical sphere packing result [5, Chapter 3] is just one example of the general connection between linear fading channels and the homogeneous minima of the corresponding forms.

In Sect. 5.2, we begin by defining a general linear fading model, which captures several channels of interest for practical applications. In Sect. 5.3, we recall how to obtain a finite signal constellation from an infinite lattice under an average power constraint. In Sect. 5.4, we review how the classical Hermite invariant can be used as a design criterion to build capacity approaching lattice codes in AWGN channels. In Sect. 5.5, we generalize the concept of Hermite invariant to linear fading channels by introducing the reduced Hermite invariant. We also show that replacing the Hermite invariant with the reduced one as a code design criterion leads to an analogous capacity result in linear fading channels.

In Sect. 5.6, we focus on channels where the fading matrices are diagonal. This brings us to consider ergodic fading single antenna channels. Following [15], we show how lattice constructions from algebraic number fields can be used to approach capacity in such channels. We begin by considering lattices arising from the canonical embedding of the ring of algebraic integers, then examine the question of improving the gap to capacity using non-principal ideals of number fields [24].[1] In particular, we show that our information-theoretic problem is actually equivalent to a certain classical problem in algebraic number theory.

Finally, in Sect. 5.7, we extend the results in [15] and show that in many relevant channel models the reduced Hermite invariant of a lattice is actually a homogeneous minimum of a certain form.

5.2 General Linear Fading Channel

In this work, we consider complex vector-valued channels, where the transmitted (and received) elements are vectors in $\mathbb{C}^k$. A *code* C is a finite set of elements in $\mathbb{C}^k$. We assume that both the receiver and the transmitter know the code.

Given a matrix $H \in M_k(\mathbb{C})$ and a row vector $x \in \mathbb{C}^k$, in order to hold on the tradition that a transmitted vector is a row, we introduce the notation

$$H[x] = (H(x^T))^T.$$

[1] More precisely, the ideal lattice construction was considered in the extended version of [24], available at http://arxiv.org/abs/1411.4591v2.

Let us assume we have an infinite sequence of random matrices H_k, $k = 1, 2 \ldots, \infty$, where for every k, H_k is a $k \times k$ matrix. Given such sequence of matrices we can define a corresponding channel. Given an input $x = (x_1, \ldots, x_k)$, we will write the channel output as

$$y = H_k[x] + w, \tag{5.1}$$

where w is a length k random vector, with i.i.d complex Gaussian entries with variance 1 and zero mean, and the random matrix H_k represents fading. We assume that the receiver always knows the channel realization of H_k and is trying to guess which was the transmitted codeword x based on y and H_k. This set-up defines a *linear fading channel* (with channel state information at the receiver), where the term "linear" simply refers to the fact that the fading can be represented as the action of a linear transform on the transmitted codeword. This type of channel (but without channel state information) has been considered before in [26].

In the following sections, we consider the problem of designing codes for this type of channels. In the remainder of the paper, we will assume the extra condition that the determinant of the random matrices H_k is non-zero with probability one. The channel model under consideration captures many communication channels of practical significance. For example, when H_k is the identity matrix, we have the classical additive Gaussian channel. Furthermore, if H_k is a diagonal matrix with i.i.d Gaussian random elements with zero mean, we obtain the Rayleigh fast fading channel. Finally, if H_k is a block diagonal matrix, we obtain a block fading MIMO channel.

5.3 Lattices and Finite Codes

As mentioned previously, our finite codes C are simply subsets of elements in $\mathbb{C}^k$. We consider the ambient space $\mathbb{C}^k$ as a metric space with the Euclidean norm.

Definition 5.1 Let $v = (v_1, \ldots, v_k)$ be a vector in $\mathbb{C}^k$. The *Euclidean norm* of v is $||v|| = \sqrt{\sum_{i=1}^{k} |v_i|^2}$.

Given a *transmission power* P, we require that every codeword $x \in C \subset \mathbb{C}^k$ satisfies the average power constraint

$$\frac{1}{k} \|x\|^2 \leq P. \tag{5.2}$$

The *rate* of the code is given by $R = \frac{\log_2 |C|}{k}$.

In this work we focus on finite codes C that are derived from *lattices*.

A full *lattice* $L \subset \mathbb{C}^k$ has the form $L = \mathbb{Z}b_1 \oplus \mathbb{Z}b_2 \oplus \cdots \oplus \mathbb{Z}b_{2k}$, where the vectors $b_1, \ldots, b_{2k}$ are linearly independent over $\mathbb{R}$, i.e., form a lattice basis.

Given an average power constraint P, the following lemma suggests that by shifting a lattice and considering its intersection with the $2k$-dimensional ball $B(\sqrt{kP})$ of radius $\sqrt{kP}$, we can have codes having roughly $\mathrm{Vol}(B(\sqrt{kP}))$ elements, where the notation Vol stands for the volume.

Lemma 5.1 (See [8]) *Suppose that L is a full lattice in $\mathbb{C}^k$ and S is a Jordan-measurable bounded subset of $\mathbb{C}^k$. Then there exists $x \in \mathbb{C}^k$ such that*

$$|(L + x) \cap S| \geq \frac{\mathrm{Vol}(S)}{\mathrm{Vol}(L)}.$$

Let α be an energy normalization constant and L a $2k$-dimensional lattice in $\mathbb{C}^k$ satisfying $\mathrm{Vol}(L) = 1$. According to Lemma 5.1, we can choose an element $x_R \in \mathbb{C}^k$ such that for the code

$$C = B(\sqrt{kP}) \cap (x_R + \alpha L) \tag{5.3}$$

we have the cardinality bound

$$|C| \geq \frac{\mathrm{Vol}(B(\sqrt{kP}))}{\mathrm{Vol}(\alpha L)} = \frac{C_k P^k}{\alpha^{2k}}, \tag{5.4}$$

where $C_k = \frac{(\pi k)^k}{k!}$. We can now see that given a lattice L with $\mathrm{Vol}(L) = 1$, the number of codewords we are guaranteed to get only depends on the size of α.

From now on, given a lattice L and power limit P, the finite codes we are considering will always satisfy (5.4). We note that while the finite codes are not subsets of the scaled lattice αL, they inherit many properties from the underlying lattice.

5.4 Hermite Invariant in the AWGN Channel

In this section, we will present the classical Hermite invariant approach to build capacity approaching codes for the AWGN channel [5, Chapter 3]. We remark that this channel can be seen as an example of our general set-up (5.1) by assuming that for every k the random matrix H_k is a $k \times k$ identity matrix with probability one. The channel equation can now be written as

$$y = x + w,$$

where $x \in C \subset \mathbb{C}^k$ is the transmitted codeword and w is the Gaussian noise vector.

After the transmission, the receiver tries to guess which was the transmitted codeword x by performing maximum likelihood (ML) decoding, and outputs

$$\hat{x} = \arg\min_{\bar{x}\in C} \|y - \bar{x}\| = \arg\min_{\bar{x}\in C} \|x - \bar{x} + w\| .$$

This suggests a simple code design criterion to minimize the error probability. Given a power limit P, the codewords of C should be as far apart as possible. As the properties of the finite code C are inherited from the underlying lattice, we should give a reasonable definition of what it means that lattice points are far apart.

Definition 5.2 The *Hermite invariant* of a $2k$-dimensional lattice $L_k \subset \mathbb{C}^k$ is defined as

$$\mathrm{h}(L_k) = \frac{\inf\{ ||x||^2 \mid x \in L_k, x \neq 0\}}{\mathrm{Vol}(L_k)^{1/k}},$$

where $\mathrm{Vol}(L_k)$ is the volume of the fundamental parallelotope of the lattice L_k.

Theorem 5.1 *Let $L_k \subset \mathbb{C}^k$ be a family of $2k$-dimensional lattice codes satisfying $\mathrm{h}(L_k) \geq 2kc$, where c is a positive constant. Then any rate*

$$R < \log_2 P - \log_2 \frac{2}{\pi e c}$$

is achievable using the lattices L_k with ML decoding.

Proof Given a power limit P, recall that the finite codes we are considering are of the form $C = B(\sqrt{kP}) \cap (x_R + \alpha L_k)$. Without loss of generality, we can assume that $\mathrm{Vol}(L_k) = 1$. Here α is a power normalization constant that we will soon solve and which will define the achievable rate. The minimum distance in the received constellation is

$$d = \min_{\substack{x,\bar{x}\in C \\ x\neq\bar{x}}} \|x - \bar{x}\| .$$

The error probability is upper bounded by

$$P_e \leq \mathbb{P}\left\{ \|w\|^2 \geq \left(\frac{d}{2}\right)^2 \right\}.$$

Note that we can lower bound the minimum distance as follows:

$$d^2 \geq \alpha^2 \min_{x\in L_k\setminus\{0\}} \|x\|^2 \geq \alpha^2 \mathrm{h}(L_k) \geq \alpha^2 2ck.$$

Therefore we have the upper bound

$$P_e \leq \mathbb{P}\left\{\|w\|^2 \geq \frac{\alpha^2 ck}{2}\right\}. \tag{5.5}$$

Let $\epsilon > 0$. Since $2\|w\|^2$ is a χ^2 random variable with $2k$ degrees of freedom, due to the law of large numbers,

$$\lim_{k\to\infty} \mathbb{P}\left\{\frac{\|w\|^2}{k} \geq 1+\epsilon\right\} = \lim_{k\to\infty} \mathbb{P}\left\{\frac{2\|w\|^2}{2k} \geq 1+\epsilon\right\} \to 0 \tag{5.6}$$

Assuming $\alpha^2 = \frac{2(1+\epsilon)}{c}$, we then have that $P_e \to 0$ when $k \to \infty$, and the cardinality bound (5.4) implies that

$$|C| \geq \frac{C_k P^k}{\alpha^{2k}} = \frac{C_k P^k c^k}{2^k (1+\epsilon)^k}.$$

For large k, $C_k \approx \frac{(\pi e)^k}{\sqrt{2\pi k}}$ using Stirling's approximation.

It follows that $\forall \epsilon > 0$ we can achieve the rate

$$R = \log_2 P - \log_2 \frac{2(1+\epsilon)}{\pi ec}.$$

Since ϵ is arbitrary, this concludes the proof. □

Remark 5.1 There exist several methods to find families of lattices satisfying the condition of Theorem 5.1. For example the Minkowski-Hlawka theorem provides a non-constructive proof of the existence of $2k$-dimensional lattices $L_k \subset \mathbb{C}^k$ having Hermite invariants $\mathrm{h}(L_k) \sim \frac{k}{\pi e}$ [5].

5.5 Hermite Invariant in General Linear Fading Models

In the previous section, we saw how the Hermite invariant can be used as a design criterion to build capacity approaching codes in the AWGN channel. Let us now define a generalization of this invariant for linear fading channels.

Suppose we have an infinite sequence of random matrices $H_k, k = 1, 2 \ldots, \infty$, where H_k is a $k \times k$ matrix. Given an input $x = (x_1, \ldots, x_k)$, we will write the channel output as

$$y = H_k[x] + w,$$

where w is a length k random vector, with i.i.d complex Gaussian entries with variance 1 per complex dimension. We assume that the receiver knows the realization of H.

Given a channel realization H, the receiver outputs the ML estimate

$$\hat{x} = \arg\min_{\bar{x} \in C} \|H[x] + w - H[\bar{x}]\|.$$

From the receiver's perspective this is equivalent to decoding the code

$$H[C] = \{H[x] \mid x \in C\}$$

over an AWGN channel.

As we assumed that the finite codes are of the form (5.3), we have

$$H[C] \subset \{H[x] \mid x \in x_R + \alpha L\} = \{z \mid z \in H[x_R] + \alpha H[L]\},$$

where

$$H[L] = \{H[x] \mid x \in L\}.$$

We can now see that the properties of $H[C]$ are inherited from the set $H[L]$.

If we assume that the matrix H has full rank with probability 1, then the linear mapping $x \mapsto H[x]$ is a bijection of $\mathbb{C}^k$ onto itself with probability 1.

Assuming that $L_k \subset \mathbb{C}^k$ has basis $\{b_1, \ldots, b_{2k}\}$ we have that

$$H[L_k] = \{H[x] \mid x \in L_k\} = \mathbb{Z}H[b_1] \oplus \cdots \oplus \mathbb{Z}H[b_{2k}],$$

is a full-rank lattice with basis $\{H[b_1], \ldots, H[b_{2k}]\}$. Since it is full-rank, we know that $h(H[L_k]) > 0$, but is it possible to choose L_k in such a way that $h(H[L_k])$ would be non-zero irrespective of the channel realization H? Let us now try to formalize this idea.

We can write the random matrix H_k in the form

$$H_k = |\det(H_k)|^{1/k} H_k'$$

where $|\det(H_k')| = 1$. Clearly, if the term $|\det(H_k)|^{1/k}$ happens to be small, it will crush the Euclidean distances of points in $H[L_k]$. However, we will show that if the random matrices H_k are "well behaving", then it is possible to design lattices that are robust against fading.

Definition 5.3 Let $\mathcal{A}$ be a set of invertible matrices such that $\forall A \in \mathcal{A}$, $|\det(A)| = 1$. The *reduced Hermite invariant* [15] of a $2k$-dimensional lattice $L \subset \mathbb{C}^k$ with respect to A is defined as

$$\mathrm{rh}_{\mathcal{A}}(L) = \inf_{A \in \mathcal{A}} \{\mathrm{h}(A[L])\}.$$

It is easy to see that

$$\inf_{A\in\mathcal{A}}\left\{\inf_{x\in L,\, x\neq 0}||A[x]||^2\right\} = \inf_{x\in L,\, x\neq 0}\left\{\inf_{A\in\mathcal{A}}||A[x]||^2\right\}. \tag{5.7}$$

This observation suggests the following definition.

Definition 5.4 We call

$$||x||_{\mathcal{A}} = \inf\{\, ||A[x]|| \mid A \in \mathcal{A}\},$$

the *reduced norm* of the vector x with respect to the set $\mathcal{A}$.

With this observation we realize that

$$\mathrm{rh}_{\mathcal{A}}(L) = \frac{\inf\{\, ||x||_{\mathcal{A}}^2 \mid x \in L,\ x \neq 0\}}{\mathrm{Vol}(L)^{1/k}}. \tag{5.8}$$

If the set $\mathcal{A}$ includes the identity matrix, we obviously have

$$\mathrm{rh}_{\mathcal{A}}(L) \leq \mathrm{h}(L).$$

Suppose that $\{H_k\}_{k\in\mathbb{N}^+}$ is a fading process such that $H_k \in M_{k\times k}(\mathbb{C})$ is full-rank with probability 1, and suppose that the weak law of large numbers holds for the random variables $\{\log\det(H_k H_k^\dagger)\}$, i.e. $\exists\mu > 0$ such that $\forall\epsilon > 0$,

$$\lim_{k\to\infty}\mathbb{P}\left\{\left|\frac{1}{k}\log\det(H_k H_k^\dagger) - \mu\right| > \epsilon\right\} = 0. \tag{5.9}$$

We denote the set of all invertible realizations of H_k with $\mathcal{A}_k^*$. Then define

$$\mathcal{A}_k = \{|\det(A)|^{-1/k}A \mid A \in \mathcal{A}_k^*\}. \tag{5.10}$$

Theorem 5.2 *Let $L_k \subset \mathbb{C}^k$ be a family of $2k$-dimensional lattice codes satisfying $\mathrm{rh}_{\mathcal{A}_k}(L_k) \geq 2kc$ for some positive constant c, and suppose that the channel satisfies (5.9). Then any rate*

$$R < \log_2 P + \mu - \log_2\frac{2}{\pi e c}$$

is achievable using the codes L_k with ML decoding.

Proof Given a power constraint P, recall that we are considering finite codes of the form (5.3), where α is a power normalization constant that we will soon solve.

The minimum distance in the received constellation is

$$d_H = \min_{\substack{x,\bar{x}\in C \\ x\neq\bar{x}}} \|H[x-\bar{x}]\| \geq \min_{\substack{x\in L_k \\ x\neq 0}} \|H[\alpha x]\|,$$

and by the hypothesis on the reduced Hermite invariant,

$$d_H^2 \geq \alpha^2 \min_{x\in L_k\setminus\{0\}} \|H[x]\|^2 \geq \alpha^2 \det(HH^\dagger)^{1/k} \mathrm{rh}_{\mathcal{A}_k}(L_k) \geq \alpha^2 \det(HH^\dagger)^{1/k} 2ck.$$

The ML error probability is bounded by

$$P_e \leq \mathbb{P}\left\{\|w\|^2 \geq \left(\frac{d_H}{2}\right)^2\right\}.$$

Fixing $\epsilon > 0$, the law of total probability implies that

$$\begin{aligned} P_e &\leq \mathbb{P}\left\{\frac{d_H^2}{4k} \geq 1+\epsilon\right\} \mathbb{P}\left\{\|w\|^2 \geq \frac{d_H^2}{4} \;\middle|\; \frac{d_H^2}{4k} \geq 1+\epsilon\right\} + \mathbb{P}\left\{\frac{d_H^2}{4k} < 1+\epsilon\right\} \\ &\leq \mathbb{P}\left\{\frac{\|w\|^2}{k} \geq 1+\epsilon\right\} + \mathbb{P}\left\{\frac{d_H^2}{4k} < 1+\epsilon\right\} \\ &\leq \mathbb{P}\left\{\frac{\|w\|^2}{k} \geq 1+\epsilon\right\} + \mathbb{P}\left\{\frac{\alpha^2 c \det(HH^\dagger)^{1/k}}{2} < 1+\epsilon\right\} \end{aligned}$$

Recall that the first term tends to zero when $k \to \infty$ due to (5.6). The second term will tend to zero as well if we choose

$$\log_2\left(\frac{2(1+\epsilon)}{\alpha^2 c}\right) = \mu - \delta$$

for some $\delta > 0$. Equation (5.4) gives us that

$$R = \frac{1}{k}\log_2|C| \leq \log_2 P - \log_2 \frac{\alpha^2}{C_k}$$

For large k, $C_k \approx \frac{(\pi e)^k}{\sqrt{2\pi k}}$. It follows that we can achieve rate

$$R = \log_2 P + \mu - \delta - \log_2 \frac{2(1+\epsilon)}{\pi e c}.$$

Since ϵ and δ are arbitrary, any rate

$$R < \log_2 P + \mu - \log_2 \frac{2}{\pi e c}$$

is achievable. □

Remark 5.2 In the case of the classical Hermite invariant, there exist several methods to build lattices with large Hermite invariant. In fact, one can prove that on average (with respect to a certain probability measure), random lattices have large Hermite invariants.

However, in the case of the reduced Hermite invariant the situation is quite different. Given a set of matrices $\mathcal{A}_k \subset M_k(\mathbb{C})$, it might be impossible to find even a single lattice for which $\mathrm{rh}_{\mathcal{A}_k}(L) > 0$. Even if we know that such lattices do exist, it might be very hard to find them. Even harder (if possible) it is to find a family of lattices satisfying the conditions of Theorem 5.2 for any c. In the following section, we will give some examples of sets $\mathcal{A}_k$ for which this is possible.

Remark 5.3 Analysing the proof of Theorem 5.2, one can see that maximizing the reduced Hermite invariant can be used as a code design criterion for the corresponding fading channel. In particular, a fixed code satisfying this criterion achieves the same rate under different fading channel statistics.

5.6 Code Design for Diagonal Fading Channels

Let us now consider a fading channel where for every k we have $H_k = \mathrm{diag}[h_1, h_2, \ldots, h_k]$. Assume that each h_i is non-zero with probability 1 and that $\{h_i\}$ forms an ergodic stationary random process. In this model, sending a single symbol x_i during the ith time unit leads to the channel equation

$$y_i = h_i \cdot x_i + w_i, \tag{5.11}$$

where w_i is a zero-mean Gaussian complex random variable with variance 1.

The corresponding set of matrices $\mathcal{A}_k$ in (5.10) is a subset of the set of diagonal matrices in $M_k(\mathbb{C})$ having determinant with absolute value 1.

The assumption that the process $\{h_i\}$ is ergodic and stationary implies that each of the random variables h_i has equal statistics. Therefore we can simply use h to refer to the statistics of all h_i. Assuming now also that $\sum_{i=1}^{k} \frac{1}{k} \log |h_i|^2$ converges in probability to some constant, we have the following.

Corollary 5.1 *Suppose that we have a family of lattices $L_k \subset \mathbb{C}^k$, where $\mathrm{rh}_{\mathcal{A}_k}(L_k) \geq 2kc$ for some positive constant c. Then any rate*

$$R < \mathbb{E}_h \left[\log_2 P\,|h|^2 \right] - \log_2 \frac{2}{\pi e c}$$

is achievable with the family L_k over the fading channel (5.11).

Proof This statement follows immediately from Theorem 5.2, where $\mu = \mathbb{E}_h\left[\log_2 |h|^2\right]$. □

Given two sets $\mathcal{A}'_k \subseteq \mathcal{A}_k$, we have for any lattice L that

$$\mathrm{rh}_{\mathcal{A}'_k}(L) \geq \mathrm{rh}_{\mathcal{A}_k}(L).$$

From now on, we will fix $\mathcal{A}_k$ to be the set of all diagonal matrices in $M_k(\mathbb{C})$ having determinant with absolute value 1. Note that with this choice, if $\mathrm{rh}_{\mathcal{A}_k}(L_k) \geq 2kc$ then Corollary 5.1 holds for any channel of the form (5.11).

Let $(x_1, x_2, \ldots, x_k) \in \mathbb{C}^k$. According to [15, Proposition 8] we have[2]

$$||(x_1, \ldots, x_k)||^2_{\mathcal{A}_k} = k|x_1 \cdots x_k|^{2/k}. \tag{5.12}$$

We can now see that a lattice with large reduced Hermite invariant must have the property that the product of the coordinates of any non-zero element of the lattice is large.

Definition 5.5 Given $x = (x_1, \ldots, x_k) \in \mathbb{C}^k$, we define its *product norm* as $n(x) = \prod_{i=1}^k |x_i|$.

Definition 5.6 Then the *normalized product distance* of L_k is

$$\mathrm{Nd}_{\mathrm{p,min}}(L_k) = \inf_{\mathbf{x} \in L_k \setminus \{0\}} \frac{n(x)}{\mathrm{Vol}(L_k)^{\frac{1}{2}}}. \tag{5.13}$$

Combining (5.12), (5.8) and (5.13) we have that

$$\mathrm{rh}_{\mathcal{A}_k}(L_k) = k(\mathrm{Nd}_{\mathrm{p,min}}(L_k))^{2/k}. \tag{5.14}$$

This result gives us a more concrete characterization of the reduced Hermite invariant and suggests possible candidates for good lattices.

5.6.1 Codes from Algebraic Number Fields

The product distance criterion in the previous section had already been derived in [3] by analyzing the pairwise error probability in the special case where the process $\{h_i\}$ is i.i.d Gaussian. The authors also pointed out that lattices that are derived from number fields have large product distance. We will now shortly present this classical construction and then study how close to the capacity we can get using

[2] More precisely, this result is slightly stronger than the statement of Proposition 8, but it is clear from its proof.

number fields. For the relevant background on algebraic number theory we refer the reader to [17].

Let $K/\mathbb{Q}$ be a totally complex extension of degree $2k$ and $\{\sigma_1, \ldots, \sigma_k\}$ be a set of $\mathbb{Q}$-embeddings, such that we have chosen one from each complex conjugate pair. Then we can define a *relative canonical embedding* of K into $\mathbb{C}^k$ by

$$\psi(x) = (\sigma_1(x), \ldots, \sigma_k(x)).$$

The following lemma is a basic result from algebraic number theory.

Lemma 5.2 *The ring of algebraic integers O_K has a $\mathbb{Z}$-basis $W = \{w_1, \ldots, w_{2k}\}$ and $\{\psi(w_1), \ldots, \psi(w_{2k})\}$ is a $\mathbb{Z}$-basis for the full lattice $\psi(O_K)$ in $\mathbb{C}^k$.*

For our purposes, the key property of the lattices $\psi(O_K)$ is that for any non-zero element $\psi(x) = (\sigma_1(x), \ldots, \sigma_k(x)) \in \psi(O_K)$, we have that

$$\left| \prod_{i=1}^{k} \sigma_i(x) \right|^2 = nr_{K/\mathbb{Q}}(x) \in \mathbb{Z},$$

where $nr_{K/\mathbb{Q}}(x)$ is the algebraic norm of the element x. In particular, it follows that $|\prod_{i=1}^{k} \sigma_i(x)| \geq 1$.

We now know that $\psi(O_K)$ is a $2k$-dimensional lattice in $\mathbb{C}^k$ with the property that $\mathrm{Nd}_{\mathrm{p,min}}(\psi(O_K)) \neq 0$ and therefore $\mathrm{rh}_{\mathcal{A}_k}(\psi(O_K)) \neq 0$. This is true for any totally complex number field. Let us now show how the value of $\mathrm{rh}_{\mathcal{A}_k}(\psi(O_K))$ is related to an algebraic invariant of the field K.

We will denote the *discriminant* of a number field K with d_K. For every number field, it is a non-zero integer.

The following lemma states some well-known results from algebraic number theory and a translation of these results into our coding-theoretic language.

Lemma 5.3 *Let $K/\mathbb{Q}$ be a totally complex extension of degree $2k$ and let ψ be the relative canonical embedding. Then*

$$\mathrm{Vol}(\psi(O_K)) = 2^{-k}\sqrt{|d_K|},$$

$$\mathrm{Nd}_{\mathrm{p,min}}(\psi(O_K)) = \frac{2^{\frac{k}{2}}}{|d_K|^{\frac{1}{4}}} \quad \text{and} \quad \mathrm{rh}_{\mathcal{A}_k}(\psi(O_K)) = \frac{2k}{|d_K|^{1/2k}}.$$

We have now translated the question of finding algebraic lattices with the largest reduced Hermite invariants into the task of finding the totally complex number fields with the smallest discriminant. Luckily this is a well-known mathematical problem with a tradition of almost a 100 years.

In [16], J. Martinet proved the existence of an infinite tower of totally complex number fields $\{K_k\}$ of degree $2k$, where $2k = 5 \cdot 2^t$, such that

$$\left|d_{K_k}\right|^{\frac{1}{k}} = G^2, \tag{5.15}$$

for $G \approx 92.368$. For such fields K_k we have that

$$\mathrm{Nd}_{\mathrm{p,min}}(\psi(O_{K_k})) = \left(\frac{2}{G}\right)^{\frac{k}{2}} \text{ and } \mathrm{rh}_{\mathcal{A}_k}(\psi(O_{K_k})) = \frac{2k}{G}.$$

Specializing Corollary 5.1 to the family of lattices $L_k = \psi(O_{K_k})$ derived from Martinet's tower, which satisfy the hypothesis with $c = 1/G$, we then have the following result:

Proposition 5.1 *Finite codes drawn from the lattices L_k achieve any rate satisfying*

$$R < \mathbb{E}_h\left[\log_2 P\,|h|^2\right] - \log_2 \frac{2G}{\pi e}.$$

Remark 5.4 We note that given a stationary and ergodic fading process $\{h_i\}$ the capacity of the corresponding channel is

$$C = \mathbb{E}_h\left[\log_2(1 + P|h|^2)\right].$$

It is easy to prove that the rate achieved in Proposition 5.1 is a constant gap from the capacity. This gap is also universal in the following sense. Let us consider all ergodic stationary channels with the same first order statistics for $|h|^2$. Then the *same* sequence of finite codes achieve the same gap to capacity in all the channels simultaneously.

Remark 5.5 We note that the number field towers we used are not the best known possible. It was shown in [9] that one can construct a family of totally complex fields such that $G < 82.2$, but this choice would add some notational complications.

Remark 5.6 The families of number fields on which our constructions are based were first brought to coding theory in [12], where the authors pointed out that the corresponding lattices have linearly growing Hermite constant. This directly implies that they are only a constant gap from the AWGN capacity. C. Xing in [27] remarked that these families of number fields provide the best known normalized product distance. Overall number field lattices in fading channels have been well-studied in the literature. However, to the best of our knowledge we were the first to prove that they actually do achieve a constant gap to capacity over fading channels.

5.6.2 Codes from Ideals

As seen in the previous section, lattice codes arising from the rings of algebraic integers of number fields with constant root discriminants will achieve a constant gap to capacity over fading channels. However, known lower bounds for discriminants [18] imply that no matter which number fields we use, the gap cannot be reduced beyond a certain threshold (at least when using our current approach to bound the error probability). It is then natural to ask whether other lattice constructions could lead us closer to capacity. The most obvious generalization is to consider additive subgroups of $\mathcal{O}_K$ and in particular ideals of $\mathcal{O}_K$, which will have non-zero reduced Hermite invariant. Most works concerning lattice codes from number fields focused on either the ring $\mathcal{O}_K$ or a principal ideal $a\mathcal{O}_K$; a more general setting was considered in [1] and [19], which addressed the question of increasing the normalized product distance using non-principal ideals I.

The problem with this approach is that while finding the reduced Hermite invariant of lattices $\psi(\mathcal{O}_K)$ or $\psi(a\mathcal{O}_K)$ is an easy task, the same is not true for $\psi(I)$ when I is non-principal. We will now show how this problem can be reduced to another well-known problem in algebraic number theory and how it can be used to study the performance limits of the lattices $\psi(I)$. Here we will follow the extended arXiv version of [24].

We note that number theoretic proofs are easier when using the equivalent product distance notation rather than the reduced Hermite invariant. Therefore, we will mostly focus on the product distance in this section.

Let K be a totally complex field of degree $2k$. We will use the notation $\mathrm{N}(I) = [\mathcal{O}_K : I]$ for the norm of an ideal I. From classical algebraic number theory, we have that $N(a\mathcal{O}_K) = |nr_{K/\mathbb{Q}}(a)|$ and $N(AB) = N(A)N(B)$.

Lemma 5.4 *Suppose that K is a totally complex field of degree $2k$ and that I is an integral ideal in K. Then $\psi(I)$ is a $2k$-dimensional lattice in $\mathbb{C}^k$ and*

$$\mathrm{Vol}\,(\psi(I)) = [\mathcal{O}_K : I]2^{-k}\sqrt{|d_K|}.$$

This well-known result allows us to compute the volume of an ideal, but computing its normalized product distance is a more complicated issue. In [1, Theorem 3.1], the authors stated the analogue of the following result for the totally real case. It is simply a restatement of the definitions.

Proposition 5.2 *Let us suppose that K is a totally complex field of degree $2k$ and that I is an integral ideal of K. We then have that*

$$\mathrm{Nd}_{\mathrm{p,min}}(\psi(I)) = \frac{2^{\frac{k}{2}}}{|d_K|^{\frac{1}{4}}}\min(I), \tag{5.16}$$

where $\min(I) := \min_{x \in I \setminus \{0\}} \sqrt{\frac{|\mathrm{nr}_{K/\mathbb{Q}}(x)|}{\mathrm{N}(I)}}$.

Proof This result follows from Lemma 5.4, the definition of the normalized product distance and from noticing that $\sqrt{|\mathrm{nr}_{K/\mathbb{Q}}(x)|} = |n(\psi(x))|$. □

Due to the basic ideal theory of algebraic numbers, $\min(I)$ is always larger or equal to 1. If I is not a principal ideal, then we have that $\min(I) \geq \sqrt{2}$. Comparing this to Lemma 5.3 we find that, given a non principal ideal domain O_K, in order to maximize the product distance we should use an ideal I which is not principal. Now there are two obvious questions. Given a non principal ideal domain O_K, which ideal I should we use and how much can we gain? Before answering these questions, we need the following.

Lemma 5.5 ([1]) *For any non-zero element $x \in K$,*

$$\mathrm{Nd}_{\mathrm{p,min}}(\psi(xI)) = \mathrm{Nd}_{\mathrm{p,min}}(\psi(I)).$$

This result proves that every ideal in a given ideal class has the same normalized product distance. It follows that given a ring of integers O_K, it is enough to consider one ideal from every ideal class. Given an ideal I we will denote with $[I]$ the ideal class to which I belongs.

Let us denote with $N_{\min}(K)$ the norm of an ideal A in K with the property that every ideal class of K contains an integral ideal with norm $N(A)$ or smaller. The question of finding the size of $N_{\min}(K)$ is a classical problem in algebraic number theory. We refer the reader to [28] for further reading. The following result is from the extended arXiv version of [24].

Proposition 5.3 *Let us suppose that K is a totally complex number field of degree $2k$ and that I is an ideal that maximizes the normalized product distance over all ideals in K. We then have that*

$$\mathrm{Nd}_{\mathrm{p,min}}(\psi(I)) = \frac{2^{k/2}\sqrt{N_{min}(K)}}{|d_K|^{\frac{1}{4}}} \text{ and } \mathrm{rh}_{A_k}(\psi(I)) = \frac{2k(N_{min}(K))^{1/k}}{|d_K|^{\frac{1}{2k}}}.$$

Proof Let L be any ideal in K, and suppose that A is an integral ideal in the class $[L]^{-1}$ with the smallest norm. Then there exists an element $y \in O_K$ such that $yO_K = AL$. As $n(\psi(y)) = \sqrt{N(L)N(A)}$ and $N(A) \leq N_{\min}(K)$ we have that $\mathrm{d}_{\mathrm{p,min}}(L) \leq \sqrt{N(L)N_{\min}(K)}$ and $\mathrm{Nd}_{\mathrm{p,min}}(\psi(L)) \leq \frac{\sqrt{N_{\min}(K)}2^{k/2}}{|d_K|^{1/4}}$.

Assume that S is an ideal such that $N(S) = N_{\min}(K)$ and choose I as an element from the class $[S]^{-1}$. For any non-zero element $x \in I$, we then have that $xO_K = IC$, for some ideal C that belongs to the class $[S]$. Therefore, we have that $n(\psi(x)) \geq \sqrt{N(I)N(C)}$. □

This result translates the question of finding the product distance of an ideal into a well-known problem in algebraic number theory. It also suggests which ideal class we should use in order to maximize the product distance.

Denote with $\mathcal{K}_{2k}$ the set of totally complex number fields of degree $2k$. Then the optimal normalized product distance over all complex fields of degree $2k$ and all ideals I is

$$\max_{K \in \mathcal{K}_{2k}} \frac{2^{k/2}\sqrt{N_{\min}(K)}}{|d_K|^{\frac{1}{4}}}. \tag{5.17}$$

As far as we know, it is an open question whether the maximum in (5.17) is always achieved when K is a principal ideal domain. Some preliminary data can be found in [1]. We point out that Proposition 5.3 makes this problem computationally much more accessible.

5.7 Reduced Hermite Invariants as Homogeneous Forms

Let us now see how different linear channels define different sets $\mathcal{A}_k$ and how the corresponding reduced norms can be seen as different *homogeneous forms*. For simplicity, we will study the case when we transmit four information symbols (x_1, x_2, x_3, x_4).

In the AWGN channel, the receiver sees

$$(x_1, x_2, x_3, x_4) + (w_1, w_2, w_3, w_4),$$

where w_i are Gaussian random variables. Here the set $\mathcal{A}_4^{(1)}$ simply consists of a single element, the 4×4 identity matrix. Therefore we obviously have

$$||(x_1, x_2, x_3, x_4)||^2_{\mathcal{A}_4^{(1)}} = |x_1|^2 + |x_2|^2 + |x_3|^2 + |x_4|^2.$$

Let us then consider a channel where the fading stays stable for 2 time units and then changes. Then the received signal will be of the form

$$(h_1 x_1, h_1 x_2, h_2 x_3, h_2 x_4) + (w_1, w_2, w_3, w_4).$$

Assuming that h_i are non-zero with probability 1, we can see that

$$\mathcal{A}_4^{(2)} = \{\mathrm{diag}[a_1, a_1, a_2, a_2] \mid |a_1 \cdot a_2| = 1, a_i \in \mathbb{C}\}.$$

Following the proof of [15, Proposition 8], we get the following result

$$||(x_1, x_2, x_3, x_4)||^2_{\mathcal{A}_4^{(2)}} = 2\sqrt{(|x_1|^2 + |x_2|^2) \cdot (|x_3|^2 + |x_4|^2)}. \tag{5.18}$$

Earlier we considered the fast fading channel in which the channel can change during every time unit giving us the following received vector:

$$(h_1x_1, h_2x_2, h_3x_3, h_4x_4) + (w_1, w_2, w_3, w_4).$$

In this case, we have that

$$\mathcal{A}_4^{(3)} = \{\text{diag}[a_1, a_2, a_3, a_4] \mid |a_1 \cdot a_2 \cdot a_3 \cdot a_4| = 1, a_i \in \mathbb{C}\}. \tag{5.19}$$

and that

$$||(x_1, x_2, x_3, x_4)||^2_{\mathcal{A}_4^{(3)}} = 4|x_1x_2x_3x_4|^{1/2}. \tag{5.20}$$

In all the previous examples, the channel could be represented as a diagonal action. On the other hand, for a 2×2 MIMO system, the channel matrix will have a block diagonal structure. In this case, the received vector can be written as

$$(h_1x_1 + h_2x_2, h_3x_1 + h_4x_2, h_1x_3 + h_2x_4, h_3x_3 + h_4x_4) + (w_1, w_2, w_3, w_4).$$

Here the set $\mathcal{A}_4^{(4)}$ consists of matrices

$$\left\{ \begin{pmatrix} h_1 & h_2 & 0 & 0 \\ h_3 & h_4 & 0 & 0 \\ 0 & 0 & h_1 & h_2 \\ 0 & 0 & h_3 & h_4 \end{pmatrix} \mid \det \left| \begin{pmatrix} h_1 & h_2 \\ h_3 & h_4 \end{pmatrix} \right| = 1 \right\}$$

According to [15, Proposition 8], we have that

$$||(x_1, x_2, x_3, x_4)||^2_{\mathcal{A}_4^{(4)}} = 2|(x_1x_2 - x_3x_4)|. \tag{5.21}$$

We immediately note that all the reduced norms share common characteristics.

Definition 5.7 A continuous function $F\colon \mathbb{C}^k \to \mathbb{R}$ is called a *homogeneous form* of degree $\sigma > 0$ if it satisfies the relation

$$|F(\alpha x)| = |\alpha|^\sigma |F(x)| \quad (\forall \alpha \in \mathbb{R}, \forall x \in \mathbb{C}^k).$$

Given a full lattice $L \in \mathbb{C}^k$ and assuming that $\text{Vol}(L) = 1$, we can define the *homogeneous minimum* of the form F as

$$\lambda(F, L) = \inf_{x \in L \setminus \{0\}} |F(x)|.$$

Setting $||\ ||^2_{A_4^{(i)}} = F_{\mathcal{A}_4^{(i)}}$, we can see that each of the squared reduced norms defined previously are homogeneous forms of degree 2.

As we saw in Theorem 5.2, given a sequence of random matrices H_k of size $k \times k$ and the corresponding sets $\mathcal{A}_k$ in (5.10), we can use $\mathrm{rh}_{\mathcal{A}_k}$ as a design criterion for building capacity-approaching lattice codes. In many cases of interest, $\|\ \|^2_{\mathcal{A}_k} = F_{\mathcal{A}_k}$ will be a homogeneous form and $\mathrm{rh}_{\mathcal{A}_k}(L) = \lambda(F, L)$. For instance, this is the case if we extend the previous examples to general size k and define

$$\mathcal{A}_k^{(1)} = I_k,$$
$$\mathcal{A}_{2k}^{(2)} = \{\mathrm{diag}[a_1, a_1, a_2, a_2, \ldots, a_k, a_k] \mid |a_1 a_2 \cdots a_k| = 1, a_i \in \mathbb{C}\},$$
$$\mathcal{A}_k^{(3)} = \{\mathrm{diag}[a_1, a_2, \ldots, a_k] \mid |a_1 a_2 \cdots a_k| = 1, a_i \in \mathbb{C}\}.$$

In the case where $\mathcal{A}_k = \{I_k\}$, we have recovered the classical connection between sphere packing and AWGN capacity, but we also proved that there exist similar connections between different channel models and the corresponding homogeneous forms.

A natural question is now how close to capacity we can get with these methods by taking the best possible lattice sequences in terms of their homogeneous minimum. We will denote with $\mathcal{L}_k$ the set of all the lattices L in $\mathbb{C}^k$ having $\mathrm{Vol}(L) = 1$. This leads us to the concept of *absolute homogeneous minimum*

$$\lambda(F) = \sup_{L \in \mathcal{L}_k} \lambda(F, L).$$

Finding the value of absolute homogeneous minima is one of the central problems in geometry of numbers. As we saw earlier it is a central problem also in the theory of linear fading channels.

In the case $\mathcal{A}_k = \{I_k\}$, $\lambda(F_{\mathcal{A}_k})$ is the *Hermite constant* γ_k. The value of the Hermite constant for different values of k has been studied in mathematics for hundreds of years and there exists an extensive literature on the topic. In particular good upper and lower bounds are available and it has been proven that we can get quite close to Gaussian capacity with this approach [5, Chapter 3].

In the case of $F_{\mathcal{A}_k^{(3)}}$, the problem of finding homogeneous minima has been considered in the context of algebraic number fields and some upper bounds have been provided. Similarly, for $F_{\mathcal{A}_k^{(2)}}$, there exists considerable literature. These and related results can be found in [8]. However, for the case of homogeneous forms arising from block diagonal structures, there seems to be very little previous research. As far as we know, the best asymptotic lower bounds are given in [15].

Remark 5.7 We note that the reduced norms in our examples are not only homogeneous forms, but multivariate polynomials and the sets $\mathcal{A}_k^{(i)}$ are groups. As we obviously have that

$$||A(x)||^2_{\mathcal{A}_k^{(i)}} = ||x||^2_{\mathcal{A}_k^{(i)}},$$

for any $A \in \mathcal{A}_k^{(i)}$, we can see that $||\ ||^2_{\mathcal{A}_k^{(i)}}$ is actually a classical polynomial invariant of the group $\mathcal{A}_k^{(i)}$. At the moment, we do not know what conditions a matrix group $\mathcal{A}_k$ should satisfy so that the corresponding reduced norm would be a homogeneous form. Just as well we do not know when some power of the reduced norm is a polynomial.

This is a nice standalone problem in mathematics but it is also an essential question from the coding theory point of view. Let us elaborate on this. While we have throughout this paper concentrated on asymptotic results and capacity questions, maximizing the reduced Hermite invariant can also be used as a code design criterion for short block lengths. In particular, codes based on this code design principle can be expected to be particularly robust. For example, the lattice that has the largest known reduced Hermite invariant $\mathrm{rh}_{\mathcal{A}_4^{(4)}}$ in [22] also has the best known performance in 2×2 quasi-static MIMO channel for most data rates. Similarly, the number fields maximizing the corresponding reduced Hermite invariant have the best performance in the fast fading SISO channel. In order to find such lattice codes, it is essential to be able to describe the reduced norm in a simple form.

Remark 5.8 We also point out that interestingly all the code constructions that maximize the reduced Hermite invariant are based on algebraic structures. For example, the lattices that maximize $\mathrm{rh}_{\mathcal{A}_4^{(4)}}$ are based on division algebras. The lattices maximizing $\mathrm{rh}_{\mathcal{A}_4^{(3)}}$ on the other hand are built using number fields, as seen in the previous section. We clearly see that $\mathcal{A}_4^{(2)} \subset \mathcal{A}_4^{(3)}$ and $\mathcal{A}_4^{(2)} \subset \mathcal{A}_4^{(4)}$. Therefore for any lattice $L \subset \mathbb{C}^4$ we have that $\mathrm{rh}_{\mathcal{A}_4^{(3)}}(L) \leq \mathrm{rh}_{\mathcal{A}_4^{(2)}}(L)$ and $\mathrm{rh}_{\mathcal{A}_4^{(4)}}(L) \leq \mathrm{rh}_{\mathcal{A}_4^{(2)}}(L)$. It follows that both the division algebra and number field construction can also be used for the block fading SISO channel.

Remark 5.9 While the definition of the reduced Hermite invariant is very natural, we have found very few previous works considering similar concepts. The first reference we have been able to locate is [21]. There the author considered matrices of type (5.19) and proved (5.12) in this special case. Our results can therefore be seen as a natural generalization of this work. The other relevant reference is [2] where the authors defined the Hermite invariant for generalized ideals in division algebras in the spirit of *Arakelov theory*. Again their definition is analogous to ours in certain special cases.

References

1. Bayer-Fluckiger, E., Oggier, F., Viterbo, E.: Algebraic lattice constellations: bounds on performance. IEEE Trans. Inf. Theory **52**(1), 319–327 (2006)
2. Bayer-Fluckiger, E., Cerri, J.-P., Chaubert, J.: Euclidean minima and central division algebras. Int. J. Number Theory **5**(7), 1155–1168 (2009)
3. Boutros, J., Viterbo, E., Rastello, C., Belfiore, J.-C.: Good lattice constellations for both Rayleigh fading and Gaussian channels. IEEE Trans. Inf. Theory **52**(2), 502–518 (1996)
4. Campello, A., Ling, C., Belfiore, J.-C.: Universal lattice codes for MIMO channels. IEEE Trans. Inf. Theory **64**(12), 7847–7865 (2018)
5. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices and Groups. Springer, New York (1988)
6. de Buda, R.: Some optimal codes have structure. IEEE J. Select. Areas Commun. **7**, 893–899 (1989)
7. Erez, U., Zamir, R.: Achieving $1/2\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. IEEE Trans. Inf. Theory **50**(10), 2293–2314 (2004)
8. Gruber, P.M., Lekkerkerker, C.G.: Geometry of Numbers. Elsevier, Amsterdam (1987)
9. Hajir, F., Maire, C.: Asymptotically good towers of global fields. In: Proceedings of the European Congress of Mathematics, pp. 207–218. Birkhäuser, Basel (2001)
10. Hindy, A., Nosratinia, A.: Approaching the ergodic capacity with lattice coding. In: IEEE Global Communications Conference (GLOBECOM), pp.1492–1496, Austin (2014)
11. Hindy, A., Nosratinia, A.: Lattice coding and decoding for multiple-antenna ergodic fading channels. IEEE Trans. Commun. **65**(5), 1873–1885 (2017)
12. Litsyn, S.N., Tsfasman, M.A.: Constructive high-dimensional sphere packings. Duke Math. J. **54**(1), 147–161 (1987)
13. Liu, L., Ling, C.: Polar codes and polar lattices for independent fading channels. IEEE Trans. Commun. **64**, 4923–4935 (2016)
14. Luzzi, L., Vehkalahti, R.: Division algebra codes achieve MIMO block fading channel capacity within a constant gap. In: IEEE International Symposium on Information Theory (ISIT), Hong Kong (2015)
15. Luzzi, L., Vehkalahti, R.: Almost universal codes achieving ergodic MIMO capacity within a constant gap. IEEE Trans. Inf. Theory **63**, 3224–3241 (2017)
16. Martinet, J.: Tours de corps de classes et estimations de discriminants. Invent. Math. **44**, 65–73 (1978)
17. Narkiewicz, W.: Elementary and Analytic Theory of Algebraic Numbers. Springer, Berlin (1980)
18. Odlyzko, A.M.: Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: a survey of recent results. Sém. Théor. Nombres Bordeaux **2**(1), 119–141 (1990)
19. Oggier, F.: Algebraic methods for channel coding. PhD Thesis, EPFL, Lausanne (2005)
20. Ordentlich, O., Erez, U.: Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap. IEEE Trans. Inf. Theory **61**(1), 323–340 (2015)
21. Skriganov, M.M.: Constructions of uniform distributions in terms of geometry of numbers. Algebra i Analiz **6**(3), 200–230 (1994)
22. Vehkalahti, R., Hollanti, C., Lahtonen, J., Ranto, K.: On the densest MIMO lattices from cyclic division algebras. IEEE Trans. Inf. Theory **55**(8), 3751–3780 (2009)
23. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space-time codes for high data rate wireless communications: performance criterion and code construction. IEEE Trans. Inf. Theory **44**, 744–765 (1998)
24. Vehkalahti, R., Luzzi, L.: Number field lattices achieve Gaussian and Rayleigh channel capacity within a constant gap. In: IEEE International Symposium on Information Theory (ISIT), Hong Kong (2015)

25. Vituri, S.: Dispersion Analysis of Infinite Constellations in Ergodic Fading Channels (2015). Available at http://arxiv.org/abs/1309.4638v2. This is a revised and extended version of the author's M.S. Thesis, Department of Electrical Engineering, Tel Aviv University (2013)
26. Weinberger, N., Feder, M.: Universal decoding for linear Gaussian fading channels in the competitive minimax sense. In: IEEE International Symposium on Information Theory (ISIT) Toronto (2008)
27. Xing, C.: Diagonal lattice space-time codes from number fields and asymptotic bounds. IEEE Trans. Inf. Theory **53**, 3921–3926 (2007)
28. Zimmert, R.: Ideale kleiner Norm in Idealklassen und eine Regulatorabschätchung. Invent. Math. **62**, 367–380 (1980)

Chapter 6
Multilevel Lattices for Compute-and-Forward and Lattice Network Coding

Yi Wang, Yu-Chih Huang, Alister G. Burr, and Krishna R. Narayanan

Abstract This work surveys the recent progresses in construction of multilevel lattices for compute-and-forward (C&F) and lattice network coding (LNC). This includes Construction π_A and elementary divisor construction (a.k.a. Construction π_D). Some important properties such as kissing numbers, nominal coding gains, goodness of channel coding, and efficient decoding algorithms of these constructions are also discussed. We then present a multilevel framework of C&F where each user adopts the same nested lattice codes from Construction π_A. The achievable computation rate of the proposed multilevel nested lattice codes under multistage decoding is analyzed. We also study the multilevel structure of LNC, which serves as the theoretical basis for solving the ring-based LNC problem in practice. Simulation results show the large potential of using iterative multistage decoding to approach the capacity.

6.1 Introduction

There has recently been a resurgence in research on lattice codes. Erez and Zamir [5] have shown that lattice codes can achieve the channel capacity with nested lattice shaping and an MMSE estimator at the receiver. This is indeed an inspiration for researchers to explore lattice code in the area of network and wireless communications. The random ensemble of nested lattice codes is capable of producing capacity-achieving lattice codes, and its structure is in particular suitable

Y. Wang (✉) · A. G. Burr
Department of Electronic Engineering, University of York, York, UK
e-mail: yi.wang@york.ac.uk; alister.burr@york.ac.uk

Y.-C. Huang
Department of Communication Engineering, National Taipei University, New Taipei City, Taiwan

K. R. Narayanan
Department of Electrical & Computer Engendering, Texas A&M University, College Station, TX, USA
e-mail: krn@tamu.edu

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*, Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_6

for many problems within network communications, which opens the window in exploiting the structure gain induced by the channels.

The use of nested lattice codes in physical layer network coding was first proposed by Nazer and Gastpar, who developed the compute-and-forward (C&F) relaying strategy as a compelling information-transmission scheme in Gaussian relay networks. Two key features of C&F have made the scheme attractive in network communications: (1) the relay computes a set of linearly independent equations instead of directly decoding/amplifying the source messages/received signals. (2) no channel state information (CSI) is required at the transmitters and no global channel-gain information is required at the destination. Feng et al. formulated a generic algebraic framework, namely lattice network coding (LNC) which employs algebraic approach to reinterpret C&F and make a direct connection between LNC and C&F. In particular, the LNC makes no assumptions on the underlying nested lattice codes, and induces an end-to-end linear network coding channel over modules.

This chapter places the fundamental theories behind LNC, which provides the design guidelines of lattice-based network coding from the aspect of practical implementation. The use of Construction A lattices provides a feasible way to implement LNC networks. However, the decoding complexity of a Construction A lattice closely relies on the underlying linear code over a prime field. When the size of the prime field increases, decoding these lattice codes within LNC networks is typically infeasible due to the exponential increase of the computational cost.

This chapter induces the concepts of multilevel lattices and correspondingly the multilevel lattice network coding (MLNC) strategy, which resolves the complexity problem and retains the property of LNC in the meantime. The theory leads to a new lattice construction approach, namely the elementary divisor construction, which is a reinterpretation of multilevel lattices construction approaches—Construction π_A and Construction π_D (introduced in the next few sections). These multilevel lattices construction methods developed subsume most of the existing construction methods, e.g. Construction A and Construction D.

Multi-stage lattice decoding, especially iteration-aided multi-stage lattice decoding, is therefore proposed in the next few sections. The simulation results reveal that with the aid of iterative decoding, the system performance is greatly improved with significantly reduced computational costs.

MLNC, multilevel lattices and multi-stage lattice decoding resolve the problem of decoding lattice codes based on fields or rings constituting a large message space, where the multilevel lattices based on these can readily be partitioned into a set of primary sublattices based on much smaller message spaces following the theory developed, so that multistage lattice decoding may be used for decoding. Note that MLNC provides the theoretical basis of a practical implementation strategy when using lattice codes within Gaussian relay networks. It is therefore a generic theory which is suitable not just for lattices constructed from channel codes, but for any lattice codes, e.g. complex low density lattice codes or signal codes.

6.2 Problem Statement

We consider the compute-and-forward relay network introduced by Nazer and Gastpar [21] which consists of L source nodes and M destination nodes as shown in Fig 6.1. Each source node has a message $w_\ell \in \{1, 2, \ldots, W\}$, $\ell \in \{1, \ldots, L\}$ which can alternatively be expressed by a length-N' vector over some finite field, i.e., $\mathbf{w}_\ell \in \mathbb{Z}_q^{N'}$ where $q \in \mathbb{N}$ with $W = q^{N'}$. This message is fed into an encoder $\mathcal{E}_\ell^N$ whose output is a length-N codeword $\mathbf{x}_\ell \in \mathbb{C}^N$. The transmitted signal is subject to an average power constraint given by

$$\mathbb{E}[\mathsf{X}^2] \leq P. \tag{6.1}$$

The signal observed at the destination m is given by

$$y_m[n] = \sum_{\ell=1}^{L} h_{m\ell} x_\ell[n] + z_m[n], \quad n \in \{1, \ldots, N\}, \tag{6.2}$$

where $h_{m\ell} \in \mathbb{C}$ is the channel coefficient between the source node ℓ and destination m, and $z_m[n] \sim \mathcal{CN}(0, 1)$. Collectively, one can also define the channel model for using the channel N times as

$$\mathbf{y}_m = \sum_{\ell=1}^{L} h_{m\ell} \mathbf{x}_\ell + \mathbf{z}_m. \tag{6.3}$$

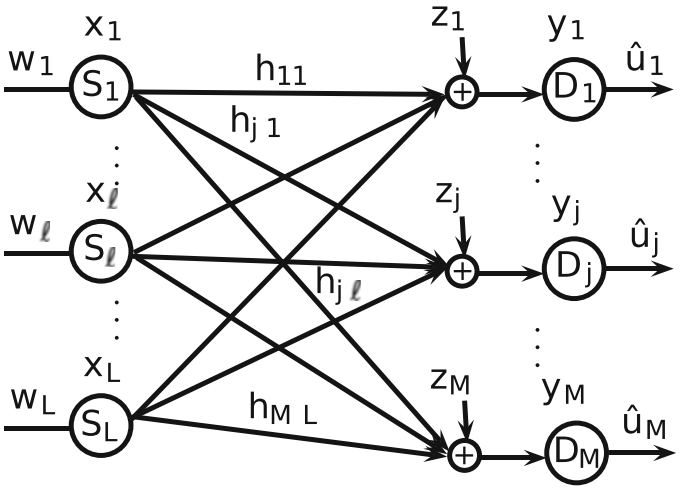


Fig. 6.1 A compute-and-forward relay network where $S_1, \ldots, S_L$ are source nodes and $D_1, \ldots, D_M$ are destination nodes

Instead of individual messages, each destination node is only interested in computing a function of messages

$$\mathbf{u}_m = f_m(w_1, \ldots, w_L). \tag{6.4}$$

Upon observing $\mathbf{y}_m$, the destination node m forms $\hat{u}_m = \mathcal{G}_m^N(\mathbf{y}_m)$ an estimate of u_m. These functions are then forwarded to the central destination which can recover all the messages given sufficiently many functions.

Definition 6.1 (Computation Codes) For a given set of functions $f_1, \ldots, f_M$, a (N, N') computation code consists of a sequence of encoding/decoding functions $(\mathcal{E}_1^N, \ldots, \mathcal{E}_L^N)/(\mathcal{G}_1^N, \ldots, \mathcal{G}_M^N)$ described above and an error probability given by

$$P_{e,m}^{(N)} \triangleq \mathsf{P}_\mathsf{e}\left(\left\{\hat{u}_m \neq u_m\right\}\right). \tag{6.5}$$

Definition 6.2 (Computation Rate at the Relay m) For a given channel vector $\mathbf{h}_m \triangleq [h_{m1}, \ldots, h_{mL}]^T$ and a given function f_m, a computation rate $R(\mathbf{h}_m, f_m)$ is achievable at the relay m if for any $\varepsilon > 0$ there is an (N, N') computation code such that

$$N' \geq NR(\mathbf{h}_m, f_m)/\log(q) \text{ and } P_{e,m}^{(N)} \leq \varepsilon. \tag{6.6}$$

Note that the first condition is equivalent to saying that $W \geq 2^{NR(\mathbf{h}_m, f_m)}$.

Throughout the paper, we consider equal power constraints and assume all the transmitters transmit at a same rate. However, similar to [20], the proposed framework can be extended to the unequal power constraint and/or unequal rate cases.

In practice, since no cooperation among the relays is assumed, a greedy protocol which mimics the behavior of random linear network coding is adopted in [21] where each relay computes and forwards the function with the highest computation rate. After that, if given those functions, the central destination is able to recover all the messages, then decoding is successful. Otherwise, the central destination declares failure. The achievable computation rate for the transmitters is then equal to $\min_m R(\mathbf{h}_m, f_m)$.

6.3 Background

6.3.1 Algebra

6.3.1.1 Ideal and Principal Ideal Domain

Let S be a commutative ring with identity 1, and $S^* = S\backslash 0$. The set of *units* $\mathcal{U}(S)$ in S refers to any element x in S such that $xs = sx = 1$ for some $s \in S$. Any root of

unity in a ring S is a unit. The set of *zero divisors* $\mathcal{Z}(S)$ in S refers to any element x in S if $xs = sx = 0$ for some $s \in S^*$. An element $p \in S$, $p \notin \mathcal{Z}(S)$, $p \notin \mathcal{U}(S)$, is called a *prime* in S when $p \mid ab$ for some $a, b \in S^*$, implies either $p \mid a$ or $p \mid b$.

An *ideal* $\mathcal{I}$ of R is a non-empty subset of R that is closed under subtraction (which implies that $\mathcal{I}$ is a group under addition), and is defined by:

1. $\forall a, b \in \mathcal{I}, a - b \in \mathcal{I}$.
2. $\forall a \in \mathcal{I}, \forall s \in S$, then $as \in \mathcal{I}$ and $sa \in \mathcal{I}$.

If $A = \{a_1, \cdots, a_m\}$ is a finite non-empty subset of S, we use $\langle a_1, \cdots, a_m \rangle$ to represent the ideal generated by A, i.e.

$$\langle a_1, \cdots, a_m \rangle = \{a_1 s_1 + \cdots + a_m s_m : s_1, \cdots, s_m \in S\}$$

Note that S has at least two ideals $\{0\}$ and $\{S\}$.

An ideal $\mathcal{I}$ of S is said to be *proper* if and only if $1 \notin \mathcal{I}$. An ideal $\mathcal{I}_{\max}$ is said to be *maximal* if $\mathcal{I}_{\max}$ is a proper ideal and the only ideals that include $\mathcal{I}_{\max}$ are S and $\mathcal{I}_{\max}$ itself. We say that an equivalence relation $a \sim b$ on the set S is defined by $\mathcal{I}$ if and only if $a - b \in \mathcal{I}$.

An ideal $\mathcal{I}$ of S is *principal* if $\mathcal{I}$ is generated by a single element $a \in \mathcal{I}$, written as $\mathcal{I} = \langle a \rangle$. A *principal ideal ring* is a ring whose every ideal is principal. If S is a principal ideal ring without zero divisors, then R forms an ideal domain, and more precisely, a *principal ideal domain* (PID). Examples of PIDs include the ring of integers, the ring of Gaussian integers $\mathbb{Z}[i]$ and the ring of Eisenstein integers $\mathbb{Z}[\omega]$.

6.3.1.2 Modules Over PID

Again, let S be a commutative ring with identity 1. An S-module M over S is an abelian group $(M, +)$ under a binary operation $+$, together with a function $\mathscr{F} : S \times M \longmapsto M$ which satisfies the same conditions as those for vector space. Note that modules over a field are the same as vector spaces. An S-submodule of M is a subgroup N of M which is closed under the action of ring elements, and hence the submodule N forms also an S-module under the restricted operations.

An S-module is said to be finitely generated (*f.g.*) if M has a finite spanning set $\{m_1, \cdots, m_n\}$ such that $\sum_i Rm_i = M$.

The annihilator $\text{Ann}_S(m)$ of an element $m \in M$ is the set of elements $s \in S$ such that $sm = 0$, which forms an ideal. The annihilator of M is the elements $s \in S$ such that $\{sm = 0 | \forall m \in M\}$, denoted by $\text{Ann}_S(M) = \bigcap \{\text{Ann}_S(m) | m \in M\}$. If M is a free S-module, then $\text{Ann}_S(M) = \langle 0 \rangle$.

We define an action of S satisfying $\forall s \in S, \forall m \in M$, and for all $m + N \in M/N$,

$$s(m + N) = sm + N,$$

then M/N is referred to as a quotient S-module.

The torsion submodule M_{Tor} of M is defined by:

$$M_{\text{Tor}} = \{m \in M : \text{Ann}_S(m) \neq \{0\}\}$$

A torsion free module is trivial.

Let M and N be two S-modules. An S-module homomorphism is a map $\phi : (M, +, \cdot) \longmapsto (N, \boxplus, \odot)$, which respects the S-module structures of M and N, i.e.,

$$\phi(s_1 m_1 + s_2 m_2) = s_1 \phi(m_1) \boxplus s_2 \phi(m_2)$$

$$\phi(s_1 m_1 \cdot s_2 m_2) = s_1 \phi(m_1) \odot s_2 \phi(m_2)$$

$\forall s_1, s_2 \in S, \forall m_1, m_2 \in M$. An S-module homomorphism $\phi : M \longmapsto N$ is called an S-module isomorphism if it is both injective and surjective, which is denoted by $M \cong N$. The kernel of ϕ denotes the elements in M which makes the image of ϕ equal to zero.

6.3.2 *Lattices and Lattice Codes*

Lattices defined within $\mathbb{R}$ are explained as follows: An N-dimensional lattice Λ^N is a discrete subgroup of $\mathbb{R}^N$ which satisfies $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 \in \Lambda^N$ and $-\boldsymbol{\lambda}_1 \in \Lambda^N$ whenever $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda^N$. One way to express a lattice is through its generator matrix $\mathbf{G}_\Lambda \in \mathbb{R}^{N \times N}$ as

$$\Lambda = \left\{ \mathbf{G}_\Lambda \mathbf{b} : \mathbf{b} \in \mathbb{Z}^N \right\}. \tag{6.7}$$

For any vector $\mathbf{x} \in \mathbb{R}^N$, a nearest neighbor quantizer associated with Λ quantizes $\mathbf{x}$ to the nearest element in Λ. That is,

$$Q_\Lambda(\mathbf{x}) \triangleq \underset{\boldsymbol{\lambda} \in \Lambda}{\text{argmin}} \|\mathbf{x} - \boldsymbol{\lambda}\|, \tag{6.8}$$

where $\|.\|$ denotes the L_2-norm and the ties are broken systematically. The fundamental Voronoi region $\mathcal{V}_\Lambda$ is the collection of all $\mathbf{x} \in \mathbb{R}^N$ that result in $Q_\Lambda(\mathbf{x}) = \mathbf{0}$. i.e.,

$$\mathcal{V}_\Lambda \triangleq \{\mathbf{x} \in \mathbb{R}^N : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}. \tag{6.9}$$

If $\mathbf{G}$ is full rank, the volume of $\mathcal{V}_\Lambda$, which we denote by $\text{Vol}(\mathcal{V}_\Lambda)$, can be easily computed as $\text{Vol}(\mathcal{V}_\Lambda) = |\det(\mathbf{G})|$. The $\mod \Lambda$ operation returns the quantization error with respect to Λ. Mathematically, it is given by

$$\mathbf{x} \mod \Lambda = \mathbf{x} - Q_\Lambda(\mathbf{x}). \tag{6.10}$$

Alternatively, this can be thought of as mapping $\mathbf{x}$ to the element of the coset $\mathbf{x}+\Lambda$ within $\mathcal{V}_\Lambda$. We shall refer $\mathbf{x} \mod \Lambda$ to as the *coset leader* of $\mathbf{x}+\Lambda$.

Let us now consider the problem of transmission over an additive white Gaussian noise channel without any power constraint. We adopt a lattice Λ as our transmission scheme and every lattice point can be sent since there is no power constraint. The signal model is given by

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \tag{6.11}$$

where $\mathbf{x} \in \Lambda$ is the transmitted signal, $\mathbf{y}$ is the received signal, and $\mathbf{z}$ is the additive noise whose elements are drawn i.i.d. from $\mathcal{N}(0, \sigma^2)$. One attempts to form $\hat{\mathbf{x}}$ an estimate of $\mathbf{x}$ based on the received $\mathbf{y}$. The decoding probability is defined as $p_e \triangleq \mathbb{P}\left\{\hat{\mathbf{x}} \neq \mathbf{x}\right\}$. A sequence of lattices Λ is said to be *good for channel coding* if $p_e \to 0$ in the limit as $n \to \infty$ as long as the volume of $\mathcal{V}_\Lambda$ is larger than the volume of the typical noise ball, i.e.,

$$\sigma^2 < \frac{\mathrm{Vol}(\mathcal{V}_\Lambda)^{\frac{2}{N}}}{2\pi e}. \tag{6.12}$$

6.3.3 Construction A

Here, we briefly review one of the most famous constructions, namely Construction A. *Construction A*[3, 16]: Let p be a prime and C be a linear code of length N and dimension r over $\mathbb{F}_p$, i.e., C is a r-dimensional subspace of the vector space $\mathbb{F}_p^N$. The Construction A lattice associated with C is given by

$$\Lambda = \left\{\boldsymbol{\lambda} \in \mathbb{Z}^N : \boldsymbol{\lambda} \mod p \in C\right\}. \tag{6.13}$$

An alternative and constructive way to describe this lattice construction is as follows. Let $\mathcal{M}$ be the natural mapping from $\mathbb{F}_p$ onto the coset leaders of $\mathbb{Z}/p\mathbb{Z}$. We denote by $\mathcal{M}^N : \mathbb{F}_p^N \to (\mathbb{Z}/p\mathbb{Z})^N$ that performs $\mathcal{M}$ elementwisely. Construction A "lifts" C to the Euclidean space $\mathbb{R}^N$ by taking the union of all the cosets $\mathcal{M}^N(\mathbf{c}) + p\mathbb{Z}^N$, $\mathbf{c} \in C$, which forms

$$\Lambda = \bigcup_{\mathbf{c}\in C} \left(\mathcal{M}^N(\mathbf{c}) + p\mathbb{Z}^N\right) = \mathcal{M}^N(C) + p\mathbb{Z}^N, \tag{6.14}$$

where $+$ above is the Minkowski sum. The fact that Construction A always produces lattices is due to the linearity of C and the isomorphism nature of the natural mapping $\mathcal{M}$.

Construction A lattices have been popular for decades due to the tight connection between lattices generated and their underlying linear codes. In [18], Loeliger

exploits the close connection between Construction A lattices and their underlying linear codes. He then uses “random coding argument" to show that Construction A can produce lattices that are good for channel coding. Erez et al. in [4] show that Construction A can produce lattices that are good in many senses simultaneously including packing, shaping, channel coding and MSE quantization. In [5], an ensemble of nested lattice codes carved from Construction A lattices is proposed and is shown to achieve the AWGN channel capacity. Ordentlich and Erez later simplify the proof of the capacity-achieving property in [22] by introducing a new ensemble of nested lattice codes which preserve the tight connection between the lattice codes and the linear codes. In this article, we will consider several multilevel lattice constructions evolved from Construction A.

6.4 Compute-and-Forward and Lattice Network Coding

In this section, we briefly review the compute-and-forward paradigm [21] and the lattice network coding framework [8]. It should be noted that although lattice network coding subsumes every lattice-based scheme (including compute-and-forward in [21]) as a special instance, we review both the frameworks. It is mainly because lattice network coding is a general framework and as a consequence, it is hard to prove the achievable computation rates as well as to construct optimal coding schemes. In contrast, compute-and-forward specifically uses nested lattice codes from Construction A lattices and thus its achievable computation rates can be derived.

6.4.1 Compute-and-Forward

In [21], the destination m aims at computing a function of the form

$$\mathbf{u}_m = a_{m1}\mathbf{w}_1 \oplus \ldots \oplus a_{mL}\mathbf{w}_L, \tag{6.15}$$

where $\mathbf{w}_\ell$ is the p-ary expansion of w_ℓ and $a_{m\ell} \in \mathbb{F}_p$. It first computes the linear combinations of codewords whose coefficients are Gaussian integers $\mathbf{a}_m = [a_{m1}, \ldots, a_{mL}]$ and then maps this integer combination back to linear combinations of messages in (6.15) where $b_{m\ell} \triangleq \sigma(a_{m\ell})$ with $\sigma(.)$ being the ring homomorphism used in Construction A for generating the underlying lattice [3, 16]. In this scenario, the function f_m is completely characterized by the coefficients $\mathbf{a}_m$ and thus the achievable computation rate is denoted as $R(\mathbf{h}_m, \mathbf{a}_m)$.

Each source node adopts an identical nested lattice code (Λ_f, Λ_c) of Erez and Zamir [5]. The source node ℓ first bijectively maps its message $\mathbf{w}_\ell$ to a lattice codeword $\mathbf{t}_\ell \in \Lambda_f \cap \mathcal{V}_{\Lambda_c}$ and sends a dithered version

$$\mathbf{x}_\ell = (\mathbf{t}_\ell - \mathbf{u}_\ell) \bmod \Lambda_c. \tag{6.16}$$

Given $\mathbf{a}_m = [a_{m1}, \ldots, a_{mL}]^T$, the relay m scales the received signal by the MMSE estimator α_m and adds the dithers back to form

$$\mathbf{y}'_m = \left(\alpha_m \mathbf{y}_m + \sum_{\ell=1}^{L} a_{m\ell}\mathbf{u}_\ell\right) \bmod \Lambda_c \tag{6.17}$$

$$= (\mathbf{t}_{eq,m} + \mathbf{z}_{eq,m}) \bmod \Lambda_c, \tag{6.18}$$

where

$$\mathbf{t}_{eq,m} = \sum_{\ell=1}^{K} a_{m\ell}\mathbf{t}_{m\ell} \bmod \Lambda_c, \tag{6.19}$$

and

$$\mathbf{z}_{eq,m} = \left(\alpha_m \mathbf{z}_m + \sum_{\ell=1}^{L} (\alpha_m h_{mL} - a_{mL})\mathbf{x}_\ell\right), \tag{6.20}$$

with

$$\begin{aligned}\sigma^2_{eq,m} &\triangleq \frac{1}{n}\mathbb{E}\|\mathbf{z}_{eq,m}\|^2 \\ &= |\alpha_m^2| + P\|\alpha_m \mathbf{h}_m - \mathbf{a}_m\|^2.\end{aligned} \tag{6.21}$$

Due to the linearity of lattice codes, $\mathbf{t}_{eq,m}$ is a codeword in $\Lambda_f \cap \mathcal{V}_{\Lambda_c}$ and hence one can directly compute this function at the relay m. Intuitively speaking, one can arbitrarily rotate and scale the received signals by α_m such that the resulting channel coefficients would be arbitrarily close to the Gaussian integer vector $\mathbf{a}_m$; however, one might also cause uncontrolled noise enhancement. It turns out that the optimal choice of α_m is the linear MMSE estimator given by

$$\alpha_{\mathrm{MMSE},m} = \frac{P\mathbf{h}_m^* \mathbf{a}_m}{1 + P\|\mathbf{h}_m\|^2}. \tag{6.22}$$

This leads us to the main result of [21] as follows.

Theorem 6.1 (Nazer-Gastpar) *For given channel coefficients* $\mathbf{h}_m$ *and Gaussian integer vector* $\mathbf{a}_m$*, the following computation rate is achievable at the relay* m*.*

$$R(\mathbf{h}_m, \mathbf{a}_m) = \log^+\left(\left(\|\mathbf{a}_m\|^2 - \frac{P|\mathbf{h}_m^* \mathbf{a}_m|^2}{1 + P\|\mathbf{h}_m\|^2}\right)^{-1}\right), \tag{6.23}$$

where $\log^+(.) \triangleq \max\{0, \log(.)\}$.

After computing $\mathbf{t}_{eq,m}$, the relay m can recover the function $\mathbf{u}_m$ in (6.15). At the central destination, one can invert the matrix $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_M]$ to recover all the messages if the matrix is invertible.

Remark 6.1 The coding scheme in [21] in fact transmits signals in the real and the imaginary parts separately and independently. Here, we describe the scheme by directly looking at the complex field and Gaussian integers. This perspective has motivated the generalization of the compute-and-forward paradigm to the ring of Eisenstein integers in [25, 26] and other rings of imaginary quadratic integers in [15] where $\mathbf{a}_m$ in (6.23) is chosen from $\mathbb{Z}[\omega]$ and other imaginary quadratic integers, respectively, instead of $\mathbb{Z}[i]$.

6.4.2 *Lattice Network Coding*

Feng et al. formulated a general algebraic framework for lattice network coding (LNC) [9], giving practical design guidelines for compute-and-forward. LNC is based on a finite lattice quotient, in which each transmitter sends an information-embedding coset through a coset representative. LNC scheme serves as a direct connection between C&F and module theory (in abstract algebra). In particular, a generic LNC makes no assumptions on the structure of the underlying nested lattice code, which makes a variety of code-design techniques available.

The key aspect of LNC is the so-called linear labelling of the points in a nested lattice code which produces a beneficial compatibility between the arithmetic operations and the linear operations in the message space that used for linear network coding. The linear labellings induces a noncoherent network coding channel with a message space having the module-theoretic algebraic structure. This provides the theoretical basis for achieving network coding over general Gaussian relay networks.

LNC specifies a map $\varphi : \Lambda \to W$ from lattice points in Λ to messages in the message space W to facilitate practical implementation of linear labelling. The map φ satisfies two conditions,

1. for any two points $\lambda_1, \lambda_2 \in \Lambda$ with $\lambda_1 - \lambda_2 \in \Lambda'$, $\varphi(\lambda_1) = \varphi(\lambda_2)$;
2. $\varphi(s_1\lambda_1 + s_2\lambda_2) = s_1\varphi(\lambda_1) + s_2\varphi(\lambda_2)$, $\forall s_1, s_2 \in S$ and $\forall \lambda_1, \lambda_2 \in \Lambda$.

As discussed above, C&F is specifically implemented by constructing lattices with Construction A. LNC generalises this and allows more powerful lattice codes to be used, e.g. low density lattice codes [29] and signal codes [7] which provides high coding gain.

Previous work, e.g. in [8, 24, 27], has given LNC design guidelines when quotient lattices are constructed from existing channel codes using complex construction A. In this book chapter, we consider a multilevel structure for lattice network coding, which provides a practical solution to the ring-based network coding problem. We also show an efficient lattice construction approach (which we term the elementary divisor construction (EDC)) based on the theorems developed, which also subsumes the most important previous lattice constructions.

6.5 Multilevel Lattices Evolved from Construction A

In this section, we review some multilevel lattices evolved from Construction A lattices including Construction D, Construction π_A, and elementary divisor construction (a.k.a. Construction π_D). Efficient decoding algorithms for these lattices are also discussed. These multilevel lattices will enable multistage compute-and-forward presented in Sect. 6.6.

6.5.1 Construction D

Construction D [1] [3, Page 232] is a multilevel lattice construction that constructs a lattice from a sequence of nested linear codes. Consider a sequence of nested linear codes constructed over $\mathbb{F}_p$, namely $C^1 \subseteq C^2 \subseteq \ldots \subseteq C^{\gamma+1}$. In this sequence of codes, $C^{\gamma+1}$ is the trivial (N, N)-code and C^l is a (N, r^l)-code for $l \in \{1, 2, \ldots \gamma\}$ with $r^1 \leq \ldots \leq r^\gamma$. One way to generate such a sequence of codes is to first choose $\{\mathbf{g}_1, \ldots, \mathbf{g}_N\}$ that spans $C^{\gamma+1}$ and then use only the subset of the first r^l vectors $\{\mathbf{g}_1, \ldots, \mathbf{g}_{r^l}\}$ to generate C^l. The procedure of Construction D is given as follows.

A multilevel lattice Λ_D with $\gamma + 1$ is given by

$$\Lambda_\mathrm{D} = \bigcup \left\{ p^\gamma \mathbb{Z}^N + \sum_{1 \leq l \leq \gamma} p^{l-1} \sum_{1 \leq i \leq r^l} a_{li} \mathbf{g}_i | a_{li} \in \{0, 1, \ldots, p-1\} \right\}, \tag{6.24}$$

where all the operations are over $\mathbb{R}^N$. In [6], an alternative presentation of Construction D is given as a extension of Construction A with coding over a finite chain ring. This is done by relating the nested linear codes with a single linear code C over $\mathbb{Z}_{p^\gamma}$. Let $\mathcal{M} : \mathbb{Z}_{p^\gamma} \rightarrow \mathbb{Z}/p^\gamma\mathbb{Z}$ be a ring isomorphism. A Construction D

lattice associated with C can be alternatively represented as

$$\Lambda_{\mathrm{D}} = \mathcal{M}^n(C) + p^{\gamma}\mathbb{Z}^n. \tag{6.25}$$

Construction D has been shown to be able to produce lattices that are good for channel coding [11]. Recently in [32], a lattice ensemble called polar lattices has been proposed which use the Construction D procedure with nested polar codes as the underlying linear codes. Thanks to the explicitness of the construction of good polar codes, polar lattices have provided an explicit construction of lattices that are good for channel coding. Later in [17], polar lattices have also been shown to achieve the rate distortion bound of memoryless Gaussian source.

6.5.2 Construction π_A

We now present Construction π_A in two different ways that are equivalent to each other: The first presentation is to regard it as a generalization of Construction A to allow the underlying codes being over $\mathbb{Z}_q$ where q is a positive squarefree integer (A number is said to be squarefree if its prime decomposition contains no repeated factors.). The second one is to think of it as a generalization of Construction A to multilevel codes in which each level's code is over a different prime field.

Construction π_A constructs lattices from linear codes over finite rings $\mathbb{Z}_q$, where $q \in \mathbb{N}$ is chosen to be squarefree and hence can be factorized into a product of primes as $q = p_1 \cdot \ldots \cdot p_L$. Let C be a linear code over $\mathbb{Z}_q$. The Construction π_A lattice associated with C is given by

$$\Lambda_{\pi_A} = \left\{ \boldsymbol{\lambda} \in \mathbb{Z}^N : \boldsymbol{\lambda} \bmod \ \in C \right\} \tag{6.26}$$

Let $\mathbf{G}$ be a generator matrix of C. From the Chinese remainder theorem, C can be uniquely decomposed into K linear codes $C^1, \ldots, C^K$ where $\mathbf{G}_l = \mathbf{G} \mod p_l$ is a generator matrix of C^l. Evidently, C^l is a linear code over the prime field $\mathbb{F}_{p_l}$.

An alternative and constructive way to describe Construction π_A is as follows. Let $p_1, \ldots, p_K$ be K distinct primes and let $q = p_1 \cdot \ldots \cdot p_K$. From Chinese remainder theorem, there exists a ring isomorphism

$$\mathcal{M} : \mathbb{F}_{p_1} \times \ldots \times \mathbb{F}_{p_K} \to \mathbb{Z}/q\mathbb{Z}. \tag{6.27}$$

We start with K linear codes. Let C^l, $l \in \{1, \ldots, K\}$, be a (N, r^l) linear code constructed over $\mathbb{F}_{p_l}$. We map the codewords in $C^1 \times \ldots \times C^K$ to $\Lambda^* \in (\mathbb{Z}/q\mathbb{Z})^N$ by applying $\mathcal{M}$ element-wise. We again denote by $\mathcal{M}^N$ for this function that maps

elements in $(\mathbb{F}_{p_1} \times \ldots \times \mathbb{F}_{p_L})^N$ to $(\mathbb{Z}/q\mathbb{Z})^N$ element-wise. After this, we tile Λ^* to $\mathbb{R}^N$. Overall, we obtain

$$\begin{aligned}\Lambda_{\pi_A} &= \mathcal{M}^N(C^1, \ldots, C^L) + q\mathbb{Z}^N \\ &= \left\{\boldsymbol{\lambda} \in \mathbb{Z}^N : \sigma(\boldsymbol{\lambda}) \in C^1 \times \ldots \times C^L\right\}, \end{aligned} \tag{6.28}$$

where $\sigma \triangleq \mathcal{M}^{-1} \circ \mod q$ is a ring homomorphism. It has been shown in [13] that there exists a sequence of Construction π_A lattices that is good for channel coding.

6.5.3 Multilevel Lattice Network Coding

We assume S is a PID over $\mathbb{C}$. Briefly if there is a matrix $\mathbf{G}_\Lambda \in \mathbb{C}^{n' \times n}$, $n' \leq n$ such that all its n' row vectors $\mathbf{g}_{\Lambda,1}, \cdots, \mathbf{g}_{\Lambda,n'} \in \mathbb{C}^n$ are linearly independent, the set of all S-linear combinations of $\mathbf{g}_{\Lambda,1}, \cdots, \mathbf{g}_{\Lambda,n'}$ forms an S-lattice $\Lambda \in \mathbb{C}^n$, written by, $\Lambda = \{\mathbf{s}\mathbf{G}_\Lambda : \mathbf{s} \in S^{n'}\}$, where $\mathbf{G}_\Lambda$ is called the lattice generator.

Following the explanation in Sect. 6.3.1.2, an n-dimensional S-lattice is precisely an S-module, and similarly the sublattice Λ' in Λ forms a S-submodule. The partition of the S-lattice, denoted by Λ/Λ' represents $|\Lambda : \Lambda'| < \infty$ (the index of Λ') equivalence classes.

Theorem 6.2 *Let Λ and Λ' be S-lattices and S-sublattices, $\Lambda' \subseteq \Lambda$, $|\Lambda : \Lambda'| < \infty$ such that Λ/Λ' has nonzero annihilators. Then Λ/Λ' is the direct sum of a finite number of quotient sublattices,*

$$\Lambda/\Lambda' = \Lambda_{p_1}/\Lambda'_{p_1} \oplus \Lambda_{p_2}/\Lambda'_{p_2} \oplus \cdots \oplus \Lambda_{p_m}/\Lambda'_{p_m} \tag{6.29}$$

where $\Lambda_{p_i}/\Lambda'_{p_i} \triangleq \{\lambda \in \Lambda/\Lambda' : p_i^\gamma \lambda = 0\}$ for some $\gamma \geqslant 1$, and every p_i, $i = 1, 2, \cdots, m$ is a distinct prime over S.

Theorem 6.2 proves that Λ/Λ' can be decomposed into the direct sum of m sublattices $\Lambda_{p_i}/\Lambda'_{p_i}$ (the primary sublattices) which itself forms a new lattice system. Hence Λ/Λ' can be regarded as an m layer quotient lattice.

Theorem 6.3 *Every primary sublattice $\Lambda_{p_i}/\Lambda'_{p_i}$ is isomorphic to a direct sum of cyclic p_i-torsion modules:*

$$\Lambda_{p_i}/\Lambda'_{p_i} \cong S/\langle p_i^{\theta_1}\rangle \oplus S/\langle p_i^{\theta_2}\rangle \oplus \cdots \oplus S/\langle p_i^{\theta_t}\rangle \tag{6.30}$$

for some integers $1 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_t$ which are uniquely determined by $\Lambda_{p_i}/\Lambda'_{p_i}$.

Theorem 6.3 implies that the quotient primary S-sublattice system $\Lambda_{p_i}/\Lambda'_{p_i}$ is isomorphic to a cyclic p_i-torsion module. The right-hand side of (6.30) can be viewed as the message space of $\Lambda_{p_i}/\Lambda'_{p_i}$ which is detailed in Lemma 6.1.

Lemma 6.1 *There exists a map:*

$$\phi_i : \Lambda_{p_i} \longmapsto \bigoplus_j S/\langle p_i^{\theta_j} \rangle \tag{6.31}$$

which is a surjective S-module homomorphism with kernel $\mathcal{K}(\phi_i) = \Lambda'_{p_i}$. To ease the abstract representation, we consider $\Lambda'_{p_i} = \Lambda'$ in the sequel. Thus, $\mathcal{K}(\phi_i) = \Lambda'$ for $i = 1, 2, \cdots, m$. If the message space is taken as the canonical decomposition of (6.30), i.e. $\mathbf{w}^i = \bigoplus_j S/\langle p_i^{\theta_j} \rangle$, there exists a surjective homomorphism $\phi : (\Lambda; +, \cdot) \longmapsto \left(\mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m; \boxplus, \odot\right)$ and also an injective map $\tilde{\phi} : (\mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m) \longmapsto \Lambda$ such that

$$\phi(\tilde{\phi}(\mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m)) = \mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m \tag{6.32}$$

Lemma 6.2 *The generator matrix of the S-sublattice Λ_{p_i} at the i*th *layer can be expressed in the form of:*

$$\mathbf{G}_{\Lambda_{p_i}} = \begin{bmatrix} \mathrm{Diag}(\underbrace{\mathbf{p}_1^{\xi_1} \cdots \mathbf{p}_{i-1}^{\xi_{i-1}}, \mathbf{I}_t, \mathbf{p}_{i+1}^{\xi_{i+1}} \cdots \mathbf{p}_m^{\xi_m}}_{k}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \mathbf{G}_\Lambda \tag{6.33}$$

and

$$\phi_i(\mathbf{w}\mathbf{G}_{\Lambda_{p_i}}) = \left(w^{i,1} + \langle p_i^{\theta_1} \rangle, \cdots, w^{i,t} + \langle p_i^{\theta_t} \rangle\right) \tag{6.34}$$

where $w^{i,t} \in S/\langle p_i^{\theta_t} \rangle$ and $\mathbf{w} \in \mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m$. $\mathbf{G}_\Lambda$ is the generator matrix of the fine lattice Λ, $\mathbf{p}_j^{\xi_j}$, $j = 1, 2, \cdots, m$ is a vector, with all elements being the same elementary divisor $p_j^{\theta_j}$ over S, and $t = \dim(\Lambda_{p_i}/\Lambda')$.

Lemma 6.2 shows a way to produce the quotient S-sublattice of each layer defined in Theorem 6.2. Λ_{p_i}/Λ' forms an independent lattice system, and the direct sum of all Λ_{p_i}/Λ', $i = 1, 2, \cdots, m$ is equal to Λ/Λ'.

6.5.4 Elementary Divisor Construction

Theorems and Lemmas developed above provide good theoretical basis for creating a general lattice construction method, namely elementary divisor construction, which is also described in another form named construction π_D [13].

Lemma 6.3 *Let Λ and Λ' be S-lattices and S-sublattices, $\Lambda' \subseteq \Lambda$, $|\Lambda : \Lambda'| < \infty$ such that Λ/Λ' has a nonzero annihilator ϖ which can be uniquely factorised into distinct powers of primes in S, $\varpi = \mathcal{U}(S) p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_m^{\gamma_m}$. Then Λ/Λ' is the direct sum of a finite number of quotient sublattices, $\Lambda_{p_i}/\Lambda' = \{\lambda \in \Lambda/\Lambda' : p_i^{\gamma_i}\lambda = 0\}$, $i = 1, 2, \cdots, m$, and given by,*

$$\Lambda/\Lambda' = \Lambda_{p_1}/\Lambda' \oplus \Lambda_{p_2}/\Lambda' \oplus \cdots \oplus \Lambda_{p_m}/\Lambda' \tag{6.35}$$

Elementary Divisor Construction (EDC) Let $p_1, p_2, \cdots, p_m$ be some distinct primes in a PID S, and $\varpi = \mathcal{U}(S) p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_m^{\gamma_m}$ is a unique factorisation, $\gamma_i \geq 1$. Let $C^1, C^2, \cdots, C^m$ be m $[n, k_i]$ linear codes over $S/\langle p_1^{\gamma_1}\rangle$, $S/\langle p_m^{\gamma_m}\rangle, \cdots, S/\langle p_m^{\gamma_m}\rangle$, respectively. The elementary divisor construction lattice is defined by:

$$\Lambda \triangleq \{\lambda \in S^n : \tilde{\sigma}(\lambda) \in C^1 \oplus C^2 \oplus \cdots \oplus C^m\} \tag{6.36}$$

and the sublattice is:

$$\Lambda' \triangleq \{\varpi\lambda : \lambda \in S^n\}$$

where $\tilde{\sigma} : S^n \longmapsto (S/\langle p_1^{\gamma_1}\rangle)^n \oplus (S/\langle p_2^{\gamma_2}\rangle)^n \oplus \cdots \oplus (S/\langle p_m^{\gamma_m}\rangle)^n$ is a natural map obtained by extending the ring homomorphism $\sigma : S \longmapsto S/\langle p_1^{\gamma_1}\rangle \times S/\langle p_2^{\gamma_2}\rangle \times \cdots \times S/\langle p_m^{\gamma_m}\rangle$ to multiple dimensions. Apparently $\Lambda' \subseteq \Lambda$. The message space under EDC is

$$W = (S/\langle p_1^{\gamma_1}\rangle)^{k_1} \oplus \cdots \oplus (S/\langle p_m^{\gamma_m}\rangle)^{k_m} \tag{6.37}$$

where k_i is the message length of the ith layer which sums up to $k = \sum_{j=1}^{m} k_j$.

The elementary divisor construction is a straightforward extension of Lemma 6.5, which defines a class of lattices constructed by m linear codes, with each operating over either a finite field or a finite chain ring. Hence the quotient Λ/Λ' must consist of m primary sublattices Λ_{p_i}/Λ', with each constructed by the ith linear code. The primary sublattices Λ_{p_i} of the ith layer is defined by:

$$\Lambda_{p_i} \triangleq \{\lambda_{p_i} \in \delta_i S : \tilde{\sigma}_i(\lambda_{p_i}) \in C^i)\} \tag{6.38}$$

where $\tilde{\sigma}_i$ is a natural map:

$$\tilde{\sigma}_i : (\delta_i S)^n \longmapsto (\delta_i S/p_i^{\gamma_i}\delta_i S)^n \cong (S/\langle p_i^{\gamma_i}\rangle)^{k_i} \tag{6.39}$$

obtained by extending the ring homomorphism $\sigma_i : \delta_i S \longmapsto \delta_i S/\langle p_i^{\gamma_i}\delta_i S\rangle$ to multiple dimensions. The scaling factor $\delta_i = \frac{\varpi}{p_i^{\gamma_i}}$ can be proved in terms of the proof in Theorem 6.2.

We consider three scenarios based on different algebraic fields which the linear codes may belong to.

Scenario 1 Assume that the primary sublattice at each layer is constructed by a linear code over a finite field, thus, $\gamma_1 = \gamma_2 = \cdots = \gamma_m = 1$. Then, $C^i \in (\delta_i S/\langle p_i \delta_i \rangle)^n$. This group of lattices corresponds to the lattices constructed from Construction π_A in Sect. 6.5.2. Since the coarse lattice Λ' is generated by a single element ϖ, Λ/Λ' forms a cyclic torsion module which allows us to produce the generator matrix of the ith layer lattice Λ_{p_i}. It will have a form described in Lemma 6.2, given by:

$$\mathbf{G}_{\Lambda_{p_i}} = \begin{bmatrix} \mathrm{Diag}\left(\mathbf{p}_1^{(k_1)} \cdots \mathbf{p}_{i-1}^{(k_{i-1})}, \mathbf{I}_{k_i}, \mathbf{p}_{i+1}^{(k_{i+1})} \cdots \mathbf{p}_m^{(k_m)}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k} \end{bmatrix} \mathbf{G} \tag{6.40}$$

where $\mathbf{p}_i^{(k_i)}$ is a length-k_i vector with each element p_i. $\mathbf{G}_{\Lambda_{p_i}}$ in (6.40) gives the generator matrix for the ith layer lattices, when the message input

$$\mathbf{w} = [\mathbf{w}^1, \mathbf{w}^2, \cdots, \mathbf{w}^m, \underbrace{\tilde{d}_1 \cdots \tilde{d}_m}_{mn-k}] \tag{6.41}$$

where $\mathbf{w}^i \in (\delta_i S/\langle p_i \delta_i \rangle)^{k_i}$, $\tilde{d}_i \in S^{n-k_i}$.

Since EDC lattices are constructed by some linear codes, the matrix $\mathbf{G}$ must include the generator matrix of each linear code C^i. Let $\tilde{\sigma}_i([\mathbf{I}_{k_i} \ \ \mathbf{B}^i_{k_i \times (n-k_i)}])$ be a generator matrix for a linear code C^i (without loss of generality, we consider that the linear code is systematic in this case.), then $\mathbf{G}$ is an $mn \times n$ matrix defined below,

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{k_1} & \mathbf{B}^1_{k_1 \times (n-k_1)} \\ \mathbf{I}_{k_2} & \mathbf{B}^2_{k_2 \times (n-k_2)} \\ \vdots & \vdots \\ \mathbf{I}_{k_m} & \mathbf{B}^m_{k_m \times (n-k_m)} \\ \mathbf{0} & \varpi \mathbf{I}_{n-k_1} \\ \vdots & \vdots \\ \mathbf{0} & \varpi \mathbf{I}_{n-k_m} \end{bmatrix} \tag{6.42}$$

Equation (6.42) follows from Lemma 6.5 and part of the proof of Theorem 6.2. The generator matrix of the coarse lattice Λ' is therefore given by,

$$\mathbf{G}_{\Lambda'} = \begin{bmatrix} \mathrm{Diag}\left(\mathbf{I}_{\sum_{j=1}^{i-1} k_j}, \mathbf{p}_i^{(k_i)}, \mathbf{I}_{\sum_{j=i+1}^{m} k_j}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k} \end{bmatrix} \mathbf{G}_{\Lambda_{p_i}} \tag{6.43}$$

It can be easily observed that these generator matrices are consistent with the Theorems and Lemmas proposed earlier. Note that the generator matrix for linear code C^i is $\tilde{\sigma}_i([\mathbf{I}_{k_i} \ \ \mathbf{B}^i_{k_i \times (n-k_i)}])$ where $\tilde{\sigma}_i$ is defined in (6.39). Theorem 6.4 establishes the theoretic fundamental for low-complexity lattice decoding (i.e. layered integer

forcing) of MLNC, and states that there exists a surjective S-module homomorphism φ_i which satisfies Lemma 6.4, with kernel $\mathcal{K}(\varphi_i) = \Lambda'_i$, which plays a key role in decoding the ith layer linearly combined messages. Its generator matrix has a form:

$$\mathbf{G}_{\Lambda'_i} = \begin{bmatrix} \mathrm{Diag}(\underbrace{\mathbf{I}, \mathbf{p}_i^{(k_i)}, \mathbf{I}}_{k}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k} \end{bmatrix} \mathbf{G} \tag{6.44}$$

We can easily verify $\Lambda/\Lambda'_i \cong (S/\langle p_i \rangle)^{k_i}$ in terms of these generator matrices.

Figure 6.2 depicts the structure of a 2-layer EDC lattice based on *scenario* 1 and Eisenstein integers, where $p_1 = 1 + 2\omega$ and $p_2 = 2$. The primary sublattice of layer-1 can be represented as $\lambda_{p_1} \in \delta_1 S = 2S$ and these sublattices are marked as green points in Fig. 6.2a. It is clear that in this case $C^1 \in (2S/\langle 2(1+2\omega)\rangle)^n \cong \mathbb{F}_3^{k_1}$. Similarly, the primary sublattice of layer-2 can be represented as $\lambda_{p_2} \in \delta_2 S = (1+2\omega)S$ and in this case $C^2 \in ((1+2\omega)S)/\langle 2(1+2\omega)\rangle)^n \cong \mathbb{F}_{2^2}^{k_2}$. The Voronoi region of Λ'_i for the ith layer is illustrated by the blue line. The red line represents the Voronoi region of coarse lattice Λ'.

Scenario 2 When $\forall i = 1, 2, \cdots, m$, $\gamma_i \neq 1$, the primary sublattice Λ_{p_i} at each layer is constructed by a linear code over a finite chain ring $T = \delta_i S/\langle p_i^{\gamma_i} \delta_i \rangle$ [10]. A finite chain ring is a finite local principal ideal ring, and the most remarkable characteristic of a finite chain ring is that its every ideal (including $\langle 0 \rangle$) is generated by the maximal ideal, which can be linearly ordered by inclusion, and hence, forms a chain. The finite chain ring T has a unique maximal ideal and hence the resultant residue field is $Q = \delta_i S/\langle p_i \delta_i \rangle$ with size $q = |\delta_i S/\langle p_i \delta_i \rangle|$. The chain length of the ideals is indeed the nil-potency index of p_i which is, in this case γ_i. We refer to T a (q, γ_i) chain ring.

At the ith layer, the generator matrix $\mathbf{G}^i_{\mathrm{FCR}}$ of a linear code over T has a standard form given in (6.45), where $\mathbf{I}_{k'_{i,t}}$ denotes an identity matrix with dimension $k'_{i,t}$,[1] $i = 1, 2, \cdots, m$ and $t = 0, 1, \cdots, \gamma_i - 1$. Hence $\mathbf{G}^i_{\mathrm{FCR}}$ has a dimension $k'_i \times n$ where $k'_i = \sum_{t=0}^{\gamma_i - 1} k'_{i,t}$. Here $Z_{t,l}$, $l = t+1, 2, \cdots, \gamma_i$, denotes a $k'_{i,t} \times k'_{i,t+1}$ ($k'_{i,\gamma_i} = n - k'_i$) matrix which is unique modulo $p_i^{\gamma_i - t}$ [19]. In (6.45), $\mathcal{I}^*_{p_i^{\gamma_i}}$ is an upper triangular matrix with dimension $k'_i \times k'_i$, and $\mathcal{B}_{k'_i, n-k'_i}$ has a dimension of $k'_i \times (n - k'_i)$. Note that the codeword is row spanned by $\mathbf{G}^i_{\mathrm{FCR}}$ and all rows of $\mathbf{G}^i_{\mathrm{FCR}}$ are linearly independent.

To study the message space of the linear codes over the finite chain ring, we first examine the kernel of the generator matrix $\mathbf{G}^i_{\mathrm{FCR}}$. This is equivalent to finding the null space for the encoder $\mathcal{E}^i : \mathbf{w}^i \longmapsto C^i$, where $\mathcal{E}^i(\mathbf{w}^i) \triangleq \mathbf{w}^i \mathbf{G}^i_{\mathrm{FCH}}$ and $\mathbf{w}^i = [\mathbf{w}_{k'_{i,0}}, \mathbf{w}_{k'_{i,1}}, \cdots, \mathbf{w}_{k'_{i,\gamma_i - 1}}]$. Here $\mathbf{w}^i$ is grouped into blocks of size $\mathbf{w}_{k'_{i,t}}$ which

[1] Here, the index i used in $k'_{i,t}$ is the indicator of layer.

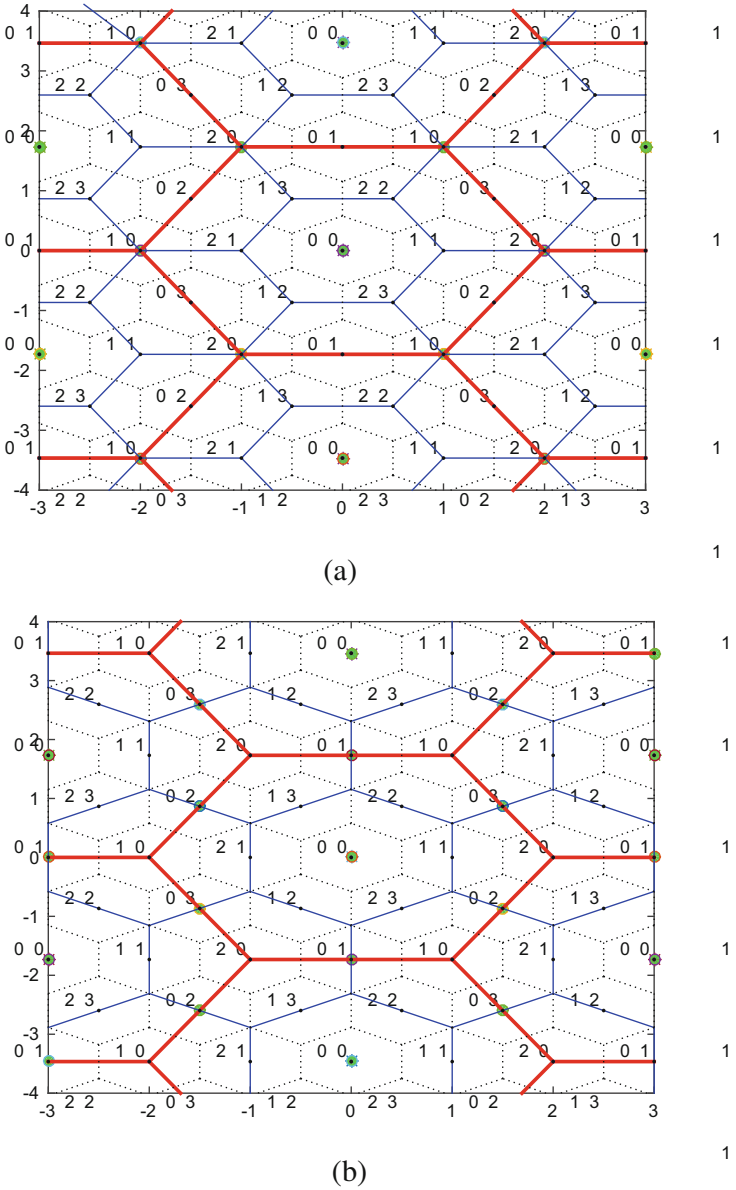


Fig. 6.2 Layer structure of a 2-layer EDC lattice. The green points and blue lines represent the primary sublattices and Voronoi region of $\mathcal{V}_{\Lambda_i'}$ for the corresponding layers, respectively. Dotted lines represent the Voronoi region of the fine lattice. (**a**) Layer 1. (**b**) Layer 2

corresponds to the row blocks defined in (6.45). In order to obtain the all-zero codeword $C^i = \mathbf{0}$, we solve the homogeneous system $\mathbf{w}^i \mathbf{G}^i_{\text{FCH}} = \mathbf{0}$, which gives $\mathbf{w}_{k'_{i,t}} \in p_i^{\gamma_i - t} T^{k'_{i,t}}$, $t = 0, 1, \cdots, \gamma_i - 1$. This result is based on the fact that if $d \in T^n$, then $p_i^t d = 0 \Longrightarrow d \in p_i^{\gamma_i - t} T^n$. The null space of the encoder $\mathcal{E}^i$ is therefore:

$$\mathbf{G}^i_{\text{FCR}} = \left[\begin{array}{ccccc|c} \mathbf{I}_{k'_{i,0}} & Z^i_{0,1} & Z^i_{0,2} & \cdots & Z^i_{0,\gamma_{i-1}} & Z^i_{0,\gamma_i} \\ 0 & p_i \mathbf{I}_{k'_{i,1}} & p_i Z^i_{1,2} & \cdots & p_i Z^i_{1,\gamma_{i-1}} & p_i Z^i_{1,\gamma_i} \\ 0 & 0 & p_i^2 \mathbf{I}_{k'_{i,2}} & \cdots & p_i^2 Z^i_{2,\gamma_{i-1}} & p_i^2 Z^i_{2,\gamma_i} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_i^{\gamma_i - 1} \mathbf{I}_{k'_{i,\gamma_i - 1}} & p_i^{\gamma_i - 1} Z^i_{\gamma_i - 1, \gamma_i} \end{array}\right] = \left[\mathcal{I}^*_{p_i^{\gamma_i}} \middle| \mathcal{B}_{k'_i, n-k'_i}\right] \tag{6.45}$$

$$\mathbf{w}' = [p_i^{\gamma_i} T^{k'_{i,0}}, \cdots, p_i T^{k'_{i,\gamma_i - 1}}] \tag{6.46}$$

According to the first isomorphism theorem, the codeword C^i is isomorphic to a direct summation:

$$\begin{aligned} C^i &\cong (T/p_i^{\gamma_i} T)^{k'_{i,0}} \oplus (T/p_i^{\gamma_i - 1} T)^{k'_{i,1}} \oplus \cdots \oplus (T/p_i T)^{k'_{i,\gamma_i - 1}} \\ &\cong (\delta_i S/\langle p_i^{\gamma_i} \delta_i \rangle)^{k'_{i,0}} \oplus (\delta_i S/\langle p_i^{\gamma_i - 1} \delta_i \rangle)^{k'_{i,1}} \oplus \cdots \oplus (\delta_i S/\langle p_i \delta_i \rangle)^{k'_{i,\gamma_i - 1}} \end{aligned} \tag{6.47}$$

The right-hand side of (6.47) denotes the message space $\boldsymbol{W}^i$ of the linear code over the finite chain ring T in terms of the generator matrix $\mathbf{G}^i_{\text{FCR}}$. Note that each component in the direct sum of (6.47) forms another module or vector space, and the size of the tth component is $q^{(\gamma_i - t)k'_{i,t}}$. This leads to the overall message size $|C| = q^{\sum_{t=0}^{\gamma_i - 1} (\gamma_i - t) k'_{i,t}}$. Of course, we can obtain this result directly from the kernel of $\mathbf{G}^i_{\text{FCR}}$, thus, $|C| = \prod_{t=0}^{\gamma_i - 1} (p_i^t T)^{k'_{i,t}}$ which gives the same result.

Let $\tilde{\boldsymbol{p}}_i^{\gamma_i}$ be a length-k'_i vector:

$$\tilde{\boldsymbol{p}}_i^{\gamma_i} \triangleq [\mathbf{p}^{\gamma_i}_{i,(k'_{i,0})}, \mathbf{p}^{\gamma_i - 1}_{i,(k'_{i,1})}, \cdots, \mathbf{p}_{i,(k'_{i,\gamma_i - 1})}]$$

where $\mathbf{p}^{\gamma_i}_{i,(k'_{i,0})}$ denotes a length-$k'_{i,0}$ vector, with each component being $p_i^{\gamma_i}$. Note that $\tilde{\boldsymbol{p}}_i^{\gamma_i}$ is closely related to (6.46). Following Lemma 6.2, the generator matrix of the primary sublattice Λ_{p_i} of the ith layer in this scenario has a form:

$$\mathbf{G}_{\Lambda_{p_i}} = \left[\begin{array}{cc} \text{Diag}\left(\tilde{\boldsymbol{p}}_1^{\gamma_1} \cdots \tilde{\boldsymbol{p}}_{i-1}^{\gamma_{i-1}}, \mathbf{I}_{k'_i}, \tilde{\boldsymbol{p}}_{i+1}^{\gamma_{i+1}} \cdots \tilde{\boldsymbol{p}}_m^{\gamma_m}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k'} \end{array}\right] \mathbf{G} \tag{6.48}$$

where $k' = \sum_{i=1}^{m} k'_i$. The EDC lattices in this scenario are constructed by some linear codes over different finite chain rings, and the matrix $\mathbf{G}$ must be associated with the generator matrix of each linear code C^i over the finite chain ring. Let $\tilde{\sigma}_i(\mathbf{d} \cdot [\tilde{\mathcal{I}}^*_{p_i^{\gamma_i}} \quad \tilde{\mathcal{B}}_{k'_i,n-k'_i}])$ be the codeword of $C^i = \mathbf{w}^i \mathbf{G}^i_{\mathrm{FCR}}$ over the finite chain ring T, $\mathbf{d} \in \delta_i S^{k'_i}$. Then, $\mathbf{G}$ in (6.48) is an $mn \times n$ matrix defined below:

$$\mathbf{G} = \begin{bmatrix} \tilde{\mathcal{I}}^*_{p_1^{\gamma_1}} & \tilde{\mathcal{B}}_{k'_1,n-k'_1} \\ \tilde{\mathcal{I}}^*_{p_2^{\gamma_2}} & \tilde{\mathcal{B}}_{k'_2,n-k'_2} \\ \vdots & \vdots \\ \tilde{\mathcal{I}}^*_{p_m^{\gamma_m}} & \tilde{\mathcal{B}}_{k'_m,n-k'_m} \\ \mathbf{0} & \varpi \mathbf{I}_{n-k'_1} \\ \vdots & \vdots \\ \mathbf{0} & \varpi \mathbf{I}_{n-k'_m} \end{bmatrix} \tag{6.49}$$

Hence, we are able to construct Λ_{p_i} and hence the EDC lattice Λ for this scenario based on the generator matrices presented above. Note that message space of each layer follows from (6.47), and $k'_{i,t}$ should be selected such that

$$\gamma_i k_i = \sum_{t=0}^{\gamma_i - 1} (\gamma_i - t) k'_{i,t} \tag{6.50}$$

in order to guarantee the consistency to the message size of the ith layer EDC lattices defined in (6.37). It is easy to prove that there exists $k'_{i,t} \in \mathbb{Z}^+, \forall t = 0, 1, \cdots, \gamma_i - 1$, satisfying (6.50).

The generator matrix of the coarse lattice Λ' is given by,

$$\mathbf{G}_{\Lambda'} = \begin{bmatrix} \mathrm{Diag}\left(\mathbf{I}_{\sum_{j=1}^{i-1} k'_j}, \tilde{\boldsymbol{p}}_i^{\gamma_i}, \mathbf{I}_{\sum_{j=i+1}^{m} k'_j}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k} \end{bmatrix} \mathbf{G}_{\Lambda_{p_i}} \tag{6.51}$$

Following (6.51), it is obvious that $\Lambda/\Lambda' \cong \boldsymbol{W}^1 \oplus \cdots \oplus \boldsymbol{W}^m$. The generator matrix for Λ'_i has a form:

$$\mathbf{G}_{\Lambda'_i} = \begin{bmatrix} \mathrm{Diag}\left(\mathbf{I}_{\sum_{j=1}^{i-1} k'_j}, \tilde{\boldsymbol{p}}_i^{\gamma_i}, \mathbf{I}_{\sum_{j=i+1}^{m} k'_j}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{mn-k} \end{bmatrix} \mathbf{G} \tag{6.52}$$

which will be used for layered integer forcing detection.

Every ideal of T is generated by the maximal ideal, which forms a chain with chain length γ_i. Hence the residue field Q plays an important role in producing the linear codes over T. We now consider a matrix in the form of:

$$\mathbf{G}_{\mathrm{D}}^{i} = \mathrm{Diag}\left(\mathbf{p}_{i,(k'_{i,0})}^{0}, \cdots, \mathbf{p}_{i,(k'_{i,\gamma_i-1})}^{\gamma_i-1}\right) \begin{bmatrix} \mathbf{g}_{k'_{i,0}}^{i} \\ \mathbf{g}_{k'_{i,1}}^{i} \\ \vdots \\ \mathbf{g}_{k'_{i,\gamma_i-1}}^{i} \end{bmatrix} \tag{6.53}$$

where $\mathbf{g}_{k'_{i,t}}^{i} \in Q_{k'_{i,t}\times n}^{*}$, and $Q_{k'_{i,t}\times n}^{*}$ is a $k'_{i,t} \times n$ matrix with each entry over the coset representative of the residue field $Q = \delta_i S/\langle p_i \delta_i \rangle$. Each row of $\mathbf{G}_{\mathrm{D}}^{i}$ must satisfy the condition that none of its rows are linear combinations of the other rows. The message space of $\mathbf{G}_{\mathrm{D}}^{i}$ could be partitioned into $\gamma_i - 1$ levels. We first define the vector $\boldsymbol{\beta}_{k'_{i,t}}^{(j)} = [\beta_1^{(j)}, \beta_2^{(j)}, \cdots, \beta_{k'_{i,t}}^{(j)}]$, when $t = 0$, where $j = 0, 1, \cdots \gamma_i - 1$, is the level indicator, and $\boldsymbol{\beta}_{k'_{i,t}}^{(j)} = [\beta_{k'_{i,t-1}+1}^{(j)}, \beta_{k'_{i,t-1}+2}^{(j)}, \cdots, \beta_{k'_{i,t}}^{(j)}]$ when $t = 1, 2, \cdots, \gamma_i - 1$. Accurately $\boldsymbol{\beta}_{k'_{i,t}}^{(j)}$ represents a length-$k'_{i,t}$ segment of the jth level message over the vector space $Q^{k'_{i,t}}$. The full message space of the jth level is given by,

$$\boldsymbol{\beta}^{(j)} = [p_i^j \boldsymbol{\beta}_{k'_{i,0}}^{(j)}, p_i^{j-1} \boldsymbol{\beta}_{k'_{i,1}}^{(j)}, p_i^{j-2} \boldsymbol{\beta}_{k'_{i,2}}^{(j)}, \underbrace{\mathbf{0}\cdots\mathbf{0}}_{k'_i - \sum_{t=0}^{j} k'_{i,t}}] \tag{6.54}$$

where the powers of p_i can not be negative integers. Hence the message space of $\mathbf{G}_{\mathrm{D}}^{i}$ is $W^i = \boldsymbol{\beta}^{(0)} + \boldsymbol{\beta}^{(1)} + \cdots + \boldsymbol{\beta}^{(\gamma_i-1)}$. The codewords C^i can be produced by

$$\begin{aligned} C^i = W^i \mathbf{G}_{\mathrm{D}}^{i} &= \left(\boldsymbol{\beta}^{(0)} + \boldsymbol{\beta}^{(1)} + \cdots + \boldsymbol{\beta}^{(\gamma_i-1)}\right) \mathbf{G}_{\mathrm{D}}^{i} \\ &= \mathbf{c}_0^i + \mathbf{c}_1^i p_i + \cdots + \mathbf{c}_{\gamma_i-1}^i p_i^{\gamma_i-1} \end{aligned} \tag{6.55}$$

Since none of the rows of $\mathbf{G}_{\mathrm{D}}^{i}$ are linear combinations of the other rows, $\mathbf{c}_t^i$ is therefore row spanned by

$$\mathbf{g}_{\mathbf{c}_t^i} = \left[\mathbf{g}_{k'_{i,0}}^{i}; \mathbf{g}_{k'_{i,1}}^{i}; \cdots; \mathbf{g}_{k'_{i,t}}^{i}\right] \tag{6.56}$$

It is obvious that $\mathbf{c}_t^i$, $t = 0, 1, \cdots, \gamma_i - 1$ forms a set of nested codes $\mathbf{c}_0^i \subseteq \mathbf{c}_1^i \subseteq \cdots \subseteq \mathbf{c}_{\gamma_i-1}^i$ over Q^*. Following the Q-adic decompostion theorem of finite chain ring [10, 19], we assert that the codeword C^i in (6.55) generated by $\mathbf{G}_{\mathrm{D}}^{i}$ is indeed over T.

In terms of (6.53) and (6.54), the message space corresponding to $\mathbf{g}^i_{k'_{i,t}}$ should be written as:

$$W^i_t = \sum_{j=t}^{\gamma_i - 1} p_i^{j-t} \boldsymbol{\beta}^{(j)}_{k'_{i,t}} \tag{6.57}$$

this complies with the Q-adic decomposition and leads to the result that the message space corresponding to $\mathbf{g}^i_{k'_{i,t}}$ is $(T/\langle p_i^{\gamma_i - t}\rangle)^{k'_{i,t}}$. This implies that the right-hand side of (6.47) is precisely the message space of $\mathbf{G}^i_{\mathrm{D}}$. Mathematically the primary sublattices Λ_{p_i} can also be represented in the form below:

$$\Lambda_{p_i} = \bigcup \left\{ \underbrace{\sum_{j=0}^{\gamma_i - 1} \sum_{\ell=1}^{\mathscr{K}^i_j} p_i^j \beta^{(j)}_\ell \mathbf{g}^i_\ell + p_i^{\gamma_i} S^n}_{(52)} | \mathbf{g}^i_\ell \in Q_{1\times n} \right\} \tag{6.58}$$

where $\mathscr{K}^i_j = k'_{i,0} + \cdots + k'_{i,j}$. It is interesting to see that (6.58) has the same structure as complex construction D. Now we conclude that the primary S-sublattices constructed by a linear code over a finite chain ring subsumes construction D.

Based on this result, we may now construct EDC lattices for this scenario using a set of nested linear codes over a finite field. Let $\mathbf{g}^i_{(n-k'_i)} \in Q^*_{n-k'_i \times n}$ be an $(n-k'_i)\times n$ matrix, then the $\mathbf{G}$ matrix is:

$$\mathbf{G} = \left[{\mathbf{G}^1_{\mathrm{D}}}^T, \cdots, {\mathbf{G}^m_{\mathrm{D}}}^T, \varpi {\mathbf{g}^1_{(n-k'_i)}}^T, \cdots, \varpi {\mathbf{g}^m_{(n-k'_i)}}^T \right]^T \tag{6.59}$$

Scenario 3 This corresponds to a hybrid case of scenario 1 and 2, and we give the following summaries:

1. $m = 1, \gamma_1 = 1$, then the EDC lattice in (6.36) is a complex construction A lattice which is indecomposable.
2. $m = 1, \gamma_1 > 1, \gamma_1 \in \mathbb{Z}^+$ then the EDC lattice in (6.36) is a complex construction D lattice which is indecomposable.
3. $m > 1, \gamma_i \in \mathbb{Z}^+, i = 1, 2, \cdots, m$, then the EDC lattice in (6.36) is decomposable, and consists of some sublattices constructed by either construction A or D.

Note that in (3), a new class of lattices over S is generated by a number of linear codes over either finite field or chain ring, which generalises the scenario 1 and 2. Scenario 3 suggests that the design of EDC lattices is very flexible, and we also give more detailed discussion about why EDC lattices are good at low-complexity decoding and throughput improvement for PNC in the next sections.

6.5.5 Nominal Coding Gain and Kissing Number

The nominal coding gain and kissing number of the EDC lattices are described in this section for all three scenarios. The definition such as the minimum-norm coset leaders and minimum Euclidean weight of the codeword follows from [9].

Scenario 1 We first study the nominal coding gain and kissing number of the ith layer primary sublattices in this scenario. Following (6.38) and (6.39), we know that C^i is a linear code of length n over $\delta_i S/p_i\delta_i S$. Thus, $\mathbf{c}^i = (c_1^i + \langle\varpi\rangle, \cdots, c_n^i + \langle\varpi\rangle) \in C^i$. We denote $\omega^{(i)}(\mathbf{c}^i)$ the Euclidean weight of a codeword $\mathbf{c}^i$ in C^i, and $\omega_{\min}^{(i)}(C^i)$ the minimum Euclidean weight of non-zero codewords in C^i. Let ϑ be a scaling factor depending on which PID is used, and $N(\omega_{\min}^{(i)}(C^i))$ be the number of codewords in C^i with the minimum Euclidean weight $\omega_{\min}^{(i)}(C^i)$.

Proposition 6.1 *Let C^i be a linear code over $\delta_i S/p_i\delta_i S$, and Λ_{p_i}/Λ' the primary quotient lattice system of the ith layer constructed by C^i, $\Lambda_{p_i} \supseteq \Lambda'$, then the nominal coding gain is given by:*

$$\varrho(\Lambda_{p_i}/\Lambda') = \frac{\omega_{\min}^{(i)}(C^i)}{\vartheta |p_i|^{2(1-\frac{k_i}{n})}|\delta_i|^2} \tag{6.60}$$

and the kissing number is:

$$K(\Lambda_{p_i}/\Lambda') = \begin{cases} N(\omega_{\min}^{(i)}(C^i))\left(\frac{\mathcal{N}_{\mathcal{U}(S)}}{|p_i|^2-1}\right)^{\frac{\omega_{\min}^{(i)}(C^i)}{|\delta_i|^2}}, & |p_i|^2 - 1 \le \mathcal{N}_{\mathcal{U}(S)} \\ N(\omega_{\min}^{(i)}(C^i)), & \text{Otherwise} \end{cases} \tag{6.61}$$

Here $\mathcal{N}_{\mathcal{U}(S)}$ represents the number of units in S.

It is of interest to study the nominal coding gain and kissing number of Λ/Λ' in terms of the m linear codes C^i. Following the proof of Theorem 6.2, and the descriptions in prior sections, $\tilde{\mathbf{c}} = \mathbf{c}^1 + \mathbf{c}^2 + \cdots + \mathbf{c}^m$, $\tilde{\mathbf{c}} \in \tilde{C}$ and $\tilde{C} \in (S/\langle\varpi\rangle)^n$. Thus, the nominal coding gain of EDC lattices is determined by the m linear codes C^i over $\delta_i S/p_i\delta_i S$, $i = 1, 2, \cdots, m$.

Proposition 6.2 *Let $C^1, \cdots, C^m$ be m linear codes over $\delta_i S/p_i\delta_i S$, $i = 1, 2, \cdots, m$, respectively. Let $\tilde{\mathbf{c}} = \mathbf{c}^1 + \mathbf{c}^2 + \cdots + \mathbf{c}^m$, $\tilde{\mathbf{c}} \in \tilde{C}$ and $\mathbf{c}^i \in C^i$. The nominal coding gain of the EDC lattices Λ/Λ' in scenario 1 is given by*

$$\varrho(\Lambda/\Lambda') = \frac{\omega_{\min}(\tilde{C})\prod_{\ell=2}^{m}|p_j|^{\frac{2(k_\ell-k_1)}{n}}}{\vartheta |p_1|^{2(1-\frac{k_1}{n})}|\delta_1|^2} \tag{6.62}$$

where $k_1 \le k_2 \le \cdots \le k_m$.

Scenario 2 This corresponds to the case where $\gamma_i > 1, \gamma_i \in \mathbb{Z}$ for $i = 1, 2, \cdots, m$. The primary sublattice of the ith layer can be constructed by a linear code C^i over a finite chain ring $\delta_i S/\langle \varpi \rangle$, where $\delta_i = \frac{\varpi}{p_i^{\gamma_i}}$. This follows immediately from (6.38) and (6.39). Here, we are more concerned with the nominal coding gain and kissing number when the ith primary sublattice is constructed by a set of nested linear codes over the residue field Q, since the linear code over a finite field is easier to generate. Let $C^{i,0} \subseteq \cdots \subseteq C^{i,\gamma_i - 1}$ be nested linear codes of length-n over Q, where $C^{i,t}$ is an $[n, \sum_{\ell=0}^{t} k'_{i,\ell}]$ linear code for the tth nested code at the ith layer, and we denote $\omega_{\min}^{(i,t)}(C^{i,t})$ the minimum Euclidean weight of non-zero codewords in $C^{i,t}$. We have:

Proposition 6.3 *Let $C^{i,0} \subseteq \cdots \subseteq C^{i,\gamma_i - 1}$ be γ_i nested linear codes of length-n over Q, and Λ_{p_i}/Λ' be the primary quotient lattice of the i*th *layer constructed from $C^{i,t}$, $t = 0, 1, \cdots, \gamma_i - 1$, then the nominal coding gain of the i*th *layer is lower bounded by*

$$\varrho(\Lambda_{p_i}/\Lambda') \geq \frac{|p_i|^{\frac{2}{n}\sum_{t=0}^{\gamma_i-1}(\gamma_i - t)k'_{i,t}} \min_{0 \leq t \leq \gamma_i - 1}\{|p_i|^{2t}\omega_{\min}^{(i,t)}(C^{i,t})\}}{\vartheta |\varpi|^2} \tag{6.63}$$

and the kissing number is upper bounded by:

$$K(\Lambda_{p_i}/\Lambda') \leq \begin{cases} \sum_{t=0}^{\gamma_i - 1} N_t(\omega_{\min}^{(i,t)}(C^{i,t})) \left(\frac{\mathcal{N}_{\mathcal{U}(S)}}{|p_i|^2 - 1}\right)^{\frac{\omega_{\min}^{(i,t)}(C^{i,t})}{|\delta_i|^2}}, & |p_i|^2 - 1 \leq \mathcal{N}_{\mathcal{U}(S)} \\ \sum_{t=0}^{\gamma_i - 1} N_t(\omega_{\min}^{(i,t)}(C^{i,t})), & \text{Otherwise} \end{cases} \tag{6.64}$$

It is of interest to study the nominal coding gain of Λ/Λ' in this scenario. If each primary sublattice is constructed via a set of nested linear codes over a finite field $Q = \delta_i S/\langle p_i \delta_i \rangle$ for the ith layer, the nominal coding gain $\varrho(\Lambda/\Lambda')$ will be related to overall $\sum_{i=1}^{m} \gamma_i$ linear codes since there are γ_i nested linear codes for each i. Let $\tilde{C}$ be a composite code such that $\tilde{\mathbf{c}} = \mathbf{c}^1 + \cdots + \mathbf{c}^m$ where $\mathbf{c}^i = \mathbf{c}^{i,0} + p_i \mathbf{c}^{i,1} + \cdots + p_i^{\gamma_i - 1}\mathbf{c}^{i,\gamma_i - 1}$. Hence $C^i \in \delta_i S/\langle \varpi \rangle$ and $\tilde{C} \in S/\langle \varpi \rangle$. We denote $\omega_{\min}(\tilde{C})$ the minimum Euclidean weight of non-zero codewords in $\tilde{C}$, then:

Proposition 6.4 *Let $C^{i,0} \subseteq \cdots \subseteq C^{i,\gamma_i - 1}$ be γ_i nested linear codes of length-n over Q, and let $\tilde{C}$ be a composite code such that $\tilde{\mathbf{c}} = \mathbf{c}^1 + \cdots + \mathbf{c}^m$ where $\mathbf{c}^i = \mathbf{c}^{i,0} + p_i \mathbf{c}^{i,1} + \cdots + p_i^{\gamma_i - 1}\mathbf{c}^{i,\gamma_i - 1}$. The nominal coding gain for Λ/Λ' in scenario 2 is given by:*

$$\begin{aligned} \varrho(\Lambda/\Lambda') &= \frac{\omega_{\min}(\tilde{C})}{(V(\mathcal{V}(\Lambda)))^{\frac{1}{n}}} \\ &= \frac{\omega_{\min}(\tilde{C}) \prod_{i=1}^{m} |p_i|^{2\sum_{t=0}^{\gamma_i - 1}(\gamma_i - t)\frac{k'_{i,t}}{n}}}{\vartheta |\varpi|^2} \end{aligned} \tag{6.65}$$

Scenario 3 As explained in the preceding section, in this case, $\gamma_i \geq 1$, $\gamma_i \in \mathbb{Z}$, and hence the EDC lattice consists of a number of primary sublattices which can be constructed by linear codes over either finite field or finite chain ring. The nominal coding gain and kissing number of the primary sublattices in each case have been derived in Propositions 6.1 and 6.3. We are more interested in the nominal coding gain of Λ/Λ' in this scenario. Again, we consider the primary sublattices of scenario 2 is constructed over a set of nested linear codes. Let $\tilde{C}$ be a composite code such that $\tilde{\mathbf{c}} = \mathbf{c}^1 + \cdots + \mathbf{c}^m$ where

$$\mathbf{c}^i = \begin{cases} \mathbf{c}^i, \ \ C^i \in \delta_i S/p_i\delta_i S, & \gamma_i = 1 \\ \mathbf{c}^{i,0} + p_i\mathbf{c}^{i,1} + \cdots + p_i^{\gamma_i-1}\mathbf{c}^{i,\gamma_i-1}; \ \ C^{i,t} \in Q, & \gamma_i > 1 \end{cases}$$

We can easily prove that $\varrho(\Lambda/\Lambda')$ has similar form as (6.65) if we set $k'_{i,0} = k_i$ for $\gamma_i = 1$.

6.6 Multistage Compute-and-Forward Over Finite Rings

In this section, we discuss the extension of the schemes in Sect. 6.4 to the multilevel setting. We first present multistage compute-and-forward [14], a generalization of compute-and-forward in [21]. We then discuss multilevel lattice network coding, a generalization of lattice network coding in [8]. Similar to their single-level counterparts in Sect. 6.4, while multilevel lattice network coding is a general framework and subsumes multistage compute-and-forward as a special case, the latter specifically narrows the codes to be those from Construction π_A lattices in Sect. 6.5.2 and thus the achievable computation rate can be analyzed.

6.6.1 Multistage Compute-and-Forward

In this subsection, we consider only the ring of integers $\mathbb{Z}$ and the real channel coefficients (where the complex channel coefficients will be discussed in the subsequent sections), i.e., $h_{j\ell} \in \mathbb{R}$. The results for the complex case can be similarly obtained by considering either $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$, or other rings of imaginary quadratic integers as the underlying ring of integers. In what follows, similar to compute-and-forward [21], we consider the asymptotic regime where we would like to know under which rates, the probability of error would vanish as the blocklength becomes large.

Let $p_1, \ldots, p_m$ be distinct primes. From the Chinese Remainder Theorem, there exists $\mathcal{M} : \times_{i=1}^m \mathbb{F}_{p_i} \to \mathbb{Z}/\Pi_{i=1}^m p_i\mathbb{Z}$ a ring isomorphism, which can be easily obtained by solving the Bézout's identity. The key enabler of our multistage compute-and-forward is to recognize the fact that from CRT, each integer $a_{j\ell} \in \mathbb{Z}$

can be uniquely represented as

$$a_{j\ell} = \bar{a}_{j\ell} + \Pi_{i=1}^{m} p_i \tilde{a}_{j\ell}, \tag{6.66}$$

with $\tilde{a}_{j\ell} \in \mathbb{Z}$ and $\bar{a}_{j\ell} \in \mathbb{Z}/\Pi_{i=1}^{m} p_i\mathbb{Z}$ which itself can be uniquely represented by its coordinate in $\times_{i=1}^{m} \mathbb{F}_{p_i}$ as

$$\bar{a}_{j\ell} = \mathcal{M}(b_{j\ell}^1, \ldots, b_{j\ell}^m). \tag{6.67}$$

With the above relationship, we can collectively write $\mathbf{a}_j = \bar{\mathbf{a}}_j + \Pi_{i=1}^{m} p_i \tilde{\mathbf{a}}_j$ where $\bar{\mathbf{a}}_j = \mathcal{M}(\mathbf{b}_j^1, \ldots, \mathbf{b}_j^m)$. In our proposed scheme, each transmitter decomposes the message w_ℓ into m sub-messages and represent each sub-message by its p_i-ary expansion $\mathbf{w}_j^i \in \mathbb{F}_{p_i}^{N'}$ for $i \in \{1, \ldots, m\}$. The functions we aim to compute at the relay j are given by

$$\mathbf{u}_j^i \triangleq b_{j1}^i \odot \mathbf{w}_1^i \oplus \ldots \oplus b_{j\ell}^i \odot \mathbf{w}_L^i, \tag{6.68}$$

for $i \in \{1, \ldots, m\}$.

The proposed multistage compute-and-forward scheme is based on the multilevel nested lattice codes from Construction π_A lattices recently proposed in [13]. To use the proposed multilevel lattices for transmission over a power-constrained system, a multilevel nested lattice code construction is proposed in [13], which tailors the nested lattice code construction of Ordentlich and Erez [22] specifically for Construction π_A lattices. In this construction, two lattices, namely the coarse and fine lattices, are constructed in such a way that the coarse lattice is a sub-lattice of the fine one. The code then consists of all the fine lattice points lying inside the fundamental Voronoi region of the coarse lattice. To this end, we first construct two sets of linear codes $C_f^1, \ldots, C_f^m$ and $C_c^1, \ldots, C_c^m$ that will later be used for constructing the fine and coarse lattices, respectively. Specifically, let

$$C_c^i = \{\mathbf{G}_c^i \odot \mathbf{w}^i | \mathbf{w}^i \in \mathbb{F}_{p_i}^{m_c^i}\}, \tag{6.69}$$

$$C_f^i = \{\mathbf{G}_f^i \odot \mathbf{w}^i | \mathbf{w}^i \in \mathbb{F}_{p_i}^{m_f^i}\}, \tag{6.70}$$

where $\mathbf{G}_c^i$ is a $n \times m_c^i$ matrix and $\mathbf{G}_f^i = \left[\mathbf{G}_c^i \ \tilde{\mathbf{G}}^i\right]$, where $\tilde{\mathbf{G}}^i$ is a $n \times (m_f^i - m_c^i)$ matrix. We then use Construction π_A with these codes to generate

$$\begin{aligned} \Lambda_f &\triangleq \gamma \left(\Pi_{i=1}^{m} p_i\right)^{-1} \mathcal{M}(C_f^1, \ldots, C_f^m) + \gamma\mathbb{Z}^n, \\ \Lambda_c &\triangleq \gamma \left(\Pi_{i=1}^{m} p_i\right)^{-1} \mathcal{M}(C_c^1, \ldots, C_c^m) + \gamma\mathbb{Z}^n, \end{aligned} \tag{6.71}$$

where γ is for the power constraint. Clearly, $C_c^i \subset C_f^i$, $i \in \{1, \ldots, m\}$ and thus $\Lambda_c \subset \Lambda_f$. We then form the nested lattice code corresponding to Λ_f/Λ_c by selecting a complete set of coset leaders with the minimum energy as codewords.

Mathematically, the multilevel nested lattice codes is given by $\Lambda_f \cap \mathcal{V}_{\Lambda_c}$. The design rate is given by

$$R_{\text{design}} = \sum_{i=1}^{m} \frac{m_f^l - m_c^i}{n} \log(p_i). \tag{6.72}$$

The design rate becomes the actual rate if every $\mathbf{G}_f^i$ is full-rank which will be fulfilled with high probability.

The transmitter ℓ first decomposes the message w_ℓ into $(\mathbf{w}_\ell^1, \ldots, \mathbf{w}_\ell^m)$, where $\mathbf{w}_\ell^i$ is a length $(m_f^i - m_c^i)$ vector over $\mathbb{F}_{p_i}$, and pads m_c^i 0 in front of $\mathbf{w}_\ell^i$ to get $\mathbf{v}_\ell^i = \begin{bmatrix} \mathbf{0} \\ \mathbf{w}_\ell^i \end{bmatrix}$, $i \in \{1, \ldots, m\}$, is then encoded via C^i to get $\mathbf{c}_\ell^i = \mathbf{G}_f^i \odot \mathbf{v}_\ell^i$. The codeword $\mathbf{t}_\ell \in \Lambda_f \cap \mathcal{V}_{\Lambda_c}$ is formed as

$$\mathbf{t}_\ell = \left(\gamma \left(\Pi_{i=1}^m p_i \right)^{-1} \mathcal{M}(\mathbf{c}_\ell^1, \ldots, \mathbf{c}_\ell^m) + \gamma \boldsymbol{\zeta}_\ell \right) \mod \Lambda_c, \tag{6.73}$$

with $\boldsymbol{\zeta}_\ell \in \mathbb{Z}^N$. It then sends a dithered version

$$\mathbf{x}_\ell = (\mathbf{t}_\ell - \mathbf{u}_\ell) \mod \Lambda_c. \tag{6.74}$$

At the receiver j, by scaling the received signal by α_j and adding the dithers back, one obtains $\mathbf{y}_j'$ in (6.17). Moreover, with the relationship $a_{j\ell} = \bar{a}_{j\ell} + \Pi_{i=1}^m p_i \tilde{a}_{j\ell}$, one can further rewrite $\mathbf{t}_{eq,j}$ in (6.19) as

$$\begin{aligned}
\mathbf{t}_{eq,j} &= \left(\sum_{\ell=1}^{L} (\bar{a}_{j\ell} + \Pi_{i=1}^m p_i \tilde{a}_{j\ell}) \mathbf{t}_\ell \right) \mod \Lambda_c \\
&= \left(\gamma (\Pi_{i=1}^m p_i)^{-1} \sum_{\ell=1}^{L} \mathcal{M}(b_{j\ell}^1, \ldots, b_{j\ell}^m) \mathcal{M}(\mathbf{c}_\ell^1, \ldots, \mathbf{c}_\ell^m) \right. \\
&\quad \left. + \gamma \sum_{\ell=1}^{L} \boldsymbol{\zeta}_\ell + \gamma \sum_{\ell=1}^{L} \tilde{a}_{j\ell} \frac{\Pi_{i=1}^m p_i}{\gamma} \mathbf{t}_\ell \right) \mod \Lambda_c \\
&\overset{(a)}{=} \left(\gamma (\Pi_{i=1}^m p_i)^{-1} \mathcal{M} \left(\bigoplus_{\ell=1}^{L} b_{j\ell}^1 \odot \mathbf{c}_\ell^1, \ldots, \bigoplus_{\ell=1}^{L} b_{j\ell}^m \odot \mathbf{c}_\ell^m \right) + \gamma \boldsymbol{\zeta}_j \right) \mod \Lambda_c,
\end{aligned} \tag{6.75}$$

where $\boldsymbol{\zeta}_j \in \mathbb{Z}^N$ and (a) holds because $\mathcal{M}(.)$ is a ring isomorphism and $\Pi_{i=1}^m p_i / \gamma \mathbf{t}_\ell \in \mathbb{Z}^N$. One can then decode the fine lattice point corresponding to $\mathbf{t}_{eq,j}$ by decoding the equivalent codeword $\bigoplus_{\ell=1}^{L} b_{j\ell}^i \odot \mathbf{c}_\ell^i$, which corresponds to

the message $\bigoplus_{\ell=1}^{L} b_{j\ell}^{i} \odot \mathbf{v}_{\ell}^{i}$, level by level. This in turn gives an estimate of $\mathbf{u}_{j}^{i}$ for $i \in \{1, \ldots, m\}$. Let $\mathsf{Z}_{eq,j}^{*}$ be a zero-mean Gaussian random variable having a same second moment with $\mathsf{Z}_{eq,j}$. Similar to [13], one can show that there exists a sequence of Λ_f whose error probability under multistage decoding can be made arbitrarily small whenever

$$\mathrm{Vol}(\Lambda_f)^{\frac{2}{n}} > 2\pi \exp(1)\sigma_{eq,j}^{2} 2^{-\frac{2}{n} D(\mathsf{Z}_{eq,j} \| \mathsf{Z}_{eq,j}^{*})}. \tag{6.76}$$

This leads to reliable computation under multistage decoding whenever

$$\begin{aligned} R &= \frac{1}{n} \log \left(\frac{\mathrm{Vol}(\Lambda_c)}{\mathrm{Vol}(\Lambda_f)} \right) \\ &= \frac{1}{n} \log(\mathrm{Vol}(\Lambda_c)) - \frac{1}{n} \log(\mathrm{Vol}(\Lambda_f)) \\ &< \frac{1}{2} \log \frac{P}{G(\Lambda_c)} - \frac{1}{2} \log 2\pi e \sigma_{eq}^{2} 2^{-\frac{2}{n} D(\mathsf{Z}_{eq} \| \mathsf{Z}_{eq}^{*})} \\ &= \frac{1}{2} \log^{+} \left(\left(\|\mathbf{a}_j\|^2 - \frac{P|\mathbf{h}_j^{*}\mathbf{a}_j|^2}{1 + P\|\mathbf{h}_j\|^2} \right)^{-1} \right) - \frac{1}{2} \log(2\pi e G(\Lambda_c)) + \frac{1}{n} D(\mathsf{Z}_{eq} || \mathsf{Z}_{eq}^{*}), \end{aligned} \tag{6.77}$$

in the limit as $N \to \infty$.

Note that if Λ_c is good for MSE quantization, then $G(\Lambda_c) \to \infty$ and $\frac{1}{n} D(\mathsf{Z}_{eq} || \mathsf{Z}_{eq}^{*}) \to 0$ [34]; thus, (6.77) becomes the achievable computation rate of Nazer and Gastpar in Theorem 6.1. However, this will require a sequence of Construction π_A lattices that is good for MSE quantization, whose existence has not been proved yet. We also note that for the hypercubic shaping, i.e., $\Lambda_c = \gamma \mathbb{Z}^n$, $G(\Lambda_c) = 1/12$ and thus,

$$R > \frac{1}{2} \log^{+} \left(\left(\|\mathbf{a}_j\|^2 - \frac{P|\mathbf{h}_j^{*}\mathbf{a}_j|^2}{1 + P\|\mathbf{h}_j\|^2} \right)^{-1} \right) - \frac{1}{2} \log \left(\frac{\pi e}{6} \right), \tag{6.78}$$

which is 1.53 dB away to the achievable computation rate in Theorem 6.1 in the high signal-to-noise ratio (SNR) regime.

6.6.2 *Layered Integer Forcing*

Based on the Theorems developed above, we show in this section an efficient way of decoding the linear combination of the multi-source messages within multilevel network decoding, named layered integer forcing (LIF), with greatly reduced complexity. The traditional layered integer forcing can be found by papers e.g. [35].

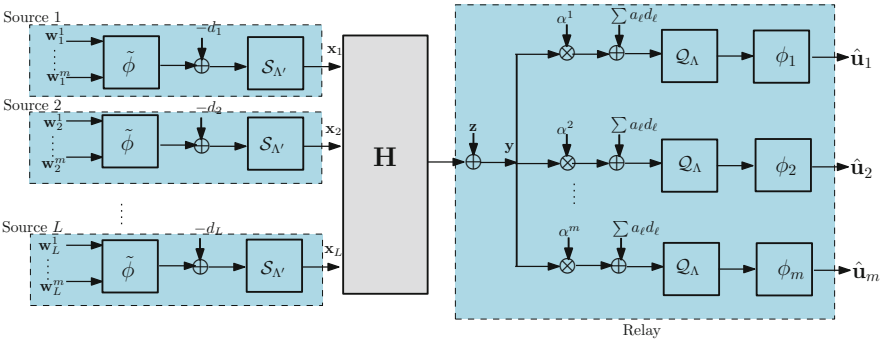


Fig. 6.3 System diagram of the multilevel lattice network coding and multistage decoding. The right-hand side of **H** represents the decoding for a single relay

Traditional Approach Theorems 6.2, 6.3, and Lemma 6.1 imply that the message space with large cardinality may be expressed as a set of smaller message spaces over the hybrid finite field and finite chain ring. Figure 6.3 depicts a multilevel lattice network coding architecture, with L sources and a single relay. The encoder $\mathcal{E}_\ell$ at the ℓth source maps the original message $\mathbf{w}_\ell = \mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m$ to a fine lattice point Λ (assuming n-dimension) via the injective map $\tilde{\phi}$ defined in Lemma 6.1. Then we add a dither $\boldsymbol{d}_\ell \in \mathbb{C}^n$ which is uniformly distributed over the fundamental Voronoi region $\mathcal{V}_{\Lambda'}$ of Λ'. The dithered lattices pass through a nested shaping operator in order to restrain the power consumption. This operation is performed via the sublattice quantization:

$$\lambda'_\ell = Q_{\Lambda'}(\tilde{\phi}(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m) + \boldsymbol{d}_\ell) \tag{6.79}$$

where $\lambda'_\ell \in \Lambda'$, and $Q_{\Lambda'}(\cdot) : \mathbb{C}^n \longmapsto \Lambda'$ is a coarse lattice quantizer. The output of the ℓth source is given by:

$$\begin{aligned} \mathbf{x}_\ell &= \mathcal{E}_\ell(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m) \\ &= \tilde{\phi}(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m) + \boldsymbol{d}_\ell - \lambda'_\ell \end{aligned} \tag{6.80}$$

Note that $\mathbf{x}_\ell$ is uniformly distributed over $\mathcal{V}_{\Lambda'}$ due to the effect of the dither. The average power of the transmitted signal $\mathbf{x}_\ell$ is given by:

$$P = \frac{1}{n\text{Vol}(\mathcal{V}_{\Lambda'})} \int_{\mathcal{V}_{\Lambda'}} \| \mathbf{x}_\ell \|^2 \, \mathrm{d}\mathbf{x}_\ell \tag{6.81}$$

which is the second moment per dimension of $\mathbf{x}_\ell$ over $\mathcal{V}_{\Lambda'}$. The message space at each source consists of a direct sum of m small message spaces (assuming there are m levels) over different finite fields or chain rings. The encoder $\mathcal{E}_\ell$ constructs a one-to-one relation between the message space and the coset system Λ/Λ'.

At the relay, given the received signals $\mathbf{y}$ and an S-integer vector $\tilde{\mathbf{a}} = [\tilde{a}_1, \tilde{a}_2, \cdots, \tilde{a}_L]^T \in S^L$, the decoder aims at computing a new lattice point which is regarded as an S-linear combination of transmitted lattice points from all sources. The homomorphism designed for the coset system will be used for decoding the lattice point to a linear combination of the original messages. We assume in this paper that the fading coefficients $\mathbf{h} = [h_1, h_2, \cdots, h_L]$, and dithers are perfectly known at the relay. The decoder can be described, generally, by:

$$\mathcal{D} : (\mathbb{C}^n, \mathbb{C}^L, S^L, \mathbb{C}) \longmapsto W, \quad \hat{\mathbf{u}} = \mathcal{D}(\mathbf{y}|\mathbf{h}, \tilde{\mathbf{a}}, \alpha, \mathbf{d}) \tag{6.82}$$

Thus, the output of $\mathcal{D}(\mathbf{y}|\mathbf{h}, \tilde{\mathbf{a}}, \alpha, \mathbf{d})$ is the estimates of the linear combination of the original messages of each source. Here α is a scaling factor [21] which maximises the computation rate. Note that the aforementioned decoder (6.82) may vary according to the specific problem. There may be additional information available to the decoder, and the decoder may also have extra outputs. However, basically the core idea for the decoding remains the same. Based on the quotient lattice Λ/Λ', we have:

$$\begin{aligned}
\hat{\mathbf{u}} &= \mathcal{D}(\mathbf{y}|\mathbf{h}, \mathbf{a}, \alpha, \mathbf{d}) \\
&\overset{(a)}{=} \phi\left(Q_\Lambda\left(\alpha\mathbf{y} - \sum_{\ell=1}^{L} \tilde{a}_\ell \boldsymbol{d}_\ell\right)\right) && (6.83)\\
&\overset{(b)}{=} \phi\left(Q_\Lambda\left(\sum_{\ell=1}^{L} \tilde{a}_\ell\left(\tilde{\phi}(\mathbf{w}_\ell) - \lambda'_\ell\right) + \mathbf{n}_{\text{eff}}\right)\right) && (6.84)\\
&\overset{(c)}{=} \phi\left(\sum_{\ell=1}^{L} \tilde{a}_\ell \tilde{\phi}(\mathbf{w}_\ell) + Q_\Lambda(\mathbf{n}_{\text{eff}})\right) && (6.85)\\
&\overset{(d)}{=} \bigoplus_{\ell=1}^{L} a_\ell \mathbf{w}_\ell \boxplus \phi\left(Q_\Lambda(\mathbf{n}_{\text{eff}})\right) && (6.86)
\end{aligned}$$

where (a) follows from the fact that we expect to quantize a set of scaled received signals which are subtracted from the corresponding dithers. (b) follows from the manipulation of:

$$\alpha\mathbf{y} = \sum_{\ell=1}^{L} \tilde{a}_\ell \mathbf{x}_\ell + \sum_{\ell=1}^{L} \tilde{a}_\ell \boldsymbol{d}_\ell + \overbrace{\sum_{\ell=1}^{L} (\alpha h_\ell - \tilde{a}_\ell)\mathbf{x}_\ell + \alpha\mathbf{z}}^{\mathbf{n}_{\text{eff}}} \tag{6.87}$$

(c) follows from the definition of the lattice quantizer, and (d) follows from the properties of a surjective module homomorphism, and also Lemma 6.1. Note that here $\phi(\tilde{a}_\ell) = a_\ell \in \mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m$.

Equations (6.83)–(6.86) reveal the decoding operations for the traditional lattice-based PNC. We are able to decode a linear combination of messages $\bigoplus_{\ell=1}^{L} a_\ell \mathbf{w}_\ell$ over all sources without errors provided that $\phi\left(Q_\Lambda\left(\mathbf{n}_{\text{eff}}\right)\right) = 0$. Thus, the successful decoding is guaranteed if and only if the effective noise is quantized to the kernel of ϕ, $\mathcal{K}(\phi)$.

The problems left unsolved are: (1) how to exploit rich ring features in order to make it practically applicable in lattice-based network coding. (2) when the cardinality (the coset representatives) of Λ/Λ' is large, the complexity of the lattice quantizer becomes unmanageable, which restricts the application of LNC. What is the practical lattice network decoding approach that could greatly relieves the decoding load in LNC. We study a new decoding solution which is specifically designed in terms of MLNC, and which relaxes the two problems mentioned.

Layered Integer Forcing The breakthrough of MLNC (based on Theorems and Lemmas developed) is that

- The original message space over Λ/Λ' can be decomposed into a direct sum of m smaller message spaces in terms of Λ_{p_i}/Λ', $i = 1, 2, \cdots, m$.
- The relay can decode each layer independently; thus the decoder tries to infer and forward a linear combination of messages of each layer *separately* over the message subspace defined in Theorem 6.3.

Let us recall the traditional decoding operations explained in (6.83)–(6.86). If we are only concerned with the linear combination of a particular layer, the quantization of the effective noise need not necessarily be the kernel of ϕ. There must exist other lattice points in Λ/Λ' such that the homomorphism of these points does not interfere with the linear combination of that layer following the aforementioned theorems.

Theorem 6.4 *There exists a quotient S-lattice Λ/Λ_i' with generator matrices $\mathbf{G}_\Lambda$ for Λ, and $\mathbf{G}_{\Lambda_i'}$ for Λ_i', which satisfies:*

$$\mathbf{G}_{\Lambda_i'} = \begin{bmatrix} \text{Diag}(\underbrace{\mathbf{I}, p_i^{\theta_1}, \cdots, p_i^{\theta_t}}_{k}, \mathbf{I}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \mathbf{G}_\Lambda \tag{6.88}$$

and there is a surjective S-module homomorphism φ_i:

$$\varphi_i : \Lambda \longmapsto S/\langle p_i^{\theta_1} \rangle \oplus S/\langle p_i^{\theta_2} \rangle \oplus \cdots \oplus S/\langle p_i^{\theta_t} \rangle \tag{6.89}$$

whose kernel $\mathcal{K}(\varphi_i) = \Lambda_i'$. The quotient S-lattice Λ/Λ_i' is isomorphic to the direct sum of cyclic modules:

$$\Lambda/\Lambda_i' \cong S/\langle p_i^{\theta_1} \rangle \oplus S/\langle p_i^{\theta_2} \rangle \oplus \cdots \oplus S/\langle p_i^{\theta_t} \rangle \tag{6.90}$$

Note that although both Λ_{p_i}/Λ' and Λ/Λ'_i are isomorphic to $S/\langle p_i^{\theta_1}\rangle \oplus S/\langle p_i^{\theta_2}\rangle \oplus \cdots \oplus S/\langle p_i^{\theta_t}\rangle$, *they belong to different coset systems*. Λ_{p_i}/Λ' is related to the construction of lattices that have multilevel structure, whereas Λ/Λ'_i is related to the decoding issues, i.e. LIF.

Theorem 6.4 defines a new sublattice Λ'_i which plays a key role in decoding MLNC, as it is the kernel of the quotient S-lattice that possesses a surjective homomorphism φ_i for the ith layer. Hence it is possible to decode an S-linear combination of fine lattice points to an S-linear combination of the original messages of the ith layer. This is explained in Lemma 6.3.

Lemma 6.4 *Given the embedding injective map* $\tilde{\phi} : (\mathbf{w}^1, \cdots, \mathbf{w}^m) \longmapsto \Lambda$, *there exists a surjective S-module homomorphism* φ_i, $i = 1, 2, \cdots, m$, *defined in (6.89), satisfying:*

$$\varphi_i\big(\tilde{\phi}(\mathbf{w}^1 \oplus \cdots \oplus \mathbf{w}^m)\big) = \begin{cases} \mathbf{w}^i, \mathbf{w}^i \notin \langle p_i^{\theta_1}\rangle \oplus \cdots \oplus \langle p_i^{\theta_t}\rangle \\ 0, \quad \mathbf{w}^i \in \langle p_i^{\theta_1}\rangle \oplus \cdots \oplus \langle p_i^{\theta_t}\rangle \end{cases} \tag{6.91}$$

Based on Lemma 6.3, it is now possible to decode the linear combination of the messages of each layer separately and independently. Assuming the messages at the ith layer is of interest, the relay computes:

$$\hat{\mathbf{u}}^i = \mathcal{D}^i(\mathbf{y}|\mathbf{h}, \mathbf{a}^i, \alpha^i, \mathbf{d}) \tag{6.92}$$

$$= \varphi_i\left(Q_\Lambda\left(\alpha^i \mathbf{y} - \sum_{\ell=1}^{L} \tilde{a}_\ell^i \boldsymbol{d}_\ell\right)\right) \tag{6.93}$$

where

$$\mathcal{D}^i : (\mathbb{C}^n, \mathbb{C}^L, S^L, \mathbb{C}, \mathbb{C}^{n\times L}) \longmapsto W^i \tag{6.94}$$

and $\alpha^i \in \mathbb{C}$ and $\mathbf{a}^i$ are the scaling parameter and S-integer coefficients of the ith layer, respectively, which are determined by some optimisation criterion in terms of the quotient S-lattice Λ/Λ'_i.

Theorem 6.4 and Lemma 6.4 lay the foundation of the layered integer forcing. The linear combination of $\hat{\mathbf{u}}^i$ can be recovered in terms of LIF by:

$$\hat{\mathbf{u}}^i \overset{(d)}{=} \varphi_i\left(Q_\Lambda\Big(\sum_{\ell=1}^{L} \tilde{a}_\ell^i\big(\tilde{\phi}(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m) - \lambda'_\ell\big) + \mathbf{n}_{\text{eff}}\Big)\right)$$

$$\overset{(e)}{=} \varphi_i\left(\sum_{\ell=1}^{L} \tilde{a}_\ell^i \tilde{\phi}(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m) - \lambda'_\ell - \lambda'_{i,\ell} + Q_\Lambda(\mathbf{n}_{\text{eff}})\right)$$

$$\stackrel{(f)}{=} \varphi_i\left(\sum_{\ell=1}^{L} \tilde{a}_\ell^i \tilde{\phi}(\mathbf{w}_\ell^1 \oplus \cdots \oplus \mathbf{w}_\ell^m)\right) \boxplus \varphi_i\left(Q_\Lambda(\mathbf{n}_{\text{eff}})\right)$$

$$\stackrel{(g)}{=} \bigoplus_{\ell=1}^{L} a_\ell^i \mathbf{w}_\ell^i \boxplus \varphi_i\left(Q_\Lambda(\mathbf{n}_{\text{eff}})\right) \tag{6.95}$$

where (d) follows from (6.80) and basic arithmetic manipulations; (e) follows from the definition of the lattice quantizer Q_Λ, and also the S-linear combination of the lattice points is restricted in $\mathcal{V}_{\Lambda_i'}$; (f) follows from the property of a surjective S-module homomorphism, and also the fact that $\lambda' \subseteq \lambda_i'$ and $\mathcal{K}(\varphi_i) = \lambda_i'$. (g) follows from Lemma 6.3, and note that $\varphi_i(\tilde{a}_\ell^i) = a_\ell^i \in W^i$.

Lemma 6.5 *The linear combination of the messages at the* ith *layer* $\hat{\mathbf{u}}^i = \bigoplus_{\ell=1}^{L} a_\ell^i \mathbf{w}_\ell^i$ *can be recovered if and only if* $Q_\Lambda(\mathbf{n}_{\text{eff}}) \in \Lambda_i'$. *Thus,* $\Pr(\hat{\mathbf{u}}^i \neq \mathbf{u}^i) = \Pr(Q_\Lambda(\mathbf{n}_{\text{eff}}) \notin \Lambda_i')$.

Lemma 6.5 reveals that the lattice Λ_i' defined in Theorem 6.4 plays a key role in decoding the messages of the ith layer.

6.6.3 Multistage Iterative Decoding Algorithm for EDC Lattices

In this section, we present an iteration-aided multistage decoding approach specifically designed for EDC, which provides a feasible way of improving the performance of decoding the linear combinations, and also of increasing the overall rate with low decoding-complexity. We consider S to be a ring of Eisenstein integers $\mathbb{Z}[\omega]$ in the sequel.

We have clearly revealed the possible encoding structure for EDC. Recalling the definition for EDC, we know that the map $\tilde{\sigma} : S^n \longmapsto (S/\langle p_1^{\gamma_1}\rangle)^n \oplus (S/\langle p_2^{\gamma_2}\rangle)^n \oplus \cdots \oplus (S/\langle p_m^{\gamma_m}\rangle)^n$ is a natural projection of a surjective ring homomorphism $\sigma : S \longmapsto S/\langle p_1^{\gamma_1}\rangle \times S/\langle p_2^{\gamma_2}\rangle \times \cdots \times S/\langle p_m^{\gamma_m}\rangle \longleftrightarrow \mathbb{F}_{\tilde{p}_1} \times \cdots \times \mathbb{F}_{\tilde{p}_m}$ by applying it element-wise [3] ($\gamma_i = 1, \forall i = 1, 2, \cdots, m$). Note that in this case, σ is actually an *f.g.* abelian group homomorphism. It is easy to see that each level $S/\langle p_i\rangle$ is coded by an $[n, k_i]$ linear code C^i over $\mathbb{F}_{\tilde{p}_i}$ (a finite field or finite chain ring determined by $\tilde{p}_i$).

6.6.3.1 Soft Detector for EDC

A general decoding method LIF for MLNC has been developed in the prior sections, based on the optimised scaling factor α, S-integer coefficient vectors $\tilde{\mathbf{a}}_i$, and a good EDC lattice quantizer, e.g. a Viterbi decoder with modified metrics (see appendix).

Thus, when EDC is employed in MLNC, LIF is also feasible. In this section, we explore another detection approach designed specifically for the EDC-based MLNC (which follows from the structure of the EDC lattices). Especially an iterative detector is developed, which exploits the multilevel structure gain of EDC by using multistage decoding.

First, we consider the non-iterative multistage decoding. The detector tries to decode the linear function of each level stage-by-stage, with the aid of the a priori information from the preceding layers. The detection structure is similar to the point-to-point multilevel codes, e.g. [2, 28] whereas here the a priori information is the soft estimation. We develop a layered soft detector (LSD) which calculates the posteriori L-vector (a vector of Log-likelihood ratio) for each layer with the aid of the multiple a priori L-vectors.

The LSD decodes the linear function of each layer over the corresponding non-binary finite field, and hence the a priori information of each layer is no longer a scalar value. We define the a priori information $\mathbf{A}^i$ to be a vector-based random variable with realization:

$$\mathbf{a}^i = \left[\log\left(\frac{\Pr(\xi|V^i = v_1^i)}{\Pr(\xi|V^i = 0)}\right) \cdots \log\left(\frac{\Pr(\xi|V^i = v_{\tilde{p}_i-1}^i)}{\Pr(\xi|V^i = 0)}\right) \right] \tag{6.96}$$

where V^i denotes the possible linear combinations at the ith level, which is a uniformly distributed random variable whose kth realization is $v_k^i \in \mathbb{F}_{\tilde{p}_i}$, $k = 1, 2, \cdots, \tilde{p}_i - 1$. $\Pr(\xi|V^i = v_k^i)$ is the probability of the a priori channel outputs $\Xi = \xi$ given the event $V^i = v_k^i$. Assume that $w_j^i \in \mathbb{F}_{\tilde{p}_i}$, $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, L$ to be the message of the ith level and the jth source, the linear function is defined by $f^i(w_1^i, \cdots, w_L^i) = \bigoplus_{\ell=1}^{L} a_\ell^i w_\ell^i$ over $\mathbb{F}_{\tilde{p}_i}$. Note that the integer coefficient a_ℓ^i can be determined either by the lattice reduction approach as introduced in [9, 12] over the ith quotient lattice Λ/Λ_i' as defined in Theorem 6.4, or by the maximum mutual information criterion as described later.

In the multistage iterative decoding, the proposed LSD outputs the extrinsic L-vector $\mathbf{e}^i$ for the ith level, based on the a priori L-vector $\mathbf{a}^j$, $j \in \{1, \cdots, m\}$, $j \neq i$. Assume that there is a two level EDC and the decoding proceeds from layer 1 (which is regarded as the first stage decoding) to layer 2 (the second stage decoding). The extrinsic outputs of layer 1 feed into layer 2 to assist the second stage decoding. With the aid of the a priori L-value, layer 2 estimates and forwards the extrinsic information (which serves as the a priori information of layer 1) to layer 1. The process is repeated and all layers are activated in turn for the second and subsequent iterations. We refer to this approach as the iterative MSD (IMSD) scheme for MLNC. The detection process is similar to iterative decoding of multilevel codes, e.g. [30] whereas the nature of the detection is different. As the iteration proceeds, each layer will produce more reliable extrinsic L-vector $\mathbf{e}^i$ which also serves as the a priori information of the soft-in soft-out non-binary decoder for the corresponding C^i. Figure 6.4 illustrates the multistage iterative decoding with two stages.

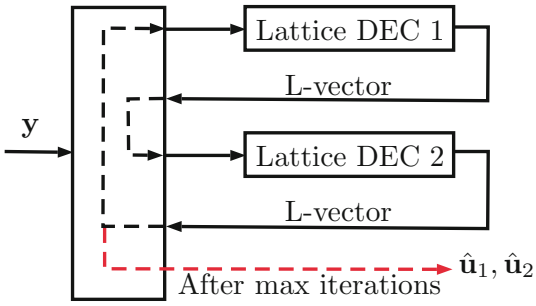


Fig. 6.4 Two-stage LSD iterative decoding model. Interleaver and de-interleaver are not shown

Table 6.1 Code type and code rate assigned for each level

i	$\mathbf{g}(D)$
1	$[-2\omega^2 + 2\omega^2 D^3, 2\omega^2 + (-2\omega^2)D + 2\omega^2 D^3]$
2	$\begin{bmatrix} -2 + (1-\omega)D^2 + (-2)D^3 \\ -2 + (-2)D + (-2)D^2 + (1-\omega^2)D^3 \end{bmatrix}$

It is seen that the maximum achievable rates for the network coded linear combinations are $\mathcal{R}^{(1)} = \log_2 3$ and $\mathcal{R}^{(2)} = \log_2 4$ for level 1 and 2. The allowable rate at a certain level is higher when the a priori information from another layer is available. We assume two memory 3, 1/2-rate convolutional codes are used at both levels (over $\mathbb{F}_3$ and $\mathbb{F}_{2^2}$ respectively). EDC lattices achieve overall rate $\frac{1}{2}\log_2(12)$, with the number of trellis states 27 and 64 at the corresponding levels. However, a single convolutional code over ring R_{12} needs 1728 trellis states. The complexity reduction is obvious.

6.6.4 Simulation Results

We focus mainly on the applications of EDC lattices in MLNC. Note that MLNC design applies in principle to any lattice codes (e.g. complex low density lattice codes [31, 33], Signal codes [23]) but we use EDC lattices to verify the theory developed and decoding performance.

We are mainly concerned with the performance of the multiple access channel (MAC) such as the two-way relay channel (TWRC), which can be viewed as the building block for more complicated network topologies. All simulations are based on a two-layer EDC lattice which has the same configuration. Thus, the two layers are constructed via linear codes $C^1 \in \mathbb{F}_3$ and $C^2 \in \mathbb{F}_{2^2}$. The linear codes at both layers are non-binary convolutional codes, with their generator polynomials defined in Table 6.1. Note that the decoder of the non-binary convolutional codes is based on the maximum a posterior (MAP) probability criteria and modified BCJR algorithm, where the soft output of the component symbols is produced.

We examine the performance of MSD based on the asymmetric coding rates over each level, where the rate of layer 2 is set to $\mathcal{R}^{(2)} = \frac{3}{4}$ and layer 1 is set to $\mathcal{R}^{(1)} = \frac{1}{2}$. Thus, the sublattice Λ_{p_2} is constructed via a higher rate linear code. The overall message rate is given by

$$\mathcal{R}_{\text{mes}} \approx \frac{1}{2}\log_2 3 + \frac{3}{4}\log_2 4 \text{ bits/symbol}$$

Note that the SER curve of level 1 (red dashed circle) without MSD should closely match that with MSD (red solid circle) when multistage decoding is used in layer 1. Simulations in Fig. 6.5 confirm this. Based on the increased coding rate, we are more concerned with the SER performance of layer 2. It is observed from Fig. 6.5 that the SER performance of layer 2 is greatly degraded if MSD is not employed, with approximately 3 dB loss at 10^{-5} compared to the half-rate code used at this level. However, when MSD is used, the SER (blue solid square) of layer 2 has more than 3 dB gain over the non-MSD case (blue dashed square) as a result of the reliable a priori feedback from layer 1. The overall performance of MSD-based detection is determined mainly by layer 1, whereas for non-MSD-based detection, the overall performance is dominated by layer 2. That is the reason why the overall SER of the MSD-based scheme performs better than the non-MSD scenario, with 2 dB gain obtained at 10^{-5}. It is interesting to note that when the decoding of the Λ_{p_i}/Λ'

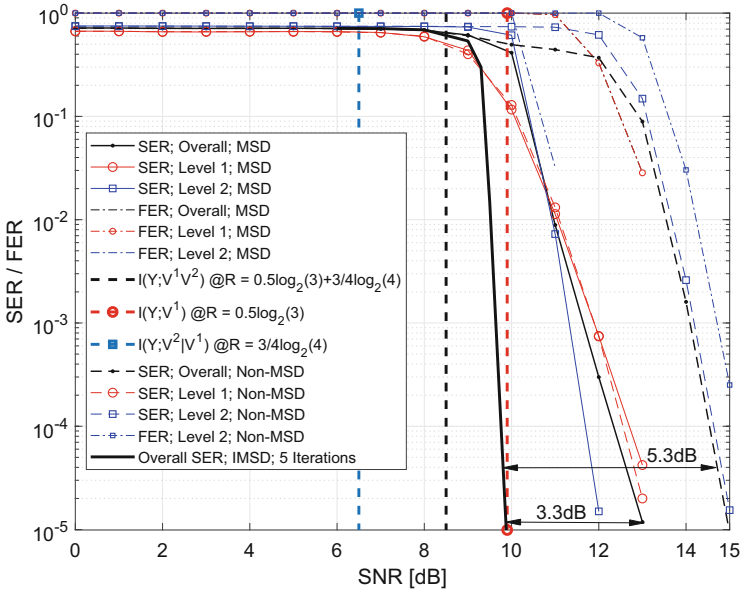


Fig. 6.5 SER and FER performance for an MLNC constructed from a two layer EDC lattices; soft detection; multistage decoding/non-multistage decoding/IMSD; frame length: 10^3. Asymmetric coding rate: $\mathcal{R}^{(1)}_{\text{mes}} = \frac{1}{2}\log_2(3)$; $\mathcal{R}^{(2)}_{\text{mes}} = \frac{3}{4}\log_2(4)$; $h_1 = h_2 = 1$

which is constructed from a higher rate linear code occurs at a later stage of MSD, the overall SER performance of MSD over non-MSD performs better. Hence, MSD is particularly suitable for the detection of EDC lattices in terms of MLNC design, since each layer of EDC operates over an asymmetric finite field or finite chain ring. Now the overall SER is 4.5 dB from the capacity. Note that the measure of SER is based on the correct recovery of the linear combinations of original messages at each source over the respective algebraic field.

Iterative Multistage Decoding We believe that there is room to improve SER and FER performance further. Based on the soft detector developed, and also the soft decoder developed for the non-binary convolutional codes, we propose to apply the iterative technique to EDC-lattice-based MLNC. Note that a pair of pseudorandom interleaver and de-interleaver has been employed in the iterative systems.

Figure 6.5 depicts the result when IMSD is used. It is observed that with 5 iterations, the SER curve (black solid thick line) has a sharp turbo cliff reaching SER $= 10^{-5}$ at 10 dB, which is only 1.4 dB from the capacity. Thus, iterative decoding gives 3.3 dB gain over the traditional MSD decoding, and 5.3 dB gain over non-MSD decoding, as shown in the figure. When sufficient iterations are given, the L-value outputs from the soft detector at both layers are sufficiently reliable that the decoder can make the estimation with small probability of error. The simulation result also validates the soft detector algorithm specifically developed for EDC-based MLNC, and implies that there is large potential in employing iterative decoding in the multilevel lattice network coding.

In Fig. 6.6, we also show the performance of the LSD when the fixed fading is considered. The channel fading vector is set to $\mathbf{h} = [-1.17 + 2.15i, 1.25 - 1.63i]$, which is the same as the fading vector used in scenario 1 of [23]. We employ a half-rate code for layer 1, and $\frac{3}{4}$-rate code for layer 2. We employ multistage decoding with 5 iterations between the two layers. A sharp turbo cliff occurs, which reaches SER $= 10^{-5}$ at 3.9 dB, approximately 1.7 dB from the capacity. When no iteration is employed, there is more than 5 dB loss. This implies that small number of additional iterations to generate more reliable values is worthwhile in improving the overall SER performance. The iterative multistage soft detection for EDC lattices achieves the overall rate of $\mathcal{R}_{\text{mes}} \approx 2.29$ bits/symbol at 3.9 dB. This demonstrates the potential of iterative decoding in improving the performance of physical layer network coding.

6.7 Conclusions

In this chapter, we have reviewed the recent progresses in using structured codes for harnessing interference. Our focus was mainly on reducing the complexity of compute-and-forward. Two frameworks have been reviewed, namely multi-stage compute-and-forward and the multilevel lattice network coding. For the multi-stage compute-and-forward, the achievable computation rate under multi-stage decoding

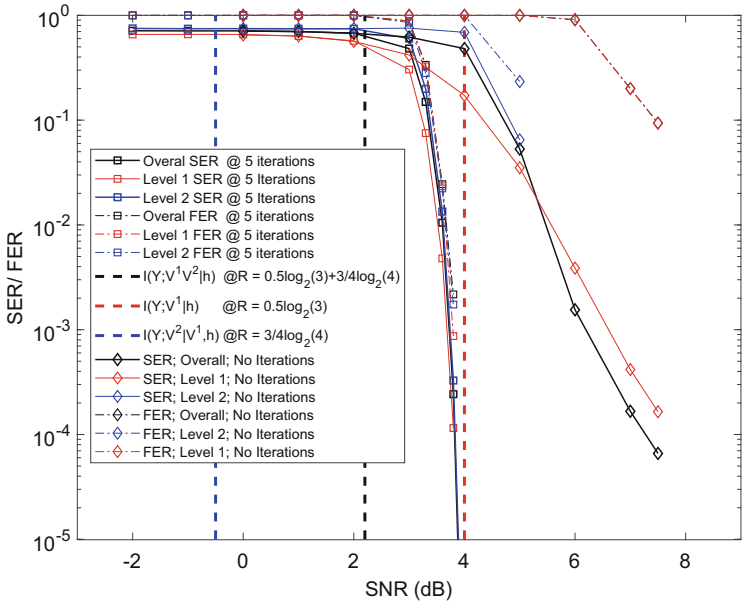


Fig. 6.6 SER and FER performance for an MLNC constructed from a two layer EDC lattices; soft detection; LIF; frame length: 10^3. $\mathcal{R}_{\text{mes}}^{(1)} = \frac{1}{2}\log_2(3)$; $\mathcal{R}_{\text{mes}}^{(2)} = \frac{3}{4}\log_2(4)$; $\mathbf{h} = [-1.17 + 2.15i, 1.25 - 1.63i]$

has been analysed. EDC lattices are proposed based on the multilevel lattice network coding theory and reinterprets the Construction π_A and Construction π_D whose goodness is discussed. Some useful properties for the lattices such as generator matrix, kissing numbers and nominal coding gains have been discussed. Low-complex decoding algorithms of these two frameworks have also been discussed and iterative detection is used to support the decoding performance. The performance of these algorithms have been evaluated through computer simulations.

Appendix: LIF Quantizer

We show here a LIF quantizer $Q_{\text{LIF}}^{(i)}$ implemented via a modified Viterbi decoder. The quantization problem for the ith layer can be mathematically expressed as:

$$\arg\min_{\mathbf{c}_i} ||\alpha^i \mathbf{y} - (\tilde{\sigma}^{-1}(\mathbf{c}^i) + \lambda_i')||^2 \tag{6.97}$$

$$= \arg\min_{\mathbf{c}_i} ||(\alpha^i \mathbf{y} - \tilde{\sigma}^{-1}(\mathbf{c}^i)) - Q_{\Lambda_i'}((\alpha^i \mathbf{y} - \tilde{\sigma}^{-1}(\mathbf{c}^i))||^2 \tag{6.98}$$

$$\text{subject to:} \quad \mathbf{c}^i \in C^i, \quad \lambda_i' \in \Lambda_i', \tag{6.99}$$

$$\tilde{\sigma}(\lambda) \in C^1 \oplus \cdots \oplus (C^i = \mathbf{c}^i) \oplus \cdots C^m \tag{6.100}$$

where $Q_{\Lambda_i'}(\mathbf{x})$ is the coarse lattice quantizer for the ith layer and can be expressed as a modulo operation $\mathbf{x} \mod \Lambda_i'$ (as defined in Theorem 6.4). $\tilde{\sigma}^{-1}(\cdot)$ is the inverse operation of $\tilde{\sigma}$ which produces a set of lattice points λ.

We can construct a trellis for the non-binary convolutional code C^i. Assume that the states of the kth and $(k+1)$th time slots are s_k and s_{k+1}, respectively. The code corresponding to the branch that exists from s_k and arrives at s_{k+1} is denoted as $c^i_{s_k \to s_{k+1}}$. The metric for each branch is given by

$$||(\alpha^i \mathbf{y} - \sigma^{-1}(c^i_{s_k \to s_{k+1}})) - Q_{\Lambda_i'}((\alpha^i \mathbf{y} - \sigma^{-1}(c^i_{s_k \to s_{k+1}}))||^2 \tag{6.101}$$

where $\sigma^{-1}(\cdot)$ is the inverse operation of $\sigma(\cdot)$. We employ the Viterbi algorithm to estimate the best possible outcome $\mathbf{c}^i$. This implements the LIF quantizer $Q^{(i)}_{\text{LIF}}$ for EDC-based MLNC.

References

1. Barnes, E.S., Sloane, N.J.A.: New lattice packings of spheres. Can. J. Math. **35**(1), 117–130 (1983)
2. Burr, A., Lunn, T.: Block-coded modulation optimized for finite error rate on the white Gaussian noise channel. IEEE Trans. Inf. Theory **43**(1), 373–385 (1997)
3. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices, and Groups. Springer, Berlin (1999)
4. Erez, U., Litsyn, S., Zamir, R.: Lattices which are good for (almost) everything. IEEE Trans. Inf. Theory **51**(10), 3401–3416 (2005)
5. Erez, U., Zamir, R.: Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding. IEEE Trans. Inf. Theory **50**(10), 2293–2314 (2004)
6. Feng, C., Silva, D., Kschischang, F.R.: Lattice network coding over finite rings. In: Proceedings of the CWIT, pp. 78–81 (2011)
7. Feng, C., Silva, D., Kschischang, F.R.: Lattice network coding via signal codes. In: 2011 IEEE International Symposium on Information Theory Proceedings, pp. 2642–2646 (2011)
8. Feng, C., Silva, D., Kschischang, F.R.: An algebraic approach to physical-layer network coding. IEEE Trans. Inf. Theory **59**(11), 7576–7596 (2013)
9. Feng, C., Silva, D., Kschischang, F.: An algebraic approach to physical-layer network coding. IEEE Trans. Inf. Theory **59**(11), 7576–7596 (2013)
10. Feng, C., Nobrega, R., Kschischang, F., Silva, D.: Communication over finite-chain-ring matrix channels. IEEE Trans. Inf. Theory **60**(10), 5899–5917 (2014)
11. Forney, G.D., Trott, M.D., Chung, S.Y.: Sphere-bound-achieving coset codes and multilevel coset codes. IEEE Trans. Inf. Theory **46**(3), 820–850 (2000)
12. Gan, Y.H., Ling, C., Mow, W.H.: Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection. IEEE Trans. Signal Process. **57**(7), 2701–2710 (2009)
13. Huang, Y.C., Narayanan, K.: Construction π_A and π_D lattices: construction, goodness, and decoding algorithms. IEEE Trans. Inf. Theory **63**(9), 5718–5733 (2017)
14. Huang, Y.C., Narayanan, K.R.: Multistage compute-and-forward with multilevel lattice codes based on product constructions. In: Proceedings of the IEEE International Symposium on Information Theory (2014). ArXiv:1401.2228 [cs.IT]
15. Huang, Y.C., Narayanan, K., Wang, P.C.: Lattices over algebraic integers with an application to compute-and-forward. IEEE Trans. Inf. Theory **64**(10), 6863–6877 (2018)

16. Leech, J., Sloane, N.J.A.: Sphere packing and error-correcting codes. Can. J. Math. **23**(4), 718–745 (1971)
17. Liu, L., Ling, C.: Polar lattices for lossy compression (2015). arXiv:1501.05683v3 [cs.IT]
18. Loeliger, H.A.: Averaging bounds for lattices and linear codes. IEEE Trans. Inf. Theory **43**(6), 1767–1773 (1997)
19. McDonald, B.R.: Finite Rings with Identity, vol. 28. Marcel Dekker, New York (1974)
20. Nazer, B., Cadambe, V., Ntranos, V., Caire, G.: Expanding the compute-and-forward framework: unequal powers, signal levels, and multiple linear combinations. IEEE Trans. Inf. Theory **62**(9), 4879–4909 (2016)
21. Nazer, B., Gastpar, M.: Compute-and-forward: harnessing interference through structured codes. IEEE Trans. Inf. Theory **57**(10), 6463–6486 (2011)
22. Ordentlich, O., Erez, U.: A simple proof for the existence of good pairs of nested lattices. IEEE Trans. Inf. Theory **62**(8), 4439–4453 (2016)
23. Shalvi, O., Sommer, N., Feder, M.: Signal codes: convolutional lattice codes. IEEE Trans. Inf. Theory **57**(8), 5203–5226 (2011)
24. Sun, Q., Yuan, J., Huang, T., Shum, K.: Lattice network codes based on Eisenstein integers. IEEE Trans. Commun. **61**(7), 2713–2725 (2013)
25. Sun, Q.T., Yuan, J., Huang, T., Shum, K.W.: Lattice network codes based on Eisenstein integers. IEEE Trans. Commun. **61**(7), 2713–2725 (2013)
26. Tunali, N.E., Huang, Y.C., Boutros, J., Narayanan, K.: Lattices over Eisenstein integers for compute-and-forward. IEEE Trans. Inf. Theory **61**(10), 5306–5321 (2015)
27. Tunali, N.E., Narayanan, K., Boutros, J., Huang, Y.C.: Lattices over Eisenstein integers for compute-and-forward. In: Proceedings of the Allerton Conference on Communication, Control, and Computing, pp. 33–40 (2012)
28. Wachsmann, U., Fischer, R., Huber, J.: Multilevel codes: theoretical concepts and practical design rules. IEEE Trans. Inf. Theory **45**(5), 1361–1391 (1999)
29. Wang, Y., Burr, A.: Physical-layer network coding via low density lattice codes. In: 2014 European Conference on Networks and Communications (EuCNC), pp. 1–5 (2014)
30. Wang, Y., Burr, A.: Code design for iterative decoding of multilevel codes. IEEE Trans. Commun. **63**(7), 2404–2419 (2015)
31. Wang, Y., Burr, A.: Complex low density lattice codes to lattice network coding. In: IEEE International Communications Conference (ICC) (2015)
32. Yan, Y., Ling, C., Wu, X.: Polar lattices: where Arikan meets Forney. In: Proceedings of the IEEE International Symposium on Information Theory, pp. 1292–1296 (2013)
33. Yona, Y., Feder, M.: Complex low density lattice codes. In: 2010 IEEE International Symposium on Information Theory Proceedings (ISIT), pp. 1027–1031 (2010)
34. Zamir, R., Feder, M.: On lattice quantization noise. IEEE Trans. Inf. Theory **42**(4), 1152–1159 (1996)
35. Zhan, J., Nazer, B., Erez, U., Gastpar, M.: Integer-forcing linear receivers. IEEE Trans. Inf. Theory **60**(12), 7661–7685 (2014)

Chapter 7
Nested Linear/Lattice Codes Revisited

Renming Qi and Chen Feng

Abstract Random nested linear/lattice codes have played an important role in network information theory. However, the proofs associated with these codes are sometimes involved, making them less accessible than conventional random codes. Recently, several attempts have been made towards simplifying the proofs related to nested linear/lattice codes. In this chapter, we review these recent developments with a particular focus on presenting a unified approach.

7.1 Introduction

In 1948, Claude E. Shannon established the maximum rate at which information can be transmitted reliably over a noisy channel [41]. The mathematical setup is shown in Fig. 7.1, where the channel is modeled as a probabilistic mapping from the input to the output, and the encoder and decoder are to be designed. Under this setup, Shannon proved a remarkable "phase transition" result: There is a fundamental rate limit—referred to as the channel capacity—under which one can design the encoder and decoder to achieve an arbitrarily small probability of error, but above which the probability of error is bounded away from zero (i.e., it cannot be made arbitrarily small no matter how we design the encoder and decoder) [41].

Shannon's channel coding theorem consists of two parts. The *achievability* part says that the probability of error can be made arbitrarily small for any rate below the channel capacity. The *converse* part states that the probability of error is bounded away from zero for any rate above the capacity. While the converse part applies to *any* decoder, the achievability part often involves several specific decoders, such as the maximum-likelihood (ML) decoder [14, p.37] and the joint typicality decoder [41][4, p.199]. These decoders, together with a random coding argument where the encoder generates independent and identically distributed (i.i.d.) codewords

R. Qi · C. Feng (✉)
The School of Engineering, University of British Columbia, Kelowna, BC, Canada
e-mail: renming.qi@alumni.ubc.ca; chen.feng@ubc.ca

V. Beresnevich et al. (eds.), *Number Theory Meets Wireless Communications*,
Mathematical Engineering, https://doi.org/10.1007/978-3-030-61303-7_7

Fig. 7.1 Model of a point-to-point communication system

according to some codeword distribution, are used to prove the existence of good codes (without explicitly constructing them).

Practical communication systems are subject to complexity constraint. To control the computational complexity of encoding and decoding operations, codes with (algebraic) structures are used in practice. This motivates a study of structured codes, such as linear codes [2] and lattice codes [12, 13, 53]. In the sequel, we formally present the system setup and then discuss the use of structured codes in this setup.

7.1.1 System Setup

Here we describe Shannon's mathematical model of a point-to-point communication system depicted in Fig. 7.1. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets, respectively. The channel maps an input sequence (of length n) $\boldsymbol{x} = (x_1, \ldots, x_n)$ to an output sequence (of length n) $\boldsymbol{y} = (y_1, \ldots, y_n)$ in a symbol-by-symbol manner. For example, when $\mathcal{X}$ and $\mathcal{Y}$ are finite, the conditional probability for the channel to output $\boldsymbol{y} \in \mathcal{Y}^n$ given $\boldsymbol{x} \in \mathcal{X}^n$ is

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} p(y_i|x_i),$$

where $p(y|x)$ is a conditional probability mass function (pmf). This channel model is called a discrete memoryless channel (DMC). When $\mathcal{X}$ and $\mathcal{Y}$ are continuous alphabets, the conditional probability density function (pdf) $f(y|x)$ should be used instead of $p(y|x)$. In particular, when

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}},$$

the corresponding channel model is called an additive white Gaussian noise (AWGN) channel.

The encoder maps a message $m \in \{1, \ldots, M\}$ to its corresponding codeword $\boldsymbol{x}(m)$ from a codebook $\mathsf{C} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(M)\}$. The decoder receives an output sequence $\boldsymbol{y}$ from the channel, and finds an "estimate" $\hat{m}$ of m according to certain decoding rule (such as ML decoding or joint typicality decoding).

We say an error occurs if $\hat{m} \neq m$ and denote this error probability as

$$\mathsf{P}_e(m; \mathsf{C}) \triangleq \mathsf{P}(\hat{m} \neq m),$$

where the randomness comes from the channel noise. We define the average error probability as

$$\mathsf{P}_e(\mathsf{C}) \triangleq \frac{1}{M} \sum_{m=1}^{M} \mathsf{P}_e(m; \mathsf{C}).$$

A rate R is said to be *achievable* if there exists a sequence of codebooks $\mathsf{C}^{(n)}$ of length n and size $M^{(n)}$ such that $M^{(n)} \geq 2^{nR}$ and $\mathsf{P}_e\left(\mathsf{C}^{(n)}\right) \to 0$ as $n \to \infty$. Achievable rates are often derived using a random coding argument. For a DMC with $p(y|x)$, we can fix a pmf $p(x)$ and construct a random i.i.d. ensemble in which each symbol of each codeword is generated independently according to $p(x)$. More specifically, we randomly and independently generate $M^{(n)} = \lceil 2^{nR} \rceil$ codewords $\boldsymbol{x}(m)$ for $m \in \{1, \ldots, M^{(n)}\}$, each according to $p(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i)$. Hence, the probability of generating a particular codebook $\mathsf{C}^{(n)}$ in the ensemble is

$$p\left(\mathsf{C}^{(n)}\right) = \prod_{m=1}^{M^{(n)}} p\left(\boldsymbol{x}(m)\right).$$

The key idea behind Shannon's random coding argument is the following. Although the error probability $\mathsf{P}_e\left(\mathsf{C}^{(n)}\right)$ for a particular codebook $\mathsf{C}^{(n)}$ is often hard to evaluate, the expected error probability averaged over all the codebooks in the ensemble is much simpler to analyze. In other words, the random coding argument is an instance of the probabilistic method [1]. Using the random coding argument, Shannon proved that random i.i.d. ensembles achieve both DMC capacity and AWGN channel capacity under joint typicality decoding in his 1948 paper [41].

7.1.2 Structured Codes

Instead of random i.i.d. ensembles, we can make use of random structured ensembles (such as random linear codes and random lattice codes) for the achievability proof. For example, Elias used random linear codes to establish the achievable rate for the binary symmetric channel (which is a special case of the DMC) in 1955 [10]. Perhaps surprisingly, in their seminal work [17], Körner and Marton demonstrated that random linear codes yield better achievable rates than random i.i.d. ensembles for a multi-user source coding problem. Modern developments along this direction include coding problems from relay networks [16, 28, 29, 31, 40, 44, 49], interference channels [3, 26, 30, 32, 34, 37, 42], distributed source coding

[18, 19, 45, 48, 51], and physical-layer secrecy [15, 47, 50], where random structured codes achieve better rates than random i.i.d. codes.

The use of random structured codes is also of practical value. For instance, random linear codes allow for computationally efficient encoding (since the encoding operation is essentially a matrix-vector multiplication), and random lattice codes allow for lattice decoding (which enjoys lower complexity than ML decoding and joint typicality decoding). Hence, the following two questions naturally arise

1. Can random linear codes achieve the DMC capacity?
2. Can random lattice codes achieve the AWGN channel capacity?

Unlike random i.i.d. codes, random structured codes are much less well understood. For example, it is only recently that Padakandla and Pradhan have demonstrated nested linear code ensembles that achieve DMC capacity under joint typicality encoding and decoding [35–37]. In an independent work, Miyake and Muramatsu showed that nested linear code ensembles with special structures based on sparse matrices can also achieve DMC capacity under ML decoding [24, 25, 27]. In 2004, Erez and Zamir showed that nested lattice code ensembles achieve the AWGN channel capacity under lattice encoding and decoding [11]. See [5–7, 11, 21–23, 38, 46] for a history of this long standing problem and Zamir's book [52] for a survey of recent results.

Despite these exciting developments, the achievability proofs associated with random structured codes are sometimes involved, making them much less accessible than their counterparts—random i.i.d. codes. Very recently, several attempts have been made towards simplifying the proofs related to random nested linear/lattice codes [20, 33, 39]. In this chapter, we will review these new developments and simplifications, with a particular focus on presenting a unified approach based on elementary probability, linear algebra, and number theory.

Here, we would like to point out that this chapter is written for a broad audience including those who are less familiar with information theory. Those who are already familiar with information theory can skip many parts in Sects. 7.2 and 7.3.

7.1.3 Notations

We closely follow the notations in [9]. We use the notation $\mathbb{F}, \mathbb{R}, \mathbb{F}_{\mathsf{q}}$ to denote a (general) field, the real numbers, and the field of order q, respectively. We use $\mathcal{X}, \mathcal{Y}$ to denote the alphabets. We use lowercase letters x, y, ... to denote constants. We use bold lowercase letters $\boldsymbol{x}$, $\boldsymbol{y}$, ... to denote constant row vectors. The i-th component of $\boldsymbol{x}$ is denoted as x_i. An all-zero vector $(0, \ldots, 0)$ with a specified dimension is denoted as $\mathbf{0}$. The i-th unit vector is denoted as $\boldsymbol{e}_i$. We use uppercase, sans-serif font letters to denote constant matrix and codebooks, e.g., a linear code C, and a matrix $\mathsf{G} \in \mathbb{F}_{\mathsf{q}}^{k \times n}$. We use uppercase letters $X, Y, \ldots$ to denote random variables. We use bold uppercase letters $\boldsymbol{X}, \boldsymbol{Y}$ to denote random row vectors. The i-th component of $\boldsymbol{X}$ is denoted as X_i. We use bold, uppercase, sans-serif font letters to denote random

Table 7.1 Summary of key notations

Notation	Definition
$\mathbb{F}, \mathbb{R}, \mathbb{F}_{\mathsf{q}}$	A field, the real numbers and a field of order q, respectively
$\mathcal{X}, \mathcal{Y}$	The alphabets
x, X	Constant and random variable, respectively
$\boldsymbol{x}, \boldsymbol{X}$	Constant and random row vector, respectively
$\mathsf{C}, \mathbf{C}$	Constant and random linear code, respectively
$\Lambda, \mathbf{\Lambda}$	Constant and random lattices, respectively
$\mathcal{V}(\Lambda), V(\Lambda)$	The Voronoi region of lattice Λ and the volume of $\mathcal{V}(\Lambda)$, respectively
$\mathsf{G}, \mathbf{G}$	Constant and random matrice, respectively
$\mathcal{B}(\boldsymbol{s}, r)$	The ball centered at $\boldsymbol{s}$ with radius r
$\mathbb{I}(\cdot)$	The indicator function
$p_X(\cdot), \mathsf{E}(X), \mathrm{Var}(X)$	The pmf, expectation and variance of X
$\pi(x \mid \boldsymbol{x})$	The empirical pmf of $\boldsymbol{x}$
$H(\cdot)$	The entropy
$I(X; Y)$	The mutual information between X and Y
$\mathcal{T}_{\epsilon}^{(n)}(X)$	The typical set
$\mathcal{T}_{\epsilon}^{(n)}(X, Y)$	The joint typical set
$\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})$	The conditional typical set
$Q_{\Lambda}(\cdot)$	The nearest neighbor quantizer with respect to Λ

matrix, e.g., a random linear code $\mathbf{C}$ and a random matrix $\mathbf{G}$. A summary of our key notations is provided in Table 7.1.

7.2 Preliminaries

7.2.1 Nested Linear Codes

An (n, k) *linear code* over $\mathbb{F}_{\mathsf{q}}$ is a k-dimensional subspace of the vector space $\mathbb{F}_{\mathsf{q}}^n$. Such a code can be expressed as

$$\mathsf{C} = \{\boldsymbol{a}\mathsf{G} : \boldsymbol{a} \in \mathbb{F}_{\mathsf{q}}^k\}$$

for some full-rank matrix $\mathsf{G} \in \mathbb{F}_{\mathsf{q}}^{k \times n}$ (called a *generator matrix* of C).

A *nested linear code* is a pair of linear codes $(\mathsf{C}_f, \mathsf{C}_c)$ such that $\mathsf{C}_c \subset \mathsf{C}_f$, i.e., each codeword of C_c is also a codeword of C_f. For convenience, C_f is called the *fine code* and C_c is called the *coarse code*. A *coset* of C_c in C_f is defined as

$$\boldsymbol{c}_f + \mathsf{C}_c = \{\boldsymbol{c}_f + \boldsymbol{c} : \boldsymbol{c} \in \mathsf{C}_c\},$$

where $\boldsymbol{c}_f$ is some codeword of C_f. Two cosets are either identical or disjoint [8]. The number of (distinct) cosets of C_c in C_f is called the *index* of C_c in C_f and is denoted by $[\mathsf{C}_f : \mathsf{C}_c]$. By Lagrange's theorem [8],

$$[\mathsf{C}_f : \mathsf{C}_c] = \frac{|\mathsf{C}_f|}{|\mathsf{C}_c|},$$

where $|\mathsf{C}_f|$ and $|\mathsf{C}_c|$ denote the cardinalities of C_f and C_c, respectively.

Suppose that a nested linear code consists of an (n, k_f) fine code C_f and an (n, k_c) coarse code C_c. Then the index $[\mathsf{C}_f : \mathsf{C}_c]$ is $\mathsf{q}^{k_f - k_c}$, since $|\mathsf{C}_f| = \mathsf{q}^{k_f}$ and $|\mathsf{C}_c| = \mathsf{q}^{k_c}$. Moreover, there exist two generator matrices $\mathsf{G}_f \in \mathbb{F}_{\mathsf{q}}^{k_f \times n}$ and $\mathsf{G}_c \in \mathbb{F}_{\mathsf{q}}^{k_c \times n}$ for C_f and C_c, respectively, such that

$$\mathsf{G}_f = \begin{bmatrix} \mathsf{G}_c \\ \mathsf{G}' \end{bmatrix},$$

where G' is a matrix of size $(k_f - k_c) \times n$.

7.2.2 Nested Lattice Codes

A *lattice* is a discrete subgroup (under vector addition) of $\mathbb{R}^n$. Any (full-rank) lattice Λ in $\mathbb{R}^n$ can be expressed in terms of some (full-rank) $n \times n$ generator matrix $\mathsf{G}_\Lambda \in \mathbb{R}^{n \times n}$ as

$$\Lambda = \{\boldsymbol{a}\mathsf{G}_\Lambda : \boldsymbol{a} \in \mathbb{Z}^n\}.$$

That is, Λ is the set of all integer combinations of the rows of G_Λ.

A *nearest neighbour quantizer* $Q_\Lambda : \mathbb{R}^n \to \Lambda$ associated with the lattice Λ maps a vector in $\mathbb{R}^n$ to the closest lattice point

$$Q_\Lambda(\boldsymbol{x}) = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \|\boldsymbol{x} - \boldsymbol{\lambda}\|, \tag{7.1}$$

where ties in (7.1) are broken systematically. The *Voronoi region* of Λ, denoted by $\mathcal{V}(\Lambda)$, is the set of all vectors in $\mathbb{R}^n$ which are quantized to $\mathbf{0}$, i.e., $\mathcal{V}(\Lambda) = \{\boldsymbol{x} \in \mathbb{R}^n : Q_\Lambda(\boldsymbol{x}) = \mathbf{0}\}$. The volume of the Voronoi region is denoted by $V(\Lambda)$.

A *nested lattice* is a pair of lattices (Λ_c, Λ_f) such that $\Lambda_c \subset \Lambda_f$. Similar to nested linear codes, Λ_f is called the *fine lattice* and Λ_c is called the *coarse lattice*. A coset of Λ_c in Λ_f is defined as

$$\boldsymbol{\lambda}_f + \Lambda_c = \{\boldsymbol{\lambda}_f + \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_c\}.$$

A *nested lattice code* $\mathcal{L}(\Lambda_c, \Lambda_f)$ consists of the lattice points of Λ_f in the Voronoi region $\mathcal{V}(\Lambda_c)$, i.e.,

$$\mathcal{L}(\Lambda_c, \Lambda_f) = \Lambda_f \cap \mathcal{V}(\Lambda_c).$$

For this reason, $\mathcal{L}(\Lambda_c, \Lambda_f)$ is also known as a Voronoi codebook. The number of codewords in $\mathcal{L}(\Lambda_c, \Lambda_f)$ is

$$|\mathcal{L}(\Lambda_c, \Lambda_f)| = \frac{V(\Lambda_c)}{V(\Lambda_f)}.$$

Intuitively, each lattice point of Λ_f "occupies" a Voronoi region of volume $V(\Lambda_f)$, and so the number of lattice points inside $\mathcal{V}(\Lambda_c)$ is $V(\Lambda_c)/V(\Lambda_f)$.

There is an alternative characterization of nested lattice codes: $\mathcal{L}(\Lambda_c, \Lambda_f)$ *consists of the shortest vectors of distinct cosets.* To see this, for each coset $\boldsymbol{\lambda}_f + \Lambda_c$, let us take a particular coset representative $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$. First, $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$ is the shortest vector in the coset $\boldsymbol{\lambda}_f + \Lambda_c$ by the definition of $Q_{\Lambda_c}(\cdot)$. Second, $\boldsymbol{\lambda}_f - Q_{\Lambda_c}(\boldsymbol{\lambda}_f)$ is in the Voronoi region $\mathcal{V}(\Lambda_c)$ of Λ_c (Fig. 7.2).

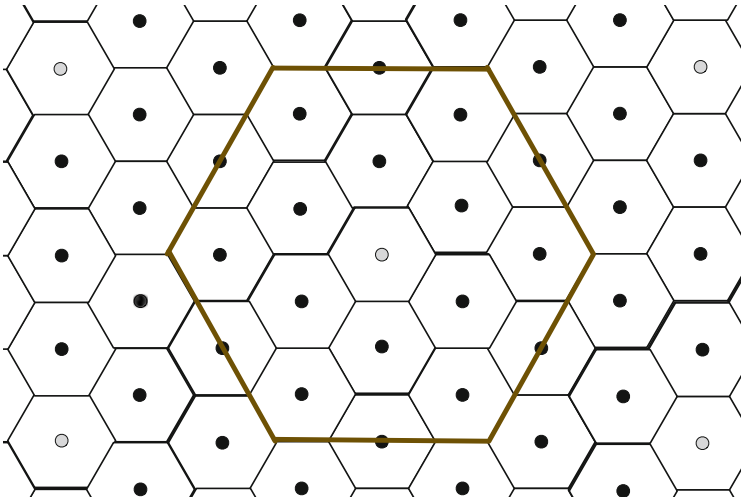


Fig. 7.2 Black (grey) points belong to the fine (coarse) lattice. The small (large) hexagon area is the Voronoi region of the fine (coarse) lattice. The lattice points inside the large hexagon form the Voronoi codebook (the ties on the boundaries are broken systematically). There are 16 lattice points in the codebook due to the tie breaking. Also note that the volume of the large hexagon is 16 times the volume of the small one

7.2.3 Nested Construction A

A nested lattice code can be constructed from a nested linear code. Consider two linear codes C_1 and C_2 over the field $\mathbb{Z}_\mathsf{p} = \{0, 1, \ldots, \mathsf{p}-1\}$, where each code C_i is determined by a (full-rank) $k_i \times n$ generator matrix G_i for $i = 1, 2$. Suppose that the generator matrices are related as

$$\mathsf{G}_1 = \begin{bmatrix} \mathsf{G}_2 \\ \mathsf{G}' \end{bmatrix}, \tag{7.2}$$

where G' is a matrix of size $(k_1 - k_2) \times n$. Clearly, we have $\mathsf{C}_2 \subset \mathsf{C}_1 \subset \mathbb{Z}_\mathsf{p}^n$. By "lifting" these linear codes to $\mathbb{Z}^n$ via Construction A, we obtain two lattices

$$\Lambda_1 = \{\boldsymbol{x} \in \mathbb{Z}^n : \boldsymbol{x} \bmod \mathsf{p} \in \mathsf{C}_1\}$$

and

$$\Lambda_2 = \{\boldsymbol{x} \in \mathbb{Z}^n : \boldsymbol{x} \bmod \mathsf{p} \in \mathsf{C}_2\}$$

with $\Lambda_2 \subset \Lambda_1 \subset \mathbb{Z}^n$.

Finally, we apply some positive scaling factor γ to obtain a fine lattice

$$\Lambda_f = \gamma\Lambda_1 \triangleq \{\gamma\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_1\}$$

and a coarse lattice

$$\Lambda_c = \gamma\Lambda_2 \triangleq \{\gamma\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_2\}$$

with $\Lambda_c \subset \Lambda_f \subset \gamma\mathbb{Z}^n$. The volumes of the Voronoi regions of Λ_f and Λ_c are $V(\Lambda_f) = \gamma^n \mathsf{p}^{n-k_1}$ and $V(\Lambda_c) = \gamma^n \mathsf{p}^{n-k_2}$, respectively.

To facilitate encoding and decoding operations, we "label" each (discrete) point of $\gamma\mathbb{Z}^n$ as follows. Let $\varphi : \gamma\mathbb{Z}^n \to \mathbb{Z}_\mathbb{P}^n$ be a map from points in $\gamma\mathbb{Z}^n$ to vectors in $\mathbb{Z}_\mathbb{P}^n$ given by

$$\varphi(\boldsymbol{x}) = \frac{1}{\gamma}\boldsymbol{x} \bmod \mathsf{p}.$$

Clearly, a point $\boldsymbol{x}$ is in Λ_f (or Λ_c, respectively) if and only if its label $\varphi(\boldsymbol{x})$ is a codeword in C_1 (or C_2, respectively). Moreover, the map φ is homomorphic, i.e.,

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \gamma\mathbb{Z}^n,\ \varphi(\boldsymbol{x} + \boldsymbol{y}) = \varphi(\boldsymbol{x}) + \varphi(\boldsymbol{y}).$$

It is also convenient to define an inverse operation that maps a vector in $\mathbb{Z}_\mathbb{P}^n$ to a point in $\gamma\mathbb{Z}^n$. This can be done through an embedding map $\tilde{\varphi} : \mathbb{Z}_\mathbb{P}^n \to \gamma\mathbb{Z}^n$: for

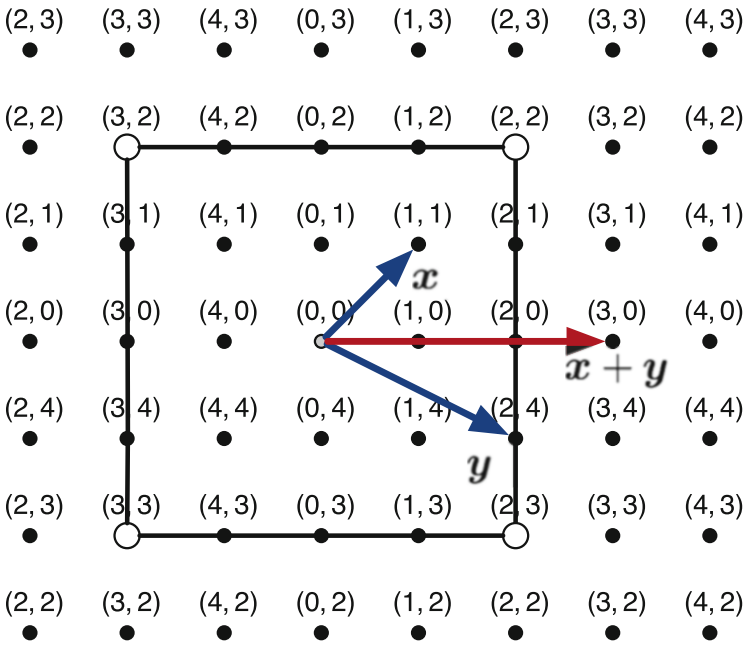


Fig. 7.3 A visualization of $\varphi(\cdot)$ when $\mathsf{p} = 5$. The labels of the points in $\gamma\mathbb{Z}^n$ can be obtained by periodically shifting the labels in the rectangle

any $\boldsymbol{c}$ in $\mathbb{Z}_{\mathbb{P}}^n$, we choose a point $\boldsymbol{x}$ in $\gamma\mathbb{Z}^n$ of the shortest Euclidean norm such that $\varphi(\boldsymbol{x}) = \boldsymbol{c}$. Clearly, such a point $\boldsymbol{x} = \tilde{\varphi}(\boldsymbol{c})$ must live in the grid $\gamma\mathbb{Z}^n \cap [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$ (Fig. 7.3).

In fact, the embedding map $\tilde{\varphi}$ can be viewed as a *Euclidean embedding* for the vector space $\mathbb{Z}_{\mathbb{P}}^n$, which connects the nested lattice codes with the underling nested linear codes.

7.2.4 Results from Number Theory

Several results from number theory will be used in this chapter and they are listed below.

Let $\mathbf{G}$ be a random matrix uniform over $\mathbb{Z}_{\mathbb{P}}^{k\times n}$, i.e., each entry of $\mathbf{G}$ is drawn uniformly and independently from $\mathbb{Z}_{\mathbb{P}}$.

Lemma 7.1 (Uniformity) *For any fixed non-zero vector $\boldsymbol{a}$, $\boldsymbol{a}\mathbf{G}$ is uniform over $\mathbb{Z}_{\mathbb{P}}^n$.*

Proof We leave it as an exercise to our readers. □

Lemma 7.2 (Linear Independence ⇒ Statistical Independence) *For any linearly independent vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the random vectors $\boldsymbol{a}\mathsf{G}$ and $\boldsymbol{b}\mathsf{G}$ are statistically independent.*

Proof Since $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly independent, there exists a full rank matrix $\mathsf{A} \in \mathbb{Z}_{\mathbb{P}}^{k\times n}$ whose first row vector is $\boldsymbol{a}$, and the second row vector is $\boldsymbol{b}$, i.e., $\boldsymbol{e}_1\mathsf{A} = \boldsymbol{a}, \boldsymbol{e}_2\mathsf{A} = \boldsymbol{b}$. For any fixed vectors $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{Z}_{\mathbb{P}}^{k}$, $\boldsymbol{e}_1\mathsf{AG} = \boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1$ and $\boldsymbol{e}_2\mathsf{AG} = \boldsymbol{b}\mathsf{G} = \boldsymbol{c}_2$, if and only if the first and second row vector of AG are $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. Let $S_{\boldsymbol{c}_1,\boldsymbol{c}_2} = \{\mathsf{B} \in \mathbb{Z}_{\mathbb{P}}^{k\times n} \mid \boldsymbol{e}_1\mathsf{B} = \boldsymbol{c}_1, \boldsymbol{e}_2\mathsf{B} = \boldsymbol{c}_2\}$, then $|S_{\boldsymbol{c}_1,\boldsymbol{c}_2}| = \mathsf{p}^{(k-2)n}$. Hence

$$\mathsf{P}(\boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1, \boldsymbol{b}\mathsf{G} = \boldsymbol{c}_2) = \sum_{\mathsf{B}\in S_{\boldsymbol{c}_1,\boldsymbol{c}_2}} \mathsf{P}(\mathbf{G} = \mathsf{A}^{-1}\mathsf{B}) = \frac{1}{p^{2n}}$$

Hence, $\mathsf{P}(\boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1, \boldsymbol{b}\mathsf{G} = \boldsymbol{c}_2) = \mathsf{P}(\boldsymbol{a}\mathsf{G} = \boldsymbol{c}_1)\,\mathsf{P}(\boldsymbol{b}\mathsf{G} = \boldsymbol{c}_2)$, which means $\boldsymbol{a}\mathsf{G}$ and $\boldsymbol{b}\mathsf{G}$ are statistically independent. □

Lemma 7.3 (Crypto Lemma) *Let Λ be a lattice. Let $\boldsymbol{D}$ be a random variable uniformly distributed over $\mathcal{V}(\Lambda)$. Let $\boldsymbol{T}$ be a random variable over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{D}$, then $\boldsymbol{X} = \boldsymbol{D} + \boldsymbol{T} \bmod \Lambda$ is uniformly distributed over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{T}$.*

Remark 7.1 This lemma is a discrete parallel of [11, Lemma 1].

Proof Note that $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t}) = \mathsf{P}(\boldsymbol{D} = [\boldsymbol{x} - \boldsymbol{t}] \bmod \Lambda \mid \boldsymbol{T} = \boldsymbol{t})$. By the fact that $\boldsymbol{D}$ and $\boldsymbol{T}$ are independent, we obtain $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t}) = \mathsf{P}(\boldsymbol{D} = [\boldsymbol{x} - \boldsymbol{t}] \bmod \Lambda)$. Since $\boldsymbol{D}$ is uniform over $\mathcal{V}(\Lambda)$, $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{T} = \boldsymbol{t})$ is constant for all possible combinations of $\boldsymbol{x}$ and $\boldsymbol{t}$. Hence, $\boldsymbol{X}$ is uniformly distributed over $\mathcal{V}(\Lambda)$, and is independent from $\boldsymbol{T}$. □

Let $\mathcal{B}(\boldsymbol{s}, r)$ denote a ball of radius $r > 0$ centered at the point $\boldsymbol{s} \in \mathbb{R}^n$, i.e., $\mathcal{B}(\boldsymbol{s}, r)$ is the set $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{s}\| \le r\}$. For convenience, we denote $\mathcal{B}(\mathbf{0}, r)$ as $\mathcal{B}(r)$. The volume of $\mathcal{B}(r)$ is given by $r^n V_n$, where V_n is the volume of the unit-radius ball.

Lemma 7.4 (Integer Points Inside a Ball [33, Lemma 1]) *For any $\boldsymbol{s} \in \mathbb{R}^n$, the number of points of $\mathbb{Z}^n$ inside $\boldsymbol{s} + \mathcal{B}(r)$ can be bounded as*

$$V_n\left(\max\left\{r - \frac{\sqrt{n}}{2}, 0\right\}\right)^n \le |\mathbb{Z}^n \cap \mathcal{B}(\boldsymbol{s}, r)| \le V_n\left(r + \frac{\sqrt{n}}{2}\right)^n.$$

Lemma 7.5 (Bertrand's Postulate [43]) *For any integer n that is larger than 3, there exists a prime p such that $n < \mathsf{p} < 2n - 2$.*

7.3 Achievable Rate of Nested Linear Codes

7.3.1 Performance Analysis of a Nested Linear Code

We begin with the analysis of a (pre-determined) nested linear code.

Codebook Generation Given a pair of linear codes $(\mathsf{C}_f, \mathsf{C}_c)$ and a dither vector $\boldsymbol{d} \in \mathbb{F}_{\mathsf{q}}^n$, we construct a codebook whose codewords are shifted cosets of the form $\{\boldsymbol{c}_f + \boldsymbol{d} + \mathsf{C}_c : \boldsymbol{c}_f \in \mathsf{C}_f\}$. The number of (distinct) codewords is $[\mathsf{C}_f : \mathsf{C}_c]$, which does not depend on the dither vector $\boldsymbol{d}$. These codewords can be expressed using generator matrices as follows.

Let $\mathsf{G}_f \in \mathbb{F}_{\mathsf{q}}^{k_f \times n}$ and $\mathsf{G}_c \in \mathbb{F}_{\mathsf{q}}^{k_c \times n}$ be two generator matrices for C_f and C_c, respectively, such that

$$\mathsf{G}_f = \begin{bmatrix} \mathsf{G}_c \\ \mathsf{G}' \end{bmatrix}.$$

Then all the codewords (i.e., the shifted cosets) can be expressed as

$$\left\{ \boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c} \right\}.$$

Note that there is a one-to-one correspondence between the vectors in $\mathbb{F}_{\mathsf{q}}^{k_f - k_c}$ and the shifted cosets of C_c. Hence, $\boldsymbol{m}$ can be viewed as the "index" of the shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$, and the codebook contains $\mathsf{q}^{k_f - k_c}$ (distinct) codewords.

Encoding To send a message vector $\boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c}$, the encoder first finds an "information-carrying" shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$. The encoder then checks the intersection

$$\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X).$$

If the intersection is nonempty, the encoder transmits a vector $x \in \mathbb{F}_{\mathsf{q}}^n$ chosen uniformly at random from the intersection. Otherwise, the encoder declares a failure and then transmits a vector $\boldsymbol{x} \in \mathbb{F}_{\mathsf{q}}^n$ chosen uniformly at random from the shifted coset $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c$ (which is not in $\mathcal{T}_{\epsilon'}^{(n)}(X)$).

Decoding Upon receiving $\boldsymbol{y} \in \mathbb{F}_{\mathsf{q}}^n$, the decoder searches for a unique index $\hat{\boldsymbol{m}} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c}$ such that the corresponding shifted coset

$$\hat{\boldsymbol{m}}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) \neq \emptyset.$$

If there is none or more than one such vector, the decoder declares a failure.

Analysis For any given message vector $\boldsymbol{m}$, we say the decoding is successful if the unique index $\hat{\boldsymbol{m}} = \boldsymbol{m}$. This occurs if all of the following events happen

- $\boldsymbol{m}\mathsf{G}' + \boldsymbol{D} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) \neq \emptyset$;
- $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)$ (which implies that $\boldsymbol{m}\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) \neq \emptyset$);
- $\forall \boldsymbol{m}' \neq \boldsymbol{m} : \boldsymbol{m}'\mathsf{G}' + \boldsymbol{d} + \mathsf{C}_c \cap \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) = \emptyset$.

7.3.2 Average Performance Analysis of Nested Linear Codes

We then proceed to the average performance analysis, which allows us to apply the probabilistic method.

Random Codebook Generation Randomly generate a matrix $\mathbf{G}_f \in \mathbb{F}_{\mathsf{q}}^{k_f \times n}$ and a vector $\boldsymbol{D} \in \mathbb{F}_{\mathsf{q}}^{n}$ where each entry of $\mathbf{G}_f$ and $\boldsymbol{D}$ is drawn independently and uniformly from $\mathbb{F}_{\mathsf{q}}$. As before, let

$$\mathbf{G}_f = \begin{bmatrix} \mathbf{G}_c \\ \mathbf{G}' \end{bmatrix}.$$

If $\mathbf{G}_f$ is full rank, then $\mathbf{G}_c$ is also full rank and, in particular, they are valid generator matrices. In this case, the codebook consists of $\mathsf{q}^{k_f - k_c}$ shifted cosets of the form

$$\left\{ \boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathbf{C}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{q}}^{k_f - k_c} \right\}.$$

If $\mathbf{G}_f$ is not full rank, we declare a codebook failure.

Encoding The same as before.

Decoding The same as before.

Analysis of the Probability of Error For any given message vector $\boldsymbol{m}$, successful decoding occurs upon receiving $\boldsymbol{Y}$ if all of the following events happen

- $\mathbf{G}_f$ is full rank;
- $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathbf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) \neq \emptyset$;
- $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(\boldsymbol{X}, \boldsymbol{Y})$;
- $\forall \boldsymbol{m}' \neq \boldsymbol{m}, \boldsymbol{l} : (\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y)$.

To conduct the error analysis, we define the following events

- $\mathcal{E}_1 = \{\mathbf{G}_f \text{ is not full rank}\}$;
- $\mathcal{E}_2(\boldsymbol{m}) = \{\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathbf{C}_c \cap \mathcal{T}_{\epsilon'}^{(n)}(X) = \emptyset\}$;
- $\mathcal{E}_3(\boldsymbol{m}) = \{(\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y)\}$;
- $\mathcal{E}_4(\boldsymbol{m}) = \{\exists \boldsymbol{m}' \neq \boldsymbol{m}, \boldsymbol{l} : \ (\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)\}$.

Let $\mathsf{P}_e(\boldsymbol{m})$ be the error probability for message $\boldsymbol{m}$. Then, by the union bound, we have

$$\mathsf{P}_e(\boldsymbol{m}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) + \mathsf{P}(\mathcal{E}_3(\boldsymbol{m})) + \mathsf{P}(\mathcal{E}_4(\boldsymbol{m})).$$

7.3.2.1 Bounding $\mathsf{P}(\mathcal{E}_1)$

Note that $\mathbf{G}_f$ is full rank if and only if the rows of $\mathbf{G}_f$ are linearly independent. Hence, we have

$$\mathsf{P}(\mathcal{E}_1) = 1 - \prod_{i=0}^{k_f-1} \left(1 - \frac{\mathsf{q}^i}{\mathsf{q}^n}\right).$$

Moreover, we have

$$\begin{aligned}\prod_{i=0}^{k_f-1} \left(1 - \frac{\mathsf{q}^i}{\mathsf{q}^n}\right) &\ge 1 - \sum_{i=0}^{k_f-1} \frac{\mathsf{q}^i}{\mathsf{q}^n} \\ &= 1 - \frac{1}{\mathsf{q}^n}\frac{\mathsf{q}^{k_f}-1}{\mathsf{q}-1} \\ &\ge 1 - \frac{1}{\mathsf{q}-1}\frac{1}{\mathsf{q}^{n-k_f}}.\end{aligned}$$

This implies that $\mathsf{P}(\mathcal{E}_1) \le \frac{1}{\mathsf{q}-1}\frac{1}{\mathsf{q}^{n-k_f}}$. Hence, $\mathsf{P}(\mathcal{E}_1) \to 0$ as $\mathsf{p} \to \infty$ or $(n-k_f) \to \infty$.

7.3.2.2 Bounding $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m}))$

Note that $\mathcal{E}_2(\boldsymbol{m})$ is equivalent to

$$\sum_{\boldsymbol{l} \in \mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right) = 0.$$

Since $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ is uniformly distributed over $\mathbb{F}_{\mathsf{q}}^n$, we have

$$\mathsf{E}\left(\mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}$$

and

$$\mathrm{Var}\left(\mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\left(1 - \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\right).$$

Note that for any $\boldsymbol{l}' \neq \boldsymbol{l}$, $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}'\mathbf{G}_c$ and $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ are independent. Hence,

$$\mathsf{E}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \mathsf{q}^{k_c}\frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}$$

and

$$\mathrm{Var}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right) = \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\mathsf{q}^{k_c}\left(1 - \frac{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}{\mathsf{q}^n}\right).$$

Finally, by Chebyshev's inequality, we have

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) &= \mathsf{P}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right) = 0\right) \\
&\leq \frac{\mathrm{Var}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right)}{\mathsf{E}\left(\sum_{\boldsymbol{l}\in\mathbb{F}_{\mathsf{q}}^{k_c}} \mathbb{I}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon'}^{(n)}(X)\right)\right)^2} \\
&\leq \frac{\mathsf{q}^{n-k_c}}{|\mathcal{T}_{\epsilon'}^{(n)}(X)|}.
\end{aligned}$$

7.3.2.3 Bounding $\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}))$

By the law of total probability, we have

$$\begin{aligned}
&\mathsf{P}\left((\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y)\right) \\
&= \mathsf{P}(\boldsymbol{X} \in \mathcal{T}_{\epsilon'}^{(n)}(X))\,\mathsf{P}((\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y) | \boldsymbol{X} \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \\
&\quad + \mathsf{P}(\boldsymbol{X} \notin \mathcal{T}_{\epsilon'}^{(n)}(X))\,\mathsf{P}((\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y) | \boldsymbol{X} \notin \mathcal{T}_{\epsilon'}^{(n)}(X)) \\
&\leq \mathsf{P}((\boldsymbol{X}, \boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X, Y) | \boldsymbol{X} \in \mathcal{T}_{\epsilon'}^{(n)}(X)) + \mathsf{P}(\boldsymbol{X} \notin \mathcal{T}_{\epsilon'}^{(n)}(X)).
\end{aligned}$$

By the conditional typicality lemma [9, p. 27], $\mathsf{P}((\boldsymbol{X},\boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X,Y)|\boldsymbol{X} \in \mathcal{T}_{\epsilon'}^{(n)}(X)) \to 0$, as $n \to \infty$. Finally, note that $\boldsymbol{X} \notin \mathcal{T}_{\epsilon'}^{(n)}(X)$ is equivalent to the event $\mathcal{E}_2(\boldsymbol{m})$. Hence, we obtain $\mathsf{P}\left((\boldsymbol{X},\boldsymbol{Y}) \notin \mathcal{T}_{\epsilon}^{(n)}(X,Y)\right) \to 0$, as long as $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) \to 0$.

7.3.2.4 Bounding $\mathsf{P}(\mathcal{E}_4(\boldsymbol{m}))$

By the union of events bound, we have

$$\mathsf{P}(\mathcal{E}_4(\boldsymbol{m})) \le \sum_{\boldsymbol{m}' \neq \boldsymbol{m}} \sum_{\boldsymbol{l}} \mathsf{P}((\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(X,Y)).$$

For each term, by the law of total probability, we have

$$\mathsf{P}((\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c, \boldsymbol{Y}) \in \mathcal{T}_{\epsilon}^{(n)}(X,Y)) = \sum_{\boldsymbol{y}} \mathsf{P}(\boldsymbol{Y} = \boldsymbol{y})\, \mathsf{P}\left(\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y}) \middle| \boldsymbol{Y} = \boldsymbol{y}\right).$$

Note that, for any $\boldsymbol{m}' \neq \boldsymbol{m}$ and any $\boldsymbol{l}$, the random vector $\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ is independent of the random shifted coset $\boldsymbol{m}\mathbf{G}' + \boldsymbol{D} + \mathbf{C}_c$. This implies that $\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c$ is independent of $\boldsymbol{Y}$. Hence,

$$\mathsf{P}(\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})|\boldsymbol{Y} = \boldsymbol{y}) = \mathsf{P}(\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})).$$

Since $\mathsf{P}(\boldsymbol{m}'\mathbf{G}' + \boldsymbol{D} + \boldsymbol{l}\mathbf{G}_c \in \mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})) = \frac{|\mathcal{T}_{\epsilon}^{(n)}(X|\boldsymbol{y})|}{\mathsf{q}^n}$, we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_4(\boldsymbol{m}) \mid \boldsymbol{Y} = \boldsymbol{y}) &\le \left(\mathsf{q}^{k_f - k_c} - 1\right) \mathsf{q}^{k_c} \frac{|\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})|}{\mathsf{q}^n} \\ &< \mathsf{q}^{k_f} \frac{|\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})|}{\mathsf{q}^n}.\end{aligned}$$

Hence, we have

$$\mathsf{P}(\mathcal{E}_4(\boldsymbol{m})) \le \sum_{\boldsymbol{y}} \mathsf{P}(\boldsymbol{Y} = \boldsymbol{y}) \frac{|\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})|}{\mathsf{q}^{n-k_f}}.$$

7.3.2.5 Putting Everything Together

Our goal is to select k_c and k_f (as functions of n) such that

$$n - k_f \to \infty \tag{7.3}$$

$$\frac{\mathsf{q}^{n-k_c}}{|\mathcal{T}_{\epsilon'}^{(n)}(X)|} \to 0 \tag{7.4}$$

$$\forall \boldsymbol{y} : \frac{|\mathcal{T}_{\epsilon}^{(n)}(X \mid \boldsymbol{y})|}{\mathsf{q}^{n-k_f}} \to 0. \tag{7.5}$$

Let $\delta > 0$ be some constant. We choose $\mathsf{q}^{n-k_c} = 2^{n(1-\epsilon'-\delta)H(X)}$ and $\mathsf{q}^{n-k_f} = 2^{n(1+\epsilon+\delta)H(X|Y)}$. More precisely, we choose

$$k_c = \left\lceil n - \frac{(1-\epsilon'-\delta)H(X)}{\log_2 \mathsf{q}} n \right\rceil$$

and

$$k_f = \left\lfloor n - \frac{(1+\epsilon+\delta)H(X|Y)}{\log_2 \mathsf{q}} n \right\rfloor.$$

We can easily verify that conditions (7.3), (7.4) are satisfied. The inequality (7.5) is also satisfied by (A.1) in Appendix 2. Finally, we calculate the achievable rate

$$\frac{1}{n}\log_2 \mathsf{q}^{k_f-k_c} \geq I(X;Y) - (\epsilon'+\delta)H(X) - (\epsilon+\delta)H(X|Y) - 2\frac{\log_2 \mathsf{q}}{n}.$$

Since ϵ, ϵ' and δ can be arbitrarily small, any rate below $I(X;Y)$ is achievable as $n \to \infty$.

7.4 Achievable Rate of Nested Lattice Codes

7.4.1 Performance Analysis of a Nested Lattice Code

Codebook Generation Given a pair of lattice codes (Λ_f, Λ_c) and a dither vector $\boldsymbol{u} \in \mathbb{R}^n$, we construct a codebook whose codewords are shifted cosets of the form $\{\boldsymbol{\lambda}_f + \boldsymbol{u} + \Lambda_c : \boldsymbol{\lambda}_f \in \Lambda_c\}$. The number of codewords is $V(\Lambda_c)/V(\Lambda_f)$, which does not depend on the dither vector $\boldsymbol{u}$.

Suppose that the pair (Λ_f, Λ_c) is constructed via Nested Construction A using generating matrices $(\mathsf{G}_f, \mathsf{G}_c)$ and a scaling factor γ. Then all the codewords (i.e., the shifted cosets) can be expressed as

$$\left\{\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}\right\}.$$

Note that there is a one-to-one correspondence between the vectors in $\mathbb{F}_{\mathsf{p}}^{k_f - k_c}$ and the shifted cosets of Λ_c. Hence, $\boldsymbol{m}$ can be viewed as the "index" of the shifted coset $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$, and the codebook contains $\mathsf{p}^{k_f - k_c}$ (distinct) codewords.

Encoding To send a message vector $\boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}$, the encoder first finds an "information-carrying" shifted coset $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$. The encoder then transmits a shortest vector $\boldsymbol{x} \in \mathbb{R}^n$ in the shifted coset, i.e.,

$$\boldsymbol{x} = \tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} \mod \Lambda_c.$$

Decoding Upon receiving $\boldsymbol{y} \in \mathbb{R}^n$, the decoder searches for a unique index $\hat{\boldsymbol{m}} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c}$ such that the distance between its corresponding shifted coset $\tilde{\varphi}(\hat{\boldsymbol{m}}\mathsf{G}') + \boldsymbol{u} + \Lambda_c$ and $\alpha \boldsymbol{y}$ is the shortest among all the shifted cosets, where $\alpha = \frac{P}{P+N}$ is some scaling factor (whose role will be explained later). P and N are the average power of the codeword and the noise per dimension, respectively. That is,

$$\hat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha \boldsymbol{y}\right).$$

In fact, one can easily show that the unique shifted coset with the shortest distance is given by $Q_{\Lambda_f}(\alpha \boldsymbol{y} - \boldsymbol{u}) + \boldsymbol{u} + \Lambda_c$ (Fig. 7.4).

Analysis For any given message vector $\boldsymbol{m}$, the average power constraint is satisfied if

- $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c \cap \mathcal{B}\left(\sqrt{nP}\right) \neq \emptyset$;

The decoding is successful if

- $\forall \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha \boldsymbol{y}\right) > \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \boldsymbol{u} + \Lambda_c, \alpha \boldsymbol{y}\right)$.

7.4.2 Average Performance Analysis of a Nested Lattice Codes

We then proceed to the average performance analysis, which also allows us to apply probabilistic methods.

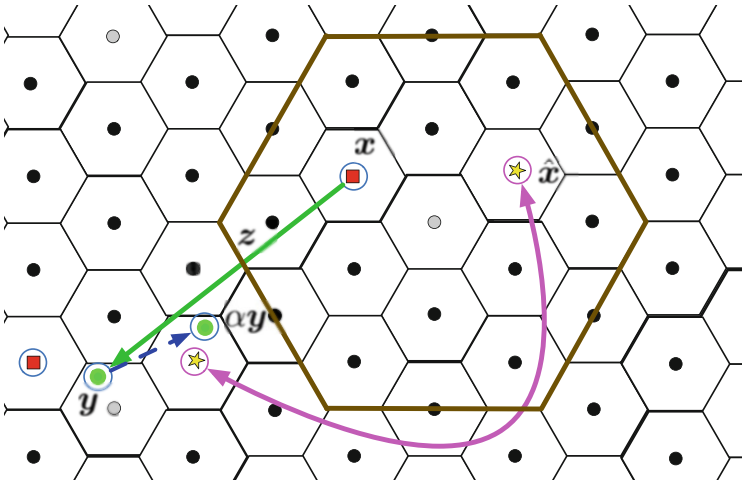


Fig. 7.4 The transmitted vector is $\boldsymbol{x}$, which is then "shifted" by the Gaussian noise $\boldsymbol{z}$ to $\boldsymbol{y}$. The received signal $\boldsymbol{y}$ is scaled by α to $\alpha\boldsymbol{y}$. The decoder will find the nearest coset to $\alpha\boldsymbol{y}$. In this example, the nearest coset to $\alpha\boldsymbol{y}$ is the coset containing $\hat{\boldsymbol{x}}$ (the star points) instead of the one containing $\boldsymbol{x}$ (the rectangle points). Hence, a decoding failure happens

Random Codebook Generation Randomly generate a matrix $\mathbf{G}_f \in \mathbb{Z}_{\mathbb{P}}^{k_f \times n}$ and a vector $\boldsymbol{U} \in \mathbb{Z}_{\mathbb{P}}^{n}$ where each entry of $\mathbf{G}_f$ and $\boldsymbol{U}$ is drawn independently and uniformly over $\mathbb{Z}_{\mathbb{P}}$. As before, let

$$\mathbf{G}_f = \begin{bmatrix} \mathbf{G}_c \\ \mathbf{G}' \end{bmatrix},$$

and if $\mathbf{G}_f$ is full rank, so is $\mathbf{G}_c$. In this case, the codebook consists of $\mathsf{p}^{k_f - k_c}$ shifted cosets of the form

$$\left\{ \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c : \boldsymbol{m} \in \mathbb{F}_{\mathsf{p}}^{k_f - k_c} \right\}.$$

If $\mathbf{G}_f$ is not full rank, we declare a codebook failure.

Encoding The same as before.

Decoding The same as before.

Analysis of the Codebook Failure Let $\mathcal{E}_1 = \{\mathbf{G}_f \text{ is not full rank}\}$. As before

$$\mathsf{P}(\mathcal{E}_1) \le \frac{1}{\mathsf{p}-1} \frac{1}{\mathsf{p}^{n-k_f}}.$$

Hence, $\mathsf{P}(\mathcal{E}_1) \to 0$, as $\mathsf{p} \to \infty$ or $(n - k_f) \to \infty$.

Analysis of Encoding Failure Recall that $\|\boldsymbol{X}\|^2 \leq nP$ if and only if $\tilde{\varphi}(\boldsymbol{m}\mathsf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c \cap \mathcal{B}\left(\sqrt{nP}\right) \neq \emptyset$, where $\mathcal{B}\left(\sqrt{nP}\right)$ is the ball centred at the origin with radius $\sqrt{nP}$. Let

$$\mathcal{E}_2(\boldsymbol{m}) = \{\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c \cap \mathcal{B}\left(\sqrt{nP}\right) = \emptyset\}.$$

We will show that $\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) \to 0$ under certain condition.

Note that when $\mathcal{B}\left(\sqrt{nP}\right) \subset [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$, $\mathcal{E}_2(\boldsymbol{m})$ is equivalent to

$$\sum_{\boldsymbol{l}\in\mathbb{Z}_{\mathbb{P}}^{k_c}} \mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathsf{G}' + \boldsymbol{u} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right) = 0,$$

because the set $\{\tilde{\varphi}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{u} + \boldsymbol{l}\mathbf{G}_c\right) : \boldsymbol{l} \in \mathbb{Z}_{\mathbb{P}}^{k_c}\}$ generates all the points of $\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c$ inside the cube $[-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$.

Since $\tilde{\varphi}\left(\boldsymbol{m}\mathbf{G}' + \boldsymbol{u} + \boldsymbol{l}\mathbf{G}_c\right)$ is uniformly distributed over the grid $\gamma\mathbb{Z}^n \cap [-\frac{\gamma\mathsf{p}}{2}, \frac{\gamma\mathsf{p}}{2}]^n$, we have

$$\mathsf{E}\left(\mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}' + \boldsymbol{u} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right)\right) = \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}$$

and

$$\begin{aligned}&\mathrm{Var}\left(\mathbb{I}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}' + \boldsymbol{u} + \boldsymbol{l}\mathbf{G}_c) \in \mathcal{B}\left(\sqrt{nP}\right)\right)\right)\\ &= \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}\left(1 - \frac{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}{\mathsf{p}^n}\right).\end{aligned}$$

Similar to the case of nested linear codes, we have

$$\mathsf{P}(\mathcal{E}_2(\boldsymbol{m})) \leq \frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}. \tag{7.6}$$

Analysis of the Decoding Failure Recall that successful decoding occurs upon receiving $\boldsymbol{Y}$ if

$$\forall \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) > \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right).$$

Let

$$\begin{aligned}\mathcal{E}_3(\boldsymbol{m}) = \{&\exists \boldsymbol{m}' \neq \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha \boldsymbol{Y}\right) \\ &\leq \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha \boldsymbol{Y}\right)\}.\end{aligned}$$

Recall that $\boldsymbol{X} = \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) \mod \boldsymbol{\Lambda}_c$, and, in particular, $\boldsymbol{X} \in \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c$. Hence,

$$\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha \boldsymbol{Y}\right) \leq \|\boldsymbol{X} - \alpha \boldsymbol{Y}\| = \|(\alpha - 1)\boldsymbol{X} + \alpha \boldsymbol{Z}\|.$$

Let $\boldsymbol{W} = (\alpha - 1)\boldsymbol{X} + \alpha \boldsymbol{Z}$ be the "effective noise". By the Total Probability Theorem, we have

$$\begin{aligned}&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) | \mathbf{G}_c = \mathsf{G}_c) \\ \leq\ &\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e) | \mathbf{G}_c = \mathsf{G}_c) \\ &+ \mathsf{P}(\boldsymbol{W} \in \mathcal{B}(r_e) | \mathbf{G}_c = \mathsf{G}_c)\, \mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) | \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c),\end{aligned}$$

where $\mathcal{B}(r_e)$ is the "typical ball" for the effective noise $\boldsymbol{W}$ with radius r_e. It will be specified in Sect. 7.4.2.1. It follows that

$$\begin{aligned}\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})) \leq\ &\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e)) \\ &+ \sum_{\mathsf{G}_c} \mathsf{P}(\boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c)\, \mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) | \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c).\end{aligned}$$

7.4.2.1 Bounding $\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e))$

Let ϵ be a small positive constant. We set $\alpha = \frac{P}{P+N}$ and set the radius

$$\begin{aligned}r_e &= \sqrt{(1+\epsilon)n((\alpha-1)^2 P + \alpha^2 N)} \\ &= \sqrt{(1+\epsilon)\frac{nPN}{P+N}}.\end{aligned}$$

Let

$$\begin{aligned}\mathcal{E}_{\boldsymbol{X}} &= \{\|\boldsymbol{X}\| > \sqrt{nP}\}, \\ \mathcal{E}_{\boldsymbol{Z}} &= \{\|\boldsymbol{Z}\| > \sqrt{(1+\epsilon/2)nN}], \\ \mathcal{E}_P &= \{\|\boldsymbol{X}\boldsymbol{Z}^T\| > n^{\frac{1}{4}}\sqrt{nPN}\}.\end{aligned}$$

It is clear that when n is large, $\mathcal{E}_X^c \cap \mathcal{E}_Z^c \cap \mathcal{E}_P^c$ implies $\|\boldsymbol{W}\| \le r_e$. Hence,

$$\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e)) \le \mathsf{P}(\mathcal{E}_X) + \mathsf{P}(\mathcal{E}_Z) + \mathsf{P}(\mathcal{E}_P).$$

Note that $\mathcal{E}_X$ is the same event as $\mathcal{E}_2$, which is bounded via (7.6). Since $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, NI_n)$, we obtain $\mathsf{P}(\mathcal{E}_Z) \le 8\epsilon^2 n^{-1}$ by Chebyshev's inequality. $\mathcal{E}_P$ represents the event that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are "almost orthogonal". We bound its probability by

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_P) &\le \mathsf{P}(\mathcal{E}_P \mid \|\boldsymbol{X}\| \le \sqrt{nP}) + \mathsf{P}(\|\boldsymbol{X}\| > \sqrt{nP}) \\
&= \mathsf{P}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 > n^{\frac{3}{2}} PN \mid \|\boldsymbol{X}\| \le \sqrt{nP}) + \mathsf{P}(\mathcal{E}_2) \\
&\le \frac{\mathsf{E}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 \mid \|\boldsymbol{X}\| \le \sqrt{nP})}{n^{\frac{3}{2}} PN} + \mathsf{P}(\mathcal{E}_2)
\end{aligned}$$

where the last inequality follows from the Markov's inequality. Note that for any given $\boldsymbol{X} = \boldsymbol{x}$ with $\|\boldsymbol{x}\| \le \sqrt{nP}$, $\boldsymbol{x}\boldsymbol{Z}^T \sim \mathcal{N}(0, \|\boldsymbol{x}\|^2 N)$, we then obtain $\mathsf{E}(\|\boldsymbol{X}\boldsymbol{Z}^T\|^2 \mid \|\boldsymbol{X}\| \le \sqrt{nP}) \le nPN$. Hence,

$$\mathsf{P}(\mathcal{E}_P) \le n^{-\frac{1}{2}} + \mathsf{P}(\mathcal{E}_2).$$

Therefore,

$$\mathsf{P}(\boldsymbol{W} \notin \mathcal{B}(r_e)) \le 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}} + 2 \times \frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}.$$

7.4.2.2 Bounding $\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathsf{G}_c = \mathsf{G}_c)$

Note that

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m}) \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathsf{G}_c = \mathsf{G}_c) \\
&\le \mathsf{P}\left(\exists \boldsymbol{m}' \ne \boldsymbol{m} : \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \Lambda_c, \alpha\boldsymbol{Y}\right) \le \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathsf{G}_c = \mathsf{G}_c\right) \\
&\le \sum_{\boldsymbol{m}' \ne \boldsymbol{m}} \mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \Lambda_c, \alpha\boldsymbol{Y}\right) \le \|\boldsymbol{W}\| \mid \boldsymbol{W} \in \mathcal{B}(r_e), \mathbf{G}_c = \mathsf{G}_c\right).
\end{aligned}$$

Note also that

$$\begin{aligned}
&\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \alpha\boldsymbol{Y}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \boldsymbol{X} + (\alpha - 1)\boldsymbol{X} + \alpha\boldsymbol{Z}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') + \tilde{\varphi}(\boldsymbol{U}) + \boldsymbol{\Lambda}_c, \boldsymbol{X} + \boldsymbol{W}\right) \\
&= \mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}') - \tilde{\varphi}(\boldsymbol{m}\mathbf{G}') + \boldsymbol{\Lambda}_c, \boldsymbol{W}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c)\\
\le&\sum_{\boldsymbol{m}'\neq\boldsymbol{m}}\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{W}\right)\le\|\boldsymbol{W}\|\mid\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c\right)\\
\le&\sum_{\boldsymbol{m}'\neq\boldsymbol{m}}\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{W}\right)\le r_e\mid\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c\right).
\end{aligned}$$

Next, we observe that $\mathbf{G}'$ and $\boldsymbol{W}=(\alpha-1)\boldsymbol{X}+\alpha\boldsymbol{Z}$ are conditionally independent when given $\mathbf{G}_c=\mathsf{G}_c$. To see this, note that conditioned on $\mathbf{G}_c=\mathsf{G}_c$, $\boldsymbol{X}$ is uniformly distributed over $\gamma\mathbb{Z}^n\cap\mathcal{V}(\Lambda_c)$ and is independent of $\mathbf{G}'$ by Lemma 7.3. By the total probability theorem, we have

$$\begin{aligned}
&\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{W}\right)\le r_e|\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c\right)\\
=&\int_{\boldsymbol{w}\in\mathcal{B}(r_e)}\tilde{f}_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w}\mid\mathsf{G}_c)\,\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{w}\right)\le r_e|\mathbf{G}_c=\mathsf{G}_c\right)d\boldsymbol{w}
\end{aligned}$$

where

$$\tilde{f}_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w}\mid\mathsf{G}_c)=\frac{f_{\boldsymbol{W}|\mathsf{G}_c}(\boldsymbol{w}\mid\mathsf{G}_c)}{\mathsf{P}(\boldsymbol{W}\in\mathcal{B}(r_e)\mid\mathbf{G}_c=\mathsf{G}_c)}.$$

It turns out that the term $\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,w\right)\le r_e|\mathbf{G}_c=\mathsf{G}_c\right)$ can be bounded following Loeliger's approach [23].

Since $\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{w}\right)\le r_e$ implies

$$[\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')]\bmod\Lambda_c\in[\boldsymbol{w}+\mathcal{B}(r_e)]\bmod\Lambda_c,$$

we have

$$\begin{aligned}
&\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\Lambda_c,\boldsymbol{w}\right)\le r_e|\mathbf{G}_c=\mathsf{G}_c\right)\\
\le&\,\mathsf{P}\left([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')]\bmod\Lambda_c\in[\boldsymbol{w}+\mathcal{B}(r_e)]\bmod\Lambda_c\mid\mathbf{G}_c=\mathsf{G}_c\right).
\end{aligned}$$

On the other hand, $([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')]\bmod\Lambda_c)$ is uniformly distributed over $\gamma\mathbb{Z}^n\cap\mathcal{V}(\Lambda_c)$, and so

$$\begin{aligned}
&\mathsf{P}\left([\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')-\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')]\bmod\Lambda_c\in([\boldsymbol{w}+\mathcal{B}(r_e)]\bmod\Lambda_c)|\mathbf{G}_c=\mathsf{G}_c\right)\\
=&\frac{|\gamma\mathbb{Z}^n\cap\mathcal{V}(\Lambda_c)\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_c}}\\
\le&\frac{|\gamma\mathbb{Z}^n\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_c}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathsf{P}\left(\mathrm{d}\left(\tilde{\varphi}(\boldsymbol{m}'\mathbf{G}')+\tilde{\varphi}(\boldsymbol{U})+\Lambda_c,\alpha\boldsymbol{Y}\right)\le\|\boldsymbol{W}\|\mid\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c\right)\\&\le\max_{\boldsymbol{w}\in\mathcal{B}(r_e)}\frac{|\gamma\mathbb{Z}^n\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_c}}\end{aligned}$$

and

$$\begin{aligned}\mathsf{P}(\mathcal{E}_3(\boldsymbol{m})|\boldsymbol{W}\in\mathcal{B}(r_e),\mathbf{G}_c=\mathsf{G}_c)&\le\mathsf{p}^{k_f-k_c}\max_{\boldsymbol{w}\in\mathcal{B}(r_e)}\frac{|\gamma\mathbb{Z}^n\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_c}}\\&\le\max_{\boldsymbol{w}\in\mathcal{B}(r_e)}\frac{|\gamma\mathbb{Z}^n\cap(\boldsymbol{w}+\mathcal{B}(r_e))|}{\mathsf{p}^{n-k_f}}.\end{aligned}$$

7.4.3 Putting Everything Together

By the union bound, the error probability P_e of the coding scheme is bounded by

$$\mathsf{P}\le\mathsf{P}(\mathcal{E}_1)+\mathsf{P}(\mathcal{E}_2)+\mathsf{P}_e(\mathcal{E}_3),\tag{7.7}$$

because the decoding is successful if $\mathbf{G}_c$ is full rank, $\|\boldsymbol{X}\|^2\le nP$, and the shifted coset containing $\tilde{\varphi}(\boldsymbol{m}\mathbf{G}')+\tilde{\varphi}(\boldsymbol{U})$ is the closest coset to $\alpha\boldsymbol{Y}$. In Sect. 7.4.3.2, we will show that, for any $\epsilon>0$, we can select parameters $k_f,k_c,\mathsf{p},\gamma$ as functions of n such that a rate of

$$R=\frac{1}{2}\log_2\left(\frac{1+P/N}{1+\epsilon}\right)$$

is achievable with error probability $\mathsf{P}_e\to0$ as $n\to\infty$.

However, the above result *does not* imply that our random ensemble achieves the AWGN capacity, because the power constraint is not always satisfied. In fact, the power constraint is violated with probability $\mathsf{P}(\mathcal{E}_2)$. To address this issue, we introduce a spherical shaping strategy, which is in parallel with the minor change introduced in [9, p.47] for proving the channel coding theorem with input cost constraint.

7.4.3.1 Spherical Shaping

We apply a "truncated" spherical shaping to $\boldsymbol{X}$ as follows

$$\boldsymbol{X}_S=\begin{cases}\boldsymbol{X}, & \text{if } \|\boldsymbol{X}\|\le nP,\\ \boldsymbol{0}, & \text{otherwise.}\end{cases}$$

Clearly, the power constraint is always satisfied for the new coding scheme. Note that the error probability for the new coding scheme is still bounded by $\mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)$, because the spherical shaping "converts"İ an encoding failure to a decoding failure.

7.4.3.2 The Selection of Parameters

To complete the proof that our random ensemble achieves the AWGN capacity with lattice encoding and decoding, we carefully select the values of $k_f, k_c, \mathsf{p}, \gamma$ so that P_e goes to zero and the rate of our coding scheme goes to the AWGN capacity as n goes to infinity.

We have already bounded the error probability as

$$\begin{aligned}\mathsf{P}_e &\le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)\\ &\le \frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} + 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}} + 3 \times \frac{\mathsf{p}^{n-k_c}}{|\gamma\mathbb{Z}^n \cap \mathcal{B}\left(\sqrt{nP}\right)|}\\ &\quad + \max_{\boldsymbol{w}\in\mathcal{B}(r_e)} \frac{|\gamma\mathbb{Z}^n \cap (\boldsymbol{w} + \mathcal{B}(r_e))|}{\mathsf{p}^{n-k_f}}.\end{aligned}$$

Using Lemma 7.4, we obtain

$$\begin{aligned}\mathsf{P}_e &\le \frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} + 8\epsilon^2 n^{-1} + n^{-\frac{1}{2}}\\ &\quad + 3 \times \frac{\mathsf{p}^{n-k_c}}{\left(\max\left\{\frac{\sqrt{nP}}{\gamma} - \frac{\sqrt{n}}{2}, 0\right\}\right)^n V_n} + \frac{\left(\frac{r_e}{\gamma} + \frac{\sqrt{n}}{2}\right)^n V_n}{\mathsf{p}^{n-k_f}}.\end{aligned}$$

Now our goal is to select p, γ, k_c and k_f (as functions of n) such that

$$\frac{1}{\mathsf{p}-1}\frac{1}{\mathsf{p}^{n-k_f}} \to 0 \tag{7.8}$$

$$\frac{\mathsf{p}^{n-k_c}}{\left(\max\left\{\frac{\sqrt{nP}}{\gamma} - \frac{\sqrt{n}}{2}, 0\right\}\right)^n V_n} \to 0 \tag{7.9}$$

$$\frac{\left(\frac{r_e}{\gamma} + \frac{\sqrt{n}}{2}\right)^n V_n}{\mathsf{p}^{n-k_f}} \to 0 \tag{7.10}$$

Under the constraint $\mathcal{B}\left(\sqrt{nP}\right) \subset [-\frac{\gamma \mathsf{p}}{2}, \frac{\gamma \mathsf{p}}{2}]^n$, which is equivalent to

$$\gamma \mathsf{p} \geq 2\sqrt{nP}. \tag{7.11}$$

Let $\eta > 0$ and $\delta \in (0, 1)$ be two constants. Then let $\gamma = n^{-\frac{1}{2}\eta}$ and let p be the smallest prime number satisfying $\mathsf{p} \geq n^{\frac{1}{2}+\eta}$. By Bertrand's Postulate [43], $\mathsf{p} \leq 2n^{\frac{1}{2}+\eta}$. Hence, we can denote $\mathsf{p} = \mu n^{\frac{1}{2}+\eta}$, where $\mu \in [1, 2]$. We then assign

$$k_c = \left\lceil n\left(1 - \frac{2\log_2(\sqrt{P}n^{\frac{1}{2}\eta} - \frac{1}{2}) + \log_2((1-\delta)nV_n^{\frac{2}{n}})}{(1+2\eta)\log_2 n + 2\log_2 \mu}\right)\right\rceil,$$

and

$$k_f = \left\lfloor n\left(1 - \frac{2\log_2(\sqrt{\frac{1}{n}r_e^2}n^{\frac{1}{2}\eta} + \frac{1}{2}) + \log_2(\frac{1}{1-\delta}nV_n^{\frac{2}{n}})}{(1+2\eta)\log_2 n + 2\log_2 \mu}\right)\right\rfloor.$$

Since $\gamma p \geq n^{\frac{1}{2}+\frac{1}{2}\eta}$, it grows faster than $n^{\frac{1}{2}}$ and then the constraint (7.11) is met when n is large. By the facts that $\lim_{n\to\infty} nV_n^{\frac{2}{n}} = 2\pi e$ from [33, (2)] and that $\frac{1}{n}r_e^2 < P$ for small ϵ, one can verify that $1 \leq k_c < k_f < n$ when n is large. We now substitute p, k_1 and k_2 into (7.8),(7.9) and (7.10). It is clear (7.8),(7.9) and (7.10) vanish as $n \to \infty$.

Finally, we calculate the achievable rate

$$\lim_{n\to\infty} \frac{1}{n}\log_2 p^{k_f - k_c} = \lim_{n\to\infty} \frac{1}{2}\log_2\left(\frac{nP}{r_e^2}\right) = \frac{1}{2}\log_2\left(\frac{1+P/N}{1+\epsilon}\right),$$

where ϵ can be arbitrarily small.

7.5 Conclusions

In this chapter, we review the recent developments towards simplifying the achievability proofs related to nested linear/lattice codes. In Sect. 7.1, we introduce the model of the communication system and the motivation of using nested linear/lattice codes. In Sect. 7.2, we present definitions related to nested linear/lattice codes and introduce several elementary results from number theory that we use in our proofs. In Sect. 7.3, we prove that nested linear codes achieve the DMC channel capacity. In Sect. 7.4, we prove that nested lattice codes achieve the AWGN channel capacity. We make a particular effort in keeping these two proofs in parallel. Potential future

work includes optimizing the exponent of the growth rate of the prime p as a function of n, extending the results to the multi-user setting such as compute-and-forward, as well as providing achievability proofs without the random dither.

Appendix 1: Entropy

We briefly introduce various definitions related to entropy.

Entropy Let X be a discrete random variable with probability mass function (pmf) $p(x)$. The "uncertainty" about the outcome of X is measured by its entropy

$$H(X) = -\mathsf{E}_X(\log p(X)).$$

Conditional Entropy Let X, Y be two discrete random variables. Since $p(y|x)$ is a pmf, we can define $H(Y|X = x)$ for every x. The conditional entropy is the average of $H(Y|X = x)$ over every X, i.e.,

$$H(Y|X) = \sum_x H(Y|x)p(x) = -\mathsf{E}_{X,Y}(\log(p(Y|X))).$$

Joint Entropy Let (X, Y) be a pair of discrete random variables with pmf $p(x, y)$. The joint entropy is

$$H(X, Y) = -\mathsf{E}(\log p(X, Y)).$$

Mutual Information The mutual information between X and Y is

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y}{p(x)p(y)}.$$

It can be shown

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y).$$

Appendix 2: Typical Sequences

Let $\mathcal{X}$ be a discrete alphabet. For a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, we define its *empirical pmf* as

$$\pi(x \mid \boldsymbol{x}) = \frac{|\{i : x_i = x\}|}{n} \quad \text{for } x \in \mathcal{X}.$$

For $X \in \mathcal{X} \sim p_X(x_i)$ and $\epsilon \in (0, 1)$, define the set of ϵ-typical n-sequences $\boldsymbol{x} \in \mathcal{X}^n$ (or the typical set in short) as

$$\mathcal{T}_\epsilon^{(n)}(X) = \{\boldsymbol{x} : |\pi(x \mid \boldsymbol{x}) - p_X(x)| \le \epsilon p_X(x) \quad \text{for all } x \in \mathcal{X}\}.$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector in $\mathcal{X}^n$ whose elements are i.i.d. random variables with each element $x_i \sim p_X(x_i)$, $i \in [1, n]$. Then by the weak law of large numbers, for each $x \in \mathcal{X}$,

$$\pi(x \mid \boldsymbol{X}) \to p_X(x) \quad \text{in probability.}$$

Hence,

$$\lim_{n\to\infty} \mathsf{P}(\boldsymbol{X} \in \mathcal{T}_\epsilon^{(n)}(X)) = 1.$$

Intuitively, for any $\boldsymbol{x} \in \mathcal{T}_\epsilon^{(n)}(X)$, the empirical average $\frac{1}{n}\sum_{i=1}^n x_i$ should be close to the expectation $\mathsf{E}(X)$. In fact, we have a more general result as follows.

Lemma G.6 (Typical Average Lemma) *Let $\boldsymbol{x} \in \mathcal{T}_\epsilon^{(n)}(X)$. Then for any non-negative function $g(\cdot)$ on $\mathcal{X}$,*

$$(1-\epsilon)\,\mathsf{E}(g(X)) \le \frac{1}{n}\sum_{i=1}^n g(x_i) \le (1+\epsilon)\,\mathsf{E}(g(X)).$$

The proof is direct by noting $\frac{1}{n}\sum_{i=1}^n g(x_i) = \sum_{x\in\mathcal{X}} \pi(x \mid \boldsymbol{x}) g(x)$. Let $g(x) = -\log p_X(x)$ and note that $\mathsf{E}(-\log p_X(x)) = H(X)$, we obtain

$$2^{-n(1+\epsilon)H(X)} \le p_{\boldsymbol{X}}(\boldsymbol{x}) \le 2^{-n(1-\epsilon)H(X)}.$$

Equipped with this, we can bound the size of $\mathcal{T}_\epsilon^{(n)}(X)$. Note that the $\sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^{(n)}(X)} p_{\boldsymbol{X}}(\boldsymbol{x}) \le 1$, we obtain

$$|\mathcal{T}_\epsilon^{(n)}(X)| \le 2^{n(1+\epsilon)H(X)}.$$

Also note that by the law of large numbers,

$$\lim_{n\to\infty} \mathsf{P}(X \in \mathcal{T}_\epsilon^{(n)}(X)) = 1.$$

That is to say when n is sufficiently large, $\mathsf{P}(X \in \mathcal{T}_\epsilon^{(n)}(X)) \ge 1 - \epsilon$. Hence,

$$|\mathcal{T}_\epsilon^{(n)}(X)| \ge (1-\epsilon)2^{n(1-\epsilon)H(X)}.$$

The notion of the typical set can be extended to multiple random variables. For $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, define their *joint empirical pmf* as

$$\pi(x, y \mid \boldsymbol{x}, \boldsymbol{y}) = \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n} \quad \text{for } (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Let $(X, Y) \sim p_{X,Y}(x, y)$. The set of jointly ϵ-typical n-sequences is defined as

$$\begin{aligned}&\mathcal{T}_\epsilon^{(n)}(X, Y)\\ &\quad = \{(\boldsymbol{x}, \boldsymbol{y}) : |\pi(x, y \mid \boldsymbol{x}, \boldsymbol{y}) - p_{X,Y}(x, y)| \le \epsilon p_{X,Y}(x, y) \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\}.\end{aligned}$$

Also define the set of conditionally ϵ-typical n-sequences as

$$\mathcal{T}_\epsilon^{(n)}(X \mid \boldsymbol{y}) = \{\boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\}.$$

It can be shown that for sufficiently large n,

$$\forall \boldsymbol{y} \in \mathcal{Y}^n : |\mathcal{T}_\epsilon^{(n)}(X \mid \boldsymbol{y})| \le 2^{n(1+\epsilon)H(X|Y)}. \tag{A.1}$$

References

1. Alon, N., Spencer, J.H.: The Probabilistic Method. John Wiley & Sons, Hoboken (2004)
2. Berlekamp, E.R.: Algebraic Coding Theory. World Scientific Publishing Co., Singapore (2015)
3. Bresler, G., Parekh, A., Tse, D.N.C.: The approximate capacity of the many-to-one and one-to-many Gaussian interference channel. IEEE Trans. Inf. Theory **56**(9), 4566–4592 (2010)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, Hoboken (2006)
5. de Buda, R.: The upper error bound of a new near-optimal code. IEEE Trans. Inf. Theory **21**(4), 441–445 (1975)
6. de Buda, R.: Some optimal codes have structure. IEEE J. Sel. Areas Commun. **7**(6), 893–899 (1989)
7. di Pietro, N., Zémor, G., Boutros, J.J.: LDA lattices without dithering achieve capacity on the Gaussian channel. arXiv:1603.02863 [cs, math] (2016)
8. Dummit, D.S., Foote, R.M.: Abstract Algebra, vol. 3. Wiley, Hoboken (2004)
9. El Gamal, A., Kim, Y.H.: Network Information Theory. Cambridge University Press, Cambridge (2011)
10. Elias, P.: Coding for noisy channels. In: IRE Convention Record, vol. 3, Part 4, pp. 37–46 (1955)
11. Erez, U., Zamir, R.: Achieving $\frac{1}{2}\log(1 + \text{SNR})$ on the AWGN channel with lattice encoding and decoding. IEEE Trans. Inf. Theory **50**(10), 2293–2314 (2004)
12. Forney, G.D.: Coset codes. I. Introduction and geometrical classification. IEEE Trans. Inf. Theory **34**(5), 1123–1151 (1988)
13. Forney, G.D.: Coset codes. II. Binary lattices and related codes. IEEE Trans. Inf. Theory **34**, 1152–1187 (1988)
14. Gallager, R.G.: Information Theory and Reliable Communication. Wiley, New York (1968)
15. He, X., Yener, A.: Providing secrecy with structured codes: tools and applications to two-user Gaussian channels. IEEE Trans. Inf. Theory **60**(4), 2121–2138 (2014)

16. Hong, S.N., Caire, G.: Compute-and-forward strategies for cooperative distributed antenna systems. IEEE Trans. Inf. Theory **59**(9), 5227–5243 (2013)
17. Körner, J., Marton, K.: How to encode the modulo-two sum of binary sources. IEEE Trans. Inf. Theory **25**(2), 219–221 (1979)
18. Krithivasan, D., Pradhan, S.S.: Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function. IEEE Trans. Inf. Theory **55**(12), 5628–5651 (2009)
19. Krithivasan, D., Pradhan, S.S.: Distributed source coding using Abelian group codes. IEEE Trans. Inf. Theory **57**(3), 1495–1519 (2011)
20. Lim, S.H., Feng, C., Pastore, A., Nazer, B., Gastpar, M.: A joint typicality approach to algebraic network information theory (2016). Preprint. arXiv:1606.09548
21. Linder, T., Schlegal, C., Zeger, K.: Corrected proof of de Buda's theorem (lattice channel codes). IEEE Trans. Inf. Theory **39**(5), 1735–1737 (1993)
22. Loeliger, H.A.: On the basic averaging arguments for linear codes. Kluwer International Series in Engineering and Computer Science, pp. 251–251 (1994)
23. Loeliger, H.A.: Averaging bounds for lattices and linear codes. IEEE Trans. Inf. Theory **43**(6), 1767–1773 (1997)
24. Miyake, S.: Coding theorems for point-to-point communication systems using sparse matrix codes. Ph.D. Thesis, University of Tokyo, Tokyo (2010)
25. Miyake, S., Muramatsu, J.: A construction of channel code, joint source-channel code, and universal code for arbitrary stationary memoryless channels using sparse matrices. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **92**(9), 2333–2344 (2009)
26. Motahari, A.S., Gharan, S.O., Maddah-Ali, M.A., Khandani, A.K.: Real interference alignment: exploring the potential of single antenna systems (2009). Preprint available at http://arxiv.org/abs/0908.2282/
27. Muramatsu, J., Miyake, S.: Hash property and coding theorems for sparse matrices and maximum-likelihood coding. IEEE Trans. Inf. Theory **56**(5), 2143–2167 (2010)
28. Nam, W., Chung, S.Y., Lee, Y.H.: Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit. IEEE Trans. Inf. Theory **56**(11), 5488–5494 (2010)
29. Nazer, B., Gastpar, M.: Compute-and-forward: Harnessing interference through structured codes. IEEE Trans. Inf. Theory **57**(10), 6463–6486 (2011)
30. Niesen, U., Maddah-Ali, M.A.: Interference alignment: from degrees-of-freedom to constant-gap capacity approximations. IEEE Trans. Inf. Theory **59**(8), 4855–4888 (2013)
31. Niesen, U., Whiting, P.: The degrees-of-freedom of compute-and-forward. IEEE Trans. Inf. Theory **58**(8), 5214–5232 (2012)
32. Ntranos, V., Cadambe, V.R., Nazer, B., Caire, G.: Integer-forcing interference alignment. In: Proceedings of the IEEE International Symposium on Information Theory. Istanbul (2013)
33. Ordentlich, O., Erez, U.: A simple proof for the existence of "good" pairs of nested lattices. IEEE Trans. Inf. Theory **62**(8), 4439–4453 (2016)
34. Ordentlich, O., Erez, U., Nazer, B.: The approximate sum capacity of the symmetric Gaussian-user interference channel. IEEE Trans. Inf. Theory **60**(6), 3450–3482 (2014)
35. Padakandla, A., Pradhan, S.S.: Achievable rate region for three user discrete broadcast channel based on coset codes (2012). Preprint available at http://arxiv.org/abs/1207.3146
36. Padakandla, A., Pradhan, S.S.: Achievable rate region based on coset codes for multiple access channel with states (2013). Preprint available at http://arxiv.org/abs/1301.5655
37. Padakandla, A., Sahebi, A.G., Pradhan, S.S.: An achievable rate region for the three-user interference channel based on coset codes. IEEE Trans. Inf. Theory **62**(3), 1250–1279 (2016)
38. Poltyrev, G.: On coding without restrictions for the AWGN channel. IEEE Trans. Inf. Theory **40**(2), 409–417 (1994)
39. Qi, R., Feng, C., Huang, Y.C.: A simpler proof for the existence of Capacity-Achieving nested lattice codes. In: 2017 IEEE Information Theory Workshop (ITW) (IEEE ITW 2017), Kaohsiung (2017)
40. Ren, Z., Goseling, J., Weber, J.H., Gastpar, M.: Maximum throughput gain of compute-and-forward for multiple unicast. IEEE Commun. Lett. **18**(7), 1111–1113 (2014)

41. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423; **27**(4), 623–656 (1948)
42. Shomorony, I., Avestimehr, S.: Degrees of freedom of two-hop wireless networks: Everyone gets the entire cake. IEEE Trans. Inf. Theory **60**(5), 2417–2431 (2014)
43. Sondow, J.: Ramanujan primes and Bertrand's postulate. Am. Math. Mon. **116**, 630–635 (2009)
44. Song, Y., Devroye, N.: Lattice codes for the Gaussian relay channel: Decode-and-forward and compress-and-forward. IEEE Trans. Inf. Theory **59**(8), 4927–4948 (2013)
45. Tse, D.N.C., Maddah-Ali, M.A.: Interference neutralization in distributed lossy source coding. In: Proceedings of the IEEE IEEE International Symposium on Information Theory, Austin, TX (2010)
46. Urbanke, R., Rimoldi, B.: Lattice codes can achieve capacity on the AWGN channel. IEEE Trans. Inf. Theory **44**(1), 273–278 (1998)
47. Vatedka, S., Kashyap, N., Thangaraj, A.: Secure compute-and-forward in a bidirectional relay. IEEE Trans. Inf. Theory **61**(5), 2531–2556 (2015)
48. Wagner, A.B.: On distributed compression of linear functions. IEEE Trans. Inf. Theory **57**(1), 79–94 (2011)
49. Wilson, M.P., Narayanan, K., Pfister, H.D., Sprintson, A.: Joint physical layer coding and network coding for bidirectional relaying. IEEE Trans. Inf. Theory **56**(11), 5641–5654 (2010)
50. Xie, J., Ulukus, S.: Secure degrees of freedom of one-hop wireless networks. IEEE Trans. Inf. Theory **60**(6), 3359–3378 (2014)
51. Yang, Y., Xiong, Z.: Distributed compression of linear functions: Partial sum-rate tightness and gap to optimal sum-rate. IEEE Trans. Inf. Theory **60**(5), 2835–2855 (2014)
52. Zamir, R.: Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation and Multiuser Information Theory. Cambridge University Press, Cambridge (2014)
53. Zamir, R., Shamai, S., Erez, U.: Nested linear/lattice codes for structured multiterminal binning. IEEE Trans. Inf. Theory **48**(6), 1250–1276 (2002)