CFMA (Compute-Forward Multiple Access) and its Applications in Network Information Theory

THÈSE N^{O} 6996 (2016)

PRÉSENTÉE LE 27 MAI 2016

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE D'INFORMATION DANS LES SYSTÈMES EN RÉSEAUX
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Jingge ZHU

acceptée sur proposition du jury:

Prof. B. Rimoldi, président du jury
Prof. M. C. Gastpar, directeur de thèse
Prof. R. Zamir, rapporteur
Prof. G. Caire, rapporteur
Prof. R. Urbanke, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2016

Abstract

While both fundamental limits and system implementations are well understood for the point-to-point communication system, much less is developed for general communication networks. This thesis contributes towards the design and analysis of advanced coding schemes for multi-user communication networks with structured codes.

The first part of the thesis investigates the usefulness of lattice codes in Gaussian networks with a generalized compute-and-forward scheme. As an application, we introduce a novel multiple access technique — Compute-Forward Multiple Access (CFMA), and show that it achieves the capacity region of the Gaussian multiple access channel (MAC) with low receiver complexities. Similar coding schemes are also devised for other multi-user networks, including the Gaussian MAC with states, the two-way relay channel, the many-to-one interference channel, etc., demonstrating improvements of system performance because of the good interference mitigation property of lattice codes.

As a common theme in the thesis, computing the sum of codewords over a Gaussian MAC is of particular theoretical importance. We study this problem with nested linear codes, and improve upon the currently best known results obtained by nested lattice codes.

Inspired by the advantages of linear and lattice codes in Gaussian networks, we make a further step towards understanding intrinsic properties of the sum of linear codes. The final part of the thesis introduces the notion of typical sumset and presents asymptotic results on the typical sumset size of linear codes. The results offer new insight to coding schemes with structured codes.

Keywords: Compute-and-forward, compute-forward multiple access, CFMA, computation rate, Gaussian multiple access channel, Gaussian interference channel, lattice code, linear code, many-to-one interference channel, nested linear code, typical sumset.

Résumé

Si les limites fondamentales et les implémentations de systèmes de communication point-à-point sont désormais bien maîtrisées, les connaissances au sujet des réseaux de communication à terminaux multiples sont quant à elles bien moins développées. Cette thèse apporte de nouvelles contributions à la conception et à l'analyse de schémas de codage avancés pour les réseaux de communication à utilisateurs multiples employant des codes structurés.

La première partie de cette thèse étudie les avantages des schémas de calcul-et-transmission (angl. *compute-and-forward*) basés sur des codes à réseau (angl. *lattice codes*) dans les configurations à terminaux multiples de canaux gaussiens. Comme exemple d'application, une technique novatrice d'accès multiple est présentée — l'accès multiple par calcul-et-transmission (angl. *Compute-Forward Multiple Access*, abrégé CFMA), et il est démontré que cette technique permet d'atteindre la région de capacité du canal gaussien à accès multiple (angl. *Multiple Access Channel*, abrégé MAC) avec un récepteur d'une faible complexité. Des schémas de codage similaires sont aussi conçus pour d'autres configurations de canal à utilisateurs multiples, dont le MAC gaussien avec états, le canal à relai bidirectionnel, le canal à interférence plusieurs-à-un, etc., démontrant ainsi les améliorations de la performance du système obtenues grâce aux propriétés avantageuses des codes à réseau pour la mitigation des phénomènes d'interférence.

En tant que thème récurrent de cette thèse, le calcul de la somme de mots-code au travers d'un MAC gaussien est d'une importance théorique particulière. Nous étudions ce problème au moyen de codes linéaires imbriqués (angl. *nested linear codes*), et parvenons à surpasser les meilleures résultats connus à ce jour et obtenus grâce aux codes à réseau.

En nous inspirant des avantages que présentent les codes linéaires et les codes à réseau dans les configurations à terminaux multiples de canaux gaussiens, nous faisons un pas supplémentaire vers une compréhension complète des caractéristiques propres à la somme de codes linéaires. La dernière partie de cette thèse introduit la notion de somme d'ensembles typique et présente des résultats asymptotiques concernant la cardinalité de l'ensemble-somme typique pour les codes linéaires. Les resultats offrent un nouvel angle de vue sur les schémas de codage basés sur les codes structurés.

Mots-clés : calcul-et-transmission, CFMA, canal gaussien à accès multiple, canal gaussien à interférence, code à réseau, code linéaire, canal à interférence plusieurs-à-un, codes linéaires imbriqués, somme d'ensembles typique.

Acknowledgements

This thesis could not have been accomplished without the guidance of my supervisor, Michael Gastpar. Michael has his unique and admirable way of thinking about research problems and assessing the values and potentials of the results, which has to a great extent influenced my research style. I also learned a lot from his superb presentation skills. During these years, he gives me complete freedom so I can pursue research directions that interest me, and offers invaluable advice and encouragement after usual frustrations. I would like to express my deepest respect and gratitude to him.

It was a great honor to have Giuseppe Caire, Bixio Rimoldi, Rüdiger Urbanke and Ram Zamir on my thesis committee and I am very thankful for their helpful comments on the thesis. I also enjoyed classes taught by Olivier Lévêque, Nicolas Macris and Emre Telatar at EPFL. Emre's comments on my research problems have always been a source of inspiration.

Life in Lausanne would be much harder without our secretary, France Faille, who is always there whenever needed, with her great affection.

Although being a small group, the LINX family leaves me many memorable moments. Many thanks to Giel Op 't Veld, not only for being an ideal office mate, but also for the nice photos he took for us around the world. It is very enjoyable to talk with Saeid Sahraei who has a great sense of humor, and I am also happy to had many discussions with Chien-Yi Wang about almost everything. It is always a great learning experience when discussing with Sung Hoon Lim, who taught me lots of interesting things in Information Theory. Special thanks go to Adriano Pastore, for helping me with French and a Schubert four-hands fantasy. I also appreciated various interactions with former members of the group: Sangwoon Jeon, Naveen Goela and Chen Feng, as well as other members in the IPG group, after we joined the big family.

Finally, I want to thank my family for their continuous supports throughout the years. This thesis is dedicated to Ye, for her love and patience which makes me a very happy man every day.

Contents

1 Introduction

With rapid progress on wireless communication technologies and the growing demands of multimedia applications, the number of wireless devices has increased drastically in recent years along with ever increasing request for higher data rates. These changes have shifted the challenges of communications in network: from combating *noise* to mitigating *interference*.

The classical Information Theory established by Shannon [1] provides definite answers to fundamental limits of point-to-point communications, where the main challenge is to deal with noise in the channel. After the birth of the mathematical theory of communication, it has taken several decades for researchers to find practical error-correcting codes along with efficient encoding and decoding algorithms. In particular, Turbo codes [2], LDPC codes [3] and recently proposed polar codes [4] are exemplary results of capacity-approaching/achieving codes amenable to practical implementations. After six decades of research, it can be argued that the simple point-to-point communication systems are well understood and the theory developed sofar is sufficient to guide the design for commercial communication systems.

On the other hand, communication systems in real life are far more complicated than idealized point-to-point models. We often deal with situations where a large number of mobile devices are active simultaneously in a relatively small space, such as wireless hotspots. In these scenarios there exist complicated interactions among different devices and we do not yet have a satisfying theory for such complex systems. Network Information Theory, also started by Shannon in [5], is an extension of classical Information Theory to communication networks, and allows us to study the fundamental limits of communication in networks to some degree [6]. However, most communication problems in networks are still wide open, including the very basic system consisting of two transceiver pairs which models the simplest interference channel.

Despite the fact that the optimal communication scheme for most multi-user communication systems are unknown, recent progress in Network Information Theory shows that certain classes of codes are particularly suited for some communication networks. These are the so-called Gaussian networks where the transmitted signals are linearly added up at receivers along with additive noises. The codes of

interest have certain algebraic structure (hence referred to as *structured codes* in the sequel) which matches the additive channel well and makes it much easier for receivers to mitigate interference. Furthermore, recent research shows that for communications in a network, it is essential to let intermediate devices (or "relays" in a network) process the information in an intelligent way. For example the celebrated Network Coding [7] result shows that in a wired (noiseless) network, mixing two information flows in intermediate relays (for example performing summation of two information symbols) can increase data throughput in the network, if there exist more than one transceiver pair in the network. For a general noisy communication network, structured codes are shown to be very useful for such intermediate information processing. Specialized to the Gaussian network, lattice codes and associated new schemes give new perspectives of channel coding in communication networks. Roughly speaking, instead of directly decoding transmitted codewords reliably, relays can choose to decode the sum of codewords, or more generally, to *compute* a function of codewords from different users reliably, and this function will be used in subsequent steps for further process.

A noticeable contribution towards computation over network is the *compute-and-forward* scheme introduced in [8]. The idea is to let intermediate nodes decode integer combinations of codewords, and in the end if receivers obtain enough integer combinations, its desired message can be extracted by solving a set of linear equations. As one of the main topics in this thesis, we will introduce a generalized compute-and-forward scheme which incorporates channel state information at transmitters (CSIT) in a meaningful way, and show that it gives new perspectives for multiple access problems. More specifically, a novel technique called Compute-Forward Multiple Access (CFMA) is introduced for the Gaussian multiple access channel (MAC) with the advantage of achieving the capacity of Gaussian MAC with relatively simple receiver structures.

Contributions

- **Generalized Compute-and-Forward scheme.** We develop a generalized compute-and-forward scheme using nested lattice codes which can utilize the channel state information (CSI) at transmitters in a beneficial way [9]. In particular, instead of using fixed lattice codes at every transmitters, we proposed to use differently scaled lattices at different transmitters. The scaling is chosen according the channel coefficients in the network so that the channel gain is fully exploited. This could enlarge the computation rate considerably in networks with asymmetric channel gains and has immediately applications on many scenario.

- **Applications: CFMA and lattice codes in Gaussian networks.** As an application of the generalized compute-and-forward scheme, a multiple access technique is developed for the classical Gaussian multiple access channel (MAC) [10] [11]. In this scheme, the receiver will first recover integer combinations of messages and solve each message individually afterwards. One attractive feature of this multiple access technique is that the receiver is equipped with a *single-user decoder* of low-complexity. This is compared to the conventional optimal decoder for Gaussian MAC, which either performs multi-user

detection (high complexity) on the received signal, or requires time-sharing between two users (extra constraints on transmitters). With the recent growing interests on non-orthogonal multiple access techniques, this novel approach may attract interests in industry-related research and help innovate the next generation communication technologies. A similar coding scheme is proposed for Gaussian MAC with states non-causally known to transmitters (the Gaussian Dirty MAC) and shown to give new achievable rate regions [11]. Various coding schemes based on lattice codes are also studied on other networks. For the Gaussian many-to-one interference channel, a lattice based scheme is shown to outperform conventional coding strategies, and establishes new constant gap or capacity results which are independent of the number of users [12]. Novel coding schemes are developed for two-user interference channels, two-way relay channels and MIMO channels, which either improve upon best known results, or recover known results with simpler decoder architectures.

- **Nested linear codes for computation.** Like lattice codes to Gaussian networks, the recently proposed *nested linear codes* can be used for computation over general discrete-time memoryless networks, including the well studied Gaussian networks. We investigate the achievable computation rates with this code for a simple two-user Gaussian MAC [13]. The results not only recover the best known results with nested lattice codes, but also show theoretical improvements with nested linear codes.

- **Typical sumsets of linear codes.** Motivated by the applications of lattice codes in wireless networks, we study the sumset of linear codes. Given two identical linear codes $\mathcal{C}$ over $\mathbb{F}_q$ of length n, we independently pick one codeword from each codebook uniformly at random. A *sumset* is formed by adding these two codewords entry-wise as integer vectors and a sumset is called *typical*, if the sum falls inside this set with high probability. We ask the question: how large is the typical sumset for most codes? We show that when the rate R of the linear code is below a certain threshold D, the typical sumset size is roughly $|\mathcal{C}|^2 = 2^{2nR}$ for most codes while when R is above this threshold, most codes have a typical sumset whose size is roughly $|\mathcal{C}| \cdot 2^{nD} = 2^{n(R+D)}$ due to the linear structure of the codes. The threshold D depends solely on the alphabet size q and takes value in $[1/2, \log\sqrt{e})$.

Notations

Vectors and matrices are denoted using bold letters such as $\mathbf{a}$ and $\mathbf{A}$, respectively. The i-th entry of a vector $\mathbf{a}$ is denoted as $\mathbf{a}_i$ and $\mathbf{A}_i$ denotes the i-th column of the matrix $\mathbf{A}$. We often use $[a : b]$ to denote the set of integers $\{a, a+1, \ldots, b-1, b\}$. Logarithm log is with base 2 and we use the shorthand notation $\log^+(x) := \max\{0, \log(x)\}$ for $x > 0$. Sets are usually denoted using calligraphic letters such as $\mathcal{A}$ and their cardinality are denoted by $|\mathcal{A}|$. We often deal with quantities depending on the codeword length n. The notation $o_n(1)$ denotes a quantity that approaches 0 as $n \to \infty$. We say $a \doteq 2^{nb}$ for some constant b if there exists some $\epsilon_n \searrow 0$ such that $2^{n(b-\epsilon_n)} \leq a \leq 2^{n(b+\epsilon_n)}$. We also consider the probability of events in the limit when the codeword length n goes to infinity. For any event H, we say the event H occurs *asymptotically almost surely* (a.a.s.) if $\mathbb{P}\{H\} \to 1$ as $n \to \infty$.

Given a probability distribution P_U over the alphabet $\mathcal{U}$, we use $\mathcal{A}_{[U]}^{(n)}$ to denote the set of typical sequences defined as:

$$\mathcal{A}_{[U]}^{(n)} := \left\{ \mathbf{m} : \left| P_U(a) - \frac{1}{n} N(a|\mathbf{m}) \right| \leq \delta, \text{ for all } a \in \mathcal{U} \right\} \tag{1.1}$$

where $N(a|\mathbf{m})$ is the occurrence count of the symbol a in sequence $\mathbf{m} = (\mathbf{m}_1, \ldots, \mathbf{m}_n)$. Similarly we can define the conditional typical sequences $\mathcal{A}_{[Z|U]}^{(n)}(\mathbf{u})$ as well as the typical sequences $\mathcal{A}_{[ZU]}^{(n)}$ determined by a joint distribution P_{ZU} as in [14, Ch. 2].

2 Preliminaries

Lattices and lattice codes are important ingredients to communication schemes studied in this thesis. This chapter is devoted to give necessary background on lattices in Euclidean space and nested lattice codes built out of it. Materials in this chapter can be found in the comprehensive treatment [15] on this topic. In particular all definitions in this chapter follow those in [15]. At the end of this chapter we also review the fundamental tool to many of the advanced communication schemes, the compute-and-forward scheme [8].

2.1 Lattices and Lattice Codes

A lattice Λ is a discrete subgroup of $\mathbb{R}^n$ with the property that if $\mathbf{t}_1, \mathbf{t}_2 \in \Lambda$, then $\mathbf{t}_1 + \mathbf{t}_2 \in \Lambda$. An n-dimensional lattice Λ can be generated by n linearly independent basis vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$ in $\mathbb{R}^n$ as

$$\Lambda = \left\{ \mathbf{t} = \sum_{i=1}^{n} a_i \mathbf{g}_i : a_i \in \mathbb{Z} \right\}.$$

The lattice quantizer $Q_\Lambda : \mathbb{R}^n \to \Lambda$ is defined as

$$Q_\Lambda(\mathbf{x}) = \text{argmin}_{\mathbf{t}\in\Lambda} \left|\left|\mathbf{t} - \mathbf{x}\right|\right|. \tag{2.1}$$

The fundamental Voronoi region of a lattice Λ is defined to be

$$\mathcal{V} := \{\mathbf{x} \in \mathbb{R}^n : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}. \tag{2.2}$$

The modulo operation gives the quantization error with respect to the lattice:

$$[\mathbf{x}] \text{mod } \Lambda := \mathbf{x} - Q_\Lambda(\mathbf{x}). \tag{2.3}$$

The following definitions describe properties of a lattice.

Definition 2.1 (Second moment)**.** *The second moment of the lattice Λ with Voronoi region $\mathcal{V}$ is defined to be*

$$\sigma^2(\Lambda) := \frac{1}{n\,Vol\,(\mathcal{V})} \int_{\mathcal{V}} \left|\left|\mathbf{x}\right|\right|^2 d\mathbf{x}. \tag{2.4}$$

Definition 2.2 (Normalized second moment)**.** *The normalized second moment of a lattice Λ with Voronoi region $\mathcal{V}$ is defined to be*

$$G(\Lambda) := \frac{\sigma^2(\Lambda)}{(Vol\ (\mathcal{V}))^{2/n}}.$$

Later in this chapter, we will construct codes using lattices for the additive white-Gaussian noise (AWGN) channel of the form

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

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{R}^n$ are the channel input, channel output and additive noise, respectively. The Gaussian noise $\mathbf{Z}$ is assumed to be independent from the channel input $\mathbf{X}$ and its probability density function is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi N_0)^{n/2}} e^{-\frac{\|\mathbf{z}\|^2}{2N_0}}$$

where N_0 is the variance per dimension. Given $\mathbf{Y}$, an estimation of $\mathbf{X}$ can be given by simply quantizing $\mathbf{Y}$ with respect to the lattice Λ. This is called *lattice decoding* (or *nearest-neighbor decoding*) in the literature and the estimate is given as

$$\hat{\mathbf{X}} = Q_\Lambda(\mathbf{Y}) = \mathrm{argmin}_{\mathbf{t}\in\Lambda} \left|\left|\mathbf{Y} - \mathbf{t}\right|\right|.$$

The following definitions are important for measuring the performance of lattice codes in an AWGN channel.

Definition 2.3 (Error probability)**.** *The error probability in lattice decoding of the lattice Λ, in the presence of AWGN $\mathbf{Z}$ with variance N_0, is defined as*

$$P_e(\Lambda, N_0) := \mathbb{P}\left\{\mathbf{Z} \notin \mathcal{V}\right\}$$

where $\mathcal{V}$ is the Voronoi region of Λ.

Definition 2.4 (Normalized volume to noise ratio)**.** *The normalized volume to noise ratio (NVNR) of a lattice Λ, at a target error probability $0 < P_e < 1$, is defined as*

$$\mu(\Lambda, P_e) := \frac{(Vol\ (\mathcal{V}))^{2/n}}{N_0(P_e)}$$

where $N_0(\epsilon)$ is the value of N_0 such that $P_e(\Lambda, N_0)$ is equal to ϵ.

Notice that NVNR is a dimensionless number and is invariant to scaling or rotation of the lattice. Now we are ready to define the asymptotic goodness of lattices.

Definition 2.5 (Good for AWGN channel)**.** *A sequence of n-dimensional lattices $\Lambda^{(n)}$ is said to be good for AWGN channel if for all $P_e > 0$, the normalized volume to noise ratios satisfy*

$$\lim_{n\to\infty} \mu(\Lambda^{(n)}, P_e) = 2\pi e$$

Definition 2.6 (Good for quantization)**.** *A sequence of n-dimensional lattices $\Lambda^{(n)}$ is said to be good for quantization if the normalized second moments satisfy*

$$\lim_{n\to\infty} G(\Lambda^{(n)}) = \frac{1}{2\pi e} \tag{2.6}$$

We will see later that these goodness properties are desirable for constructing lattice codes with good performance. An important result from [16] shows that there exists a sequence of lattices such that they are asymptotically good both for quantization and AWGN channel.

Theorem 2.1 (Simultaneous goodness [16])**.** *There exists a sequence of lattices of increasing dimension $\Lambda^{(n)}$ which satisfy $\mu(\Lambda^{(n),P_e}) \to 2\pi e$ and $G(\Lambda_n) \to 1/2\pi e$ as $n \to \infty$ for all $P_e > 0$.*

Two lattices Λ and Λ' are said to be nested if $\Lambda' \subseteq \Lambda$. A nested lattice code $\mathcal{C}$ can be constructed using the coarse Λ' for *shaping* and the fine lattice Λ as codewords:

$$\mathcal{C} := \{\mathbf{t} \in \mathbb{R}^n : \mathbf{t} \in \Lambda \cap \mathcal{V}'\} \tag{2.7}$$

where $\mathcal{V}'$ is the Voronoi region of Λ'. It can be shown [15, Cor. 8.2.1] that the size of the codebook $|\mathcal{C}|$ is given by $\Gamma^n := \text{Vol}\,(\mathcal{V}')/\text{Vol}\,(\mathcal{V})$ where Γ is called the *nesting ratio* and the *rate* of this nested lattice code is defined to be

$$R := \frac{1}{n}\log\frac{\text{Vol}\,(\mathcal{V}')}{\text{Vol}\,(\mathcal{V})}. \tag{2.8}$$

The following result shows that there also exists a sequences of nested lattices which are simultaneously good.

Theorem 2.2 (Good nested lattices)**.** *For any nesting ratio Γ, there exists a sequence of nested lattices $(\Lambda^{(n)}, \Lambda'^{(n)})$ with $\Lambda'^{(n)} \subseteq \Lambda^{(n)}$, such that each lattice is good for quantization and good for AWGN channel.*

A proof of the above result can be found in Erez and Zamir [17] or [15, Thm. 8.5.1]. Nam et al. [18, Theorem 2] extend the results to the case when there are multiple nested lattice codes.

Given a lattice code belonging to a lattice Λ, it is shown in [19] that the code can be used (with a spherical shaping region) to achieve the capacity of AWGN channels with ML decoding. A more interesting question is if lattice codes can achieve the capacity of AWGN channels with lattice decoding. That is, the decoder estimates the transmitted codeword by simply quantizing the (possibly pre-processed) channel output with respect to the lattice Λ. This question is studied in [20] and finally settled by Erez and Zamir [17] using nested lattice codes.

Theorem 2.3 (Capacity-achieving lattice codes with lattice decoding [17])**.** *Consider the AWGN channel in (2.5) with capacity $C = \frac{1}{2}\log(1+P)$. For any $\epsilon > 0$, there exists a sequence of nested lattice codes with rate defined in (2.8) grater than $C - \epsilon$, that achieve the capacity of this channel using lattice decoding.*

This theorem, in particular the performance of lattice codes under lattice decoding, is a key result to many advanced coding schemes to be studied in the rest of the thesis.

2.2 Multiple Access Channels

The multiple access channel (MAC) is a basic building block for many communication networks. It is also one of the few examples in network information theory whose optimal transmission strategy is known. In this section we review the results for a general K-user discrete memoryless MAC.

A multiple access channel with K users is specified by a conditional probability mass function $p_{Y|X_1X_2\ldots,X_K}(y|x_1,x_2,\ldots,x_k)$ with channel inputs $x_k \in \mathcal{X}_k, k = 1,\ldots,K$ and channel output $y \in \mathcal{Y}$. Each transmitter is equipped with an encoder $\mathcal{E}_k$ which maps a message M_k from the set $\mathcal{M}_k := \{1,\ldots,2^{nR_k}\}$ to a channel input in $\mathcal{X}_k^n$ with length n, and the receiver is equipped with a decoder $\mathcal{D}$ which maps the channel output in $\mathcal{Y}^n$ to K estimated messages. The receiver is interested in decoding all messages from all transmitters reliably. More specifically, let M_k denote the randomly chosen message in the message set $\mathcal{M}_k$ of user K, the average error probability of decoding all messages is given by

$$P_e^{(n)} := \bigcup_{k=1}^{K} \mathbb{P}\left\{\hat{M}_k \neq M_k\right\}$$

where $\hat{M}_k$ denotes the estimated messages at the receiver. We say the *achievable message rate tuple*[1] $(R_1,\ldots,R_K)$ is achievable, if there exist encoders and a decoder such that the above error probability P_e can be made arbitrarily small for large enough n. The *capacity region* of the MAC is the closure of the set of achievable rate tuples.

The capacity region of the multiple access channel is found in [21] [22].

Theorem 2.4. *The capacity region of the K-user discrete memoryless multiple access channel is the set of rate tuples $(R_1,R_2,\ldots,R_k)$ such that*

$$\sum_{j\in\mathcal{J}} R_j \leq I(X(\mathcal{J}),Y|X(\mathcal{J}^c),Q) \text{ for every } \mathcal{J}\subseteq[1:K]$$

for some pmf $p_Q(q)\prod_{j=1}^{K} p_j(x_j|q)$ with the cardinality of Q bounded as $|\mathcal{Q}| \leq K$. Furthermore $X(\mathcal{J})$ denotes the set $\{X_j, j\in\mathcal{J}\}$ and $\mathcal{J}^c$ is the complement set of $\mathcal{J}$.

In particular, the capacity region of the 2-user MAC is the set of rate pairs (R_1,R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2,Q) \\ R_2 &\leq I(X_2;Y|X_1,Q) \\ R_1+R_2 &\leq I(X_1,X_2;Y|Q) \end{aligned}$$

for some pmf $p_Q(q)p_q(x_1|q)p_2(x_2|q)$ with $|\mathcal{Q}| \leq 2$.

2.3 Achievable Computation Rates

The term "achievable rate" is widely used in the information and communication theory literature and has a straightforward meaning in the context of conventional

[1] or simply *achievable rate tuple.*

communication networks, where messages of source nodes are to be decoded reliably at the intended destinations as in the previous section. But some interesting (or even optimal) communication schemes in network will require the receiver not to decode individual messages, but to process the incoming information in some other way. Hence if the goal of a communication scheme is not to decode (individual) messages, the term "achievable rate" should be used with caution. The purpose of this section to make clear distinctions between these concepts. In this section, definitions are given with multiple access channels for the sake of simplicity, but the ideas carry over easily to general networks.

Now we consider a K-user MAC where the receiver wishes to decode a function of incoming messages reliably. In its most general form, let g be a function which maps K messages from $\mathcal{M}_1 \times \ldots \times \mathcal{M}_K$ to an element in a set $\mathcal{G}$. The goal at the receiver is not to decode individual messages M_k, but to decode a function of the messages. Then the error probability of this coding scheme is given by

$$P_{e,g}^{(n)} := \mathbb{P}\left\{\hat{G}(Y^n) \neq g(M_1, \ldots, M_K)\right\}$$

where $\hat{G}(Y^n)$ denotes the estimated function value using channel output Y^n. For computing such a function of messages, the achievable computation rates are defined as follows.

Definition 2.7 (Computation rate tuple)**.** *Consider a K-user multiple access channel. We say a computation rate tuple $(R_1, \ldots, R_K)$ with respect to the function $g : \mathcal{M}_1 \times \cdots \times \mathcal{M}_K \mapsto \mathcal{G}$ is achievable, if there exist encoders and a decoder, such that the decoding error probability $P_{e,g}^{(n)}$ can be made arbitrarily small for large enough n.*

Notice that the achievable computation rates depend not only on the channel, but also on the function to be computed. But in slight abuse of notation, the dependence on the function g is suppressed in the notation for the computation rate R_k. The term *computation rates* are also often used without explicitly mentioning the function to be computed, tacitly assuming that it is clear from the context. We should point out that the concept of achievable computation rates can be viewed as a generalization of the conventional achievable (message) rates. Indeed, if we let g to be the identity function, i.e., $g(M_1, M_2) = (M_1, M_2)$, then the two definitions coincide. However we shall see in subsequent chapters that for a given channel, the achievable computation rates (for certain function) can be higher than achievable message rates.

2.4 The Compute-and-Forward Scheme

We will briefly review the compute-and-forward scheme proposed by Nazer and Gastpar in [8], which considers computing the *sum* of codewords via a Gaussian network. Although the scheme discussed in [8] is applicable to a general Gaussian network with multiple transmitters and multiple receivers, we will only restrict our attention to the Gaussian multiple access channel in this section for the sake of brevity. Applications in later chapter will consider more general settings.

To illustrate the basic idea of the compute-and-forward scheme, we consider the canonical K-user Gaussian MAC. The discrete-time real Gaussian MAC has the following vector representation

$$\mathbf{y} = \sum_{k=1}^{K} h_k \mathbf{x}_k + \mathbf{z} \tag{2.9}$$

with $\mathbf{y}, \mathbf{x}_k \in \mathbb{R}^n$ denoting the channel output at the receiver and channel input of transmitter k. The white Gaussian noise with unit variance per entry is denoted by $\mathbf{z} \in \mathbb{R}^n$. A fixed real number h_k denotes the channel coefficient from user k. Notice that in the original compute-and-forward scheme, transmitters do not need to have the knowledge of channel coefficients. We can assume without loss of generality that every user has the same power constraints on the channel input as $\mathbb{E}\{||\mathbf{x}_k||^2\} \leq nP$.

As described in Section 2.1, given two simultaneously good lattices $\Lambda' \subseteq \Lambda$, and a nested lattice code is constructed as $\mathcal{C} := \Lambda \cap \mathcal{V}'$. For user k, each message M_k is mapped to a codeword $\mathbf{t}_k(M_k)$ in $\mathcal{C}$ in a one-to-one fashion. The way to construct this mapping is called *Construction A* and is discussed in details in [8]. The function to be computed at the receiver is of the form:

$$g(M_1, \ldots, M_k) := \left[\sum_{k=1}^{K} a_k \mathbf{t}_k(M_k) \right] \bmod \Lambda' \tag{2.10}$$

where a_k are integers for all $k = 1, \ldots, K$.

Theorem 2.5 (Compute-and-forward [8])**.** *For the K-user Gaussian MAC in (2.9), the computation rate tuple $(r_1, \ldots, r_K)$ with respect to the modulo sum $g(M_1, \ldots, M_k)$ defined in (2.10) is achievable if*

$$r_k < \log^+ \left(||\mathbf{a}|| - \frac{P(\mathbf{h}^T \mathbf{a})^2}{1 + P\,||\mathbf{h}||^2} \right)^{-1}, k = 1, \ldots K$$

where $\mathbf{a} := [a_1, \ldots, a_K] \in \mathbb{Z}^K$ and $\mathbf{h} := (h_1, \ldots, h_K)$.

The key property that the sum of two lattice points is still a lattice point is the rational behind choosing lattice codes for computation. Namely, the possible sums of codewords from a structured code (lattice code for example) are much fewer than that from an unstructured (randomly chosen) code. Hence intuitively it should be easier to decode the sum with structured codes. More concrete results on the sum of codes will be presented in Chapter 7.

We point out that in the original formulation of the compute-and-forward scheme [8], the achievable computation rates are the same for all users if the power constraints are the same, regardless of the channel coefficients (notice that the expression in Theorem 2.5 does not depend on k). For the case when power constraints are different, the authors in [23] have shown achievable computation rate tuples with different rates for different users, using similar nested lattice codes construction (although the ratio of the rates is determined by their power constraints). However it is known that for a Gaussian MAC, one can always absorb the power constraints into the channel coefficients and assume without loss of generality that the power constraints are the same. This suggests that the results in [8] and [23] are special cases of a more general scheme, which we shall discuss in the next chapter.

Computation with Lattice Codes over Gaussian Networks

3

In this chapter we will extend the compute-and-forward strategy to the scenario where the channel state information is known at transmitters (CSIT)[1]. We will show that with this information, a modified compute-and-forward scheme incorporating CSIT will significantly enlarge the computation rate regions in some cases.

3.1 A General Compute-and-Forward Scheme

We first introduce the generalized compute-and-forward scheme for the K-user Gaussian MAC

$$\mathbf{y} = \sum_{k=1}^{K} h_k \mathbf{x}_k + \mathbf{z} \tag{3.1}$$

with $\mathbf{y}, \mathbf{x}_k \in \mathbb{R}^n$ denoting the channel output at the receiver and channel input of transmitter k, respectively. The channel coefficient from user k to the receiver is denoted by h_k, and is assumed to be known at the transmitter k. We can assume without loss of generality that every user has the same power constraints on the channel input as $\mathbb{E}\{||\mathbf{x}_k||^2\} \leq nP$.

To construct the nested lattice codes in our scheme, let $\beta_k, k = 1, \ldots, K$ be K nonzero real numbers. For each user we choose a lattice Λ_k which is simultaneously good in the sense of Definition 2.5 and 2.6. These K lattices $\Lambda_k, k = 1, \ldots, K$ are chosen to form a nested lattice chain according to a certain order which will be determined later (We do not exclude the possibility that these K lattices are the same). We let Λ_c denote the coarsest lattice among them, i.e., $\Lambda_c \subseteq \Lambda_k$ for all $k = 1, \ldots, K$. We will construct another K nested lattices $\Lambda_k^s \subseteq \Lambda_c$ where all lattices are also simultaneously good, and with second moment

$$\frac{1}{n\text{Vol}\left(\mathcal{V}_k^s\right)} \int_{\mathcal{V}_k^s} ||\mathbf{x}||^2 \, d\mathbf{x} = \beta_k^2 P$$

[1]The material of this chapter has appeared in

J. Zhu and M. Gastpar, "Asymmetric Compute-and-Forward with CSIT", in *Proc. International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2014

where $\mathcal{V}_k^s$ denotes the Voronoi region of the lattice Λ_k^s. The lattice Λ_k^s is used as the *shaping region* for the codebook of user k. For each transmitter k, we construct the codebook as

$$\mathcal{C}_k = \Lambda_k \cap \mathcal{V}_k^s \tag{3.2}$$

and the *rate* of the codebook $\mathcal{C}_k$ is defined to be

$$r_k := \frac{1}{n}\log|\mathcal{C}_k| = \frac{1}{n}\log\frac{\text{Vol}\,(\mathcal{V}_k^s)}{\text{Vol}\,(\mathcal{V}_k)} \tag{3.3}$$

Furthermore, messages M_k of user k are bijectively mapped to codewords in $\mathcal{C}_k$.

Similar to the original compute-and-forward scheme, the function to be computed at the receiver is given by

$$g(M_1,\ldots,M_k) := \left[\sum_{k=1}^{K} a_k \mathbf{t}_k(M_k)\right] \bmod \Lambda_f^s \tag{3.4}$$

where $\mathbf{t}_k(M_k)$ is the codeword from user k and Λ_f^s denotes the finest lattice among $\Lambda_k^s, k=1,\ldots,K$ and a_k are integers for all $k=1,\ldots,K$.

Theorem 3.1 (General compute-and-forward for the Gaussian MAC)**.** *Consider a K-user Gaussian MAC with channel coefficients $\mathbf{h}=(h_1,\ldots,h_K)$ and equal power constraint P. Let $\beta_1,\ldots,\beta_K$ be K nonzero real numbers, the computation rate tuple $(r_1,\ldots,r_K)$ with respect to the modulo sum in (3.4) is achievable if*

$$r_k < \left[\frac{1}{2}\log\left(||\tilde{\mathbf{a}}||^2 - \frac{P(\mathbf{h}^T\tilde{\mathbf{a}})^2}{1+P\,||\mathbf{h}||^2}\right)^{-1} + \frac{1}{2}\log\beta_k^2\right]^+ \tag{3.5}$$

for all k where $\tilde{\mathbf{a}} := [\beta_1 a_1, ..., \beta_K a_K]$ and $a_k \in \mathbb{Z}$ for all $k \in [1:K]$.

Proof. A proof is given in the Appendix of this chapter. □

We have the following remarks regarding this general compute-and-forward scheme.

- By setting $\beta_k = 1$ for all k we recover the original compute-and-forward formula given in Theorem 2.5.
- The usefulness of the parameters $\beta_1,\ldots,\beta_K$ lies in the fact that they can be chosen according to the channel coefficients h_k and power P.
- In the case that each transmitter has power P_k, replace h_k by $h_k' := \sqrt{P_k/P}h_k$ for all k in (3.5).

Before moving on, it is instructive to inspect formula (3.5) in some details. We rewrite (3.5) in the following expression

$$\frac{1}{2}\log\left(\beta_i(1+P\|\mathbf{h}\|^2)\right) - \frac{1}{2}\log\left(\|\tilde{\mathbf{a}}\|^2 + P(\|\mathbf{h}\|^2\|\tilde{\mathbf{a}}\|^2 - (\mathbf{h}^T\tilde{\mathbf{a}})^2)\right). \tag{3.6}$$

As already pointed out in [24], the term $\|\mathbf{h}\|^2\|\tilde{\mathbf{a}}\|^2 - (\mathbf{h}^T\tilde{\mathbf{a}})^2$ in the second log has a natural interpretation – it measures how the coefficient $\tilde{\mathbf{a}}$ differs from the channel $\mathbf{h}$, in other words the rate loss occurred because of the mismatch between the chosen

coefficient and channel gains. Cauchy-Schwartz Inequality implies that this term is always nonnegative and is zero if and only if $\tilde{\mathbf{a}}$ is colinear with the channel coefficient $\mathbf{h}$. Notice that in the original compute-and-forward scheme, where $\tilde{\mathbf{a}} = \mathbf{a}$ by setting all β_k to be 1, this term is not necessarily zero because $\mathbf{a}$ is an integer vector while $\mathbf{h}$ can take all possible values in $\mathbb{R}^K$. However in this generalized scheme we are given the freedom to tune parameters $\beta_k \in \mathbb{R}^K$, and the rate loss due to the mismatch can be completely eliminated by choosing β_k to align $\tilde{\mathbf{a}}$ with $\mathbf{h}$. In general, the lattice scaling coefficients β_k allow us to adjust the codebook rate freely and is essential to our coding scheme for Gaussian MAC discussed in the sequel.

Lastly we comment again on the difference between achievable (message) rates and achievable computation rates defined in Definition 2.7. We give an example of computation rate pairs for a 2-user Gaussian MAC in Figure 3.1. It is worth noting that the achievable computation rate region can be strictly larger than the achievable message rate region.

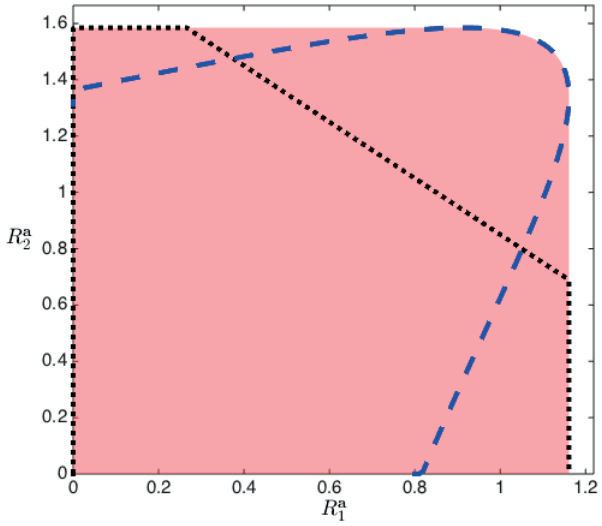


Figure 3.1 – In this figure we show an achievable computation rate region for computing the sum $[\mathbf{t}_1 + \mathbf{t}_2] \mod \Lambda_f^s$ over a 2-user Gaussian MAC where $h_1 = 1, h_2 = \sqrt{2}$ and $P = 4$. The dotted black line shows the capacity region of this MAC. The dashed blue line depicts the computation rate pairs given by (3.5) in Theorem 3.1. Points along this curve are obtained by choosing different β_1, β_2. The shaded region shows the whole computation rate region, in which all the computation rate pairs are achievable. Notice that in this case the computation rate region contains the whole capacity region of this Gaussian MAC and is strictly larger than the latter.

As studied in [8], the compute-and-forward scheme can be used in a Gaussian network with more than one receivers. More precisely, we can consider a Gaussian network with K transmitters and M relays as

$$\mathbf{y}_m = \sum_{k=1}^{K} h_{mk}\mathbf{x}_k + \mathbf{z}_m, m = 1, \ldots, M \tag{3.7}$$

where each relay wants to decode one integer combination of codewords in the form

$$g_m(M_1,\dots,M_K) := \left[\sum_{k=1}^{K} a_{mk}\mathbf{t}_k(M_k)\right] \mod \Lambda_f^s, m = 1,\dots M \tag{3.8}$$

with $a_{mk} \in \mathbb{Z}$ for all m,k. As before Λ_f^s denotes the finest lattice among $\Lambda_k^s, k = 1,\dots,K$. Notice that in this case, the computation rate tuple is defined under the condition that *all* modulo sum $g_m, m = 1,\dots,M$ should be decoded reliably at the intended relay.

Theorem 3.2 (General compute-and-forward with multiple receivers)**.** *Consider a network with K transmitters and M relays in (3.7) with channel coefficients $\mathbf{h}_m := (h_{m1},\dots,h_{mK})$ and equal power constraints P. Let $\beta_1,\dots,\beta_K$ be K nonzero real numbers, the computation rate tuple $(r_1,\dots,r_K)$ with respect to the M modulo sums in (3.8) is achievable if*

$$r_k < \min_{m\in[1:M]} R_k(\mathbf{a}_m,\mathbf{h}_m)$$

where $R_k(\mathbf{a}_m,\mathbf{h}_m)$ is defined as

$$R_k(\mathbf{a}_m,\mathbf{h}_m) := \left[\frac{1}{2}\log\left(||\tilde{\mathbf{a}}_\mathbf{m}||^2 - \frac{P(\mathbf{h}_m^T\tilde{\mathbf{a}}_\mathbf{m})^2}{1+P\,||\mathbf{h}_m||^2}\right)^{-1} + \frac{1}{2}\log\beta_k^2\right]^+$$

with $\tilde{\mathbf{a}}_m := [a_{m1}\beta_1,\dots,a_{mK}\beta_K]$.

Proof. The codes constructions are given in (3.2). Unlike the special case of Gaussian MAC with one receiver, the fine lattices $\Lambda_k, k = 1,\dots,K$ in this network are in general different (but still nested). We use Λ_f to denote the finest lattice among $\Lambda_k, k = 1,\dots K$, and each relay m decodes the function g_m with respect to Λ_f in the same way as in the proof of Theorem 3.1. The decoding procedure at relay m imposes a constraint on the rate of the codebook $\mathcal{C}_k$, i.e., it should hold that $r_k \le R_k(\mathbf{a}_m,\mathbf{h}_m)$ for all k. If all relays want to decode the sum successfully, each transmitter has to construct its codebook such that it meets the above constraints at *all* relays. Therefore when the codebook is constructed as in (3.2), the fine lattice Λ_k for $\mathcal{C}_k$ should be chosen such that the message rate R_k does not exceed $R_k(\mathbf{a}_m,\mathbf{h}_m)$ for any m, hence the rate of the codebook $\mathcal{C}_k$ is given by $\min_{m\in[1:M]} R_k(\mathbf{a}_m,\mathbf{h}_m)$. □

3.2 Appendix

We give the proof of Theorem 3.1.

Proof of Theorem 3.1. The codes constructions are given in (3.2). In fact for the Gaussian MAC with one receiver, we can choose all the fine lattices $\Lambda_k, k = 1,\dots,K$ to be the same lattice, denoted as Λ. Its Voronoi region is denoted by $\mathcal{V}$. When the message M_k of user k is chosen, the encoder finds the corresponding codeword $\mathbf{t}_k$ and forms its channel input as follows

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k] \mod \Lambda_k^s/\beta_k \tag{3.9}$$

where the dither $\mathbf{d}_k$ is a random vector uniformly distributed in the scaled Voronoi region $\mathcal{V}_k^s/\beta_k$. As pointed out in [17], $\mathbf{x}_k$ is independent from $\mathbf{t}_k$ and also uniformly in Λ_k^s/β_k hence has average power P for all k.

At the decoder we form

$$\begin{aligned}\tilde{\mathbf{y}} &:= \alpha\mathbf{y} - \sum_k a_k\beta_k\mathbf{d}_k \\ &= \sum_k a_k\left(\beta_k(\mathbf{t}_k/\beta_k + \mathbf{d}_k) - \beta_k Q_{\Lambda_k^s/\beta_k}(\mathbf{t}_k/\beta_k + \mathbf{d}_k)\right) - \sum_k a_k\beta_k\mathbf{d}_k + \tilde{\mathbf{z}} \\ &\overset{(a)}{=} \tilde{\mathbf{z}} + \sum_k a_k(\mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k)) \\ &:= \tilde{\mathbf{z}} + \sum_k a_k\tilde{\mathbf{t}}_k\end{aligned}$$

with $\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k)$ and the equivalent noise

$$\tilde{\mathbf{z}} := \sum_k(\alpha h_k - a_k\beta_k)\mathbf{x}_k + \alpha\mathbf{z} \tag{3.10}$$

which is independent of $\sum_k a_k\tilde{\mathbf{t}}_k$ since all $\mathbf{x}_k$ are independent of $\sum_k a_k\tilde{\mathbf{t}}_k$ thanks to the dithers $\mathbf{d}_k$. The step (a) follows because it holds $Q_\Lambda(\beta X) = \beta Q_{\frac{\Lambda}{\beta}}(X)$ for any $\beta \neq 0$.

The decoder obtains the sum $\sum_k a_k\tilde{\mathbf{t}}_k$ using lattice decoding with respect to the lattice Λ. That is, the decoder quantizes $\tilde{\mathbf{y}}$ to its nearest neighbor in Λ. Notice we have $\tilde{\mathbf{t}}_k \in \Lambda$ for all k because $\mathbf{t}_k \in \Lambda$ and $\Lambda_k^s \subseteq \Lambda$ due to the nested codes construction. Hence the sum $\sum_k a_k\tilde{\mathbf{t}}_k$ also belongs to the lattice Λ. The decoding error probability is equal to the probability that the equivalent noise $\tilde{\mathbf{z}}$ leaves the Voronoi region surrounding the lattice point $\sum_k a_k\tilde{\mathbf{t}}_k$. Since the fine lattice Λ is good for AWGN channel, the probability $\Pr(\tilde{\mathbf{z}} \notin \mathcal{V})$ goes to zero exponentially as long as

$$\frac{\text{Vol }(\mathcal{V})^{2/n}}{N(\alpha)} > 2\pi e \tag{3.11}$$

where

$$N(\alpha) := \mathbb{E}\,||\tilde{\mathbf{z}}||^2/n = ||\alpha\mathbf{h} - \tilde{\mathbf{a}}||^2 P + \alpha^2 \tag{3.12}$$

denotes the average power per dimension of the equivalent noise. Recall that the shaping lattice Λ_k^s is good for quantization hence we have

$$G(\Lambda_k^s)2\pi e < (1+\delta) \tag{3.13}$$

for any $\delta > 0$ if n is large enough. Together with the rate expression in (3.3) we can see that lattice decoding is successful if

$$\beta_k^2 P 2^{-2r_k}/G(\Lambda_k^s) > 2\pi e N$$

for every k, or equivalently

$$r_k < \frac{1}{2}\log\left(\frac{P}{N(\alpha)}\right) + \frac{1}{2}\log\beta_k^2 - \frac{1}{2}\log(1+\delta)$$

By choosing δ arbitrarily small and optimizing over α we conclude that the lattice decoding of $\sum_k a_k \tilde{\mathbf{t}}_k$ will be successful if

$$r_k < \max_{\alpha} \frac{1}{2} \log \left(\frac{P}{N(\alpha)} \right) + \frac{1}{2} \log \beta_k^2 \tag{3.14}$$

$$= \frac{1}{2} \log \left(||\tilde{\mathbf{a}}||^2 - \frac{P(\mathbf{h}^T \tilde{\mathbf{a}})^2}{1 + P\,||\mathbf{h}||^2} \right)^{-1} + \frac{1}{2} \log \beta_k^2 \tag{3.15}$$

Lastly the modulo sum is obtained by

$$\left[\sum_k a_k \tilde{t}_k \right] \mod \Lambda_f^s = \left[\sum_k a_k \mathbf{t}_k - \sum_k a_k Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k \mathbf{d}_k) \right] \mod \Lambda_f^s$$

$$= \left[\sum_k a_k t_k \right] \mod \Lambda_f^s$$

where the last equality holds because Λ_f^s is the finest lattice among $\Lambda_k^s, k = 1, \ldots, K$. □

Application: Compute-Forward Multiple Access (CFMA)

4

Lattice codes used under the compute-and-forward paradigm suggest an alternative strategy for the standard Gaussian multiple-access channel (MAC): The receiver successively decodes integer linear combinations of the messages until it can invert and recover all messages.[1] As it is entirely based on the compute-forward scheme, this type of multiple-access technique will be called *compute-forward multiple access* (CFMA). In this chapter, we will show that how CFMA can achieve the capacity region of the two-user Gaussian MAC, with the advantage that simple *single-user decoders* can be used at the receiver. Coding strategies with the general compute-and-forward scheme are also applied to other networks, including the general K-user Gaussian MAC, the two-user Gaussian MAC with states non-causally known to transmitters, the Gaussian two-way relay channel, and the point-to-point Gaussian MIMO channel.

4.1 The Two-user Gaussian MAC

The Gaussian multiple access channel is a well-understood communication system. To achieve its entire capacity region, the receiver can either use joint decoding (a multi-user decoder), or a single-user decoder combined with successive cancellation decoding and time-sharing [25, Ch. 15]. An extension of the successive cancellation decoding called Rate-Splitting Multiple Access is developed in [26] where only single-user decoders are used to achieve the whole capacity region without time-sharing, but at the price that messages have to be split to create more virtual users.

In this section we show that without time-sharing, the entire capacity region can be attained with a single-user decoder with CFMA as soon as the signal-to-noise ratios are above $1 + \sqrt{2}$. For the 2-user Gaussian MAC, the receiver first

[1] The material of this chapter has appeared in

1. J. Zhu and M. Gastpar, "Asymmetric Compute-and-Forward with CSIT", in *Proc. International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2014

2. J. Zhu and M. Gastpar, "Gaussian (dirty) multiple access channels: A compute-and-forward perspective", in *Proc. 2014 IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014

3. J. Zhu and M. Gastpar, "Multiple Access via Compute-and-Forward", in *arXiv: 1407.8463*.

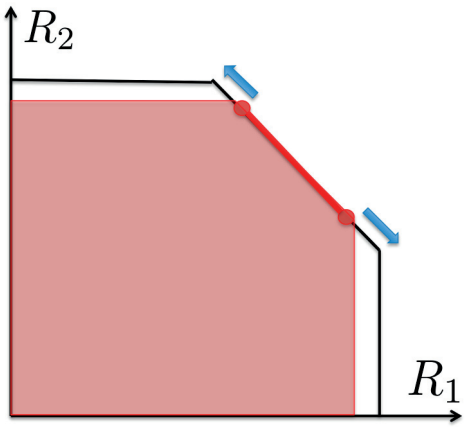


Figure 4.1 – An illustration of an achievable rate region for a 2-user Gaussian MAC with the proposed scheme. The rate pairs in the shaded region can be achieved using a single-user decoder without time-sharing. As SNR increases, the end points of the line segment approach the corner points and the whole capacity region becomes achievable. A sufficient condition for achieving the whole capacity region is that the SNR of both users are above $1+\sqrt{2}$.

decodes the sum of the two transmitted codewords, and then decodes either one of the codewords, using the sum as side information. As an example, Figure 4.1 gives an illustration of an achievable rate region for a *symmetric* 2-user Gaussian MAC with our proposed scheme. When the *signal-to-noise ratio* (SNR) of both users is below 1.5, the proposed scheme cannot attain rate pairs on the dominant face of the capacity region. If the SNR exceeds 1.5, a line segment on the capacity boundary is achievable. As SNR increases, the end points of the line segment approach the corner points, and the whole capacity region is achievable as soon as the SNR of both users is larger than $1+\sqrt{2}$. We point out that the decoder used in our scheme is a single-user decoder since it merely performs lattice quantizations on the received signal. Hence this novel approach allows us to achieve rate pairs in the capacity region using only a single-user decoder without time-sharing or rate splitting.

We should point out that a related result in [27] shows that using a similar idea of decoding multiple integer sums, the sum capacity of the Gaussian MAC can be achieved within a constant gap. Furthermore, it is also shown in [28] that under certain conditions, some isolated (non-corner) points of the capacity region can be attained. To prove these results, the authors use fixed lattices which are independent of channel gains. Here we close these gaps by showing that if the lattices are properly scaled in accordance with the channel gains, the full capacity region can be attained.

Recall that the 2-user Gaussian MAC is given by

$$\mathbf{y} = h_1\mathbf{x}_1 + h_2\mathbf{x}_2 + \mathbf{z} \tag{4.1}$$

with equal power constraints $||\mathbf{x}_k||^2 \leq nP, k = 1, 2$. We use nested lattice codes for two users with the same construction described in Section 3.1. The encoding and decoding procedures are given as follows.

- Encoding: For user k, given the message and the unique corresponding code-

word $\mathbf{t}_k$, the channel input is generated as

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k] \text{mod } \Lambda_k^s/\beta_k, k = 1, 2. \tag{4.2}$$

where $\mathbf{d}_k$ is called a *dither* which is a random vector uniformly distributed in the scaled Voronoi region $\mathcal{V}_k^s/\beta_k$.

- Decoding: To decode the first sum with coefficient (a_1, a_2), let Λ_f denote the finer lattice between Λ_1, Λ_2 if $a_1, a_2 \neq 0$. Otherwise set $\Lambda_f = \Lambda_1$ if $a_2 = 0$, or $\Lambda_f = \Lambda_2$ if $a_1 = 0$. Let α_1 be a real number to be determined later and form $\tilde{\mathbf{y}}_1 := \alpha_1 \mathbf{y} - \sum_k a_k \beta_k \mathbf{d}_k$, the first sum with coefficient $\mathbf{a}$ is decoded by performing the lattice quantization

$$Q_{\Lambda_f}(\tilde{\mathbf{y}}_1) \tag{4.3}$$

Define Λ_f' in the similarly way for the second sum with coefficient (b_1, b_2), the second sum is obtained by performing the lattice quantization

$$Q_{\Lambda_f'}(\tilde{\mathbf{y}}_2) \tag{4.4}$$

where the construction of $\tilde{\mathbf{y}}_2$ is given the proof of the following theorem.

Theorem 4.1 (Achievable message rate pairs for the 2-user Gaussian MAC)**.** *Consider the 2-user multiple access channel in (4.1). Let β_1, β_2 be two nonzero real numbers and we collect them into one vector $\underline{\beta} := (\beta_1, \beta_2)$. The following message rate pair is achievable*

$$R_k = \begin{cases} r_k(\mathbf{a}, \underline{\beta}) & \text{if } b_k = 0 \\ r_k(\mathbf{b}|\mathbf{a}, \underline{\beta}) & \text{if } a_k = 0 \\ \min\{r_k(\mathbf{a}, \underline{\beta}), r_k(\mathbf{b}|\mathbf{a}, \underline{\beta})\} & \text{otherwise} \end{cases}$$

for any linearly independent $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^2$ and $\underline{\beta} \in \mathbb{R}^2$ if it holds $r_k(\mathbf{a}, \underline{\beta}), r_k(\mathbf{b}|\mathbf{a}, \underline{\beta}) \geq 0$ for $k = 1, 2$, where we define

$$r_k(\mathbf{a}, \underline{\beta}) := \frac{1}{2} \log \frac{\beta_k^2 (1 + h_1^2 P + h_2^2 P)}{K(\mathbf{a}, \underline{\beta})} \tag{4.5}$$

$$r_k(\mathbf{b}|\mathbf{a}, \underline{\beta}) := \frac{1}{2} \log \frac{\beta_k^2 K(\mathbf{a}, \underline{\beta})}{\beta_1^2 \beta_2^2 (a_2 b_1 - a_1 b_2)^2} \tag{4.6}$$

with

$$K(\mathbf{a}, \underline{\beta}) := \sum_k a_k^2 \beta_k^2 + P(a_1 \beta_1 h_2 - a_2 \beta_2 h_1)^2 \tag{4.7}$$

Proof. Recall that the transmitted signal for user k is given by

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k] \text{mod } \Lambda_k^s/\beta_k \tag{4.8}$$

Notice that $\mathbf{x}_k$ is independent of $\mathbf{t}_k$ and uniformly distributed in Λ_k^s/β_k hence has average power P_k for $k = 1, 2$.

Given two integers a_1, a_2 and some real number α_1, we can form

$$\begin{aligned}
\tilde{\mathbf{y}}_1 &:= \alpha_1 \mathbf{y} - \sum_k a_k \beta_k \mathbf{d}_k \\
&= \underbrace{\sum_k (\alpha_1 h_k - a_k \beta_k)\mathbf{x}_k + \alpha_1 \mathbf{z}_1}_{\tilde{\mathbf{z}}_1} + \sum_k a_k \beta_k \mathbf{x}_k - \sum_k a_k \beta_k \mathbf{d}_k \\
&\overset{(a)}{=} \tilde{\mathbf{z}}_1 + \sum_k a_k \left(\beta_k (\mathbf{t}_k/\beta_k + \mathbf{d}_k) - \beta_k Q_{\Lambda_k^s/\beta_k}(\mathbf{t}_k/\beta_k + \mathbf{d}_k) \right) - \sum_k a_k \beta_k \mathbf{d}_k \\
&\overset{(b)}{=} \tilde{\mathbf{z}}_1 + \sum_k a_k (\mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k \mathbf{d}_k)) \\
&= \tilde{\mathbf{z}}_1 + \sum_k a_k \tilde{\mathbf{t}}_k
\end{aligned} \tag{4.9}$$

with the notation

$$\tilde{\mathbf{z}}_1 := \sum_k (\alpha_1 h_k - \beta_k a_k)\mathbf{x}_k + \alpha_1 \mathbf{z} \tag{4.10}$$

$$\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k \mathbf{d}_k) \tag{4.11}$$

Step (a) follows from the definition of $\mathbf{x}_k$ and step (b) uses the identity $Q_\Lambda(\beta \mathbf{x}) = \beta Q_{\Lambda/\beta}(\mathbf{x})$ for any real number $\beta \neq 0$. Note that $\tilde{\mathbf{t}}_k$ lies in Λ due to the nested construction $\Lambda_k^s \subseteq \Lambda$. The term $\tilde{\mathbf{z}}_1$ acts as an equivalent noise independent of $\sum_k a_k \tilde{\mathbf{t}}_k$ (thanks to the dithers) and has an average variance per dimension

$$N_1(\alpha_1) = \sum_k (\alpha_1 h_1 - \beta_k a_k)^2 P + \alpha_1^2 \tag{4.12}$$

The decoder obtains the sum $\sum_k a_k \tilde{\mathbf{t}}_k$ from $\tilde{\mathbf{y}}_1$ using *lattice decoding*: it quantizes $\tilde{\mathbf{y}}_1$ to its closest lattice point in Λ. Using the same argument in the proof of Theorem 3.1, we can show this decoding process is successful if the rate of the transmitter k satisfies

$$r_k < r_k(\mathbf{a}, \underline{\beta}) := \max_{\alpha_1} \frac{1}{2} \log^+ \frac{\beta_k^2 P}{N_1(\alpha_1)} \tag{4.13}$$

Optimizing over α_1 we obtain the claimed expression in (4.5). In other words we have the computation rate pair $(r_1(\mathbf{a}, \underline{\beta}), r_2(\mathbf{a}, \underline{\beta}))$ for computing the sum [2] $a_1 \tilde{\mathbf{t}}_1 + a_2 \tilde{\mathbf{t}}_2$. We remark that the expression (4.5) is exactly the general compute-and-forward formula given in Theorem 3.1 for $K = 2$.

To decode a second integer sum with coefficients $\mathbf{b}$ we use the idea of successive cancellation [8][29]. If $r_k(\mathbf{a}, \underline{\beta}) > 0$ for $k = 1, 2$, i.e., the sum $\sum_k a_k \tilde{\mathbf{t}}_k$ can be decoded, we can reconstruct the term $\sum_k a_k \beta_k \mathbf{x}_k$ as $\sum_k a_k \beta_k \mathbf{x}_k = \sum_k a_k \tilde{\mathbf{t}}_k + \sum_k a_k \beta_k \mathbf{d}_k$.

[2] Notice that in Theorem 3.1, the computation rate tuple is defined with respect to the modulo sum $[\sum_k a_k \mathbf{t}_k] \mod \Lambda_f^s$. Here we decode the sum $\sum_k a_k \tilde{\mathbf{t}}_k$ without the modulo operation. However this will not affect the achievable message rate pair, because we can also recover the two codewords $\mathbf{t}_1$ and $\mathbf{t}_2$ using the two sums $\sum_k a_k \tilde{\mathbf{t}}_k$ and $\sum_k b_k \tilde{\mathbf{t}}_k$, as shown in the proof.

Similar to the derivation of (4.9), we can use $\sum_k a_k\beta_k\mathbf{x}_k$ to form

$$\tilde{\mathbf{y}}_2 := \alpha_2\mathbf{y} + \lambda(\sum_k a_k\beta_k\mathbf{x}_k) - \sum_k b_k\beta_k\mathbf{d}_k \tag{4.14}$$

$$= \sum_k(\alpha_2 h_k - (b_k + \lambda a_k)\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z} + \sum_k b_k\tilde{\mathbf{t}}_k \tag{4.15}$$

$$= \tilde{\mathbf{z}}_2 + \sum_k b_k\tilde{\mathbf{t}}_k \tag{4.16}$$

where the equivalent noise

$$\tilde{\mathbf{z}}_2 := \sum_k(\alpha_2 h_k - (b_k + \lambda a_k)\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z} \tag{4.17}$$

has average power per dimension

$$N_2(\alpha_2, \lambda) = \sum_k(\alpha_2 h_k - (b_k + \lambda a_k)\beta_k)^2 P + \alpha_2^2. \tag{4.18}$$

Under lattice decoding, the term $\sum_k b_k\tilde{\mathbf{t}}_k$ can be decoded if for $k = 1, 2$ we have

$$r_k < r_k(\mathbf{b}|\mathbf{a}, \underline{\beta}) = \max_{\alpha_2,\lambda} \frac{1}{2}\log^+ \frac{\beta_k^2 P}{N_2(\alpha_2, \lambda)} \tag{4.19}$$

Optimizing over α_2 and λ gives the claimed expression in (4.6). In other words we have the computation rate pair $(r_1(\mathbf{b}|\mathbf{a}, \underline{\beta}), r_2(\mathbf{b}|\mathbf{a}, \underline{\beta}))$ for computing the sum $b_1\tilde{\mathbf{t}}_1 + b_2\tilde{\mathbf{t}}_2$.

A simple yet important observation is that if $\mathbf{a}, \mathbf{b}$ are two linearly independent vectors, then $\tilde{\mathbf{t}}_1$ and $\tilde{\mathbf{t}}_2$ can be solved using the two decoded sums, and consequently two messages $\mathbf{t}_1, \mathbf{t}_2$ are found by

$$\mathbf{t}_k = [\tilde{\mathbf{t}}_k] \mod \Lambda_k^s$$

This means that if two vectors $\mathbf{a}$ and $\mathbf{b}$ are linearly independent, the message rate pair (R_1, R_2) is achievable with

$$R_k = \min\{r_k(\mathbf{a}, \underline{\beta}), r_k(\mathbf{b}|\mathbf{a}, \underline{\beta})\} \tag{4.20}$$

Another important observation is that when we decode a sum $\sum_k a_k\tilde{\mathbf{t}}_k$ with the coefficient $a_i = 0$, the lattice point $\tilde{\mathbf{t}}_i$ does not participate in the sum $\sum_k a_k\tilde{\mathbf{t}}_k$ hence the rate R_i will not be constrained by this decoding procedure as in (4.13). For example if we decode $a_1\tilde{\mathbf{t}}_1 + a_2\tilde{\mathbf{t}}_2$ with $a_1 = 0$, the computation rate pair is actually $(\infty, r_1(\mathbf{a}, \underline{\beta}))$, since the rate of user 1 in this case can be arbitrarily large. The same argument holds for the case $b_k = 0$. Combining (4.20) and the special cases when a_k or b_k equals zero, we have the claimed result. □

The achievability scheme described in the above theorem is based on the compute-and-forward scheme hence is called compute-forward multiple access (CFMA). Now we state the main theorem in this section showing it is possible to use CFMA to achieve non-trivial rate pairs satisfying $R_1 + R_2 = C_{sum} := \frac{1}{2}\log(1 + h_1^2 P + h_2^2 P)$. Furthermore, we show that the whole capacity region is achievable under certain conditions on h_1, h_2 and P.

Theorem 4.2 (Capacity achieving for the 2-user Gaussian MAC)**.** *We consider the two-user Gaussian MAC in (4.1) where two sums with coefficients* $\mathbf{a}$ *and* $\mathbf{b}$ *are decoded. We assume that* $a_k \neq 0$ *for* $k = 1, 2$ *and define*

$$A := \frac{h_1 h_2 P}{\sqrt{1 + h_1^2 P + h_2^2 P}}. \tag{4.21}$$

Case I): *If it holds that*

$$A < 3/4, \tag{4.22}$$

the sum capacity cannot be achieved with CFMA.

Case II): *If it holds that*

$$A \geq 3/4, \tag{4.23}$$

the sum rate capacity can be achieved by decoding two integer sums using $\mathbf{a} = (1, 1), \mathbf{b} = (0, 1)$ *with message rate pairs*

$$R_1 = r_1(\mathbf{a}, \beta_2), R_2 = r_2(\mathbf{b}|\mathbf{a}, \beta_2), \text{ with some } \quad \beta_2 \in [\beta_2', \beta_2''] \tag{4.24}$$

or using $\mathbf{a} = (1, 1), \mathbf{b} = (1, 0)$ *with message rate pairs*

$$R_1 = r_1(\mathbf{b}|\mathbf{a}, \beta_2), R_2 = r_2(\mathbf{a}, \beta_2), \text{ with some } \quad \beta_2 \in [\beta_2', \beta_2''] \tag{4.25}$$

where β_2', β_2'' *are two real roots of the quadratic equation*

$$f(\beta_2) := K(\mathbf{a}, \beta_2) - \beta_2 \sqrt{1 + h_1^2 P + h_1^2 P} = 0 \tag{4.26}$$

The expressions $r_k(\mathbf{a}, \beta_2)$, $r_k(\mathbf{b}|\mathbf{a}, \beta_2)$ *and* $K(\mathbf{a}, \beta_2)$ *are given in (4.5), (4.6) and (4.7) by setting* $\beta_1 = 1$*, respectively.*

Case III: *If it holds that*

$$A \geq 1, \tag{4.27}$$

by choosing $\mathbf{a} = (1, 1)$ *and* $\mathbf{b} = (0, 1)$ *or* $\mathbf{b} = (1, 0)$*, the achievable rate pairs in (4.24) and (4.25) cover the whole dominant face of the capacity region.*

Proof. It is easy to see from the rate expressions (4.5) and (4.6) that we can without loss of generality assume $\beta_1 = 1$ in the following derivations. We do not consider the case when $a_k = 0$ for $k = 1$ or $k = 2$, which is just the classical interference cancellation decoding. Also notice that it holds:

$$r_1(\mathbf{a}, \beta_2) + r_2(\mathbf{b}|\mathbf{a}, \beta_2) = r_2(\mathbf{a}, \beta_2) + r_1(\mathbf{b}|\mathbf{a}, \beta_2) = \frac{1}{2} \log \frac{1 + (h_1^2 + h_2^2)P}{(a_2 b_1 - a_1 b_2)^2} \tag{4.28}$$

$$= C_{sum} - \log |a_2 b_1 - a_1 b_2| \tag{4.29}$$

We start with **Case I)** when the sum capacity cannot be achieved. This happens when

$$r_k(\mathbf{a}, \beta_2) < r_k(\mathbf{b}|\mathbf{a}, \beta_2), k = 1, 2 \tag{4.30}$$

for *any* choice of β_2, which is equivalent to

$$f(\beta_2) > 0 \tag{4.31}$$

where $f(\beta_2)$ is given in (4.26). To see this, notice that Theorem 4.1 implies that in this case the sum message rate is

$$R_1 + R_2 = r_1(\mathbf{a}, \beta_2) + r_2(\mathbf{a}, \beta_2) \tag{4.32}$$

for $a_k \neq 0$. Due to Eqn. (4.29) we can upper bound the sum message rate by

$$R_1 + R_2 < r_1(\mathbf{a}) + r_2(\mathbf{b}|\mathbf{a}, \beta_2) \leq C_{sum} \tag{4.33}$$
$$R_1 + R_2 < r_2(\mathbf{a}) + r_1(\mathbf{b}|\mathbf{a}, \beta_2) \leq C_{sum}, \tag{4.34}$$

meaning the sum capacity is not achievable. It remains to characterize the condition under which the inequality $f(\beta_2) > 0$ holds. It is easy to see the expression $f(\beta_2)$ is a quadratic function of β_2 with the leading coefficient $a_2^2(1+h_1^2P)$. Hence $f(\beta_2) > 0$ always holds if the equation $f(\beta_2) = 0$ does not have any real root. The solutions of $f(\beta_2) = 0$ are given by

$$\beta_2' := \frac{2a_1a_2h_1h_2P + S - \sqrt{SD}}{2(a_2^2 + a_2^2h_1^2P)} \tag{4.35a}$$
$$\beta_2'' := \frac{2a_1a_2h_1h_2P + S + \sqrt{SD}}{2(a_2^2 + a_2^2h_1^2P)} \tag{4.35b}$$

with

$$S := \sqrt{1 + (h_1^2 + h_2^2)P} \tag{4.36}$$
$$D := S(1 - 4a_1^2a_2^2) + 4Pa_1a_2h_1h_2 \tag{4.37}$$

Inequality $f(\beta_2) > 0$ holds for all real β_2 if $D < 0$ or equivalently

$$\frac{h_1h_2P}{\sqrt{1 + (h_1^2 + h_2^2)P}} < \frac{4a_1^2a_2^2 - 1}{4a_1a_2} \tag{4.38}$$

The R.H.S. of the above inequality is minimized by choosing $a_1 = a_2 = 1$ which yields the condition (4.22). This is shown in Figure 4.2a: in this case the computation rate pair of the first sum $\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2$ is too small and it cannot reach the sum capacity.

In **Case II)** we require $r_k(\mathbf{a}, \beta_2) \geq r_k(\mathbf{b}|\mathbf{a}, \beta_2)$ or equivalently $f(\beta_2) \leq 0$ for some β_2. By the derivation above, this is possible if $D \geq 0$ or equivalently

$$\frac{h_1h_2P}{\sqrt{1 + (h_1^2 + h_2^2)P}} \geq \frac{4a_1^2a_2^2 - 1}{4a_1a_2} \tag{4.39}$$

If we choose the coefficients to be $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (0, b_2)$ for some nonzero integers a_1, a_2, b_2, Theorem 4.1 implies the sum rate is

$$R_1 + R_2 = r_1(\mathbf{a}, \beta_2) + r_2(\mathbf{b}|\mathbf{a}, \beta_2) = C_{sum} - \log|a_2b_1 - a_1b_2| \tag{4.40}$$

If the coefficients satisfy $|a_2b_1 - a_1b_2| = 1$, the sum capacity is achievable by choosing $\beta_2 \in [\beta_2', \beta_2'']$, with which the inequality (4.39) holds. Notice that if we choose

$\beta_2 \notin [\beta_2', \beta_2'']$, then $r_k(\mathbf{a}, \beta_2) < r_k(\mathbf{b}|\mathbf{a}, \beta_2)$ and we are back to Case I). The condition $|a_2 b_1 - a_1 b_2| = 1$ is satisfied if the coefficients are chosen to be $\mathbf{a} = (1, 1), \mathbf{b} = (0, 1)$. For simplicity we collect these two vectors and denote them as $\mathbf{A}_1 := (\mathbf{a}^T, \mathbf{b}^T)^T$.

In general, not the whole dominant face of the capacity region can be achieved by varying $\beta_2 \in [\beta_2', \beta_2'']$. One important choice of β_2 is $\beta_2^{(1)} := \frac{h_1 h_2 P}{1+h_1^2 P}$. With this choice of β_2 and coefficients $\mathbf{A}_1$ we have

$$R_1 = r_1(\mathbf{a}, \beta_2^{(1)}) = \frac{1}{2}\log(1 + h_1^2 P) \tag{4.41}$$

$$R_2 = r_2(\mathbf{b}|\mathbf{a}, \beta_2^{(1)}) = \frac{1}{2}\log(1 + \frac{h_2^2 P}{1 + h_1^2 P}) \tag{4.42}$$

which is one corner point of the capacity region. Similarly with $\beta_2^{(2)} := \frac{1+h_2^2 P}{h_1 h_2 P}$ and coefficients $\mathbf{A}_2$ we have

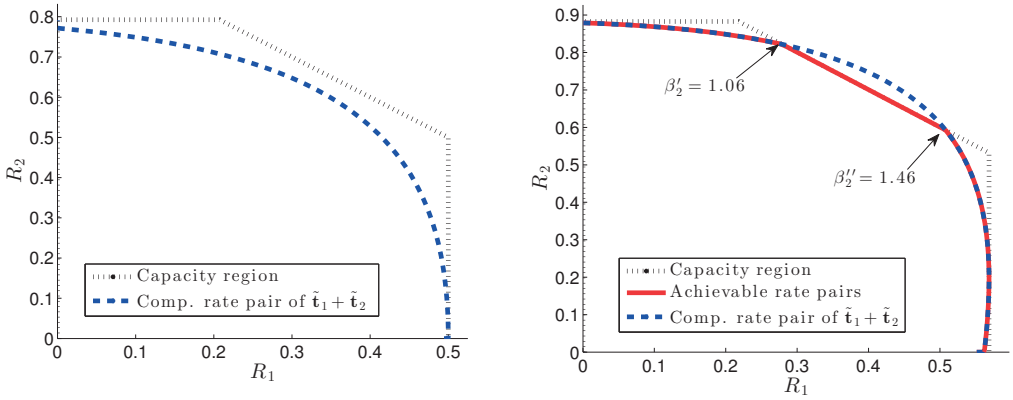
$$R_2 = r_2(\mathbf{a}, \beta_2^{(2)}) = \frac{1}{2}\log(1 + h_2^2 P) \tag{4.43}$$

$$R_1 = r_1(\mathbf{b}|\mathbf{a}, \beta_2^{(2)}) = \frac{1}{2}\log(1 + \frac{h_1^2 P}{1 + h_2^2 P}) \tag{4.44}$$

which is another corner point of the capacity region. If the condition $\beta_2^{(1)}, \beta_2^{(2)} \notin [\beta_2', \beta_2'']$ is not fulfilled, we cannot choose β_2 to be $\beta_2^{(1)}$ or $\beta_2^{(2)}$ hence cannot achieve the corner points of the capacity region. In Figure 4.2b we give an example in this case where only part of rate pairs on the dominant face can be achieved.

In **Case III)** we require $\beta_2^{(1)}, \beta_2^{(2)} \in [\beta_2', \beta_2'']$. In Appendix 4.6.1 we show that $\beta_2^{(1)}, \beta_2^{(2)} \in [\beta_2', \beta_2'']$ if and only if the condition (4.27) is satisfied. With the coefficients $\mathbf{A}_1$, the achievable rate pairs $(r_1(\mathbf{a}, \beta_2), r_2(\mathbf{b}|\mathbf{a}, \beta_2))$ lies on the dominant face by varying β_2 in the interval $[\beta_2^{(1)}, \beta_2'']$ and in this case we do not need to choose β_2 in the interval $[\beta_2', \beta_2^{(1)})$, see Figure 4.3a for an example. Similarly with coefficients $\mathbf{A}_2$, the achievable rate pairs $(r_1(\mathbf{b}|\mathbf{a}, \beta_2), r_2(\mathbf{a}, \beta_2))$ lie on the dominant face by varying β_2 in the interval $[\beta_2', \beta_2^{(2)}]$ and we do not need to let β_2 take values in the interval $(\beta_2^{(2)}, \beta_2'']$, see Figure 4.3b for an example. Since we always have $r_1(\mathbf{a}, \beta_2') \geq r_1(\mathbf{b}|\mathbf{a}, \beta_2'')$ and $r_2(\mathbf{b}|\mathbf{a}, \beta_2') \geq r_2(\mathbf{a}, \beta_2'')$, the achievable rate pairs with coefficients $\mathbf{A}_1$ and $\mathbf{A}_2$ cover the whole dominant face of the capacity region.

As mentioned previously, a similar idea is developed in [28] showing that certain isolated points on the capacity boundary are achievable under certain condition. Before ending the proof, we comment on two main pints in the proposed scheme, which also us to improve upon the previous result. The first point is the introduction of the scaling parameters β_k which allow us to adjust the rates of two users. More precisely, equations (4.13) and (4.19) show that the scaling parameters not only affect the equivalent noise $N_1(\alpha_1)$ and $N_2(\alpha_2, \lambda)$, but also balance the rates of different users (as they also appear in the numerators). We need to adjust the rates of two users carefully through these parameters to make sure that the rate pair lie on the capacity boundary. The second point is that in order to achieve the whole capacity boundary, it is very important to choose the right coefficient of the sum. In particular for the two-user Gaussian MAC, the coefficient for the second sum should be $(1, 0)$ or $(0, 1)$. More discussions on this point is given in the next section. □



(a) Case I with $h_1 = 1, h_2 = \sqrt{2}, P = 1$ (b) Case II with $h_1 = 1, h_2 = \sqrt{2}, P = 1.2$

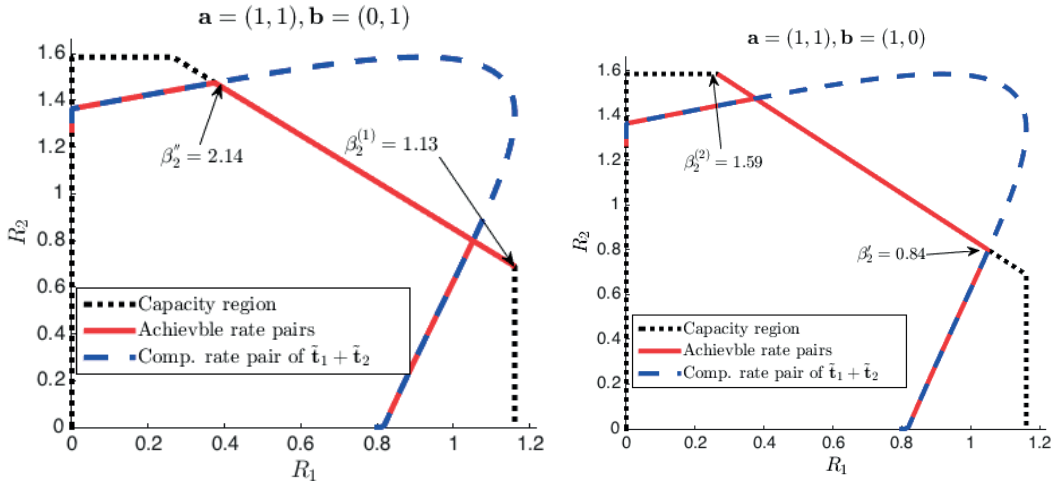
Figure 4.2 – Plot (a) shows the achievable rate pairs in Case I. In this case the condition (4.22) is satisfied and the computation rate pair of the first sum is too small. It has no intersection with the dominant face hence cannot achieve sum rate capacity. Notice that the (message) rate pairs contained in the computation rate region are achievable. Plot (b) shows the situation in Case II. In this case the condition (4.23) is fulfilled and the computation rate pair of the first sum is larger. It intersects with the dominant face. hence the sum capacity is achievable. In this example the condition (4.27) is not satisfied hence only part of the dominant face can be achieved, as depicted in the plot. The rate pair segement on the dominant face can be achieved by choosing $\mathbf{a} = (1, 1)$, $\mathbf{b} = (1, 0)$ or $\mathbf{b} = (0, 1)$ and varying $\beta_2 \in [\beta_2', \beta_2'']$. Choosing β_2 to be β_2', β_2'' gives the end points of the segement. We emphesize that if we choose $\mathbf{a} = (1, 0), \mathbf{b} = (0, 1)$ or $\mathbf{a} = (0, 1), \mathbf{b} = (1, 0)$, i.e., the conventional successive cancellation decoding, we can always achieve the whole capacity region, irrespective of the condition (4.22) or (4.23).

Figure 4.4 shows the achievability of our scheme for different values of received signal-to-noise ratio $h_k^2 P$. In Region III (a sufficient condition is $h_k^2 P \geq 1 + \sqrt{2}$ for $k = 1, 2$), we can achieve any point in the capacity region. In Region I and II the proposed scheme is not able to achieve the entire region. However, we should point out that if we choose the coefficients to be $\mathbf{a} = (1, 0), \mathbf{b} = (0, 1)$ or $\mathbf{a} = (0, 1), \mathbf{b} = (1, 0)$, the CFMA scheme reduces to the conventional successive cancellation decoding, and is *always* able to achieve the corner point of the capacity region, *irrespective of* the values of h_1, h_2 and P.

4.1.1 On the choice of coefficients

In Theorem 4.2 we only considered the coefficients $\mathbf{a} = (1, 1)$, $\mathbf{b} = (1, 0)$ or $\mathbf{b} = (0, 1)$. It is natural to ask whether choosing other coefficients could be advantageous. We first consider the case when the coefficients $\mathbf{a}$ of the first sum is chosen differently.

Lemma 4.1 (Achieving capacity with a different $\mathbf{a}$)**.** *Consider a 2-user Gaussian MAC where the receiver decodes two integer sums of the codewords with coefficients* $\mathbf{a} = (a_1, a_2)$ *and* $\mathbf{b} = (0, 1)$ *or* $\mathbf{b} = (1, 0)$*. Certain rate pairs on the dominant face*



(a) Case III with $h_1 = 1, h_2 = \sqrt{2}, P = 4$ (b) Case III with $h_1 = 1, h_2 = \sqrt{2}, P = 4$

Figure 4.3 – Achievable rate pairs in Case III. The capacity region and the computation rate pairs in the two plots are the same. In this case the condition (4.27) is satisfied hence the computation rate pair of the first sum is large enough to achieve the whole capacity region by decoding two nontrivial integer sums. Plot (a) shows the achievable rate pairs by choosing $\mathbf{a} = (1,1), \mathbf{b} = (0,1)$ and varying $\beta_2 \in [\beta_2^{(1)}, \beta_2'']$. Plot (b) shows the achievable rate pairs by choosing $\mathbf{a} = (1,1), \mathbf{b} = (1,0)$ and varying $\beta_2 \in [\beta_2', \beta_2^{(2)}]$. The union of the achievable rate pairs with coefficients cover the whole dominant face of the capacity region. Recall that we have studied the achievable computation rate region for this channel in Figure 3.1.

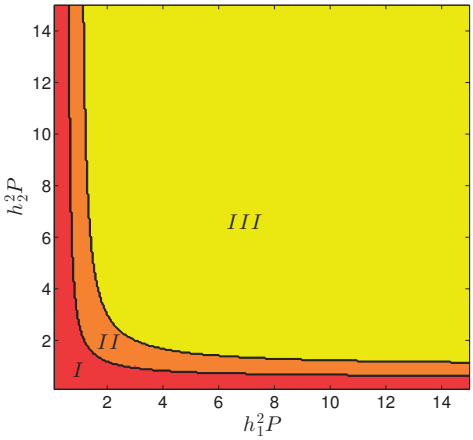


Figure 4.4 – The plane of the received SNR $h_1^2 P, h_2^2 P$ is divided into three regions. Region I corresponds to Case I when the condition (4.22) holds and the scheme cannot achieve points on the boundary of the capacity region. In Region II the condition (4.23) is met but the condition (4.27) is not, hence only part of the points on the capacity boundary can be achieved. Region III corresponds to Case III where (4.27) are satisfied and the proposed scheme can achieve any point in the capacity region.

are achievable if it holds that

$$\frac{h_1 h_2 P}{\sqrt{1+(h_1^2+h_2^2)P}} \geq \frac{4a_1^2a_2^2-1}{4a_1a_2} \tag{4.45}$$

Furthermore the corner points of the capacity region are achievable if it holds that

$$\frac{h_1 h_2 P}{\sqrt{1+(h_1^2+h_2^2)P}} \geq a_1a_2 \tag{4.46}$$

Proof. The proof of the first statement is given in the proof of Theorem 4.2, see Eqn. (4.38). The proof of the second statement is omitted as it is the same as the proof of Case III in Theorem 4.2 with a general $\mathbf{a}$. □

This result suggests that for any $\mathbf{a}$, it is always possible to achieve the sum capacity if the SNR of users are large enough. However the choice $\mathbf{a} = (1, 1)$ is the best, in the sense that it requires the lowest SNR threshold, above which the sum capacity or the whole capacity region is achievable.

To illustrate this, let us reconsider the setting of Figure 4.3, but with coefficients $\mathbf{a}$ different from $(1, 1)$. As can be seen in Figure 4.5a, it is not possible to achieve the sum capacity with $\mathbf{a} = (1, 2)$ or $\mathbf{a} = (2, 1)$. If we increase the power from $P = 4$ to $P = 10$, a part of the capacity boundary is achieved, as shown in Figure 4.5b. We remark that in this case we cannot achieve the whole capacity region with $\mathbf{a} = (1, 2)$ and $\mathbf{a} = (2, 1)$.

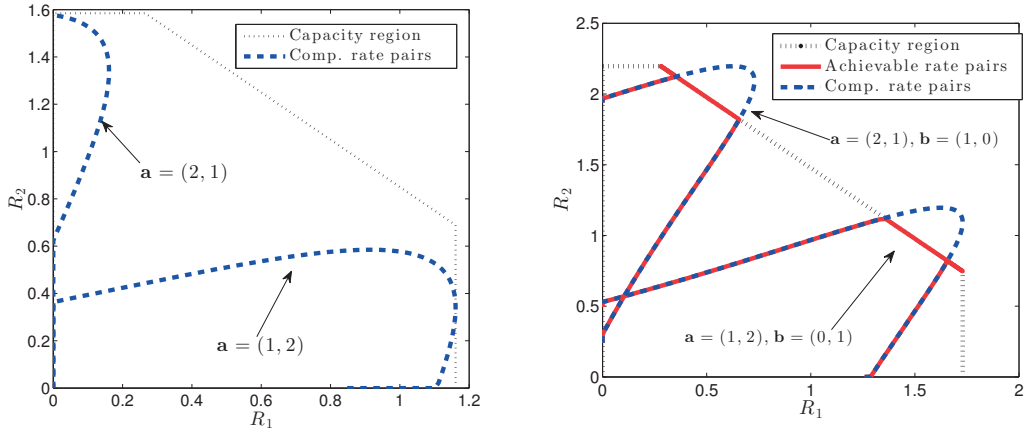
Now we consider a different choice on the coefficients $\mathbf{b}$ of the second sum. Although from the perspective of solving equations, having two sums with coefficients $\mathbf{a} = (1, 1), \mathbf{b} = (1, 0)$ or $\mathbf{a} = (1, 1), \mathbf{b} = (1, 2)$ is equivalent, here it is very important to choose $\mathbf{b}$ such that it has one zero entry. Recall the result in Theorem 4.1 that if $b_k \neq 0$ for $k = 1, 2$, both message rates R_1, R_2 will have two constraints, resulting from the two sums decoded. This extra constraint will diminish the achievable rate region, and in particular it only achieves some isolated points on the dominant face. This is illustrated by the example in Figure 4.6.

As a rule of thumb, the receiver should always decode the sums whose coefficients are as small as possible in a Gaussian MAC.

4.1.2 A comparison with other multiple access techniques

Here we lay out the limitations and possible advantages of CFMA, and compare it with other existing multiple access techniques.

- We have mentioned that one advantage of CFMA scheme is that the decoder used for lattice decoding is a single-user decoder since it only requires performing lattice quantizations on the received signal. Compared to a MAC decoder with joint-decoding, it permits a simpler receiver architecture. In other words, a lattice codes decoder for a point-to-point Gaussian channel can be directly used for a Gaussian MAC with a simple modification. But a joint-decoder needs to perform estimations simultaneously on both messages hence generally has higher complexity.



(a) Achievable (computation) rate pairs with $h_1 = 1, h_2 = \sqrt{2}, P = 4$ and $\mathbf{a} = (1,2)$ or $\mathbf{a} = (2,1)$.

(b) Achievable rate pairs with $h_1 = 1, h_2 = \sqrt{2}, P = 10$ and $\mathbf{a} = (1,2)$ or $\mathbf{a} = (2,1)$.

Figure 4.5 – In the left plot we show the computation rate pairs with parameters $h_1 = 1, h_2 = \sqrt{2}, P = 4$ where the coefficients of the first sum are chosen to be $\mathbf{a} = (1,2)$ or $\mathbf{a} = (2,1)$. In this case the condition (4.45) is not satisfied hence no point on the dominant face can be achieved for the first sum. Compare it to the example in Figure 4.3a or 4.3b where $\mathbf{a} = (1,1)$ and the whole capacity region is achievable. We also note that the achievable computation rate pairs depicted in the Figure are also achievable message rate pairs, which can be shown using Theorem 4.1. In the right plot we show the achievable rate pairs with parameters $h_1 = 1, h_2 = \sqrt{2}, P = 10$ where the coefficient of the first sum is chosen to be $\mathbf{a} = (1,2)$ or $\mathbf{a} = (2,1)$. It can be checked with Lemma 4.1 that we can achieve the sum capacity with the given system parameters. Notice that only parts of the capacity boundary are achievable and we cannot obtain the whole dominant face in this case. In contrast, choosing $\mathbf{a} = (1,1)$ achieves the whole dominant face.

- Compared to the successive cancellation decoding scheme with time sharing, CFMA also performs successive cancellation decoding but does not require time-sharing for achieving the desired rate pairs in the capacity region (provided that the mild condition on SNR is fulfilled).

- The rate-splitting scheme also permits a single-user decoder at the receiver. As shown in [26], $2K-1$ single-user decoders are enough for the rate-splitting scheme in a K-user Gaussian MAC. One disadvantage of this approach is that the messages need to be split into smaller sub-messages and then re-emerged at the receiver. On the other hand, CFMA requires a matrix inversion operation to solve individual messages after collecting different sums which could be computationally expensive. However as shown in an example in Section 4.2.2, we can often choose the matrix to have very special structure and make it very easy to solve for individual messages. Furthermore, CFMA can be combined with rate-splitting where sums of several splitted messages can be decoded. However the combination is not needed in this particular case.

- We also point out that in certain communication scenarios, conventional single-

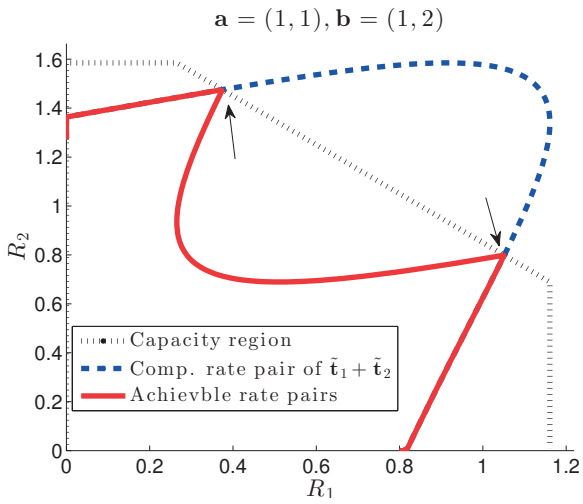


Figure 4.6 – The achievable rate pairs with parameters $h_1 = 1, h_2 = \sqrt{2}, P = 4$. In this case the condition (4.27) is satisfied hence the first sum is chosen properly. But as we choose $\mathbf{b} = (1, 2)$, only two isolated points (indicated by arrows) on the dominant face can be achieved. This is due to the fact non-zero entries in $\mathbf{b}$ will give an extra constraint on the rate, cf. Theorem 4.1. Compare it with the example in Figure 4.3b.

user decoding with time-sharing or rate splitting is not able to achieve the optimal performance. An example for such scenario is the Gaussian interference channel with strong interference. Detailed discussions will be given in the next chapter.

4.2 The K-user Gaussian MAC

In this section, we extend the CFMA scheme to the general K-user Gaussian MAC of the form

$$\mathbf{y} = \sum_{k=1}^{K} h_k \mathbf{x}_k + \mathbf{z} \tag{4.47}$$

with power constraints $||\mathbf{x}_k||^2 \leq nP$. Continuing with the coding scheme for the 2-user Gaussian MAC, in this case the receiver decodes K integer sums with linearly independent coefficients and uses them to solve for the individual messages. The coefficients of the K sums will be denoted by a *coefficient matrix* $\mathbf{A} \in \mathbb{Z}^{K\times K}$

$$\mathbf{A} := (\mathbf{a}_1^T \dots \mathbf{a}_K^T)^T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{22} & a_{22} & \dots & a_{2K} \\ \dots & \dots & \dots & \dots \\ a_{K1} & a_{K2} & \dots & a_{KK} \end{pmatrix} \tag{4.48}$$

where the row vector $\mathbf{a}_\ell := (a_{\ell 1}, \dots, a_{\ell K}) \in \mathbb{Z}^{1\times K}$ denotes the coefficients of the ℓ-th sum, $\sum_{k=1}^{K} a_{\ell k}\tilde{\mathbf{t}}_k$.

The following theorem gives an achievable message rate tuple for the general K-user Gaussian MAC. It is an extension of [28, Thm. 2] as the scaling parameters β_k in CFMA allow a larger achievable rate region.

Theorem 4.3 (Achievable message rate tuples for the K-user Gaussian MAC)**.** *Consider the K-user Gaussian MAC in (4.47). Let $\mathbf{A}$ be a full-rank integer matrix and $\beta_1, \ldots, \beta_K$ be K non-zero real numbers. We define $\mathbf{B} := diag(\beta_1, \ldots, \beta_K)$ and*

$$\mathbf{K}_{\mathbf{Z}'} := P\mathbf{A}\mathbf{B}(\mathbf{I} + P\mathbf{h}\mathbf{h}^T)^{-1}\mathbf{B}^T\mathbf{A}^T \tag{4.49}$$

Let the matrix $\mathbf{L}$ be the unique Cholesky factor of the matrix $\mathbf{A}\mathbf{B}(\mathbf{I}+P\mathbf{h}\mathbf{h}^T)^{-1}\mathbf{B}^T\mathbf{A}^T$, i.e.

$$\mathbf{K}_{\mathbf{Z}'} = P\mathbf{L}\mathbf{L}^T \tag{4.50}$$

The message rate tuple $(R_1, \ldots, R_K)$ is achievable with

$$R_k = \min_{\ell \in [1:K]} \left\{ \frac{1}{2} \log^+ \left(\frac{\beta_k^2}{L_{\ell\ell}^2} \right) \cdot \chi(a_{\ell k}) \right\}, k = 1, \ldots, K \tag{4.51}$$

where we define

$$\chi(x) = \begin{cases} +\infty & \text{if } x = 0, \\ 1 & \text{otherwise.} \end{cases} \tag{4.52}$$

Furthermore if $\mathbf{A}$ is a unimodular ($|\mathbf{A}| = 1$) and R_k is of the form

$$R_k = \frac{1}{2} \log \left(\frac{\beta_k^2}{L_{\Pi(k)\Pi(k)}^2} \right), k = 1, \ldots, K \tag{4.53}$$

for some permutation Π of the set $\{1, \ldots, K\}$, then the sum rate satisfies

$$\sum_k R_k = C_{sum} := \frac{1}{2} \log(1 + \sum_k h_k^2 P) \tag{4.54}$$

Proof. To proof this result, we will adopt a more compact representation and follow the proof technique given in [28]. We rewrite the system in (4.47) as

$$\mathbf{Y} = \mathbf{h}\mathbf{X} + \mathbf{z} \tag{4.55}$$

with $\mathbf{h} = (h_1, \ldots, h_K) \in \mathbb{R}^{1\times K}$ and $\mathbf{X} = (\mathbf{x}_1^T \ldots \mathbf{x}_K^T)^T \in \mathbb{R}^{K\times n}$ where each $\mathbf{x}_k \in R^{1\times n}$ is the transmitted signal sequence of user k given by

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k] \text{mod } \Lambda_k/\beta_k \tag{4.56}$$

Similar to the derivation for the 2-user case, we multiply the channel output by a matrix $\mathbf{F} \in \mathbb{R}^{K\times 1}$ and it can be shown that the following equivalent output can be obtained

$$\tilde{\mathbf{Y}} = \mathbf{A}\mathbf{T} + \tilde{\mathbf{Z}} \tag{4.57}$$

where $\mathbf{T} := (\tilde{\mathbf{t}}_1^T \dots \tilde{\mathbf{t}}_K^T)^T \in \mathbb{R}^{K\times n}$ and the lattice codeword $\tilde{\mathbf{t}}_k \in R^{n\times 1}$ of user k is the same as defined in (4.11). Furthermore the noise $\tilde{\mathbf{Z}} \in \mathbb{R}^{K\times n}$ is given by

$$\tilde{\mathbf{Z}} = (\mathbf{Fh} - \mathbf{AB})\mathbf{X} + \mathbf{Fz} \tag{4.58}$$

where $\mathbf{B} := \mathrm{diag}(\beta_1, \dots, \beta_K)$. The matrix $\mathbf{F}$ is chosen to minimize the variance of the noise:

$$\mathbf{F} := P\mathbf{ABh}^T \left(\frac{1}{P}\mathbf{I} + \mathbf{hh}^T\right)^{-1} \tag{4.59}$$

As shown in the proof of [8, Thm. 5], when analyzing the lattice decoding for the system given in (4.57), we can consider the system

$$\tilde{\mathbf{Y}} = \mathbf{AT} + \mathbf{Z}' \tag{4.60}$$

where $\mathbf{Z}' \in \mathbb{R}^{K\times n}$ is the equivalent noise and each row $\mathbf{z}_k$ is a n-sequence of i.i.d Gaussian random variables z_k for $k = 1, \dots, K$. The covariance matrix of the Gaussians $z_1, \dots, z_K$ is the same as that of the original noise $\tilde{\mathbf{Z}}$ in (4.57). It is easy to show that the covariance matrix of the equivalent noise $z_1, \dots, z_K$ is given in Eq. (4.49).

Now instead of doing the successive interference cancellation as in the 2-user case, we use an equivalent formulation which is called "noise prediction" in [28]. Because the matrix $\mathbf{AB}(\mathbf{I}+P\mathbf{hh}^T)^{-1}\mathbf{B}^T\mathbf{A}^T$ is positive definite, it admits the Cholesky factorization hence the covariance matrix $\mathbf{K}_{\mathbf{Z}'}$ can be rewritten as

$$\mathbf{K}_{\mathbf{Z}'} = P\mathbf{LL}^T \tag{4.61}$$

where $\mathbf{L}$ is a lower triangular matrix.

Using the Cholesky decomposition of $\mathbf{K}_{\tilde{\mathbf{Z}}}$, the system (4.60) can be represented as

$$\begin{aligned}\tilde{\mathbf{Y}} &= \mathbf{AT} + \sqrt{P}\mathbf{LW} \\ &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ a_{K1} & a_{K2} & \dots & a_{KK} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{t}}_1 \\ \tilde{\mathbf{t}}_2 \\ \vdots \\ \tilde{\mathbf{t}}_K \end{pmatrix} + \sqrt{P} \begin{pmatrix} L_{11} & 0 & 0 & \dots & 0 \\ L_{21} & L_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ L_{K1} & L_{K2} & L_{K3} & \dots & L_{KK} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_K \end{pmatrix}\end{aligned} \tag{4.62}$$

with $\mathbf{W} = [\mathbf{w}_1^T, \dots, \mathbf{w}_K^T] \in \mathbb{R}^{K\times n}$ where $\mathbf{w}_i \in \mathbb{R}^{n\times 1}$ is an n-length sequence whose components are i.i.d. zero-mean white Gaussian random variables with unit variance. This is possible by noticing that $\sqrt{P}\mathbf{LW}$ and $\mathbf{Z}'$ have the same covariance matrix. Now we apply lattice decoding to each row of the above linear system. The first row of the equivalent system in (4.62) is given by

$$\tilde{\mathbf{y}}_1 := \mathbf{a}_1\mathbf{T} + \sqrt{P}L_{11}\mathbf{w}_1 \tag{4.63}$$

Using lattice decoding, the first integer sum $\mathbf{a}_1\mathbf{T} = \sum_k a_{1k}\tilde{\mathbf{t}}_k$ can be decoded reliably if

$$r_k < \frac{1}{2}\log^+ \frac{\beta_k^2 P}{PL_{11}^2} = \frac{1}{2}\log^+ \frac{\beta_k^2}{L_{11}^2}, k = 1, \dots, K \tag{4.64}$$

Notice that if a_{1k} equals zero, the lattice point $\tilde{\mathbf{t}}_k$ does not participate in the sum $\mathbf{a}_1\mathbf{T}$ hence r_k is not constrained as above.

The important observation is that knowing $\mathbf{a}_1\mathbf{T}$ allows us to recover the noise term $\mathbf{w}_1$ from $\tilde{\mathbf{y}}_1$. This "noise prediction" is equivalent to the successive interference cancellation, see also [28]. Hence we could eliminate the term $\mathbf{w}_1$ in the second row of the system (4.62) to obtain

$$\tilde{\mathbf{y}}_2 := \mathbf{a}_2\mathbf{T} + \sqrt{P}L_{22}\mathbf{w}_2 \tag{4.65}$$

The lattice decoding of $\mathbf{a}_2\mathbf{T}$ is successful if

$$r_k < \frac{1}{2}\log^+ \frac{\beta_k^2 P}{PL_{22}^2} = \frac{1}{2}\log^+ \frac{\beta_k^2}{L_{22}^2}, k = 1, \ldots, K \tag{4.66}$$

Using the same idea we can eliminate all noise terms $\mathbf{w}_1, \ldots, \mathbf{w}_{\ell-1}$ when decode the ℓ-th sum. Hence the rate constraints on k-th user when decoding the sum $\mathbf{a}_\ell\mathbf{T}$ is given by

$$r_k < \frac{1}{2}\log^+ \frac{\beta_k^2 P}{PL_{\ell\ell}^2} = \frac{1}{2}\log^+ \frac{\beta_k^2}{L_{\ell\ell}^2}, k = 1, \ldots, K \tag{4.67}$$

When decoding the ℓ-th sum, the constraint on r_k will be active only if the coefficient of $\tilde{\mathbf{t}}_k$ is not zero. Otherwise this decoding will not constraint r_k. This fact is captured by introducing the χ function in the statement of this theorem.

In the case when the achievable message rate R_k is of the form

$$R_k = \frac{1}{2}\log\left(\frac{\beta_k^2}{L_{\Pi(k)\Pi(k)}^2}\right) \tag{4.68}$$

The sum rate is

$$\sum_k R_k = \sum_k \frac{1}{2}\log\frac{\beta_k^2}{L_{\Pi(k)\Pi(k)}^2} = \frac{1}{2}\log\prod_k \frac{\beta_k^2}{L_{kk}^2} \tag{4.69}$$

$$= \frac{1}{2}\log\frac{\prod_k \beta_k^2}{|\mathbf{L}\mathbf{L}^T|} = \frac{1}{2}\log\frac{\prod_k \beta_k^2}{|\mathbf{A}\mathbf{B}(\mathbf{I} + P\mathbf{h}\mathbf{h}^T)^{-1}\mathbf{B}^T\mathbf{A}^T|} \tag{4.70}$$

$$= \frac{1}{2}\log|\mathbf{I} + P\mathbf{h}\mathbf{h}^T| + \frac{1}{2}\log\prod_k \beta_k^2 - \log|\mathbf{A}| - \frac{1}{2}\log|\mathbf{B}^T\mathbf{B}| \tag{4.71}$$

$$= \frac{1}{2}\log|\mathbf{I} + P\mathbf{h}\mathbf{h}^T| - \log|\mathbf{A}| \tag{4.72}$$

$$= C_{sum} - \log|\mathbf{A}| \tag{4.73}$$

If $\mathbf{A}$ is unimodular, i.e., $|\mathbf{A}| = 1$, the sum rate is equal to the sum capacity. □

The above theorem says that to achieve the sum capacity, $\mathbf{A}$ needs to be unimodular and R_k should have the form $R_k = \frac{1}{2}\log\frac{\beta_k^2}{L_{\Pi(k)\Pi(k)}^2}$, whose validity also depends on the choice of $\mathbf{A}$. It is difficult to characterize the class of $\mathbf{A}$ for which this holds. In the case when $\mathbf{A}$ is upper triangular with non-zero diagonal entries and $L_{11}^2 \leq \ldots \leq L_{KK}^2$, this condition holds and in fact in this case we have $R_k = \frac{1}{2}\log\frac{\beta_k^2}{L_{kk}^2}$. It can be seen that we are exactly in this situation when we study the 2-user MAC in Theorem 4.2.

4.2.1 An example of a 3-user MAC

It is in general difficult to analytically characterize the achievable rate using our scheme of the K-user MAC. We give an example of a 3-user MAC in Figure 4.7 to help visualize the achievable region. The channel has the form $\mathbf{y} = \sum_{k=1}^{3} \mathbf{x}_3 + \mathbf{z}$ and the receiver decodes three sums with coefficients of the form

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ & \mathbf{e}_i & \\ & \mathbf{e}_j & \end{pmatrix} \tag{4.74}$$

for $i, j = 1, 2, 3$ and $i \neq j$ where $\mathbf{e}_i$ is a row vector with 1 in its i-th and zero otherwise. It is easy to see that there are in total 6 matrices of this form and they all satisfy $|\mathbf{A}| = 1$ hence it is possible to achieve the capacity of this MAC according to Theorem 4.3. For power $P = 8$, most parts of the dominant face are achievable except for three triangular regions. For smaller power $P = 2$, the achievable part of the dominant face shrinks and particularly the symmetric capacity point is not achievable. It can be checked that in this example, no other coefficients will give a larger achievable region.

Unlike the 2-user case, even with a large power, not the whole dominant face can be obtained in this symmetric 3-user MAC. To obtain some intuition why it is the case, we consider one edge of the dominant face indicated by the arrow in Figure 4.7a. If we want to achieve the rate tuple on this edge, we need to decode user 1 last because R_1 attains its maximum. Hence a reasonable choice of the coefficients matrix would be

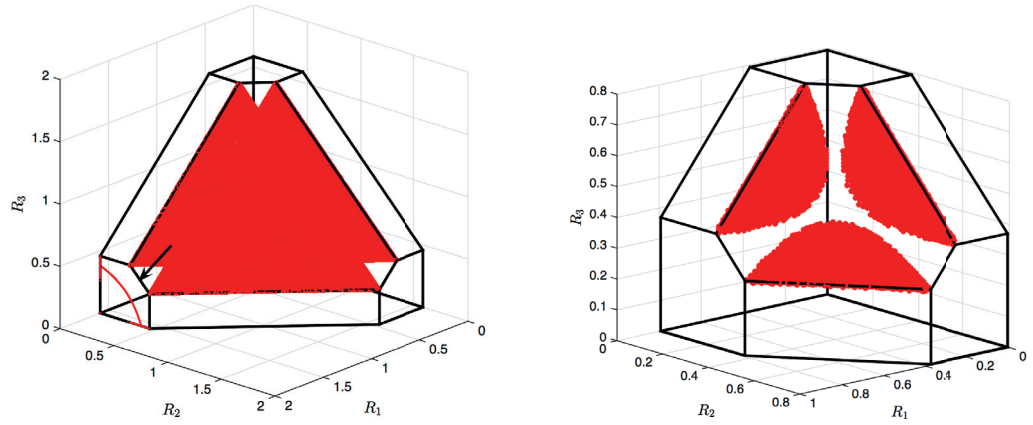
$$\mathbf{A}' = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ or } \mathbf{A}' = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{4.75}$$

Namely we first decode two sums to solve both $\mathbf{t}_2$ and $\mathbf{t}_3$, and then decode $\mathbf{t}_1$ without any interference. When decoding the first two sums, we are effectively dealing with a 2-user MAC while treating $\mathbf{t}_1$ as noise. But the problem is that with $\mathbf{t}_1$ as noise, the signal-to-noise ratio of user 2 and 3 are too high, such that computation rate pair cannot reach the dominant face of the effective 2-user MAC with $\mathbf{t}_1$ being noise. This is the same situation as the Case I considered in Theorem 4.2. In Figure 4.7a we also plot the achievable rates with the coefficients $\mathbf{A}'$ above, on the side face. On the side face where R_1 attains its maximal value, we see the achievable rates cannot reach the dominant face, as a reminiscence of the 2-user example in Figure 4.2a.

4.2.2 The symmetric capacity for the symmetric Gaussian MAC

As it is difficult to obtain a complete description of the achievable rate region for a K-user MAC, in this section we investigate the simple symmetric channel where all the channel gains are the same. In this case we can absorb the channel gain into the power constraint and assume without loss of generality the channel model to be

$$\mathbf{y} = \sum_k \mathbf{x}_k + \mathbf{z} \tag{4.76}$$



(a) $h_k = 1, P = 8$ (b) $h_k = 1, P = 2$

Figure 4.7 – The achievable rate region (red part) in Theorem 4.3 for a symmetric 3-user Gaussian MAC with $h_k = 1$ for $k = 1, 2, 3$ and different powers P.

where the transmitted signal $\mathbf{x}_k$ has an average power constraint P. We want to see if CFMA can achieve the symmetric capacity

$$C_{sym} = \frac{1}{2K} \log(1 + KP) \tag{4.77}$$

For this specific goal, we will fix our coefficient matrix to be

$$\mathbf{A} := \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.78}$$

Namely we first decode a sum involving all codewords $\sum_k \mathbf{t}_k$, then decode the individual codewords one by one. Due to symmetry the order of the decoding procedure is irrelevant and we fix it to be $\mathbf{t}_2, \dots, \mathbf{t}_K$. As shown in Theorem 4.3, the analysis of this problem is closely connected to the Cholesky factor $\mathbf{L}$ defined in (4.50). This connection can be made more explicit if we are interested in the symmetric capacity for the symmetric channel.

We define

$$\mathbf{C} := \begin{pmatrix} 1 & \beta_2 & \beta_3 & \dots & \beta_K \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{4.79}$$

and $\mathbf{E}$ to be the all-one matrix. Let the lower triangular matrix $\tilde{\mathbf{L}}$ denote the unique Cholesky factorization of the matrix $\mathbf{C}(\mathbf{I} - \frac{P}{1+KP}\mathbf{E})\mathbf{C}^T$, i.e.,

$$\mathbf{C}\left(\mathbf{I} - \frac{P}{1+KP}\mathbf{E}\right)\mathbf{C}^T = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \tag{4.80}$$

Proposition 4.1 (Symmetric capacity)**.** *If there exist real numbers $\beta_2, \ldots, \beta_K \geq 1$ with $|\beta_k| \geq 1$ such that the diagonal entries of $\tilde{\mathbf{L}}$ given in (4.80) are equal in amplitude i.e., $|\tilde{L}_{kk}| = |\tilde{L}_{jj}|$ for all k, j, then the symmetric capacity, i.e., $R_k = C_{sym}$ for all k, is achievable for the symmetric K-user Gaussian MAC.*

Proof. Recall we have $\mathbf{B} = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_K)$. Let $\mathbf{A}$ be as given in (4.78) and the channel coefficients $\mathbf{h}$ be the all-one vector. Substituting them into (4.49), (4.50) gives

$$P\tilde{\mathbf{C}}\left(\mathbf{I} - \frac{P}{1+KP}\mathbf{E}\right)\tilde{\mathbf{C}}^T = P\mathbf{L}\mathbf{L}^T \tag{4.81}$$

where

$$\tilde{\mathbf{C}} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \ldots & \beta_K \\ 0 & \beta_2 & 0 & \ldots & 0 \\ 0 & 0 & \beta_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \beta_K \end{pmatrix} \tag{4.82}$$

In this case the we are interested in the Cholesky factorization $\mathbf{L}$ above. Due to the special structure of $\mathbf{A}$ chosen in (4.78), Theorem 4.3 implies that the following rates are achievable

$$R_1 = \frac{1}{2}\log\frac{\beta_1^2}{L_{11}^2} \tag{4.83}$$

$$R_k = \min\left\{\frac{1}{2}\log\frac{\beta_k^2}{L_{11}^2}, \frac{1}{2}\log\frac{\beta_k^2}{L_{kk}^2}\right\}, k \geq 2 \tag{4.84}$$

Using the same argument in the proof of Theorem 4.3, it is easy to show that the sum capacity is achievable if $L_{kk}^2 \geq L_{11}^2$ for all $k \geq 2$. In the case of symmetric capacity we further require that

$$\frac{\beta_k^2}{L_{kk}^2} = \frac{\beta_j^2}{L_{jj}^2} \tag{4.85}$$

for all k, j. This is the same as requiring $\mathbf{B}^{-1}\mathbf{L}$ to have diagonals equal in amplitude with $\mathbf{L}$ given in (4.81), or equivalently requiring the matrix $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}(\mathbf{I} + P\mathbf{h}\mathbf{h}^T)^{-1}\mathbf{B}^T\mathbf{A}^T\mathbf{B}^{-T}$ having Cholesky factorization whose diagonals are equal in amplitude. We can let $\beta_1 = 1$ without loss of generality and it is straightforward to check that in this case $\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \mathbf{C}$. Now the condition in (4.85) is equivalently represented as

$$\tilde{L}_{kk}^2 = \tilde{L}_{jj}^2 \tag{4.86}$$

and the requirement $L_{kk}^2 \geq L_{11}^2$ for $k \geq 2$ can be equivalently written as $\beta_k^2 \geq \beta_1^2 = 1$. □

We point out that the value of power P plays a key role in Proposition 4.1. It is not true that for any power constraint P, there exists $\beta_2, \ldots, \beta_K$ such that the equality condition in Proposition 4.1 can be fulfilled. For the two user case analyzed

in Section 4.1, we can show that for the symmetric channel, the equality condition in Proposition 4.1 can be fulfilled if the condition (4.23) holds, which in turn requires $P \geq 1.5$ for the symmetric channel. In general for a given K, we expect that there exists a threshold $P^*(K)$ such that for $P \geq P^*(K)$, we can always find $\beta_2, \ldots, \beta_K$ which satisfy the equality condition in Proposition 4.1 hence achieve the symmetric capacity. This can be formulated as follows.

Conjecture 4.1 (Achievablity of the symmetric capacity)**.** *For any $K \geq 2$, there exists a positive number $P^*(K)$, such that for $P \geq P^*(K)$, we can find real numbers $\beta_2, \ldots, \beta_K$, where $|\beta_k| \geq 1$ with which the diagonal entries of $\tilde{\mathbf{L}}$ given in (4.80) are equal in amplitude i.e., $|\tilde{L}_{kk}| = |\tilde{L}_{jj}|$ for all k, j.*

We have not been able to prove this claim. Table 4.1 gives some numerical results for the choices of $\underline{\beta}$ which achieve the symmetric capacity in a K-user Gaussian MAC with power constraint $P = 15$ and different values of K. With this power constraint the claim in Conjecture 4.1 is numerically verified with K up to 6. Notice that the value β_k decreases with the index k for $k \geq 2$. This is because with the coefficient matrix $\mathbf{A}$ in (4.78), the decoding order of the individual users is from 2 to K (and user 1 is decoded last). The earlier the message is decoded, the larger the corresponding β will be.

Table 4.1 – The choice of $\underline{\beta}$ for a K-user Gaussian MAC with power $P = 15$.

K	β_1	β_2	β_3	β_4	β_5	β_6
2	1	1.1438				
3	1	1.5853	1.2582			
4	1	1.6609	1.3933	1.1690		
5	1	1.6909	1.4626	1.2796	1.1034	
6	1	1.6947	1.4958	1.3361	1.1980	1.0445

Some numerical results for $P^*(K)$ for K up to 5 is given in Table 4.2. As we have seen $P^*(2) = 1.5$. For other K we give the interval which contains $P^*(K)$ by numerical evaluations.

Table 4.2 – The intervals containing $P^*(K)$

K	$P^*(K)$
2	1.5
3	[2.23, 2.24]
4	[3.74, 3.75]
5	[7.07, 7.08]

4.3 The Two-user Gaussian Dirty MAC

We have shown in previous sections that as an alternative multiple access technique, CFMA enjoys the advantage of being able to achieve the capacity region with low-complexity decoders. In this and subsequent sections we show that CFMA is also useful for others communication systems besides usual multiple access channels.

We first consider the Gaussian MAC with interfering signals which are non-causally known at the transmitters. This channel model is called Gaussian "dirty MAC" and is studied in [30]. Some related results are given in [31], [32], [33]. A two-user Gaussian dirty MAC is given by

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{z} \tag{4.87}$$

where the channel input $\mathbf{x}_1, \mathbf{x}_2$ are required to satisfy the power constraints $\mathbb{E}\{||\mathbf{x}_k||\}^2 \leq P_k, k = 1, 2$ and $\mathbf{z}$ is the white Gaussian noise with unit variance per entry. The interference $\mathbf{s}_k$ is a zero-mean i.i.d. Gaussian random sequence with variance Q_k for each entry, $k = 1, 2$. An important assumption is that the interference signal $\mathbf{s}_k$ is only non-causally known to transmitter k. Two users need to mitigate two interference signals in a distributed manner, which makes this problem challenging. By letting $Q_1 = Q_2 = 0$ we recover the standard Gaussian MAC.

This problem can be seen as an extension of the well-known dirty-paper coding problem [34] to the multiple-access channels. However as shown in [30], a straightforward extension of the usual Gelfand-Pinsker scheme [35] is not optimal and in the limiting case when interference is very strong, the achievable rates are zero. Although the capacity region of this channel is unknown in general, it is shown in [30] that lattice codes are well-suited for this problem and give better performance than the usual random coding scheme.

Now we will extend our coding scheme in previous sections to the dirty MAC. The basic idea is still to decode two linearly independent sums of the codewords. The new ingredient is to mitigate the interference $\mathbf{s}_1, \mathbf{s}_2$ in the context of lattice codes. For a point-to-point AWGN channel with interference known non-causally at the transmitter, it has been shown that capacity can be attained with lattice codes [36]. Our coding scheme is an extension of the schemes in [36] and [30].

Theorem 4.4 (Achievability for the Gaussian dirty MAC)**.** *For the dirty multiple access channel given in (4.87), the following message rate pair is achievable*

$$R_k = \begin{cases} r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}) & \text{if } b_k = 0 \\ r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta}) & \text{if } a_k = 0 \\ \min\{r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}), r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta})\} & \text{otherwise} \end{cases} \tag{4.88}$$

for any linearly independent integer vectors $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^2$ *and* $\underline{\gamma}, \underline{\beta} \in \mathbb{R}^2$ *if* $r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}) > 0$ *and* $r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta}) > 0$ *for* $k = 1, 2$*, whose expressions are given as*

$$r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}) := \max_{\alpha_1} \frac{1}{2} \log^+ \frac{\beta_k^2 P_k}{N_1(\alpha_1, \underline{\gamma}, \underline{\beta})} \tag{4.89}$$

$$r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta}) := \max_{\alpha_2, \lambda} \frac{1}{2} \log^+ \frac{\beta_k^2 P_k}{N_2(\alpha_2, \underline{\gamma}, \underline{\beta}, \lambda)} \tag{4.90}$$

with

$$N_1(\alpha_1, \underline{\gamma}, \underline{\beta}) = \alpha_1^2 + \sum_k \Big((\alpha_1 - a_k\beta_k)^2 P_k + (\alpha_1 - a_k\gamma_k)^2 Q_k \Big) \tag{4.91}$$

$$N_2(\alpha_2, \underline{\gamma}, \underline{\beta}, \lambda) = \alpha_2^2 + \sum_k \Big((\alpha_2 - \lambda a_k\gamma_k - b_k\gamma_k)^2 Q_k + (\alpha_2 - \lambda a_k\beta_k - b_k\beta_k)^2 P_k \Big) \tag{4.92}$$

Proof. Let $\mathbf{t}_k$ be the lattice codeword of user k and $\mathbf{d}_k$ the dither uniformly distributed in $\mathcal{V}_k^s/\beta_k$. The channel input is given as

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k \mathbf{s}_k/\beta_k] \mod \Lambda_k^s/\beta_k$$

for some γ_k to be determined later. In Appendix 4.6.2 we show that with the channel output $\mathbf{y}$ we can form

$$\tilde{\mathbf{y}}_1 := \tilde{\mathbf{z}}_1 + \sum_k a_k \tilde{\mathbf{t}}_k + \sum_k (\alpha_1 - a_k\gamma_k)\mathbf{s}_k \tag{4.93}$$

where α_1 is some real numbers to be optimized later and we define $\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \gamma_k\mathbf{s}_k)$ and $\tilde{\mathbf{z}}_1 := \sum_k(\alpha_1 - a_k\beta_k)\mathbf{x}_k + \alpha_1\mathbf{z}$. Due to the nested lattice construction we have $\tilde{\mathbf{t}}_k \in \Lambda$. Furthermore the term $\tilde{\mathbf{z}}_1 + \sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k$ is independent of the sum $\sum_k a_k\tilde{\mathbf{t}}_k$ thanks to the dither and can be seen as the equivalent noise having average power per dimension $N_1(\alpha, \underline{\gamma}, \underline{\beta})$ in (4.91) for $k = 1, 2$.

In order to decode the integer sum $\sum_k a_k\tilde{\mathbf{t}}_k$ we require

$$r_k < r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}) := \max_{\alpha_1} \frac{1}{2} \log^+ \frac{\beta_k^2 P_k}{N_1(\alpha_1, \underline{\gamma}, \underline{\beta})} \tag{4.94}$$

Notice this constraint on R_k is applicable only if $a_k \neq 0$.

If we can decode $\sum_k a_k\tilde{\mathbf{t}}_k$ with positive rate, the idea of successive interference cancellation can be applied. We show in Appendix 4.6.2 that for decoding the second sum we can form

$$\tilde{\mathbf{y}}_2 := \tilde{\mathbf{z}}_2 + \sum_k (\alpha_2 - \lambda a_k\gamma_k - b_k\gamma_k)\mathbf{s}_k + \sum_k b_k\tilde{\mathbf{t}}_k \tag{4.95}$$

where α_2 and λ are two real numbers to be optimized later and we define $\tilde{\mathbf{z}}_2 := \sum_k(\alpha_2 - \lambda a_k\beta_k - b_k\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z}$. Now the equivalent noise $\tilde{\mathbf{z}}_2 + \sum_k(\alpha_2 - \lambda a_k\gamma_k - b_k\gamma_k)\mathbf{s}_k$ has average power per dimension $N_2(\alpha_2, \underline{\gamma}, \underline{\beta}, \lambda)$ given in (4.92). Using lattice decoding we can show the following rate pair for decoding $\sum_k b_k\tilde{\mathbf{t}}_k$ is achievable

$$r_k < r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta}) := \max_{\alpha_2, \lambda} \frac{1}{2} \log^+ \frac{\beta_k^2 P_k}{N_2(\alpha_2, \underline{\gamma}, \underline{\beta}, \lambda)} \tag{4.96}$$

Again the lattice points $\tilde{\mathbf{t}}_k$ can be solved from the two sums if $\mathbf{a}$ and $\mathbf{b}$ are linearly independent, and $\mathbf{t}_k$ is recovered by the modulo operation $\mathbf{t}_k = [\tilde{\mathbf{t}}_k] \mod \Lambda_k^s$ even if $\mathbf{s}_k$ is not known at the receiver. If we have $b_k = 0$, the above constraint does not apply to R_k. □

4.3.1 Decoding one integer sum

We revisit the results obtained in [30] and show they can be obtained in our framework in a unified way.

Theorem 4.5 ([30] Theorem 2, 3)**.** *For the dirty multiple access channel given in (4.87), we have the following achievable rate region:*

$$R_1+R_2 = \begin{cases} \frac{1}{2}\log(1+\min\{P_1,P_2\}) & \text{if } \sqrt{P_1P_2}-\min\{P_1,P_2\}\geq 1 \\ \frac{1}{2}\log^+\left(\frac{P_1+P_2+1}{2+(\sqrt{P_1}-\sqrt{P_2})^2}\right) & \text{otherwise} \end{cases}$$

Proof. A proof is given in Appendix 4.6.3. □

In [30], the above rate region was obtained by considering the transmitting scheme where only one user transmits at a time. In our framework, it is the same as assuming one transmitted signal, say $\mathbf{t}_1$, is set to be $\mathbf{0}$ and known to the decoder. In this case we need only one integer sum to decode $\mathbf{t}_2$. Here we give a proof to show the achievability for

$$R_2 = \begin{cases} \frac{1}{2}\log(1+P_2) & \text{for } P_1 \geq \frac{(P_2+1)^2}{P_2} \\ \frac{1}{2}\log(1+P_1) & \text{for } P_2 \geq \frac{(P_1+1)^2}{P_1} \\ \frac{1}{2}\log^+\left(\frac{P_1+P_2+1}{2+(\sqrt{P_1}-\sqrt{P_2})^2}\right) & \text{otherwise} \end{cases} \tag{4.97}$$

while $R_1 = 0$. Theorem 4.5 is obtained by showing the same result holds when we switch the two users and a time-sharing argument.

An outer bound on the capacity region given in [30, Corollary 2] states that the sum rate capacity should satisfy

$$R_1 + R_2 \leq \frac{1}{2}\log(1+\min\{P_1,P_2\}) \tag{4.98}$$

for strong interference (both Q_1, Q_2 go to infinity). Hence in the strong interference case, the above achievability result is either optimal (when P_1, P_2 are not too close) or only a constant away from the capacity region (when P_1, P_2 are close, see [30, Lemma 3]). However the rates in Theorem 4.5 are strictly suboptimal for general interference strength as we will show in the sequel.

4.3.2 Decoding two integer sums

Now we consider decoding two sums for the Gaussian dirty MAC by evaluating the achievable rates stated in Theorem 4.4. Unlike the case of the clean MAC studied in Section 4.1, here we need to optimize over $\underline{\gamma}$ for given $\mathbf{a}, \mathbf{b}$ and $\underline{\beta}$, which does not have a closed-form solution due to the $\min\{\cdot\}$ operation. Hence in this section we resort to numerical methods for evaluations. To give an example of the advantage for decoding two sums, we show achievable rate regions in Figure 4.8 for a dirty MAC where $P_1 = Q_1 = 10$ and $P_2 = Q_2 = 2$. We see in the case when the transmitting power and interference strength are comparable, decoding two sums gives a significantly larger achievable rate region. In this example we choose the coefficients to be $\mathbf{a} = (a_1, 1), \mathbf{b} = (1, 0)$ or $\mathbf{a} = (1, a_2), \mathbf{b} = (1, 0)$ for $a_1, a_2 = 1, \ldots, 5$

and optimize over parameters $\underline{\gamma}$. We also point out that unlike the case of the clean MAC where it is best to choose a_1, a_2 to be 1, here choosing coefficients a_1, a_2 other than 1 gives larger achievable rate regions in general.

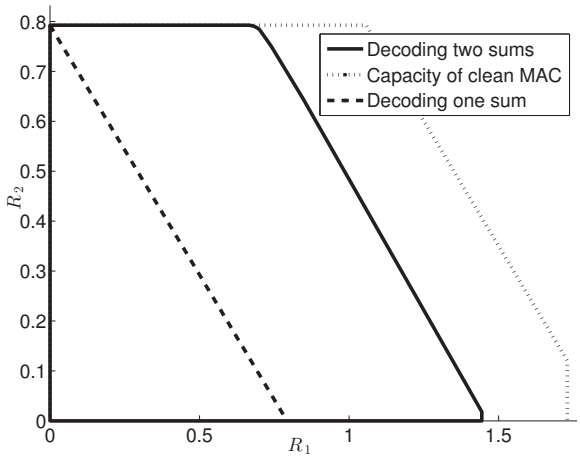


Figure 4.8 – We consider a dirty MAC with $P_1 = Q_1 = 10$ and $P_2 = Q_2 = 2$. The dashed line is the achievable rate region given in Theorem 4.5 from [30] which corresponds to decoding only one sum. The solid line gives the achievable rate region in Theorem 4.4 by decoding two sums with the coefficients $\mathbf{a} = (a_1, 1), \mathbf{b} = (1, 0)$ or $\mathbf{a} = (1, a_2), \mathbf{b} = (1, 0)$ for $a_1, a_2 = 1, \ldots, 5$ and optimizing over parameters $\underline{\gamma}$.

Different from the point-to-point Gaussian channel with interference known at the transmitter, it is no longer possible to eliminate all interference completely without diminishing the capacity region for the dirty MAC. The proposed scheme provides us with a way of trading off between eliminating the interference and treating it as noise. Figure 4.9 shows the symmetric rate of the dirty MAC as a function of interference strength. When the interference is weak, the proposed scheme balances the residual interference $\mathbf{s}_1, \mathbf{s}_2$ in N_1 and N_2 by optimizing the parameters $\underline{\gamma}$, see Eqn. (4.91) and Eqn. (4.92). This is better than only decoding one sum in which we completely cancel out the interference.

As mentioned in the previous subsection, decoding one integer sum is near-optimal in the limiting case when both interference signals $\mathbf{s}_1, \mathbf{s}_2$ are very strong, i.e., $Q_1, Q_2 \to \infty$. It is natural to ask if we can do even better by decoding two sums in this case. It turns out in the limiting case we are not able to decode two linearly independent sums with this scheme.

Lemma 4.2 (Only one sum for high interference)**.** *For the 2-user dirty MAC in (4.87) with $Q_1, Q_2 \to \infty$, we have $r_k(\mathbf{a}, \underline{\gamma}, \underline{\beta}) = r_k(\mathbf{b}|\mathbf{a}, \underline{\gamma}, \underline{\beta}) = 0, k = 1, 2$ for any linearly independent $\mathbf{a}, \mathbf{b}$ where $a_k \neq 0$, $k = 1, 2$.*

Proof. The rate expressions in (4.94) and (4.96) show that we need to eliminate all terms involving Q_k in the equivalent noise N_1 in (4.91) and N_2 in (4.92), in order to have a positive rate when $Q_1, Q_2 \to \infty$. Consequently we need $\alpha_1 - a_k\gamma_k = 0$ and

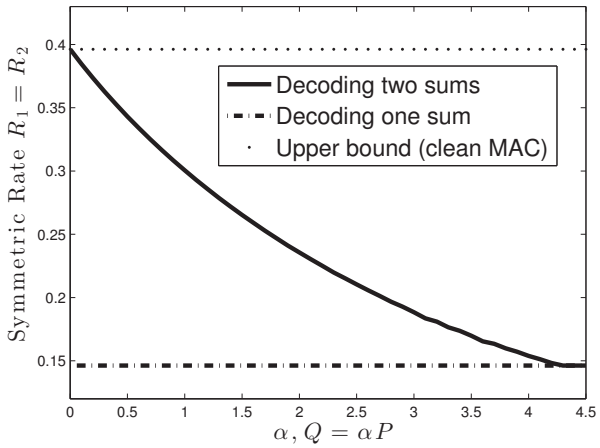


Figure 4.9 – We consider a dirty MAC with $P_1 = P_2 = 1$ and $Q_1 = Q_2 = \alpha P_1$ with different α varying from $[0, 4.5]$. The vertical axis denotes the maximum symmetric rate $R_1 = R_2$. The dotted line is the maximum symmetric rate $1/4 \log(1+P_1+P_2)$ for a clean MAC as an upper bound. The dashed line gives the achievable symmetric rate in Theorem 4.5 from [30] and the solid line depicts the symmetric rate in Theorem 4.4 by decoding two sums.

$\alpha_2 - \lambda a_k \gamma_k - b_k \gamma_k = 0$ for $k = 1, 2$. or equivalently

$$\begin{pmatrix} 1 & 0 & -a_1 & 0 \\ 0 & 1 & -\lambda a_1 - b_1 & 0 \\ 1 & 0 & 0 & -a_2 \\ 0 & 1 & 0 & -\lambda a_2 - b_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \mathbf{0} \tag{4.99}$$

Performing elementary row operations gives the following equivalent system

$$\begin{pmatrix} 1 & 0 & -a_1 & 0 \\ 0 & 1 & -\lambda a_1 - b_1 & 0 \\ 0 & 0 & a_1 & -a_2 \\ 0 & 0 & 0 & \frac{a_2(\lambda a_1 + b_1)}{a_1} - \lambda a_2 - b_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \mathbf{0} \tag{4.100}$$

To have non-trivial solutions of $\underline{\alpha}$ and γ with $a_1 \neq 0$, we must have $\frac{a_2(\lambda a_1 + b_1)}{a_1} - \lambda a_2 - b_2 = 0$, which simplifies to $a_2 b_1 = a_1 b_2$, meaning $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent. □

This observation suggests that when both interference signals are very strong, the strategy in [30] to let only one user transmit at a time (section 4.3.1) is the best thing to do within this framework. However we point out that in the case when only one interference is very strong (either Q_1 or Q_2 goes to infinity), we can still decode two independent sums with positive rates. For example consider the system in (4.87) with $\mathbf{s}_2$ being identically zero, $\mathbf{s}_1$ only known to User 1 and $Q_1 \to \infty$. In this case we can decode two linearly independent sums with $\mathbf{a} = (1, 1), \mathbf{b} = (1, 0)$ or $\mathbf{a} = (1, 0), \mathbf{b} = (0, 1)$. The resulting achievable rates with Theorem 4.4 is the same as that given in [30, Lemma 9]. Moreover, the capacity region of the dirty MAC with only one interference signal commonly known to both users [30, VIII] can also be achieved using Theorem 4.4, by choosing $\mathbf{a} = (1, 0), \mathbf{b} = (0, 1)$ for example.

4.4 The Gaussian Two-Way Relay Channel

In this section we consider the Gaussian two-way relay channel shown in Figure 4.10 where two transceivers wish to send their messages to each other via a relay, with a similar approach studied in [23], [37]. Two encoders have different power constraints P_1 and P_2 and the channel gain from both transmitters is 1. The relay has power constraint P_R. All noises are Gaussian with unit variance.

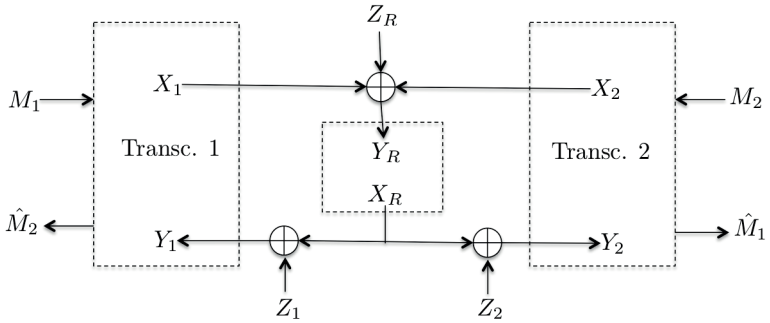


Figure 4.10 – A Gaussian two-way relay channel.

Already shown in [23], [37], it can be beneficial for the relay to decode a linear combination of the two messages rather than decoding the two messages individually. They give the following achievable rate for this network

$$R_1 \leq \min\left\{\frac{1}{2}\log^+\left(\frac{P_1}{P_1+P_2}+P_1\right), \frac{1}{2}\log(1+P_R)\right\} \tag{4.101a}$$

$$R_2 \leq \min\left\{\frac{1}{2}\log^+\left(\frac{P_2}{P_1+P_2}+P_2\right), \frac{1}{2}\log(1+P_R)\right\} \tag{4.101b}$$

where the relay decodes the function $\mathbf{t}_1+\mathbf{t}_2$ and broadcasts it to two decoders. With the general compute-and-forward scheme we also ask the relay to decode a linear combination of the form $\sum_{k=1}^{2} a_k\mathbf{t}_k$ where $a_1, a_2 \neq 0$, with which each decoder can solve for the desired message.

Theorem 4.6. *For the Gaussian two-way relay channel where user k has power P_k and relay has power P_R, the following rate pair is achievable:*

$$R_1 \leq \min\left\{\frac{1}{2}\log^+\left(\frac{P_1\beta_1^2}{\tilde{N}(\beta_1,\beta_2)}\right), \frac{1}{2}\log\frac{\beta_1^2 P_1(1+P_R)}{P_R}\}\right\}$$

$$R_2 \leq \min\left\{\frac{1}{2}\log^+\left(\frac{P_2\beta_2^2}{\tilde{N}(\beta_1,\beta_2)}\right), \frac{1}{2}\log\frac{\beta_2^2 P_2(1+P_R)}{P_R}\}\right\}$$

where

$$\tilde{N}(\beta_1,\beta_2) := \frac{P_1P_2(a_1\beta_1-a_2\beta_2)^2+(a_1\beta_1)^2P_1+(a_2\beta_2)^2P_2}{P_1+P_2+1}$$

for any positive β_1, β_2 satisfying $\max\{\beta_1^2P_1, \beta_2^2P_2\} \leq P_R$.

Proof. A proof of this theorem is given in Appendix 4.6.4. □

Now we show the achievable rate region in Theorem 4.6 and compare to the existing results in [23], [37] with the help of an example in Figure 4.11.

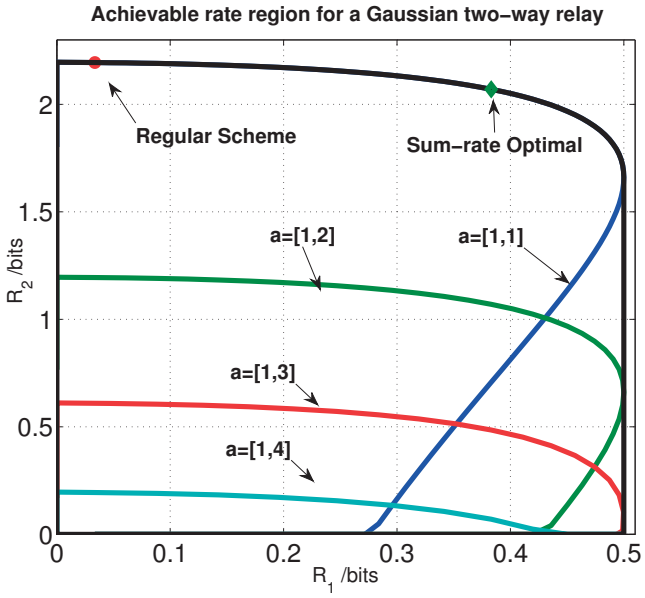


Figure 4.11 – Achievable rate region for the Gaussian two-way relay in Figure 4.10 with unequal power constraints $P_1 = 1$, $P_2 = 20$ and equal channel gain $\mathbf{h} = [1, 1]$. The relay has power $P_R = 20$. Color curves show different achievable rate region when the relay decodes different linear functions as marked in the plot. The red dot denotes the achievable rate pair given in (4.101) when relay decodes $\mathbf{t}_1 + \mathbf{t}_2$ using regular compute-and-forward (other function will give worse rate pair). Notice this point is not sum-rate optimal. The achievable rate region given by the black convex hull is strictly larger than the regular scheme since the CSI can be used at the transmitters.

4.5 Linear Integer-Forcing Receivers for MIMO Channels

We now apply the same idea to the MIMO system with an integer-forcing linear receiver [38]. We consider a point-to-point MIMO system with channel matrix $\mathbf{H} \in \mathbb{R}^{M\times K}$ which is full rank. It is shown in [38] that the following rate is achievable using an integer-forcing receiver

$$R_{IF} \leq \min_{m\in[1:k]} -\frac{K}{2} \log \mathbf{a}_m^T \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{a}_m$$

for any full rank integer matrix $\mathbf{A} \in \mathbb{Z}^{K\times K}$ with its m-th row as $\mathbf{a}_m$ and $\mathbf{V} \in \mathbb{R}^{K\times K}$ is composed of the eigenvectors of $\mathbf{H}^T\mathbf{H}$. The matrix $\mathbf{D} \in \mathbb{R}^{K\times K}$ is diagonal with element $\mathbf{D}_{k,k} = 1/(P\lambda_k^2 + 1)$ and λ_k is the k-th singular value of $\mathbf{H}$.

Applying the general compute-and-forward to the integer-forcing receiver gives the following result. We note that a similar idea also appears in [39] where a precoding matrix is used at the encoder.

Theorem 4.7. *For a $K \times M$ real MIMO system with full rank channel matrix $\mathbf{H} \in \mathbb{R}^{M\times K}$, the following rate is achievable using an integer-forcing linear receiver*

for any $\beta_1, \ldots, \beta_K$

$$R < \leq \sum_{k=1}^{K} \min_{m \in [1:K]} \left(-\frac{1}{2} \log \frac{\tilde{\mathbf{a}}_m^T \mathbf{V} \mathbf{D} \mathbf{V}^T \tilde{\mathbf{a}}_m}{\beta_k^2} \right) \tag{4.102}$$

for any full rank $\mathbf{A} \in \mathbb{Z}^{K \times K}$ *with its* m*-th row being* $\mathbf{a}_m$. $\tilde{\mathbf{a}}$ *is defined as* $\tilde{\mathbf{a}}_m := [\beta_1 a_{m1}, ..., \beta_K a_{mK}]$ *for* $m = 1, \ldots, K$ *and* $\mathbf{V}, \mathbf{D}$ *are defined as above.*

In Figure 4.12 we compare the achievable rates of two schemes on a 2×2 MIMO system with the channel matrix $\mathbf{H} = \begin{bmatrix} 0.7 & 1.3 \\ 0.8 & 1.5 \end{bmatrix}$.

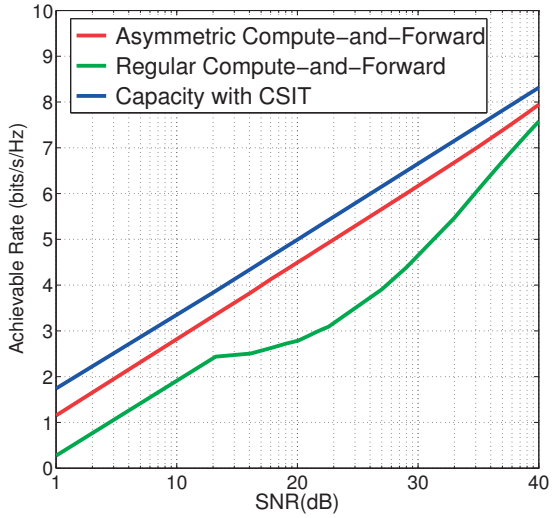


Figure 4.12 – Achievable rates for a 2×2 MIMO system $\mathbf{H} = [0.7, 1.3; 0.8, 1.5]$. At SNR = 40dB, the best coefficients for regular scheme are $\mathbf{a}_1 = [1, 2]$ and $\mathbf{a}_2 = [7, 13]$, while for the modified scheme we have the best parameters as $\beta_1 = 1, \beta_2 = 4.887, \mathbf{a}_1 = [8, 3]$ and $\mathbf{a}_2 = [13, 5]$.

Lastly we give another example where the integer-forcing receiver with the general compute-and-forward scheme performs arbitrarily better than the original scheme. Consider the 2×2 MIMO channel with channel matrix $\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 0 & \epsilon \end{bmatrix}$ where $0 < \epsilon < 1$. It has been shown in [38, Section V, C] that the achievable rate of integer forcing is upper bounded as $R_{IF} \leq \log(\epsilon^2 P)$ which is of order $O(1)$ if $\epsilon \sim \frac{1}{\sqrt{P}}$ while the joint ML decoding can achieve a rate at least $\frac{1}{2} \log(1 + 2P)$. With the modified scheme we can show the following result.

Lemma 4.3. *For the channel* $\mathbf{H}$ *above, the rate expression* R *in (4.102) scales as* $\log P$ *for any* $\epsilon > 0$.

To see this, we can show (assuming w. l. o. g. $\beta_1 = 1$)

$$R_{mIF} \geq \min_{m=1,2} \frac{1}{2} \log^+ \left(\frac{P}{a_{m1}^2 + (a_{m2}\beta_2 - a_{m1})^2 \frac{1}{\epsilon^2}} \right) + \min_{m=1,2} \frac{1}{2} \log^+ \left(\frac{\beta_2^2 P}{a_{m1}^2 + (a_{m2}\beta_2 - a_{m1})^2 \frac{1}{\epsilon^2}} \right)$$

Based on the standard results on simultaneous Diophantine approximation [40], for any given a_{m2} and $Q > 0$ there exists $\beta_2 < Q$ and a_{m1} such that $|a_{m2}\beta_2 - a_{m1}| < Q^{-1/2}$ for $m = 1, 2$. Hence the we have the achievable rate

$$\min_{m=1,2} \frac{1}{2} \log^+ \left(\frac{P}{a_{m1}^2 + Q^{-1} \frac{1}{\epsilon^2}} \right) + \min_{m=1,2} \frac{1}{2} \log^+ \left(\frac{\beta_2^2 P}{a_{m1}^2 + Q^{-1} \frac{1}{\epsilon^2}} \right)$$

If we choose $Q \sim \epsilon^{-2}$, and notice that we also have $\beta_2, a_{m1} \sim Q$, then the second term above scales as $\frac{1}{2} \log P$ for P large. Consequently R_{mIF} also scales as $\frac{1}{2} \log P$ for any ϵ, hence can be arbitrarily better than the regular scheme.

4.6 Appendix

4.6.1 Derivations in the proof of Theorem 4.2

Here we prove the claim in Theorem 4.2 that $\beta_2^{(1)}, \beta_2^{(2)} \in [\beta_2', \beta_2'']$ if and only if the Condition (4.27) holds. Recall we have defined $\beta_2^{(1)} := \frac{h_1 h_2 P}{1+h_1^2 P}$, $\beta_2^{(2)} := \frac{1+h_2^2 P}{h_1 h_2 P}$ and β_2', β_2'' in Eqn. (4.35).

With the choice $\mathbf{a} = (1, 1)$ we can rewrite (4.35) as

$$\beta_2' := \frac{2h_1 h_2 P + S - \sqrt{SD}}{2(1 + h_1^2 P)} \tag{4.103}$$

$$\beta_2'' := \frac{2h_1 h_2 P + S + \sqrt{SD}}{2(1 + h_1^2 P)} \tag{4.104}$$

with $S := \sqrt{1 + h_1^2 P + h_2^2 P}$ and $D := 4Ph_1 h_2 - 3S$. Clearly the inequality $\beta_2' \le \beta_2^{(1)}$ holds if and only if $S - \sqrt{SD} \le 0$ or equivalently

$$\frac{Ph_1 h_2}{\sqrt{1 + h_1^2 P + h_2^2 P}} \ge 1 \tag{4.105}$$

which is just Condition (4.27). Furthermore notice that $\beta_2^{(1)} < \frac{h_2}{h_1} P < \beta_2^{(2)}$ hence it remains to prove that $\beta_2^{(2)} \le \beta_2''$ if and only if (4.27) holds. But this follows immediately by noticing that $\beta_2^{(2)} \le \beta_2''$ can be rewritten as

$$2S^2 \le h_1 h_2 P(S + \sqrt{SD}) \tag{4.106}$$

which is satisfied if and only if $S \le D$, or equivalently Condition (4.27) holds.

4.6.2 Derivations in the proof of Theorem 4.4

In this section we give the derivation of the expressions of $\tilde{\mathbf{y}}_1$ in (4.93) and $\tilde{\mathbf{y}}_2$ in (4.95). To obtain $\tilde{\mathbf{y}}_1$, we process the channel output $\mathbf{y}$ as

$$\begin{aligned} \tilde{\mathbf{y}}_1 &:= \alpha_1 \mathbf{y} - \sum_k a_k \beta_k \mathbf{d}_k \\ &= \underbrace{\sum_k (\alpha_1 - a_k \beta_k) \mathbf{x}_k + \alpha_1 \mathbf{z}}_{\tilde{\mathbf{z}}_1} + \alpha_1 \sum_k \mathbf{s}_k + \sum_k a_k \beta_k \mathbf{x}_k - \sum_k a_k \beta_k \mathbf{d}_k \end{aligned}$$

$$
\begin{aligned}
&= \tilde{\mathbf{z}}_1 + \alpha_1 \sum_k \mathbf{s}_k + \sum_k a_k\beta_k(\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k\mathbf{s}_k/\beta_k) \\
&- \sum_k a_k\beta_k Q_{\Lambda_k^s/\beta_k}(\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k\mathbf{s}_k/\beta_k) - \sum_k a_k\beta_k\mathbf{d}_k \\
&= \tilde{\mathbf{z}}_1 + \sum_k a_k(\mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \alpha_1\mathbf{s}_k)) + \sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k \\
&= \tilde{\mathbf{z}}_1 + \sum_k a_k\tilde{\mathbf{t}}_k + \sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k
\end{aligned}
$$

When the sum $\sum_k a_k\tilde{\mathbf{t}}_k$ is decoded, the term $\tilde{\mathbf{z}}_1 + \sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k$ which can be calculated using $\tilde{\mathbf{y}}_1$ and $\sum_k a_k\tilde{\mathbf{t}}_k$. For decoding the second sum we form the following with some numbers α_2' and λ:

$$
\begin{aligned}
\tilde{\mathbf{y}}_2 &:= \alpha_2'\mathbf{y} + \lambda\left(\tilde{\mathbf{z}}_1 + \sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k\right) - \sum_k b_k\beta_k\mathbf{d}_k \\
&= \alpha_2'(h_1\mathbf{x}_1 + h_2\mathbf{x}_2 + \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{z}) + \sum_k(\lambda\alpha_1 h_k - \lambda a_k\beta_k)\mathbf{x}_k + \lambda\alpha_1\mathbf{z} + \lambda\sum_k(\alpha_1 - a_k\gamma_k)\mathbf{s}_k \\
&= \sum_k(\alpha_2' + \lambda\alpha_1 - \lambda a_k\beta_k)\mathbf{x}_k + (\alpha_2' + \lambda\alpha_1)\mathbf{z} + \sum_k(\alpha_2' + \lambda\alpha_1 - \lambda a_k\gamma_k)\mathbf{s}_k - \sum_k b_k\beta_k\mathbf{d}_k \\
&:= \sum_k(\alpha_2 - \lambda a_k\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z} + \sum_k(\alpha_2 - \lambda a_k\gamma_k)\mathbf{s}_k - b_k\beta_k\mathbf{d}_k
\end{aligned}
$$

by defining $\alpha_2 := \alpha_2' + \lambda\alpha_1$. In the same way as deriving $\tilde{\mathbf{y}}_1$, we can show

$$
\begin{aligned}
\tilde{\mathbf{y}}_2 &= \underbrace{\sum_k(\alpha_2 - \lambda a_k\beta_k - b_k\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z}}_{\tilde{\mathbf{z}}_2} + \sum_k(\alpha_2 - \lambda a_k\gamma_k)\mathbf{s}_k + \sum_k b_k\beta_k\mathbf{x}_k - \sum_k b_k\beta_k\mathbf{d}_k \\
&= \tilde{\mathbf{z}}_2 + \sum_k(\alpha_2 - a_k\gamma_k)\mathbf{s}_k \\
&+ \sum_k b_k\left(\beta_k(\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k\mathbf{s}_k/\beta_k) - \beta_k Q_{\Lambda_k^s/\beta_k}(\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k\mathbf{s}_k/\beta_k)\right) - \sum_k b_k\beta_k\mathbf{d}_k \\
&= \tilde{\mathbf{z}}_2 + \sum_k(\alpha_2 - \lambda a_k\gamma_k - b_k\gamma_k)\mathbf{s}_k + \sum_k b_k\tilde{\mathbf{t}}_k
\end{aligned}
$$

by defining $\alpha_2 := \alpha_2' + \lambda\alpha_1$ and $\tilde{\mathbf{z}}_2 := \sum_k(\alpha_2 - \lambda a_k\beta_k - b_k\beta_k)\mathbf{x}_k + \alpha_2\mathbf{z}$.

4.6.3 Proof of Theorem 4.5

Proof. Choosing $\mathbf{a} = (1, 1)$ and $\gamma_1 = \gamma_2 = \alpha_1$ in (4.94), we can decode the integer sum $\sum_k \tilde{\mathbf{t}}_k$ if

$$
r_2 < r_2(\mathbf{a}, \underline{\beta}) = \frac{1}{2}\log\frac{P_2(1 + P_1 + P_2)}{r^2P_1 + P_2 + P_1P_2(r-1)^2} \tag{4.107}
$$

by choosing the optimal $\alpha_1^* = \frac{\beta_1P_1 + \beta_2P_2}{P_1 + P_2 + 1}$ and defining $r := \beta_1/\beta_2$. An important observation is that in order to extract $\mathbf{t}_2$ from the integer sum (assuming $\mathbf{t}_1 = \mathbf{0}$)

$$
\sum_k \tilde{\mathbf{t}}_k = \mathbf{t}_2 - Q_{\Lambda_2^s}(\mathbf{t}_2 + \beta_2\mathbf{d}_2 - \gamma_2\mathbf{s}_2) - Q_{\Lambda_1^s}(\beta_1\mathbf{d}_1 - \gamma_1\mathbf{s}_1),
$$

one sufficient condition is $\Lambda_1^s \subseteq \Lambda_2^s$. Indeed, due to the fact that $[\mathbf{x}] \bmod \Lambda_2^s = 0$ for any $\mathbf{x} \in \Lambda_1^s \subseteq \Lambda_2^s$, we are able to recover $\mathbf{t}_2$ by performing $[\sum_k \tilde{\mathbf{t}}_k] \bmod \Lambda_2^s$ if $\Lambda_1^s \subseteq \Lambda_2^s$. This requirement amounts to the condition $\beta_1^2 P_1 \geq \beta_2^2 P_2$ or equivalently $r \geq \sqrt{P_2/P_1}$. Notice if we can extract $\mathbf{t}_2$ from just one sum $\sum_k \tilde{\mathbf{t}}_k$ (with $\mathbf{t}_1$ known), then the computation rate $R_2^{\mathbf{a}} = r_2(\mathbf{a}, \underline{\beta})$ will also be the message rate $R_2 = r_2(\mathbf{a}, \underline{\beta})$.

Taking derivative w. r. t. r in (4.107) gives the critical point

$$r^* = \frac{P_2}{P_2 + 1} \tag{4.108}$$

If $r^* \geq \sqrt{P_2/P_1}$ or equivalently $P_1 \geq \frac{(P_2+1)^2}{P_2}$, substituting r^* in (4.107) gives

$$R_2 = \frac{1}{2} \log(1 + P_2)$$

If $r^* \leq \sqrt{P_2/P_1}$ or equivalently $P_1 \leq \frac{(P_2+1)^2}{P_2}$, R_2 is non-increasing in r hence we should choose $r = \sqrt{P_2/P_1}$ to get

$$R_2 = \frac{1}{2} \log^+ \left(\frac{1 + P_1 + P_2}{2 + (\sqrt{P_2} - \sqrt{P_1})^2} \right) \tag{4.109}$$

To show the result for the case $P_2 \geq \frac{(P_1+1)^2}{P_1}$, we set the transmitting power of user 2 to be $P_2' = \frac{(P_1+1)^2}{P_1}$ which is smaller or equal to its full power P_2 under this condition. In order to satisfy the nested lattice constraint $\Lambda_1^s \subseteq \Lambda_2^s$ we also need $\beta_1^2 P_1 \leq \beta_2^2 P_2'$ or equivalently $r \geq \sqrt{P_2'/P_1}$. By replacing P_2 by the above P_2' and choosing $r = \sqrt{P_2'/P_1}$ in (4.107) we get

$$R_2 = \frac{1}{2} \log(1 + P_1) \tag{4.110}$$

Interestingly under this scheme, letting the transmitting power to be P_2' gives a larger achievable rate than using the full power P_2 in this power regime. □

4.6.4 Proof of Theorem 4.6

Proof. Let Λ_1^s, Λ_2^s be simultaneously good nested lattices with second moment $\beta_1^2 P_1, \beta_2^2 P_2$, respectively, which are nested in another simultaneously good lattice Λ. The lattice codes for user $k, k = 1, 2$ is constructed as $\mathcal{C}_k = \Lambda \cap \mathcal{V}_k^s$ with $\mathcal{V}_k^s$ denoting the Voronoi region of Λ_k^s. The channel input is given by

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k] \mod \Lambda_k^s/\beta_k \tag{4.111}$$

$$= \frac{1}{\beta_k}[\mathbf{t}_k + \mathbf{d}_k \beta_k] \mod \Lambda_k^s \tag{4.112}$$

Using lattice decoding with respect to Λ, the relay obtains the sum $\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2$ where

$$\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k \mathbf{d}_k), k = 1, 2 \tag{4.113}$$

with the computation rate

$$R_k^s := \frac{1}{2} \log^+ \left(\frac{P_k \beta_k^2}{\tilde{N}(\beta_1, \beta_2)} \right), k = 1, 2$$

where

$$\tilde{N}(\beta_1, \beta_2) := \frac{P_1 P_2 (a_1\beta_1 - a_2\beta_2)^2 + (a_1\beta_1)^2 P_1 + (a_2\beta_2)^2 P_2}{P_1 + P_2 + 1}.$$

If the computation rate pair (R_1^s, R_2^s) satisfies $R_1^s \geq R_2^s$, we should have Vol $(\mathcal{V}_1^s) \geq$ Vol $(\mathcal{V}_2^s)$ because of the relationship $R_k^s = \frac{1}{n} \log \frac{V(\mathcal{V}_k^s)}{V(\mathcal{V})}$. Hence if Λ_1^s, Λ_2^s are nested lattice, then the nesting relationship is that

$$\Lambda_1^s \subseteq \Lambda_2^s. \tag{4.114}$$

Let Λ_R be the shaping lattice at the relay which satisfies $\Lambda_R \subseteq \Lambda_1^s \subseteq \Lambda_2^s$. The Voronoi region $\mathcal{V}_R$ of Λ_R is denoted by $\mathcal{V}_R$ which has second moment P_R, in order to satisfy the power constraint of the relay. This implies we need to choose β_1 such that it holds

$$P_R \geq \beta_1^2 P_1, \tag{4.115}$$

so that the lattice chain above can be formed. Similarly if it holds that $R_2^s \geq R_1^s$, then we will construct the lattice such that $\Lambda_R \subseteq \Lambda_2^s \subseteq \Lambda_1^s$. Combining these two cases we require β_1, β_2 are chosen such that

$$P_R \geq \max\{\beta_1^2 P_1, \beta_2^2 P_2\} \tag{4.116}$$

After decoding $\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2$, the relay will form

$$\mathbf{x}_R = [\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 + \mathbf{d}_r] \mod \Lambda_R \tag{4.117}$$

where $\mathbf{d}_r$ is a dither uniformly distributed in $\mathcal{V}_R$. The received signal $\mathbf{y}_k$ at Rx k is processed to form

$$\tilde{\mathbf{y}}_k := (\alpha \mathbf{z}_k + (\alpha - 1)\mathbf{x}_R) + \mathbf{x}_R \tag{4.118}$$

$$= \tilde{\mathbf{z}}_k + \tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 - Q_{\Lambda_R}(\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 + \mathbf{d}_r) \tag{4.119}$$

where $\tilde{\mathbf{z}}_k$ is equivalent noise with average power $N = P_R/(1 + P_R)$ for $k = 1, 2$ (assuming both noises at two receivers have unit variance). Rx k quantizes $\tilde{\mathbf{y}}_k$ with respect to Λ to get

$$\mathbf{u} := \tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 - Q_{\Lambda_R}(\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 + \mathbf{d}_r) \tag{4.120}$$

Notice that $\mathbf{u} \in \Lambda$ due to the construction. This decoding will be successful if

$$\frac{Vol(\mathcal{V})^{2/n}}{N} > 2\pi e \tag{4.121}$$

which is equivalent to

$$R_k \leq \frac{1}{2} \log \frac{\beta_k^2 P_k (1 + P_R)}{P_R} \tag{4.122}$$

Now we need to show that with $\mathbf{u}$, each receiver can decode the desired signal. We will only analyze the case when $R_1^s \geq R_2^s$. The other situation follows similarly. The main observation is that if we have $\Lambda_A \subseteq \Lambda_B$, then it holds that

$$[[x] \mod \Lambda_A] \mod \Lambda_B = [x] \mod \Lambda_B \tag{4.123}$$

For the case $\Lambda_R \subseteq \Lambda_1^s \subseteq \Lambda_2^s$, Rx 1 proceeds as

$$[\mathbf{u} - \mathbf{t}_1] \mod \Lambda_1^s = [\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2 - \mathbf{t}_1] \mod \Lambda_1^s \tag{4.124}$$

$$= [\tilde{\mathbf{t}}_2] \mod \Lambda_1^s \tag{4.125}$$

Performing $[\cdot] \mod \Lambda_2^s$ again on the above quantity gives $\mathbf{t}_2$. Rx 2 proceeds as

$$[\mathbf{u} - \tilde{\mathbf{t}}_2] \mod \Lambda_1^s = \mathbf{t}_1 \tag{4.126}$$

Combining all rate constrains we have the following achievable rate pair

$$R_1 \leq \min \left\{ R_1^s, \frac{1}{2} \log \frac{\beta_1^2 P_1 (1 + P_R)}{P_R} \right\} \tag{4.127}$$

$$R_2 \leq \min \left\{ R_2^s, \frac{1}{2} \log \frac{\beta_2^2 P_2 (1 + P_R)}{P_R} \right\} \tag{4.128}$$

with the constraint $P_R \geq \max\{\beta_1^2 P_1, \beta_2^2 P_2\}$. □

Application: Lattice Codes on Interference Channels

5

The usefulness of lattice codes are investigated in this chapter for various models based on the interference channel.[1] The celebrated results on interference alignment [41] show that if a receiver experiences interference from more than one undesired transmitters, the optimal transmission strategy should confine all undesired interference signals in a subspace. With examples in this chapter, we wish to convey the point that using structured codes *is* a form of interference alignment. When the interfering codewords are summed up linearly by the channels, the interference signal (more precisely, the sumset of the interfering codewords) seen by the unintended receiver is much "smaller" when structured codes are used than when the codewords are chosen randomly. Hence the interference is "aligned" due to the linear structure of the codebook. This property gives powerful interference mitigation ability at the signal level.

5.1 The Many-to-One Channel with Cognitive Messages

The concept of *cognitive radio* has been intensively studied and as one of its information-theoretic abstractions, a model of the cognitive radio channel of two users was proposed and analyzed in [42], [43], [44]. In this model, the cognitive user is assumed to know the message of the primary user non-causally before transmissions take place. The capacity region of this channel with additive white Gaussian noise is known for most of the parameter region, see for example [45] for an overview of the results.

Here we extend this cognitive radio channel model to include many cognitive users. We consider the simple many-to-one interference scenario with K cognitive

[1] The material of this chapter has appeared

1. J. Zhu and M. Gastpar, "Lattice Codes for Many-to-One Interference Channels With and Without Cognitive Messages", in *IEEE Transactions on Information Theory*, vol. 61, no. 3, 2015.

2. J. Zhu and M. Gastpar, "Lattice codes for many-to-one cognitive interference networks", in *Proc. 2013 IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jun. 2013.

3. J. Zhu and M. Gastpar, "On lattice codes for Gaussian interference channels", in *Proc. 2015 IEEE International Symposium on Information Theory (ISIT)*, HongKong, China, Jun. 2015.

users illustrated in Figure 5.1. The message W_0 (also called the *cognitive message*) of the primary user is given to all other K users, who could help the transmission of the primary user.

Existing coding schemes for the cognitive interference channel exploit the usefulness of cognitive messages. For the case $K = 1$, i.e., a single cognitive user, the strategy consists in letting the cognitive user spend part of its resources to help the transmission of this message to the primary receiver. At the same time, this also appears as interference at the cognitive receiver, but dirty-paper coding can be used at the cognitive transmitter to cancel (part of) this interference. A new challenge arises when there are many cognitive users. The primary user now benefits from the help of all cognitive users, but at the same time suffers from their collective interference. This inherent tension is more pronounced when the channels from cognitive transmitters to the primary receiver are strong. In the existing coding scheme, the interference from cognitive users is either decoded or treated as noise. As we will show later, direct extensions of this strategy to the many-to-one channel have significant shortcomings, especially when the interference is relatively strong.

Similar systems have been studied in the literature. For the case $K = 2$, the system under consideration is studied in [46]. A similar cognitive interference channel with so-called cumulative message sharing is also studied in [47] where each cognitive user has messages of multiple users. We note that those existing results have not exploited the possibility of using structured codes in cognitive interference networks. The many-to-one channel without cognitive message is studied in [48], where a similar idea of aligning interference based on lattice codes was used. We also point out that the method of compute-and-forward is versatile and beneficial in many network scenarios. For example it has been used in [23], [37] to study the Gaussian two-way relay channel, in [49] to study the K-user symmetric interference channel and in [38] to study the multiple-antenna system.

5.1.1 System model and problem statement

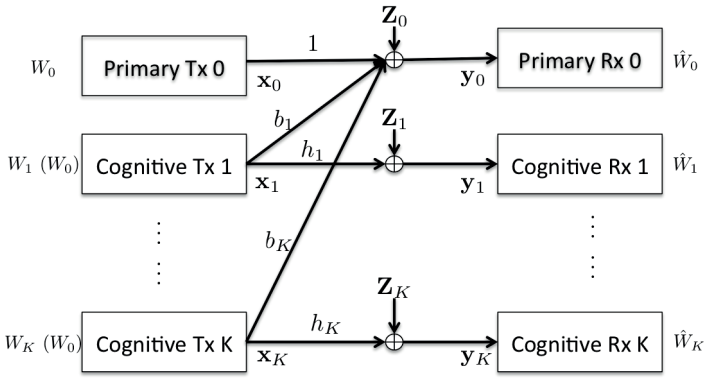


Figure 5.1 – A many-to-one interference channel. The message of the first user W_0 (called cognitive message) may or may not be present at other user's transmitter.

We consider a multi-user channel consisting of $K+1$ transmitter-receiver pairs as

shown in Figure 5.1. The real-valued channel has the following vector representation:

$$\mathbf{y}_0 = \mathbf{x}_0 + \sum_{k=1}^{K} b_k \mathbf{x}_k + \mathbf{z}_0, \tag{5.1}$$

$$\mathbf{y}_k = h_k \mathbf{x}_k + \mathbf{z}_k, \quad k \in [1:K], \tag{5.2}$$

where $\mathbf{x}_k$, $\mathbf{y}_k \in \mathbb{R}^n$ denote the channel input and output of the transmitter-receiver pair k, respectively. The noise $\mathbf{z}_k \in \mathbb{R}^n$ is assumed to be i.i.d. Gaussian with zero mean and unit variance for each entry. Let $b_k \geq 0$ denote the channel gain from Transmitter k to the Receiver 0 and h_k denote the direct channel gain from Transmitter k to its corresponding receiver for $k \in [1:K]$. We assume a unit channel gain for the first user without loss of generality. This system is sometimes referred to as the *many-to-one interference channel* (or *many-to-one channel* for simplicity), since only Receiver 0 experiences interference from other transmitters.

We assume that all users have the same power constraint, i.e., the channel input $\mathbf{x}_k$ is subject to the power constraint

$$\mathbb{E}\{||\mathbf{x}_k||^2\} \leq nP, \quad k \in [1:0]. \tag{5.3}$$

Since channel gains are arbitrary, this assumption is without loss of generality. We also assume that all transmitters and receivers know their own channel coefficients; that is, b_k, h_k are known at Transmitter k, h_k is known at Receiver k, and $b_k, k \geq 1$ are known at Receiver 0.

Now we introduce two variants of this channel according to different message configurations.

Definition 5.1 (Cognitive many-to-one channel)**.** *User* 0 *is called the primary user and User* k *a cognitive user (for* $k \geq 1$*). Each user has a message* W_k *from a set* $\mathcal{W}_k$ *to send to its corresponding receiver. Furthermore, all the cognitive users also have access to the primary user's message* W_0 *(also called cognitive message).*

Definition 5.2 (Non-cognitive many-to-one channel)**.** *Each user* k *has a message* W_k *from a set* $\mathcal{W}_k$ *to send to its corresponding receiver. The messages are not shared among users.*

For the cognitive many-to-one channel, each transmitter has an encoder $\mathcal{E}_k : \mathcal{W}_k \to \mathbb{R}^n$ which maps the message to its channel input as

$$\mathbf{x}_0 = \mathcal{E}_k(W_0) \tag{5.4}$$

$$\mathbf{x}_k = \mathcal{E}_k(W_k, W_0), \quad k \in [1:K]. \tag{5.5}$$

Each receiver has a decoder $\mathcal{D}_k : \mathbb{R}^n \to \mathcal{W}_k$ which estimates message $\hat{W}_k$ from $\mathbf{y}_k$ as

$$\hat{W}_k = \mathcal{D}_k(\mathbf{y}_k), \quad k \in [1:K]. \tag{5.6}$$

The rate of each user is

$$R_k = \frac{1}{n} \log |\mathcal{W}_k| \tag{5.7}$$

under the average error probability requirement

$$\Pr\left(\bigcup_{k=0}^{K}\{\hat{W}_k \neq W_k\}\right) \to \epsilon \tag{5.8}$$

for any $\epsilon > 0$.

For the non-cognitive many-to-one channel, the encoder takes the form

$$\mathbf{x}_k = \mathcal{E}_k(W_k), \quad k \in [0:K] \tag{5.9}$$

and other conditions are the same as in the cognitive channel.

As mentioned earlier, we find it convenient to first treat the general model—the cognitive many-to-one channel where we derive a novel coding scheme which outperforms conventional strategies. We will show that the coding scheme for the cognitive channel can be extended straightforwardly to the non-cognitive channel, which also gives new results for this channel.

5.1.2 Extensions of conventional coding schemes

In this section we revisit existing coding schemes for the two-user cognitive interference channel and extend them to our cognitive many-to-one channel. The extensions are straightforward from the schemes which can be found, for example, in [42], [50] and [45] proposed for the two-user cognitive channel. Throughout the paper, many schemes can be parametrized by letting cognitive transmitters split their power. For each cognitive user, we introduce a power splitting parameter $0 \leq \lambda_k \leq 1$. For convenience, we also define the vector $\underline{\lambda} := \{\lambda_1, \ldots, \lambda_K\}$.

In the first coding scheme, the cognitive users split the power and use part of it to transmit the message of the primary user. Luckily this part of the signal will not cause interference to the cognitive receiver since it can be completely canceled out using dirty-paper coding (DPC). We briefly describe this coding scheme:

- **Primary encoder.** For each possible message W_0, User 0 generates a codeword $\mathbf{x}_0$ with i.i.d. entries according to the Gaussian distribution $\mathcal{N}(0, P)$.

- **Cognitive encoders.** User k generates a sequence $\hat{\mathbf{x}}_k$ with i.i.d. entry according to the Gaussian distribution $\mathcal{N}(0, \bar{\lambda}_k P)$ for any given λ_k and form

$$\mathbf{u}_k = h_k \hat{\mathbf{x}}_k + \gamma h_k \sqrt{\lambda_k}\mathbf{x}_0 \tag{5.10}$$

  with $\gamma = \bar{\lambda}_k h_k^2 P/(1 + \bar{\lambda}_k h_k^2 P)$, $k \geq 1$. The channel input is given by

$$\mathbf{x}_k = \sqrt{\lambda_k}\mathbf{x}_0 + \hat{\mathbf{x}}_k, \quad k \in [1:K]. \tag{5.11}$$

- **Primary decoder.** Decoder 0 decodes $\mathbf{x}_0$ from $\mathbf{y}_0$ using typicality decoding.

- **Cognitive decoders.** Decoder k ($k \geq 1$) decodes $\mathbf{u}_k$ from $\mathbf{y}_k$ using typicality decoding.

This coding scheme gives the following achievable rate region.

Proposition 5.1 (DPC)**.** *For the cognitive many-to-one channel, the above dirty paper coding scheme achieves the rate region:*

$$R_0 \leq \frac{1}{2}\log\left(1+\frac{(\sqrt{P}+\sum_{k\geq 1} b_k\sqrt{\lambda_k P})^2}{\sum_{k\geq 1} b_k^2\bar{\lambda}_k P+1}\right) \tag{5.12}$$

$$R_k \leq \frac{1}{2}\log\left(1+\bar{\lambda}_k h_k^2 P\right), \quad k\in[1:K] \tag{5.13}$$

for any power-splitting parameter $\underline{\lambda}$*.*

It is worth noting that this scheme achieves the capacity in the two-user case ($K=1$) when $|b_1|\leq 1$, see [50, Theorem 3.7] for example.

Another coding scheme which performs well in the two-user case when $|b_1|>1$, is to let the primary user decode the message of the cognitive user as well [45]. We extend this scheme by enabling *simultaneous nonunique decoding* (SND) [6, Ch. 6] at the primary decoder. SND improves the cognitive rates over uniquely decoding the messages $W_k, k\geq 1$ at primary decoder. We briefly describe the coding scheme

- **Primary encoder.** For each possible message W_0, User 0 generates a codewords $\mathbf{x}_0$ with i.i.d. entries according to the distribution $\mathcal{N}(0,P)$.
- **Cognitive encoders.** Given the power splitting parameters λ_k, user k generates $\hat{\mathbf{x}}_k$ with i.i.d. entry according to the distribution $\mathcal{N}(0,\bar{\lambda}_k P)$ for its message $W_k, k\geq 1$. The channel input is given by

$$\mathbf{x}_k = \sqrt{\lambda_k}\mathbf{x}_0 + \hat{\mathbf{x}}_k \tag{5.14}$$

- **Primary decoder.** Decoder 0 simultaneously decodes $\mathbf{x}_0, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K$ from $\mathbf{y}_1$ using typicality decoding. More precisely, let $T^{(n)}(Y_0, X_0, \hat{X}_1\ldots, \hat{X}_K)$ denotes the set of n-length typical sequences (see, for example [6, Ch. 2]) of the joint distribution $(\prod_{i=1}^K P_{\hat{X}_i})P_{X_0}P_{Y_0|X_0\ldots\hat{X}_K}$. The primary decoder decodes its message $\mathbf{x}_0$ such that

$$(\mathbf{x}_0, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K) \in T^{(n)}(Y_0, X_0, \hat{X}_1\ldots, \hat{X}_K) \tag{5.15}$$

  for unique $\mathbf{x}_0$ and *some* $\hat{\mathbf{x}}_k, k\geq 1$.

- **Cognitive decoders.** Decoder k decodes $\hat{\mathbf{x}}_k$ from $\mathbf{y}_k$ for $k\geq 1$.

We have the following achievable rate region for the above coding scheme.

Proposition 5.2 (SND at Rx 0)**.** *For the cognitive many-to-one channel, the above simultaneous nonunique decoding scheme achieves the rate region:*

$$R_0 \leq \frac{1}{2}\log\left(1+\left(\sqrt{P}+\sum_{k\geq 1} b_k\sqrt{\lambda_k P}\right)^2\right), \tag{5.16}$$

$$R_0+\sum_{k\in\mathcal{J}} R_k \leq \frac{1}{2}\log\left(1+\sum_{k\in\mathcal{J}} b_k^2\bar{\lambda}_k P+\left(\sqrt{P}+\sum_{k\geq 1} b_k\sqrt{\lambda_k P}\right)^2\right) \tag{5.17}$$

$$R_k \leq \frac{1}{2}\log\left(1+\frac{\bar{\lambda}_k h_k^2 P_k}{1+\lambda_k h_k^2 P_k}\right) \tag{5.18}$$

for any power-splitting parameter $\underline{\lambda}$ *and every subset* $\mathcal{J}\subseteq[1:K]$*.*

We point out that if instead of using simultaneous nonunique decoding at the primary decoder but require it to decode all messages of the cognitive users $W_k, k \geq 1$, we would have the extra constraints

$$\sum_{k\in\mathcal{J}} R_k \leq \frac{1}{2}\log\left(1+\sum_{k\in\mathcal{J}} b_k^2\bar{\lambda}_k P\right) \tag{5.19}$$

for every subset $\mathcal{J}\subseteq[1:K]$, which may further reduce the achievable rate region.

For the two-user case ($K=1$), the above scheme achieves the capacity when $|b_1|\geq\sqrt{1+P+P^2}+P$, see [45, Theorem V.2] for example.

We can further extend the above coding schemes by combining both dirty paper coding and SND at Rx 0, as it is done in [45, Theorem IV.1]. However this results in a very cumbersome rate expression in the multiple-user system but gives little insight to the problem. We will show in the sequel that our proposed scheme combines the ideas in the above two schemes in a unified framework.

5.1.3 Lattice codes for cognitive many-to-one channels

We first describe how to construct lattice codes for the cognitive many-to-one channels. The lattice codes constructions are similar to the construction given in Chapter 3. Let $\underline{\beta}:=\{\beta_0,\ldots,\beta_K\}$ denotes a set of positive numbers. For each user, we choose a lattice Λ_k which is good for AWGN channel. These $K+1$ fine lattices will form a nested lattice chain [18] according to a certain order which will be determined later. We let Λ_c denote the coarsest lattice among them, i.e., $\Lambda_c\subseteq\Lambda_k$ for all $k\in[0:K]$. As shown in [18, Thm. 2], we can find another $K+1$ simultaneously good nested lattices such that $\Lambda_k^s\subseteq\Lambda_c$ for all $k\in[0:K]$ whose second moments satisfy

$$\sigma_0^2:=\sigma^2(\Lambda_0^s)=\beta_0^2 P \tag{5.20a}$$

$$\sigma_k^2:=\sigma^2(\Lambda_k^s)=(1-\lambda_k)\beta_k^2 P,\quad k\in[1:K] \tag{5.20b}$$

with given power-splitting parameters $\underline{\lambda}$. Introducing the scaling coefficients $\underline{\beta}$ enables us to flexibly balance the rates of different users and utilize the channel state information in a natural way. This point is made clear in the next section when we describe the coding scheme.

The codebook for user k is constructed as

$$\mathcal{C}_k:=\{\mathbf{t}_k\in\mathbb{R}^n:\mathbf{t}_k\in\Lambda_k\cap\mathcal{V}_k^s\},\quad k\in[0:K] \tag{5.21}$$

where $\mathcal{V}_k^s$ denotes the Voronoi region of the *shaping lattice* Λ_k^s used to enforce the power constraints. With this lattice code, the message rate of user k is also given by

$$R_k=\frac{1}{n}\log\frac{\text{Vol}\,(\mathcal{V}_k^s)}{\text{Vol}\,(\mathcal{V}_k)} \tag{5.22}$$

with $\mathcal{V}_k$ denoting the Voronoi region of the fine lattice Λ_k.

Equipped with the nested lattice codes constructed above, we are ready to specify the coding scheme. Each cognitive user splits its power and uses one part to help the primary receiver. Messages $W_k\in\mathcal{W}_k$ of user k are mapped surjectively to lattice points $\mathbf{t}_k\in\mathcal{C}_k$ for all k.

Let $\underline{\gamma} = \{\gamma_1, \ldots, \gamma_K\}$ be K real numbers to be determined later. Given all messages W_k and their corresponding lattice points $\mathbf{t}_k$, transmitters form

$$\mathbf{x}_0 = \left[\frac{\mathbf{t}_0}{\beta_0} + \mathbf{d}_0\right] \bmod \Lambda_0^s/\beta_0 \tag{5.23a}$$

$$\hat{\mathbf{x}}_k = \left[\frac{\mathbf{t}_k}{\beta_k} + \mathbf{d}_k - \frac{\gamma_k \mathbf{x}_0}{\beta_k}\right] \bmod \Lambda_k^s/\beta_k, \quad k \in [1:K] \tag{5.23b}$$

where $\mathbf{d}_k$ (called *dither*) is a random vector independent of $\mathbf{t}_k$ and uniformly distributed in $\mathcal{V}_k^s/\beta_k$. It follows that $\mathbf{x}_0$ is also uniformly distributed in $\mathcal{V}_0^s/\beta_0$ hence has average power $\beta_0^2 P/\beta_0^2 = P$ and is independent from $\mathbf{t}_0$ [17, Lemma 1]. Similarly $\hat{\mathbf{x}}_k$ has average power $\bar{\lambda}_k P$ and is independent from $\mathbf{t}_k$ for all $k \geq 1$.

Although $\mathbf{x}_0$ will act as interference at cognitive receivers, it is possible to cancel its effect at the receivers since it is known to cognitive transmitters. The dirty-paper coding idea in the previous section can also be implemented within the framework of lattice codes, see for example [36]. The parameters $\underline{\gamma}$ are used to cancel the $\mathbf{x}_0$ partially or completely at the cognitive receivers.

The channel input for the primary transmitter is $\mathbf{x}_0$ defined above and the channel input for each cognitive transmitter is

$$\mathbf{x}_k = \sqrt{\lambda_k}\mathbf{x}_0 + \hat{\mathbf{x}}_k, \quad k \in [1:K]. \tag{5.24}$$

Notice that $\mathbb{E}\{||\mathbf{x}_k||^2\}/n = \lambda_k P + \bar{\lambda}_k P = P$ hence power constraints are satisfied for all cognitive users.

We first give an informal description of the coding scheme and then present the main theorem. Let $\mathbf{a} := [a_0, \ldots, a_K] \in \mathbb{Z}^{K+1}$ be a vector of integers. We shall show that the integer sum of the lattice codewords $\sum_{k\geq 0} a_k \mathbf{t}_k$ can be decoded reliably at the primary user for certain rates R_k. As mentioned earlier, we will continue decoding further integer sums with judiciously chosen coefficients and solve for the desired codeword using these sums at the end. An important observation (also made in [8] and [29]) is that the integer sums we have decoded can be used to decode the subsequent integer sums. We also point out the new ingredients in our proposed scheme compared to the existing successive compute-and-forward schemes as in [29] and [8]. Firstly the scaling parameters introduced in (5.20) allow users to adjust there rates according to the channel gains and generally achieve larger rate regions (see [9] for more applications). They will be important for deriving constant gap and capacity results for the non-cognitive channel in Section 5.1.7. Secondly as the cognitive message acts as interference at cognitive receivers, using dirty-paper-coding against the cognitive message in general improves the cognitive rates. But its implementation within successive compute-and-forward framework is not straightforward and requires careful treatment, as shown later in our analysis.

In general, let $L \in [1:K+1]$ be the total number of integer sums[2] the primary user decodes and we represent the L sets of coefficients in the following *coefficient matrix*:

$$\mathbf{A} = \begin{pmatrix} a_0(1) & a_1(1) & a_2(1) & \ldots & a_K(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_0(L) & a_1(L) & a_2(L) & \ldots & a_K(L) \end{pmatrix}, \tag{5.25}$$

[2]There is no need to decode more than $K+1$ sums since there are $K+1$ users in total.

where the ℓ-th row $\mathbf{a}(\ell) := [a_0(\ell), \ldots, a_K(\ell)]$ represents the coefficients for the ℓ-th integer sum $\sum_k a_k(\ell)\mathbf{t}_k$. We will show all L integer sums can be decoded reliably if the rate of user k satisfies

$$R_k \le \min_{\ell} r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) \tag{5.26}$$

with

$$r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) := \max_{\alpha_1,\ldots,\alpha_\ell \in \mathbb{R}} \frac{1}{2} \log^+ \left(\frac{\sigma_k^2}{N_0(\ell)} \right). \tag{5.27}$$

The notation $\mathbf{a}_{\ell|1:\ell-1}$ emphasizes the fact that when the primary decoder decodes the ℓ-th sum with coefficients $\mathbf{a}(\ell)$, all previously decoded sums with coefficients $\mathbf{a}(1), \ldots, \mathbf{a}(\ell-1)$ are used. Recall that σ_k^2 is given in (5.20) and $N_0(\ell)$ is defined as

$$\begin{aligned} N_0(\ell) := \alpha_\ell^2 &+ \sum_{k\ge 1} \left(\alpha_\ell b_k - a_k(\ell)\beta_k - \sum_{j=1}^{\ell-1} \alpha_j a_k(j)\beta_k \right)^2 \bar{\lambda}_k P \\ &+ \left(\alpha_\ell b_0 - a_0(\ell)\beta_0 - \sum_{j=1}^{\ell-1} \alpha_j a_0(j)\beta_0 - g(\ell) \right)^2 P \end{aligned} \tag{5.28}$$

with

$$b_0 := 1 + \sum_{k\ge 1} b_k \sqrt{\lambda_k} \tag{5.29}$$

$$g(\ell) := \sum_{k\ge 1} \left(\sum_{j=1}^{\ell-1} \alpha_j a_k(j) + a_k(\ell) \right) \gamma_k. \tag{5.30}$$

For any matrix $\mathbf{A} \in \mathbb{F}_p^{L\times(K+1)}$, let $\mathbf{A}' \in \mathbb{F}_p^{L\times K}$ denote the matrix $\mathbf{A}$ without the first column. We define a set of matrices as

$$\begin{aligned} \mathcal{A}(L) := \{ \mathbf{A} \in \mathbb{F}_p^{L\times(K+1)} : &\operatorname{rank}(\mathbf{A}) = m, \operatorname{rank}(\mathbf{A}') = m-1 \\ &\text{for some integer } m, 1 \le m \le L \}. \end{aligned} \tag{5.31}$$

We will show that if the coefficients matrix $\mathbf{A}$ of the L integer sums is in this set, the desired codeword $\mathbf{t}_0$ can be reconstructed at the primary decoder. For cognitive receivers, the decoding procedure is much simpler. They will decode the desired codewords directly using lattice decoding.

Now we state the main theorem of this section.

Theorem 5.1. *For any given set of power-splitting parameters $\underline{\lambda}$, positive numbers $\underline{\beta}$, $\underline{\gamma}$ and any coefficient matrix $\mathbf{A} \in \mathcal{A}(L)$ defined in (5.31) with $L \in [1 : K+1]$, define $\mathcal{L}_k := \{\ell \in [1 : L] | a_k(\ell) \neq 0\}$. If $r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) > 0$ for all $\ell \in \mathcal{L}_k$, $k \in [0 : K]$, then the following rate is achievable for the cognitive many-to-one interference channel*

$$R_0 \le \min_{\ell\in\mathcal{L}_0} r_0(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) \tag{5.31a}$$

$$R_k \le \min \left\{ \min_{\ell\in\mathcal{L}_k} r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}), \max_{\nu_k\in\mathbb{R}} \frac{1}{2} \log^+ \frac{\sigma_k^2}{N_k(\gamma_k)} \right\} \quad \text{for } k \ge 1. \tag{5.31b}$$

The expressions $r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma})$ *and* σ_k^2 *are defined in (5.27) and (5.20) respectively, and* $N_k(\gamma_k)$ *is defined as*

$$N_k(\gamma_k) := \nu_k^2 + (\nu_k h_k - \beta_k)^2 \bar{\lambda}_k P + (\nu_k \sqrt{\lambda_k} h_k - \gamma_k)^2 P \tag{5.32}$$

Proof. A proof is given in Appendix 5.5.1. □

Several comments are made on the above theorem. We use $r_k(\mathbf{a}_{\ell|1:\ell-1})$ to denote $r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma})$ for brevity.

- In our coding scheme the primary user may decode more than one integer sums. In general, decoding the ℓ-th sum gives a constraint on R_k:

$$R_k \leq r_k(\mathbf{a}_{\ell|1:\ell-1}). \tag{5.33}$$

  However notice that if $a_k(\ell) = 0$, i.e., the codeword $\mathbf{t}_k$ is not in the ℓ-th sum, then R_k does not have to be constrained by $r_k(\mathbf{a}_{\ell|1:\ell-1})$ since this decoding does not concern Tx k. This explains the minimization of ℓ over the set $\mathcal{L}_k$ in (5.31a) and (5.31b): the set $\mathcal{L}_k$ denotes all sums in which the codeword $\mathbf{t}_k$ participates and R_k is given by the minimum of $r_k(\mathbf{a}_{\ell|1:\ell-1})$ over ℓ in $\mathcal{L}_k$.

- Notice that $r_k(\mathbf{a}_{\ell|1:\ell-1})$ is not necessarily positive and a negative value means that the ℓ-th sum cannot be decoded reliably. The whole decoding procedure will succeed only if all sums can be decoded successfully. Hence in the theorem we require $r_k(\mathbf{a}_{\ell|1:\ell-1}) > 0$ for all $\ell \in \mathcal{L}_k$ to ensure that all sums can be decoded.

- The primary user can choose which integer sums to decode, hence can maximize the rate over the number of integer sums L and the coefficients matrix $\mathbf{A}$ in the set $\mathcal{A}(L)$:

$$R_k \leq \max_{L \in [1:K+1]} \max_{\mathbf{A} \in \mathcal{A}(L)} \min_{\ell \in \mathcal{L}_k} r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}). \tag{5.34}$$

  The optimal $\mathbf{A}$ is the same for all k. To see this, notice that the denominator inside the log of the expression $r_k(\mathbf{a}_{\ell|1:\ell-1})$ in (5.27) is the same for all k and the numerator depends only on k but does not involve the coefficient matrix $\mathbf{A}$, hence the maximizing $\mathbf{A}$ will be the same for all k.

- In the expression of $r_k(\mathbf{a}_{\ell|1:\ell-1})$ in (5.27) we should optimize over ℓ parameters $\alpha_1, \ldots, \alpha_\ell$. The reason for involving these scaling factors is that there are two sources for the effective noise $N_0(\ell)$ at the lattice decoding stage, one is the non-integer channel gain and the other is the additive Gaussian noise in the channel. These scaling factors are used to balance these two effects and find the best trade-off between them, see [8, Section III] for a detailed explanation. The optimal α_ℓ can be given explicitly but the expressions are very complicated hence we will not state it here. We note that the expression $r_k(\mathbf{a}_1)$ with the optimized α_1, $\beta_k = 1$ and $\gamma_k = 0$ is the computation rate of compute-and-forward in [8, Theorem 2].

- For the cognitive users, their rates are constrained both by their direct channel to the corresponding receiver, and by the decoding procedure at the primary user. The two terms in (5.31b) reflect these two constraints. The parameters $\underline{\gamma}$ are used to (partially) cancel the interference $\mathbf{x}_0$ at the cognitive receivers. For example if we set $\gamma_k = \nu_k\sqrt{\lambda_k}h_k$, the cognitive receiver k will not experience any interference caused by $\mathbf{x}_0$. However this affects the computation rate at the primary user in a non-trivial way through $r_k(\mathbf{a}_{\ell|1:\ell-1})$ (cf. Equations (5.27) and (5.28)).

This proposed scheme can be viewed as an extension of the techniques used in the conventional schemes discussed in section 5.1.2. First of all it includes the dirty-paper coding within the lattice codes framework and we can show the following lemma.

Lemma 5.1. *The achievable rates in Proposition 5.1 can be recovered using Theorem 5.2 by decoding one trivial sum with the coefficient* $\mathbf{a}(1) = [1, 0, \ldots, 0]$.

Proof. For given power-splitting parameters $\underline{\lambda}$ we decode only one trivial sum at the primary user by choosing $\mathbf{a}(1)$ such that $a_0(1) = 1$ and $a_k(1) = 0$ for $k \geq 1$, which is the same as decoding $\mathbf{t}_0$. First consider decoding at the primary user. Using the expression (5.27) we have $R_k \leq r_k(\mathbf{a}(1)) = \frac{1}{2}\log(\sigma_k^2/N_0(1))$ with $N_0(1) = \alpha_1^2\left(1 + \sum_{k\geq 1} b_k^2\bar{\lambda}_k P\right) + (\alpha_1 b_0 - \beta_0)^2 P$ and $g(1) = 0$ with this choice of $\mathbf{a}(1)$ for any $\underline{\gamma}$. After optimizing α_1 we have

$$R_0 \leq \frac{1}{2}\log\left(1 + \frac{b_0^2 P}{1 + \sum_{k\geq 1} b_k^2\bar{\lambda}_k P}\right). \tag{5.35}$$

Notice that this decoding does not impose any constraint on R_k for $k \geq 1$.

Now we consider the decoding process at the cognitive users. Choosing $\gamma_k = \nu_k\sqrt{\lambda_k}h_k$ in (5.32) will give $N_k(\gamma_k) = \nu_k^2 + (\nu_k h_k - \beta_k)^2\bar{\lambda}_k P$ and

$$\max_{\nu_k\in\mathbb{R}} \frac{1}{2}\log^+ \frac{\sigma_k^2}{N_k(\gamma_k)} = \frac{1}{2}\log(1 + h_k^2\bar{\lambda}_k P) \tag{5.36}$$

with the optimal $\nu_k^* = \frac{\beta_k h_k \bar{\lambda}_k P}{\lambda_k h_k^2 P + 1}$. This proves the claim. □

The proposed scheme can also be viewed as an extension of simultaneous nonunique decoding (Proposition 5.2). Indeed, as observed in [51], SND can be replaced by either performing the usual joint (unique) decoding to decode all messages or treating interference as noise. The former case corresponds to decoding $K + 1$ integer sums with a full rank coefficient matrix and the latter case corresponds to decoding just one integer sum with the coefficients of cognitive users' messages being zero. Obviously our scheme includes these two cases. As a generalization, the proposed scheme decodes just enough sums of codewords without decoding the individual messages. Unfortunately it is difficult to show analytically that the achievable rates in Proposition 5.2 can be recovered using Theorem 5.1, since it would require the primary receiver to decode several non-trivial sums and the achievable rates are not analytically tractable for general channel gains. However the numerical examples in Section 5.1.5 will show that the proposed scheme generally performs better than the conventional schemes.

5.1.4 On the optimal coefficient matrix A

From Theorem 5.1 and its following comments we see that the main difficulty in evaluating the expression $r_k(\mathbf{a}_{\ell|1:\ell-1})$ in (5.31a) and (5.31b) is the maximization over all possible integer coefficient matrices in the set $\mathcal{A}(L)$. This is an integer programming problem and is analytically intractable for a system with general channel gains $b_1, \ldots, b_K$. In this section we give an explicit formulation of this problem and an example of the choice of the coefficient matrix.

The expression $r_k(\mathbf{a}_{\ell|1:\ell-1})$ in (5.27) is not directly amenable to analysis because finding the optimal solutions for the parameters $\{\alpha_\ell\}$ in (5.28) is prohibitively complex. Now we give an alternative formulation of the problem. We write $N_0(\ell)$ from Eq. (5.28) as in (5.37).

$$N_0(\ell) := \alpha_\ell^2 + \sum_{k\geq 1}\left(\alpha_\ell b_k\sqrt{\bar{\lambda}_k} - a_k(\ell)\beta_k\sqrt{\bar{\lambda}_k} - \sum_{j=1}^{\ell-1}\alpha_j a_k(j)\beta_k\sqrt{\bar{\lambda}_k}\right)^2 P$$
$$+\left(\alpha_\ell b_0 - a_0(\ell)\beta_0 - \sum_{k\geq 1}a_k(\ell)\gamma_k - \sum_{j=1}^{\ell-1}\alpha_j\left(a_0(j)\beta_0 + \sum_{k\geq 1}a_k(j)\gamma_k\right)\right)^2 P. \quad (5.37)$$

It can be further rewritten compactly as

$$N_0(\ell) = \alpha_\ell^2 + \left|\left|\alpha_\ell\mathbf{h} - \tilde{\mathbf{a}}_\ell - \sum_{j=1}^{\ell-1}\alpha_j\tilde{\mathbf{a}}_j\right|\right|^2 P \quad (5.38)$$

where we define $\mathbf{h}, \tilde{\mathbf{a}}_j \in \mathbb{R}^K$ for $j \in [1:\ell]$ as follows:

$$\mathbf{h} = \left[b_0, b_1\sqrt{\bar{\lambda}_1}, \ldots, b_K\sqrt{\bar{\lambda}_k}\right]$$
$$\tilde{\mathbf{a}}_j = \left[a_0(j)\beta_0 + \sum_{k\geq 1}a_k(j)\gamma_k, a_1(j)\beta_1\sqrt{\bar{\lambda}_1}, \ldots, a_K(j)\beta_K\sqrt{\bar{\lambda}_K}\right]. \quad (5.39)$$

We will reformulate the above expression in such a way that the optimal parameters $\{\alpha_j\}$ have simple expressions and the optimization problem on $\mathbf{A}$ can be stated explicitly. This is shown in the following proposition.

Proposition 5.3. *Given $\tilde{\mathbf{a}}_j, j \in [1:\ell-1]$ and $\mathbf{h}$ in (5.39), define*

$$\mathbf{u}_j = \tilde{\mathbf{a}}_j - \sum_{i=1}^{j-1}\tilde{\mathbf{a}}_j|_{\mathbf{u}_i}, \quad j = 1, \ldots \ell-1$$
$$\mathbf{u}_\ell = \mathbf{h} - \sum_{i=1}^{\ell-1}\mathbf{h}|_{\mathbf{u}_i} \quad (5.40)$$

where $\mathbf{x}|_{\mathbf{u}_i} := \frac{\mathbf{x}^T\mathbf{u}_i}{||\mathbf{u}_i||^2}\mathbf{u}_i$ denotes the projection of a vector $\mathbf{x}$ on $\mathbf{u}_i$. The problem of finding the optimal coefficient matrix $\mathbf{A}$ maximizing $r_k(\mathbf{a}_{\ell|1:\ell-1})$ in Theorem 5.1 can be equivalently formulated as the following optimization problem

$$\min_{\substack{L\in[1:K+1]\\ \mathbf{A}\in\mathcal{A}(L)}}\max_{\ell\in\mathcal{L}_k}\left|\left|\mathbf{B}_\ell^{1/2}\mathbf{a}(\ell)\right|\right| \quad (5.41)$$

where $\mathbf{a}(\ell)$ *is the coefficient vector of the* ℓ*-th integer sum. The set* $\mathcal{A}(L)$ *is defined in (5.31) and* $\mathcal{L}_k := \{\ell \in [1:L] | a_k(\ell) \neq 0\}$*. The notation* $\mathbf{B}_\ell^{1/2}$ *denotes a matrix satisfying*[3] $\mathbf{B}_\ell^{1/2}\mathbf{B}_\ell^{1/2} = \mathbf{B}_\ell$*, where* $\mathbf{B}_\ell$ *is given by*

$$\mathbf{B}_\ell := \mathbf{C}\left(\mathbf{I} - \sum_{i=1}^{\ell-1} \frac{\mathbf{u}_j\mathbf{u}_j^T}{||\mathbf{u}_j||^2} - \frac{(\mathbf{u}_\ell\mathbf{u}_\ell^T)P}{1+P||\mathbf{u}_\ell||^2}\right)\mathbf{C}^T. \tag{5.42}$$

The matrix $\mathbf{C}$ *is defined as*

$$\mathbf{C} := \begin{pmatrix} \beta_0 & 0 & 0 & \dots & 0 \\ \gamma_1 & \beta_1\sqrt{\bar{\lambda}_1} & 0 & \dots & 0 \\ \gamma_2 & 0 & \beta_2\sqrt{\bar{\lambda}_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_K & 0 & 0 & \dots & \beta_K\sqrt{\bar{\lambda}_K} \end{pmatrix}. \tag{5.43}$$

Proof: The proof is given in Appendix 5.5.3. ■

The above proposition makes the optimization of $\mathbf{A}$ explicit, although solving this problem is still a computationally expensive task. We should point out that this problem is related to the *shortest vector problem* (SVP) where one is to find the shortest non-zero vector in a lattice. In particular let $\mathbf{B} \in \mathbb{R}^{K\times K}$ be a matrix whose columns constitute one set of basis vectors of the lattice, the SVP can be written as

$$\min_{\mathbf{a}\in\mathbb{Z}^k, \mathbf{a}\neq\mathbf{0}} ||\mathbf{B}\mathbf{a}||. \tag{5.44}$$

Our problem in Proposition 5.3 is more complicated than solving L shortest vector problems. Because the L matrices $\mathbf{B}_\ell^{1/2}$ are related through the optimal integer vectors $\mathbf{a}(\ell)$ in a nontrivial manner and the objective in our problem is to minimize the maximal vector length $\max_\ell \left|\left|\mathbf{B}_\ell^{1/2}\mathbf{a}(\ell)\right|\right|$ of the L lattices. Furthermore the vectors $\mathbf{a}(1), \dots, \mathbf{a}(\ell)$ should lie in the set $\mathcal{A}(L)$ and the number of sums L is also an optimization variable. A low complexity algorithm has been found to solve this instance of SVP for the compute-and-forward problem in simple cases, see [52].

Here we provide an example on the optimal number of sums we need to decode. Consider a many-to-one channel with three cognitive users. We assume $b_1 = 3.5$ and vary b_2 and b_3 in the range $[0, 6]$. We set the direct channel gains $h_k = 1$ and consider four different power constraints. Now the goal is to maximize the sum rate

$$\max_{\substack{L\in[1:4] \\ \mathbf{A}\in\mathcal{A}(L)}} \sum_{k=0}^{4} \min_{\ell\in\mathcal{L}_k} r_k(\mathbf{a}_{1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) \tag{5.45}$$

with respect to $L \in [1:4]$, $\mathbf{A} \in \mathcal{A}(L)$ and $\underline{\beta} \in \mathbb{R}^4$. For simplicity we assume $\lambda_k = \gamma_k = 0$ for $k \geq 1$. Here we search for all possible $\mathbf{A}$ and are interested in the optimal L: the optimal number of sums that need to be decoded.

[3] It is shown that $N_0 = P\mathbf{a}(\ell)^T\mathbf{B}_\ell\mathbf{a}(\ell)$ hence $\mathbf{B}_\ell$ is positive semi-definite because $N_0 \geq 0$. The guarantees the existence of $\mathbf{B}_\ell^{1/2}$.

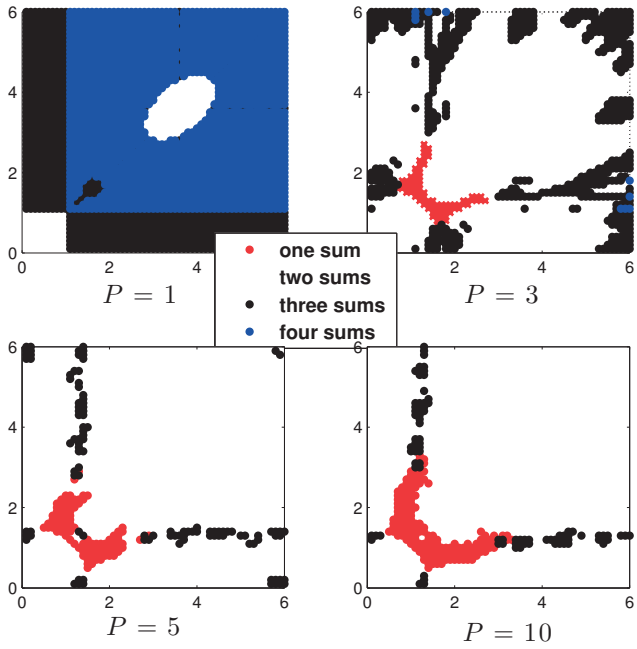


Figure 5.2 – We consider a many-to-one channel with three cognitive users and $b_1 = 3.5$. The horizontal and vertical axes are the range of b_2 and b_3, respectively. The objective is to maximize the sum rate. The red, white, black and blue areas denote the region of different channel gains, in which the number of the best integer sums (the optimal L) is one, two, three and four respectively. Here the patterns are shown for four different power constraints.

The four plots in Figure 5.2 show the *optimal number of integer sums* that the primary user will decode for different power constraints where P equals $1, 3, 5$ or 10. The red area denotes the channel gains where the optimal L equals 1, meaning we need only decode one sum to optimize the sum rate, and so on. Notice that the sign of the channel coefficients b_2, b_3 will not change the optimization problem hence the patterns should be symmetric over both horizontal and vertical axes. When power is small ($P = 1$) we need to decode more than two sums in most channel conditions. The patterns for P equals $3, 5$ or 10 look similar but otherwise rather arbitrary– reflecting the complex nature of the solution to an integer programming problem. One observation from the plots is that for P relatively large, with most channel conditions we only need to decode two sums and we do not decode four sums, which is equivalent to solving for all messages. This confirms the point we made in the previous section: the proposed scheme generalizes the conventional scheme such as Proposition 5.2 to decode just enough information for its purpose, but not more.

5.1.5 Symmetric cognitive many-to-one channels

As we have seen in Section 5.1.4, it is in general difficult to describe the optimal coefficient matrix A. However we can give a partial answer to this question if we focus on one simple class of many-to-one channels. In this section we consider a

symmetric system with $b_k = b$ and $h_k = h$ for all $k \geq 1$ and the case when all cognitive users have the same rate, i.e., $R_k = R$ for $k \geq 1$. By symmetry the parameters λ_k, β_k and γ_k should be the same for all $k \geq 1$. In this symmetric setup, one simple observation can be made regarding the optimal number of integer sums L and the coefficient matrix $\mathbf{A}$.

Lemma 5.2. *For the symmetric many-to-one cognitive interference channel, we need to decode at most two integer sums, $L \leq 2$. Furthermore, the optimal coefficient matrix is one of the following two matrices:*

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \tag{5.46}$$

or

$$\mathbf{A}_2 = \begin{pmatrix} c_0 & c & \dots & c \\ 0 & 1 & \dots & 1 \end{pmatrix} \tag{5.47}$$

for some integer c_0 and nonzero integer c.

Proof. For given $\underline{\lambda}$, $\underline{\beta}$ and $\underline{\gamma}$, to maximize the rate R_k with respect to $\mathbf{A}$ is the same as to minimize the equivalent noise variance $N_0(\ell)$ in (5.28). We write out $N_0(1)$ for decoding the first equation ($\ell = 1$) with $\beta_k = \beta$, $\lambda_k = \lambda$ and $\gamma_k = \gamma$ for all $k \geq 1$:

$$N_0(1) = \alpha_1^2 + \sum_{k \geq 1} (\alpha_1 b - a_k(1)\beta)^2 \, \bar{\lambda} P + (\alpha_1 b_0 - a_0(1)\beta_0 - \gamma \sum_{k \geq 1} a_k(1))^2 P$$

The above expression is symmetric on $a_k(1)$ for all $k \geq 1$ hence the minimum is obtained by letting all $a_k(1)$ be the same. It is easy to see that the same argument holds when we induct on ℓ, i.e., for any $\ell \in [1 : L]$, the minimizing $a_k(\ell)$ is the same for $k \geq 1$. Clearly $\mathbf{A}_1$ and $\mathbf{A}_2$ satisfy this property.

To see why we need at most two integer sums: the case with $\mathbf{A}_1$ when the primary decoder decodes one sum is trivial; now consider when it decodes two sums with the coefficients matrix $\mathbf{A}_2$. First observe that $\mathbf{A}_2$ is in the set $\mathcal{A}(2)$, meaning we can solve for $\mathbf{t}_0$. Furthermore, there is no need to decode a third sum with $a_k(3)$ all equal for $k \geq 1$, because any other sums of this form can be constructed by using the two sums we already have. We also mention that the coefficient matrix

$$\mathbf{A}_3 = \begin{pmatrix} c_0 & c & \dots & c \\ 1 & 0 & \dots & 0 \end{pmatrix} \tag{5.48}$$

is also a valid choice and will give the same result as $\mathbf{A}_2$. □

Now we give some numerical results comparing the proposed scheme with the conventional schemes proposed in Section 5.1.2 for the symmetric cognitive many-to-one channels.

Figure 5.3 shows the achievable rate region for a symmetric cognitive many-to-one channel. The dashed and dot-dash lines are achievable regions with DPC in Proposition 5.1 and SND at Rx 0 in Proposition 5.2, respectively. The solid line depicts the rate region using the proposed scheme in Theorem 5.1. Notice the achievable rates based on the simple conventional schemes in Proposition 5.1 and 5.1 are not much better than the trivial time sharing scheme in the multi-user scenario,

due to their inherent inefficiencies on interference suppression. On the other hand, the proposed scheme based on structured codes performs a kind of interference alignment in the signal level, which gives better interference mitigation ability at the primary receiver. The effect is emphasized more when we study the non-cognitive system in Section 5.1.6. The outer bound in Figure 5.3 is obtained by considering the system as a two-user multiple-antenna broadcast channel whose capacity region is known. A brief description to this outer bound is given in Appendix 5.5.4.

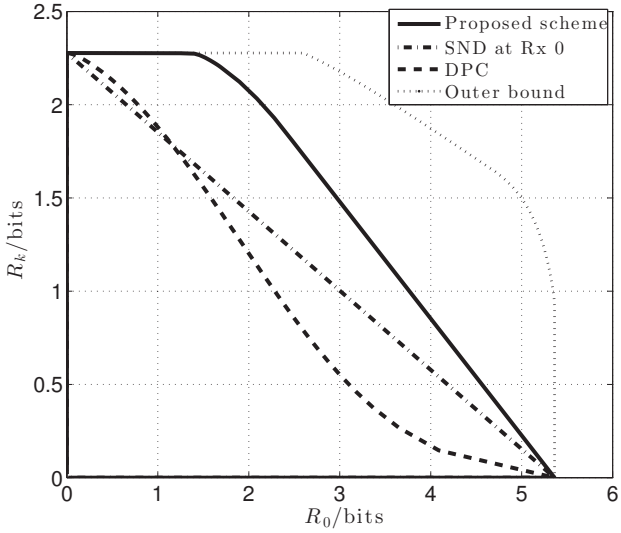


Figure 5.3 – Achievable rate region for a many-to-one symmetric cognitive many-to-one channel with power $P = 10$, channel gain $b_k = 4$, $h_k = 1.5$ for $k \geq 1$ and $K = 3$ cognitive users.The plot compares the different achievable rates for the cognitive many-to-one channel. The horizontal and vertical axis represents the primary rate R_0 and cognitive rate $R_k, k \geq 1$, respectively.

It is also instructive to study the system performance as a function of the channel gain b. We consider a symmetric channel with h fixed and varying value of b. For different values of b, we maximize the symmetric rate $R_{sym} := \min\{R_0, R\}$ where $R = R_k$ for $k \geq 1$ by choosing optimal $\mathbf{A}, \underline{\lambda}$ and $\underline{\beta}$, i.e.,

$$\max_{\substack{\mathbf{A}\in\mathcal{A}(2)\\ \underline{\lambda},\underline{\beta}}} \min\left\{\min_{\ell\in\mathcal{L}_0} r_0(\mathbf{a}_{\ell|1:\ell-1}), \min_{\ell\in\mathcal{L}_k} r_k(\mathbf{a}_{\ell|1:\ell-1}), \max_{\nu_k\in\mathbb{R}} \frac{1}{2}\log^+\frac{\sigma_k^2}{N_k(\gamma_k)}\right\} \quad (5.49)$$

where the first term is the rate of the primary user and the minimum of the second and the third term is the rate of cognitive users. Notice $\lambda_k, \beta_k, r_k(\mathbf{a}_{\ell|1:\ell-1})$ are the same for all $k \geq 1$ in this symmetric setup. Figure 5.4 shows the maximum symmetric rate of different schemes with increasing b.

5.1.6 Non-cognitive many-to-one channels

As an interesting special case of the cognitive many-to-one channel, in this section we will study the non-cognitive many-to-one channels where user $1, \ldots, K$ do not have access to the message W_0 of User 0. The many-to-one interference channel has also been studied, for example, in [48], where several constant-gap results are

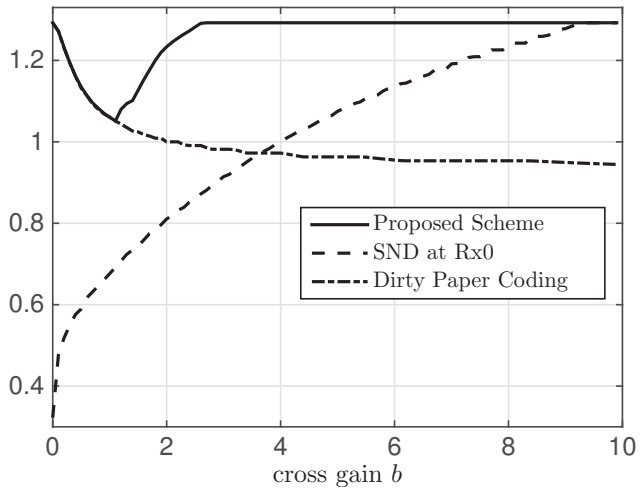


Figure 5.4 – The maximum symmetric rates R_{sym} of different schemes for a many-to-one cognitive interference network with power $P = 5$ and $K = 3$ cognitive users where $R_k = R$ for $k \geq 1$. We set $h = 1$ and vary the cross channel gain b in the interval $[0 : 10]$. Notice the maximum symmetric rate is upper bounded by $\frac{1}{2}\log(1 + h^2P)$. We see the proposed scheme performs better than the other two schemes in general. When the interference becomes larger, the proposed scheme quickly attains the maximum symmetric rate. The joint decoding method approaches the maximum symmetric rate much slower, since it requires the cross channel gain to be sufficiently large such that the primary decoder can (nonuniquely) decode all the messages of the cognitive users. The dirty paper coding approach cannot attain the maximum symmetric rate since the primary decoder treats interference as noise.

obtained. Using the coding scheme introduced here, we are able to give some refined result to this channel in some special cases.

It is straightforward to extend the coding scheme of the cognitive channel to the non-cognitive channel by letting users $1, \dots K$ not split the power for the message W_0 but to transmit their own messages only. The achievable rates are the same as in Theorem 5.1 by setting all power splitting parameters λ_k to be zero and γ_k to be zero because $\mathbf{x}_0$ will not be interference to cognitive users.. Although it is a straightforward exercise to write out the achievable rates, we still state the result formally here.

Theorem 5.2. *For any given positive numbers $\underline{\beta}$ and coefficient matrix $\mathbf{A} \in \mathcal{A}(L)$ in (5.31) with $L \in [1 : K+1]$, define $\mathcal{L}_k := \{\ell \in [1 : L] | a_k(\ell) \neq 0\}$. If $r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) > 0$ for all $\ell \in \mathcal{L}_k$, $k \in [0 : K]$, then the following rate is achievable for the many-to-one interference channel*

$$R_0 \leq \min_{\ell \in \mathcal{L}_0} \tilde{r}_0(\mathbf{a}_{\ell|1:\ell-1}, \underline{\beta}) \tag{5.49a}$$

$$R_k \leq \min\left\{\frac{1}{2}\log\left(1 + h_k^2 P\right), \min_{\ell \in \mathcal{L}_k} \tilde{r}_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\beta})\right\} \text{for } k \in [1 : K], \tag{5.49b}$$

with

$$\tilde{r}_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\beta}) := \max_{\alpha_1, \dots, \alpha_\ell \in \mathbb{R}} \frac{1}{2}\log^+\left(\frac{\beta_k^2 P}{\tilde{N}_0(\ell)}\right) \tag{5.50}$$

where $\tilde{N}_0(\ell)$ *is defined as*

$$\tilde{N}_0(\ell) := \alpha_\ell^2 + \sum_{k\geq 1}\left(\alpha_\ell b_k - a_k(\ell)\beta_k - \sum_{j=1}^{\ell-1}\alpha_j a_k(j)\beta_k\right)^2 P + \left(\alpha_\ell - a_0(\ell)\beta_0 - \sum_{j=1}^{\ell-1}\alpha_j a_0(j)\beta_0\right)^2 P. \tag{5.51}$$

Proof. The proof of this result is almost the same as the proof of Theorem 5.1 in Section 5.5.1. The only change in this proof is that the user $1,\ldots,K$ do not split the power to transmit for the primary user and all γ_k are set to be zero since $\mathbf{x}_0$ will not act as interference to cognitive receivers. We will make slight adjustment to the codes constructions. Given positive numbers $\underline{\beta}$ and a simultaneously good fine lattice Λ, we choose $K+1$ simultaneously good lattices such that $\Lambda_k^s \subseteq \Lambda_k$ with second moments $\sigma^2(\Lambda_k^s) = \beta_k^2 P$ for all $k \in [0:K]$.

Each user forms the transmitted signal as

$$\mathbf{x}_k = \left[\frac{\mathbf{t}_k}{\beta_k} + \mathbf{d}_k\right] \bmod \Lambda_k^s/\beta_k, \quad k \in [0:K] \tag{5.52}$$

The analysis of the decoding procedure at all receivers is the same as in Section 5.5.1. User 0 decodes integer sums to recover $\mathbf{t}_0$ and other users decode their message $\mathbf{t}_k$ directly from the channel output using lattice decoding. In fact, the expression $\tilde{r}_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\beta})$ in (5.50) is the same as $r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma})$ in (5.27) by letting $\lambda_k = \gamma_k = 0$ in the later expression. Furthermore we have

$$\max_{\nu_k \in \mathbb{R}} \frac{1}{2}\log\frac{\sigma_k^2}{N_k(\gamma_k = 0)} = \frac{1}{2}\log(1 + h_k^2 P) \tag{5.53}$$

for any choice of $\beta_k, k \geq 1$. □

For a simple symmetric example, we compare the achievable rate region of the cognitive many-to-one channel (Theorem 5.1) with the achievable rate region of the non-cognitive many-to-one channel (Theorem 5.2) in Figure 5.5. The parameters are the same for both channel. This shows the usefulness of the cognitive messages in the system.

5.1.7 Capacity results for non-cognitive symmetric channels

Now we consider a symmetric non-cognitive many-to-one channel where $b_k = b$ and $h_k = h$ for $k \geq 1$. In [48], an approximate capacity result is established within a gap of $(3K+3)(1+\log(K+1))$ bits per user for *any* channel gain. In this section we will give refined results for the symmetric many-to-one channel. The reason we restrict ourselves to the symmetric case is that, for general channel gains the optimization problem involving the coefficient matrix $\mathbf{A}$ is analytically intractable as discussed in Section 5.1.4, hence it is also difficult to give explicit expressions for achievable rates. But for the symmetric many-to-one channel we are able to give a constant gap result as well as a capacity result when the interference is strong. First notice that the optimal form of the coefficient matrix for the cognitive symmetric channel given in Lemma 5.2 also applies in this non-cognitive symmetric setting.

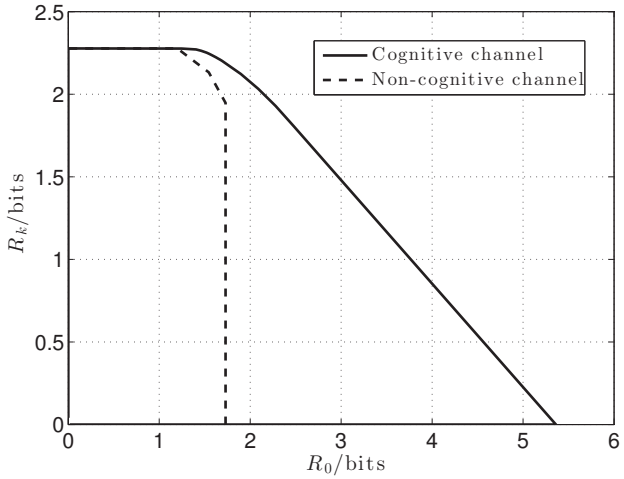


Figure 5.5 – A many-to-one symmetric interference channel with power $P = 10$, channel gain $b_k = 4$, $h_k = 1.5$ for $k \geq 1$ and $K = 3$ cognitive users. This plot compares the different achievable rate regions for the cognitive and non-cognitive channel. The horizontal and vertical axis represents the primary rate R_0 and cognitive rate $R_k, k \geq 1$, respectively. The rate region for the cognitive channel given by Theorem 5.1 is plotted in solid line. The dashed line gives the achievable rate region in Theorem 5.2 for the non-cognitive many-to-one channel.

Theorem 5.3. *Consider a symmetric (non-cognitive) many-to-one interference channel with $K+1$ users. If $|b| \geq |h|\left\lceil\sqrt{P}\right\rceil$, then each user is less than 0.5 bit from the capacity for any number of users. Furthermore, if $|b| \geq \sqrt{\frac{(1+P)(1+h^2P)}{P}}$, each user can achieve the capacity, i.e., $R_0 = \frac{1}{2}\log(1+P)$ and $R_k = \frac{1}{2}\log(1+h^2P)$ for all $k \geq 1$.*

Proof. A proof is given in Appendix 5.5.5. □

Comparing to the constant gap result in [48], our result only concerns a special class of many-to-one channel, but gives a gap which does not depend on the number of users K. We also point out that in [53], a K-user symmetric interference channel is studied where it was shown that if the cross channel gain b satisfies $|b| \geq \sqrt{\frac{(1+P)^2}{P}}$, then every user achieves the capacity $\frac{1}{2}\log(1+P)$. This result is very similar to our result obtained here and is actually obtained using the same coding technique.

5.2 The Gaussian Interference Channel with Strong Interference

Consider a two-user Gaussian IC

$$\mathbf{y}_1 = \mathbf{x}_1 + g_1\mathbf{x}_2 + \mathbf{z}_1 \tag{5.54a}$$

$$\mathbf{y}_2 = \mathbf{x}_2 + g_2\mathbf{x}_1 + \mathbf{z}_2 \tag{5.54b}$$

with $\mathbf{x}_k, \mathbf{y}_k \in \mathbb{R}^n$ denoting the channel input at transmitter (Tx) k and the channel output at receiver (Rx) k, $k = 1, 2$. The noises $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^n$ are assumed to be

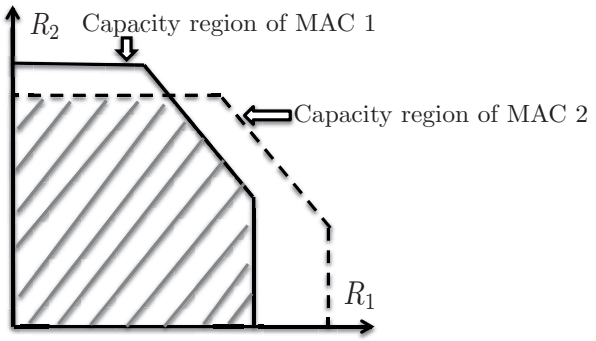


Figure 5.6 – The shaded region is the capacity region of a two-user IC under strong interference as the intersection of capacity regions of two Gaussian MAC: solid line for MAC 1 and dashed line for MAC 2. This example shows the case when $S_2+I_2 \geq S_1+I_1$, i.e., MAC 2 has a higher sum rate capacity.

Gaussian with unit variance per entry. Power constraints are imposed on the channel input as $||\mathbf{x}_k||^2 \leq nP_k$ for $k = 1, 2$. Transmitter k has a message W_k from the set $\{1, \ldots, 2^{nR_k}\}$ to send to the corresponding Rx k and it is required that both receivers can decode their intended message reliably. We denote the received signal-to-noise ratio as $S_k := P_k, k = 1, 2$ and the received interference-to-noise ratio as $I_1 := g_1^2 P_2$, $I_2 := g_2^2 P_1$.

The capacity region of this channel is known under the *strong interference* condition, i.e. when it holds that

$$I_1 \geq S_2, I_2 \geq S_1. \tag{5.55}$$

In this case the capacity region of the two-user Gaussian with strong interference is given by [54]

$$R_1 \leq \mathsf{C}(S_1), \quad R_2 \leq \mathsf{C}(S_2) \tag{5.56a}$$

$$R_1 + R_2 \leq \mathsf{C}_{\min} := \min\{\mathsf{C}(S_1 + I_1), \mathsf{C}(S_2 + I_2)\} \tag{5.56b}$$

where $\mathsf{C}(x) := \frac{1}{2}\log(1+x)$. An illustration is shown in Figure 5.6. The capacity region in this case is the intersection of capacity regions of two Gaussian multiple access channels (MAC): one composed of two Txs and Rx 1 as the receiver (call it MAC 1); and one composed of two Txs and Rx 2 as the receiver (call it MAC 2).

It is well known that for such a Gaussian IC, the capacity region can be achieved by letting both receivers perform joint decoding (also called simultaneous decoding) to decode both messages. For a Gaussian MAC, it is also well known that in addition to joint decoding, two other decoding schemes, successive cancellation decoding (SCD) with time-sharing and rate-splitting scheme [26], can achieve the capacity region with a single-user decoder. A single-user decoder is sometimes preferred in practice for various reasons including complexity issues. Since the capacity region of Gaussian IC with strong interference is the intersection of capacity regions of two MACs, we could ask if the above two low complexity methods also achieve the capacity region of a Gaussian IC. However, as it is shown in [55], the standard rate-splitting scheme is not able to achieve the whole capacity region, regardless the number of layers and the code distribution of each layer. It is also easy to

see that SCD with time-sharing fails to achieve the capacity region. To overcome this difficulty, a sliding-window superposition coding scheme is proposed in [55] which achieves the joint decoding inner bound for general interference channels. Combined with time-sharing, it achieves the capacity region of Gaussian IC with strong interference.

Here we show that for the Gaussian IC with strong interference, CFMA always achieves the corner points of the capacity region, and for some parameters, achieves the whole capacity region.

5.2.1 CFMA for the two-user Gaussian IC

In this section we show how to apply CFMA to the Gaussian interference channel. Rx k decodes two sums of the codewords in the form:

$$\mathbf{u}_{k1} = a_{k1}\mathbf{t}_1 + a_{k2}\mathbf{t}_2, \quad \mathbf{u}_{k2} = b_{k1}\mathbf{t}_1 + b_{k2}\mathbf{t}_2$$

with the *coefficient matrix* $\mathbf{A}_k = \begin{pmatrix} a_{k1} & a_{k2} \\ b_{k1} & b_{k2} \end{pmatrix}$ satisfying the requirement that $\mathbf{A}_k$ is a full rank integer matrix. Let $\hat{\mathbf{u}}_{kj}, j = 1, 2$ denote the two decoded integer sums at Rx k and define the error probability of decoding as

$$P_{e,k}^{(n)} := \mathbb{P}(\{\hat{\mathbf{u}}_{k1} \neq \mathbf{u}_{k1}\} \cup \{\hat{\mathbf{u}}_{k2} \neq \mathbf{u}_{k2}\}), k = 1, 2 \tag{5.57}$$

where n is the length of the codewords. Formally we have the following definition.

Definition 5.3 (Achievability with CFMA)**.** *For a two-user Gaussian IC, we say a rate pair (R_1, R_2) is achievable with compute-and-forward multiple access (CFMA), if the rate of codebook $\mathcal{C}_k$ is R_k and the error probability $P_{e,k}^{(n)}, k = 1, 2$ in (5.57) can be made arbitrarily small for large enough n.*

Notice that we do not include time-sharing in the above definition. This means if we say a certain rate pair is achievable using CFMA, only a single-user decoder is used at each receiver without time-sharing.

We focus on the Gaussian IC with strong but not *very strong* interference, i.e., in addition to (5.55), the sum rate constraint in (5.56b) is active. In this case the capacity region in (5.56) is a pentagon and we can identify two *corner points* (R_1, R_2) as

$$(\mathsf{C}_{\min} - \mathsf{C}(S_2), \mathsf{C}(S_2)) \text{ and } (\mathsf{C}(S_1), \mathsf{C}_{\min} - \mathsf{C}(S_1)). \tag{5.58}$$

Theorem 5.4 (CFMA for the Gaussian IC)**.** *Consider the two-user Gaussian IC in (5.54) with strong but not very strong interference. If it holds that*

$$\min\left\{\sqrt{\frac{S_1 I_1}{1+S_1+I_1}}, \sqrt{\frac{S_2 I_2}{1+S_2+I_2}}\right\} \geq 1 \tag{5.59}$$

the corner points (5.58) of the capacity region are achievable using CFMA. Furthermore if it holds that

$$I_1 \geq S_2(1+S_1) \quad \textit{or} \quad I_2 \geq S_1(1+S_2), \tag{5.60}$$

the whole capacity region is achievable with CFMA.

Proof. The codes construction and encoding/decoding procedure are exactly the same as in two-user Gaussian MAC studied in Theorem 4.2. We give the proof for the case when it holds that $S_2 + I_2 \geq S_1 + I_1$, i.e., MAC 2 has a higher sum capacity than MAC 1. The other case can be proved similarly. In this case the capacity of this Gaussian IC is depicted in Figure 5.7 as the intersection of capacity regions of two Gaussian MACs. The two corner points of the IC capacity region are marked as A and B, and the upper corner point of the MAC 2 capacity region is marked as C. We use A_1 and A_2 to denote point A's coordinates on horizontal and vertical axes, respectively, and so on.

We first consider the subcase 1 on the left side of Figure 5.7 , where $C_1 \leq B_1$. This means $(1 + S_2 + I_2)/(1 + S_2) < 1 + S_1$, or equivalently $I_2 < S_1(1 + S_2)$.

In order to achieve the corner point A, Rx 1 decodes two sums with coefficient matrix $\mathbf{A}_1 = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ or $\mathbf{A}_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. According to Theorem 4.2, depending on the position of A, at least one of the two coefficient matrices $\mathbf{A}_1$ allows Rx 1 to decode both messages at the rate $(R_1, R_2) = (A_1, A_2)$, if the condition (5.59) holds. Rx 2 decodes two sums with the coefficient matrix $\mathbf{A}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ namely the usual successive cancellation decoding. This allows Rx 2 to recover both messages if the rates satisfy $R_1 \leq C_1$ and $R_2 \leq C_2$. We point out that in order to let Rx 1 operate at point A, the scaling parameter β_1, β_2 should satisfy $\beta_1/\beta_2 = c$ for some value c *depending on* A. However, the usual SCD at Rx 2 works for *any* values of β_1, β_2. Furthermore notice that $A_1 \leq C_1$ and $A_2 \leq C_2$ due to our assumption that MAC 2 has a higher sum capacity, this guarantees that both receivers can decode both messages reliably for the rate pair A.

To achieve the corner point B, we let Rx 2 decode two sums with coefficient matrix $\mathbf{A}_2 = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ or $\mathbf{A}_2 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Due to Theorem 4.2, at least one of the two choices on coefficient matrix $\mathbf{A}$ allows Rx 2 to decode both messages at the rate $R_1 \leq B_1'$ and $R_2 \leq B_2'$, if the condition (5.59) holds and parameters β_1, β_2 are chosen accordingly. Here B' is the projection of point B on the dominant face of the MAC 2 capacity region along the vertical axis. Now Rx 1 performs the SCD to decode $\mathbf{t}_2$ and $\mathbf{t}_1$ at the rate $R_1 = B_1, R_2 = B_2$. Since $B_1' = B_1$ and $B_2' \geq B_2$, both decoders can decode both messages reliably, hence achieve the corner point B.

Now we consider the subcase 2 on the right side of Figure 5.7 when $I_2 \geq S_1(1 + S_2)$. In this case we have $C_1 \geq B_1$. The same as achieving point A, we let Rx 1 to decode two sums with coefficient matrix $\mathbf{A}_1 = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ or $\mathbf{A}_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Due to Theorem 4.2, all points on the segment AB are achievable if β_1, β_2 are chosen accordingly. Rx 2 uses SCD (equivalently $\mathbf{A}_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$) which allows it to decode both messages if the rate pair (R_1, R_2) satisfies $R_1 \leq C_1$ and $R_2 \leq C_2$. However, this is true for all rate pairs on the segment AB in this case. This means all points on the dominant face of the capacity region can be achieved using CFMA in this case.

For the case when MAC 1 has a higher sum rate, the results can be proved in the same way and we summarize the decoding operation in Table 5.1 and 5.2. Table 5.1 shows how receivers should decode to obtain the corner points. Table 5.2 shows in the case when either one of the receivers experience very strong interference, the decoding operation at receivers for achieving the whole capacity region. □

It is well known that if it holds that

$$I_1 \geq S_2(1 + S_1) \quad \textit{and} \quad I_2 \geq S_1(1 + S_2), \tag{5.61}$$

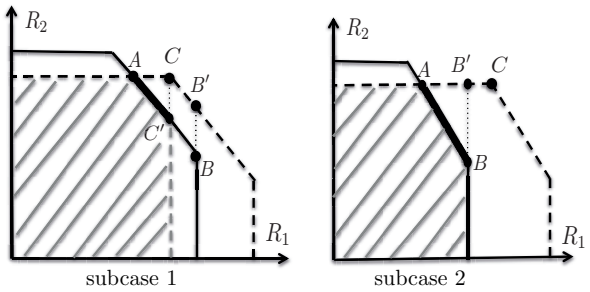


Figure 5.7 – The figure depicts the capacity region of a Gaussian IC with strong but not very strong interference as the intersection of two MAC capacity regions. The shaded regions in two subcases are achievable using CFMA (without time-sharing). In subcase 1, the line segment AC' and the point B are achievable using CFMA. Time-sharing can be used to achieve the whole capacity region. In subcase 2, CFMA can achieve the line segment AB, hence the whole capacity region without time-sharing.

the sum rate constraint in (5.56b) is inactive and the channel is said to be in *very strong* interference regime. The optimal point in its capacity region $R_k = \mathsf{C}(S_k), k = 1, 2$ can be achieved by using SCD at both receivers to first decode the other user's message. Our results show that under a weaker condition (5.60), where interference from only one transmitter is very strong, the proposed scheme can already achieve the whole capacity region using a single-user decoder without time-sharing.

We also point out that even when the one-sided very strong interference condition in (5.60) is not fulfilled, we can still achieve points other than the corner points on the capacity region with CFMA. As marked in Figure 5.7 subcase 1, using the same argument we can show that all points on segment AC' are achievable using CFMA, where C' is the projection of C along the vertical axis on the dominant face of MAC 1.

Another special case where CFMA achieves the whole capacity region without time-sharing is when $g_1 = g_2 = 1$ (equivalently $I_1 = S_2$ and $I_2 = S_1$), which is not covered in the above theorem. In this case both decoders choose the same coefficient matrix $\mathbf{A}_k = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ or $\mathbf{A}_k = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), k = 1, 2$.

5.3 The Gaussian Z-Interference Channel

As a special case of the Gaussian IC, the so-called Gaussian Z-interference channel has also been studied in, for example, [56] and [57]. In this model, the channel gain g_2 in (5.54) is set to be zero (hence $I_2 = 0$) and other setup is the same for the Gaussian IC channel. In the case $I_1 \geq S_2$ (notice it does not satisfy the strong interference condition in (5.55)), the capacity region of this channel is known to be

$$R_1 \leq \mathsf{C}(S_1), \quad R_2 \leq \mathsf{C}(S_2), \quad R_1 + R_2 \leq \mathsf{C}(S_1 + I_1) \tag{5.62}$$

We use a similar argument to show that this capacity region is achievable with CFMA.

Table 5.1 – strong but not very strong interference: Choice of coefficients for achieving corner points with CFMA

Corner point (R_1, R_2)	$\mathbf{A}_1$ at Rx 1	$\mathbf{A}_2$ at Rx 2
$\mathsf{C}_{\min} - \mathsf{C}(S_2), \mathsf{C}(S_2)$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ or $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$
$\mathsf{C}(S_1), \mathsf{C}_{\min} - \mathsf{C}(S_1)$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ or $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$

Table 5.2 – One-sided very strong interference: choice of coefficients for achieving the whole capacity region with CFMA

Condition	$\mathbf{A}_1$ at Rx 1	$\mathbf{A}_2$ at Rx 2
$I_2 \geq S_1(1+S_2)$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ or $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$
$I_1 \geq S_2(1+S_1)$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ or $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$

Theorem 5.5 (CFMA for Gaussian Z-interference channels)**.** *Consider the Gaussian Z-interference channel with strong interference, i.e., $I_1 \geq S_2$. If it holds that*

$$\sqrt{\frac{S_1 I_1}{1+S_1+I_1}} \geq 1, \tag{5.63}$$

the whole capacity region is achievable using CFMA (without time-sharing).

Proof. The capacity region (5.62) of a Gaussian Z-interference channel with strong interference is given in Figure 5.8. The solid line depicts the capacity region of MAC 1. The dominant face is the line segment AB where A denotes the rate pair $(R_1 = \mathsf{C}(S_1+I_1)-\mathsf{C}(S_2), R_2 = \mathsf{C}(S_2))$ and B denotes the rate pair $(R_1 = \mathsf{C}(S_1), R_2 = \mathsf{C}(S_1+I_1)-\mathsf{C}(S_1))$. To achieve any point on this line, Rx 1 decodes two sums with coefficient matrix $\mathbf{A}_1 = \left(\begin{smallmatrix}1 & 1\\ 1 & 0\end{smallmatrix}\right)$ or $\mathbf{A}_1 = \left(\begin{smallmatrix}0 & 1\\ 1 & 1\end{smallmatrix}\right)$. According to Theorem 4.2, given any rate pair on the line AB, at least one of the two coefficient matrices $\mathbf{A}_1$ allows Rx 1 to decode both messages if the condition (5.63) holds and parameters β_1, β_2 are chosen accordingly. Rx 2 performs usual lattice decoding as in a point-to-point channel and the decoding will be successful if $R_2 \leq \mathsf{C}(S_2)$, which is the case for any rate pair on the line AB. □

Different from the Gaussian IC, rate pairs on the dominant face of the Z-interference channel capacity region can be achieved using the rate-splitting scheme [26]. It can be seen that if Tx 1 splits its message into two parts (with an appropriate power allocation) and lets Rx 1 decode all three messages (two messages from Tx 1 and the message from Tx 2) in an appropriate order, any rate pair on the dominant face can be achieved.

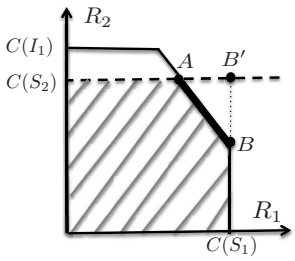


Figure 5.8 – The shaded region is the capacity region of a Gaussian Z-interference channel, which is achievable using CFMA without time-sharing.

5.4 The Two-user Gaussian IC with States

As the last example in the family of Gaussian interference channels, we consider a two-user Gaussian IC of the form

$$\mathbf{y}_1 = \mathbf{x}_1 + g_1\mathbf{x}_2 + c_1\mathbf{s} + \mathbf{z}_1 \tag{5.64a}$$

$$\mathbf{y}_2 = \mathbf{x}_2 + g_2\mathbf{x}_1 + c_2\mathbf{s} + \mathbf{z}_2 \tag{5.64b}$$

where $\mathbf{s} \in \mathbb{R}^n$ is a state sequence non-causally known to two transmitters but not to receivers. Each entry of $\mathbf{s}$ is an i.i.d. random variable with a given distribution (not necessarily Gaussian) and variance $E\,||\mathbf{s}_i||^2 = Q$ for $i = 1, \ldots, n$. The other setup is the same as for the normal Gaussian IC in (5.54). To make the model slightly more general, we use two real numbers c_1, c_2 to represent the fact that two channels may suffer from differently scaled versions of the same interference $\mathbf{s}$. This model has been studied in, for example, [58] [59] where various coding schemes are given.

In our scheme, the channel input for this channel is given by

$$\mathbf{x}_k = [\mathbf{t}_k/\beta_k + \mathbf{d}_k - \gamma_k\mathbf{s}/\beta_k] \mod \Lambda_k^s/\beta_k, k = 1, 2$$

for some real numbers γ_k to be chosen later. In addition to the Gaussian IC where β_k are used to control the rates of two users, the extra parameters γ_k are used to eliminate (partially or completely) the interference $\mathbf{s}$. For given β_k, γ_k and $\mathbf{A}_k$, the optimal α_{k1}, α_{k2} and λ_k which maximize the

We can show that Rx $k, k = 1, 2$ can form

$$\tilde{\mathbf{y}}_{k1} := \tilde{\mathbf{z}}_{k1} + \sum_{i=1}^{2} a_{ki}\tilde{\mathbf{t}}_i, \quad \tilde{\mathbf{y}}_{k2} := \tilde{\mathbf{z}}_{k2} + \sum_{i=1}^{2} b_{ki}\tilde{\mathbf{t}}_i$$

with $\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \gamma_k\mathbf{s})$. The variance N_{k1} per dimension for noise $\tilde{\mathbf{z}}_{k1}$, and variance N_{k2} for noise $\tilde{\mathbf{z}}_{k2}$ at Rx k are given as follows

$$N_{11} = (\alpha_{11} - a_{11}\beta_1)^2P_1 + (\alpha_{11}g_1 - a_{12}\beta_2)^2P_2 + \alpha_{11}^2 + (\alpha_{11}c_1 - \sum_{i=1}^{2} a_{1i}\gamma_i)^2Q \tag{5.65a}$$

$$N_{12} = (\alpha_{12} - \lambda_1 a_{11}\beta_1 - b_{11}\beta_1)^2P_1 + (\alpha_{12}g_1 - \lambda_1 a_{12}\beta_2 - b_{12}\beta_2)^2P_2 + \alpha_{12}^2 + (\alpha_{12}c_1 - \sum_{i=1}^{2}(\lambda_1 a_{1i} + b_{1i})\gamma_i)^2Q \tag{5.65b}$$

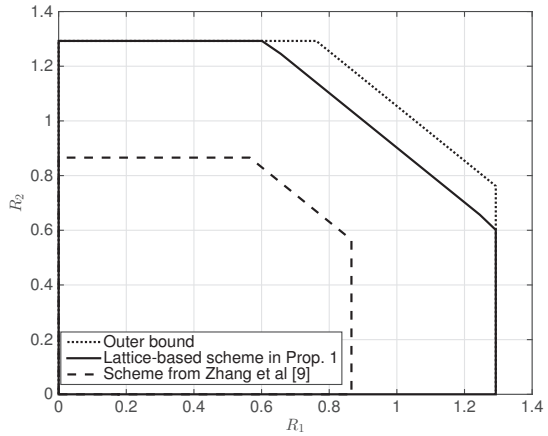


Figure 5.9 – The achievable rate regions for the a state-dependent Gaussian IC. In the case when interfering state $\mathbf{s}$ has very large power Q, the proposed scheme can outperform the best known results.

$$N_{21} = (\alpha_{21}g_2 - a_{21}\beta_1)^2 P_1 + (\alpha_{21} - a_{22}\beta_2)^2 P_2 + \alpha_{21}^2 + (\alpha_{21}c_2 - \sum_{i=1}^{2} a_{2i}\gamma_i)^2 Q \tag{5.65c}$$

$$N_{22} = (\alpha_{22}g_2 - \lambda_2 a_{21}\beta_1 - b_{21}\beta_1)^2 P_1 + (\alpha_{22} - \lambda_2 a_{22}\beta_2 - b_{22}\beta_2)^2 P_2 + \alpha_{22}^2 + (\alpha_{22}c_2 - \sum_{i=1}^{2}(\lambda_2 a_{2i} + b_{2i})\gamma_i)^2 Q \tag{5.65d}$$

Using lattice decoding, we can show the following achievable rate region for the 2-user Gaussian IC with state.

Proposition 5.4. *The following rates are achievable for the Gaussian IC with states in (5.64):*

$$R_k = \frac{1}{2}\log^+ \frac{\beta_k^2 P_k}{\max\{N_{11}\cdot \mathbf{1}_{a_{1k}}, N_{12}\cdot \mathbf{1}_{b_{1k}}, N_{21}\cdot \mathbf{1}_{a_{2k}}, N_{22}\cdot \mathbf{1}_{b_{2k}}\}}$$

for any $\alpha_{k1}, \alpha_{k2}, \lambda_k, \gamma_k, \beta_k$ and full rank integer coefficient matrices $\mathbf{A}_k = \begin{pmatrix} a_{k1} & a_{k2} \\ b_{k1} & b_{k2} \end{pmatrix}$ in (5.65), $k = 1, 2$. The indicator function $\mathbf{1}_a$ evaluates to 1 if $a \neq 0$ and to 0 otherwise. We define $\log^+ x := \max\{0, \log x\}$.

Depending on system parameters , the lattice-based scheme for the Gaussian IC with state can outperform the best known schemes, especially when the interference $\mathbf{s}$ is very strong. We show such an example in Figure 5.9. We consider a symmetric Gaussian IC with state in (5.64) with parameters $P_1 = P_2 = 5, g_1 = g_2 = 1.5, Q = 6000$ and compare our achievable rate region with the best known result from [58, Thm. 3]. We use the capacity region in (5.56) as an outer bound.

The capacity region for this channel is characterized in [59] for certain parameter regimes. However the capacity result for the following special case seems not to be present in the literature.

Lemma 5.3 (Capacity for a special case)**.** *For the Gaussian IC with states given in (5.64) with $g_1 = g_2 = 1$ and $c_1 = c_2$, if it holds that $\sqrt{\frac{P_1 P_2}{1+P_1+P_2}} \geq 1$, the capacity region is given by*

$$R_1 \leq \mathsf{C}(P1), \quad R_2 \leq \mathsf{C}(P_2), \quad R_1 + R_2 \leq \mathsf{C}(P_1 + P_2)$$

Proof. The converse is obvious. For the achivability part, note that if it holds that $g_1 = g_2 = 1$ and $c_1 = c_2$, the system is equivalent to two Gaussian MACs which are exactly the same. Indeed, notice that in this case the noises N_{k1}, N_{k2} in (5.65) at two receivers $k = 1, 2$ are identical if we choose $\alpha_{k1}, \alpha_{k2}, \lambda_k$ and $\mathbf{A}_k$ to be the same for $k = 1, 2$. Further notice that for *any* α_{k1}, α_{k2} and λ_k we can choose γ_1, γ_2 such that the terms in (5.65) involving Q vanish. Hence the interference $\mathbf{s}$ can be canceled out completely and the system is equivalent to two identical usual Gaussian MAC (without interference). Using the result in Theorem 4.2, we know that the entire capacity region of the corresponding Gaussian MAC can be achieved with the coefficient matrices $\mathbf{A}_k = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ or $\mathbf{A}_k = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), k = 1, 2$. □

5.5 Appendix

5.5.1 Proof of Theorem 5.1

In this section we provide a detailed proof for Theorem 5.1. We also discuss the choice of the fine lattices Λ_k in the codes constructions. The encoding procedure has been discussed in section 5.1.3, now we consider the decoding procedure at the primary user. The received signal $\mathbf{y}_0$ at the primary decoder is

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{x}_0 + \sum_{k\geq 1} b_k \mathbf{x}_k + \mathbf{z}_0 && (5.66)\\ &= (1 + \sum_{k\geq 1} b_k \sqrt{\lambda_k})\mathbf{x}_0 + \sum_{k\geq 1} b_k \hat{\mathbf{x}}_k + \mathbf{z}_0 && (5.67)\\ &= b_0 \mathbf{x}_0 + \sum_{k\geq 1} b_k \hat{\mathbf{x}}_k + \mathbf{z}_0 && (5.68) \end{aligned}$$

where we define $b_0 := 1 + \sum_{k\geq 1} b_k \sqrt{\lambda_k}$.

Given a set of integers $\mathbf{a}(1) := \{a_k(1) \in \mathbb{Z}, k \in [0 : K]\}$ and some scalar $\alpha_1 \in \mathbb{R}$, the primary decoder can form the following:

$$\begin{aligned} \tilde{\mathbf{y}}_0^{(1)} &= \alpha_1 \mathbf{y}_0 - \sum_{k\geq 0} a_k(1)\beta_k \mathbf{d}_k && (5.69)\\ &= (\alpha_1 b_0 - a_0(1)\beta_0)\mathbf{x}_0 + \sum_{k\geq 1} (\alpha_1 b_k - a_k(1)\beta_k)\hat{\mathbf{x}}_k + \alpha_1 \mathbf{z}_0 && (5.70)\\ &\quad + \sum_{k\geq 1} a_k(1)\beta_k \hat{\mathbf{x}}_k + a_0(1)\beta_0 \mathbf{x}_0 - \sum_{k\geq 0} a_k(1)\beta_k \mathbf{d}_k. && (5.71) \end{aligned}$$

Rewrite the last three terms in the above expression as

$$\sum_{k\geq 1} a_k(1)\beta_k\hat{\mathbf{x}}_k + a_0(1)\beta_0\mathbf{x}_0 - \sum_{k\geq 0} a_k(1)\beta_k\mathbf{d}_k \tag{5.72}$$

$$\begin{aligned}
&\overset{(b)}{=} \sum_{k\geq 1} a_k(1)\left(\beta_k(\frac{\mathbf{t}_k}{\beta_k} - \frac{\gamma_k\mathbf{x}_0}{\beta_k}) - \beta_k Q_{\frac{\Lambda_k^s}{\beta_k}}(\frac{\mathbf{t}_k}{\beta_k} + \mathbf{d}_k - \frac{\gamma_k\mathbf{x}_0}{\beta_k})\right)\\
&\quad + a_0(1)\left(\beta_0\mathbf{t}_0 - \beta_0 Q_{\frac{\Lambda_0^s}{\beta_0}}(\frac{\mathbf{t}_0}{\beta_0} + \mathbf{d}_0)\right)\\
&\overset{(c)}{=} -\sum_{k\geq 1} a_k(1)\gamma_k\mathbf{x}_0 + a_0(1)(\mathbf{t}_0 - Q_{\Lambda_0^s}(\mathbf{t}_0 + \beta_0\mathbf{d}_0))\\
&\quad + \sum_{k\geq 1} a_k(1)\left(\mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \gamma_k\mathbf{x}_0)\right)\\
&\overset{(d)}{=} -\sum_{k\geq 1} a_k(1)\gamma_k\mathbf{x}_0 + \sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k.
\end{aligned} \tag{5.73}$$

In step (b) we used the definition of the signals $\mathbf{x}_0$ and $\hat{\mathbf{x}}_k$ from Eqn. (5.23). Step (c) uses the identity $Q_\Lambda(\beta\mathbf{x}) = \beta Q_{\frac{\Lambda}{\beta}}(\mathbf{x})$ for any real number $\beta \neq 0$. In step (d) we define $\tilde{\mathbf{t}}_k$ for user k as

$$\tilde{\mathbf{t}}_0 := \mathbf{t}_0 - Q_{\Lambda_0^s}(\mathbf{t}_0 + \beta_k\mathbf{d}_0) \tag{5.74}$$

$$\tilde{\mathbf{t}}_k := \mathbf{t}_k - Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \gamma_k\mathbf{x}_0) \quad k \in [1:K]. \tag{5.75}$$

Define $g(1) := \sum_{k\geq 1} a_k(1)\gamma_k$ and substitute the expression (5.73) into $\tilde{\mathbf{y}}_0^{(1)}$ to get

$$\begin{aligned}
\tilde{\mathbf{y}}_0^{(1)} &= (\alpha_1 b_0 - a_0(1)\beta_0 - g(1))\,\mathbf{x}_0 + \sum_{k\geq 1}(\alpha_1 b_k - a_k(1)\beta_k)\hat{\mathbf{x}}_k + \alpha_1\mathbf{z}_0 + \sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k\\
&= \tilde{\mathbf{z}}_0(1) + \sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k
\end{aligned} \tag{5.76}$$

where we define the equivalent noise $\tilde{\mathbf{z}}_0(1)$ at the primary receiver as:

$$\tilde{\mathbf{z}}_0(1) := \alpha_1\mathbf{z}_0 + (\alpha_1 b_0 - a_0(1)\beta_0 - g(1))\mathbf{x}_0 + \sum_{k\geq 1}(\alpha_1 b_k - a_k(1)\beta_k)\hat{\mathbf{x}}_k \tag{5.77}$$

where $b_0 := 1 + \sum_{k\geq 1} b_k\sqrt{\lambda_k}$.

Notice that we have $\tilde{\mathbf{t}}_k \in \Lambda_k$ since $\mathbf{t}_k \in \Lambda_k$ and $\Lambda_k^s \subseteq \Lambda_c$ due to the lattice code construction (recall that Λ_c denotes the coarsest lattice among Λ_k). Furthermore because all Λ_k are chosen to form a nested lattice chain, the integer combination $\sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k$ also belongs to a Λ_k for some $k \in [0:K]$. Furthermore, the equivalent noise $\tilde{\mathbf{z}}_0(1)$ is independent of the signal $\sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k$ thanks to the dithers $\mathbf{d}_k$.

The primary decoder uses lattice decoding to decode the integer sum $\sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_\mathbf{k}$ by quantizing $\tilde{\mathbf{y}}_0^{(1)}$ to its nearest neighbor in Λ. A decoding error occurs when $\tilde{\mathbf{y}}_0^{(1)}$ falls outside the Voronoi region around the lattice point $\sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_\mathbf{k}$. The probability of this event is equal to the probability that the equivalent noise $\tilde{\mathbf{z}}_0(1)$ leaves the Voronoi region of the fine lattice, i.e., $\Pr(\tilde{\mathbf{z}}_0(1) \notin \mathcal{V})$. The same as in the proof of [8, Theorem 5], the probability $\Pr(\tilde{\mathbf{z}}_0(1) \notin \mathcal{V})$ goes to zero if the probability

$\Pr(\mathbf{z}_0^*(1) \notin \mathcal{V})$ goes to zero where $\mathbf{z}_0^*(1)$ is a zero-mean Gaussian vector with i.i.d entries whose variance equals the variance of the noise $\tilde{\mathbf{z}}_0(1)$:

$$N_0(1) = \alpha_1^2 + (\alpha_1 b_0 - a_0(1)\beta_0 - g(1))^2 P + \sum_{k\geq 1} (\alpha_1 b_k - a_k(1)\beta_k)^2 \bar{\lambda}_k P. \quad (5.78)$$

Lattice decoding will be successful if

$$R_k < r_k(\mathbf{a}_1, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) := \frac{1}{2}\log\left(\frac{\sigma_k^2}{N_0(1)}\right) - \frac{1}{2}\log(1+\delta) \quad (5.79)$$

that is

$$R_0 < \frac{1}{2}\log^+\left(\frac{\beta_0^2 P}{\alpha_1^2 + P\,||\alpha_1 \mathbf{h} - \tilde{\mathbf{a}}||^2}\right) \quad (5.80a)$$

$$R_k < \frac{1}{2}\log^+\left(\frac{(1-\lambda_k)\beta_k^2 P}{\alpha_1^2 + P\,||\alpha_1 \mathbf{h} - \tilde{\mathbf{a}}||^2}\right) \quad k \in [1:K] \quad (5.80b)$$

if we choose δ arbitrarily small and define

$$\mathbf{h} := [b_0, b_1\sqrt{\bar{\lambda}_1}, \ldots, b_K\sqrt{\bar{\lambda}_K}] \quad (5.81)$$

$$\tilde{\mathbf{a}} := [a_0(1)\beta_0 + g(1), a_1(1)\beta_1\sqrt{\bar{\lambda}_1}, \ldots, a_K(1)\beta_K\sqrt{\bar{\lambda}_K}]. \quad (5.82)$$

Notice we can optimize over α_1 to maximize the above rates.

At this point, the primary user has successfully decoded one integer sum of the lattice points $\sum_{k\geq 0} a_k \tilde{\mathbf{t}}_k$. As mentioned earlier, we may continue decoding other integer sums with the help of this sum. The method of performing *successive compute-and-forward* in [29] is to first recover a linear combination of all transmitted signals $\tilde{\mathbf{x}}_k$ from the decoded integer sum and use it for subsequent decoding. Here we are not able to do this because the cognitive channel input $\hat{\mathbf{x}}_k$ contains $\mathbf{x}_0$ which is not known at Receiver 0. In order to proceed, we use the observation that if $\sum_{k\geq 0} a_k \tilde{\mathbf{t}}_k$ can be decoded reliably, then we know the equivalent noise $\tilde{\mathbf{z}}_0(1)$ and can use it for the subsequent decoding.

In general assume the primary user has decoded $\ell - 1$ integer sums $\sum_k a_k(j)\mathbf{t}_k$, $j \in [1:\ell-1], \ell \geq 2$ with positive rates, and about to decode another integer sum with coefficients $\mathbf{a}(\ell)$. We show separately in Appendix 5.5.2 that with the previously known $\tilde{\mathbf{z}}_0(\ell-1)$ for $\ell \geq 2$, the primary decoder can form

$$\tilde{\mathbf{y}}_0^{(\ell)} = \tilde{\mathbf{z}}_0(\ell) + \sum_{k\geq 0} a_k(\ell)\tilde{\mathbf{t}}_k \quad (5.83)$$

with the equivalent noise $\tilde{\mathbf{z}}_0(\ell)$

$$\begin{aligned}\tilde{\mathbf{z}}_0(\ell) := \alpha_\ell \mathbf{z}_0 &+ \sum_{k\geq 1}\left(\alpha_\ell b_k - a_k(\ell)\beta_k - \sum_{j=1}^{\ell-1}\alpha_j a_k(j)\beta_k\right)\hat{\mathbf{x}}_k \\ &+ \left(\alpha_\ell b_0 - a_0(\ell)\beta_0 - \sum_{j=1}^{\ell-1}\alpha_j a_0(j)\beta_0 - g(\ell)\right)\mathbf{x}_0 \end{aligned} \quad (5.84)$$

where $g(\ell)$ is defined in (5.30) and the scaling factors $\alpha_1, \ldots, \alpha_\ell$ are to be optimized.

In the same vein as we derived (5.79), using $\tilde{\mathbf{y}}_0^{(l)}$ we can decode the integer sums of the lattice codewords $\sum_{k\geq 0} a_k(\ell)\tilde{\mathbf{t}}_\mathbf{0}$ reliably using lattice decoding if the fine lattice satisfy

$$\frac{(\text{Vol }(\mathcal{V}_k))^{2/n}}{N_0(\ell)} > 2\pi e \tag{5.85}$$

for k satisfying $a_k(\ell) \neq 0$ and we use $N_0(\ell)$ to denote the variance of the equivalent noise $\tilde{\mathbf{z}}_0(\ell)$ per dimension given in (5.28). Equivalently we require the rate R_k to be smaller than

$$r_k(\mathbf{a}_{\ell|1:\ell-1}, \underline{\lambda}, \underline{\beta}, \underline{\gamma}) := \max_{\alpha_1,\ldots,\alpha_\ell \in \mathbb{R}} \frac{1}{2}\log^+\left(\frac{\sigma_k^2}{N_0(\ell)}\right) \tag{5.86}$$

where σ_k^2 is given in (5.20). Thus we arrive at the same expression in (5.27) as claimed.

Recalling the definition of the set $\mathcal{A}(L)$ in (5.31), we now show that if the coefficient matrix A is in this set, the term $\tilde{\mathbf{t}}_0$ can be solved using the L integer sums with coefficients $\mathbf{a}(1), \ldots, \mathbf{a}(L)$.

For the case $\text{rank}(\mathbf{A}) = K+1$ the statement is trivial. For the case $\text{rank}(\mathbf{A}) = m \leq L < K+1$, we know that by performing Gaussian elimination on $\mathbf{A'} \in \mathbb{Z}^{L\times K}$ with rank $m-1$, we obtain a matrix whose last $L-m+1$ rows are zeros. Notice that $\mathbf{A} \in \mathbb{Z}^{L\times K+1}$ is a matrix formed by adding one more column in front of $\mathbf{A'}$. So if we perform exactly the same Gaussian elimination procedure on the matrix $\mathbf{A}$, there must be at least one row in the last $L-m+1$ row whose first entry is non-zero, since $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A'})+1$. This row will give the value of $\tilde{\mathbf{t}}_0$. Finally the true codeword $\mathbf{t}_0$ can be recovered as

$$\mathbf{t}_0 = [\tilde{\mathbf{t}}_0] \text{mod } \Lambda_0^s. \tag{5.87}$$

Now we consider the decoding procedure at the cognitive receivers, for whom it is just a point-to-point transmission problem over Gaussian channel using lattice codes. The cognitive user k can process its received signal for some ν_k as

$$\begin{aligned}
\tilde{\mathbf{y}}_k &= \nu_k \mathbf{y}_k - \beta_k \mathbf{d}_k \\
&= \nu_k(\mathbf{z}_k + \sqrt{\lambda_k}h_k\mathbf{x}_0) + (\nu_k h_k - \beta_k)\hat{\mathbf{x}}_k + \beta_k\hat{\mathbf{x}}_k - \beta_k\mathbf{d}_k \\
&= \nu_k(\mathbf{z}_k + \sqrt{\lambda_k}h_k\mathbf{x}_0) + (\nu_k h_k - \beta_k)\hat{\mathbf{x}}_k - \beta_k\mathbf{d}_k \\
&\quad + Q_{\Lambda_k^s}(\mathbf{t}_k + \beta_k\mathbf{d}_k - \gamma_k\mathbf{x}_0) + \beta_k(\frac{\mathbf{t}_k}{\beta_k} + \mathbf{d}_k - \frac{\gamma_k}{\beta_k}\mathbf{x}_0) \\
&= \tilde{\mathbf{z}}_k + \tilde{\mathbf{t}}_k.
\end{aligned}$$

In the last step we define the equivalent noise as

$$\tilde{\mathbf{z}}_k := \nu_k\mathbf{z}_k + (\nu_k h_k - \beta_k)\hat{\mathbf{x}}_k + (\nu_k\sqrt{\lambda_k}h_k - \gamma_k)\mathbf{x}_0 \tag{5.88}$$

and $\tilde{\mathbf{t}}_k$ as in (5.75).

Using the same argument as before, we can show that the codeword $\tilde{\mathbf{t}}_k$ can be decoded reliably using lattice decoding if

$$\frac{(\text{Vol }(\mathcal{V}_k))^{2/n}}{N_k(\gamma_k)} > 2\pi e \tag{5.89}$$

for all $k \geq 1$ where $N_k(\gamma)$ is the variance of the equivalent noise $\tilde{\mathbf{z}}_k$ per dimension given in (5.32). Equivalently the cognitive rate R_k should satisfy

$$R_k < \max_{\nu_k} \frac{1}{2} \log \frac{\sigma_k^2}{N_k(\gamma_k)}. \tag{5.90}$$

Similarly we can obtain $\mathbf{t}_k$ from $\tilde{\mathbf{t}}_k$ as $\mathbf{t}_k = [\tilde{\mathbf{t}}_k] \bmod \Lambda_k^s$. This completes the proof of Theorem 5.1.

We also determined how to choose the fine lattice Λ_k. Summarizing the requirements in (5.89) and (5.85) on Λ_k for successful decoding, the fine lattice Λ_0 of the primary user satisfies

$$(\text{Vol } (\mathcal{V}_0))^{2/n} > 2\pi e N_0(\ell) \tag{5.91}$$

for all ℓ where $a_0(\ell) \neq 0$ and the fine lattice Λ_k of the cognitive user k, $k \in [1:K]$, satisfies

$$(\text{Vol } (\mathcal{V}_k))^{2/n} > \max\{2\pi e N_0(\ell), 2\pi e N_k(\gamma_k)\} \tag{5.92}$$

for all ℓ where $a_k(\ell) \neq 0$. Recall that the fine lattices Λ_k are chosen to form a nested lattice chain. Now the order of this chain can be determined by the volumes of $\mathcal{V}_k$ given above.

5.5.2 Derivations in the proof of Theorem 5.1

We give the details for the claim made in Appendix 5.5.1 that we could form the equivalent channel

$$\tilde{\mathbf{y}}_0^{(\ell)} = \tilde{\mathbf{z}}_0(\ell) + \sum_{k \geq 0} a_k(\ell) \tilde{\mathbf{t}}_k$$

with $\tilde{\mathbf{z}}_0(\ell)$ defined in (5.84) when the primary decoder decodes the ℓ-th integer sum $\sum_{k\geq 0} a_k(\ell)\tilde{\mathbf{t}}_k$ for $\ell \geq 2$.

We first show the base case for $\ell = 2$. Since $\sum_{k\geq 0} a_k(1)\tilde{\mathbf{t}}_k$ is decoded, the equivalent noise $\tilde{\mathbf{z}}_0(1)$ in Eqn. (5.77) can be inferred from $\tilde{\mathbf{y}}_0$. Given α_{20}, α_{21} we form the following with $\mathbf{y}_0$ in (5.68) and $\tilde{\mathbf{z}}_0(1)$

$$\begin{aligned}
\tilde{\mathbf{y}}_0^{(2)} &:= \alpha_{20}\mathbf{y}_0 + \alpha_{21}\tilde{\mathbf{z}}_0(1) \\
&= (\alpha_{20} + \alpha_{21}\alpha_1)\mathbf{z}_0 + \sum_{k\geq 1}((\alpha_{20} + \alpha_{20}\alpha_1)b_k - \alpha_{21}a_k(1)\beta_k)\hat{\mathbf{x}}_k \\
&\quad +((\alpha_{20} + \alpha_{21}\alpha_1)b_0 - \alpha_{21}a_0(1)\beta_0 - \alpha_{21}g(1))\mathbf{x}_0 \\
&= \alpha_2'\mathbf{z}_0 + \sum_{k\geq 1}(\alpha_2' b_k - \alpha_1' a_k(1)\beta_k)\hat{\mathbf{x}}_k + (\alpha_2' b_0 - \alpha_1' a_0(1)\beta_0 - \alpha_1' g(1))\mathbf{x}_0
\end{aligned}$$

by defining $\alpha_1' := \alpha_{21}$ and $\alpha_2' := \alpha_{20} + \alpha_{21}\alpha_1$. Now following the same step for deriving $\tilde{\mathbf{y}}_0^{(1)}$ in (5.76), we can rewrite $\tilde{\mathbf{y}}_0^{(2)}$ as

$$\tilde{\mathbf{y}}_0^{(2)} = \sum_{k\geq 0} a_k(2)\tilde{\mathbf{t}}_k + \tilde{\mathbf{z}}_0(2) \tag{5.93}$$

with

$$\tilde{\mathbf{z}}_0(2) := \alpha_2'\mathbf{z}_0 + \sum_{k\geq 1}(\alpha_2' b_k - a_k(2)\beta_k - \alpha_1' a_k(1)\beta_k)\hat{\mathbf{x}}_k \tag{5.94}$$

$$+(\alpha_2' b_0 - a_0(2)\beta_0 - \alpha_1' a_0(1)\beta_0 - g(2))\mathbf{x}_0 \tag{5.95}$$

This establishes the base case by identifying $\alpha_i' = \alpha_i$ for $i = 1, 2$.

Now assume the expression (5.84) is true for $\ell - 1$ ($\ell \geq 3$) and we have inferred $\tilde{\mathbf{z}}_0(m)$ from $\tilde{\mathbf{y}}_0^{(m)}$ using the decoded sum $\sum_{k\geq 0} a_k(m)\tilde{\mathbf{t}}_k$ for all $m \leq 1, \ldots, \ell - 1$, we will form $\tilde{\mathbf{y}}_0^{(\ell)}$ with ℓ numbers $\alpha_{\ell 0}, \ldots, \alpha_{\ell\ell-1}$ as

$$\begin{aligned}\tilde{\mathbf{y}}_0^{(\ell)} &:= \alpha_{\ell 0}\mathbf{y}_0 + \sum_{m=1}^{\ell-1}\alpha_{\ell m}\tilde{\mathbf{z}}_0(m) \\ &= \alpha_\ell'\mathbf{z}_0 + \sum_{k\geq 1}\left(\alpha_\ell' b_k - \beta_k C_{\ell-1}(k)\right)\hat{\mathbf{x}}_k + \left(\alpha_\ell' b_0 - \beta_0 C_{\ell-1}(0) - \sum_{m=1}^{\ell-1}\alpha_{\ell m} g(m)\right)\mathbf{x}_0\end{aligned}$$

with

$$\alpha_\ell' := \alpha_{\ell 0} + \sum_{m=1}^{\ell-1}\alpha_{\ell m}\alpha_m \tag{5.96}$$

$$C_{\ell-1}(k) := \sum_{m=1}^{\ell-1}\alpha_{\ell m}\left(a_k(m) + \sum_{j=1}^{m-1}\alpha_j a_k(j)\right). \tag{5.97}$$

Algebraic manipulations allow us to rewrite $C_{\ell-1}(k)$ as

$$C_{\ell-1}(k) = \sum_{m=1}^{\ell-1}\left(\alpha_{\ell m} + \alpha_m \sum_{j=m+1}^{\ell-1}\alpha_{\ell j}\right)a_k(m) \tag{5.98}$$

$$= \sum_{m=1}^{\ell-1}\alpha_m' a_k(m) \tag{5.99}$$

by defining $\alpha_m' := \alpha_{\ell m} + \alpha_m \sum_{j=m+1}^{\ell-1}\alpha_{\ell j}$ for $m = 1, \ldots, \ell - 1$. Substituting the above into $\tilde{\mathbf{y}}_0^{(\ell)}$ we get

$$\tilde{\mathbf{y}}_0^{(\ell)} = \alpha_\ell'\mathbf{z}_0 + \sum_{k\geq 1}\left(\alpha_\ell' b_k - \beta_k \sum_{m=1}^{\ell-1}\alpha_m' a_k(m)\right)\hat{\mathbf{x}}_k \tag{5.100}$$

$$+\left(\alpha_\ell' b_0 - \beta_0 \sum_{m=1}^{\ell-1}\alpha_m' a_0(m) - \sum_{m=1}^{\ell-1}\alpha_{\ell m} g(m)\right)\mathbf{x}_0. \tag{5.101}$$

Together with the definition of $g(m)$ in (5.30) and some algebra we can show

$$\sum_{m=1}^{\ell-1} a_{\ell m} g(m) = \sum_{k=1}^{K}\gamma_k C_{\ell-1}(k) \tag{5.102}$$

$$= \sum_{k=1}^{K}\left(\sum_{m=1}^{\ell-1}\alpha_m' a_k(m)\right)\gamma_k. \tag{5.103}$$

Finally using the same steps for deriving $\tilde{\mathbf{y}}_0^{(1)}$ in (5.76) and identifying $\alpha'_m = \alpha_m$ for $m = 1, \ldots, \ell$, it is easy to see that we have

$$\tilde{\mathbf{y}}_0^{(\ell)} = \sum_{k \geq 0} a_k(\ell)\tilde{\mathbf{t}}_k + \tilde{\mathbf{z}}_0(\ell) \tag{5.104}$$

with $\tilde{\mathbf{z}}_0(\ell)$ claimed in (5.84).

5.5.3 Proof of Proposition 5.3

For any given set of parameters $\{\alpha_j, j \in [1 : \ell]\}$ in the expression $N_0(\ell)$ in (5.38) , we can always find another set of parameters $\{\alpha'_j, j \in [1 : \ell]\}$ and a set of vectors $\{\mathbf{u}_j, j \in [1 : \ell]\}$, such that

$$\alpha_\ell \mathbf{h} + \sum_{j=1}^{\ell-1} \alpha_j \tilde{\mathbf{a}}_j = \sum_{j=1}^{\ell} \alpha'_j \mathbf{u}_j \tag{5.105}$$

as long as the two sets of vectors, $\{\mathbf{h}, \tilde{\mathbf{a}}_j, j \in [1 : \ell - 1]\}$ and $\{\mathbf{u}_j, j \in [1 : \ell]\}$ span the same subspace. If we choose an appropriate set of basis vectors $\{\mathbf{u}_j\}$, the minimization problem of $N_0(\ell)$ can be equivalently formulated with the set $\{\mathbf{u}_j\}$ and new parameters $\{\alpha'_j\}$ where the optimal $\{\alpha'_j\}$ have simple solutions. Notice that $\{\mathbf{u}_j, j \in [1 : \ell]\}$ in Eqn. (5.40) are obtained by performing the Gram-Schmidt procedure on the set $\{\mathbf{h}, \tilde{\mathbf{a}}_j, j \in [1 : \ell - 1]\}$. Hence the set $\{\mathbf{u}_j, j \in [1 : \ell]\}$ contains orthogonal vectors and spans the same subspace as the set $\{\mathbf{h}, \tilde{\mathbf{a}}_j, j \in [1 : \ell - 1]\}$ does. For any $\ell \geq 1$, the expression $N_0(\ell)$ in (5.38) can be equivalently rewritten as

$$N_0(\ell) = \alpha'^2_\ell + \left|\left| \sum_{j=1}^{\ell} \alpha'_j \mathbf{u}_j - \tilde{\mathbf{a}}_\ell \right|\right|^2 P \tag{5.106}$$

with $\{\mathbf{u}_j\}$ defined above and some $\{\alpha'_j\}$. Due to the orthogonality of vectors $\{\mathbf{u}_j\}$, we have the following simple optimal solutions for $\{\alpha'^*_j\}$ which minimize $N_0(\ell)$:

$$\alpha'^*_j = \frac{\tilde{\mathbf{a}}_\ell^T \mathbf{u}_j}{||\mathbf{u}_j||^2}, \quad j \in [1 : \ell - 1] \tag{5.107}$$

$$\alpha'^*_\ell = \frac{P\tilde{\mathbf{a}}_\ell^T \mathbf{u}_\ell}{P\,||\mathbf{u}_\ell||^2 + 1}. \tag{5.108}$$

Substituting them back to $N_0(\ell)$ in (5.106) we have

$$N_0(\ell) = P\,||\tilde{\mathbf{a}}_\ell||^2 - \sum_{j=1}^{\ell-1} \frac{(\tilde{\mathbf{a}}_\ell^T \mathbf{u}_j)^2 P}{||\mathbf{u}_j||^2} - \frac{P^2(\mathbf{u}_\ell^T \tilde{\mathbf{a}}_\ell)^2}{1 + P\,||\mathbf{u}_\ell||^2} \tag{5.109}$$

$$= P\tilde{\mathbf{a}}_\ell^T \left(\mathbf{I} - \sum_{i=1}^{\ell-1} \frac{\mathbf{u}_j \mathbf{u}_j^T}{||\mathbf{u}_j||^2} - \frac{(\mathbf{u}_\ell \mathbf{u}_\ell^T)P}{1 + P\,||\mathbf{u}_\ell||^2} \right) \tilde{\mathbf{a}}_\ell \tag{5.110}$$

$$= P\mathbf{a}(\ell)^T \mathbf{B}_\ell \mathbf{a}(\ell) \tag{5.111}$$

with $\mathbf{B}_\ell$ given in (5.42). As we discussed before, maximizing $r_k(\mathbf{a}_{\ell|1:\ell-1})$ is equivalent to minimizing $N_0(\ell)$ and the optimal coefficients $\mathbf{a}(\ell), \ell \in [1 : L]$ are the same for all users. This proves the claim.

5.5.4 An outer bound on the capacity region

In this section we give a simple outer bound on the capacity region of the cognitive many-to-one channel, which is used for the numerical evaluation in Figure 5.3, Section 5.1.5. Notice that if we allow all transmitters $k = 0, \ldots, K$ to cooperate, and allow the cognitive receivers $k = 1, \ldots, K$ to cooperate, then the system can be seen as a 2-user broadcast channel where the transmitter has $K + 1$ antennas. The two users are the primary receiver and the aggregation of all cognitive receivers with K antennas. Obviously the capacity region of this resulting 2-user MIMO broadcast channel will be a valid outer bound on the capacity region of the cognitive many-to-one channel. The capacity region $\mathcal{C}_{BC}$ of the broadcast channel is given by (see [6, Ch. 9] for example)

$$\mathcal{C}_{BC} = \mathcal{R}_1 \bigcup \mathcal{R}_2 \tag{5.112}$$

where $\mathcal{R}_1$ is defined as

$$R_1 \leq \frac{1}{2} \log \frac{|\mathbf{H}_1(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{H}_1^T + \mathbf{I}|}{|\mathbf{H}_1\mathbf{K}_2\mathbf{H}_2^T + \mathbf{I}|} \tag{5.113}$$

$$R_2 \leq \frac{1}{2} \log |\mathbf{H}_2\mathbf{K}_2\mathbf{G}_2^T + \mathbf{I}| \tag{5.114}$$

and $\mathcal{R}_2$ defined similarly with all subscripts 1 and 2 in $\mathcal{R}_1$ swapped. The channel matrices $\mathbf{H}_1 \in \mathbb{R}^{1\times(K+1)}$ and $\mathbf{H}_2 \in \mathbb{R}^{K\times(K+1)}$ are defined as

$$\mathbf{H}_1 = \begin{bmatrix} 1 & b_1 & \ldots & b_K \end{bmatrix} \tag{5.115}$$

$$\mathbf{H}_2 = \begin{bmatrix} 0 & h_1 & 0 & \ldots & 0 \\ 0 & 0 & h_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & h_K \end{bmatrix} \tag{5.116}$$

where $\mathbf{H}_1$ denotes the channel from the aggregated transmitters to the primary receiver and $\mathbf{H}_2$ denotes the channel to all cognitive receivers. The variables $\mathbf{K}_1, \mathbf{K}_2 \in \mathbb{R}^{(K+1)\times(K+1)}$ should satisfy the condition

$$\mathrm{tr}(\mathbf{K}_1 + \mathbf{K}_2) \leq (K + 1)P \tag{5.117}$$

which represents the power constraint for the corresponding broadcast channel[4]. As explained in [6, Ch. 9], the problem of finding the region $\mathcal{C}_{BC}$ can be rewritten as convex optimization problems which are readily solvable using standard convex optimization tools.

5.5.5 Proof of Theorem 5.3

Proof. For the symmetric non-cognitive many-to-one channel, we have the following trivial capacity bound

$$R_0 \leq \frac{1}{2} \log(1 + P) \tag{5.118}$$

$$R_k \leq \frac{1}{2} \log(1 + h^2 P). \tag{5.119}$$

[4]Since each transmitter has its individual power constraint, we could give a slightly tighter outer bound by imposing a per-antenna power constraint. Namely the matrices $\mathbf{K}_1, \mathbf{K}_2$ should satisfy $(\mathbf{K}_1 + \mathbf{K}_2)_{ii} \leq P$ for $i \in [1 : K + 1]$ where $(\mathbf{X})_{ii}$ denotes the (i, i) entry of matrix $\mathbf{X}$. However this is not the focus of this paper and we will not pursue it here.

To show the constant gap result, we choose the coefficients matrix of the two sums to be

$$\mathbf{A} = \begin{pmatrix} 1 & c & \dots & c \\ 0 & 1 & \dots & 1 \end{pmatrix} \tag{5.120}$$

for some nonzero integer c. Furthermore we choose $\beta_0 = 1$ and $\beta_k = b/c$ for all $k \geq 1$. In Appendix 5.5.6 we use Theorem 5.2 to show the following rates are achievable:

$$R_0 = \frac{1}{2}\log^+ P \tag{5.121}$$

$$R_k = \min\left\{\frac{1}{2}\log^+ \frac{b^2 P}{c^2}, \frac{1}{2}\log^+ b^2, \frac{1}{2}\log(1+h^2P)\right\}. \tag{5.122}$$

If $|b| \geq |h|\left\lceil\sqrt{P}\right\rceil$, choosing $c = \left\lceil\sqrt{P}\right\rceil$ will ensure $R_k \geq \frac{1}{2}\log^+ h^2P$.

Notice that for $P \leq 1$, then $\frac{1}{2}\log(1+P) \leq 0.5$ hence the claim is vacuously true. For $P \geq 1$, we have

$$\frac{1}{2}\log(1+P) - R_0 \leq \frac{1}{2}\log\frac{1+P}{P} \leq \frac{1}{2}\log 2 = 0.5 \text{ bit} \tag{5.123}$$

With the same argument we have

$$\frac{1}{2}\log(1+h^2P) - R_k \leq 0.5 \text{ bit} \tag{5.124}$$

To show the capacity result, we set $\beta_0 = 1$ and $\beta_k = \beta$ for all $k \geq 1$. The receiver 0 decodes two sums with the coefficients matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 0 \end{pmatrix}. \tag{5.125}$$

The achievable rates using Theorem 5.2 is shown in Appendix 5.5.6 to be

$$R_0 = \frac{1}{2}\log(1+P) \tag{5.126}$$

$$R_k = \min\left\{\frac{1}{2}\log\left(\frac{Pb^2}{1+P}\right), \frac{1}{2}\log(1+h^2P)\right\}. \tag{5.127}$$

The inequality

$$\frac{Pb^2}{1+P} \geq 1 + h^2P \tag{5.128}$$

is satisfied if it holds that

$$b^2 \geq \frac{(1+P)(1+h^2P)}{P}. \tag{5.129}$$

This completes the proof. □

5.5.6 Derivations in the proof of Theorem 5.3

We give detailed derivations of the achievable rates in Theorem 5.3 with two chosen coefficient matrices.

When the primary user decodes the first equation ($\ell = 1$) in a symmetric channel, the expression (5.51) for the variance of the equivalent noise simplifies to (denoting $\beta_k = \beta$ for $k \geq 1$)

$$\tilde{N}_0(1) = \bar{\alpha}_1^2 + K(\bar{\alpha}_1 b - a_k(1)\beta)^2 P + (\bar{\alpha}_1 - a_0(1)\beta_0)^2 P. \tag{5.130}$$

For decoding the second integer sum, the variance of the equivalent noise (5.51) is given as

$$\begin{aligned}\tilde{N}_0(2) = \alpha_2^2 + K(\alpha_2 b - a_k(2)\beta - \alpha_1 a_k(1)\beta)^2 P \\ +(\alpha_2 - a_0(2)\beta_0 - \alpha_1 a_0(1)\beta_0)^2 P.\end{aligned} \tag{5.131}$$

We first evaluate the achievable rate for the coefficient matrix in (5.120). We choose $\beta_0 = 1$ and $\beta = b/c$. Using Theorem 5.2, substituting $\mathbf{a}(1) = [1, c, \ldots, c]$ and the optimal $\bar{\alpha}_1^* = 1 - \frac{1}{P(Kb^2+1)}$ into (5.130) will give us a rate constraint on R_0

$$\tilde{r}_0(\mathbf{a}_1, \underline{\beta}) = \frac{1}{2}\log^+\left(\frac{1}{1+Kb^2} + P\right) > \frac{1}{2}\log^+ P \tag{5.132}$$

$$\tilde{r}_k(\mathbf{a}_1, \underline{\beta}) = \frac{1}{2}\log^+\left(\frac{b^2 P(Kb^2P + P + 1)}{c^2(Kb^2P + P)}\right) > \frac{1}{2}\log^+\frac{b^2 P}{c^2}. \tag{5.133}$$

Notice here we have replaced the achievable rates with smaller values to make the result simple. We will do the same in the following derivation.

For decoding the second sum with coefficients $\mathbf{a}(2) = [0, 1, \ldots, 1]$, we use Theorem 5.2 and (5.131) to obtain rate constraints for R_k

$$\tilde{r}_k(\mathbf{a}_{2|1}, \underline{\beta}) = \frac{1}{2}\log^+\left(b^2 + \frac{1}{K}\right) > \frac{1}{2}\log^+ b^2 \tag{5.134}$$

with the optimal $\alpha_1^* = \frac{-b^2 K}{c(Kb^2+1)}$ and $\alpha_2^* = 0$. Notice that $\mathbf{a}_0(1) = 0$ hence decoding this sum will not impose any rate constraint on R_0. Therefore we omit the expression $\tilde{r}_0(\mathbf{a}_{2|1}, \underline{\beta})$. Combining the results above with Theorem 5.2 we get the claimed rates in the proof of Theorem 5.3.

Now we evaluate the achievable rate for the coefficient matrix in (5.125). We substitute $\beta_0 = 1$, $\beta_k = \beta$ for any β and $\mathbf{a}(1) = [0, 1, \ldots, 1]$ in (5.130) with the optimal $\bar{\alpha}_1^* = \frac{Kb\beta p}{Kb^2P+P+1}$. Notice again R_0 is not constrained by decoding this sum hence we only have the constraint on R_k as

$$\tilde{r}_k(\mathbf{a}_1, \underline{\beta}) = \frac{1}{2}\log^+\left(\frac{1}{K} + \frac{P}{1+P}b^2\right) > \frac{1}{2}\log^+\frac{Pb^2}{1+P}. \tag{5.135}$$

For the second decoding, using $\mathbf{a}(2) = [1, 0, \ldots, 0]$ in (5.131) gives

$$\tilde{r}_0(\mathbf{a}_{2|1}, \underline{\beta}) = \frac{1}{2}\log(1+P) \tag{5.136}$$

with the optimal scaling factors $\alpha_1^* = \frac{bP}{\beta(P+1)}$ and $\alpha_2^* = \frac{P}{P+1}$. Combining the achievable rates above with Theorem 5.2 gives the claimed result.

Intermezzo: on Computation Rates for the Gaussian MAC

6

We have seen in previous chapters, that the key element to all applications with lattice codes is to decode the sum of codewords via a Gaussian network.[1] In its simplest form, the receiver in a 2-user Gaussian MAC wishes to decode the sum of two codewords from the two users. This problem is of particular interests, because any improvement on the computation rates for the 2-user Gaussian MAC could immediately imply better achievable rates for many other communication networks. Unfortunately, the best computation rate region is not known even for this simple system. In this chapter we will study this problem with the help of nested linear codes.

6.1 Sum Decoding with Nested Linear Codes

Consider again the 2-user Gaussian MAC of the form

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{z} \tag{6.1}$$

where both users are assumed to have the power constraint $\mathbb{E}\,||\mathbf{x}_k||^2 \leq nP$. The white Gaussian noise with unit variance per entry is denoted by $\mathbf{z} \in \mathbb{R}^n$.

Instead of using nested lattice codes, we will equip two users with nested linear codes defined as follows. Let q be a prime number, we use vectors in $\mathbb{F}_q^\ell$ to denote the messages. The codebooks are constructed with the following steps.

- For user k. select a random variable U_k defined on $\mathbb{F}_q$ with an arbitrary probability distribution.
- Generates two matrices $\mathbf{H} \in \mathbb{F}_q^{n\times\ell}, \mathbf{G} \in \mathbb{F}_q^{n\times h}$ and two vectors $\mathbf{d}_k \in \mathbb{F}_q^n$ whose entries are chosen i.i.d. uniformly from $\mathbb{F}_q$.

[1] The material of this chapter has appeared in
J. Zhu and M. Gastpar, "Compute-and-Forward using nested linear codes for the Gaussian MAC", *Proc. Information Theory Workshop (ITW) 2015, Jerusalam, Israel.*

- For any message $\mathbf{w}_k \in \mathbb{F}_q^\ell$, user k tries to find some $\mathbf{a}_k \in \mathbb{F}_q^h$ and form

$$\mathbf{u}_k = \mathbf{H}\mathbf{w}_k \oplus \mathbf{G}\mathbf{a}_k \oplus \mathbf{d}_k$$

  such that $\mathbf{u}_k \in \mathcal{A}_{[U_k]}^{(n)}$. (Recall that $\mathcal{A}_{[U_k]}^{(n)}$ denotes the set of typical sequences with the distribution of U_k.) If this is possible, this $\mathbf{u}_k$ is included in the codebook $\mathcal{C}_k$ as the codeword for the message $\mathbf{w}_k$, otherwise an error occurs. Consequently the rate of this codebook is $\frac{1}{n}\log|\mathcal{C}_k| = \frac{k}{n}\log q$.

We will first consider the case when the two users use the same codebook, hence have the same rate. The goal of the receiver is to decode the sum of two codewords $\mathbf{s} := \mathbf{u}_1 \oplus \mathbf{u}_2$ from the channel output $\mathbf{y}$ where the sum is performed component-wise in the finite field. Let $\hat{\mathbf{s}}$ denote the decoded sum. The error event is defined as

$$P_{e,sum}^{(n)} := \mathbb{P}(\hat{\mathbf{s}} \neq \mathbf{s}) \tag{6.2}$$

where n is the length of codewords.

Let us recall two other known schemes to this problem. With the compute-and-forward scheme, the symmetric computation rate[2]

$$R_{CF}^s(P) := \frac{1}{2}\log(1/2 + P) \tag{6.3}$$

is achievable. With power allocations, the symmetric computation rate

$$\alpha R_{CF}^s(P/\alpha) = \frac{\alpha}{2}\log(1/2 + P/\alpha) \tag{6.4}$$

is achievable for any $\alpha \in [0, 1]$. This result is obtained using Theorem 2.5 by setting channel coefficients to be 1. Notice that the generalized result in Theorem 3.1 does not give a higher symmetric computation rate in the symmetric case.

We should also point out that the compute-and-forward scheme discussed in Chapter 2 and 3 concerns with decoding the modulo sum of the lattice codewords of the form $[\sum_k a_k \mathbf{t}_k] \mod \Lambda$, where $\mathbf{t}_k$ are lattice points in $\mathbb{R}^n$ as in Theorem 2.5 and 3.1. In this approach the receiver wishes to decode the sum of the codewords $\mathbf{u}_1 \oplus \mathbf{u}_2$, where $\mathbf{u}_k$ are codewords in $\mathbb{F}_q^n$. Nevertheless these two sums are equivalent as far as our applications are concerned. Moreover, it is also shown in [8] that if Construction A is used for generating the nested lattice codes, then sum of the lattice codewords $[\sum_k a_k \mathbf{t}_k] \mod \Lambda$ permits us to recover the modulo sum of messages $\mathbf{w}_1 \oplus \mathbf{w}_2$, and $\mathbf{u}_1 \oplus \mathbf{u}_2$ is readily obtainable since $\mathbf{u}_1 \oplus \mathbf{u}_2 = \mathbf{H}(\mathbf{w} \oplus \mathbf{w}_2) \oplus G(\mathbf{a}_1 \oplus \mathbf{a}_2) \oplus \mathbf{d}_1 \oplus \mathbf{d}_2$ ($\mathbf{a}_k, \mathbf{d}_k$ are known at the receiver).

Although asymptotically optimal in the high SNR regime, the result in the low SNR regime can be improved using a simple separation scheme. Namely let the receiver decode both codewords and add them up. From the results on the capacity region of the Gaussian MAC, we know that

$$R_{SEP}^s(P) := \frac{1}{4}\log(1 + 2P). \tag{6.5}$$

[2]For a two-user Gaussian MAC, achievable symmetric computation rate R simply means that the computation rate pair (R, R) is achievable.

is an achievable symmetric computation rate. Comparing to the upper bound $\frac{1}{2}\log(1+P)$, this achievable computation rate is good at low SNR but suboptimal at high SNR regime. We can further improve the rate by time-sharing the two above schemes.

Proposition 6.1 (Time-sharing)**.** *For any $P_1, P_2 \geq 0, \beta \in [0,1], \alpha \in [0,1]$ such that $\beta P_1 + (1-\beta)P_2 = P$, an achievable symmetric computation rate for the Gaussian MAC in (6.1) is*

$$R^s_{TS}(P) := \beta\alpha R^s_{CF}(P_1/\alpha) + (1-\beta)R^s_{SEP}(P_2) \tag{6.6}$$

where R^s_{CF}, R^s_{SEP} are defined in (6.3) and (6.5).

In the following we show that with nested linear codes, we can give an alternative codes construction for the compute-and-forward scheme. This construction recovers the original compute-and-forward result and more importantly, it improves upon the best known results. The encoding and decoding procedure with nested linear codes for the two-user Gaussian MAC are given as follows.

- Encoding: User k selects a conditional probability distribution $p_{X|U} : \mathbb{F}_q \to \mathbb{R}$. Given the codewords $\mathbf{u}_k$, it generates the channel input $\mathbf{x}_k$ element-wise according to $p_{X|U}(\mathbf{x}_{k,i}|\mathbf{u}_{k,i})$ where $\mathbf{u}_{k,i}$ denotes the i-th entry of $\mathbf{u}_k$ for $i = 1, \ldots, n$.
- Decoding: Define $\mathbf{u}(\mathbf{w}_s, \mathbf{a}) := \mathbf{H}\mathbf{w}_s \oplus \mathbf{G}\mathbf{a} \oplus \mathbf{d}_1 \oplus \mathbf{d}_2$. Given the channel output $\mathbf{y}$, the decoder finds a unique $\hat{\mathbf{w}}_s$ such that

  $$(\mathbf{y}, \mathbf{u}(\hat{\mathbf{w}}_s, \mathbf{a})) \in \mathcal{A}^{(n)}_{[Y, U_1 \oplus U_2]}$$

  for some $\mathbf{a} \in \mathbb{F}_q^h$. The estimated sum codeword is then formed as $\hat{\mathbf{s}} = \mathbf{u}(\hat{\mathbf{w}}_s, \mathbf{a})$. [3]

With the procedure above we can show the following result.

Theorem 6.1. *Consider the 2-user MAC in (6.1). Let random variables U_1, U_2 have the same distribution p_U over the finite field $\mathbb{F}_q$ with a prime q. The symmetric computation rate*

$$R^s_{NL}(P) := I(U_1 \oplus U_2; Y) - (H(U_1 \oplus U_2) - H(U_1)) \tag{6.7}$$

is achievable where $X_k \in \mathbb{R}$ is generated through a conditional probability distribution $p_{X|U}$ satisfying $E\,||X_k||^2 \leq P$.

Proof. The main idea of the construction can be found in [60, Thm. 1], which deals with a joint source-channel coding problem. A proof of this theorem can be deduced from [61] for the case when Y is a discrete random variable. Using a quantization argument on Y as in [6, Ch. 3], it is straightforward to extend the result for the Gaussian case with continuous output alphabet. □

[3] Here Y is understood to be a discrete random variable such that the typical sets are well-defined. Using a quantization argument on Y [6, Ch. 3.4], this construction can be extended to the Gaussian case when Y is continuous.

For simplicity of presentation, we will represent the elements in the finite field $\mathbb{F}_q$ using the set [4]

$$\mathcal{U} := \{-(q-1)/2, \ldots, (q-1)/2\} \tag{6.8}$$

The sum of two elements is given by $U_1 \oplus U_2 := (U_1 + U_2) \bmod q$, i.e. the usual modular arithmetic for integers. We also define $\mathcal{U}^+ := \{1, \ldots, (q-1)/2\}$ and $\mathcal{U}^- := \{-(q-1)/2, \ldots, -1\}$.

The achievable computation rate given in Theorem 6.1 depends on the conditional distribution $p_{X|U}$ which we have the freedom to choose according to the channel in consideration. For the Gaussian MAC, we study a simple (deterministic) function which takes the form

$$X_k = U_k \cdot \Delta \text{ for } k = 1, 2 \tag{6.9}$$

with some real number $\Delta > 0$ satisfying the power constraint

$$\sum_{u=-(q-1)/2}^{(q-1)/2} p_U(u)(\Delta u)^2 = P. \tag{6.10}$$

Given the distribution of U_1, U_2, we need the distribution of $U_1 \oplus U_2$ and the equivalent channel from $U_1 \oplus U_2$ to Y in order to evaluate the expression in Theorem 6.1.

Proposition 6.2. *Assume U_1, U_2 have the distribution p_U over the finite field $\mathbb{F}_q$ represented using the set $\mathcal{U}$ in (6.8). Define $S := U_1 \oplus U_2$ and*

$$\begin{aligned}
A(s) &:= \sum_{i=-(q-1)/2}^{-(q+1)/2+s} p_U(i)p_U(s-i-q) \\
B(s) &:= \sum_{i=-(q-1)/2+s}^{(q-1)/2} p_U(i)p_U(s-i) \\
D(s) &:= \sum_{i=-(q-1)/2}^{(q-1)/2+s} p_U(i)p_U(s-i) \\
E(s) &:= \sum_{i=(q+1)/2+s}^{(q-1)/2} p_U(i)p_U(s-i+q).
\end{aligned}$$

The distribution of S is given by

$$p_S(s) = A(s) + B(s)$$

for $s \in \mathcal{U}^+ \cup \{0\}$ and

$$p_S(s) = D(s) + E(s)$$

[4] This choice of $\mathcal{U}$ is feasible for a prime number $q \geq 3$. For $q = 2$ we can choose $\mathcal{U} := \{0, 1\}$ and the results in this paper can be adapted accordingly.

for $s \in \mathcal{U}^-$. If X_k is generated as in (6.9), the conditional density function $f_{Y|S}$ is given by

$$f_{Y|S}(y|s) = \frac{A(s)}{p_S(s)}\mathcal{N}(y;\Delta(s-q),1) + \frac{B(s)}{p_S(s)}\mathcal{N}(y;\Delta s,1)$$

for $s \in \mathcal{U}^+$,

$$f_{Y|S}(y|s) = \frac{D(s)}{p_S(s)}\mathcal{N}(y;\Delta s,1) + \frac{E(s)}{p_S(s)}\mathcal{N}(y;\Delta(s+q),1)$$

for $s \in \mathcal{U}^-$ and

$$f_{Y|S}(y|0) = \mathcal{N}(y;0,1)$$

where $\mathcal{N}(y;m,\sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-m)^2/(2\sigma^2)}$.

The proof is straightforward but tedious hence omitted. In fact the distribution p_S is the circular convolution of p_U with period q. It is easy to show that if p_U is symmetric, i.e., $p_U(u) = p_U(-u)$ for $u \in \mathcal{U}$, p_S is also symmetric. The achievable computation rate in Theorem 6.1 can be readily evaluated for any given distribution p_U. We give a few examples in the sequel.

Example 1 (Uniform distribution.) We assign a uniform distribution to U_1, U_2, i.e., $p_U(u) = 1/q$ for all $u \in \mathcal{U}$. It is easy to see that S is also uniformly distributed in $\mathcal{U}$. We can find $f_{Y|S}$ using Proposition 6.2 and evaluate the achievable rates using Theorem 6.1. Figure 6.1 shows the achievable rates with different choices of q. Notice that in this case $H(U_1 \oplus U_2) = H(U_1) = \log q$ hence R^s_{NL} is always positive. In high SNR regime, we can show that the rate only scales as $\frac{1}{2}\log\frac{6P}{\pi e}$ due to the *shaping loss*.

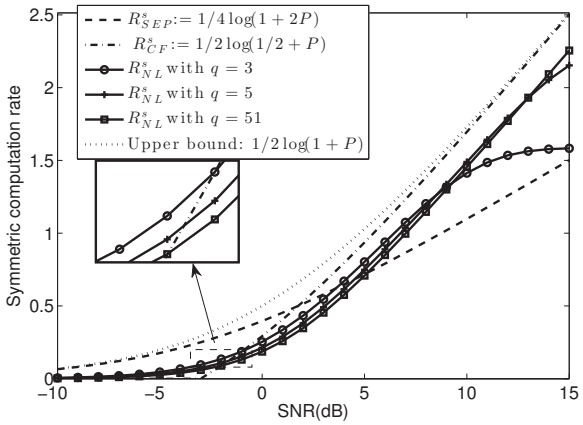


Figure 6.1 – Achievable computation rate R_{NL} with uniform input distribution and different q. It is interesting to notice that for the low SNR regime, the uniform distribution with a smaller q results in a better rate than a larger q.

Example 2 (Discretized Gaussian distribution.) In this example we show that with a proper choice of the distribution on U, the symmetric computation rate

$\log(1/2+P)$ with compute-and-forward can be recovered using Theorem 6.1. Given a prime number q and $A>0$, we consider the following distribution on $\mathcal{U}$

$$p_U(u) = \frac{1}{\alpha_{(q-1)/2}} e^{-(\Delta u)^2/2A} \tag{6.11}$$

with

$$\alpha_{(q-1)/2} := \sum_{u=-(q-1)/2}^{(q-1)/2} e^{-(\Delta u)^2/2A}$$

and Δ is chosen such that (6.10) is satisfied. In this example we will only focus on the limits

$$q \to \infty, \Delta \to 0 \text{ and } q\Delta^2 \to \infty \tag{6.12}$$

with which p_U approaches a Gaussian distribution.

Proposition 6.3 (Discretized Gaussian)**.** *Consider the 2-user Gaussian MAC in (6.1). Let p_U be the distribution given in (6.11) and choose $A=P$. In the limits of (6.12), we have the achievable symmetric computation rate*

$$R_{NL}^s = \frac{1}{2}\log(1/2+P) \tag{6.13}$$

where R_{NL}^s is given in (6.7).

Proof. In this proof we use natural logarithm for simplicity. Choosing p_U in (6.11), the entropy of U_1 is calculated to be

$$H(U_1) = \log \alpha_{(q-1)/2} + \frac{1}{2} \tag{6.14}$$

We set $A=P$ and use the lower bound on $\alpha_{(q-1)/2}$ in Lemma 6.1 in the Appendix to obtain:

$$H(U_1) > \log(\sqrt{2\pi P} - (1+\epsilon)\Delta) - \log\Delta + 1/2 \tag{6.15}$$

where $\epsilon \to 0$ in the limits (6.12). In Lemma 6.2 we show that the distribution p_S of $S := U_1 \oplus U_2$ approaches a discretized Gaussian distribution with power $2P$, i.e.

$$p_S(s) \longrightarrow \frac{\Delta}{\sqrt{4\pi P}} e^{-\frac{(\Delta s)^2}{4P}} \tag{6.16}$$

hence we have [25, Ch. 8]

$$H(S) \longrightarrow \frac{1}{2}\log(4\pi e P) - \log\Delta \tag{6.17}$$

It is also shown in Lemma 6.2 that the channel $f_{Y|S}$ approaches a point-to-point Gaussian channel in the limits (6.12)

$$f_{Y|S}(y|s) \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-(y-s\Delta)^2/2} \tag{6.18}$$

hence we have [6, Ch. 3]

$$I(Y;S) \longrightarrow \frac{1}{2}\log(1+2P) \tag{6.19}$$

This is expected because the distribution p_S is a circular convolution of p_U, and in the limit (6.12) the circular convolution approaches a usual convolution because the support size of U tends to infinity and the convolution of two Gaussian distributions is Gaussian. Finally we have our achievable computation rate

$$\begin{aligned} R &= I(Y;S) - H(S) + H(U) \\ &> I(Y;S) - H(S) + \log(\sqrt{2\pi Pe} - (1+\epsilon)\Delta\sqrt{e}) - \log\Delta \\ &\longrightarrow \frac{1}{2}\log(1/2+P) \end{aligned}$$

in the limit (6.12). □

Example 3 (Achievable rates with optimized distributions.) In this example we show that new achievable rates can be obtained with good input distributions. They are in general better than R^s_{CF} in (6.4) and are better than R^s_{SEP} in (6.5) when SNR exceeds a certain value. For example choosing $q = 3$ and $\mathcal{U} = \{-1, 0, 1\}$ gives

$$p_U(0) = p_0 \tag{6.20a}$$
$$p_U(1) = p_U(-1) = (1-p_0)/2 := p_1. \tag{6.20b}$$

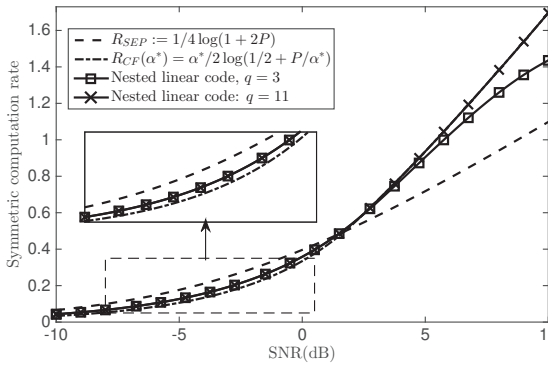
To satisfy the power constraint, the constant Δ is chosen to be $\Delta = \sqrt{P/(1-p_0)}$ and X_k takes values in the set $\{-\Delta, 0, \Delta\}$. Using Proposition 6.2, it is easy to calculate the distribution on $S := U_1 \oplus U_2$

$$\begin{aligned} p_S(0) &= p_0^2 + 2p_1^2 \\ p_S(1) &= p_S(-1) = 2p_0p_1 + p_1^2 \end{aligned}$$

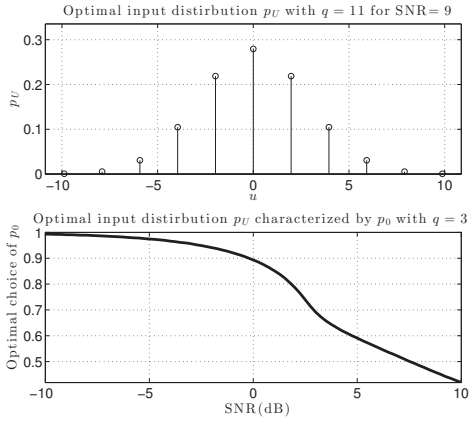
and density function for the equivalent channel from S to Y

$$\begin{aligned} f_{Y|S}(y|0) &= \mathcal{N}(y;0,1) \\ f_{Y|S}(y|1) &= \frac{p_1^2}{p_S(1)}\mathcal{N}(y;-2\Delta,1) + \frac{2p_0p_1}{P_S(1)}\mathcal{N}(y;\Delta,1) \\ f_{Y|S}(y|-1) &= \frac{p_1^2}{p_S(-1)}\mathcal{N}(y;2\Delta,1) + \frac{2p_0p_1}{P_S(-1)}\mathcal{N}(y;-\Delta,1) \end{aligned}$$

This can be extended directly to other values of q. To evaluate the achievable rate, a procedure based on the classical Blahut-Arimoto algorithm is developed in [62] to find the optimal distribution p_U which maximizes R^s_{NL}. Figure 6.2a shows that in low SNR regime, the nested linear codes with even a small value of q can outperform the compute-and-forward scheme in (6.4), which, according to Proposition 6.3, is equivalent to choosing a Gaussian distribution for nested linear codes. This in particular implies that a (discretized) Gaussian distribution is in general *suboptimal* for the computation problem with nested linear codes. The choice of power $P = 1.5$ is interesting with which the two known schemes give the



(a) Achievable computation rate with small constellations and optimized probability distribution. The achievable rates $q = 11$ is very close to the rates with $q = 3$ in low SNR while larger than the latter in high SNR.



(b) Examples of optimal distributions.

Figure 6.2 – The left plot gives achievable symmetric computation rates R^s_{NL} using nested linear codes with constellation size $q = 3$ and $q = 11$. They are better than the compute-and-forward rate R^s_{CF} in low SNR regime as shown in the zoomed-in plot. As SNR increases, R^s_{NL} can be at least as good as R^s_{CF} by choosing a large enough q and an optimized input distribution. In this plot R^s_{CF} almost coincides with R^s_{NL} using $q = 11$ for relatively high SNR. As an example, the upper plot on the right shows the optimal input distribution p_U which maximizes R^s_{NL} in (6.7) with $q = 11$ for SNR= 9. The input distribution with $q = 3$ can be characterized by a number p_0 as in (6.20). The lower plot on the right shows the optimal choice of p_0 for different SNR.

same computation rate $R^s_{CF} = R^s_{SEP} = 0.5$ bit and the optimized compute-and-forward gives $R^s_{CF}(\alpha^*) \approx 0.5020$ bit. The linear nested code gives a rate about 0.5112 bit with $q = 3$ and a rate about 0.5120 bit with $q = 11$ under the simple channel input mapping (6.9).

We do not have a complete characterization of the optimal input distribution. In the limit when P approaches zero, we have the following observation.

Proposition 6.4. *In the limit $P \to 0$, the optimal distribution p_U with the channel input mapping (6.9) which maximizes R_{NL} in (6.7) approaches a Delta function, i.e., $p_U(0) = 1 - \sigma$ where $\sigma \to 0$ as $P \to 0$.*

Proof sketch. First observe that as $P \to 0$, we have $I(U_1 \oplus U_2; Y) \to 0$ hence the optimal distribution should satisfy the property $H(U_1 \oplus U_2) - H(U_1) \to 0$. However this is only possible if p_U either approaches a uniform distribution (or p_U is a uniform distribution) or approaches a Delta function with all its mass on $u = 0$. We show that the uniform distribution cannot be optimal. Starting with a uniform distribution $p_U(u) = 1/q$ for all $u \in \mathcal{U}$, we consider the perturbation $p_U(0) = 1/q + 2\delta$, $p_U((q-1)/2) = p_U(-(q-1)/2) = 1/q - \delta$ with small $\delta > 0$. Let $R^s_{NL}(P, \delta)$ denote the achievable computation rate in (6.7) with the power P and a uniform input

distribution with perturbation δ, we have the approximation

$$R^s_{NL}(P,\delta) \approx R^s_{NL}(0,0) + P\frac{\partial R^s_{NL}}{\partial P}(0,0) + \delta\frac{\partial R^s_{NL}}{\partial \delta}(0,0)$$

for small P and δ. We can show that $\frac{\partial R^s_{NL}}{\partial \delta}(0,0)$ is strictly positive, hence a perturbation to the uniform distribution increases the achievable rates in the limit. □

Notice that the Figure 6.2b agrees with the above observation. As SNR decreases, the optimal value p_0 approaches 1. Equivalently the optimal distribution p_U approaches the Delta function.

The above result has immediate application on the Gaussian TWRC. In the symmetric setting when two transmitters have power P and the relay has power P_R, the best known symmetric rate is

$$\min\left\{R^s_{TS}(P), \frac{1}{2}\log(1+P_R)\right\}$$

with $R^s_{TS}(P)$ defined in (6.6). Theorem 6.1 shows the possibility of increasing the first term in the min expression. Namely we can achieve the symmetric rate

$$\min\left\{\tilde{R}^s_{TS}(P), \frac{1}{2}\log(1+P_R)\right\}$$

where $\tilde{R}^s_{TS}(P) := \beta\alpha R^s_{NL}(P_1/\alpha) + (1-\beta)R^s_{SEP}(P_2)$ for any $\alpha, \beta \in [0,1], P_1, P_2 \geq 0$ satisfying $\beta P_1 + (1-\beta)P_2 = P$. Since we can always ensure $R^s_{NL} > R^s_{CF}$ by choosing the optimal input distribution p_U, we will obtain a higher rate $\tilde{R}^s_{TS}(P)$ than $R^s_{TS}(P)$. However the improvement is minor.

6.2 Appendix

We study the discrete random variable U given in (6.11). Natural logarithm is used in the derivation for simplicity. Recall that the discrete random variable U taken integer values in the set $\mathcal{U} := \{-(q-1)/2, \ldots, (q-1)/2\}$ with a prime number q. Let A be some given positive real number, the probability distribution p_U on U depends on three parameters q, A, P and is defined as

$$P_U(U=i) = \frac{1}{\alpha}e^{-(\Delta i)^2/2A} \tag{6.21}$$

with

$$\alpha_{(q-1)/2} := \sum_{i\in\mathcal{U}} e^{-(\Delta i)^2/2A} \tag{6.22}$$

and Δ is chosen such that we have

$$\sum_{i\in\mathcal{U}} (\Delta i)^2 P_U(i) = P \tag{6.23}$$

for some given positive number P.

This probability distribution can be viewed as a discretized Gaussian rv and some special choices of A could be $A = 1$ or $A = P$. We are interested in the entropy of U and how it behaves in the limiting cases when q is large and P is small. A direct calculation shows that

Lemma 6.1 (Bounds on α)**.** *Let $M, A > 0$ and define*

$$\alpha_{M,A} := \sum_{i=-M}^{M} e^{-(\Delta i)^2/2A}.$$

We have the bounds

$$\max\{1, \frac{\sqrt{2\pi A}}{\Delta} - 1 - \epsilon'_{M,\Delta}\} < \alpha_{M,A} < 1 + \frac{\sqrt{2\pi A}}{\Delta}$$

where $\epsilon'_{M,\Delta} > 0$ depends on M, Δ in the way

$$\epsilon'_{M,\Delta} \to 0 \text{ as } M \to \infty \text{ and } M\Delta^2 \to \infty.$$

Proof. Let $S_{M,A} := \sum_{i=1}^{M} \frac{1}{\sqrt{2\pi A}} e^{-\frac{(\Delta i)^2}{2A}} \Delta$, we rewrite

$$\alpha_{M,A} = 1 + \frac{2\sqrt{2\pi A}}{\Delta} S_{M,A}$$

The bound $\alpha_{M,A} > 1$ is obvious. Let $f_A(x) := \frac{1}{\sqrt{2\pi A}} e^{-\frac{x^2}{2A}}$. Then $S_{M,A}$ is the (right) Riemann sum of $f_A(x)$ in the interval $[0, M\Delta]$. Hence we have

$$\begin{aligned} S_{M,A} &> \int_0^{M\Delta} f_A(x)dx - \Delta(f_A(0) - f_A(M\Delta)) \\ &= \frac{1}{2} - Q(\frac{M\Delta}{\sqrt{A}}) - \frac{\Delta}{\sqrt{2\pi A}}(1 - e^{-\frac{(M\Delta)^2}{2A}}) \end{aligned}$$

Using the bound on the Q-function $Q(x) < \frac{1}{x}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ we have

$$S_{M,A} > \frac{1}{2} - \frac{\Delta}{\sqrt{2\pi A}} - \frac{\sqrt{A}}{M\Delta\sqrt{2\pi}} e^{-\frac{M^2\Delta^2}{2A}} + \frac{\Delta}{\sqrt{2\pi A}} e^{-\frac{(M\Delta)^2}{2A}}$$

and

$$\alpha_{M,A} > \frac{\sqrt{2\pi A}}{\Delta} - 1 + (2 - \frac{2A}{M\Delta^2}) e^{-\frac{M^2\Delta^2}{2A}} \tag{6.24}$$

The lower bound follows in the limit $M\Delta^2 \to \infty$. Similarly we have

$$S_{M,A} < \int_0^{M\Delta} f_A(x)dx = \frac{1}{2} - Q(\frac{M\Delta}{\sqrt{A}}) \tag{6.25}$$

Invoking the lower bound $Q(x) > \frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ we have

$$S_{M,A} < \frac{1}{2} - \frac{M\Delta\sqrt{A}}{\sqrt{2\pi}(A + M^2\Delta^2)} e^{-\frac{M^2\Delta^2}{2A}} \tag{6.26}$$

and hence

$$\alpha_{M,A} < 1 + \frac{\sqrt{2\pi A}}{\Delta} - \frac{2MA}{A + M^2\Delta^2} e^{-\frac{M^2\Delta^2}{2A}} \tag{6.27}$$

The upper bound follows directly. □

Lemma 6.2 (Distribution of the sum and the channel)**.** *Let U_1, U_2 have the probability distribution p_U in (6.11) and $S := U_1 \oplus U_2$. In the limit (6.12), the distribution of S is*

$$p_S(s) = \frac{\Delta}{\sqrt{2\pi A}} e^{-\frac{\Delta^2 s^2}{4A}} + o(\Delta) \tag{6.28}$$

and the equivalent channel $f_{Y|S}$ in Proposition 6.2 is

$$f_{Y|S}(y|s) = \frac{B(s)}{B(s) + o(\Delta)} \frac{1}{2\pi} e^{-(y-\Delta s)^2/2} + o(\Delta) \tag{6.29}$$

Proof sketch. Due to the symmetry of p_S we only need to consider the case $s \in \{0\} \cup \mathcal{U}^+$. Choosing p_U in (6.11), $A(s)$ and $B(s)$ defined in Proposition 6.2 can be rewritten as

$$A(s) = \frac{\alpha_{(s-1)/2,A/2}}{\alpha^2_{(q-1)/2,A}} e^{-\frac{\Delta^2 (s-q)^2}{4A}}$$
$$B(s) = \frac{\alpha_{(q-s-1)/2,A/2}}{\alpha^2_{(q-1)/2,A}} e^{-\frac{\Delta^2 s^2}{4A}}$$

For $s \in \{0\} \cup \mathcal{U}^+$, we can use Lemma 6.1 to show

$$A(s) < (1 + \frac{\sqrt{\pi A}}{\Delta}) e^{-\frac{\Delta^2 (-q/2)^2}{4A}} \tag{6.30}$$

hence $A(s) = o(\Delta)$. Implied by Lemma 6.1, we can write $\alpha_{M,A}$ as $\alpha_{M,A} = \frac{\sqrt{2\pi A}}{\Delta} + a$ for some a with $|a| \leq 2$ in the limit (6.12). With some a_1, a_2 with $|a_1|, |a_2| \leq 2$ and the Taylor expansion we can show

$$\frac{\alpha_{(q-s-1)/2,A/2}}{\alpha^2_{(q-1)/2,A}} = \frac{\sqrt{\pi A}/\Delta + a_1}{(\sqrt{2\pi A}/\Delta + a_2)^2} = \frac{\Delta}{\sqrt{4\pi A}} + o(\Delta)$$

It follows that

$$\begin{aligned} p_S(s) &= A(s) + B(s) \\ &= o(\Delta) + (\frac{\Delta}{\sqrt{4\pi A}} + o(\Delta)) e^{-\frac{\Delta^2 s^2}{4A}} \end{aligned}$$

For the equivalent channel $f_{Y|S}$ given in Proposition 6.2, we can bound the ratio

$$\begin{aligned} \frac{A(s)}{A(s) + B(s)} < \frac{A(s)}{B(s)} &< \frac{(1 + \frac{\sqrt{\pi A}}{\Delta}) e^{-\frac{\Delta^2 (s-q)^2}{4A}}}{B(s)} \\ &< \frac{(\Delta + \sqrt{\pi A})\sqrt{4\pi A}}{\Delta^2} e^{-\frac{\Delta^2}{4A}(q^2 - 2qs)} \\ &\leq \frac{(\Delta + \sqrt{\pi A})\sqrt{4\pi A}}{\Delta^2} e^{-\frac{\Delta^2 q}{4A}} = o(\Delta) \end{aligned}$$

Hence

$$f_{Y|S}(y|s) = o(\Delta)\mathcal{N}(y; \Delta(s-q), 1) + \frac{B(s)}{o(\Delta) + B(s)} \mathcal{N}(y; \Delta s, 1)$$

which proves the claim. □

7 Typical Sumsets of Linear Codes

In previous chapters we have studied the problem of computing the sum of codewords via the Gaussian MAC with nested lattice codes and nested linear codes.[1] With nested lattice codes, two codewords in $\mathbb{R}^n$ are added as real-valued vectors by the Gaussian MAC directly. With nested linear codes, after lifting linear codes from the finite field to $\mathbb{R}^n$, the channel also adds two vectors in $\mathbb{F}_q^n$ as real-valued vectors instead of in a finite field. This motivates our question: what does the sum of two codebooks look like?

To put our study in perspective, it is worth pointing out that our problem is closely connected to *sumset theory*, which studies the size of the set $\mathcal{A}+\mathcal{B} := \{a+b : a \in \mathcal{A}, b \in \mathcal{B}\}$ where $\mathcal{A}, \mathcal{B}$ are two finite sets taking values in some additive group. One objective of the sumset theory is to use *sumset inequalities* to relate the cardinality of sets $|\mathcal{A}|, |\mathcal{B}|$ and $|\mathcal{A}+\mathcal{B}|$. As a simple example, for $\mathcal{A} = \{0, 1, 2, 3, 4\}$ with 5 elements we have $|\mathcal{A}+\mathcal{A}| = 9$ elements. But if let $\mathcal{A}' = \{0, 0.2, 0.8, 1.1, 2.1\}$ with 5 elements we have $|\mathcal{A}'+\mathcal{A}'| = 15$ elements. This shows that the sumset size $|\mathcal{A}+\mathcal{B}|$ depends heavily on structures of the sets. As a rule of thumb, the sumset size will be small if and only if the individual sets are "structured". Some classical results of sumset theory and inverse sumset theory can be found in, e.g. [63].

Our problem is slightly different: given two linear codes of two users, we independently pick one codeword from each codebook uniformly at random and add them as integer vectors. We would like to know how large is the sumset? This problem concerns with *sums of random variables* defined over a certain set, hence can be viewed as a sumset problem in a probabilistic setting. It shares similarity with the classical sumset problem while has its own feature. We first point out the main difference between the two problems. Given a set of integers $\mathcal{U} = \{0, 1, \ldots, q-1\}$, the sumset $\mathcal{U}+\mathcal{U}$ contains $2q-1$ elements. Now let U_1, U_2 be two independent random variables uniformly distributed in the set $\mathcal{U}$, a natural connection between the size of the set $\mathcal{U}$ and the random variables U_1, U_2 is that $H(U_1) = H(U_2) = \log|\mathcal{U}|$, i.e., the entropy of the random variable is equal to the logarithmic size of $\mathcal{U}$. On the

[1] The material of this chapter has appeared in
J. Zhu and M. Gastpar, "Typical sumsets of linear codes", in *arXiv: 1511.08435*, Nov, 2015.

other hand, although the sum variable $W := U_1 + U_2$ takes *all* possible values in $\mathcal{U}+\mathcal{U}$, it is "smaller" than $\log|\mathcal{U}+\mathcal{U}|$ because the distribution of W is non-uniform over $\mathcal{U}+\mathcal{U}$. Indeed we have $H(W) < \log|\mathcal{U}+\mathcal{U}|$ in this case but the difference between $H(W)$ and $\log|\mathcal{U}+\mathcal{U}|$ is small. However this phenomenon is much more pronounced in high dimensional spaces as we shall see later in this paper. Nevertheless, it is also important to realize that in the probabilistic setting, the structure of the random variable still has decisive impact on the sumset "size", which can be partially characterized by the entropy of the sum variable. Using the examples in the preceding paragraph, if the identical independent random variables U_1, U_2 are uniformly distributed in $\mathcal{A}$, we have $H(U_1+U_2) \approx 2.99$ bit while if U_1', U_2' uniformly distributed in $\mathcal{A}'$, it gives $H(U_1'+U_2') \approx 3.84$ bit.

7.1 Typical Sumsets of Linear Codes

In this section we formally define and study typical sumsets of linear codes. We use $[a:b]$ to denote the set of integers $\{a, a+1, \ldots, b-1, b\}$ and define two sets $\mathcal{U} := [0:q-1]$ and $\mathcal{W} := [0:2q-2]$. We also define P_U to be the uniform probability distribution over the set $\mathcal{U}$ i.e.,

$$P_U(a) = 1/q \text{ for all } a \in \mathcal{U}. \tag{7.1}$$

If U_1, U_2 are two independent random variables with distribution P_U, the sum $W := U_1 + U_2$ is a random variable distributed over the set $\mathcal{W}$. Let P_W denote the probability distribution of this random variable. A direct calculation shows that

$$P_W(a) = \begin{cases} \frac{a+1}{q^2} & a \in [0:q-1] \\ \frac{2q-1-a}{q^2} & a \in [q:2q-2] \end{cases} \tag{7.2}$$

and the entropy of W is given as

$$H(W) = 2\log q - \frac{1}{q^2}(2\sum_{i=1}^{q} i\log i - q\log q). \tag{7.3}$$

Recall the definition of typical sequences in Chapter 1 and the standard results regarding the typical sequences.

Lemma 7.1 (Typical sequences [14])**.** *Let U^n be a n-length random vector with each entry i.i.d. according to P_U. Then for every $\delta > 0$ in (1.1), it holds that*

$$\mathbb{P}\left\{U^n \in \mathcal{A}_{[U]}^{(n)}\right\} \geq 1 - 2|\mathcal{U}|e^{-2n\delta^2} \tag{7.4}$$

Furthermore, the size of set of typical sequences is bounded as

$$2^{n(H(U)-\epsilon_n)} \leq |\mathcal{A}_{[U]}^{(n)}| \leq 2^{n(H(U)+\epsilon_n)} \tag{7.5}$$

for some $\epsilon_n \searrow 0$ as $n \to \infty$.

In this chapter, the notations $\mathbf{Ab}$ or $\mathbf{a}^T\mathbf{b}$ are understood as matrix multiplication *modulo* q, or the matrix multiplication over the corresponding finite field. Modulo addition is denoted with $\oplus$ and $+$ means the usual addition over integers.

7.1.1 Problem statement and main results

Given two positive integers k, n satisfying $k < n$, a (n, k) linear code over $\mathbb{F}_q$ is a k-dimensional subspace in $\mathbb{F}_q^n$ where q is a prime number. The *rate* of this code is given by $R := \frac{k}{n} \log q$. Any (n, k) linear code can be constructed as

$$\mathcal{C} = \left\{ \mathbf{t} : \mathbf{t} = \mathbf{G}\mathbf{m}, \text{ for all } \mathbf{m} \in \mathbb{F}_q^k \right\} \tag{7.6}$$

with a *generator matrix* $\mathbf{G} \in \mathbb{F}_q^{n \times k}$. A (n, k) linear code $\mathcal{C}$ over $\mathbb{F}_q$ is called *systematic* if it can be constructed as

$$\mathcal{C} = \left\{ \mathbf{t} : \mathbf{t} = \begin{bmatrix} \mathbf{I}_{k\times k} \\ \mathbf{Q} \end{bmatrix} \mathbf{m}, \text{ for all } \mathbf{m} \in \mathbb{F}_q^k \right\} \tag{7.7}$$

with some $\mathbf{Q} \in \mathbb{F}_q^{(n-k)\times k}$ where $\mathbf{I}_{k\times k}$ is the $k \times k$ identity matrix.

From now on we will view $\mathcal{C}$ as a set of n-length vectors taking values in $\mathcal{U}^n$ where $\mathcal{U} := \{0, \ldots, q-1\}$. The sumset of two linear codes is

$$\mathcal{C} + \mathcal{C} := \{\mathbf{t} + \mathbf{v} : \mathbf{t}, \mathbf{v} \in C\} \tag{7.8}$$

where the sum is performed element-wise between the two n-length vectors as integer addition. Namely each element in $\mathcal{C}+\mathcal{C}$ takes value in $\mathcal{W}^n$ where $\mathcal{W} := \{0, \ldots, 2q-2\}$. When the code $\mathcal{C}$ is systematic, the sumset contains sums of two codewords $\mathbf{t}, \mathbf{v} \in \mathcal{C}$ of the form

$$\mathbf{t} + \mathbf{v} = \begin{pmatrix} \mathbf{m} + \mathbf{n} \\ \mathbf{Q}\mathbf{m} + \mathbf{Q}\mathbf{n} \end{pmatrix} := \begin{pmatrix} \mathbf{s}(\mathbf{m}, \mathbf{n}) \\ \mathbf{p}(\mathbf{m}, \mathbf{n}) \end{pmatrix} \tag{7.9}$$

for some $\mathbf{m}, \mathbf{n} \in \mathcal{U}^k$. We call $\mathbf{s}(\mathbf{m}, \mathbf{n})$ and $\mathbf{p}(\mathbf{m}, \mathbf{n})$ defined above as the *information-sum* and *parity-sum*, respectively. We shall omit their dependence on $\mathbf{m}, \mathbf{n}$ and use $\mathbf{s}, \mathbf{p}$ if the context is clear. For a systematic code, $\mathbf{m}$ and $\mathbf{n}$ can be viewed as two messages taking all possible values in $\mathcal{U}^k$ from two users.

We are interested in the scenario where two independent users are equipped with the same linear code $\mathcal{C}$ and they choose their messages uniformly at random. To model this situation, we use T to denote the random variable taking values in the code $\mathcal{C}$ with uniform distribution, i.e.

$$\mathbb{P}\{T^n = \mathbf{t}\} = q^{-k} \text{ for all } \mathbf{t} \in \mathcal{C} \tag{7.10}$$

Now let T_1^n, T_2^n be two independent copies of T^n, the sum codewords $T_1^n + T_2^n$ is also a random variable taking values in $\mathcal{C} + \mathcal{C}$. There is a natural distribution on $\mathcal{C} + \mathcal{C}$ induced by T_1^n, T_2^n, which is formally defined as follows.

Definition 7.1 (Induced distribution on $\mathcal{C}_1 + \mathcal{C}_2$)**.** *Given a codebook $\mathcal{C}$ and assume T_1^n, T_2^n are two independent random vectors which are uniformly distributed as in (7.10). We use $P_{\mathcal{S}}$ to denote the distribution on $\mathcal{C} + \mathcal{C}$ which is induced from the distribution of T_1^n, T_2^n.*

The object of interest in this chapter is given in the following definition.

Definition 7.2 (Typical sumset)**.** *Let $\mathcal{C}^{(n)}$ be a sequence of linear codes indexed by their dimension. Let T_1^n, T_2^n be two independent random variables uniformly distributed in $\mathcal{C}^{(n)}$ as in (7.10). A sequence of subsets $\mathcal{K}^{(n)} \subseteq \mathcal{C}^{(n)} + \mathcal{C}^{(n)}$ is called typical sumsets of $\mathcal{C}^{(n)}$, if $T_1^n + T_2^n \in \mathcal{K}^{(n)}$ asymptotically almost surely, i.e.,*

$$\mathbb{P}\left\{T_1^n + T_2^n \in \mathcal{K}^{(n)}\right\} \to 1 \quad \text{as } n \to \infty.$$

To make notations easier, we will often drop the dimension n and say $\mathcal{K}$ is a typical sumset of $\mathcal{C}$, with the understanding that a sequence of codes are considered as in Definition 7.2. Clearly the sumset $\mathcal{C} + \mathcal{C}$ is always a typical sumset according to the above definition because all possible $T_1^n + T_2^n$ must fall inside it. However we will show that for almost all linear codes, most sum codewords $T_1^n + T_2^n$ will fall into a subset $\mathcal{K}$ which could be much smaller than $\mathcal{C} + \mathcal{C}$ by taking the probability distribution of T_1^n and T_2^n into account.

Theorem 7.1 (Normal typical sumsets)**.** *Let $\mathcal{C}^{(n)}$ be a sequence of linear codes in the form (7.6) indexed by their dimension. The rate of the code is given by $R = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{C}^{(n)}|$ and let T_1^n, T_2^n be two independent random variables uniformly distributed on $\mathcal{C}^{(n)}$. We assume each entry of the generator matrix $\mathbf{G}$ is independent and identically distributed according to the uniform distribution in $\mathbb{F}_q$. Then a.a.s. there exists a sequence of typical sumsets $\mathcal{K}_N \subseteq \mathcal{C}^{(n)} + \mathcal{C}^{(n)}$ whose size satisfies*

$$|\mathcal{K}_N^{(n)}| \doteq \begin{cases} 2^{2nR} & R \le D(q) \\ 2^{n(R+D(q))} & R > D(q) \end{cases} \tag{7.11}$$

$$D(q) := H(U_1 + U_2) - \log q. \tag{7.12}$$

where U_1, U_2 are independent with distribution P_U. Furthermore for all $\mathbf{w} \in \mathcal{K}_N$

$$P_{\mathcal{S}}(\mathbf{w}) \doteq \begin{cases} 2^{-2nR} & R \le D(q) \\ 2^{-n(R+D(q))} & R > D(q) \end{cases} \tag{7.13}$$

where $P_{\mathcal{S}}$ is the induced distribution defined in Definition 7.1.

Proof. A proof of the theorem is given in Section 7.1.4. In Appendix 7.2.1 we show that $D(q)$ is an increasing function of q and

$$1/2 \le D(q) < \log \sqrt{e} \approx 0.7213 \tag{7.14}$$

where the lower bound holds for $q = 2$ and the upper bound is approached with $q \to \infty$. □

Remark: For any fixed vector $\mathbf{d} \in \mathbb{F}_q^n$ and define $\mathcal{C}'(n) := \mathcal{C}^{(n)} \oplus \mathbf{d} = \{\mathbf{t} \oplus \mathbf{d} | \mathbf{t} \in \mathcal{C}^{(n)}\}$. We can show that the same results hold for $\mathcal{C}^{(n)} + \mathcal{C}'^{(n)}$.

Figure 7.2 provides a generic plot showing the code rate R vs. normalized size $\frac{1}{n} \log |\mathcal{K}_N|$ of the normal typical sumset size. We see there exists a threshold $D(q)$ on the rate R of the code, above or below which the typical sumset K behave differently. First notice that for the low rate regime $R < D(q)$, almost every different codeword pair T_1^n, T_2^n gives a distinct sum codeword, hence the sumset size $|\mathcal{K}_N|$ is essentially

$|\mathcal{C}|^2$ (up to the exponent), corresponding to the part of the linear function in Figure 7.2 with slope 2. This result shows that for almost all codes, the linear structure of the code does not manifest itself in this low rate regime.

For the high rate regime $R \geq D(q)$, due to the linear structure of the code, there are (exponentially) many different codeword pairs T_1^n, T_2^n giving the same sum codeword, and the normal typical sumset size $|\mathcal{K}_N|$ grows only as $2^{nD(q)}|\mathcal{C}|$ where $D(q)$ does not depend on R, corresponding to the part of the affine function in Figure 7.2 with slope 1. In this regime the code $\mathcal{C}$ has a typical sumset which is exponentially smaller than $\mathcal{C}+\mathcal{C}$. The interesting fact is that, on contrary to the low dimension cases, the codewords are *uniformly* distributed in the typical sumset $\mathcal{K}_N$ as shown by (7.13) in Theorem 7.1. This is reminiscent of classical typical sequences with asymptotic equipartition property (AEP), i.e., the typical sumset occurs a.a.s. but is uniformly filled up with a small subset of all possible sequences. Now we can also give a pictorial description of the sum codewords $T_1^n + T_2^n$ in Figure 7.1. Notice that the sum codewords $T_1^n + T_2^n$ are essentially uniformly distributed in the typical sumset $\mathcal{K}_N$ in high dimensional spaces.

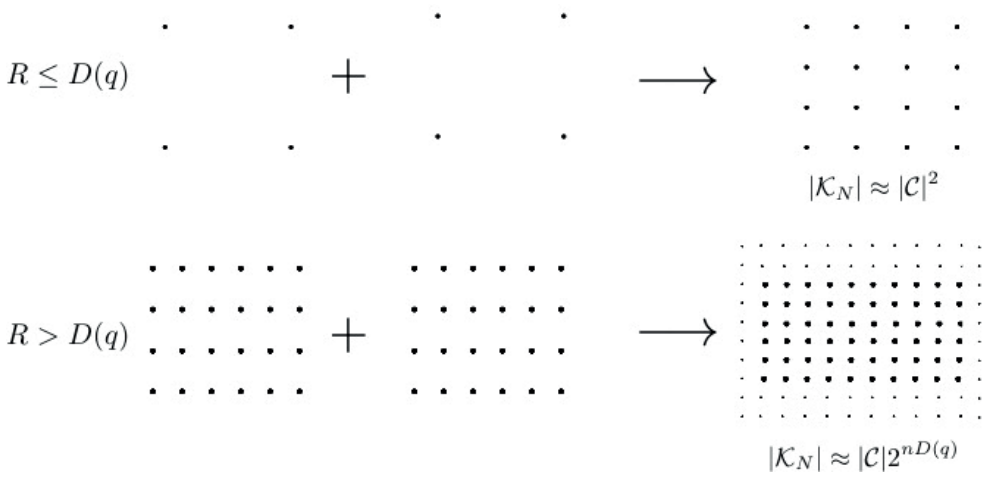


Figure 7.1 – An illustration of the sum codewords $T_1^n + T_2^n$. For rate $R \leq D(q)$, each pair (T_1^n, T_2^n) will give a different sum and typical sumset $\mathcal{K}_N$ is essentially the same as $\mathcal{C} + \mathcal{C}$. The sum codeword is hence uniformly distributed in $\mathcal{C} + \mathcal{C}$. For rate $R > D(q)$, many pairs (T_1^n, T_2^n) give the same sum codeword and the typical sumset $\mathcal{K}_N$ is much smaller than $\mathcal{C} + \mathcal{C}$. Interestingly in the n-dimensional space with $n \to \infty$, the sum codewords $T_1^n + T_2^n$ is basically *uniformly* distributed in the typical sumset $\mathcal{K}_N$ (represented by thick dots in the plot). The other sum codewords in $(\mathcal{C} + \mathcal{C}) \setminus \mathcal{K}_N$ (represented by the small dots) have only negligible probability.

7.1.2 Comparison with $|\mathcal{C} + \mathcal{C}|$

To emphasize the distinction between the classical sumset theory and our study of typical sumsets in probabilistic setting, we compare the size of a normal typical sumset $\mathcal{K}_N$ with the size of the usual sumset $\mathcal{C} + \mathcal{C}$. Before doing this, we first introduce a useful result relating the sumsets of general linear codes with that of systematic linear codes.

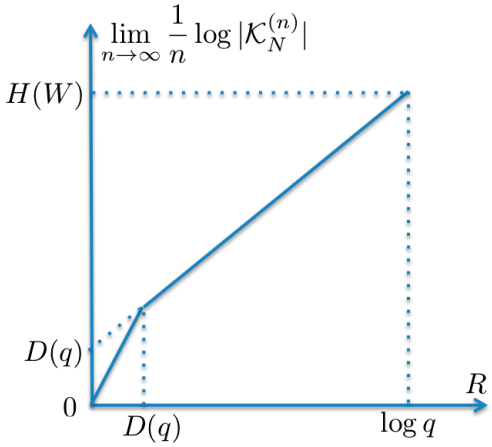


Figure 7.2 – An illustration of the size of normal typical sumsets of linear codes. $H(W)$ and $D(q)$ are given in (7.3) and (7.12), respectively. The piece-wise linear function has slope 2 for low rate regime and slope 1 for medium-to-high rate regime.

Lemma 7.2 (Equivalence between systematic and non-systematic codes)**.** *Given any linear code $\mathcal{C}$, there exists a systematic linear code $\mathcal{C}'$ with a one-to-one mapping $\phi : \mathcal{C} \longrightarrow \mathcal{C}'$ such that for any pair $\mathbf{t}, \mathbf{v} \in \mathcal{C}$ satisfying $\mathbf{t}+\mathbf{v} = \mathbf{s}$, we have $\phi(\mathbf{t})+\phi(\mathbf{v}) = \phi(\mathbf{s})$.*

Proof. Let π denote a permutation over the set $\{1, \ldots, n\}$. A code $\mathcal{C}$ is said to be *equivalent* to another code $\mathcal{C}'$ if every codeword $\mathbf{t}'$ in $\mathcal{C}'$ can be obtained by permuting the coordinates of some codeword in $\mathcal{C}$ using π, i.e.,

$$\mathbf{t}' := (\mathbf{t}'_1, \mathbf{t}'_2, \ldots, \mathbf{t}'_n) = (\mathbf{t}_{\pi(1)}, \mathbf{t}_{\pi(2)}, \ldots, \mathbf{t}_{\pi(n)}) \tag{7.15}$$

for some $\mathbf{t} := (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n) \in \mathcal{C}$. It is known that any linear code $\mathcal{C}$ is equivalent to some systematic linear code (see [64, Ch. 4.3] for example). We define the mapping ϕ to be the permutation needed to transform the given linear code $\mathcal{C}$ to its systematic counterpart $\mathcal{C}'$. Clearly this permutation is a one-to-one mapping.

For two different pairs $(\mathbf{t}, \mathbf{v})$ and $(\tilde{\mathbf{t}}, \tilde{\mathbf{v}})$ from code $\mathcal{C}$ such that $\mathbf{t} + \mathbf{v} = \tilde{\mathbf{t}} + \tilde{\mathbf{v}} = \mathbf{s}$, it holds that

$$\begin{aligned} \phi(\mathbf{t}) + \phi(\mathbf{v}) &= (\mathbf{t}_{\pi(1)} + \mathbf{v}_{\pi(1)}, \mathbf{t}_{\pi(2)} + \mathbf{v}_{\pi(2)}, \ldots, \mathbf{t}_{\pi(n)} + \mathbf{v}_{\pi(n)}) && (7.16)\\ &= (\tilde{\mathbf{t}}_{\pi(1)} + \tilde{\mathbf{v}}_{\pi(1)}, \tilde{\mathbf{t}}_{\pi(2)} + \tilde{\mathbf{v}}_{\pi(2)}, \ldots, \tilde{\mathbf{t}}_{\pi(n)} + \tilde{\mathbf{v}}_{\pi(n)}) && (7.17)\\ &= \phi(\tilde{\mathbf{t}}) + \phi(\tilde{\mathbf{v}}) = \phi(\mathbf{s}) && (7.18) \end{aligned}$$

□

This lemma shows that for any linear code $\mathcal{C}$, there exists a corresponding systematic code $\mathcal{C}'$ whose sumset structure is the same as the former. Now we can show the following simple bounds on the size of the sumset $\mathcal{C} + \mathcal{C}$.

Lemma 7.3 (Simple sumset estimates)**.** *Let $\mathcal{C}$ be a (n, k) linear code over $\mathbb{F}_q$. The size of the sumset $\mathcal{C} + \mathcal{C}$ is upper bounded as*

$$|\mathcal{C} + \mathcal{C}| \leq q^{2k} \tag{7.19}$$

and lower bounded as

$$|\mathcal{C}+\mathcal{C}| \geq (2q-2)^k \tag{7.20}$$

Proof. The upper bound follows simply from the fact that $|\mathcal{C}+\mathcal{C}| \leq |\mathcal{C}|^2$ for any set $\mathcal{C}$. To establish the lower bound, Lemma 7.2 shows that for *any* linear code $\mathcal{C}$, we can find a corresponding systematic linear codes $\mathcal{C}'$ whose sumset size $|\mathcal{C}'+\mathcal{C}'|$ equals to $|\mathcal{C}+\mathcal{C}|$. Then the lower bound holds by noticing that the information-sums $\mathbf{s}$ in (7.9) take all possible values in $\mathcal{W}^k$ with cardianality $(2q-2)^k$. □

Notice that $|\mathcal{K}_N|$ can be smaller than the cheap lower bound given in (7.20) for certain rate range. The intuition for this phenomenon is clear: some of the sum codewords $T_1^n+T_2^n$ occurs very rarely if T_1^n and T_2^n are chosen uniformly. Those sum codewords will be counted in the sumset $\mathcal{C}+\mathcal{C}$ but are probabilistically negligible. Particularly in the case $R > D(q)$, $|\mathcal{K}_N|$ can be exponentially smaller than $|\mathcal{C}+\mathcal{C}|$. For a comparison, we see the lower bound in (7.20) states that

$$|\mathcal{C}+\mathcal{C}| \geq 2^{nR\log(2q-2)/\log q}. \tag{7.21}$$

Then Eq. (7.11) implies that $|\mathcal{K}_N|$ is smaller than $|\mathcal{C}+\mathcal{C}|$ for the rate range

$$R > \frac{D(q)}{\log(2q-2)/\log q - 1}, \tag{7.22}$$

(Notice that the RHS is always larger than $D(q)$ for $q \geq 2$ but is only meaningful if it is smaller than $\log q$). For example $|\mathcal{K}_N|$ is smaller than the lower bound in (7.20) for $R > 2.85$ bits with $q = 11$ and for $R > 4.87$ bits for $q = 101$.

7.1.3 Entropy of sumsets

Often we are interested in inequalities involving entropy of a random variables X and entropy of the sum of two i.i.d. random variables X_1+X_2. One classical result is the *entropy power inequality* involving differential entropy. There are recent results on entropy sumset inequalities which relate the entropy $H(X)$ of some random variable X with the entropy of the sum $H(X+X)$, see [65] [66] for example. If a code $\mathcal{C}$ has a normal typical sumset and let T^n be a random variable uniformly distributed in $\mathcal{C}$, we are able to relate $H(T^n)$ to $H(T_1^n+T_2^n)$ directly where T_1^n, T_2^n are two independent copies of T^n.

Theorem 7.2 (Entropy of sumsets)**.** *Let $\mathcal{C}^{(n)}$ be a sequence of linear codes with normal typical sumsets $\mathcal{K}_N$ as in Theorem 7.1. Let T^n be a random n-length vector uniformly distributed in the code $\mathcal{C}^{(n)}$ and T_1^n, T_2^n two independent copies of T^n. In the limit $n \to \infty$ we have*

$$\lim_{n\to\infty} H(T_1^n+T_2^n)/n = \begin{cases} 2H(T^n)/n = 2R & \text{if } R \leq D(q) \\ H(T^n)/n + D(q) = R + D(q) & \text{if } R > D(q) \end{cases} \tag{7.23}$$

where as before, $D(q) := H(W) - \log q$ with W distributed according to P_W in (7.2).

Proof. As T^n is uniformly distributed in the (n,k) linear code $\mathcal{C}$ with rate R, we have $H(T^n) = nR$. Recall that $P_{\mathcal{S}}$ denote the distribution on $\mathcal{C}+\mathcal{C}$ induced by T_1^n, T_2^n, we have

$$H(T_1^n + T_2^n) = -\sum_{\mathbf{w}\in C+C} P_{\mathcal{S}}(\mathbf{w}) \log P_{\mathcal{S}}(\mathbf{w}) \tag{7.24}$$

$$\geq -\sum_{\mathbf{w}\in \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \log P_{\mathcal{S}}(\mathbf{w}) \tag{7.25}$$

As Theorem 7.1 shows that for $\mathbf{w} \in \mathcal{K}_N$ it holds that $P_{\mathcal{S}}(\mathbf{w}) \leq 2^{-2n(R-\epsilon_n)}$ for $R \leq D(q)$, hence

$$H(T_1^n + T_2^n) \geq -\log 2^{-2n(R-\epsilon_n)} \sum_{\mathbf{w}\in K} P_{\mathcal{S}}(\mathbf{w}) \tag{7.26}$$

$$= 2n(R-\epsilon_n)(1-\delta_n) \tag{7.27}$$

with $\delta_n \to 0$ because $\mathcal{K}_N$ is a typical sumset. It follows that

$$\lim_{n\to\infty} H(T_1^n + T_2^n)/n \geq \lim_{n\to\infty} 2(R-\epsilon_n)(1-\delta_n) \tag{7.28}$$

$$= 2R = 2H(T)/n \tag{7.29}$$

as both $\delta_n, \epsilon_n \to 0$.

On the other hand, we have

$$H(T_1^n + T_2^n) = -\sum_{\mathbf{w}\in \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \log P_{\mathcal{S}}(\mathbf{w}) - \sum_{\mathbf{w}\notin \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \log P_{2C}(\mathbf{w}) \tag{7.30}$$

For $\mathbf{w} \in \mathcal{K}_N$ it holds $P_{\mathcal{S}}(\mathbf{w}) \geq 2^{-2n(R+\epsilon_n)}$ in the case $R \leq D(q)$ as shown in Theorem 7.1, hence the first term above is bounded as

$$-\sum_{\mathbf{w}\in \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \log P_{\mathcal{S}}(\mathbf{w}) \leq -\log 2^{-2n(R+\epsilon_n)} \sum_{\mathbf{w}\in \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \tag{7.31}$$

$$\leq 2n(R+\epsilon_n) \tag{7.32}$$

To bound the second term, using *log sum inequatliy* [14, Lemma 3.1] gives

$$-\sum_{\mathbf{w}\notin \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w}) \log P_{\mathcal{S}}(\mathbf{w}) \leq -\left(\sum_{\mathbf{w}\notin \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w})\right) \log \frac{\sum_{\mathbf{w}\notin \mathcal{K}_N} P_{\mathcal{S}}(\mathbf{w})}{|\overline{\mathcal{K}_N}|} \tag{7.33}$$

$$= -P_{\mathcal{S}}(\overline{\mathcal{K}_N}) \log P_{\mathcal{S}}(\overline{\mathcal{K}_N}) + P_{\mathcal{S}}(\overline{\mathcal{K}_N}) \log |\overline{\mathcal{K}_N}| \tag{7.34}$$

where $\overline{\mathcal{K}_N}$ denotes the complementary set of K^*. We use the fact that $P_{\mathcal{S}}(\overline{\mathcal{K}_N}) \leq Ae^{-n(2\delta^2/\log q)}$ in proved later in Lemma 7.4, Eq. (7.43). For $n \to \infty$, the first term above approaches zero as $P_{\mathcal{S}}(\overline{\mathcal{K}_N}) \to 0$. The second term is bounded as

$$P_{\mathcal{S}}(\overline{\mathcal{K}_N}) \log |\overline{\mathcal{K}_N}| \leq Ae^{-n(2R\delta^2/\log q)} \log 2^{2nR} \tag{7.35}$$

$$= 2nRAe^{-n(2R\delta^2/\log q)} \tag{7.36}$$

approaches zero as well for large enough n. Overall we have

$$\lim_{n\to\infty} H(T_1^n + T_2^n) \leq \lim_{n\to\infty} 2(R+\epsilon_n) + o_n(1) \tag{7.37}$$

$$= 2nR = 2H(T^n)/n \tag{7.38}$$

This shows in the limit we have $H(T_1^n + T_2^n) \to 2H(Y^n)/n$ for $R \leq D(q)$ and the claim for the case $R > D(q)$ can be proved in the same way. □

7.1.4 Proof of Theorem 7.1

We prove Theorem 7.1 in a few steps. Lemma 7.2 already shows that for any linear code $\mathcal{C}$, there exists a corresponding systematic code $\mathcal{C}'$ whose sumset structure is the same as the former. Hence we first concentrate on systematic linear codes and establish a similar result.

Theorem 7.3 (Normal typical sumset - systematic linear codes)**.** *Let $\mathcal{C}^{(n)}$ be a sequence of systematic linear codes in the form (7.7). We assume each entry of the matrix $\mathbf{Q}$ is independent and identically distributed according the uniform distribution in $\mathbb{F}_q$. Then a.a.s. there exists a sequence of typical sumsets $\mathcal{K}_N^{(n)} \subseteq \mathcal{C}^{(n)} + \mathcal{C}^{(n)}$ with size given in (7.11). Furthermore, the induced probability distribution $P_{\mathcal{S}}$ on $\mathcal{C}^{(n)} + \mathcal{C}^{(n)}$ satisfies (7.13).*

We point out that there exist linear codes with a smaller typical sumset than $|\mathcal{K}_N|$. As a simple example consider a systematic linear codes with generator matrix $[\mathbf{I}; \mathbf{0}]$, i.e., the $\mathbf{Q}$ matrix is the zero matrix. Since the sum codewords are essentially k-length sequences with each entry i.i.d. with distribution P_W, it is easy to see that the set of typical sequences $\mathcal{A}_{[W]}^k$ is actually a typical sumset for this code with size $2^{kH(W)} = 2^{nRH(W)/\log q}$ where W has the distribution in (7.2). This code has a typical sumset which is smaller than the normal typical sumset as demonstrated in Figure 7.3. However this kind of codes are rare and the above theorem states that a randomly picked systematic linear code has a normal typical sumset a.a.s..

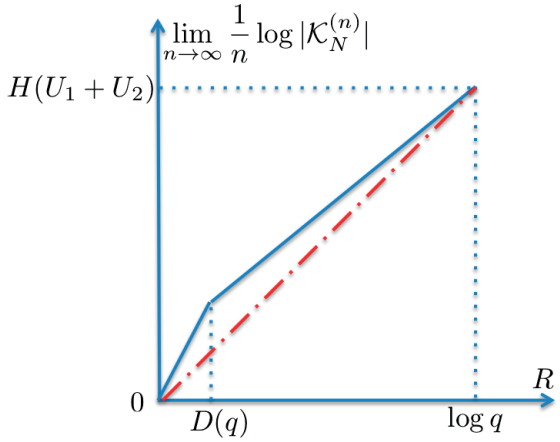


Figure 7.3 – Linear code with a typical sumset which is not normal: The solid line shows the size of the normal typical sumset and the dot-dashed line shows the size of a typical sumset of the example given above. This code has a small typical sumset with size $2^{nRH(W)/\log q}$ but is uninteresting for error correction.

We first prove Theorem 7.3. Let T_1^n, T_2^n be two independent random variables uniformly distributed in a systematic linear code $\mathcal{C}$ generated by the generator matrix $[\mathbf{I}; \mathbf{Q}]$, and $\mathbf{t}$ and $\mathbf{v}$ realizations of T_1^n and T_2^n, respectively. We choose the set $\mathcal{K}_N$ to contain sum codewords whose information-sums $\mathbf{s}$ defined in (7.9) are typical:

$$\mathcal{K}_N := \left\{ \mathbf{t} + \mathbf{v} \middle| \mathbf{t} + \mathbf{v} = \begin{bmatrix} \mathbf{s} \\ \mathbf{p} \end{bmatrix} \text{ where } \mathbf{s} \in \mathcal{A}_{[W]}^{(k)} \right\} \tag{7.39}$$

For all pairs of codewords $(\mathbf{t}, \mathbf{v})$ whose information-sum equals to a common value $\mathbf{s}$ we define the set of all possible parity-sums as

$$\mathcal{P}_{\mathbf{Q}}(\mathbf{s}) := \{\mathbf{Qm} + \mathbf{Qn} : \mathbf{m}, \mathbf{n} \in \mathcal{U}^k \text{ such that } \mathbf{m} + \mathbf{n} = \mathbf{s}\}. \tag{7.40}$$

Lemma 7.4. *(Simple estimates of $|\mathcal{K}_N|$) Let T_1^n, T_2^n be two independent copies of T^n uniformly distributed in a systematic linear code $\mathcal{C}$ as in (7.7) with any matrix $\mathbf{Q}$. Asymptotically almost surely, we have*

$$T_1^n + T_2^n \in \mathcal{K}_N \tag{7.41}$$

with $\mathcal{K}_N$ defined in (7.39). Furthermore we have

$$2^{k(H(U_1+U_2)+o(1))} \cdot \min_{\mathbf{s}\in\mathcal{A}_{[W]}^{(k)}} |\mathcal{P}_{\mathbf{Q}}(\mathbf{s})| \le |\mathcal{K}_N| \le 2^{k(H(U_1+U_2)+o(1))} \cdot \max_{\mathbf{s}\in\mathcal{A}_{[W]}^{(k)}} |\mathcal{P}_{\mathbf{Q}}(\mathbf{s})| \tag{7.42}$$

where U_1, U_2 are two independent random variables with distribution p_U in (7.1).

Proof. Recall that we defined $\mathcal{K}_N$ to be the set containing all sum codewords whose information-sum $\mathbf{s}$ is a typical sequence in $\mathcal{A}_{[W]}^{(k)}$. As T_1^n and T_2^n are independently and uniformly chosen from $\mathcal{C}$, the first k entries of both T_1^n and T_2^n are independent and they are in fact i.i.d. random variables with distribution P_U, due the the systematic form of the code $\mathcal{C}$. Hence by definition of $\mathcal{K}_N$ we have

$$\mathbb{P}\left\{T_1^n + T_2^n \in \mathcal{K}_N\right\} = \mathbb{P}\left\{S^k \in \mathcal{A}_{[W]}^{(k)}\right\} \ge 1 - 2|\mathcal{W}|e^{-2k\delta^2} = 1 - 2(2q-2)e^{-n(2\delta^2/\log q)} \tag{7.43}$$

where S^k is a k-length random vector with each entry i.i.d. according to P_W and the inequality follows from the property of typical sequences in Lemma 7.1. For the choice δ ensuring $n\delta^2 \to \infty$, we have that $T_1^n + T_2^n \in \mathcal{K}_N$ a.a.s. for n large enough, and particularly

$$\mathbb{P}\left\{T_1^n + T_2^n \in \mathcal{K}_N\right\} \ge 1 - Ae^{-n(2\delta^2/\log q)}. \tag{7.44}$$

if we define $A := 2(2q-2)$. To bound the size of $\mathcal{K}_N$, we can write the set $\mathcal{K}_N$ as the disjoint union

$$\mathcal{K}_N = \bigcup_{\mathbf{s}\in\mathcal{A}_{[W]}^{(k)}} \mathcal{P}_{\mathbf{Q}}(\mathbf{s}).$$

Then the claim follows from the fact that $|\mathcal{A}_{[W]}^{(k)}| = 2^{k(H(U_1+U_2)+o(1))}$, also shown in Lemma 7.1.

□

Now we are only interested in message pairs $(\mathbf{m}, \mathbf{n})$ if they sum up to a typical sequence $\mathbf{s} \in \mathcal{A}_{[W]}^{(k)}$. For a fixed such sequence $\mathbf{s}$, we can explicitly characterize all the pairs $(\mathbf{m}, \mathbf{n})$ such that $\mathbf{m} + \mathbf{n} = \mathbf{s}$.

Lemma 7.5 (Characterization of information-sum)**.** *Given a k-length sequence $\mathbf{s} \in \mathcal{A}_{[W]}^{(k)}$, there are L different pairs $(\mathbf{m}, \mathbf{n})$ satisfying $\mathbf{m} + \mathbf{n} = \mathbf{s}$ where*

$$L = 2^{k(\log q - D(q) + o(1))} \tag{7.45}$$

Proof. Observe that for a given entry $\mathbf{s}_i \in \mathcal{W}$, we can write out all possible $(\mathbf{m}_i, \mathbf{n}_i)$ explicitly as the following

$$\begin{aligned}
\mathbf{s}_i &: (\mathbf{m}_i, \mathbf{n}_i) \text{ such that } \mathbf{m}_i + \mathbf{n}_i = \mathbf{s}_i \\
0 &: (0,0) \\
1 &: (0,1), (1,0) \\
2 &: (1,1), (2,0), (0,2) \\
3 &: (0,3), (3,0), (1,2), (2,1) \\
&\vdots \\
q-1 &: (0,q-1), (q-1,0), (1,q-2), (q-2,1), \ldots, ((q-1)/2, (q-1)/2) \\
&\vdots \\
2q-3 &: (q-1,q-2), (q-2,q-1) \\
2q-2 &: (q-1,q-1)
\end{aligned}$$

To prove the claim, we show that the number of different pairs $\mathbf{m}, \mathbf{n}$ satisfying $\mathbf{m} + \mathbf{n} = \mathbf{s}$ is

$$L = \cdot 2^{(2/q^2+o(1))k} 3^{(3/q^2+o(1))k} \cdots q^{(q/q^2+o(1))k} (q-1)^{((q-1)/q^2+o(1))k} \cdots 2^{(2/q^2+o(1))k} \tag{7.46}$$

$$= \frac{1}{2} \prod_{a=1}^{q} a^{(a/q^2+o(1))k} \prod_{a=1}^{q-1} a^{(a/q^2+o(1))k} \tag{7.47}$$

$$= 2^{k(\log q - D(q) + o(1))} \tag{7.48}$$

To see why this is the case, there are for example $(2/q^2 + o(1))k$ entries (denoted by $I(\mathbf{s}, 1)$) in $\mathbf{s}$ taking value 1, as implied by the definition of typical sequences in (1.1). The pair $(\mathbf{m}_i, \mathbf{n}_i)$ can take value $(1,0)$ or $(0,1)$ for the indices $i \in I(\mathbf{s}, 1)$. Hence there are $2^{(2/q^2+o(1))k}$ different choices on the pair $(\mathbf{m}, \mathbf{n})$ for those entries $i \in I(\mathbf{s}, 1)$. The same argument goes for other entries taking values $2, \ldots, 2q-2$ using number of possible values of $(\mathbf{m}_i, \mathbf{n}_i)$ shown in the above list. □

As there are L different pairs $(\mathbf{m}, \mathbf{n})$ for a given $\mathbf{s}$, we use $\mathbf{p}(\ell)$ to denote the parity-sum in (7.9) resulting from the ℓ-th pair $(\mathbf{m}, \mathbf{n})(\ell)$, i.e.

$$\mathbf{p}(\ell) := \mathbf{Q}\mathbf{m} + \mathbf{Q}\mathbf{n} \text{ for the } \ell\text{-th pair } (\mathbf{m}, \mathbf{n}), \mathbf{m} + \mathbf{n} = \mathbf{s}, \ell \in [1 : L] \tag{7.49}$$

Now we set out to characterize the parity-sum $\mathbf{p}(\mathbf{m}, \mathbf{n})$. One key observation is that for a fixed $\mathbf{Q}$ and any pair $(\mathbf{m}, \mathbf{n})$ sum up to $\mathbf{s}$, the i-th entry of all parity-sums $\mathbf{p}(\mathbf{m}, \mathbf{n})$ takes only one or two values.

Lemma 7.6. *(Key property of parity-sums) For any given $\mathbf{Q}$, let $\mathbf{p}_i(\ell)$ be the i-th entry of $\mathbf{p}(\ell)$ defined in (7.49) with $\mathbf{m} + \mathbf{n} = \mathbf{s}$. We have*

$$\mathbf{p}_i(\ell) \in \{a, a+q\} \text{ with some } a \in [0 : q-1] \text{ for all } \ell \in [1 : L] \tag{7.50}$$

Equivalently, define a subset in $\mathcal{W}^{n-k}$ with a vector $\mathbf{a} \in \mathcal{U}^{n-k}$ as

$$\mathcal{F}(\mathbf{a}) := \{\mathbf{h} : \mathbf{h}_i \in \{a_i, a_i + q\}, i \in [1 : n-k]\}, \tag{7.51}$$

we always have

$$\mathcal{P}_{\mathbf{Q}}(\mathbf{s}) \subseteq \mathcal{F}(\mathbf{a}) \tag{7.52}$$

with some $\mathbf{a} \in \mathcal{U}^{n-k}$ *depending only on* $\mathbf{s}$ *and* $\mathbf{Q}$.

Proof. Using the notation $\langle \mathbf{m}, \mathbf{n} \rangle$ to denote the inner product of two vectors in $\mathbb{R}^k$ and $\mathbf{Q}_i$ to denote the i-th column of $\mathbf{Q}$, we have

$$\begin{aligned}
\mathbf{p}_i(\ell) &= \mathbf{Q}_i^T \mathbf{m} + \mathbf{Q}_i^T \mathbf{n} && (7.53)\\
&= \langle \mathbf{Q}_i, \mathbf{m} \rangle + q n_1 + \langle \mathbf{Q}_i, \mathbf{n} \rangle + q n_2 && (7.54)\\
&\overset{(a)}{=} \langle \mathbf{Q}_i, \mathbf{s} \rangle + q(n_1 + n_2) && (7.55)\\
&= q n_3 + a + q(n_1 + n_2) && (7.56)\\
&= a + q(n_1 + n_2 + n_3) && (7.57)
\end{aligned}$$

for some $n_1, n_2, n_3 \in \mathbb{Z}$ and $a \in [0 : q-1]$. In step (a) we used the fact that $\mathbf{m}+\mathbf{n} = \mathbf{s}$ in the assumption. On the other hand we know $\mathbf{p}_i(\ell)$ only takes value in $[0 : 2q-2]$, the above expression implies $\mathbf{p}_i$ can only equal to a or $a+q$ for some $a \in [0, q-1]$, namely $n_1 + n_2 + n_3$ can only equal to 0 or 1. In particular if $a = q-1$, we must have $n_1 + n_2 + n_3 = 0$ and $\mathbf{p}_i = q-1$. We can use the same argument for all entries $p_i(\ell), i = 1, \ldots, n-k$ and show that the entry $p_i(\ell)$ can take at most two different values for any ℓ. As a consequence for a fixed $\mathbf{s}$ and any given $\mathbf{Q}$, all the parity-sums $\mathcal{P}_{\mathbf{Q}}(\mathbf{s})$ belong to the set $\mathcal{F}(\mathbf{a})$ with $\mathbf{a}$ depending on $\mathbf{s}$ and $\mathbf{Q}$. □

Since there are q^{n-k} different choice of $\mathbf{a}$, we can partition the whole space $\mathcal{W}^{n-k}$ into q^{n-k} disjoint subsets $\mathcal{F}(\mathbf{a})$. For a given $\mathbf{Q}$ and information sum $\mathbf{s}$, all the parity-sums $\mathcal{P}_{\mathbf{Q}}(\mathbf{s})$ are confined in a subset $\mathcal{F}(\mathbf{a})$. This is the key property for characterizing the sumset structure. To lighten the notation, for a given $\mathbf{s}$ we define

$$F(\mathbf{a}) := \{\mathcal{P}_{\mathbf{Q}}(\mathbf{s}) \subseteq \mathcal{F}(\mathbf{a})\} \tag{7.58}$$

to denote the event when all parity-sums are contained in the set $\mathcal{F}(\mathbf{a})$ due to the choice of $\mathbf{Q}$. As each row $\mathbf{Q}_i$ of $\mathbf{Q}$ is chosen independently, we have

$$\begin{aligned}
\mathbb{P}\{F(\mathbf{a})\} &= \mathbb{P}\{\mathbf{p}_i(\ell) \in \{\mathbf{a}_i, \mathbf{a}_i + q\} \text{ for all } i \in [1 : n-k]\} && (7.59)\\
&= \prod_{i=1}^{n-k} \mathbb{P}\{\mathbf{p}_i(\ell) \in \{\mathbf{a}_i, \mathbf{a}_i + q\}\} && (7.60)\\
&= q^{-(n-k)} && (7.61)
\end{aligned}$$

where the last step uses Lemma 7.9 on the distribution of $\mathbf{p}_i(\ell)$. Notice that $\mathbb{P}\{F(\mathbf{a})\}$ is independent from the actual value $\mathbf{a}$. Also notice that the notations $\mathbf{p}(\ell), F(\mathbf{a})$ and the notations in the sequel all concern the parity sums of a given information sum $\mathbf{s}$, which is omitted in the notations for the sake of brevity. The results should hold for all typical $\mathbf{s}$.

The estimates in Lemma 7.4 do not depend on the specific choice of the matrix $\mathbf{Q}$, namely the code $\mathcal{C}$. Now we study $|\mathcal{P}_{\mathbf{Q}}(\mathbf{s})|$ with randomly chosen $\mathbf{Q}$. In this case we use $\mathcal{P}(\mathbf{s})$ to denote a *random set* resulting from a randomly chosen matrix $\mathbf{Q}$. We need more notations to facilitate our arguments. Notice that the dependence

on the sum $\mathbf{s}$ is omitted in the notation. For a given vector $\mathbf{h} \in \mathcal{W}^{(n-k)}$, we define random variables $Z_{\ell,i}(\mathbf{h}), i \in [1 : n-k]$ to be the indicator function

$$Z_{\ell,i}(\mathbf{h}) := \mathbf{1}_{\mathbf{p}_i(\ell)=\mathbf{h}_i} \tag{7.62}$$

i.e., $Z_{\ell,i}(\mathbf{h})$ equals 1 when the i-th entry of the parity-sum $\mathbf{p}(\ell)$ is equal to the entry $\mathbf{h}_i$. Furthermore we define

$$Z_\ell(\mathbf{h}) := \prod_{i=1}^{n-k} Z_{\ell,i}(\mathbf{h}), \tag{7.63}$$

$$Z(\mathbf{h}) := \sum_{\ell=1}^{L} Z_\ell(\mathbf{h}). \tag{7.64}$$

We see $Z_\ell(\mathbf{h})$ is also an indicator function and is equal to 1 if the ℓ-th pair sum up to the parity-sum $\mathbf{h}$. Furthermore $Z(\mathbf{h})$ counts the number of different pairs $(\mathbf{m}, \mathbf{n})$ summing up to $\mathbf{s}$ which give a parity-sum $\mathbf{p}(\mathbf{m}, \mathbf{n})$ equal to $\mathbf{h}$. With this notation the event $\{\mathbf{p}(\ell) = \mathbf{h}\}$ is equivalent to the event $\{Z_\ell(\mathbf{h}) = 1\}$ and the two following events

$$\{\mathbf{h} \in \mathcal{P}(\mathbf{s})\} = \{\mathbf{p}(\ell) = \mathbf{h} \text{ for some } \ell \in [1 : L]\} \tag{7.65}$$

are equivalent to the event $\{Z(\mathbf{h}) \geq 1\}$.

Lemma 7.7 (Size of parity-sums for $R \leq D(q)$)**.** *Consider a systematic linear code $\mathcal{C}$ in (7.7) with rate R. We assume that each entry for its matrix $\mathbf{Q}$ is i.i.d. according to P_U. For any information-sum $\mathbf{s} \in \mathcal{A}^{(k)}_{[W]}$, the size of the parity-sums $\mathcal{P}(\mathbf{s})$ defined in (7.40) satisfies*

$$|\mathcal{P}(\mathbf{s})| \doteq 2^{k(\log q - D(q))} \quad a.a.s. \tag{7.66}$$

if $R \leq D(q)$.

Proof. We show in Appendix 7.2.2 that each entry of any parity-sum $\mathbf{p}$ in $\mathcal{P}(\mathbf{s})$ is i.i.d. according to P_W hence the probability that a parity-sum $\mathbf{p}$ being atypical is negligible.

In Lemma 7.5 we showed that there are L different pairs $(\mathbf{m}, \mathbf{n})$ pairs sum up to $\mathbf{s}$. Here we show that each pair will give a different parity sum $\mathbf{Qm} + \mathbf{Qn}$ a.a.s., hence the size of the set $\mathcal{P}(\mathbf{s})$ is equal to L. This is done by showing that given the fact that the information sum is equal to some $\mathbf{s}$, $\mathbb{P}\{Z(\mathbf{h}) > 1\}$ can be made arbitrarily small for any typical $\mathbf{h}$. In Appendix 7.2.3, we claim that for a typical sequence $\mathbf{h} \in \mathcal{F}(\mathbf{a})$, the expectation and variance of $Z(\mathbf{h})$ conditioned on the event $F(\mathbf{a})$ have the form

$$\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] = 2^{n(R-D(q)+o(1))} \tag{7.67a}$$

$$\mathrm{Var}[Z(\mathbf{h})|F(\mathbf{a})] < \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] \tag{7.67b}$$

This implies that we have

$$\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] \leq 2^{n(R-D(q)+\epsilon_n)} \tag{7.68}$$

with $\epsilon_n \searrow 0$ as $n \to \infty$.

Recall that $Z(\mathbf{h})$ denotes the number of pairs $(\mathbf{m}, \mathbf{n}')$ which sum up to the sequence $\mathbf{h}$. By Markov inequality we have

$$\mathbb{P}\{Z(\mathbf{h}) > 1 | F(\mathbf{a})\} \leq \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] \leq 2^{n(R-D(q)+\epsilon_n)}$$

which can be arbitrarily small with sufficiently large n ensuring that $R < D(q) - \epsilon_n$. As $Z(\mathbf{h})$ denotes the number of pairs $(\mathbf{m}, \mathbf{n})$ which give a parity-sum part equal to $\mathbf{h}$. This means a.a.s. any typical sequence $\mathbf{h}$ can be formed by at most one pair $(\mathbf{m}, \mathbf{n})$ conditioned on $F(\mathbf{a})$. In other words, every pair gives a distinct $\mathbf{p}$ a.a.s. hence the size of $\mathcal{P}(\mathbf{s})$ equals the total number of pairs L. □

Lemma 7.8 (Size of parity-sums for $R > D(q)$)**.** *Consider a systematic linear code $\mathcal{C}$ in (7.7) with rate R. We assume that each entry for its matrix $\mathbf{Q}$ is i.i.d. according to P_U. For any information-sum $\mathbf{s} \in \mathcal{A}_{[W]}^{(k)}$, asymptotically almost surely the size of the parity-sums $\mathcal{P}(\mathbf{s})$ defined in (7.40) satisfies*

$$|\mathcal{P}(\mathbf{s})| \doteq 2^{(n-k)D(q)}$$

if $R > D(q)$.

Proof. The same as in the proof of Lemma 7.7, we will only concentrate on typical sequence as the probability of parity-sum $\mathbf{p}$ being atypical is negligible. We first show that for rate $R > D(q)$ and a typical sequence $\mathbf{h}$, the random variable $Z(\mathbf{h})$ concentrates around $\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})]$ conditioned on the event $F(\mathbf{a})$. Recall from (7.67) that we have

$$2^{n(R-D(q)-\epsilon_n)} \leq \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] \leq 2^{n(R-D(q)+\epsilon_n)} \tag{7.69}$$

with some $\epsilon_n \searrow 0$ as $n \to \infty$. Hence for some $\epsilon'_n > 0$ depending on n, by (conditional version of the) Chebyshev inequality (see [67, Ch. 23.4] for example) we have

$$\begin{aligned}
\mathbb{P}\left\{|Z(\mathbf{h}) - \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] \geq 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)}|F(\mathbf{a})\right\} &\leq \frac{\mathrm{Var}[Z(\mathbf{h})|F(\mathbf{a})]}{2^{2\cdot\frac{n}{2}(R-D(q)+\epsilon'_n)}} && (7.70)\\
&\leq \frac{\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})]}{2^{n(R-D(q)+\epsilon'_n)}} && (7.71)\\
&\leq 2^{-n(\epsilon'_n-\epsilon_n)} && (7.72)
\end{aligned}$$

where we used the inequality $\mathrm{Var}[Z(\mathbf{h})|F(\mathbf{a})] \leq \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})]$ proved in Appendix 7.2.3. If we choose $\epsilon'_n > \epsilon_n$ and n such that $n(\epsilon'_n - \epsilon_n) \to \infty$ and $\epsilon'_n \to 0$ (this is possible because $\epsilon_n \searrow 0$), then a.a.s. $Z(\mathbf{h})$ satisfies

$$\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] - 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)} \leq Z(\mathbf{h}) \leq \mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] + 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)} \tag{7.73}$$

conditioned on the event $F(\mathbf{a})$. Furthermore we have the following identity regarding the total number of pairs $(\mathbf{m}, \mathbf{n})$ sum up to $\mathbf{s}$:

$$\sum_{\mathbf{h}\in\mathcal{P}(\mathbf{s})} Z(\mathbf{h}) = L \tag{7.74}$$

Combining (7.73) and (7.74), conditioned on the event $F(\mathbf{a})$ for any $\mathbf{a} \in \mathcal{U}^{n-k}$, the following estimates hold a.a.s.

$$\frac{L}{\mathbf{E}\left[Z(\mathbf{h})|F(\mathbf{a})\right] + 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)}} \leq |\mathcal{P}(\mathbf{s})| \leq \frac{L}{\mathbf{E}\left[Z(\mathbf{h})|F(\mathbf{a})\right] - 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)}} \quad (7.75)$$

Using L from Lemma 7.5, Eq. (7.69) and the above expression, $\mathcal{P}(\mathbf{s})$ is bounded a.a.s. as

$$\frac{2^{(n-k)(D(q)+o(1))}}{1+2^{-\frac{n}{2}(R-D(q)+2\epsilon_n-\epsilon'_n)}} \leq |\mathcal{P}(\mathbf{s})| \leq \frac{2^{(n-k)(D(q)+o(1))}}{1-2^{-\frac{n}{2}(R-D(q)-2\epsilon_n-\epsilon'_n)}} \quad (7.76)$$

Assume $R = D(q) + \sigma$ for some $\sigma > 0$ for now, we have

$$2^{-\frac{n}{2}(R-D(q)+2\epsilon_n-\epsilon'_n)} = 2^{-\frac{n}{2}(\sigma+2\epsilon_n-\epsilon')} \quad (7.77)$$

$$2^{-\frac{n}{2}(R-D(q)-2\epsilon_n-\epsilon'_n)} = 2^{-\frac{n}{2}(\sigma-2\epsilon_n-\epsilon'_n)} \quad (7.78)$$

and both terms approaches 0 if $\sigma > 2\epsilon_n + \epsilon'_n$. Since both ϵ_n and ϵ'_n are chosen to approach 0, we can have σ arbitrarily close to 0 as well. This proves that for $R > D(q)$ and n large enough we have a.a.s.

$$\frac{2^{(n-k)(D(q)+o(1))}}{1+o_n(1)} \leq |\mathcal{P}(\mathbf{s})| \leq \frac{2^{(n-k)(D(q)+o(1))}}{1-o_n(1)} \quad (7.79)$$

or equivalently $\mathcal{P}(\mathbf{s}) \doteq 2^{(n-k)D(q)}$ for $R > D(q)$ a.a.s. if n is sufficiently large. As this estimates holds conditioned a.a.s. under any event $F(\mathbf{a})$, and each $F(\mathbf{a})$ occurs with the same probability for all $\mathbf{a}$ (see Eq. (7.61)), we conclude that the claimed estimate holds a.a.s. unconditionally. □

With the foregoing lemmas we can finalize the proof of Theorem 7.3.

Proof of Theorem 7.3. Notice that the asymptotic estimates on $\mathcal{P}(\mathbf{s})$ in Lemma 7.7 and 7.8 hold for *all* information-sum $\mathbf{s} \in \mathcal{A}^{(k)}_{[W]}$, in particular they also hold for $\min_{\mathbf{s}\in\mathcal{A}^{(k)}_{[W]}} \mathcal{P}(\mathbf{s})$ and $\max_{\mathbf{s}\in\mathcal{A}^{(k)}_{[W]}} \mathcal{P}(\mathbf{s})$. Hence combining Lemma 7.4, 7.7 and 7.8, we conclude that for $R \leq D(q)$ we have

$$|\mathcal{K}_N| \doteq 2^{k(H(U_1+U_2))} \cdot 2^{k(\log q - D(q))} \quad (7.80)$$

$$= 2^{2k\log q} = 2^{2nR} \quad \text{a.a.s.} \quad (7.81)$$

and for $R > D(q)$ we have

$$|\mathcal{K}_N| \doteq 2^{k(H(U_1+U_2))} \cdot 2^{(n-k)D(q)} \quad (7.82)$$

$$= 2^{nD(q)+k\log q} = 2^{n(R+D(q))} \quad \text{a.a.s.} \quad (7.83)$$

Now we prove the asymptotic equipartion property (AEP) of the normal typical sumset $\mathcal{K}_N$ in (7.13). Assume the code $\mathcal{C}$ has a normal typical sumset $\mathcal{K}_N$ and define M^k, N^k to be two independent random variables uniformly distributed on $\mathcal{U}^k$. If we view M^k, N^k as two independent messages and let $T_1^n = \mathbf{G}M^k$, $T_2^n = \mathbf{G}N^k$ where $\mathbf{G}$ is a generator matrix in the form (7.7), then T_1^n, T_2^n are two independent

random variables uniformly distributed on $\mathcal{C}$. Recall that $P_{\mathcal{S}}$ denotes the probability distribution on the sumset $\mathcal{C}+\mathcal{C}$ induced by T_1^n, T_2^n.

We first consider the low rate regime when $R \le D(q)$. Recall that a sum codeword in $\mathcal{K}_N$ has the form $\mathbf{w} = \binom{\mathbf{s}}{\mathbf{p}}$ where $\mathbf{s}$ is a typical sequence in $\mathcal{A}^k_{[W]}$. Lemma 7.5 shows that there are L different pairs $(\mathbf{m}, \mathbf{n})$ sum up to $\mathbf{s}$ and Lemma 7.7 shows that each pair gives a unique parity-sum $\mathbf{p}$. In other words any $\mathbf{w} = \binom{\mathbf{s}}{\mathbf{p}} \in \mathcal{K}_N$ is formed by a unique pair $(\mathbf{m}_0, \mathbf{n}_0)$, i.e., $\mathbf{s} = \mathbf{m}_0 + \mathbf{n}_0$ and $\mathbf{p} = \mathbf{Q}\mathbf{m}_0 + \mathbf{Q}\mathbf{n}_0$. Hence

$$P_{\mathcal{S}}(\mathbf{w}) = \mathbb{P}\left\{M^k = \mathbf{m}_0, N^k = \mathbf{n}_0\right\} \tag{7.84}$$

$$= \mathbb{P}\left\{M^k = \mathbf{m}_0\right\}\mathbb{P}\left\{N^k = \mathbf{n}_0\right\} \tag{7.85}$$

$$= q^{-2k} = 2^{-2nR} \tag{7.86}$$

Now consider the case when $R > D(q)$. For any $\mathbf{w} = \binom{\mathbf{s}}{\mathbf{p}} \in \mathcal{K}_N$, Lemma 7.5 shows that there are L different pairs $(\mathbf{m}, \mathbf{n})$ sum up to $\mathbf{s}$ and Lemma 7.8 shows that within these L pairs, many pairs give the same parity-sum $\mathbf{p}$. More precisely, the number of pairs sum up to a particular parity-sum $\mathbf{p}$ in $\mathcal{P}(\mathbf{s})$ is bounded in (7.73) as

$$2^{n(R-D-\epsilon_n)} - 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)} \le Z(\mathbf{p}) \le 2^{n(R-D+\epsilon_n)} + 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)} \tag{7.87}$$

for some $\epsilon_n, \epsilon'_n \to \infty$. Hence for a sum codeword $\mathbf{w} = \binom{\mathbf{s}}{\mathbf{p}} \in \mathcal{K}_N$, we have

$$P_{\mathcal{S}}(\mathbf{w}) = \sum_{\substack{(\mathbf{m},\mathbf{n}) \\ \mathbf{m}+\mathbf{n}=\mathbf{s}, \mathbf{Q}\mathbf{m}+\mathbf{Q}\mathbf{n}=\mathbf{p}}} \mathbb{P}\left\{M^k = \mathbf{m}, N^k = \mathbf{n}\right\} \tag{7.88}$$

$$= \sum_{\substack{(\mathbf{m},\mathbf{n}) \\ \mathbf{m}+\mathbf{n}=\mathbf{s}, \mathbf{Q}\mathbf{m}+\mathbf{Q}\mathbf{n}=\mathbf{p}}} q^{-2k} \tag{7.89}$$

$$\le (2^{n(R-D(q)+\epsilon_n)} + 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)})2^{-2k\log q} \tag{7.90}$$

$$= 2^{-n(R+D(q)-\epsilon_n)}(1 + 2^{-\frac{n}{2}(R-D(q)-\epsilon'_n+2\epsilon_n)}) \tag{7.91}$$

$$\le 2^{-n(R+D(q)-\epsilon_n)}(1 + 2^{-n(-\epsilon'_n/2+\epsilon_n)}) \tag{7.92}$$

for $R > D(q)$. If we furthermore require $\epsilon'_n \le 2\epsilon_n$ (notice in the proof of Lemma 7.8 we required that $\epsilon'_n > \epsilon_n$), then we can find a $\sigma_n \searrow 0$ such that $2^{n\sigma_n} \ge 1 + 2^{-n(-\epsilon'_n/2+\epsilon_n)} \to 1$ for n large enough. Hence we have

$$P_{\mathcal{S}}(\mathbf{w}) \le 2^{-n(R+D(q)-\epsilon_n-\sigma_n)} \tag{7.93}$$

On the other hand we have

$$P_{\mathcal{S}}(\mathbf{w}) = \sum_{\substack{(\mathbf{m},\mathbf{n}) \\ \mathbf{m}+\mathbf{n}=\mathbf{s}, \mathbf{Q}\mathbf{m}+\mathbf{Q}\mathbf{n}=\mathbf{p}}} \mathbb{P}\{M = \mathbf{m}, N = \mathbf{n}\} \tag{7.94}$$

$$= \sum_{\substack{(\mathbf{m},\mathbf{n}) \\ \mathbf{m}+\mathbf{n}=\mathbf{s}, \mathbf{Q}\mathbf{m}+\mathbf{Q}\mathbf{n}=\mathbf{p}}} q^{-2k} \tag{7.95}$$

$$\ge (2^{n(R-D(q)-\epsilon_n)} + 2^{\frac{n}{2}(R-D(q)+\epsilon'_n)})2^{-2k\log q} \tag{7.96}$$

$$= 2^{-n(R+D(q)+\epsilon_n)}(1 + 2^{-\frac{n}{2}(R-D(q)-\epsilon'_n-2\epsilon_n)}) \tag{7.97}$$

$$\ge 2^{-n(R+D(q)+\epsilon_n)} \tag{7.98}$$

This proves that for $R > D(q)$ we have

$$P_{\mathcal{S}}(\mathbf{w}) \doteq 2^{-n(R+D(q))} \tag{7.99}$$

and concludes the proof Theorem 7.3. □

With the results established for systematic linear codes, we can finally prove the results for general linear codes.

Proof of Theorem 7.1. In Theorem 7.3 we considered the ensemble of codes where all possible full-rank systematic generator matrices $[\mathbf{I}; \mathbf{Q}]$ is chosen with equal probability. It is known that the systematic generator matrix for a systematic linear code is unique. Furthermore, as we can identify a linear code with the k-dimensional subspace spanned by its generator matrix, each systematic generator matrix thus gives a distinct code hence a distinct k-dimensional subspace. It is known that the total number of k-dimensional subspaces in $\mathbb{F}_q^n$ is given by the so-called Gaussian binomial coefficient (see [68] for example):

$$\binom{n}{k}_q := \frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})} \tag{7.100}$$

As shown by Lemma 7.2 there is a one-to-one mapping between two codes and their sumsets if two codes are equivalent, hence if a code $\mathcal{C}$ is equivalent to some systematic linear code $\mathcal{C}'$ with a normal typical sumset $\mathcal{K}_N$, the code $\mathcal{C}$ also has a normal typical sumset. But since every linear code (equivalently every k-dimensional subspace) is equivalent to some systematic code, Theorem 7.3 then shows that almost all of the k-dimensional subspaces correspond to (n, k) codes who have a normal typical sumset. Formally the number of codes which have a normal typical sumset is $(1 - o(1))\binom{n}{k}_q$.

Now consider the codes ensemble in Theorem 7.1 where we choose all possible q^{nk} generator matrices with equal probability. Clearly some of the generator matrices give the same code if they span the same k-dimensional subspace. We will show most of these generator matrices will give codes which have a normal typical sumsets. To show this, notice that each distinct k-dimensional subspace can be generated by $(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})$ different generator matrices (because there are this many different choices of basis in a k-dimensional subspace). Hence the fraction of the generator matrices with a normal typical sumset is

$$\begin{aligned}
\rho &:= \frac{(1-o(1))\binom{n}{k}_q \cdot (q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}{q^{nk}} \\
&= (1-o(1))\frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{q^{nk}} \\
&= (1-o(1))(1-q^{-n})(1-q^{-n+1})\cdots(1-q^{-n+k-1}) \\
&> (1-o(1))(1-q^{-n+k})^k
\end{aligned}$$

Assume $k = \beta n$ for some $\beta \in [0, 1)$, L'Hôpital's rule shows the logarithm of the term $(1 - q^{-n+k})^k$ has limit

$$\lim_{n\to\infty} \beta n \ln(1 - q^{-n(1+\beta)}) = \lim_{n\to\infty} \frac{\ln(1 - q^{-n(1+\beta)}}{1/\beta n} \tag{7.101}$$

$$= \lim_{n\to\infty} \frac{-\beta n^2}{1 - q^{-n(1+\beta)}} q^{-n(1+\beta)}(1 + \beta) \ln q \tag{7.102}$$

$$= 0 \tag{7.103}$$

Hence the fraction ρ of codes with a normal typical sumset is arbitrarily close to 1 for sufficiently large n. This proves that for the code ensemble in Theorem 7.1, codes will have a normal typical sumset a.a.s.. The proof of AEP property of the normal typical sumset is the same as in the proof of Theorem 7.3 by noticing that every linear code is equivalent to some systematic linear code, and we shall not repeat it. □

7.1.5 The weakness of certain structured codes

The results on typical sumsets can offer enlightening results in certain multi-user communication scenarios. We give a simple example in this section by considering the following multiple access channel

$$Y = X_1 + X_2 \tag{7.104}$$

where X_1, X_2 take values in the set of integers $\{0, \ldots, q-1\}$ for some prime number q. As formally described in Section 2.2, the decoder wishes to decode both messages of the two users.

The sum capacity of this channel is easily shown to be

$$C_{sum} := \max_{P_{X_1}, P_{X_2}} I(X_1, X_2; Y) = \log(2q - 2) \tag{7.105}$$

which can be achieved if both users independently generate their codes.

What is the achievable rates if linear codes are used? Here we assume that linear codes $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{F}_q^n$ are generated, and the codewords (vectors in $\{0, \ldots, q-1\}^n$) are directly inserted to the channel. It is not hard to show that, if both users *independently* generate linear codes $\mathcal{C}_1, \mathcal{C}_2$, the achievable sum rate is give by

$$R_{sum} = I(X_{1,U}, X_{2,U}; Y) = \log q + c$$

with some constant $c \geq 0.5$. Here $X_{1,U}, X_{2,U}$ are two independent random variables which are uniformly distributed in the set $\{0, \ldots, q-1\}$. The restriction to a uniform channel input distribution is due the fact that linear codes are used.

Now we ask the question: what is the achievable rates if two users use the same linear code $\mathcal{C}$? For simplicity we consider the symmetric case when $R_1 = R_2 = R$. Let $P_e^{(n)}$ denote the decoding error probability, Fano's inequality states

$$H(X_1^n, X_2^n | Y^n) \leq 2nRP_e^{(n)} + 1,$$

and it can be rewritten for the example in (7.104) as

$$P_e^{(n)} \geq 1 - \frac{H(X_1^n + X_2^n)}{2nR} - \epsilon_n$$

The above expression shows that for large n, the error probability $P_e^{(n)}$ is bounded away from zero if $H(X_1^n+X_2^n)$ is smaller than $2nR$. Theorem 7.2 shows that for most linear codes, $H(X_1^n + X_2^n)$ is equal to $2R$ for $R \leq D(q)$ and is equal to $R + D(q)$ (hence smaller than $2R$) in the regime $R > D(q)$, if X_1^n, X_2^n are drawn from the same linear codes. This means that if the same linear codes are used by both users, $P_e^{(n)}$ is bounded away from 0 for $R > D(q)$, hence the symmetric achievable rate for the channel (7.104) cannot be higher $D(q)$. Furthermore the same results on $H(X_1^n + X_2^n)$ hold for the case when X_1^n, X_2^n are independently chosen from two codes which is coset to each other, i.e., $\mathcal{C}_1 = \mathcal{C}_2 \oplus \mathbf{d}$ for any $\mathbf{d} \in \mathbb{F}_q^n$. We then conclude that the symmetric achievable rate cannot be higher than $D(q)$ if both users use the same linear codes (up to cosets).

Recall that we have $D(q) < \log\sqrt{e}$, which is in contrast to the achievable rates in (7.105) and (7.106), which scale with q.

7.2 Appendix

7.2.1 Properties of $D(q)$

The sum $\sum_{i=1}^{q} i \log i$ can be bounded as

$$\int_1^q x \log x dx + 1 \cdot \log 1 \leq \sum_{i=1}^{q} i \log i \leq \int_1^q x \log x dx + q \cdot \log q \tag{7.106}$$

which evaluates to

$$\frac{q^2}{2}\log q - \log e(q^2/4 + 1/4) \leq \sum_{i=1}^{q} i \log i \leq \frac{q^2}{2}\log q - \log e(q^2/4 + 1/4) + q \log q$$

Using the expression in (7.3) we have

$$\log q + \log\sqrt{e} - \frac{1 + q\log q}{q^2} \leq H(U_1 + U_2) \leq \log q + \log\sqrt{e} - \frac{1 - q\log q}{q^2}.$$

This shows that for $q \to \infty$ we have $H(U_1+U_2) \to \log q + \log\sqrt{e}$ and $D(q) \to \log\sqrt{e}$.

The fact that $D(q)$ is increasing with q can be checked straightforwardly.

7.2.2 On the distribution of parity-sums

For a message pair $(\mathbf{m}, \mathbf{n})$, we want to analyze the distribution of the parity-sum $\mathbf{p} := \mathbf{Q}\mathbf{m} + \mathbf{Q}\mathbf{n}$ with randomly chosen generator matrix $\mathbf{Q}$. Since each column of $\mathbf{Q}$ is chosen independently, we can without loss of generality study only one entry of $\mathbf{p}$.

Lemma 7.9 (Distribution of parity-sum)**.** *Let* $\mathbf{q}$ *be a* k*-length vectors taking values in* $\mathbb{F}_q^k$ *and* $p := \mathbf{q}^T\mathbf{m} + \mathbf{q}^T\mathbf{n}$ *for any pair* $(\mathbf{m}, \mathbf{n})$ *such that* $\mathbf{m} + \mathbf{n} \in \mathcal{A}_{[W]}^{(k)}$*. If each entry of* $\mathbf{q}$ *is i.i.d. random variable with distribution* p_U*, then* p *has distribution* P_W *defined in (7.2).*

Proof. For any $\mathbf{m}+\mathbf{n} \in \mathcal{A}_{[W]}^{(k)}$, we write out the expressions explicitly

$$\mathbf{q}^T\mathbf{m} = \mathbf{q}_1\mathbf{m}_1 \oplus \cdots \oplus \mathbf{q}_k\mathbf{m}_k \tag{7.107}$$

$$\mathbf{q}^T\mathbf{n} = \mathbf{q}_1\mathbf{n}_1 \oplus \cdots \oplus \mathbf{q}_k\mathbf{n}_k. \tag{7.108}$$

First observe that since each entry $\mathbf{q}_i$ is chosen independently with the uniform distribution p_U, each term $\mathbf{q}_i\mathbf{m}_i$ and $\mathbf{q}_i\mathbf{n}_i$ also have a uniform distribution for nonzero $\mathbf{m}_i, \mathbf{n}_i$. Hence both $\mathbf{q}^T\mathbf{m}$ and $\mathbf{q}^T\mathbf{n}$ have distribution p_U as long as $\mathbf{m}, \mathbf{n}$ are not zero vectors, which is always the case here.

We can also show that $\mathbf{q}^T\mathbf{m}$ and $\mathbf{q}^T\mathbf{n}$ are independent. We denote $\mathbf{s} := \mathbf{m}+\mathbf{n}$ and since $\mathbf{s} \in \mathcal{A}_{[W]}^{(k)}$, there are about $2k/q^2$ entries of $\mathbf{s}$ taking value 1. Let $I(\mathbf{s}, 1)$ denote the set of indices of these entries. As shown in Lemma 7.5, we should have $(\mathbf{m}_i, \mathbf{n}_i) = (0, 1)$ or $(\mathbf{m}_i, \mathbf{n}_i) = (1, 0)$ for all $i \in I(\mathbf{s}, 1)$. Hence $\mathbf{q}^T\mathbf{m}$ and $\mathbf{q}^T\mathbf{n}$ always have the form

$$\mathbf{q}^T\mathbf{m} = \mathbf{q}_1\mathbf{m}_1 \oplus \cdots \oplus \mathbf{q}_i\mathbf{m}_i \oplus \cdots \oplus \mathbf{q}_k\mathbf{m}_k \tag{7.109}$$

$$\mathbf{q}^T\mathbf{n} = \mathbf{q}_1\mathbf{m}_1 \oplus \cdots \oplus \mathbf{q}_i\mathbf{n}_i \oplus \cdots \oplus \mathbf{q}_k\mathbf{n}_k \tag{7.110}$$

with $(\mathbf{m}_i, \mathbf{n}_i)$ equals $(0, 1)$ or $(1, 0)$ for $i \in I(\mathbf{s}, 1)$. Hence there exists at least one term $\mathbf{q}_i$ which is either in the summation of $\mathbf{q}^T\mathbf{m}$ or in the summation of $\mathbf{q}^T\mathbf{n}$, but not in both. As each $\mathbf{q}_i$ is chosen independently according to the uniform distribution p_U, we conclude that $\mathbf{q}^T\mathbf{m}$ and $\mathbf{q}^T\mathbf{n}$ are independent with distribution p_U. It follows immediately that the sum p has the distribution p_W. □

7.2.3 Conditional expectation and variance

We calculate the conditional expectation $\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})]$ and conditional variance $\mathrm{Var}[Z(\mathbf{h})|F(\mathbf{a})]$ for typical sequence $\mathbf{h} \in \mathcal{F}(\mathbf{a})$. Notice we have $\mathbf{h}_i \in \{\mathbf{a}_i, \mathbf{a}_i + q\}$ conditioned on the event $F(\mathbf{a})$ where $\mathbf{a}_i \in [0 : q-1]$.

Now for a sequence $\mathbf{h} \in \mathcal{F}(\mathbf{a})$, by definition we have

$$\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] = \sum_{\ell=1}^{L} \mathbf{E}\left[\prod_{i=1}^{n-k} Z_{\ell,i}(\mathbf{h}) \middle| F(\mathbf{a})\right] \tag{7.111}$$

$$\overset{(a)}{=} \sum_{\ell=1}^{L} \prod_{i=1}^{n-k} \mathbf{E}[Z_{\ell,i}(\mathbf{h})|F(\mathbf{a})] \tag{7.112}$$

$$= \sum_{\ell=1}^{L} \prod_{i=1}^{n-k} \mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i|F(\mathbf{a})\} \tag{7.113}$$

where step (a) follows since each row $\mathbf{Q}_i$ is picked independently, hence $Z_{\ell,i}$ are also independent for different i.

Recall that the set $I(\mathbf{h}, a)$ contains all indices of entries of $\mathbf{h}$ taking value a. For a given $\mathbf{h}$, we can rewrite the product term as:

$$\prod_{i}^{n-k} \mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i|F(\mathbf{a})\} = \prod_{a=0}^{2q-2} \prod_{i\in I(\mathbf{h},a)} \mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a|F(\mathbf{a})\} \tag{7.114}$$

From Lemma 7.9 we know that each $\mathbf{p}_i(\ell)$ has distribution P_W. Hence for any $i \in I(\mathbf{h}, b)$ and any $\ell \in [1 : L]$:

$$\begin{aligned}
&\mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a | F(\mathbf{a})\} \\
&= \frac{\mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a, F(\mathbf{a})\}}{\mathbb{P}\{F(\mathbf{a})\}} \\
&= \frac{\mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a, \mathbf{p}_j(\ell) \in \{\mathbf{a}_j, \mathbf{a}_j + q\} \text{ for all } j \in [1 : n-k]\}}{\mathbb{P}\{\mathbf{p}_j(\ell) \in \{\mathbf{a}_j, \mathbf{a}_j + q\} \text{ for all } j \in [1 : n-k]\}} \\
&\overset{(a)}{=} \frac{\mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a, \mathbf{p}_i(\ell) \in \{\mathbf{a}_i, \mathbf{a}_i + q\}\}}{\mathbb{P}\{\mathbf{p}_i(\ell) \in \{\mathbf{a}_i, \mathbf{a}_i + q\}\}} \cdot \frac{\mathbb{P}\{\mathbf{p}_j(\ell) \in \{\mathbf{a}_j, \mathbf{a}_j + q\} \text{ for all } j \neq i\}}{\mathbb{P}\{\mathbf{p}_j(\ell) \in \{\mathbf{a}_j, \mathbf{a}_j + q\} \text{ for all } j \neq i\}} \\
&= \frac{\mathbb{P}\{\mathbf{p}_i(\ell) = a\}}{\mathbb{P}\{\mathbf{p}_i(\ell) \in \{a, a + q\}\}} \\
&= P_W(a) \cdot q
\end{aligned}$$

where step (a) follows from the fact that $\mathbf{h} \in \mathcal{F}(\mathbf{a})$ and $Z_{\ell,i}$ are independent for different i. The last step follows from the fact that $\mathbf{p}_i(\ell)$ has distribution P_W (established in Lemma 7.9) and it is easy to see that $\mathbb{P}\{\mathbf{p}_i(\ell) \in \{a, a + q\}\} = 1/q$ for all $a \in [0 : q - 1]$.

The interesting case is when $\mathbf{h}$ is a typical sequence in $\mathcal{A}_{[W]}^{(n-k)}$ hence $|I(\mathbf{h}, a)| = (n - k)(P_W(a) + o(1))$. We can continue as

$$\begin{aligned}
E(Z_\ell(\mathbf{h})|F(\mathbf{a})) &= \prod_{i=1}^{n-k} \mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i | F(\mathbf{a})\} && (7.115)\\
&= \prod_{a=0}^{2q-2} \mathbb{P}\{\mathbf{p}_i(\ell) = \mathbf{h}_i = a | F(\mathbf{a})\}^{|I(\mathbf{h},a)|} && (7.116)\\
&= \prod_{a=0}^{2q-2} (P_W(a) \cdot q)^{|I(\mathbf{h},a)|} && (7.117)\\
&= q^{\sum_{a=0}^{2q-2} |I(\mathbf{h},a)|} \prod_{a=0}^{2q-2} P_W(a)^{(n-k)(P_W(a)+o(1))} && (7.118)\\
&= q^{n-k} 2^{(n-k)(-H(W)+o(1))} && (7.119)\\
&= 2^{(n-k)(\log q - H(W)+o(1))} && (7.120)
\end{aligned}$$

Notice that $\mathbf{E}[Z_\ell(\mathbf{h})|F(\mathbf{a})]$ does not depend on ℓ asymptotically. Using Lemma 7.5 we have:

$$\begin{aligned}
\mathbf{E}[Z(\mathbf{h})|F(\mathbf{a})] &= \sum_{\ell=1}^{L} \mathbf{E}[Z_\ell(\mathbf{h})|F(\mathbf{a})] && (7.121)\\
&= L 2^{(n-k)(\log q - H(W)+o(1))} && (7.122)\\
&= 2^{k(2\log q - H(W)+o(1))+(n-k)(\log q - H(W)+o(1))} && (7.123)\\
&= 2^{n(R - H(W)+\log q+o(1))} && (7.124)\\
&= 2^{n(R - D(q)+o(1))} && (7.125)
\end{aligned}$$

To evaluate the variance, we first observe that (here we drop $\mathbf{h}$ for simplicity)

$$Z^2 = \left(\sum_{\ell=1}^{L} Z_\ell\right)^2 \tag{7.126}$$

$$= \sum_{\ell=1}^{L} Z_\ell^2 + \sum_{\ell \neq j} Z_\ell Z_j \tag{7.127}$$

$$= \sum_{\ell=1}^{L} Z_\ell + \sum_{\ell \neq j} Z_\ell Z_j \tag{7.128}$$

$$= Z + \sum_{\ell \neq j} Z_\ell Z_j \tag{7.129}$$

as $Z_\ell^2 = \prod_i Z_{\ell,i}^2 = \prod_i Z_{\ell,i} = Z_\ell$ for indicator functions. Furthermore

$$\mathbf{E}\left[Z^2 \middle| F(\mathbf{a})\right] = \mathbf{E}\left[Z|F(\mathbf{a})\right] + \sum_{\ell \neq j} \mathbf{E}\left[Z_\ell Z_j | F(\mathbf{a})\right] \tag{7.130}$$

$$\overset{(a)}{=} \mathbf{E}\left[Z|F(\mathbf{a})\right] + \sum_{\ell \neq j} \mathbf{E}\left[Z_\ell | F(\mathbf{a})\right] \mathbf{E}\left[Z_j | F(\mathbf{a})\right] \tag{7.131}$$

$$\leq \mathbf{E}\left[Z|F(\mathbf{a})\right] + \mathbf{E}\left[Z|F(\mathbf{a})\right]^2 \tag{7.132}$$

where step (a) follows since Z_ℓ, Z_j are conditionally independent for $\ell \neq j$, conditioned on the event $F(\mathbf{a})$. Hence we have

$$\mathbf{E}\left[(Z - \mathbf{E}\left[Z|F(\mathbf{a})\right])^2 | F(\mathbf{a})\right] = \mathbf{E}\left[Z^2 | F(\mathbf{a})\right] - \mathbf{E}\left[Z|F(\mathbf{a})\right]^2 \tag{7.133}$$

$$\leq \mathbf{E}\left[Z|F(\mathbf{a})\right] + \mathbf{E}\left[Z|F(\mathbf{a})\right]^2 - \mathbf{E}\left[Z|F(\mathbf{a})\right]^2 \tag{7.134}$$

$$= E(Z|F(\mathbf{a})) \tag{7.135}$$

8 Conclusion

In this thesis, we studied coding techniques with structured codes in communication networks. For Gaussian networks, we generalized the compute-and-forward scheme to incorporated CSI at transmitters, and proposed the novel compute-forward multiple access (CFMA) scheme, as a low-complexity alternative to other multiple access techniques. Various coding schemes based on lattice codes are also devised for several communication networks. These schemes either improve upon best known results for such networks, or recover known results with simpler decoder architectures. Since the main theme of the thesis concerns with decoding the sum of codewords of structured codes, the typical sumset of linear codes is introduced and several asymptotic results are given.

We conclude the thesis with two general research directions:

- **Beyond linear functions.** Computing the sum of codewords is a natural choice for additive channels with linear codes, but it is by no means the only meaningful choice for general communication networks. For example, computing the product of two codewords could be a preferred choice, if the channel is multiplicative than additive. Based on the existing results on typical sumsets of linear codes and proof techniques, we could characterize the asymptotic size of "typical images" under other (more general) functions. These results could be useful when we analyze possible coding schemes which involve computing nonlinear functions.

- **Converse on computation rates.** The definition of computation rates is subtle because it involves the function to be computed, making this concept more complicated (and less elegant) compared to the usual definition of achievable rates. In particular, a subset of functions should be specified, if we want to give any meaningful outer bound (or converse results) on the computation rates (otherwise computing a constant function always has infinite computation rates). However, which subset of functions should be chosen is an open-ended question and has not been studied carefully. Even with a specific set of functions, there is no standard technique to prove converse results for achievable computation rates. In particular, new inequalities relating the entropy of

random variables and entropy of functions should be established and will be crucial to the problem.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Tehnical Journal*, vol. 27, pp. 379–423, 1948.

[2] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *Communications, IEEE Transactions on*, vol. 44, no. 10, pp. 1261–1271, 1996.

[3] R. G. Gallager, "Low-density parity-check codes," *Information Theory, IRE Transactions on*, vol. 8, no. 1, pp. 21–28, 1962.

[4] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3051–3073, 2009.

[5] C. E. Shannon and others, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, vol. 1. Citeseer, 1961, pp. 611–644.

[6] A. El Gamal and Y. H. Kim, *Network information theory*. Cambridge University Press, 2011.

[7] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, 2000.

[8] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, 2011.

[9] J. Zhu and M. Gastpar, "Asymmetric compute-and-forward with CSIT," in *International Zurich Seminar on Communications*, 2014.

[10] ——, "Gaussian (dirty) multiple access channels: A compute-and-forward perspective," in *2014 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 2949–2953.

[11] ——, "Multiple Access via Compute-and-Forward," *submitted to IEEE Transactions on Information Theory*, Jul. 2014, in revision, arXiv: 1407.8463.

[12] ——, "Lattice Codes for Many-to-One Interference Channels With and Without Cognitive Messages," *Information Theory, IEEE Transactions on*, vol. 61, no. 3, pp. 1309–1324, 2015.

[13] ——, "Compute-and-forward using nested linear codes for the Gaussian MAC," in *IEEE Information Theory Workshop (ITW)*, 2015.

[14] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[15] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge University Press, 2014.

[16] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3401–3416, 2005.

[17] U. Erez and R. Zamir, "Achieving 1/2 log (1+ SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2293–2314, 2004.

[18] W. Nam, S.-Y. Chung, and Y. H. Lee, "Nested lattice codes for Gaussian relay networks with interference," *IEEE Trans. Inf. Theory*, vol. 57, 2011.

[19] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 273–278, Jan. 1998.

[20] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

[21] R. Ahlswede, "Multi-way communication channels," in *Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2-8, 1971*, 1973.

[22] H. H.-J. Liao, "Multiple access channels." Ph.D. dissertation, Dept. Elec. Eng., Univ. of Hawai, Tech. Rep., 1972.

[23] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within 1/2 bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, 2010.

[24] U. Niesen and P. Whiting, "The degrees of freedom of compute-and-forward," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5214–5232, 2012.

[25] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.

[26] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, 1996.

[27] O. Ordentlich, U. Erez, and B. Nazer, "The Approximate Sum Capacity of the Symmetric Gaussian K-User Interference Channel," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3450–3482, Jun. 2014.

[28] ——, "Successive integer-forcing and its sum-rate optimality," in *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, 2013.

[29] B. Nazer, "Successive compute-and-forward," in *International Zurich Seminar on Communications*, 2012, p. 103.

[30] T. Philosof, R. Zamir, U. Erez, and A. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, 2011.

[31] A. Somekh-Baruch, S. Shamai, and S. Verdu, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct. 2008.

[32] S. Kotagiri and J. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, no. 1, Mar. 2008.

[33] I.-H. Wang, "Approximate capacity of the dirty multiple-access channel with partial state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2781–2787, May 2012.

[34] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[35] S. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problemy Pered. Inf. (Probl. Inf. Trans.)*, vol. 9, 1980.

[36] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[37] M. Wilson, K. Narayanan, H. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, 2010.

[38] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-Forcing Linear Receivers," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[39] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," arXiv e-print, Jan. 2013.

[40] J. W. S. Cassels, *An introduction to Diophantine approximation.* University Press Cambridge, 1957.

[41] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, 2008.

[42] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *Information Theory, IEEE Transactions on*, vol. 52, pp. 1813–1827, May 2006.

[43] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with one cooperating transmitter," *European Transactions on Telecommunications*, vol. 19, pp. 405–420, 2008.

[44] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, pp. 3945–3958, 2009.

[45] S. Rini, D. Tuninetti, and N. Devroye, "Inner and outer bounds for the Gaussian cognitive interference channel and new capacity results," *Information Theory, IEEE Transactions on*, vol. 58, pp. 820–848, 2012.

[46] K. G. Nagananda, P. Mohapatra, C. R. Murthy, and S. Kishore, "Multiuser cognitive radio networks: an information-theoretic perspective," *International Journal of Advances in Engineering Sciences and Applied Mathematics*, vol. 5, no. 1, pp. 43–65, Mar. 2013.

[47] D. Maamari, D. Tuninetti, and N. Devroye, "Approximate sum-capacity of k-user cognitive interference channels with cumulative message sharing," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 3, pp. 654–666, Mar. 2014.

[48] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *Information Theory, IEEE Transactions on*, vol. 56, pp. 4566–4592, 2010.

[49] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *arXiv:1206.0197 [cs, math]*, Jun. 2012.

[50] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 53, pp. 4391–4399, 2007.

[51] S. S. Bidokhti, V. M. Prabhakaran, and S. Diggavi, "Is non-unique decoding necessary?" in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, 2012, p. 398–402.

[52] S. Sahraei and M. Gastpar, "Compute-and-forward: Finding the best equation," in *52nd Annual Allerton Conference on Communication, Control, and Computing, Champaign, Illinois, USA*, 2014.

[53] S. Sridharan, A. Jafarian, S. Vishwanath, and S. Jafar, "Capacity of symmetric K-user Gaussian very strong interference channels," in *IEEE Global Telecommunications Conference*, 2008.

[54] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, Jan. 1981.

[55] L. Wang, E. Sasoglu, and Y.-H. Kim, "Sliding-window superposition coding for interference networks," in *2014 IEEE ISIT*.

[56] N. Liu and S. Ulukus, "On the capacity region of the gaussian z-channel," in *IEEE GLOBECOM '04*, vol. 1.

[57] S. Vishwanath, N. Jindal, and A. Goldsmith, "The "z" channel," in *IEEE GLOBECOM '03*, vol. 3.

[58] L. Zhang, J. Jiang, and S. Cui, "Gaussian interference channel with state information," *IEEE Trans. on Wireless Communications*, Aug. 2013.

[59] R. Duan, Y. Liang, and S. Shamai, "On the capacity region of gaussian interference channels with state," in *2013 IEEE ISIT*.

[60] A. Padakandla and S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," in *ISIT*, Jul. 2013.

[61] B. Nazer and M. Gastpar, "Compute-and-forward for discrete memoryless networks," in *Information Theory Workshop (ITW)*, 2014.

[62] E. Sula, "Optimal channel input distributions for function computation in wireless network," in *Semester project report. EPFL-LINX*, 2014.

[63] I. Z. Ruzsa, "Sumsets and structure," *Combinatorial number theory and additive group theory*, pp. 87–210, 2009.

[64] D. Welsh, *Codes and cryptography*. Oxford University Press, 1988.

[65] T. Tao, "Sumset and Inverse Sumset Theory for Shannon Entropy," *Combinatorics, Probability and Computing*, vol. 19, no. 04, pp. 603–639, Jul. 2010.

[66] I. Kontoyiannis and M. Madiman, "Sumset and Inverse Sumset Inequalities for Differential Entropy and Mutual Information," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4503–4514, Aug. 2014.

[67] B. E. Fristedt and L. F. Gray, *A Modern Approach to Probability Theory*. Springer Science & Business Media, 1997.

[68] S. Roman, *Advanced linear algebra*. Springer, 2005.

Curriculum Vitae

Jingge Zhu

Education

2011 - 2016 **École Polytechnique Fédéral de Lausanne**, Lausanne, Switzerland
Docteur ès sciences
Advisor: Prof. Michael C. Gastpar

2008 - 2011 **Technische Universität Berlin**, Berlin, Germany
Diplom-Ingenieur, Technische Informatik

2008 - 2011 **Shanghai Jiao Tong University**, Shanghai, China
Master of Engineering, Telecommunication and Information System

2004 - 2008 **Shanghai Jiao Tong University**, Shanghai, China
Bachelor of Engineering, Information Engineering

Research experience

2011 - 2016 Laboratory for Information in Networked Systems, EPFL
Doctoral assistant

2008 - 2010 Fraunhofer Institute, Mobile Communication Institute, Berlin
Research assistant

Publications

Journal Papers and Manuscripts

- **Jingge Zhu** and Michael Gastpar, "Typical sumsets of linear codes", November, 2015, submitted, in *arXiv: 1511.08435*.

- **Jingge Zhu** and Michael Gastpar, "Multiple Access via Compute-and-Forward", July 2014, submitted to IEEE Transactions on Information Theory, in revision, in *arXiv: 1407.8463*.

- **Jingge Zhu** and Michael Gastpar, "Lattice Codes for Many-to-One Interference Channels With and Without Cognitive Messages", in IEEE Transactions on Information Theory, vol. 61, March, 2015.

- **Jingge Zhu**, Jianhua Mo and Meixia Tao, "Cooperative Secret Communication with Artificial Noise in Symmetric Interference Channel", *IEEE Comm. Letters*, vol. 14, no. 10, pp. 885-887, Oct. 2010.

  [2013 IEEE Heinrich Hertz Award for Best Communications Letters]

Conference Papers

- **Jingge Zhu** and Michael Gastpar, "On Lattice Codes for Gaussian Interference Channels", 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 2015.

- **Jingge Zhu** and Michael Gastpar, "Compute-and-Forward using nested linear codes for the Gaussian MAC", IEEE Information Theory Workshop (ITW), Jerusalem, Israel, 2015.

- **Jingge Zhu** and Michael Gastpar, "Gaussian (dirty) multiple access channels: a compute-and-forward perspective", IEEE International Symposium on Information Theory (ISIT), Honolulu, USA, 2014.

- **Jingge Zhu** and Michael Gastpar, "Asymmetric Compute-and-Forward with CSIT". International Zurich Seminar on Communications (IZS), Zurich, Switzerland, 2014.

- **Jingge Zhu** and Michael Gastpar, "Lattice codes for many-to-one cognitive interference networks", 2013 IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, 2013.

- **Jingge Zhu**, Sławomir Stańczak and Gunther Reißig, "Stabilization of Linear Dynamical Systems with Scalar Quantizers under Communication Constraints" , *Proc. 44th Annual Conf. Information Sci. and Systems (CISS)*, Princeton, NJ, U.S.A., Mar 17-19, 2010,