Fundamental Limits of Network Communication with General Message Sets: A Combinatorial Approach

by

Henry Paul Romero

B.S.,University of Colorado, 2007

M.S.,University of Colorado, 2012

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

2014

This thesis entitled:
Fundamental Limits of Network Communication with General Message Sets: A Combinatorial Approach
written by Henry Paul Romero
has been approved for the Department of Applied Mathematics

_____

Prof. Mahesh Varanasi

_____

Prof. Juan Restrepo

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Romero, Henry Paul (Ph.D., Applied Mathematics)

Fundamental Limits of Network Communication with General Message Sets: A Combinatorial Approach

Thesis directed by Prof. Mahesh Varanasi

The classical theoretical framework for communication networks is based on the simplifying assumption that each message to be sent is known to a single transmitter and intended for a single receiver. Modern communication protocols reflect this framework by treating the physical layer as a network of individual links. However, this wireline view of wireless communications fails to account for the heterogeneous nature of network demands, consisting of both unicast and multicast services, and can fail to leverage the inherent broadcast advantage of the wireless medium.

This thesis extends the classical framework of a private-message interface to the physical layer to one with both private and common messages. A key difficulty, in both the description and analysis of a communication model with general messages sets, is that there are combinatorially many message possibilities. With order-theoretic language and tools from combinatorial optimization and graphical models, we develop a general framework for characterizing the fundamental limits of information transfer over large many-to-one (multiple access) and one-to-many (broadcast) communication channels with general message sets. In particular, achievable regions are proposed for arbitrary such channels. For the multiple-access channel, the achievable region is optimal, and the order-theoretic perspective both unifies and extends previous results. For the broadcast channel, the region is specialized to an inner bound to the Degree of Freedom region, a setting where it is provably optimal in select cases.

This thesis provides fresh insights into the long-standing random coding technique of superposition coding to arrive at these results. Governing constraints on reliable communication through superposition coding are shown to be polymatroidal over a lattice of subsets that may not be the boolean lattice of all subsets. Permissible input distributions for superposition coding are concisely

characterized through directed graphical models of conditional dependencies. The two-user interference channel is also addressed, where the state-of-the art is extended from the case with two private messages to one with an additional common message.

Dedication

To my family,

for their unwavering support and patience,

and for inspiration to aim high.

Acknowledgements

I am fortunate to have met many intelligent and inspiring people during my graduate studies. First, I would like to thank my advisor, Prof. Mahesh Varanasi, for introducing me to the deep world of information theory and for his guidance over the past four and a half years in both the selection of research questions and the pursuit of meaningful answers. The many hours he spent with me helped to crystallize vague research ideas into a clear research agenda. His patient support and belief in me helped push me to discover more and understand better.

I am grateful to the members of my committee—J. Corcoran, V. Dukic, C. Mullis, and J. Restrepo—for their time and efforts in seeing this thesis through to completion. The Applied Mathematics department as a whole has been a wonderful academic home, with attentive faculty and a friendly and intelligent group of fellow graduate students, who helped both inside and outside the classroom from the start to the finish. The many teachers I had along the way taught beautiful courses, which revealed the breadth and depth of applied mathematics and were beneficial to me in many ways. Special thanks are also in order for the DSP group of graduate students, who helped me maintain focus and inspiration through the final years of graduate study.

I am fortunate to have received financial support during my graduate career from various sources: from the departments of Applied Mathematics and Electrical Engineering, from the National Science Foundation (NSF), and from the GAANN and PREP Fellowships offered through the Department of Education and the National Institute of Standards and Technology, respectively.

Last but not least, I would like to thank my family for their encouragement and support, especially to my wife Margaret, my parents Paolo and Marian, and my brother Charles.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction and Preliminaries

1.1 Motivation

In the modern age, with the abundance of wireless devices offering nearly ubiquitous high-speed connectivity to the internet and telephony network, it may be hard to imagine a time when high-speed wireless communication seemed improbable. Yet in the early part of the 20th century, this was the case: communication over radio waves, which are subject to obstructions, reflections, and thermal noise at the sending and receiving antennas, appeared to be intrinsically unreliable, as any transmitted **analog** signal could not be perfectly reconstructed at the receiver.

In an influential paper in 1948, the Bell Labs mathematician Claude Shannon changed the discussion by observing that the picture is considerably more optimistic when one considers only **digital** signals [95]. In the same way that the outcome of 10,000 fair coin flips can be more reliably estimated than the outcome of 10 fair coin flips, the corruptions introduced in wireless communications are entirely predictable when the wireless channel is used repetitively. Moreover, the mathematics assured that a communication link can be designed in a modular fashion, with **bits** serving as a universal interface between modules.

This insight provided the courage and vision for academics and engineers since to construct high-speed **digital** and **modular** communications networks in the 65 years since this publication. The networks which underlie both the telephony and the internet are layered—as the bit is a universal interface for the purposes of point-to-point communication, we know that each layer can be designed independently with the assurance that the separate layers will work together reliably if

the bit is the interface of choice. In wireless communications, improvements within the physical layer have allowed the mobile telephony network to grow from a small, low-rate, voice data transmission network into a large, high-speed, general-purpose data transmission network. Modern channel codes now approach the fundamental physical limits for a single wireless channel discovered through Shannon's work: improvement cannot come by focusing on each communication link individually. The uncertainty of wireless wave propagation and reception, with little competition for wireless resources, no longer serves as a rate-limiting factor.

In its place, scarcity of wireless resources and competition for these resources serve as the principal modern rate-limiting factors. Spectrum is no longer an open frontier and attention must be paid to efficiently using available spectrum. Though wireless communication is well-understood at the the level of a single isolated point to point communication link, much less is understood when viewing communication from a broader perspective—that of a network communication problem. The modern wireless telephony network breaks the network architecture into modular components known as cells, where a central base station communicates to and from mobile users in its corresponding cell:

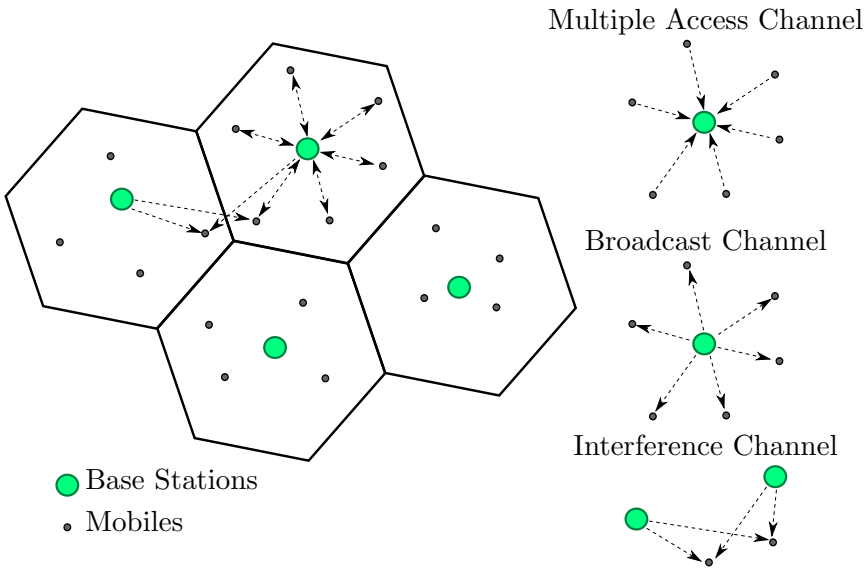


Figure 1.1: Centralized communication network

In each cell, there are two communication directions: the uplink, or multiple access channel (MAC), where the mobile users communicate simultaneously to the base station, and the downlink, or the broadcast channel (BC), where the base station communicates simultaneously to the mobile users. Two mobile users at the fringe of two neighboring cells may have their transmissions to and from their respective base stations interfere with each other; the mathematical model for this four-terminal situation (two base stations, two mobiles) is known as the interference channel (IC).

In the tradition of Shannon's work, the laws of information flow over these channels are studied with the universal interface of a bit. With multiple terminals, there are multiple interfaces. A practical and simplifying assumption is to assign an independent interface to each independent mobile user. In the MAC, this equates to assuming that each mobile user has a private message, unknown to the other users, to send to the base station. In the BC, it equates to the base station having distinct messages for distinct mobile users, not to be desired by more than one mobile user. This private message view has filtered into networking, where the default interface to the physical layer is one of private messages.

In the language of networking, a messaging service with only one source and one destination is a unicast service. An alternative, where messaging is one-to-many or many-to-one, is known as a multicast service. Currently, the task of switching between the two is relegated to higher layers than the physical layer. However, doing so incurs a loss in optimality. As wireless signals are broadcasted, the wireless setting offers intrinsic advantages for multicast services; this is known as the broadcast advantage. There is potentially much to be gained by considering a richer physical layer interface, with both unicast and multicast services offered natively and simultaneously. For example,

- Streaming media (such as mobile TV), peer-to-peer services, and large-scale software updates [56] indicate that an interface for the downlink should accommodate multicasting in addition to the unicasting currently natively offered at the physical layer. While upcoming specifications, such as MBMS (Multimedia Broadcast Multicast Services) for 3GPP cellular

networks, provision resources for separate multicasting and unicasting modes, it would be better yet to have a interface that could accommodate both modes simultaneously.

- Cooperation, or feedback, allows mobile users to infer information about each other's messages. Such iterative communication can be constructed as a multi-step procedure, with the initial step serving to construct common messages and with the final step serving to transmit the constructed common jointly with what parts of the original messages remain private. Cooperative schemes allow higher data rates than are achievable without cooperation or feedback.

- If correlated data, as produced with sensor networks, have correlations which take the special form of common parts, then a transmission scheme which accounts for this common data performs better than one that does not.

The objective of this thesis is to address such a general interface by considering the classical K-to-one, one-to-K, and two-to-two models of communication (the MAC, BC and IC, respectively) from an under-developed perspective: that of general message sets.

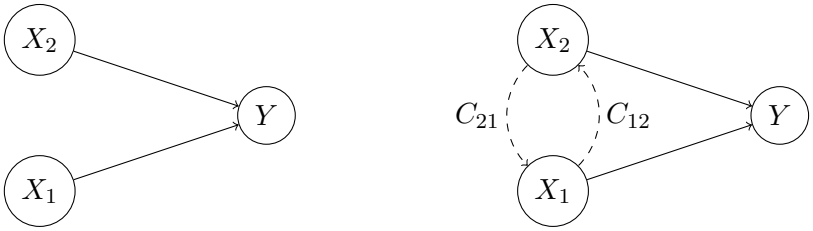
1.2 Prior Work

While a great deal of scholarly attention has been directed towards the study of either multiple-unicast or single multicast, much less work has focused on the consideration of both unicast and multicast together.

Early work in network information theory took up the task, with hopes for a general theory of network information flow to be tractable. Perhaps the earliest example is in 1973, when Slepian and Wolf [99] consider the the two-user MAC where each user has a private message, unknown to the other user, and both share knowledge of a third common message. They provide matching inner and outer bounds to the capacity region, and thus establish the capacity region. Achievability is shown through superposition coding and the analysis of error is provided with Gallager's error exponent guarantees [35], an approach too burdensome to permit a generalization to more than two users.

Though Slepian and Wolf postulate a three-user generalization, this guess was later demonstrated to be incorrect (by Gel'fand, see Prelov [80]). In 1979, Han [48] focuses on the simpler arguments afforded by Cover's weak typicality [22] to correctly deduce a capacity generalization to the K-user MAC. He is the first to observe that the resultant capacity characterization is expressible as a union of **polymatroids** [48], which are polytopes with special properties. Polymatroids were defined in the 1970's as a means of unifying, and more deeply understanding, the theory on why some combinatorial optimization problems permit greedy algorithmic solutions [28]. This connection between information theory and combinatorial optimization bore fruit immediately with the introduction in 1981 of Han and Kobayashi's inner bound to the IC [47], which remains to this day the largest largest known inner bound to the IC with two unicast services corresponding to disjoint transmitter-receiver pairs. Despite its early promise, this connection was not to be further investigated until the mid 1990s and early 2000s, when scholarly work turned to characterizing the capacity boundary of multiple unicast communication over the Gaussian MAC and BC.

Slepian and Wolf's results found application as part of a more complex cooperative communication scheme in the work of Willems' [119]. There, two users of a MAC are modeled as having two noiseless, but rate-limited, conferencing links between them, which operate independently of each other and of the channel. By first sharing data over the conferencing links and collecting



Two-user MAC without cooperation. Two-user MAC with cooperation

Figure 1.2: Willems' conferencing encoders for the two-user MAC

the shared knowledge into a common message, and then transmitting with common and private data, communication rates strictly higher than those available without cooperation are attained. Moreover, the channel affords a tight converse argument to provide that the proposed inner bound

is the capacity region.

In 1980, Gel'fand and Pinsker [36] consider a general message set for the two-user BC. Achievability is shown at each vertex of their achievable region, where superposition and binning, as in Marton's bound for the private message case [69], are used. Though this inner bound remains the largest inner bound known, there is no general argument for its optimality, a stark contrast to the situation with the MAC. It took 27 years for the Gaussian specialization of Marton's bound (for the multiple unicast case) to be proven to be optimal for the K-user Gaussian BC where each terminal may have multiple antennas [116]. Moreover, only within the past year did it become known that two-user Gel'fand-Pinsker inner bound is also optimal for the Gaussian case where each terminal may have multiple antennas [37]. In both cases, the outer bounds required quite a bit of ingenuity and used tools specific to the Gaussian case.

For analytical simplicity, scholarly work drifted away from consideration of general message sets and to a focus on the multiple-unicast, or private message, case. This is in part as the tools were easily generalizable. Ahlswede and Liao [1] characterize the capacity region for the MAC with private messages. In the mid 1990s, it was observed that to achieve any maximal weighted sum-rate (the total capacity) on the boundary of the multiple access channel with private messages, it suffices to use successive decoding at the decoder [106, 123]. This has practical importance, as successive decoding can be implemented with much lower complexity, as well as analytical importance, as the explicit formula for the vertices attainable by successive decoding lead to convex efficient formulations for the computation of optimal input distributions.

A similar observation can be made for largest known inner bound the broadcast channel with private messages, Marton's region. It too is a polymatroid, and its vertices have an explicit representation. Depending on the choice of input distribution, these vertices may be attainable by successive Gel'fand Pinkser encoding [123]. The connection between high complexity joint decoding (or encoding) schemes and low-complexity successive decoding (or encoding) schemes can be more systematically developed through polymatroids, and the dual notion of contra-polymatroids [123]. Under this umbrella, we may also fit the Berger-Tung inner bound [7, 108] for the lossy source

coding, which is contra-polymatroidal.

Recent work on Gaussian channel models with multiple antennas at input and output terminals (referred to as **M**ultiple **I**nput **M**ultiple **O**utput, or MIMO) have further exploited the fact that capacity of the MAC with private messages is expressible as a union of polymatroids. Efficient optimal resource allocations for the fading MIMO MAC [71] build on the work of Tse and Hanly [106]. The recent observation of a MAC-BC duality theory [111, 55, 110] in the private message case has permitted many of these observations to extend to the BC [71].

Recently, there has been work on the multicast situation in isolation for the broadcast channel. A seminal work is that of Jafar and Zhi-Quan [54], who compare and contrast isotropic, beamforming, and orthogonal multicast transmission schemes. The latter corresponds to relegating multicasting to higher network layers, and the authors find this to be strictly suboptimal. The other choices instead correspond to different strategies for native physical layer multicasting, and are optimal in certain regimes. Another seminal work is that of Wiesthelier et al [117], who demonstrate that multi-hop multicast can be done with less transmit power if the protocol takes into account the broadcast nature of wireless communication.

A lesson from the broadcast channel is that exact capacity characterizations for general multi-terminal networks are elusive. Despite all the years of progress, there remain two-user broadcast channels with unknown capacity region. Another similarly simple network which has unknown capacity region is the two-user interference channel, even though its study dates back to Shannon in 1961 [96]. For the interference channel, consideration of just two unicast messages with disjoint transmitter-receiver pairs has proven to be immensely difficult, thus discouraging consideration of more general message sets.

Recent developments have broken this gridlock by observing that searching for capacity approximations can be far more fruitful than searching for exact capacity characterizations. While capacity is unknown for the two-user Gaussian interference channel, a "strong" approximate characterization of capacity is known [33, 57], indicating that Han-Kobayashi's inner bound is at least nearly equal to capacity, if not equal to it. Moreover, all interference channels of a semi-determinisitic

character [103] exhibit this "strong" characterization of capacity. With the confidence engendered by the revelation that capacity approximations can be much more tractable than capacity itself, scholarly work on interference networks has begun to shift towards more general settings, with more users (e.g., the K-user interference channel) or with more general message sets. An example with more general message sets is a setting with a common multicast message, known to both transmitters and desired by both receivers, in addition to the two unicast messages with disjoint transmit-receiver pairs. Initially studied in 1980 by Tan [101], Jiang-Xin-Garg [53] strictly improved on Tan's inner bound by incorporating ideas from Chong et al's modern treatment [17] of the classic Han-Kobayashi [47] achievability scheme.

1.3 Organization and Original Contributions

In Chapter 2, we introduce the essentials of our mathematical models of communication as well as the relevant concepts from order theory, graphical models, and combinatorial optimization.

In Chapter 3, we revisit the classical theoretical technique of superposition coding through the lens of order theory. Doing so reveals a succinct and useful combinatorial structure in the resultant conditions for reliable communication: that of a polymatroid. The language and tools here form the basis for much of the work in this thesis. We apply the order-theoretic perspective to form an inner bound to the discrete memoryless MAC with an arbitrary number of users, K, and an arbitrary collection of messages. The formulation allows for a simultaneous treatment of different capacity formulations and of varied message sets - the full boolean lattice of all possible messages, or a chain of degraded messages, or an antichain of private messages.

In Chapter 4, we specialize the results from the discrete memoryless case to the Gaussian setting, of practical interest for wireless communications. The principal technical result therein, a generalization of the maximum entropy, may be of independent interest to other information theoretic problems. Further, we leverage the beneficial polymatroid structure induced by superposition coding, and the convex properties of the mutual information bounds, to provide a schematic for efficient numerical methods to calculate the optimal operating covariances.

In Chapter 5, we study the Broadcast Channel, where the situation is more subtle involving difficulties in the development of both the inner and outer bounds. To gain insight, attention is first focused on the approximate measure of capacity known as Degrees of Freedom (DoF). There, an inner bound based on recursive row selection is proposed, where constraints imposed on the recursive procedure are polymatroidal. When the recursive selection leads to a matrix factorization without antenna selection, the proposed inner bound is optimal; but when the inner bound requires antenna selection, the inner bound can be suboptimal. To improve on this bound, focus is shifted to the general discrete memoryless channel, where a generalization of the largest known inner bound for the two-user channel is extended to the K-user case. The prior theory for superposition coding is leveraged, and the technique of **binning** is extended allow recursive codeword generation. This bound is demonstrated to replicate the capacity of the 2 or 3-user deterministic combination networks, and improves on the prior DoF region.

In Chapter 6, a two-user semi-deterministic IC with two private messages and one common message is studied. Determining the capacity region of the general discrete memoryless IC and its Gaussian specialization is a long-standing open problem in information theory. Recent progress has focused on constant gap results. To work towards a general message interface for the IC, we extend a constant gap result for the semi-deterministic model with two private messages to one with two private messages and one common message.

Chapter 2

Mathematical Models and Tools

2.1 Introduction

This chapter introduces the mathematical abstractions of one-to-one, many-to-one, and one-to-many communication. We introduce communication in a general setting, with the goal of characterizing fundamental limits that cannot be broken no matter how smart the transmitters or receivers are.

Our framework includes general message sets, containing both private messages (which are known at a single transmitter and desired at a single receiver) and common messages (which are either known at many transmitters or desired by many receivers). A well-known strategy for coding with common and private messages is to code with superposition, as originally introduced by Cover [21] for the BC. In superposition coding, a common message is viewed as carrying coarse details of a total message set while a private message carries finer details of that message set. To paraphrase, the message sources of the total message set are partially ordered according to the level of detail each source carries about the total message set. Though seemingly simple, this partial order ties together ideas from graphical models of conditional independence, combinatorial optimization, and superposition coding to yield insights that to the best of our knowledge have gone unnoticed within the information theory community.

2.1.1 Point-to-Point Channel

An abstract model of communication from a single transmitter $X \in \mathcal{X}$ to a single receiver $Y \in \mathcal{Y}$ can be modeled via a probability transition function $P(Y = y|X = x)$, succinctly referred to with $p(y|x)$. If the output of a single channel use is a corrupted version of the input, then at best we can have an unreliable estimate of the original input.

To combat this unreliability, consider using the channel repetitively. In this case, the set of total information to be sent accumulates as a stream of information. For simplicity, take the unit of information to be a bit, so that our stream is a string of bits `0101110`.... These bits are each assumed to carry meaningful information: each new bit carries information which refines the knowledge to be sent. If there are M bits to send, then the sending the bits sequence $b_1, \ldots, b_M$ is equivalent to sending an integer within $\{1, \ldots, 2^M\}$. If we can recover any M bit sequence after n channel uses, then the effective communication rate is $R = M/n$. Sending information at a constant rate R implies that the total number of messages grows exponentially in the number of channel uses. With bit as the unit of measure, there are $W = \lfloor 2^{nR} \rfloor$ messages to send.

To inquire into fundamental limits, we formalize the notion of a arbitrary transmission strategy as a code consisting of mappings that may be arbitrarily complex. Formally, a code of block length n consists of a mapping from the message set to an input sequence (the encoder) and from the channel output sequence and a message estimate (decoder) as summarized below.

Point-to-point code		
Encoder:	$e_n : [1 : W] \mapsto \mathcal{X}^n$	mapping messages to inputs
Decoder:	$d_n : \mathcal{Y}^n \mapsto [1 : W]$	mapping channel outputs to message estimates

Table 2.1: Interface for the Point-to-Point Channel

Let M be a random variable uniformly distributed over $\mathcal{W} = [1 : W]$. The figure of merit for a code will be the average error probability $P(M \neq \hat{M})$, where $\hat{M} = d_n(Y^n)$ when $X^n = e_n(M)$. A communication rate R is said to be achievable if, for every $\epsilon > 0$, there exists a block length n and corresponding code such that $W \geq 2^{n(R-\epsilon)}$ and $P(M \neq \hat{M}) < \epsilon$.

Index the channel inputs and outputs of the tth channel use as X_t, Y_t. A reasonable assumption is that the channel uses are independent of each other (that is, the channel is memoryless), in which case the probability transition function is

$$p(y^n|x^n) = \prod_{t=1}^{n} p_{Y|X}(y_t|x_t).$$

The classic result of information theory is that the set of non-negative achievable rate satisfy [23]

$$R \leq \sup_{p(x)} I(X;Y),$$

where $I(X;Y)$ is the mutual information between X and Y:

$$\text{(discrete)} \qquad I(X;Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

$$\text{(continuous)} \qquad I(X;Y) = \int_{x,y} f(x,y) \log \left(\frac{f(x,y)}{f(x)f(y)} \right) \mathrm{d}x\mathrm{d}y \ ,$$

where $p(x,y)$ is a probability mass function and $f(x,y)$ is a probability density function.

2.1.2 Random Coding

How do we interpret the above result? From the definition of a code, we see that we have set the problem into one of high-dimensional geometry (specifically, n-dimensional, where n may be arbitrarily large). Sending data reliably amounts to picking a set of codewords (the encoder's output) within the space of all possible inputs $\mathcal{X}^n$ so that the high probability images of the received codewords overlap with low probability. A constructive procedure for choosing such codewords is not at all obvious: doing so has has been the subject of coding theory for more than a half century. But an argument for the existence of such codewords can be had relatively easily through random coding. Picking the encoder randomly leads to an average performance that is good enough, thus ensuring that at least one choice performs well.

Let's formalize this in the discrete memoryless (DM) context (where we take the input and output alphabets to be finite). Fix an input distribution $p(x)$ over the input alphabet $\mathcal{X}$. For each

message $m \in [1 : W]$, pick codewords independently and identically according to

$$x^n(m) \sim \prod_{i=1}^{n} p(x_i) \ .$$

Though the best decoding rule would be a maximum likelihood decoder, another rule—joint typicality decoding—sufficies to attain capacity. A sequence $z^n \in \mathcal{Z}^n$ is typical with respect to a distribution $p(z)$ if its empirical distribution closely matches it (see Figure 2.1). Formally, z^n is

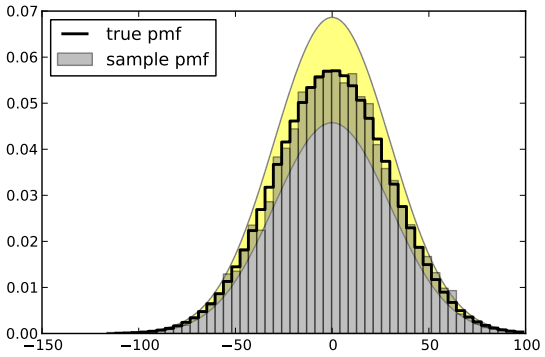


Figure 2.1: Depiction of the notion of a typical sequence.

ϵ-typical if

$$|\pi(z|Z^n) - p(z)| < \epsilon p(z) \quad \forall z \in \mathcal{Z} \tag{2.1}$$

where $\pi(z|Z^n) = \frac{1}{n}\sum_{i=1}^{n} 1(Z_i = z)$ is the sample histogram, and where $\epsilon > 0$ is a small parameter. Let $\mathcal{T}_\epsilon^{(n)}(Z)$ be the set of all sequences z^n which satisfy (2.1). An i.i.d. sequence $z^n \sim \prod_{i=1}^{n} p_Z(z_i)$ is likely to

- be ϵ-typical with their generating distribution: $P(Z^n \in \mathcal{T}_\epsilon^{(n)}(Z)) \to 1$, and
- not be ϵ-typical with a different distribution: $P(Z^n \in \mathcal{T}_\epsilon^{(n)}(W)) \leq 2^{-nD(W\|Z)(1-\epsilon)} \to 0$

where $D(W\|Z) = \sum_z p_W(z) \log(p_W(z)/p_Z(z))$ is the Kullback-Leilber distance between the generating distribution on Z and a different distribution on W.

These two facts are the crux of the argument that random codes perform well. In particular, if the codeword $x^n(m)$ is sent over the channel, then the joint distribution of $(x^n(m), y^n)$ has

distribution $\prod_{i=1}^{n} p_X(x_i)p_{Y|X}(y_i|x_i)$. By contrast, for any $m' \neq m$, the joint distribution of the codeword $x^n(m')$ and the output y^n is $\prod_{i=1}^{n} p_X(x_i)p_Y(y_i)$. We decode the message by declaring $\hat{m}$ to be the message sent only if it is the only message m such that $(x^n(m), Y^n)$ is ϵ-jointly typical. Let E_m be the event that the codeword $x^n(m)$ is ϵ-jointly typical with the channel output sequence Y^n. Then this decoding rule fails whenever the sent message has a codeword $x^n(m)$ is not ϵ-jointly typical with the output y^n or when a message that was not sent has a codeword $x^n(m')$ which is ϵ-jointly typical with the output y^n. That is, the error event is

$$\text{Error} = E_m^c \bigcup \left(\cup_{m' \neq m} E_{m'}\right).$$

By the properties of typical sequences, it will be highly likely that the desired codeword is jointly typical with the output,

$$P(E_m) \to 1 \qquad (\text{ as } n \to \infty),$$

while no undesired codeword will be likely to be jointly typical with the output if $R < I(X;Y)(1-\epsilon)$, where $I(X;Y) = D(p(x,y)\|p(x)p(y))$ as

$$P(\cup_{m' \neq m} E_{m'}) \leq \sum_{m' \neq m} P(E_{m'}) \leq \sum_{m' \neq m} 2^{-nI(X;Y)(1-\epsilon)} \leq 2^{n(R-I(X;Y)(1-\epsilon))} \to 0 \qquad (\text{ as } n \to \infty)\ .$$

Thus, in expectation, random coding achieves any point $R < I(X;Y)$. Maximizing over input distributions then yields the largest achievable rate region. This last point provides an important guideline that those that design codes have tried to emulate - a good code should have codewords which have empirical statistics closely matching this maximizing input distribution. In particular, for the Gaussian point-to-point channel,

$$y = hx + z \qquad z \sim \mathcal{N}(0, \sigma^2), \tag{2.2}$$

where the input codeword $x_1, \ldots, x_n$ is constrained to have finite average power ($\frac{1}{n}\sum_{i=1}^{n} |x_i|^2 \leq P$), and where the receiver has knowledge of the multiplicative coefficient h (known as the fading state), the maximizing distribution is known to be itself Gaussian. As Gaussian channel (2.2) is a good model for wireless communication, a guiding light for many code designers has been to design codes

to have Gaussian empirical distributions so that achievable rates approach the channel's capacity, given by the well-known formula

$$C = \log(1 + \mathtt{snr}) \qquad \mathtt{snr} = |h|^2 P/\sigma^2.$$

2.1.3 Multiple Access-Channel Model with General Message Sets

The above model of communication can be extended easily to many-to-one and one-to-many models of communication. The former case is the MAC, where K transmitters which wish to communicate simultaneously to a common receiver. The channel has K finite input alphabets $\mathcal{X}_j, j \in \{1, \ldots, K\}$, a finite output alphabet $\mathcal{Y}$, and a probability transition function

$$p(y^n|x_1^n, \ldots, x_K^n) = \prod_{t=1}^{n} p_{Y|X_1,\ldots,X_K}(y_t|x_{1t}, \ldots, x_{kt})$$

where $x_{j,t}$ is the jth user's input at the tth channel use and y_t is the output of the tth channel use.

Classically, this channel has been studied under the assumption that each user only has knowledge of a private message, in which case the capacity region of the MAC may be characterized via the set of probability distributions that factorize as the product of the input distributions of the users [1]. A more general model is to consider not only private messages, each known only to single user, but also common messages, with each known to potentially many transmitters. A general such message set would then be a collection of M independent common messages, where each unique message is revealed to a unique subset of the K users. Though previously studied in [99, 48, 43], we adopt a slightly different notation than adopted previously: rather than simply enumerating the available messages, we index each message by the subset of transmitters which is cognizant of that message. This subtle difference emphasizes the partial order inherent in knowledge of common messages at the transmitters and provides for a more fluent presentation of the associated capacity results.

To be precise, we consider a collection E of M distinct non-empty subsets $[1:K]$,

$$E = \{S_1, \ldots, S_M\} = \{S_j \subseteq [1:K] : \emptyset \neq S_j \neq S_i \ \forall i, j \in [1:M]\}, \tag{2.3}$$

where to each element $S \in E$, we assign an independent message source $M_S \in [1 : W_S]$, which we reveal only to the transmitters listed in S. A code for the transmitters and receiver are a collection of K encoders, with the jth encoder mapping messages known to jth user to a input sequence for the jth user and a decoder, mapping the observed channel output to an estimate for all of the sent messages, as summarized in Table 2.2.

Multiple Access code	
Encoder 1:	$e_1 : \prod_{1 \in S \in E} [1 : W_S] \mapsto \mathcal{X}_1^n$
$\vdots$	$\vdots$
Encoder K:	$e_K : \prod_{K \in S \in E} [1 : W_S] \mapsto \mathcal{X}_K^n$
Decoder:	$d : \mathcal{Y}^n \mapsto \prod_{S \in E} [1 : W_S]$

Table 2.2: Interface for the Multiple Access Channel with General Message Sets

For each $S \in E$, let $\hat{M}_S$ be the decoder's estimate for the message M_S. Then an error occurs if $M_S \neq \hat{M}_S$ for any $S \in E$. With each $M_S \sim \text{Uniform}([1 : W_S])$ independently for all $S \in E$, define the average probability of error for a given block length n, decoder, and set of encoders to be $P_e^{(n)} = P(\cup_{S \in E} \{M_S \neq \hat{M}_S\})$. Then, a rate tuple $(R_S : S \in E)$ is achievable if, for every $\epsilon > 0$, there is a block length n, set of encoders, and decoder such that $W_S \geq 2^{n(R_S - \epsilon)}$ for all $S \in E$ and $P_e^{(n)} < \epsilon$. The capacity region, in turn, is defined as the closure of the set of achievable rate tuples.

2.1.4 Broadcast Channel Model with General Message Sets

As with the MAC, a similar definition can be made for the discrete memoryless BC. It is defined in terms of a single finite input alphabet $\mathcal{X}$ and K finite ouput alphabets $\mathcal{Y}_j, j \in \{1, \ldots, K\}$, a finite output alphabet, and a probability transition function

$$p(y_1^n, \ldots, y_k^n | x^n) = \prod_{t=1}^{n} p_{Y_1, \ldots, Y_K | X}(y_{1t}, \ldots, y_{kt} | x_t)$$

where x_t is the input at the tth channel use and y_{jt} is the jth users' output at the tth channel use.

As in the MAC, we consider a general message set, with both private and common messages. With E as a set of M distinct non-empty subsets $[1 : K]$, each element $S \in E$ is assigned to an independent message source $M_S \in [1 : W_S]$ desired only to the transmitters listed in S. A code

for the transmitters and receiver are a single encoder, mapping the set of all messages to an input sequence, and a collection of K decoders, with the jth decoder mapping the output sequence observed by the jth user to a set of message estimates for those messages desired by the jth user. This is summarized in Table 2.3.

Broadcast code	
Encoder:	$e : \prod_{S \in E} [1 : W_S] \mapsto \mathcal{X}^n$
Decoder 1:	$d_1 : \mathcal{Y}_1^n \mapsto \prod_{S \in E : 1 \in S} [1 : W_S]$
$\vdots$	$\vdots$
Decoder K:	$d_K : \mathcal{Y}_K^n \mapsto \prod_{S \in E : K \in S} [1 : W_S]$

Table 2.3: Interface for the Broadcast Channel with General Message Sets

2.2 Superposition Coding Preliminaries

For both the MAC and BC, a strategy to prove the achievability of certain rate tuples relies on superposition coding, where some message codewords are superposed on top of other message codewords. Intrinsic to this is a notion of order. For example, consider two distinct messages sources $M_S, M_{S'}$ with indices ordered by inclusion: $S \subset S'$. In the MAC, wherever M_S is known among the transmitters, so too is $M_{S'}$. Hence, one may choose to construct the codeword of M_S with the knowledge of the codeword constructed for $M_{S'}$ in mind. Analogously, if the context is the BC, then wherever the message M_S is demanded among the receivers, so too is $M_{S'}$. Hence, as decoding of M_S always involves decoding $M_{S'}$, one may choose to construct the codeword for M_S in a manner such that its correct decoding is dependent on the correct decoding of the codeword for $M_{S'}$. In both cases, the suggestion is that the construction of the codewords for the messages corresponding to the subset S and S' could occur in an ordered fashion: the codewords corresponding to S' could be constructed before the codewords corresponding to S. Motivated by these observations, we introduce a framework from the general theory of order [24].

2.2.1 Order Theory

For a set P, an order on P is a binary relation $\leq$ on P such that for all $x, y, z \in P$

(i) $x \leq x$ (ii) $x \leq y$ and $y \leq x$ imply $x = y$ (iii) $x \leq y$ and $y \leq z$ imply $x \leq z$.

A strict order $<$ is the relation defined by $x < y$ iff $x \leq y$ and $x \neq y$. When equipped with an order, we call P an ordered set, and explicitly denote the pair via $(P; \leq)$.

Finite order relations can be visualized through Hasse diagrams, which are defined in terms of covering relations. For an ordered set P and two elements $x, y \in P$, if $x < y$ and $x \leq z < y$ implies that $z = x$, then $x \prec y$ (in words, y covers x). The Hasse diagram of P is an assignment of a point $p(x)$ in the plane $\mathbb{R}^2$ to each element x within P such that if $x \prec y$, then x is lower than y. For each covering pair $x \prec y$ in P, connect the point $p(x)$ with $p(y)$ so that the connecting line does not lie over another point $p(z)$.

For the MAC or BC with general message sets, we will treat the message index set as an ordered set[1] , and for superposition coding, two message indices can be comparable only if they are comparable via set inclusion.

Definition 1 (Superposition Order)**.** *An order $\leq$ on a message index set E satisfying*

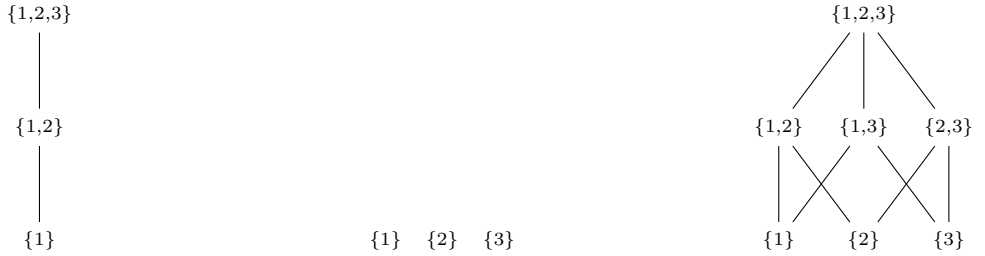
$$S \leq S' \qquad \textit{only if} \qquad S \subseteq S' \tag{2.4}$$

for $S, S' \in E$.

Examples of some message index sets for the three user MAC (or BC), along with the inclusion order (i.e., $S \leq S'$ iff $S \subseteq S'$) are in Figure 2.2:

While $M_{S'}$ is available as side information wherever M_S is known, if $S \subset S'$, we may or may not use it in the process of constructing the codeword for the message M_S. The superposition order formally encodes what decision we make per pair $\{S, S'\} \subseteq E$ with $S \subset S'$. If $S < S'$, then we superpose the codeword of M_S over the codeword of $M_{S'}$. By contrast, if S, S' are not comparable with respect to the chosen superposition order, then we do not superpose the codeword

[1] Recall that we denote the message index set by E, and that it consists of subsets of $[1 : K]$.



(a) Degraded Message Sets (b) Private Messages (c) All possible messages

Figure 2.2: Hasse diagrams for subsets of the set of subsets under the inclusion order $\subseteq$.

of M_S over the codeword of $M_{S'}$. Both options have benefits: coding with superposition achieves a larger rate region while coding without superposition reduces the necessary codebook size, and hence potentially reducing the computational burden on the encoder and decoder. Two options for superposition orders that play an important role in our development are the following:

(1) The inclusion $\subseteq$ order: put $S \leq S'$ iff $S \subseteq S'$.

(2) The discrete $=$ order: put $S \leq S'$ iff $S = S'$.

The former is leads to "maximally structured" codes while the latter leads to "minimally structured" codes. An illustration of these two particular choices, for either the two-user MAC or two-user BC, is depicted in Figure 2.3.

(a) $\subseteq$: the inclusion order. (b) $=$: the discrete order.

Figure 2.3: Inclusion and discrete order for the message index set $E = \{\{1\}, \{2\}, \{1, 2\}\}$.

Once we have settled on a choice of superposition order, there are special subsets of the message index set E which play an important role in superposition coding: the up(or down)- sets:

Definition 2. *Let P be an ordered set and let $Q \subseteq P$.*

*(i) Q is a **up-set** if whenever $x \in Q, y \in P$ and $y \geq x$, then $y \in Q$.*

*(ii) Dually, Q is a **down-set** if whenever $x \in Q, y \in P$ and $y \leq x$, then $y \in Q$.*

For each $x \in P$, the smallest up- (or down-) set containing x are the principal up- (or down-) sets. Respectively, these are $\uparrow x = \{y \in P : x \leq y\}$ and $\downarrow x = \{y \in P : x \geq y\}$. Each up-set (principal or not) is the union of the principle up-sets of its elements, $Q = \bigcup_{x \in Q} \uparrow x$. Dually, an arbitrary down-set is the union of the principal down-sets it contains.

Set in the context of the MAC (or BC) with E as the message index set equipped with a superposition order, the principal up- (or down-) sets are given by

$$\uparrow S = \{S' \in E : S \leq S'\} \ , \tag{2.5}$$

$$\downarrow S = \{S' \in E : S \geq S'\} \ . \tag{2.6}$$

These sets have an important operational meaning for superposition coding. The principal up-set in (2.5) corresponds to the set of codewords on which the codeword for the message M_S will be superposed. Dually, while decoding, the principal down-set in (2.6) corresponds to the set of codewords which were superposed over the codeword for the message M_S (and hence dependent on the correct decoding of M_S). Moreover the set of all up-sets, $\mathcal{F}_\uparrow(P; \leq)$, or down-sets, $\mathcal{F}_\downarrow(P; \leq)$, is particularly interesting[2] , as they are lattices:

Definition 3 (Lattice Set Family)**.** *A set of sets $\mathcal{F}$ satisfying*

$$A \cap B, A \cup B \in \mathcal{F} \quad \text{for all } A, B \in \mathcal{F}. \tag{2.7}$$

Maps which preserve the defining property (2.7) are also of interest:

Definition 4 (Lattice Homomorphism)**.** *Let $\mathcal{F}'$ be a lattice set family and $\mathcal{F}$ be a set of sets. Then an onto map $\mathcal{Z}_\mathcal{F} : \mathcal{F}' \mapsto \mathcal{F}$ is said to be a lattice homomorphism if for all $A, B \in \mathcal{F}'$*

$$\mathcal{Z}_\mathcal{F}(A \cap B) = \mathcal{Z}_\mathcal{F}(A) \cap \mathcal{Z}_\mathcal{F}(B) \tag{2.8}$$

$$\mathcal{Z}_\mathcal{F}(A \cup B) = \mathcal{Z}_\mathcal{F}(A) \cup \mathcal{Z}_\mathcal{F}(B). \tag{2.9}$$

[2] We further abbreviate these to $\mathcal{F}_\uparrow$ or $\mathcal{F}_\downarrow$ if the set and its order are clear from the context.

In particular, by (2.8)*, the map's range* $\mathcal{F}$ *must be a lattice set family as well.*

2.2.2 Up-set Lattice Conditional Independence

Superposition coding is a random coding strategy, and requires a distribution to code with respect to. A natural choice is to assign one auxiliary random variable per message source with joint distribution factoring in the same successive manner in which superposition coding proceeds. That order will be along the up-sets, and random tuples with distributions that factor in this way under the purview of the up-set lattice conditional independence model:

Definition 5 (Up-set lattice conditional independence model)**.** *Contains every random tuple* $(U_x : x \in P)$*, indexed by an ordered set* P*, whose distribution factors as*[3]

$$p(u_P) = \prod_{x \in P} p(u_x | u_{\uparrow x \setminus \{x\}}) = \prod_{x \in P} p(u_x | u_{x'} : x' > x). \tag{2.10}$$

There are several equivalent characterizations of this model of conditional independence. One is through the language of graphical models, where the graph is the transitive completion of the Hasse diagram of the ordered set P, with downward direction on all edges. Another is through the set of all conditional independence relations implied by the factorization above [2]:

Definition 6 (Lattice conditional independence (LCI) model)**.** *Contains every random tuple* $(U_x : x \in P)$*, indexed by a finite set* P*, which obey the Markov (or conditional independence) relations*

$$U_B \multimap U_{B \cap B'} \multimap U_{B'} \qquad \forall B, B' \in \mathcal{F}, \tag{2.11}$$

where $\mathcal{F}$ *is a sub-lattice of the boolean lattice of all subsets of* P *containing the empty set and* P *itself.*

In the case of the up-set lattice conditional independence model, the lattice is the up-set lattice $\mathcal{F}_\uparrow$. While the definition of lattice conditional independence is seemingly more general as it applies to any lattice set family, rather than only up-set lattice set families, it is not: by Birkhoff's

[3] The notation U_L refers to the tuple of random variables $(U_x : x \in L)$ indexed by the index set L.

representation theorem for finite distributive lattices, each lattice set family $\mathcal{F}$ is associated with an order on P whose up-set lattice is equal to $\mathcal{F}$ [2],[24]. That the recursive factorization (2.10) implies the Markov relations (2.11) is easy to see, and the reverse direction follows by Theorem 4.1 in [2].

2.2.3 Polymatroids

For a given choice of order and up-set lattice conditionally independent auxiliary random tuple, the rate region achieved by superposition coding will turn out to have a special structure: that of a polymatroid. All capacity regions are convex, and many capacity regions are described as unions of bounded polyhedra, which are known as polytopes. All polytopes are the convex hull of a finite set of vertices. But among all possible polytopes, only polymatroids are those for which all of its vertices can be found explicitly and quickly, a result of a "discete convex" property to be shortly described. Formally, a polymatroid is

Definition 7 (Polymatroid)**.** *Given a finite ground set E with M elements, the polytope* [4]

$$\mathcal{P}(f) = \left\{ x \in \mathbb{R}_+^E : \qquad \sum_{e \in B} x_e \leq f(B) \ \forall \ B \subseteq E \right\} \tag{2.12}$$

is a polymatroid if the set function[5] $f : 2^E \mapsto \mathbb{R}_+$ *satisifes*

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B) \qquad \textit{(submodular).} \tag{2.13a}$$

$$f(A) \leq f(B) \qquad \textit{if } A \subseteq B \qquad \textit{(nondecreasing)} \tag{2.13b}$$

$$f(\emptyset) = 0 \qquad \textit{(normalized)} \tag{2.13c}$$

for any two subsets $A, B \subseteq E$. [6]

[4] $\mathbb{R}_+^E = \{x \in \mathbb{R}^E : x_e \geq 0 \ \forall \ e \in E\}$ is the positive orthant of $\mathbb{R}^E$, the real vector space with coordinates indexed by the elements of E. If E consists of M elements, then $\mathbb{R}^E$ may be identified with $\mathbb{R}^M$.

[5] $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ refers to the non-negative real numbers.

[6] Set functions which satisfy (2.13) go by many names; Edmonds [28] and others [34, 48] refer to them as β-functions, while Lovasz [66] calls them polymatroid functions, which we adopt. When f is the rank of a matroid within E [59], then such a set function is known as a rank function; in keeping with this, they may be referred to as rank functions [106, 123]. The rank of a matroid in turn is a reference to the origins of matroid theory as an abstraction of linear independence relations among the columns of a matrix [59].

Together, the submodular and monotonic constraints constitute a "discrete convex" condition (a short proof can be found in Lemma II.3 of [27]):

$$f(B \cup C) - f(B) \leq f(A \cup C) - f(A) \qquad \text{if } A \subseteq B \ . \tag{2.14}$$

This is a diminishing returns property, and is a discrete analog to the notion of a concave function on the real line, which satisfies

$$g(y+h) - g(y) \leq g(x+h) - g(x) \qquad \text{if } x \leq y$$

when $h \geq 0$; see Figure 2.4.[7]

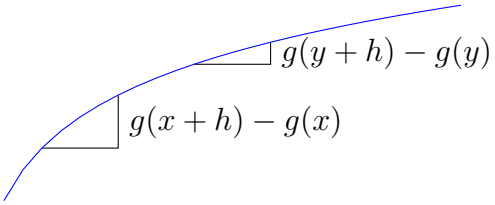


Figure 2.4: Diminishing returns: concave functions

Concave functions are functions over which maximization is simple - a greedy hill-climbing procedure will find the global maximum. Submodular function provide an analogous guarantee in a discrete setting —a greedy hill-climbing procedure attains the maximum of the weighted sum rate

$$\text{maximize } \sum\nolimits_{e \in E} \mu_e x_e \qquad \text{subject to } x \in \mathcal{Q}. \tag{2.15}$$

Any convex set $\mathcal{Q}$ is fully characterized by the set of solutions to the linear programs (2.15), for all choices of real weights $(\mu_e : e \in E)$: it is a dual description of $\mathcal{Q}$. The distinguishing property of polymatroids, relative to the wider classes of polytopes, is that they are the **unique** polytopes for which the following greedy algorithm finds a vertex which attains the maximum weighted linear sum (2.15) for all choices of weights $(\mu_e : e \in E)$ (for a proof, see [59]):

(Step 1) Enumerate E so that $\mu_1 \geq \ldots \geq \mu_k > 0 \geq \mu_{k+1} \geq \cdots \geq \mu_M$ and set $x_1 = \cdots = x_M = 0$.

[7] The reader may ask why (2.14) is labeled as a convex, rather than concave, condition. For many years, researchers within combinatorial optimization could not decide whether submodularity was closer to concavity or convexity, see [66]. After further study, consensus has shifted towards thinking of convexity as the more appropriate analogy [73].

(Step 2) For $j = 1, \ldots, k$, Increase x_j until a constraint becomes tight.

The output of this greedy algorithm is

$$x_1 = f(\{e_1\})$$

$$x_i = f(\{e_1, \ldots, e_i\}) - f(\{e_1, \ldots, e_{i-1}\}) \text{ for } 1 \leq i \leq k$$

$$x_i = 0 \text{ for } i > k.$$

Hence, there are at most $M!$ vertices that are maximal with respect to the ordering $y \leq x \leftrightarrow (y_e \leq x_e : e \in E)$. These $M!$ dominating vertices needn't be distinct: some of the defining bounds (2.12) may be redundant. In fact, we may still define a polymatroid function, and a polymatroid, if we only require that the definitions (2.13) and the sum-rate bounds (2.12) to hold over a lattice set family $\mathcal{F}$ on E (rather than over the entire power set of E). The resultant polytope is a polymatroid in the sense of Definition 7 as substantiated by the following lemma. While this appears to be well-understood within combinatorial optimization, we provide a short proof here for completeness and clarity.[8]

Lemma 2.2.1 (Polymatroid over a Lattice Set Family [66][94])**.** *Given a finite ground set E with M elements and a lattice set family $\mathcal{F}$ defined on E, the polytope*

$$\mathcal{P}_{\mathcal{F}}(f) = \left\{ x \in \mathbb{R}_+^E : \qquad \sum_{e \in B} x_e \leq f(B) \ \forall \ B \in \mathcal{F} \right\}$$

is a polymatroid if $f : \mathcal{F} \mapsto \mathbb{R}_+$ is a polymatroid function over $\mathcal{F}$ (that is, f is normalized, nondecreasing, and submodular over the elements of $\mathcal{F}$).

Proof. For each $B \subseteq E$, define

$$\mathcal{Z}_{\mathcal{F}}(B) = \bigcap \left\{ B' \in \mathcal{F} : B \subseteq B' \right\} \tag{2.16}$$

to be the smallest element of $\mathcal{F}$ containing B. It is a lattice homomorphism (recall Definition 4) with its fixed points on 2^E being the elements of $\mathcal{F}$. Extend f to **all** subsets B of E with [9]

$$\tilde{f}(B) = f \circ \mathcal{Z}_{\mathcal{F}}(B). \tag{2.17}$$

[8] Jack Edmonds understood this - it is in his original exposition in 1970 [28], but without details.

[9] $f \circ g(x) = f(g(x))$ is the composition of the functions f and g.

As the map $\mathcal{Z}_{\mathcal{F}} : 2^E \mapsto \mathcal{F}$ is a lattice homomorphism, $\tilde{f}$ is again a polymatroid function. Consider the polymatroid $\mathcal{P}(\tilde{f})$ and consider the corresponding inequality $\sum_{S \in B} R_S \leq \tilde{f}(B)$ for each $B \subseteq E$. If $B \notin \mathcal{F}$, then this inequality is redundant given the corresponding inequality for $B' = \mathcal{Z}_{\mathcal{F}}(B)$ as $\tilde{f}(B) = \tilde{f}(B')$ and $B \subset B'$. If $B \in \mathcal{F}$, then $\tilde{f}(B) = f(B)$ and this inequality is a defining inequality for the polytope $\mathcal{P}_{\mathcal{F}}(f)$. Hence, $\mathcal{P}_{\mathcal{F}}(f) = \mathcal{P}(\tilde{f})$. □

2.3 Common Notation

We collect all of the previously introduced notation into Table 2.4 so that it may serve as a quick reference. Through the thesis, we adopt further notation that will be of use.

We refer to random variables in upper case (e.g., A or B) and to specific values they take in lower case (e.g., a). When demonstrating the distribution we refer to, we provide a subscript (e.g., the probability that $B = a$ is $p_B(a)$). If the distribution is clear from the context, as in $p_A(a)$, we further abbreviate to $p(a)$. We use the notation $A \multimap\!\circ\!\!-\, B \,-\!\!\circ\!\multimap\, C$ to denote a Markov Chain (which we can take to be ordered either as $A \to B \to C$ or as $C \to B \to A$). We will refer to vectors and matrices in bold face (e.g. $\mathbf{A}$), with upper case and lower case again denoting the difference between the random variable and the specific value it may take. If $\mathcal{I}$ indexes a set of variables $(x_i : i \in \mathcal{I})$, let $x_{\mathcal{I}} = (x_i : i \in \mathcal{I})$ denote the collection of these variables and let $x(\mathcal{I}) = \sum_{i \in \mathcal{I}} x_i$ denote their sum. If we write $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$), we imply that both matrices are Hermitian and that $\mathbf{A} - \mathbf{B}$ is positive semidefinite (positive definite). Blank entries in a matrix are intended to interpreted as being equal to zero; e.g.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ & \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \\ & \\ & 1 \end{bmatrix}.$$

We use $|\cdot|$ for different quantities depending on the argument. If $\mathbf{X} \succeq \mathbf{0}$, then $|\mathbf{X}| = \det(\mathbf{X})$. If S is a set, $|S|$ is its cardinality. Finally, if $x \in \mathbb{C}$, then $|x|$ is its magnitude.

Abbreviation	Meaning	Notes
$\lfloor M \rfloor$	largest integer N s.t. $N \leq M$	
$[1:M]$	$\{1, \ldots, \lfloor M \rfloor\}$ for any $M \geq 1$	
2^X	Set of all subsets of a finite set X	
E	A subset of $2^{[1:K]} \backslash \emptyset$	Represents a message index set
$\mathbb{R}_+$	Non-negative real numbers	
$\mathbb{R}_+^E$	Non-negative real numbers indexed by elements of E	
$\uparrow S$	$\{S' \in E : S \leq S'\}$: Up-set of S	Unless otherwise noted, the order $\leq$ is the inclusion order $\subseteq$.
$\downarrow S$	$\{S' \in E : S \geq S'\}$: Down-set of S	
$\uparrow' S$	$\uparrow S \backslash S$	
$\downarrow' S$	$\downarrow S \backslash S$	
$\mathcal{F}_\downarrow(P; \leq)$	Lattice of all down-sets in a set P w.r.t $\leq$	
$\mathcal{F}_\uparrow(P; \leq)$	Lattice of all up-sets in a set P w.r.t $\leq$	
$\mathcal{F}_\downarrow$	Lattice of all down-sets in E w.r.t $\subseteq$	
$\mathcal{F}_\uparrow$	Set of all up-sets in E w.r.t $\subseteq$	
$\mathcal{P}(f)$	Polymatroid defined by submodular set function f (see (2.12))	Ground set is taken to be E
$L(P; \leq)$	Set of distributions satsifying the lattice conditional independence model over the ordered set $(P; \leq)$.	

Table 2.4: Abbreviations and Notation

Chapter 3

DM Multiple Access Channel with General Message Sets

3.1 Introduction

In this chapter, we use the lens of order theory to provide a fresh perspective on the classical random coding technique of superposition coding. This perspective unifies and extends prior results on the K-user DM[1] MAC with general message sets, a setting for which a fundamental partial order exists among the message sources. The approach is general and reproduces all of the following results with a single argument.

- In 1973 and 1972, Ahlswede [1] and Liao [64] characterized capacity when only private messages exist, where each message source is available to a single transmitter.
- In 1979, Han [48] characterized the capacity when all possible private and common messages exist (of which there are $2^K - 1$). That is, to each subset of users, there is an independent message source known to those users and unknown to the other users.
- In 1984, Prelov [80] characterized the capacity region for the degraded message setting, where all message sources are accessible to the first encoder, all sources except the first are accessible to the second encoder, all sources except the first and second are accessible to the third encoder, and so forth.
- In 2010, Gündüz and Simeone [43] devised a tailored capacity representation for the settings

[1] Recall that DM refers to discrete memoryless, indicating that the channel is memoryless and has finite input and output alphabets.

The classical setting, with only private messages	$E = \{\{1\}, \{2\}, \ldots, \{K\}\}$
The degraded message setting	$E = \{\{1\}, \{1,2\}, \ldots, \{1,2,\ldots,K\}\}$
All possible messages	$E = 2^{[1:K]} \backslash \emptyset$

Table 3.1: Message Set Possibilities

where some, but not all, of the possible messages exist. Their method extends to the case with all possible messages, but their development is most beneficial when there the order of K, rather than 2^K, messages.

All of these previous results, in fact, correspond to superposition coding, and differ in the choice of superposition coding **order** and of message index set. In general, the coding order can be any superposition order, per Definition 1, and any message index set E, which consists of distinct subsets of the set of users, as described in Section 2.1.3. Possible selections of the message index set E are given in Table 3.1 and possible selections of the superposition order[2] , which range between either always coding without superposition or always coding with superposition, are depicted in Figure 3.1.

Beyond establishing capacity, the previous results also determine a number of interesting properties of the capacity region:

- Capacity corner points are achievable through successive decoding [48].
- Capacity is attainable with small auxiliary codebooks [48] per input distribution.[3] However, there are a large number of possible input distributions to search over.

[2] Recall that this is defined as an order with $S \leq S'$ only if $S \subseteq S'$

[3] Specifically, each message is assigned an auxiliary codeword independently of the other messages and input codewords are a channel use by channel use deterministic function of these auxiliary codewords.

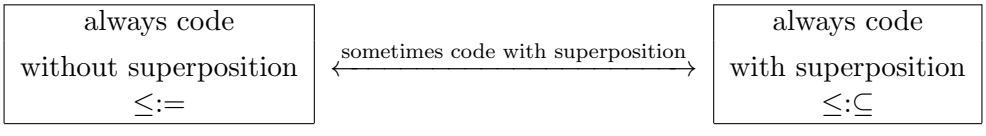


Figure 3.1: Superposition coding strategies, corresponding to different parital orders

- Capacity can be described by a small number of possible input distributions, at the expense of higher encoding and decoding complexity per input distribution [43].[4]

The general superposition coding strategy described herein simply reproduces all of these properties. The latter two properties follow by the flexibility of choice in the partial order by which superposition proceeds. The first point follows as, per input distribution, the achievable rate region afforded by superposition is a **polymatroid**, regardless of the choice of choice of message index set or of superposition order. Though this has been known for coding without superposition, we uncover this polymatroidal structure is also present for coding with superposition. Exploiting polymatroidal structure and the partial order on the message index set sheds further light onto the structure of the capacity region:

- Concatenating a no-superposition coding inner code with an outer code consisting of elementary down-set rate transfer operations achieves the same rate region as can be achieved through superposition coding; see Figure 3.2.

- Full joint decoding is unnecessary—successive group decoding suffices to attain capacity, without time-sharing, and the points achievable through successive group decoding are shared by among all possible superposition coding strategies.

- While the set of admissible distributions differs for different superposition orders, they are all equivalent in a certain rate-preserving sense.

The rate transfer result is of special mention: it is an example of an efficient projection from a high-dimensional polytope onto a lower dimensional polytope that proceeds without appeal to Fourier-Motzkin Elimination, which would be unwieldy to apply directly. Moreover, it highlights that one partial order, the order corresponding to set inclusion on the message index set, is truly fundamental to the MAC with general message sets, as it may concatenated to any achievable scheme to provide a larger achievable rate region.

[4] Specifically, each message is assigned an auxiliary codeword partially dependent on the other messages and each input codewords equal to one of these auxiliary codewords.

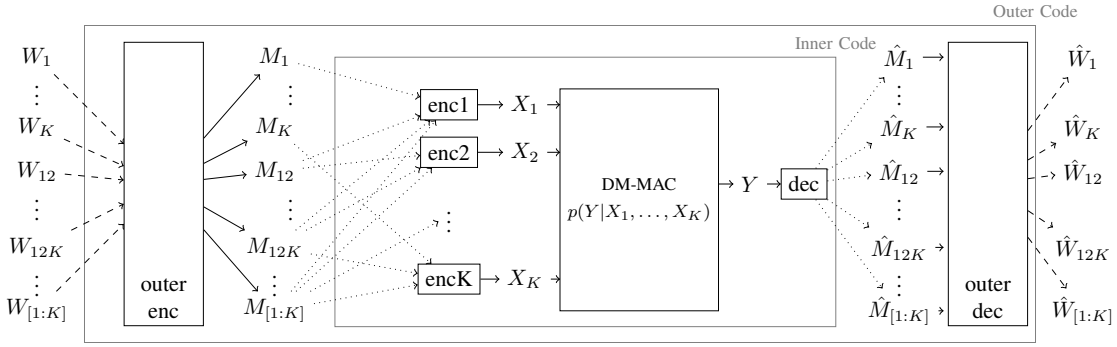


Figure 3.2: Code concatenation: Down-set rate splitting can be universally applied over any other achievable scheme for a MAC with general message sets.

3.2 To Superpose or To Not Superpose

To illustrate that random coding with superposition and random coding without superposition are two strategies cut from the same cloth, we review two classical results for the two-user case with two private and one common message (so that the message index set is $E = \{1, 2, 12\}$[5]). In 1973, Slepian and Wolf [99] characterize capacity by coding with superposition, while in 1979, Han [48] characterizes capacity by coding without superposition. Despite the different means of attaining capacity, there similarities between the two approaches that have been hitherto unnoticed:

- Both characterizations involve union of polymatroids,[6] and the set of admissible input distributions for each case are equivalent in a certain sense.
- Full joint decoding is unnecessary—successive group decoding suffices, and the points achievable through successive group decoding are shared by among all possible superposition coding strategies.

3.2.1 Prior results: Two-user case

In both cases, the distributions which we code with respect to involve an auxiliary tuple (U_1, U_2, U_{12}) which are related to the inputs via two deterministic functions $X_j = x_j(U_j, U_{12})$.

[5] For brevity, we abbreviate $\{1, 2\}$ to 12, $\{1\}$ to 1 and $\{2\}$ to 2

[6] While known for the Han rate region, this fact appears to have been unnoticed for the Slepian-Wolf rate region.

Each auxiliary random variable is paired with the message source of the same index. Should we code with or without superposition?

3.2.1.1 Coding without superposition: Han's strategy

Coding without superposition does not impose any dependence among the codewords, or auxiliary random variables, of the different message sources. As such, attention is restricted to auxiliary random variables which are independent and to codebooks where each message is assigned an auxiliary codebook at random,

$$u_S^n(m_S) \sim \prod_{t=1}^{n} p(u_{S,t}) \qquad \forall\ m_S \in \left[1 : 2^{nR_S}\right],$$

independently of the other message sources. At the jth input, the codeword to be sent is computed by applying the deterministic function x_j to the auxiliary codewords u_j^n, u_{12}^n on a channel-use by channel-use basis: $x_{j,t}(m_j, m_{12}) = x_j(u_{j,t}(m_j), u_{12,t}(m_{12}))$. Decoding with joint typicality, we can reliably estimate the sent message if

$$\sum_{S\in B} R_S \le I(U_B; Y|U_{E\setminus B}) \qquad \forall B \in \Big\{\emptyset, \{1\}, \{2\}, \{12\}, \{1,2\}, \{2,12\}, \{1,12\}, \{1,2,12\}\Big\}. \tag{3.1}$$

Each of these conditions, of which there is one per subset B of E, has a simple interpretation. Suppose that the message tuple $(m_S : S \in E)$ is the sent message. Then the inequality corresponding to B assures that the probability of some wrong message tuple $(\hat{m}_S : S \in E)$ satisfying (3.2) being jointly typical with the receiver is vanishingly small.

$$\begin{aligned} \hat{m}_S &\neq m_S \qquad S \in B \\ \hat{m}_S &= m_S \qquad S \notin B. \end{aligned} \tag{3.2}$$

The polyhedral condition (3.1) has a special character: the bounds $\rho(B) = I(U_B; Y|U_{E\setminus B})$ are submodular, monotonic, and increasing [7] . Thus, by the canonical polymatroid definition (7), which involves an inequality for every element of the boolean lattice of all subsets of E, the region (3.1) is a polymatroid.

[7] As noted by Han [48]

3.2.1.2 Coding with superposition: Slepian and Wolf's strategy

Now suppose that we do build in dependence among the codewords of the different message sources via superposition coding. In particular, we allow the the private message codewords, and private auxiliary random variables, to depend on their common message counterpart. That is, the auxiliary random variables may factor recursively as $p(U_{12}, U_1, U_2) = p(U_{12})p(U_1|U_{12})p(U_2|U_{12})$, and we select the auxiliary codewords iteratively and at random as

$$\begin{aligned} u_{12}^n(m_{12}) &\sim \prod_{t=1}^{n} p(u_{12,t}) && \forall\ m_S \in \left[1 : 2^{nR_{12}}\right] \\ u_j^n(m_j|m_{12}) &\sim \prod_{t=1}^{n} p(u_{j,t}|u_{12,t}) && \forall\ m_j, m_{12} \in \prod_{S\in\{j,12\}} \left[1 : 2^{nRS}\right]. \end{aligned}$$

The input codewords are generated in the same channel use-by-channel use basis as before. Decoding with joint typicality, we can reliably estimate the sent message if

$$\sum_{S\in B} R_S \leq I(U_B; Y|U_{E\setminus B}) \qquad \forall B \in \Big\{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,2,12\}\Big\}. \tag{3.3}$$

As with coding with superposition, each bound corresponds to an error event of the form (3.2), but in contrast to the previous situation, the conditions corresponding to $B \in \{\{12\}, \{12,1\}, \{12,2\}\}$ aren't enforced. Because of the dependence among message sources built into the codewords, all three of these error events are dominated in probability by the error event $B = \{12,1,2\}$. All four correspond to the case where the receiver has misdecoded the common message m_{12}. But, as the private message auxiliary codeword choice depends common message as well, misdecoding the common message assures that probability of a false message estimate being joint typical with the output is always as low as it would be when the private messages are miscoded, irrespective of whether or not the private message estimates are actually correct.

The conditions in (3.3) are similar to those in (3.1), with each of the four bounds matching one of the seven bounds in (3.1). Coding without superposition lead to reliable communication conditions that are polymatroidal. Does this continue to be true when we code with superposition? Yes! As

- the defining inequalities of (3.3) are over the down-set lattice set family $\mathcal{F}_{\downarrow}(E;\subseteq)$, and
- the defining bounds of (3.3) are submodular, monotonic, and normalized over this lattice set family (a consequence of Lemma 3.2.1 to be shown shorty),

the Polymatroid over a Lattice Set Family Lemma[8] assures that the corresponding set of inequalities do define a polymatroid, albeit with redundant inequalities. To the best of the author's knowledge, this is an unnoticed fact within information theory literature. The requisite lemma for this conclusion will be useful for the general case, and we state and prove it in generality here:

Lemma 3.2.1. *Equip E with an order and fix a random tuple $U \equiv (U_S : S \in E)$ in the corresponding up-set lattice conditional independence model. Then irrespective of the conditional probability mass function $p(y|u_E)$, the mutual information*

$$\rho(B) = I(U_B; Y|U_{E\setminus B}),$$

viewed as a set function, is a polymatroid function over the down-set lattice of E.

Proof. Recall that ρ is polymatroidal if it is normalized (2.13c), increasing (2.13b), and submodular (2.13a) with respect to all the down sets of E. Regardless of the distributional assumptions on U, the mutual information bound ρ is normalized and increasing, where the latter is a consequence of the simple fact that conditioning reduces entropy. To show that ρ is submodular, we follow Han [48] and for any two down-sets A, B we write

$$\begin{aligned}\rho(A\cup B)+\rho(A\cap B) &= I(U_{A\cup B}; Y|U_{E\setminus(A\cup B)}) + I(U_{A\cap B}; Y|U_{E\setminus(A\cap B)}) \\ &= H(U_E) - H(U_{A^c\cap B^c}) + H(U_E) - H(U_{A^c\cup B^c}) \\ &\quad - H(U_{A\cup B}|U_{A^c\cap B^c}, Y) - H(U_{A\cap B}|U_{A^c\cup B^c}, Y), \end{aligned} \tag{3.4}$$

where we write $X^c = E\setminus X$ for any subset $X \subseteq E$. Now, for **any** distribution on U we have [48]

$$H(U_{A\cup B}|U_{A^c\cup B^c}, Y) = H(U_A|U_{A^c}, Y) + H(U_{A\setminus B}|U_{A^c\cup B^c}, Y)$$

[8] Lemma 2.2.1

$$\geq H(U_A|U_{A^c}, Y) + H(U_{A\setminus B}|U_{B^c}, Y) \tag{3.5}$$

$$H(U_{A\cap B}|U_{A^c\cup B^c}, Y) = H(U_B|U_{B^c}, Y) - H(U_{A\setminus B}|U_{B^c}, Y). \tag{3.6}$$

As A and B are down-sets, their complements, A^c and B^c are up-sets. Hence, if U is in the up-set lattice conditional independence model, then the implied Markov conditions (2.11) assure that

$$H(U_{A^c}) + H(U_{B^c}) - H(U_{A^c\cap B^c}) - H(U_{A^c\cup B^c}) = I(U_{A^c}; U_{B^c}|U_{A^c\cap B^c}) = 0. \tag{3.7}$$

Substituting (3.5), (3.6), and (3.7) into (3.4) provides that ρ satisfies (2.13a) and hence is submodular. Together with the previous assertions that ρ is normalized and increasing, we find that ρ is a polymatroid function over the down-set lattice of E. □

3.2.2 Common Order-theoretic Framework

Despite their differences, the strategies of coding with superposition or coding without superposition share a remarkable fact: they both lead to a polymatroidal inner bound. A unifying view of both of the strategies which explains this similarity is to view each as an instantiation of superposition coding that differ in the choice of partial **order** by which coding proceeds. Equipping with the message index set with a superposition order, the two strategies are both instantiations of the following framework:

- Choose an auxiliary random tuple which factors successively along this order:

  $$p(u_1, u_2, u_{12}) = p(u_1|u_{\uparrow 1\setminus 1}p(u_1|u_{\uparrow 1\setminus 1})p(u_{12})$$

  and a pair of functions x_j that map the auxiliary codewords to input codewords.

- Construct codewords successively along this order as

  $$u_{12}^n(m_{12}) \sim \prod_{t=1}^{n} p(u_{12,t}) \qquad \forall\ m_S \in \left[1 : 2^{nR_{12}}\right]$$

  $$u_j^n(m_j|m_{\uparrow j\setminus j}) \sim \prod_{t=1}^{n} p(u_{j,t}|u_{\uparrow j\setminus j,t}) \qquad \forall\ m_{\uparrow j} \in \prod_{S\in\uparrow j} \left[1 : 2^{nRS}\right]$$

  and map these auxiliary codewords to inputs via $x_j(u_{j,t}(m_{\uparrow j}, u_{12,t}(m_{12}))$ for each channel use $t \in [1 : n]$.

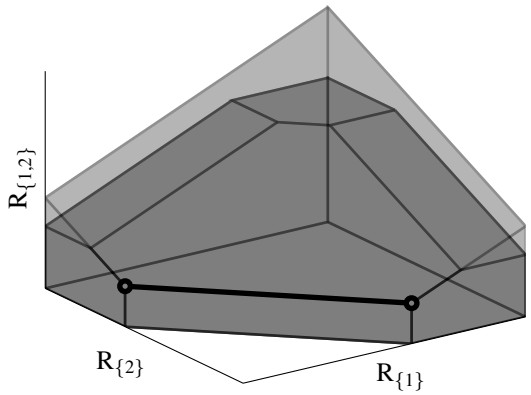


Figure 3.3: Comparison of the Slepian-Wolf (light grey) and Han (dark grey) polymatroids for a fixed input. The edge of the dominant face shared by both polymatroids is highlighted.

- Decode by joint typicality decoding, which reliably estimates the sent message if

$$\sum_{S \in B} R_S \leq I(U_B; Y | U_{E \setminus B}) \qquad \forall \text{ down-sets } B \in \mathcal{F}_{\downarrow}(E; \leq) \tag{3.8}$$

Coding with superposition fits this framework when the partial order is the inclusion order. Perhaps less clear is that coding without superposition also fits in this framework with the order as the discrete order. In this case, the inputs must factor as a product distribution. Similarly, the codebooks corresponding to different message sources are generated independently. Moreover, the down-set (and up-set) lattice of the discrete order is simply the power-set of E:

$$\mathcal{F}_{\downarrow}(E; =) = \Big\{\emptyset, \{1\}, \{2\}, \{12\}, \{12\}, \{2, 12\}, \{1, 12\}, \{1, 2, 12\}\Big\} = \mathcal{F}_{\uparrow}(E; =) = 2^E \ .$$

For a fixed input distribution permissible to use for both strategies[9] , coding with superposition achieves a larger rate region that does coding without superposition; see Figure 3.3. Yet despite this, in both cases the union of the achievable polymatroids over all permissible auxiliary random variables and input functions x_1, x_2 yields the capacity region [99, 48]. The reason why this is so has to do with the importance of the inclusion order for the MAC with general message sets, which leads to a couple of novel conclusions: first, that down-set rate splitting always accounts for the difference between coding with respect to discrete or inclusion order, and secondly, that the

[9] That is, a tuple (U_1, U_2, U_{12}) of independent random variables and fixed pair of mappings from the auxiliary random variables to the inputs.

points achievable through successive group decoding characterize capacity, and are the polymatroid faces shared by all permissible superposition coding strategies. We further detail these conclusions in the subsequent section.

The results of Slepian and Wolf include a further specialization: it suffices to only consider trivial input functions $x_j(u_j, u_{12}) = u_j$ for both $j \in \{1, 2\}$. The common framework above can be further extended to accommodate this, by observing that a mapping exists between the various sets of admissible input distributions. The differences between the admissible input distributions for coding with, and without, superposition are provided in Table 3.2:

Superposition Order	$\subseteq$	$=$
Auxiliary RV dependencies	$\underbrace{p(U_1\|U_{12})}_{\uparrow 1}\underbrace{p(U_2\|U_{12})}_{\uparrow 2}\underbrace{p(U_{12})}_{\uparrow 12}$	$\underbrace{p(U_1)}_{\uparrow 1}\underbrace{p(U_2)}_{\uparrow 2}\underbrace{p(U_{12})}_{\uparrow 12}$
Functional dependencies	$X_j = x_j(U_j, U_{12}) = U_j$	$X_j = x_j(U_j, U_{12})$
Input Dependencies	through Markov conditions	through Shannon strategies

Table 3.2: Different representation of input dependencies: two-users

Notably the different sets of input distributions represent channel input dependences by different means: either fully through Markov dependences, or fully through the deterministic functions x_1, x_2, which we call Shannon strategies, in analogy to the coding with state for the point to point channel. These sets of input distributions are equivalent in a certain rate-preserving sense: for the polymatroid (3.8) defined by the inclusion superposition order, we may equivalently express a correlated input by either of the two means above while preserving all of the defining mutual information bounds

$$\rho_{x,U}(B) = I(U_B; Y|U_{E\setminus B}) \qquad B \in \Big\{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,2,12\}\Big\}. \tag{3.9}$$

To see this, we use the following result.

Lemma 3.2.2 (Functional Representation Lemma: Appendix B of [31]). *For any tuple (A, B, C), there exists a random variable D independent of A and B such that we may represent C as a function of (B, D).*

To map a tuple of the Markov type to a tuple of Shannon strategy type while preserving the bounds (3.9), consider $(A, B, C) = (X_i, U_{12}, X_j)$ for one of the two configurations $(i, j) \in \{(1, 2), (2, 1)\}$. Apply the Functional Representation Lemma and label the resultant random variable as $U_{\{i\}}$ and the resultant deterministic function as x_i. By construction, U_1 and U_2 will be independent of each other and of U_{12}. Moreover, by the data processing inequality, replacing X_j by U_j in the bounds (3.9) does not change their values. The reverse direction, of mapping a tuple of the Shannon strategy type to a tuple of Markov type, is trivial: simply reassign the variables U_1 and U_2 to be X_1 and X_2, respectively.

3.2.2.1 Importance of the Inclusion Order

With an abundance of possible capacity descriptions, a natural question to ask is which is the most fundamental. A partial answer in favor of the inclusion order are the following two points:

- Down-set rate-splitting along the inclusion order accounts the difference in achievable rates between coding with and without superposition.
- Successive group decoding along the inclusion order suffices to attain capacity, though encoding may occur with or without superposition.

Rate transfer Per input distribution, coding without superposition achieves a smaller region that afforded by coding with superposition; see Figure 3.3 for example. A simple strategy to enlarge any rate region is to relabel parts of the common message as though they were part of a private message. This corresponds to rate-splitting the common message rate as

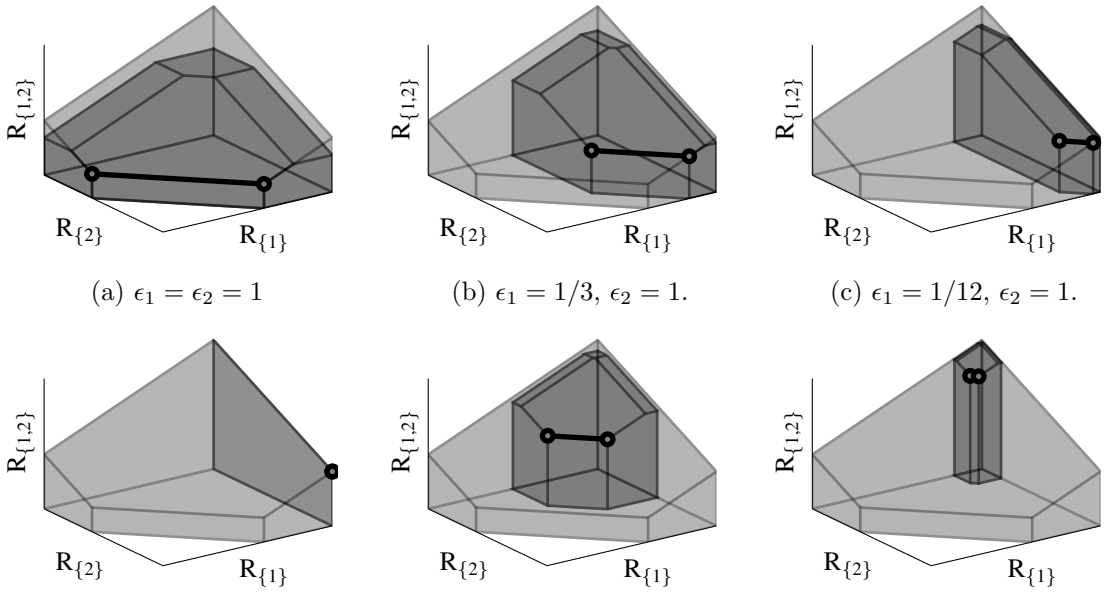
$$R_{12} = r_{12\to 12} + r_{12\to 1} + r_{12\to 2},$$

where $r_{12\to S}$ denotes the part of the message source m_{12} that will be re-labeled as belonging to the message source m_S. Thus, after relabeling, one obtains a new set of messages to be transmitted at the rates

$$\tilde{R}_1 = R_1 + r_{12\to 1} \qquad \tilde{R}_2 = R_2 + r_{12\to 2} \qquad \tilde{R}_{12} = r_{12\to 12}.$$

By restricting these new rates to satisfy the no-superposition conditions (3.1), and un-doing the relabeling process at the receiver, we can reliably transmit a message at rate $R = (R_1, R_2, R_{12})$. The set of rates thus achievable are of the form (3.3), and can be shown with an argument reliant on polymatroid properties. A general argument, which accounts for the K-user extension of this idea, is demonstrated in Appendix B.2.[10]

Sufficiency of Successive Group Decoding Why are there so many permissible capacity descriptions? A partial answer to this is that any capacity description which contains a special set of points suffices. There are a collection of faces that are shared among all possible polymatroids achieved through superposition: these faces correspond to the set of points achievable through successive group decoding along the inclusion order: namely the common message is decoded first, and then successively, the group of the two private messages are jointly decoded conditioned on the knowledge of the common message.



(a) $\epsilon_1 = \epsilon_2 = 1$ (b) $\epsilon_1 = 1/3$, $\epsilon_2 = 1$. (c) $\epsilon_1 = 1/12$, $\epsilon_2 = 1$.

(d) $\epsilon_1 = 0$, $\epsilon_2 = 1$. (e) $\epsilon_1 = \epsilon_2 = 1/3$. (f) $\epsilon_1 = \epsilon_2 = 1/12$.

Figure 3.4: Each point on dominant face of a Slepian-Wolf polymatroid (light grey) is on the successive group decoding edge of a Han polymatroid (dark grey) corresponding to a different input distribution.

[10] Han demonstrated this in Han[48], with a hand-crafted argument reliant on submodularity, but without the larger picture of polymatroidal structure that permits the generalization to K-users.

To show this, we fix an specific input distribution and corresponding achievable polymatroid. With the discrete convex properties of defining polymatroid bounds, one can show that every point on the dominant face of a polymatroid corresponds to the successive group decoding edge of another poylmatroid achievable by a different input distribution. The relationship between the two distributions is one of a variable split, to be defined shortly, where randomness is "shifted" from one variable to another. Figure 3.4 depicts this procedure, by demonstrating that each point on a Slepian-Wolf polymatroid corresponds to a successive group decoding edge of another Han-polymatroid for another input distribution, characterized by two parameters ϵ_1, ϵ_2.

A technical tool will be the notion of a variable split, which we borrow from [41]. The idea is that randomness is "shifted" from one variable to another, and the direction of the shift corresponds to the partial order on the message index set.

Definition 8 (Split)**.** *A split for a random variable $U \in \mathcal{U}$ is a family of triples (f, p_W, p_V) parameterized by a real number $\epsilon \in [0,1]$. The function $f : \mathcal{U} \times \mathcal{U}$ is called a splitting function and p_W, p_V are pmfs, each dependent on ϵ. Admissible choices for (f, p_W, p_V) satisfy*

(i) $f(W,V) \sim p_U$, where $p_{W,V}(w,v) = p_W(w)p_V(v)$.

(ii) For fixed values of u and w, the pmf $p_{f(W,V)}(u|w)$ is continuous function of ϵ. For $\epsilon = 0$, $p_{f(W,V)}(u|w) = p_U$ (so $f(W,V)$ is independent of W) while for $\epsilon = 1$, $p_{f(W,V)}(u|w)$ puts all of its mass on one element (so $f(W,V)$ is completely determined by W).

As per Example 3 in [41], it is always possible to find such a family of (f, p_W, p_V) with the desired properties for any discrete random variable $U \in \mathcal{U}$. The sufficiency of successive group decoding is provided by the following.

Theorem 3.2.3. *Consider the two-user DM MAC with two private messages and one common*

message. Then the capacity region is dominated by the rate points which satisfy

$$\begin{aligned} R_1 &\leq I(U_1; Y|U_2, U_{12}) \\ R_2 &\leq I(U_2; Y|U_1, U_{12}) \\ R_1 + R_2 &= I(U_1, U_2; Y|U_{12}) \\ R_1 + R_2 + R_{12} &= I(U_1, U_2, U_{12}; Y) \end{aligned}, \tag{3.10}$$

where the inputs may be expressible either of the Markov type or of the Shannon strategy type.

Proof. By the Slepian-Wolf description of capacity, every achievable rate tuple is in a polymatroid (3.8), where the superstition order is the inclusion order, for an input tuple of the Markov type, which is expressible as an input tuple of the Shannon strategy type. By the greedy algorithm for maximizing the weighted sum-rate over polymatroids, we know that all rate points in this polymatroid are dominated by those that satisfy $R_{12} + R_1 + R_2 = I(U_1, U_2, U_{12}; Y)$. Pick one such dominating rate tuple R'. To show the desired result, we construct a **new** tuple $(U_1', U_2', U_{12}', x_1', x_2')$ such that the bounds in (3.10) are satisfied with respect to this new input tuple.

Let (f_1, W_1, V_1) be a split of the auxiliary random variable U_1, parametrized by $\epsilon_1 \in [0, 1]$ and (f_2, W_2, V_2) be a split of the auxiliary random variable U_2, parametrized by $\epsilon_2 \in [0, 1]$. Consider the "re-assigned" auxiliary random tuple $U_1' = V_1$,$U_2' = V_2$,$U_{12}' = (W_1, W_2, U_{12})$, whose distribution is parametrized by ϵ_1, ϵ_2. In this manner, as ϵ_i increases, it "shifts" the randomness from $U'_{\{i\}}$ towards U_{12}'. For every choice of ϵ_1, ϵ_2 define

$$\rho_{x', U'}(B) = I(U_B'; Y|U'_{E\setminus B}) \quad \forall B \subseteq E, \tag{3.11}$$

which we know to be a polymatroid function over the power set of E (i.e. the down-set lattice set family with respect to the discrete order $=$) per Lemma (3.2.1). By assumption,

$$R_1' \leq I(U_1'; Y|U_2', U_{12}') \tag{3.12a}$$

$$R_2' \leq I(U_2'; Y|U_1', U_{12}') \tag{3.12b}$$

$$R_1' + R_2' \leq I(U_1', U_2'; Y|U_{12}') \tag{3.12c}$$

$$R_1' + R_2' + R_{12}' = I(U_1', U_2', U_{12}'; Y) \tag{3.12d}$$

for $\epsilon_1 = \epsilon_2 = 0$. Let's focus on the split for U_1 initially and note that for all choices $\epsilon_1 \in [0,1]$, the bounds (3.12b) and (3.12d) are unchanged. By contrast, both the bounds (3.12a) and (3.12c) do change with ϵ_1. As mutual information is a continuous function of its input pmf's, both of these bounds are continuous functions of ϵ_1. Moreover, as a function of ϵ_1, the bound (3.12a) is a continuous function with

$$\begin{aligned} \epsilon_1 = 1: &\qquad I(U_2'; Y|U_1', U_{12}') = 0 \\ \epsilon_1 = 0: &\qquad I(U_2'; Y|U_1', U_{12}') = I(U_1; Y|U_2, U_{12}). \end{aligned}$$

Hence, by the intermediate value theorem, as we increase ϵ_1 from zero to one, there is a value at which one of the two constraints (3.12a) and (3.12c) becomes tight while the other continues to hold. Let ϵ_1^* be this value.

If the private sum-rate bound on $R_1' + R_2'$ is tight after the prior step, then we are done. So suppose instead that the bound on R_1', rather than the private sum-rate bound, is tight. Hereafter, assume $\epsilon_1 = \epsilon_1^*$ and let U^* be the random tuple U' with $\epsilon_2 = 0$. If the private sum-rate bound holds and the bound on R_1' is tight, then necessarily the bound on R_2' holds:

$$\begin{aligned} R_2 = (R_1 + R_2) - R_1 &\leq \rho_{x',U'}(\{1,2\}) - \rho_{x',U'}(\{1\}) = I(U_2'; Y|U_{12}') \\ &\overset{(i)}{\leq} \rho_{x',U'}(\{2\}) - \rho_{x',U'}(\emptyset) = I(U_2'; Y|U_1', U_{12}'), \end{aligned} \tag{3.13}$$

where (i) follows by submodularity. Hence, with the benefit of knowing that the bound on R_1' is tight, we need only assure that the private sum-rate bound holds to assure that the bound on R_2' continues to hold. Now, $I(U_2'; Y|U_{12}')$ is a continuous function of ϵ_2 with

$$\begin{aligned} \epsilon_2 = 1: &\qquad I(U_2'; Y|U_{12}') = 0 \\ \epsilon_2 = 0: &\qquad I(U_2'; Y|U_{12}') = I(U_2^*; Y|U_{12}^*). \end{aligned}$$

Hence, again by the intermediate value theorem, there must be a choice of ϵ_2 such that the bound (3.13), and consequently private sum-rate bound (3.12c), is tight. □

The construction above, with variables being split is analogous to the rate-split argument. In both cases, the set of feasible rate (or variable) splits is determined by the inclusion order on

the message set index E. Further, per input distribution, the dominating rate points in (3.10) are shared by **both** the Slepian-Wolf and Han achievability schemes. An illustration of the set of rate points that achieve the rate conditions above with equality in Figure 3.4a. The above result may also be interpreted as clarifying the relationship between different constituent polymatroids of different capacity descriptions. In particular, for a fixed distribution with $(U_1, U_2, U_{12}) \in L(E;=)$ with two deterministic strategies x_1, x_2, let $\mathcal{C}(x, U)$ of tuples (x', U') derivable from (x, U) by the above procedure through the parameters $\epsilon_1, \epsilon_2 \in [0, 1]$. Then the above result implies that

$$\mathcal{P}_{\mathcal{F}_\downarrow(E;\subseteq)}(\rho_{x,U}) = \bigcup_{(x,'U')\in\mathcal{C}(x,U)} \mathcal{P}(\rho_{x',U'}).$$

That all points on the boundary of the capacity region are attainable by successive group decoding monotonically along the inclusion order $\subseteq$ on E for the two-user MAC was observed in the context of the two-user scalar Gaussian channel in [65]. Theorem 3.2.3 generalizes this to the discrete memoryless case.

3.3 Generalized Superpositon Coding

Our principal contribution is that the polymatroidal structure observed in the two-user case persists in all of the various K-user capacity regions reported thus far in the literature, and we provide a common framework which identifies that in fact a **class** of polymatroidal capacity descriptions, some of which have not yet been reported in the literature, exists. While previously observed in Han's characterization of the K-user capacity, such an observation has gone unnoticed for generalizations of the Slepian-Wolf description to more than two-users.

3.3.1 K-user Capacity Region

Theorem 3.3.1. *For the K-user DM MAC with general message sets messages, let E to be the index set of all messages to be sent (a subset of all non-empty subsets of $[1:K]$) and fix some superposition order $\leq$ over E. Then the capacity region of this channel is the convex closure of all*

rate tuples that lie in a polytope

$$\left\{R \in \mathbb{R}_+^E : \quad \sum_{S \in B} R_S \leq I\left(U_B; Y | U_{E\backslash B}\right) \ \forall\ B \in \mathcal{F}_{\downarrow}(E; \leq)\right\} \tag{3.14}$$

for some auxiliary tuple $(U_S : S \in E) \in L(E; \leq)$ *and some set of* K *deterministic relations* $X_j = x_j((U_S : j \in S \in E))$

Proof. The interesting direction is the forward direction, which follows through random coding and joint typicality decoding [31]: our contribution is an emphasis on partial order to correctly and efficiently keep track of the combinatorially many error events and of the order by which superposition proceeds. Each choice of a superposition order corresponds to a unique superposition coding strategy, where the codeword for the message M_S is superposed on the codewords indexed by $(\uparrow S)\backslash S = \{S' \in E : S < S'\}$. With such a superposition coding scheme, an error in decoding M_S results in all of the codewords that were generated dependently on the codeword for M_S (i.e. those listed in down-set $\downarrow S$) to be independent of the output. Suppose a wrong message tuple estimate has only those message estimates listed in $B \subseteq E$ differing from the correct message. By the above dependencies just discussed, its probability of being jointly typical with the output is dependent on $I\left(U_{\mathcal{Z}_{\downarrow}(B)}; Y | U_{E\backslash \mathcal{Z}_{\downarrow}(B)}\right)$, where $\mathcal{Z}_{\downarrow}(B)$ is the smallest down-set containing B. Removing redundant inequalities leads to (3.14). Rigorous details of both achievability and of the converse are provided in Appendices B.1 and B.3, respectively. □

Corollary 3.3.2. *All capacity descriptions in Theorem 3.3.1 have constituent polytopes which are polymatroids.*

Proof. Each polytope is defined over a lattice set family of inequalities, where the bounds are submodular, monotonic, and normalized (per Lemma 3.2.1). Thus, by the Polymatroid over a Lattice Set Family Lemma (i.e. Lemma 2.2.1), each polytope is a polymatroid. □

We remark the setting of Theorem 3.3.1 is quite a bit more general than the special case presented in Section 3.2: not only may the number of users be arbitrary, but the message index set and "superposition" scheme dictated by superposition order $\leq$ may be arbitrary. That is, E

may be any set of subsets of the K-users while the "superposition" scheme may be any valid choice between that of using all available side information (i.e. coding with respect to the inclusion order) and that of using none of the available side information (i.e. coding without superposition, or equivalently, coding with respect to the discrete order). It includes as special cases those listed in Table 3.1. Our formalism subsumes the K-user capacity description of Han [48] and of Gündüz and Simeone [43], and reproduces the MAC capacity region provided by Rini and Goldsmith [81].

3.3.2 Relationship to Previous Results

The results of Rini and Goldsmith [81] fits under the purview of Theorem (3.3.1) when the message index set contains all possible messages is equipped with the inclusion order, the most structured order possible. Han's general K-user description of capacity [48] also fits under the umbrella of Theorem (3.3.1), with the message index set again containing all possible messages, but with the superposition order taken to be the discrete order, the least structured superposition order possible. The up- and down-set lattices for the discrete order are both equal to the power set of E: $\mathcal{F}_{\downarrow}(=) = \mathcal{F}_{\uparrow}(=) = 2^E$. As the principal down-sets of this order are the singletons in E, Han's coding scheme does not permit either the codebooks or the auxiliary random variables of different message sources to depend on each other. [11] Though we do not provide cardinality bounds on the auxiliary random variable alphabets, Han [48] was able to determine such bounds. In particular, $\mathcal{U}_S$ can be bounded as $|\mathcal{U}_S| \leq \prod_{j \in S} |\mathcal{X}_j| + |E|$. This bounding approach appears to be easily tractable only for the capacity representation corresponding to the discrete order.

Gündüz and Simeone's description fits partially the purview of Theorem 3.3.1, with an additional caveat: for special message index sets, one may dispense with the deterministic functions $x_1, \ldots, x_K$ and simply model the inputs as belonging to the collection of auxiliary random variables (specifically, where each input is either equal to, or part of, a single auxiliary random variable). This approach is especially advantageous when the number of messages to send is small and it is

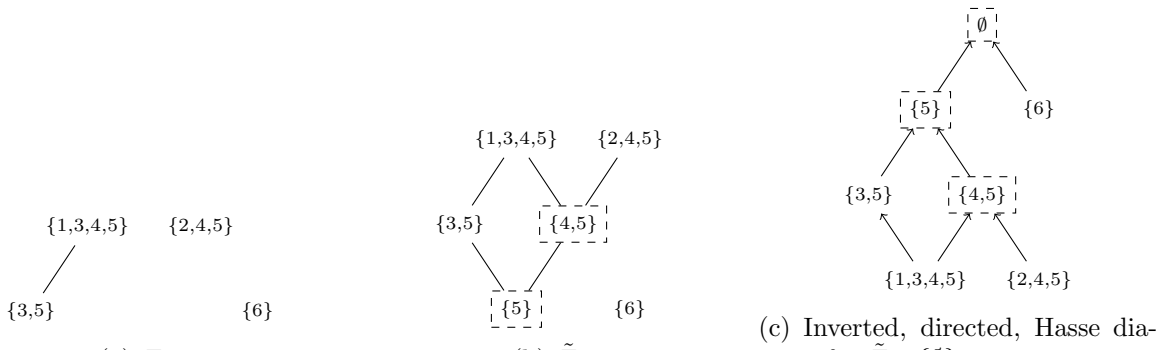
[11] By definition of the up-set lattice conditional independence model, when the order is the discrete order, the auxiliary tuple $(U_S : S \in E)$ has a product distribution $p(U_E) = \prod_{S \in E} p(U_S)$.

possible to either greatly reduce or obviate the need for auxiliary random variables other than the inputs in the description of the capacity region.

This can be done whenever the message index set contains all sets of the form

$$\mathcal{S}(j) = \bigcup \{S : j \in S \in E\} \tag{3.15}$$

By appending zero-rate messages if necessary, add all sets of the form (3.15) to E to create a message index set $\tilde{E}$. To describe the form of the auxiliary random variables, consider the **message graph** of [43], which is a directed version of the inverted Hasse Diagram for the appended message index set $\tilde{E} \cup \emptyset$, where the empty set is included so that all "down-sets" can be thought of as rooted trees; see Figure 3.5. Adopting the language of [43] for this induced message graph, define a parent of a



(a) E. (b) $\tilde{E}$ (c) Inverted, directed, Hasse diagram for $\tilde{E} \cup \{\emptyset\}$.

Figure 3.5: Gündüz and Simeone's $K = 6$-user example with the original message index set (a), the smallest intersection-closed message index set containing E (a), and its rooted version (c).

node $S \in \tilde{E}$ any node $S' \in \tilde{E}$ such that $S' \prec S$. Then define $\mathcal{M}$ to be the set of vertices $S \in \tilde{E}$ with more than one parent. Notably, if $S \notin \mathcal{M}$, then there must be at least one $j \in S$ for which $\mathcal{S}(j) = S$. For each $S \in \tilde{E}$, let

$$U_S = \begin{cases} (X_j : \mathcal{S}(j) = S) & S \notin \mathcal{M} \\ \tilde{U}_S, (X_j : \mathcal{S}(j) = S) & S \in \mathcal{M} \end{cases}. \tag{3.16}$$

Then the capacity region of [43] is the one provided by Theorem 3.3.1, with $\tilde{E}$ equipped with the inclusion order in place of E, with $R_S = 0$ for $S \in \tilde{E} \backslash E$, and with up-set lattice conditionally independent $(U_S : S \in \tilde{E})$ that are the form (3.16).

3.3.3 On the Description Complexity

An advantage of polymatroid structure is that despite the extraordinarily large number of constraining inequalities on the achievable rate region, the underlying structure of the rate region is low, as the vertices are attainable through a greedy algorithm. For the moment, consider just how many bounds there are: if, for example, we code with respect to the discrete order, then there is a defining bound for each subset of E: this is the boolean lattice (minus the empty set) of subsets of E, which contains $2^{|E|} - 1$ elements. The inclusion order affords much smaller number of defining bounds; the question is, how much smaller?

In the case that the message set index is as large as possible, where $E = 2^{[1:K]} \backslash \emptyset$, the number of bounds for the defining polymatroids, under the inclusion and discrete orders, is tabulated in Table 3.3. Counting reveals that there are $2^{2^K - 1} - 1$ bounds for the polymatroid which arise under

K	# Sources: $\lvert E\rvert$	# bounds (inclusion order): $\lvert\mathcal{F}_{\downarrow}\rvert$	# bounds (discrete order): $\lvert 2^E \backslash \emptyset\rvert$
2	3	4	7
3	7	18	127
4	15	166	32,767
5	31	7,579	$\approx 2 \times 10^9$
6	63	7,828,352	$\approx 8 \times 10^{18}$

Table 3.3: Counting antichains: number of defining polymatroid bounds under the discrete or the inclusion order.

the discrete order. How many are there for the polymatroid under the inclusion order? Counting the number of these defining bounds is equivalent to counting the number of anti-chains in the set E, which happens to be an old problem with a long history dating back to Dedekind (see [13]). Progressively better bounds have been proposed, and the easiest to state match to first order in the exponent. In particular, the (base-2) logarithm of the number of anti-chains in E is approximately $N = \binom{K}{\lfloor K/2 \rfloor}$, the size of the middle layer $\{S : |S| = \lfloor K/2 \rfloor\})$ of E. As there are 2^N anti-chains available by picking only from this middle layer, it is clear that N is a lower bound to the logarithm of the total number of anti-chains in E. Remarkably, N serves also as an upper bound, with the first resolution of this provided in 1967 [13], and with a short modern proof provided through

Pippenger's information theoretic method [79].

Thus, to compare the number of bounds arising between the different choices of the inclusion order or the discrete order, we have

$$\log(|\mathcal{F}_{\downarrow}|) \sim \binom{K}{\lfloor K/2 \rfloor} \overset{(ii)}{\sim} \frac{\sqrt{2\pi K}(K/\mathrm{e})^K}{\pi K(K/(2\mathrm{e}))^K} = \sqrt{\frac{2}{\pi}}\frac{2^K}{\sqrt{K}}$$
$$\log(|2^E|) \sim 2^K \quad .$$

where (ii) by follows by Stirling's approximation. Thus, there are far less defining bounds when coding with respect to the inclusion order than with respect to the discrete order, but there are still a super-exponential number of bounds.

3.3.4 On Permissible Input Distributions

To establish a common notation for the different classes of partially correlated input tuples, for a fixed superposition order $\leq$ on E, define $\mathcal{Q}(E;\leq)$ to consist of all pairs $x = (x_1, \ldots, x_K)$ and $U = (U_S : S \in E)$, where U is a random tuple with distribution that factors according to the up-set lattice conditional independence model,

$$\prod_{S \in E} p(u_S | u_{S'} : S < S'), \tag{3.17}$$

and where each x_j is a deterministic function relating the jth input to the tuple U as

$$X_j = x_j(U_S : j \in S \in E) \quad j \in [1:K]\,. \tag{3.18}$$

The above characterization of a partially correlated input provides much flexibility - namely, there are two potential avenues by which to model dependency among the inputs:

(i) Through Markov dependencies: If there are dependencies among the auxiliary random variables $(U_S : S \in E)$, then the channel inputs are partially correlated.

(ii) Through Shannon strategies: if the deterministic functions $x_1, \ldots, x_K$ have shared function inputs, then the channel inputs are partially correlated.

When the message set E includes all sets of the form (3.15) (by adding zero-rate messages, one can always append such sets to E), then either description is sufficiently general to encompass all possible input correlations. That is we can either a) subsume the inputs into the set of auxiliary random variables and represent channel input correlations through Markov dependences in the auxiliary random variables themselves, **or** b) have independent auxiliary variables and represent all channel input correlations through complex Shannon strategies.

To demonstrate this argument concretely, for a given superposition order on E, and input pair $(x, U) \in \mathcal{Q}(E; \leq)$, define the polymatroid

$$\mathcal{P}_{\leq}(x, U) = \left\{ R \in \mathbb{R}_+^E : \qquad \sum_{S \in B} R_S \leq I(U_B; Y | U_{E \setminus B}),\ \forall\ B \in \mathcal{F}_{\downarrow}(E; \leq) \right\},$$

where (Y, U) has joint distribution induced by passing the $X_j = x_j(U_S : j \in S \in E)$ through the multiple access channel. Then input dependencies can be completely characterized through Shannon strategies:

Lemma 3.3.3 (Correlation through Shannon strategies)**.** *Let $\leq$ be any superposition order on E. Take any input $(x, U) \in \mathcal{Q}(E; \leq)$, with correlations possibly present through both Markov dependences and Shannon strategies. Then there is an alternate input $(x', U') \in \mathcal{Q}(E; =)$, with correlations **only** present through Shannon strategies that preserves the rate region achievable under the superposition strategy defined by $\leq$; that is,*

$$\mathcal{P}_{\leq}(x, U) = \mathcal{P}_{\leq}(x', U').$$

Proof. Let Y, U have joint distribution induced by passing the inputs $X_j = x_j(U_S : j \in S \in E)$ through the multiple access channel. We will show that, without affecting the joint distribution on (Y, U), we may assume

$$U_S = g_S(U'_{\uparrow S}) \quad \forall S \in E \tag{3.19}$$

where each g_S is a deterministic function and where the random variables $(U'_S : S \in E)$ are independent. Then, as the tuple U has distribution which factors according to a superposition

order, it follows that

$$U'_{E\setminus B} \multimap U_{E\setminus B} \multimap Y \tag{3.20}$$

for every **down-set** B of E and so

$$\begin{aligned} I(U'_B; Y|U'_{E\setminus B}) &= H(Y|U'_{E\setminus B}) - H(Y|X_1,\ldots,X_K) \\ &\overset{(i)}{=} H(Y|U_{E\setminus B}, U'_{E\setminus B}) - H(Y|X_1,\ldots,X_K) \\ &\overset{(ii)}{=} H(Y|U_{E\setminus B}) - H(Y|X_1,\ldots,X_K) = I(U_B; Y|U_{E\setminus B}), \end{aligned} \tag{3.21}$$

where (i) follows as each $U_{E\setminus B}$ is a function of $U'_{E\setminus B}$ per (3.19) and (ii) follows by the Markov relation (3.20). As the defining bounds of the achievable rate region are preserved, so too is the rate region itself.

It remains to substantiate our claim that (3.19) holds without loss of generality for some collection of deterministic functions and independent random variables. To see this, let $S_1, S_2, \ldots, S_M$ be an exhaustive, never-increasing listing of E. By definition, U has distribution which recursively factors as

$$p(u_{S_1},\ldots,u_{S_r}) = \prod_{i\le r} p(u_{S_i}|u_{\uparrow S_i\setminus S_i}). \tag{3.22}$$

We proceed by induction. For the root case, set $U'_{S_1} = U_{S_1}$ for which (3.19) trivially holds with S_i in place of S. Now let $r > 1$ and assume there is a secondary tuple $(U'_{S_1},\ldots,U'_{S_{r-1}})$ of independent random variables such that for each $i < r$, (3.19) holds with S_i in place of S. There are two cases.

- If $S_r =\!\uparrow S_r$, set $U'_{S_r} = U_{S_r}$. By the factorization (3.22), U'_{S_r} is independent of all $(U_{S_1},\ldots,U_{S_{r-1}})$ and trivially (3.19) holds with S_r in place of S.

- Suppose $S_r \subset\!\uparrow S_r$. Apply the Functional Representation Lemma (Lemma 3.2.2) to yield that we may represent U_{S_r} as $U_{S_r} = f_r(U'_{S_r}, U_{\uparrow S_r\setminus S_r})$ for some function f_r and random variable U'_{S_r} independent of $(U'_{S_1},\ldots,U'_{S_{r-1}})$. Composing f_r with the prior functions g_{S_i} for $i < r$ yields a function g_{S_r} for which $U_{S_r} = g_{S_r}(U'_{\uparrow S_r})$.

After creating the tuple U', compose the K original mappings $x_1,\ldots,x_K$ with the functions g_S to produce K new mappings $X_j = x'_j(U'_S : j \in S \in E)$ for each $j \in [1:K]$. □

If E contains the sets $\mathcal{S}(j)$ defined in (3.15) and we only consider the capacity characterizations corresponding to the down-set lattice set families, then we may do the reverse: subsume the input X_j into the auxiliary random variable $U_{\mathcal{S}(j)}$ and completely characterized input correlations through Markov dependencies.

Lemma 3.3.4 (Correlation through Markov dependencies)**.** *Assume, by appending zero-rate messages if necessary, that E contains all sets $\{\mathcal{S}(j) : 1 \leq j \leq K\}$. Fix an input distribution $(x, U) \in \mathcal{Q}(E, \subseteq)$, with correlations possibly present through both Markov dependences and Shannon strategies. Then there is an input distribution $(x'', U'') \in \mathcal{Q}(E, \subseteq)$, with correlations **only** present through Markov dependencies that preserves the rate region achievable under the superposition strategy defined by the inclusion order $\subseteq$; that is,*

$$\mathcal{P}_{\subseteq}(x, U) = \mathcal{P}_{\subseteq}(x'', U'').$$

Proof. Let $U'_S = (U_{S'} : S \subseteq S' \in E)$ for each $S \in E$ and observe that necessarily U'_S has distribution that factors recursively as $\prod_{S \in E} p(u_S | u_{S'} : S \subset S')$. By construction, we may take x'_j to depend only on the single auxiliary random variable $U_{\mathcal{S}(j)}$ and not on $U_{E \setminus \mathcal{S}(j)}$ so that the inputs are unchanged. Moreover, for each $B \in \mathcal{F}_{\downarrow}(E; \subseteq)$, as $U_{E \setminus B} \mapsto U'_{E \setminus B}$ is a one-to-one map, $I(U'_B; Y | U'_{E \setminus B}) = I(U_B; Y | U_{E \setminus B})$. Hence, $\mathcal{P}_{\subseteq}(x, U) = \mathcal{P}_{\subseteq}(x', U')$. With the language of Section **??**, define a new auxiliary random tuple via

$$U''_S = \begin{cases} (X_j : \mathcal{S}(j) = S) & S \notin \mathcal{M} \\ U'_S, (X_j : \mathcal{S}(j) = S) & S \in \mathcal{M} \end{cases}.$$

for each $S \in E$. Notably, the inputs are subsumed into the auxiliary random tuple. By construction, the auxiliary random tuple $U'' = (U''_S : S \in E)$ has distribution that recursively factors as $\prod_{S \in E} p(u''_S | u''_{\uparrow S \setminus S})$; that is, $U'' \in L(E; \subseteq)$. Moreover, for every $B \in \mathcal{F}_{\downarrow}(E; \subseteq)$, as

$$U'_{E \setminus B} \multimap U''_{E \setminus B} \multimap Y$$

$$U''_{E \setminus B} \text{ is a function of } U'_{E \setminus B},$$

the applying the same argument leading up to (3.21) provides that $\mathcal{P}_{\subseteq}(x', U') = \mathcal{P}_{\subseteq}(x'', U'')$. $\square$

3.4 Importance of the Inclusion Order

3.4.1 Rate Delegation

As in the two-user case, we can add power to the general K-user case of Han's coding scheme by delegating the rate load of the "more common" rate loads among "less common" rate loads. Specifically, with a fixed input distribution, we can achieve any rate tuple satisfying just the inequalities corresponding to the down-set lattice $\mathcal{F}_{\downarrow}(\subseteq)$ (but possibly violating some inequalities in $\mathcal{F}_{\downarrow}(=)\backslash\mathcal{F}_{\downarrow}(\subseteq)$), by delegating "more common" message rate loads as

$$R_S = \sum\nolimits_{S' \in E : S' \subseteq S} r_{(S',S)} \quad S \in E \tag{3.23}$$

among "less common" message rate loads to form effective rates

$$\tilde{R}_{S'} = \sum\nolimits_{S \in E : S' \subseteq S} r_{(S',S)} \quad S' \in E \tag{3.24}$$

which satisfy all of the inequalities corresponding to the power set of E. The technical challenge of such a result is in eliminating the rate-splits $(r_{(S',S)} : S', S \in E, S' \subseteq S)$ to leave only the target rates R_S. While the standard prescription in the literature to eliminating extraneous rates is to use the Fourier-Motzkin procedure, we find that such an approach is unwieldy in this case. Instead, we eliminate the rate-splits in an alternative manner: via the properties of polymatroids. While it requires polymatroidal structure, whenever polymatroidal structure is present, such a method might provide a much more **scalable** and **tractable** approach to projection. The details of this projective step are relegated to Appendix B.2.

3.4.2 Sufficiency of Successive Group Decoding

As in the two-user case, any point on the boundary of the K-user capacity region are achievable through successive group decoding along the inclusion order on the message index set. For the general K-user case, successive group decoding process along successive decoding chains:

Definition 9 (Successive Decoding Chain)**.** *Any chain $\emptyset = B_1 \subset B_2 \subset \cdots \subset B_k = E$ where, with respect to the inclusion order on E, the B_i are down-sets and their differences $G_i = B_{i+1}\backslash B_i$*

contain incomparable elements

$$S, S' \in G_i \implies S \not\subset S', S' \not\subset S. \tag{3.25}$$

Successive group decoding proceeds by first jointly decoding the messages in G_1, then jointly decoding the messages in G_2, and so forth. Points achievable through successive group decoding lie on a specific face of the achievable polymatroids. To show that these points are a face of the achievable polymatroid, we require the following two lemmas, where the latter one depends critically on the diminishing returns property of the defining polymatroid bounds.

Lemma 3.4.1. *For any successive decoding chain, the incomparability condition* (3.25) *assures that* $(B \cap B_{i+1}) \cup B_i$ *is a down-set with respect to the inclusion order for every subset* $B \subseteq E$ *and every* $i \in [1:k]$.

Lemma 3.4.2. *Let* $f : 2^E \mapsto \mathbb{R}_+$ *be a polymatroid function. Suppose there is a family* $\mathcal{F}$ *of subsets of* E *containing the length-*k *chain*

$$\emptyset = B_1 \subset B_2 \subset \cdots \subset B_k = E \tag{3.26}$$

and a point x *satisfying* [12]

$$x(B_i) = f(B_i) \text{ for all } i \in [1:k] \tag{3.27a}$$

$$x(B') \leq f(B') \text{ for all } B' \in \mathcal{F} \text{ satisfying} \tag{3.27b}$$

$$B_i \subseteq B' \subseteq B_{i+1} \text{ for some } i \in [1:k-1]\,. \tag{3.27c}$$

Then,

$$x(B'') \leq f(B'') \text{ for all } B'' \subseteq E \text{ satisfying} \tag{3.27d}$$

$$(B'' \cap B_{i+1}) \cup B_i \in \mathcal{F} \quad \text{for } i \in [1:k]\,. \tag{3.27e}$$

Proof. Suppose $B'' \subseteq E$ satisfies (3.27e). Let $T_i = (B'' \cap B_{i+1}) \backslash B_i$. Then for each $i = 0, 1, \ldots, k$,

$$x(T_i) = x(T_i \cup B_i) - x(B_i) = x((B' \cap B_{i+1}) \cup B_i) - x(B_i) \overset{(i)}{\leq} f((B' \cap B_{i+1}) \cup B_i) - x(B_i) \tag{3.28}$$

[12] In this section, we adopt the notation $x(B) = \sum_{S \in B} x_S$.

where (i) follows by (3.27e) and (3.27c). Hence,

$$\begin{aligned} x(B'') &= x(T_k) + \sum_{j=0}^{k-1} x(T_j) \\ &\overset{(i)}{\leq} f(T_k \cup B_k) - f(B_k) + \sum_{j=0}^{k-1} x(T_j) \\ &\overset{(ii)}{\leq} f(T_k \cup T_{k-1} \cup B_{k-1}) - f(T_{k-1} \cup B_{k-1}) + \sum_{j=0}^{k-1} x(T_j) \\ &\overset{(iii)}{\leq} f(T_k \cup T_{k-1} \cup B_{k-1}) - f(B_{k-2}) + \sum_{j=0}^{k-2} x(T_j) \\ &\;\;\vdots \\ &\leq f(T_k \cup T_{k-1} \cup \cdots \cup T_0) = f(B''), \end{aligned}$$

where $(i), (iii)$ follow by (3.28) and (ii) follows by the diminishing returns property of submodular and monotonic functions (i.e. $f(A \cup B) - f(B) \geq f(A \cup B \cup C) - f(A \cup B)$ for any subsets A, B, C). The remaining steps follow by induction. □

An implication of this lemma is that if the $\mathcal{F}$ is the down-set lattice set family under the inclusion order, (3.26) is a Successive Decoding chain, and the point x satisfies the conditions (3.27), then Lemmas 3.4.1 and 3.4.2 assure that this point is in the polymatroid $\mathcal{P}(f) = \{x \in \mathbb{R}_+^E : x(B) \leq f(B) \ \forall \ B \subseteq E\}$. With this implication in hand, we demonstrate that all points in the capacity region lie on some successive group decoding face and are achievable without superposition coding:

Theorem 3.4.3. *Consider the K-user DM MAC with message index set E. Then the capacity region is dominated by rate points that satisfy*

$$\begin{aligned} &R(B_i) = I(U_{B_i}; Y | U_{E \setminus B_i}) \quad \textit{for all } i \in [1:K] \\ &R(B') \leq I(U_{B'}; Y | U_{E \setminus B'}) \quad \textit{for all down-sets } B' \textit{ w.r.t.}^{13} \textit{ the } \textbf{\textit{inclusion}} \textit{ order satisfying} \\ &\qquad\qquad\qquad\qquad\qquad\qquad B_i \subseteq B' \subseteq B_{i+1} \textit{ for some } i \in [1:k-1]. \end{aligned}$$

for some input distribution (x, U) in the up-set lattice conditional model with respect to the ***discrete*** *order and for some Successive Decoding chain $\{B_1, \ldots, B_k\}$.*

[13] with respect to

Proof. Equip E with a superposition order. By Theorem 3.3.1, the capacity is given as the union of all polymatroids $\mathcal{P}_{\leq}(x'', U'') = \left\{ R \in \mathbb{R}_+^E : R(B) \leq I\left(U''_B; Y | U''_{E\setminus B}\right) \ \forall\ B \in \mathcal{F}_{\downarrow}(E; \leq) \right\}$, over all possible input tuples $(x'', U'') \in \mathcal{Q}(E; \leq)$. Take one specific such polymatroid by fixing an input tuple $(x'', U'') \in \mathcal{Q}(E; \leq)$. By Lemma 3.3.3, there is an input tuple $(x, U) \in \mathcal{Q}(E, =)$ so that $\mathcal{P}_{\leq}(\rho_{x'', U''}) = \mathcal{P}_{\leq}(\rho_{x,U})$, where we define

$$\rho_{x,U}(B) = I(W_B; Y | U_{E\setminus B}). \tag{3.29}$$

As the auxiliary random variables $(U_S : S \in E)$ are independent, the bound (3.29) is a submodular and monotonic over the power set of E.[14] Now, consider some maximal point R within the above polymatroid. By the greedy algorithm for maximizing the weighted sum-rate over polymatroids, this point lies on face given by $R(E) = I(U_E; Y)$. We will construct by induction a new input distribution (x', U') such that this maximal rate point is on the successive group decoding face of the polymatroid corresponding to (x', U').

- **Root Case** By assumption, the down-set chain $\emptyset = B_1 \subset B_2 = E$ has

  $$R(B_i) = \rho(B_i) \text{ for all } i \in \{1, 2\}$$
  $$R(B') \leq \rho(B') \text{ for all } B' \in \mathcal{F} \text{ satisfying } B_1 \subseteq B' \subseteq B_2.$$

- **Inductive Step** Assume the length-k down-set chain $\emptyset = B_1 \subset B_2 \subset \cdots \subset B_k = E$ is not a successive decoding chain and R' satisfies

  $$R(B_i) = \rho(B_i) \text{ for all } i \in [1 : k]$$
  $$R(B') \leq \rho(B') \text{ for all } B' \in \mathcal{F} \text{ satisfying } B_i \subseteq B' \subseteq B_{i+1} \text{ for some } i \in [1 : k-1].$$

  Fix i to be the smallest index such that $G_i = B_{i+1} \setminus B_i$ has comparable elements under the inclusion order. Then there is a pair $(S, S') \in G_i$, ordered as $S \subset S'$, so that S is minimal with respect to the inclusion order in G_i (that is, no $S'' \in G_i$ has $S'' \subset S$). For this pair, let $B'_1, \ldots, B'_m$ be the down-sets containing S but not S', and satisfying $B_i \subset B'_k \subset B_{i+1}$. This list contains at least one element, $B'_1 = B_i \cup \{S\}$. There are two cases

[14] Per Lemma 3.2.1

(1) If there is a B'_k with $R(B'_k) = \rho_{x,U}(B'_k)$, continue to the next inductive step with the chain $\{B_1, \ldots, B_i, B'_k, B_{i+1}, \ldots, B_k\}$ in place of $\{B_1, \ldots, B_k\}$.

(2) Suppose $R(B'_k) < \rho_{x,U}(B'_k)$ for all $k \in [1:m]$. Let $(f_{S,S'}, V_S, V_{S'})$ be a split as in Definition 8 with parameter $\epsilon_{S,S'}$. Set $U_{S''} = U_{S''}$ for $S'' \in E \backslash \{S, S'\}$ and $U'_S = V_S$, $U'_{S'} = (U_{S'}, V_{S'})$, Update the functions x_j with $j \in S$ by composing them with the function $f_{S,S'}$; call these new composed function x'_j. For $j \notin S$, simply set $x'_j = x_j$. Each $\rho_{x',U'}(B'_k) = I(V_S, U_{B'_k \backslash \{S\}}; Y | V_{S'}, U_{E \backslash (B'_k \cup \{S'\})})$ is a continuous function of $\epsilon_{S,S'}$ with $\rho_{x',U'}(B'_k) = \rho_{x,U}(B'_k)$ for $\epsilon_{S,S'} = 1$. Moreover, as

$$\begin{aligned} R_S &= R(B'_1) - R(B_i) \\ &< \rho_{x,U}(B_i \cup \{S\}) - \rho_{x,U}(B_i) \\ &= I(U_{B_i}; Y | U_{E \backslash B_i}) - I(U_{B_i \backslash \{S'\}}; Y | U_{E \backslash (B_i \backslash \{S'\})}) \\ &= I(U_S; Y | U_{E \backslash B_i}) \end{aligned} \tag{3.30}$$

and $\rho_{x',U'}(B_i \cup \{S\}) - \rho_{x',U'}(B_i) = I(V_S; Y | U_{E \backslash B_i})$ is a continuous function of $\epsilon_{S,S'}$ onto the interval $[0, I(U_S; Y | U_{E \backslash B_i})]$, we know by the intermediate value theorem that there exists a $\epsilon_{S,S'} > 0$ such that for some $k \in [1:m]$,

$$\begin{aligned} R(B'_k) &= \rho(B'_k) \\ R(B'_j) &\leq \rho(B'_j). \qquad j \neq k, j \in [1:m] \end{aligned}$$

By Lemma 3.4.2, we are assured that all other which held at the beginning of this step continue to hold. Continue to the next inductive step with (x', U') in place of (x, U) and $B_1, \ldots, B_i, B'_k, B_{i+1}, \ldots, B_k$ in place of $\{B_1, \ldots, B_k\}$.

□

Chapter 4

MIMO Multiple Access Channel with General Message Sets

4.1 Introduction

While the discrete memoryless channel permits a clean development of the theory related to superposition coding, a particular class of channels is of practical importance: linear, additive Gaussian noise channels, which model the wireless channel. Of particular interest is the scenario where the transmitters and receiver are equipped with multiple antennas (i.e. the so-called multiple-input multiple-output (MIMO) setting).

For guidelines into practical design, we seek a characterization of the optimal input and auxiliary random tuples. Reasonably, we expect such optimal distributions to be Gaussian - for example, in the non-cooperative Gaussian MAC, the set of optimal inputs are Gaussian as a result of the principle that, subject to a covariance constraint, the Gaussian distribution maximizes entropy [20] (and conditional entropy [104]). However demonstrating this rigorously for the MAC with common messages is more subtle - in addition to covariance constraints on the channel inputs, there also Markov constraints on the auxiliary random variables that need be satisfied. Simply replacing the joint ensemble of the channel inputs and auxiliary random variables with a corresponding Gaussian random variable of the same covariance in general fails to preserve the desired Markov relations.

As observed in the two- and three-user cases, Bross **et al.** [10] and Wigger **et al.** [118], this difficulty can be circumvented by first replacing the auxiliary tuple with a clever choice of an intermediate auxiliary tuple and then subsequently replacing this intermediate auxiliary tuple with

a Gaussian tuple of the same covariance. Our key contribution is to use the formalism of order and lattice conditional independence to provide a simpler perspective on these prior maximum entropy methods for the two- and three-user case that generalizes naturally to the K-user case. In doing so, we find that the K-user capacity region capacity can be simply parameterized as a union of polymatroids over a convex set of admissible covariance matricies. Moreover, as the mutual information bounds which define the constituent polymatroids are concave in the admissible covariance matrices, the computation of the optimal input covariances can now be cast as a convex optimization problem, which in principle may be solved with efficient computation routines.

4.2 System Model and Preliminaries

4.2.1 Channel and Source Messages

The K-user Gaussian MIMO multiple access channel is defined, for the t-th channel use, to be

$$\mathbf{Y}_t = \underbrace{\begin{bmatrix}\mathbf{H}_1 & \dots & \mathbf{H}_K\end{bmatrix}}_{\mathbf{H}} \begin{bmatrix}\mathbf{X}_{1,t} \\ \vdots \\ \mathbf{X}_{K,t}\end{bmatrix} + \mathbf{Z}_t,$$

where the channel output $\mathbf{Y}_t$ is a $r \times 1$ vector, $\mathbf{H}_j$ is a channel gain matrix of size $r \times t_j$, the channel input of the j-th user is $\mathbf{X}_{j,t}$, and the circularly symmetric Gaussian noise $\mathbf{Z}_t$ is of size $r \times 1$ with identity covariance. The channel is assumed to be memoryless so that additive noise sequence $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ has elements that are distributed independently and identically. We adopt an average total power constraint, where when communicating over a block of n channel uses, the input sequence must obey

$$\frac{1}{n}\sum_{t=1}^{n} \|\mathbf{X}_{j,t}\|^2 \leq nP_j \tag{4.1}$$

for some $P_j > 0$ and for each transmitter $j \in [1 : K]$.

A code is defined as in (2.1.3). In this setting with a cost constraint on the inputs, we require an achievable rate tuple to not only have a vanishing average probability of error, but to also satisfy

the power constraint (4.1) for each n. Define the capacity region $\mathcal{C}_K(P_1, \ldots, P_K)$ to be the closure of the set of all such achievable rate tuples.

4.2.2 Interlocking Mulitvariate Gaussian Distributions

As in the discrete memoryless case, capacity will be in terms of auxiliary random tuples, with one auxiliary variable per message. In the Gaussian case, it will suffice to consider auxiliary random variables which are jointly Gaussian with each other and the channel inputs. Moreover, it suffices to represent these Gaussian auxiliary variables in the following specific manner. Throughout this chapter, we equip the message index set E with the inclusion order so that $S \leq S'$ if and only if $S \subseteq S'$.

If a vector $\mathbf{w}_S$ is indexed by a non-empty subset $S \subseteq [1:K]$, then we take it to have size $(\sum_{j \in S} t_j) \times 1$. We partition this vector in a specific manner: with $\{j_1, \ldots, j_{|S|}\} = S$ as an increasing enumeration of S (that is, $j_1 < j_2 < \ldots < j_{|S|}$), we partition the vector w_S as

$$\mathbf{w}_S = \begin{bmatrix} \mathbf{w}_{S,j_1} \\ \vdots \\ \mathbf{w}_{S,j_{|S|}} \end{bmatrix} \qquad \mathbf{w}_{S,j} \text{ is of size } t_j \times 1. \tag{4.2}$$

Such a vector could be used to cooperativey beamform across the joint array of antennas offered by the $|S|$ users listed in S to send a common message M_S. To compare the beams of different common messages to each other, it is useful to interpret $\mathbf{w}_S$ as a cooperative beam form across the joint array of antennas offered by all the transmitters through by the vector $\mathbf{P}_S \mathbf{w}_S$ of size $(t_1 + t_2 + \cdots + t_K) \times 1$, where for each $S \subseteq [1:K]$, $\mathbf{P}_S$ is the coordinate embedding $\mathbb{C}^{\sum_{j \in S} t_j} \mapsto \mathbb{C}^{t_1 + \cdots + t_K}$ defined implicitly by

$$\mathbf{P}_S \mathbf{w}_S = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_K \end{bmatrix} \qquad \mathbf{v}_j = \begin{cases} \mathbf{w}_{S,j} & j \in S \\ 0 & \text{else} \end{cases}$$

for every $(\sum_{j \in S} t_j) \times 1$ complex vector $\mathbf{w}_S$ (see Figure 4.1).

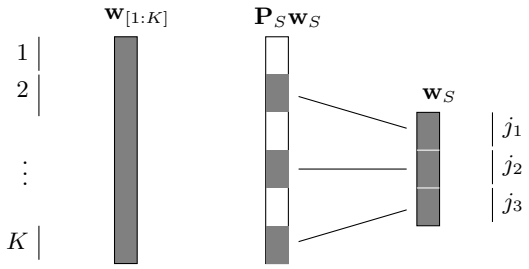


Figure 4.1: Common message beamforming where $S = \{j_1, j_2, j_3\} \subseteq [1:K]$ is enumerated in ascending order: $j_1 < j_2 < j_3$. The gray regions represent the vector partitions that may be non-zero.

Henceforth, any random variable $\mathbf{W}_S$ (with $S \subseteq [1:K]$) is tacitly taken to have the dimensions and partitions defined by (4.2). Similarly, if $\mathbf{K}_S = \mathrm{Cov}(\mathbf{W}_S, \mathbf{W}_S)$, let $\mathbf{K}_{S,jk} = \mathrm{Cov}(\mathbf{W}_{S,j}, \mathbf{W}_{S,k})$ be the block partition of this covariance corresponding to the covariance among the j-th and k-th partitions of the vector $\mathbf{W}_S$. If $\mathbf{W}_S \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_S)$, then a Shur decomposition $\mathbf{K}_S = \mathbf{B}_S \Lambda_S \mathbf{B}_S^T$ where $\mathbf{B}_S$ has orthonormal columns and Λ_S is diagonal, positive definite, and of size $r_S \times r_S$, provides that an equivalent representation is

$$\mathbf{W}_S = \sum_{i=1}^{r_S} \mathbf{b}_S^{(i)} \tilde{w}_{S,i} \qquad \mathbf{B}_S = \begin{bmatrix} \mathbf{b}_S^{(1)} & \cdots & \mathbf{b}_S^{(r_S)} \end{bmatrix}$$

where $w_{S,i} \sim \mathcal{CN}(0, (\Lambda_S)_{ii})$ for each $1 \leq i \leq r_S$, the $\tilde{w}_{S,1}, \ldots, \tilde{w}_{S,r_S}$ are independent, and the ith column vector $\mathbf{b}_S^{(i)}$ of $\mathbf{B}_S$ has the dimensions and partitions of (4.2).

Suppose that $(\mathbf{W}_S : S \in E)$ is a tuple of independent zero-mean jointly Gaussian random variables. Define the auxiliary random tuple $(\mathbf{U}_S : S \in E)$ via

$$\mathbf{U}_S = \mathbf{P}_S^T \left(\sum_{S' \in \uparrow S} \mathbf{P}_{S'} \mathbf{W}_{S'} \right) \tag{4.3}$$

for each $S \in E$. By construction, the distribution for the random tuple $(\mathbf{U}_S : S \in E)$ factors as $\prod_{S \in E} p(\mathbf{U}_S | \mathbf{U}_{\uparrow S \setminus S})$ and hence is a member of the up-set lattice conditional independence model corresponding the inclusion order. When the message index set E is closed under intersections, a simple pair-wise covariance characterization of lattice conditional independence exists. This follows as for any pair $R, S \in E$ either $(\uparrow R) \cap (\uparrow S) = \emptyset$ or $(\uparrow R) \cap (\uparrow S) = \uparrow T$ where

$$\bigcap \left\{ T' :\in (\uparrow S) \cap (\uparrow R) \right\} = T \in E.$$

So, for any two users $s, r \in [1:K]$ with $s \in S$ and $r \in R$,

$$\begin{aligned}\mathrm{Cov}(\mathbf{U}_{S,s}, \mathbf{U}_{R,r}) &= \sum_{S' \in (\uparrow S) \cap (\uparrow R)} \mathrm{Cov}(\mathbf{W}_{S',s}, \mathbf{W}_{S',r}) \\ &= \begin{cases} \sum_{S' \in \uparrow T} \mathrm{Cov}(\mathbf{W}_{S',s}, \mathbf{W}_{S',r}) & (\uparrow R) \cap (\uparrow S) \neq \emptyset \\ 0 & (\uparrow R) \cap (\uparrow S) = \emptyset \end{cases} \\ &= \begin{cases} \mathrm{Cov}(\mathbf{U}_{T,s}, \mathbf{U}_{T,r}) & (\uparrow R) \cap (\uparrow S) \neq \emptyset \\ 0 & (\uparrow R) \cap (\uparrow S) = \emptyset \end{cases}. \end{aligned} \tag{4.4}$$

Conversely, if we know that (4.4) holds for all $S, R \in E$ and $s \in S, r \in R$, $E[\mathbf{U}_S] = \mathbf{0}$ for all $S \in E$, and $(\mathbf{U}_S : S \in E)$ is jointly Gaussian, then as the mean and covariance uniquely specify Gaussian variables, the tuple $(\mathbf{U}_S : S \in E)$ must be a member of the up-set lattice conditional independence model with respect to the inclusion order. Moreover, the tuple must be associated with jointly Gaussian independent variables $(\mathbf{W}_S : S \in E)$ through (4.3), which are recoverable through Gram-Schmidt orthogonalization: if $S_1, \ldots, S_M$ is an exhaustive, never-increasing enumeration of the elements of E, then we successively re-construct, for $i = 1, \ldots, M$,

$$\begin{aligned}\mathbf{W}_{S_i,s} &= \mathbf{U}_{S_i,s} - E[\mathbf{U}_{S_i,s} | (\mathbf{U}_{S_j,s} : S_j \in \uparrow S_i)] \\ &= \mathbf{U}_{S_i,s} - E[\mathbf{U}_{S,s} | (\mathbf{W}_{S_j,s} : S_j \in \uparrow S_i)] \\ &= \mathbf{U}_{S_i,s} - \sum\nolimits_{S_j \in \uparrow S_i} \mathbf{W}_{S_j,s} \end{aligned} \tag{4.5}$$

for each $s \in S_i$.

4.3 Static Channel: K-User Capacity Region

Assume, without loss of generality, that the message index set E is closed under intersections (zero-rate messages may always be appended if need be).

Theorem 4.3.1. *The capacity region of the K-user MAC with common and private messages*

indexed by E is the set of rate tuples $(R_S : S \in E)$ satisfying

$$\sum_{S \in B} R_S \leq \log \det \left(\mathbf{I} + \mathbf{H} \left(\sum_{S \in B} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^* \right) \mathbf{H}^* \right)$$

for all down-sets B of E with respect to the inclusion order and for some set of covariances satisfying

$$\sum_{j \in S \in E} \operatorname{tr}(\mathbf{K}_{S,jj}) \leq P_j \qquad j \in [1:K], \tag{4.6}$$

where the structure of the covariance matrices $(\mathbf{K}_S : S \in E)$ is as specified in Section 4.2.2.

The key technical challenge in in the converse, where we demonstrate that **Guassian** inputs suffice, a step is accomplished by a maximum entropy lemma. Details of the proof of this theorem are provided in Section 4.3.3. The relationship between the admissible covariance matrices in Theorem 4.3 and the corresponding Gaussian input is precisely that as described in (4.2.2) where to each $S \in E$ we assign an independent $\mathbf{W}_S \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_S)$ and the inputs are related to these auxiliary random variables via

$$\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} = \begin{bmatrix} \sum_{1 \in S \in E} \mathbf{W}_{S,1} \\ \vdots \\ \sum_{K \in S \in E} \mathbf{W}_{S,K} \end{bmatrix} = \sum_{S \in E} \mathbf{P}_S \mathbf{W}_S. \tag{4.7}$$

The Gaussian restriction has an important advantage. The results from the previous chapter provide that above rate region is a union of **polymatroids**. While the discrete memoryless channel is described over a non-convex set of permissible input distributions, the Gaussian channel over the convex set of permissible jointly Gaussian input tuples. The result is that the calculation of the optimal input covariances reduces to the maximization of a concave function over a convex set, at task that is theoretically computationally efficient.

The coding complexity of Theorem is low: as in the discrete memoryless case, joint decoding is uncesssary, and successive group decoding suffices to attain all points in the capacity region. This is substantiated by the following counterpart to Theorem 3.4.3 for the DM MAC:

Theorem 4.3.2. *The capacity of the K-user MAC with common and private messages indexed by E is the set of rate tuples $(R_S : S \in E)$ satisfying*

$$\sum_{S \in B' \setminus B_i} R_S \leq \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S \in B' \setminus B_i} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right) - \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S \in B_i} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right)$$

for every $1 \leq i \leq k$ and every $B' \subseteq E$ satisfying $B_i \subset B' \subseteq B_{i+1}$ for some Successive Decoding Chain $B_1 \subset \cdots \subset B_k$ and some set of covariances satisfying (4.6).

Proof. Achievability follows simply, and we outline this below. Fix a Successive Decoding Chain $B_1 \subset \cdots \subset B_k$ and a tuple of independent and jointly Gaussian $(\mathbf{W}_S : S \in E)$ where, for each $S \in E$, $\mathbf{W}_S$ has covariance $\mathbf{K}_S$. Induce a joint distribution between the channel inputs and auxiliary variables by constructing the channel inputs as sums of the appropriate partitions of the auxiliary codewords as in (4.7). Such a strategy can be interpreted as treating the MAC as though each message index $S \in E$ corresponds to a distinct user with private message M_S, $(\sum_{j \in S} t_j)$-dimensional input $\mathbf{U}_S$, and channel gain matrix $\mathbf{HP}_S$ effectively yielding the following special case of the non-cooperative MAC:

$$\mathbf{Y} = \sum_{S \in E} (\mathbf{HP}_S)\mathbf{W}_S + \mathbf{Z}.$$

Then by standard random coding it is clear that we can achieve any rate tuple satisfying the hypothesis of the theorem by decoding, successively, the groups of messages in $B_{i+1} \setminus B_i$ for $i \in \{1, \ldots, k-1\}$.

To show that these conditions are necessary for reliable communication, we use Theorem 4.3 and mimic the proof of Theorem 3.4.3 for the Gaussian setting. Details are in Appendix C.1. □

4.3.1 Single Antenna (SISO) Specialization

In the single-input single-output case (i.e. where $t_1 = \cdots = t_K = r = 1$), an application of Cauchy-Schwartz demonstrates that it suffices to only consider rank-one common message covariance matrices in the characterization in Theorem 4.3:

Corollary 4.3.3. *The capacity of the SISO MAC with common messages is given by the set of rate tuples satisfying*

$$\sum_{S\in\mathcal{B}} R_S \le \log\left(1+\sum_{S\in\mathcal{B}} p_S\right)$$

for all down-sets $\mathcal{B}$ of E with respect to the inclusion order for some set of received powers $(p_S : S \in E)$ and load-balance vectors $(\rho_S : S \in E)$ satisfying the power constraint

$$\sum_{j\in S\in E} \frac{\rho_{S,jj}^2 p_S}{|h_j|^2} \le P_j \qquad j \in [1:K] \tag{4.8}$$

and having $\rho_S = \begin{bmatrix}\rho_{S,j_1} & \cdots & \rho_{S,j_{|S|}}\end{bmatrix}$ be on the $|S|$-dimensional simplex for each $S \in E$.

Proof. Consider an admissible set of covariances in Theorem 4.3. In the SISO case, each block diagonal $\mathbf{K}_{S,jj}$ is simply a real, non-negative scalar. For each $S \in E$, define the (maximal) received power for the signal corresponding to this auxiliary random variable as

$$p_S = \sum_{i\in S}\sum_{j\in S} |h_i||h_j|\sqrt{\mathbf{K}_{S,(ii)}\mathbf{K}_{S,(jj)}} \tag{4.9a}$$

$$\rho_{S,j}^2 p_S = |h_j|^2 \mathbf{K}_{S,(jj)} \tag{4.9b}$$

for each $j \in S \in E$. By (4.9), for each $S \in E$ we have $\left(\sum_{j\in S}\rho_{S,j}\right)^2 = \frac{1}{p_S}\left(\sum_{j\in S}|h_j|\sqrt{\mathbf{K}}_{S,(jj)}\right)^2 = 1$ so that $\rho_S = \begin{bmatrix}\rho_{S,j_1} & \cdots & \rho_{S,j_{|S|}}\end{bmatrix}$ is on the $|S|$-dimensional simplex. Define, for each $S \in E$, the unit-rank covariance matrices

$$\tilde{\mathbf{K}}_S = \tilde{\mathbf{B}}_S\tilde{\mathbf{B}}_S^* \qquad\qquad \tilde{\mathbf{B}}_S = \begin{bmatrix}\rho_{S,j_1}\sqrt{p_S}/h_{j_1}\\ \vdots \\ \rho_{S,j_{|S|}}\sqrt{p_S}/h_{j_{|S|}}\end{bmatrix}. \tag{4.10}$$

As $\mathbf{K}_{S,jj} = \tilde{\mathbf{K}}_{S,jj}$ for all pairs (j,S) with $j \in S \in E$, the covariance matrices $\tilde{\mathbf{K}}_S$ also satisfy the power constraint (4.6) and are thus also admissible. Moreover, by Cauchy-Schwartz, $\mathbf{h}\mathbf{P}_S\mathbf{K}_S\mathbf{P}_S^T\mathbf{h}^* \le p_S = \mathbf{h}\mathbf{P}_S\tilde{\mathbf{K}}_S\mathbf{P}_S^T\mathbf{h}^*$ for every $S \in E$ and

$$\begin{aligned}\log\left(1+\mathbf{h}\left(\sum\nolimits_{S\in\mathcal{B}}\mathbf{P}_S\mathbf{K}_S\mathbf{P}_S^T\right)\mathbf{h}^*\right) &\le \log\left(1+\sum\nolimits_{S\in\mathcal{B}}p_S\right)\\ &= \log\left(1+\mathbf{h}\left(\sum\nolimits_{S\in\mathcal{B}}\mathbf{P}_S\tilde{\mathbf{K}}_S\mathbf{P}_S^T\right)\mathbf{h}^*\right).\end{aligned}$$

for each down-set B. □

4.3.2 Relationship to Previous Results

Our successive group decoding result was inspired by, and generalizes, an observation for the two-user SISO MAC with two private and one common message studied by Liu and Ulukus [65]. There, it was shown that the capacity region can be achieved by independent Gaussian codebooks per message and by first decoding the common message and then jointly decoding the two private messages.

For the three-user channel, our results reproduce those of Wigger and Kramer [118], and uncover polymatroid structure in their capacity characterization that they appear to have not noticed. Succintly, [118] states that the three-user MAC with all possible common and private messages (that is, with message index set $E = \{1, 2, 3, 12, 13, 23, 123\}$[1]), has capacity region given by the set of rate tuples $(R_{123}, R_{12}, R_{13}, R_{23}, R_1, R_2, R_3)$ which satisfy

$$\sum_{S \in B} R_S \leq I\left(\mathbf{V}_B; \mathbf{Y} | \mathbf{V}_{E \setminus B}\right) \qquad \text{for all down-sets } B \in \mathcal{F}_{\downarrow}(E; \subseteq) \tag{4.11}$$

for some jointly Gaussian input tuple $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{V}_{12}, \mathbf{V}_{13}, \mathbf{V}_{23}, \mathbf{V}_{123}$ having $\mathbf{X}_j = \mathbf{V}_j$ for $j \in [1:3]$ and satisfying the Markov and independence constraints

$$\mathbf{X}_1 \multimap \mathbf{V}_{12}, \mathbf{V}_{13}, \mathbf{V}_{123} \multimap \mathbf{X}_2, \mathbf{X}_3, \mathbf{V}_{23} \tag{4.12a}$$

$$\mathbf{X}_2 \multimap \mathbf{V}_{12}, \mathbf{V}_{23}, \mathbf{V}_{123} \multimap \mathbf{X}_1, \mathbf{X}_3, \mathbf{V}_{13} \tag{4.12b}$$

$$\mathbf{X}_3 \multimap \mathbf{V}_{13}, \mathbf{V}_{23}, \mathbf{V}_{123} \multimap \mathbf{X}_1, \mathbf{X}_2, \mathbf{V}_{12} \tag{4.12c}$$

$$\mathbf{V}_{12}, \mathbf{V}_{13}, \mathbf{V}_{23}, \mathbf{V}_{123} \text{ are independent.} \tag{4.12d}$$

Moreover, each auxiliary random random variable $\mathbf{V}_S$ can be assumed to have $(\sum_{j \in S} t_j)$-dimensions. To connect this with our framework, let us reflect on what the jointly Gaussian assumption and the dependency constraints in (4.12) entail. In particular, we must have that the conditional covariance

$$\operatorname{Cov}\left(\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\\ \mathbf{X}_3\end{bmatrix}, \begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\\ \mathbf{X}_3\end{bmatrix} \middle| \mathbf{V}_{12}, \mathbf{V}_{13}, \mathbf{V}_{23}, \mathbf{V}_{123}\right) = \begin{bmatrix}\mathbf{Q}_1 & & \\ & \mathbf{Q}_2 & \\ & & \mathbf{Q}_3\end{bmatrix}$$

[1] For brevity of notation, we append the elements of S in a string rather than in the usual set notation; e.g. 12 rather than $\{1, 2\}$ so that we write U_{12} rather than $U_{\{1,2\}}$.

is block-diagonal where $\mathbf{Q}_j$ is a $t_j \times t_j$ positive semi-definite matrix for each $j \in \{1,2,3\}$. Hence, we may introduce three independent random vectors $(\mathbf{V}_{j,j} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_j) : j \in \{1,2,3\})$ that are jointly Gaussian with, and independent of, $(\mathbf{V}_S : S \in E, |S| > 1)$ such that

$$\mathbf{X}_1 = \mathbf{V}_{1,1} + \mathbf{A}_{1,12}\mathbf{V}_{12} + \mathbf{A}_{1,13}\mathbf{V}_{13} + \mathbf{A}_{1,123}\mathbf{V}_{123} \tag{4.13a}$$

$$\mathbf{X}_2 = \mathbf{V}_{2,2} + \mathbf{A}_{2,23}\mathbf{V}_{23} + \mathbf{A}_{2,21}\mathbf{V}_{21} + \mathbf{A}_{2,123}\mathbf{V}_{123} \tag{4.13b}$$

$$\mathbf{X}_3 = \mathbf{V}_{3,3} + \mathbf{A}_{3,31}\mathbf{V}_{31} + \mathbf{A}_{3,32}\mathbf{V}_{32} + \mathbf{A}_{3,123}\mathbf{V}_{123} \tag{4.13c}$$

for some a set of matrices $(\mathbf{A}_{jS} \in \mathbb{C}^{t_j \times (\sum_{i \in S} t_i)} : j \in S \in E)$. By assigning $\mathbf{W}_S = (\mathbf{A}_{jS}\mathbf{V}_S : j \in S)$ for each $S \in E$, we may interpret these relations as

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{1,1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{V}_{2,2} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{V}_{3,3} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_{12,1} \\ \mathbf{V}_{12,2} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_{13,1} \\ \mathbf{0} \\ \mathbf{V}_{13,3} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{V}_{23,2} \\ \mathbf{V}_{23,3} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_{123,1} \\ \mathbf{V}_{123,2} \\ \mathbf{V}_{123,3} \end{bmatrix} \tag{4.14}$$

Moreover, in this case we may write the bounds in (4.11) as

$$I(\mathbf{V}_B; \mathbf{Y} | \mathbf{V}_{E\setminus B}) = I\left(\mathbf{W}_B; \mathbf{Y} \middle| \mathbf{W}_{E\setminus B}\right) = \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S \in B} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right).$$

In this form, it is easy to see that the bounds are submodular, monotonic, and normalized in subsets $B \subseteq E$. By the development in the previous chapter, as the defining inequalities above are over a lattice set family (the down-set lattice), they constitute a polymatroid.

4.3.3 Achievability and Maximum Entropy

Let's return to the proof of Theorem 4.3.2. To prove achievability, discretizing the corresponding results for the DM case. To this end, consider auxiliary random tuples $U_E = (U_S : S \in E)$ which satisfy

$$\mathbf{X}_j = x_j((U_S : j \in S \in E)) \text{ for each } j \in [1:K] \tag{4.15a}$$

$$(U_S : S \in E) \in L(E; \subseteq), \tag{4.15b}$$

$$E[\|\mathbf{X}_j\|^2] \leq P_j \text{ for each } j \in [1:K], \tag{4.15c}$$

where $\mathbf{X}_j$ is a $t_j \times 1$ vector, define $\mathcal{R}_K(\mathbf{X}_{[1:K]}, U_E)$ to be the polytope [2]

$$\left\{ R \in \mathbb{R}_+^E : \sum_{S \in B} R_S \leq I\left(U_B; \mathbf{Y} \middle| U_{E \setminus B}\right) \ \forall B \in \mathcal{F}_{\downarrow}(E; \subseteq) \right\},$$

where $\mathbf{Y} = \sum_{j=1}^K \mathbf{H}_j \mathbf{X}_j + \mathbf{Z}$ with $\mathbf{Z}$ as a zero-mean circularly symmetric r-dimensional Gaussian variable with identity covariance and independent of the random tuple $(\mathbf{X}_j : j \in [1:K]), (U_S : S \in E)$. Then from previous results for the discrete memoryless case [86], with appropriate modifications for channels with cost, we know that if we consider a series of increasingly finer and larger finite quantizations of the complex field $\mathbb{C}$ (as in Chapter 3, Section 4.1 of [31]), the capacity of the K-user MIMO MAC is given by

$$\mathcal{C}_K(P_1, \ldots, P_K) = \bigcup_{\substack{\mathbf{X}_{[1:K]}, U_E \\ \text{satisfying (4.15)}}} \mathcal{R}_K(\mathbf{X}_{[1:K]}, U_E).$$

Left as is, this characterization of capacity provides no insight into what distributions, among all admissible distributions, are optimal. Theorem 4.3 states that we can restrict without loss of generality the union above to be over a much smaller set and can be paraphrased as

$$\mathcal{C}_K(P_1, \ldots, P_K) = \bigcup_{\substack{\mathbf{X}_{[1:K]}, U_E \\ \text{satisfying (4.15)} \\ \text{and are jointly Gaussian}}} \mathcal{R}_K(\mathbf{X}_{[1:K]}, U_E)$$

We prove this in Section 4.3.3.1, but before doing so, we demonstrate how the above formulation matches the formulation in Wigger and Kramer [118] for the three-user MIMO MAC with common messages. A key simplification in our formulation is a more explicit description of Markov dependencies which characterize the input distributions relevant to the characterization of the capacity region.

4.3.3.1 Max Entropy Lemma

Assume that the message index set E is closed under intersections, implying that both

$$\mathcal{S}(j) = \bigcap \{S : j \in S \in E\} \text{ and } \bigcap \{T' \in (\uparrow S) \cap (\uparrow S')\}$$

[2] $\mathbb{R}_+^E = \{x \in \mathbb{R}^E : x_S \geq 0 \ \forall \ S \in E\}$ is the positive orthant of $\mathbb{R}^E$, the real vector space with coordinates indexed by the elements of E. If E consists of M elements, then $\mathbb{R}^E$ may be identified with $\mathbb{R}^M$.

are elements of E. We remind the reader that within this Chapter, we have equipped E with the partial order of set inclusion.

Lemma 4.3.4 (Maximum Entropy subject to Lattice Conditional Independence and Covariance Constraints)**.** *Let $(U_S : S \in E)$ be an auxiliary tuple such that*

(P1) $\mathbf{X}_j = x_j((U_S : j \in S \in E))$ *for each* $j \in [1 : K]$

(P2) $(U_S : S \in E)$ *are up-set lattice conditionally independent* [3]

(P3) $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_j) = \mathbf{K}_j$ *for each* $j \in [1 : K]$.

Then there is a ***jointly Gaussian*** *choice $(\mathbf{U}_S^G : S \in E)$ satisfying (P1)-(P3) with conditional entropy $h(Y|\mathbf{U}_{E\setminus B}^G)$ at least as large as $h(Y|U_{E\setminus B})$ for every down-set B with respect to the inclusion order. Moreover, this choice satisfies the more restrictive (Q1)-(Q2) in addition to same covariance constraint (Q3) as in (P3).*

(Q1) $\mathbf{X}_j^G = \mathbf{U}_{S(j),j}^G$ *for each* $j \in [1 : K]$

(Q2) $(\mathbf{U}_S^G : S \in E)$ *are up-set lattice conditionally independent satisfying*

$$\mathbf{U}_S^G = \mathbf{P}_S^T \left(\textstyle\sum_{S' \in \uparrow S} \mathbf{P}_{S'} \mathbf{W}_{S'}^G \right) \quad \text{for each } S \in E$$

where the $\mathbf{W}_S^G \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_S)$ *are independent.*

(Q3) $\mathrm{Cov}(\mathbf{X}_j^G, \mathbf{X}_j^G) = \mathbf{K}_j$ *for each* $j \in [1 : K]$.

Hence,

$$\begin{aligned} I(U_B; Y|U_{E\setminus B}) = h\Big(\mathbf{Y}|U_{E\setminus B}\Big) - h(\mathbf{Z}) &\le h\Big(\mathbf{Y}^G \Big| \mathbf{U}_{E\setminus B}^G\Big) - h(\mathbf{Z}) \\ &= \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S\in B} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right) \end{aligned} \tag{4.16}$$

for every down-set B.

[3] Recall that this implies that the distribution factors as $p(U_E) = \prod_{S\in E} p(U_{\uparrow S})$.

Proof. Assume without loss of generality that $E[\mathbf{X}_j] = \mathbf{0}$. Define, for each $j \in S \in E$, $\mathbf{U}_{S,j} = E[\mathbf{X}_j|U_{\uparrow S}]$, which is a vector of size $t_j \times 1$. Let $\mathbf{U}_S$ be the vector obtained by stacking the partitions $\mathbf{U}_{S,j}$ as described in Section 4.2.2. We make key three observations:

(1) Each $\mathbf{U}_S$ is a function of the tuple $U_{\uparrow S} = (U_{S'} : S \subseteq S')$. Consequently, when B is an element of the up-set lattice, the tuple $\mathbf{U}_B = (\mathbf{U}_S : S \in B)$ is a function of the tuple $U_B = (U_S : S \in B)$.

(2) $\mathbf{X}_j = E[\mathbf{X}_j|U_{\uparrow \mathcal{S}(j)}] = \mathbf{U}_{\mathcal{S}(j),j}$ as each $\mathbf{X}_j$ is a function of $U_{\uparrow \mathcal{S}(j)} = (U_j : j \in S \in E)$.

(3) Pick any pair of users $s, r \in [1 : K]$ and any pair of message set indices $S, R \in E$ such that $s \in S$ and $r \in R$ and consider

$$\mathrm{Cov}(\mathbf{U}_{S,s}, \mathbf{U}_{R,r}) = E\Big[E[\mathbf{X}_s|U_{\uparrow S}]E[\mathbf{X}_r|U_{\uparrow R}]^*\Big].$$

Suppose that $(\uparrow S) \cap (\uparrow R) = \emptyset$. Then by the lattice conditional independence of the tuple $(U_S : S \in E)$ with respect to the up-set lattice, we have $U_{\uparrow S}$ is independent of $U_{\uparrow R}$. Hence,

$$E\Big[E[\mathbf{X}_s|U_{\uparrow S}]E[\mathbf{X}_r|U_{\uparrow R}]^*\Big] = E\Big[E[\mathbf{X}_s|U_{\uparrow S}]\Big]E\Big[E[\mathbf{X}_r|U_{\uparrow R}]^*\Big] = E[\mathbf{X}_s]E[\mathbf{X}_r^*] = 0$$

Suppose instead that $(\uparrow S) \cap (\uparrow R) \neq \emptyset$. Then, with $T = \bigcap\{T' \in (\uparrow S) \cap (\uparrow S')\}$ (which is in the message index set E by assumption),

$$\begin{aligned} E\Big[E[\mathbf{X}_s|U_{\uparrow S}]E[\mathbf{X}_r|U_{\uparrow R}]^*\Big] &= E\Bigg[\Big[E[\mathbf{X}_s|U_{\uparrow S}]E[\mathbf{X}_r|U_{\uparrow R}]^*\Big]\Big|U_{\uparrow T}\Bigg] \\ &= E\Big[\Big[E[\mathbf{X}_s|U_{\uparrow T}]E[\mathbf{X}_r|U_{\uparrow T}]^*\Big] = \mathrm{Cov}(\mathbf{U}_{T,s}, \mathbf{U}_{T,r}), \end{aligned}$$

where the first step follows by the tower property of conditional independence and the second step by the Markov relation $U_{\uparrow S} \multimap U_{\uparrow T} \multimap U_{\uparrow R}$ implied by the lattice conditional independence of the tuple $(U_S : S \in E)$. Hence, the covariance condition (4.4) holds, which is a sufficient condition for a Gaussian tuple of the same covariance as $(\mathbf{U}_S : S \in E)$ to be lattice conditionally independent over the up-set lattice.

Let $(\mathbf{U}_S^G : S \in E), \mathbf{X}_1^G, \ldots, \mathbf{X}_K^G, \mathbf{Y}^G$ be the jointly Gaussian tuple with the same joint covariance as $(\mathbf{U}_S : S \in E), \mathbf{X}_1, \ldots, \mathbf{X}_K, \mathbf{Y}$. By the discussion in Section 4.2.2, if the joint covariance of $(\mathbf{U}_S^G : S \in E)$ satisfies (4.4) for all (s, r) and (S, R) satisfying $s \in S \in E$ and $r \in R \in E$, then property (Q1) is satisfied, where the variables $(\mathbf{W}_S^G : S \in E)$ are recoverable through the iterative Gram-Schmidt orthogonalization provided in (4.5). The inputs are thus related to the independent $(\mathbf{W}_S^G : S \in E)$ via

$$\begin{bmatrix} \mathbf{X}_1^G \\ \vdots \\ \mathbf{X}_K^G \end{bmatrix} = \begin{bmatrix} \sum_{1 \in S \in E} \mathbf{W}_{S,1}^G \\ \vdots \\ \sum_{K \in S \in E} \mathbf{W}_{S,K}^G \end{bmatrix} = \sum_{S \in E} \mathbf{P}_S \mathbf{W}_S^G.$$

By construction, $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_j) = \mathrm{Cov}(\mathbf{X}_j^G, \mathbf{X}_j^G)$ and so (Q2) is satisfied. Moreover for any down-set B, $E \backslash B$ is an up-set and so

$$\begin{aligned} h\Big(\mathbf{Y} | (U_S : \ S \in E \backslash B)\Big) &\overset{(i)}{=} h\left(\mathbf{Y} \middle| (\mathbf{U}_S : S \in E \backslash B)\right) \\ &\overset{(ii)}{=} h\left(\mathbf{Y}^G \middle| (\mathbf{U}_S^G : S \in E \backslash B)\right) \\ &= \log\det\left((2\pi)\left(\mathbf{I} + \mathbf{H}\left(\sum_{S \in B} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right)\right) \end{aligned}$$

where (i) follows by the data-processing inequality, and (ii) follows as the Gaussian distribution maximizes conditional entropy subject to a covariance constraint [104]. □

4.4 Static Channel: Optimal Covariance

A fortuitous property of the capacity characterization in Theorem 4.3 is that the computation of optimal covariances can be carried out through convex programming. Several key properties enable this:

(1) The capacity region is a union of **polymatroids** and hence

(2) The polymatroid bounds are **concave** in the admissible covariances

(3) The admissible covariances lie in a convex set.

The former property enables the efficient optimization with respect to multiple criteria: for the weighted sum-rate capacity, the explicit formula for its vertices leads to a simple convex formulation of the boundary of the capacity region, where the convexity of the formulation follows by the remaining two properties. An alternative manner in which the polymatroidal property may be exploited, but for which we have not developed, would be to consider optimality criterion for "fairness" of users as in [68].

We first present a generic observation of the **convex geometry** of the capacity region that applies to any MAC with implicit cooperation (be it Gaussian or discrete memoryless). While this may be deduced from Theorem 4.3, we find it more enlightening to see that this can be characterized **a priori** with only the operational definition of the capacity region.

Lemma 4.4.1 (Dual Characterization of Capacity)**.** *The capacity region of the K-user MIMO MAC with implicit cooperation is convex and given by*

$$\bigcap_{\mu\in\mathcal{O}} \left\{ R \in \mathbb{R}_+^E : \sum_{S\in E} \mu_S R_S \le V_\mu \right\}, \tag{4.17}$$

the intersection of all of the supporting hyperplanes defined by the weights in

$$\mathcal{O} = \{\mu \in \mathbb{R}_+^E : \mu_S > \mu_{S'} \text{ only if } S \subset S'\}.$$

Proof. As time-sharing is permissible, the capacity region is convex and expressible as an intersection of half-planes

$$\mathcal{C}(P_1,\ldots,P_K) = \bigcap_{\mu\in\mathbb{R}_+^E} \left\{ R \in \mathbb{R}_+^E : \sum_{S\in E} \mu_S R_S \le V_\mu \right\}$$

for some collection of constants $\{V_\mu : \mu \in \mathbb{R}_+^E\}$. The key point is that the bounds corresponding to $\mu \in \mathbb{R}_+^E \backslash \mathcal{O}$ are redundant given those bounds in $\mu \in \mathcal{O}$ in the dual description (4.17). Fix a $\mu \in \mathbb{R}_+^E \backslash \mathcal{O}$ and consider the bound

$$V_\mu = \max_{R\in\mathcal{C}(P_1,\ldots,P_K)} \sum_{S\in E} \mu_S R_S.$$

Take an achievable code with rate R^* which achieves this maximum. Let $\mathcal{V} = \{S' \in E : \mu_S > \mu_{S'} \text{ for some } S \text{ with } S' \subset S \in E\}$ index the entries of μ which violate the defining property of the set $\mathcal{O}$.

Then for any $S' \in \mathcal{V}$, $R^*_{S'} = 0$. Else, were $R^*_{S'} > 0$, we could repurpose the codewords for M_S (where $S' \subset S \in E$) to send more common messages for the message index S' to achieve the rate tuple $\tilde{R}$ given by

$$\tilde{R}_{S'} = R^*_S + R^*_{S'} \qquad \tilde{R}_S = 0$$
$$\tilde{R}_{S''} = R^*_{S''} \text{ for all } S'' \in E\backslash\{S, S'\}$$

But then $\tilde{R}$ would achieve a larger weighted sum-rate point, $\sum_{S\in E} \mu_S R^*_S < \sum_{S\in E} \mu_S \tilde{R}_S$, a contradiction.

Hence, R^* remains a maximizing sum rate point for the weighted sum-rate $\sum_{S\in E} \tilde{\mu}_S R^*_S$ for each $\tilde{\mu}$ in the set $\mathcal{I}(\mu)$ defined by

$$\left\{ \tilde{\mu} \in \mathbb{R}^E_+ : \begin{array}{l} \tilde{\mu}_S = \mu_S \text{ for } S \in E\backslash\mathcal{V} \\ \\ \tilde{\mu}_{S'} < \mu_S \text{ for each } S' \in \mathcal{V} \text{ with } S' \subset S \in E \end{array} \right\}.$$

Hence $V_\mu = V_{\tilde{\mu}}$ for all $\tilde{\mu} \in \mathcal{I}(\mu)$. With the component-wise order over $\mathbb{R}^E_+$ defined by $\tilde{\mu} \leq \mu$ iff $\tilde{\mu}_S \leq \mu_S$ for each $S \in E$, there is a unique $\tilde{\mu}^* = \sup \mathcal{I}(\mu)$. By continuity, $V_\mu = V_{\tilde{\mu}^*}$. Moreover, as each rate tuple R in the capacity region satisfies

$$\sum_{S\in E} \tilde{\mu}_S R_S \leq \sum_{S\in E} \tilde{\mu}^*_S R_S \leq V_{\tilde{\mu}^*} = V_{\tilde{\mu}},$$

the inequalities corresponding to $\tilde{\mu} \in \mathcal{I}(\mu)$ are redundant. □

Remark For each $\mu \in \mathcal{O}$, there is an enumeration $\{S_1, \ldots, S_M\}$ of the message index set E such that $\mu_{S_1} \geq \mu_{S_2} \geq \cdots \geq \mu_{S_M}$ and there are k numbers $i_1 < i_2 < \cdots < i_k$ such that

$$B_j = \{S_{i_j}, S_{i_j+1} \ldots, S_{i_{j+1}}\}$$

forms a Successive Decoding chain.

Thus, to compute the maximal weighted-sum rates, it suffices to only consider those weights belonging to $\mathcal{O}$. Pick one such μ. Enumerate the message index set $E = \{S_1, \ldots, S_M\}$ so that $\mu_{S_1} \geq \cdots \geq \mu_{S_M}$. As $\mu \in \mathcal{O}$ we have that $B_i = \{S_1, \ldots, S_i\} \in \mathcal{F}_\downarrow(E; \subseteq)$ for each $i \in [1 : M]$. By

the capacity characterization in Theorem 4.3, and the explicit characterization of the vertices of a polymatroid, we know that for each admissible choice of $\mathbf{K}_E$ (that is, a set of covariances satisfying (4.1)),

$$\max_{R\in\mathcal{P}_{\mathcal{F}_\downarrow(E;\subseteq)}} \sum_{i=1}^{M} \mu_{S_i} R_{S_i} = \sum_{i=1}^{M} \mu_{S_i}(\rho(\mathbf{K}_E; B_i) - \rho(\mathbf{K}_E; B_{i-1})) = \sum_{i=1}^{M} (\mu_{S_i} - \mu_{S_{i+1}})\rho(\mathbf{K}_E; B_i)$$

where for convenience we wrote $B_0 = \emptyset$ and $\mu_{S_{M+1}} = 0$. Hence the optimization problem

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{M} \mu_{S_i} R_{S_i} \\ &\text{such that} && R \in \mathcal{C}_K(P_1, \ldots, P_K) \end{aligned}$$

is equivalent to maximizing, with $\delta_i = \mu_{S_i} - \mu_{S_{i+1}} \geq 0$ for $i \in [1 : M]$,

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{M} \delta_i \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{k=1}^{i} \mathbf{P}_{S_k}\mathbf{K}_{S_k}\mathbf{P}_{S_k}^*\right)\mathbf{H}^*\right) \\ &\text{such that} && \sum_{j\in S\in E} \operatorname{tr}(\mathbf{K}_{S,jj}) \leq P_j \qquad j \in [1 : K]\,. \end{aligned}$$

The constraints, given by (4.6), are linear and hence the set of the admissible covariance matrices is convex. Moreover, as $\log\det(\mathbf{I} + \mathbf{X})$ is a concave function over positive semidefinite matricies $\mathbf{X} \succeq \mathbf{0}$, the objective is concave. Thus, in principle, the above formulation is efficiently solvable by standard semi-definite programming techniques [9].

4.5 Fading Channel: K-User Capacity Region

4.5.1 Channel Model

We study the fading MAC where the base station is equipped with r antennas and the K mobile users are equipped with $t_1, \ldots, t_K$ antennas, respectively. The mobile users transmit synchronously on a time-block basis, while the fading process is taken to be stationary and ergodic. Let ν be a random variable representing the stationary distribution of the channel fading state with cumulative density function $F(\nu)$ and support $\mathcal{H}$. Thus, at each state $\nu \in \mathcal{H}$, the fading MAC can

be considered as a discrete time channel represented by

$$\mathbf{Y}(\nu) = \sum_{j=1}^{K} \mathbf{H}_j(\nu)\mathbf{X}_j(\nu) + \mathbf{Z}(\nu), \tag{4.18}$$

where $\mathbf{Y}(\nu) \in \mathbb{C}^{r\times 1}$ denotes the received signal vector. $\mathbf{X}_j(\nu) \in \mathbb{C}^{t_j\times 1}$ and $\mathbf{H}_j(\nu) \in \mathbb{C}^{r\times t_j}$ denote, respectively, the transmitted signal vector and channel matrix of user j, $j \in [1:K]$. $\mathbf{Z}(\nu) \in \mathbb{C}^{r\times 1}$ denotes the additive Gaussian noise at the receiver and it is assumed that $\mathbf{Z}(\nu) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

We assume that the transmitters have complete channel state information and so may adapt their transmission strategy to the channel state. In particular, they may dynamically allocate their available powers to fully exploit the benefitial fading states and avoid the detrimental fading states. For each state $\nu \in \mathcal{H}$, let $\mathbf{K}_j(\nu) = E[\mathbf{X}_j(\nu)\mathbf{X}_j(\nu)^*]$, where the expectation is taken over the codebook. Then under the long term power constraints considered here, any codebook for user j must satisfy

$$\int_{\mathcal{H}} \mathrm{tr}(\mathbf{K}_j(\nu))\mathrm{d}F(\nu) \le P_j \qquad \forall j \in [1:K] \tag{4.19}$$

for some vector $P = [P_1, \ldots, P_K] \in \mathbb{R}_+^K$ of average-power constraints for all the users.

More precisely, to transmit its known messages, transmitter j may send over n consecutive uses of the channel (which we index by the channel use t, rather than the channel state ν in (4.18)). Complete channel state information refers to a non-causal knowledge of the channel state sequence $\mathbf{H}^n = \{\mathbf{H}_j(1), \ldots, \mathbf{H}_j(n) : j \in [1:K]\}$ so that to send its message, transmitter j selects which symbol to send via an encoding mapping $\mathbf{X}_j(t) \leftarrow (M_S : j \in S \in E), \mathbf{H}^n$, where $A \leftarrow B$ denotes that A is a function of B, for each channel use. To decode these messages, the receiver maps the output sequence $\{\mathbf{Y}(1), \ldots, \mathbf{Y}(n)\} = \mathbf{Y}^n$ and channel state sequence to an estimate $\hat{M}_S$ of each message M_S for $S \in E$. Our metric of reliable communication is that of average probability: with each M_S independently and uniformly distributed over $\mathcal{M}_S = \left[1 : 2^{nR_S}\right]$, we declare a rate tuple $(R_S : S \in E)$ to be achievable if there exists a set of encoder sequences which satisfy the power constraint (4.19) on average[4] and a sequence of decoding functions with the probability of error

[4] That is, $E\left[\frac{1}{n}\sum_{t=1}^{n} \|\mathbf{X}_j(t)\|^2\right] \le P_j$ for each transmitter $j \in [1:K]$, where the expectation is over all channel state sequences $\mathbf{H}^n$.

$P(\cup_{S\in E}\{M_S \neq \hat{M}_S\}) \to 0$ vanishing as the block length n tends to infinity. The capacity region $\mathcal{C}(P)$ is the closure of the set of all achievable rate tuples.

4.5.2 Capacity Region Under Dynamic Resource Allocation

Suppose that the transmitters and receiver have complete knowledge of the current state of the channels of every user. With this knowledge, the codewords and decoding scheme can depend on the current state of the channels. In particular, for the Gaussian case, the Gaussian inputs can depend on the channel state $\nu \in \mathcal{H}$ so that the received vector may be written in the form

$$\mathbf{Y}(\nu) = \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{U}_S(\nu) + \mathbf{Z}(\nu), \tag{4.20}$$

where for each state $\nu \in \mathcal{H}$ and message index $S \in E$, $\mathbf{U}_S(\nu) \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_S(\nu))$. We call the set of covariance choices $\{\mathbf{K}_S(\mu)\}$, for each fading state μ and message index $S \in E$, a **covariance allocation**. For any given covariance allocation consider the set of rates given by

$$C_f(\{\mathbf{K}_S(\mu)\}) = \Bigg\{ R \in \mathbb{R}_+^E : \sum_{S\in B} R_S \leq \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{K}_S(\nu)\mathbf{H}_S^*(\nu)\right) \mathrm{d}F(\nu)\ \forall B \in \mathcal{F}_\downarrow \Bigg\}. \tag{4.21}$$

Comparing this with the capacity region given in Theorem 4.3, one can heuristically think of $C_f(\{\mathbf{K}_S(\mu)\})$ as the set of achievable rates when the power and linear signaling strategy are dynamically allocated according to the covariance allocation $\{\mathbf{K}_S(\mu)\}$. The following theorem substantiates such an interpretation.

Theorem 4.5.1. *The capacity of the ergodic fading MIMO MAC with implicit cooperation and with complete channel state information at the transmitters and receiver is given by*

$$C(P) = \bigcup_{\{\mathbf{K}_S(\mu)\}\in\mathcal{F}} C_f(\{\mathbf{K}_S(\mu)\})$$

where $\mathcal{F}$ is the set of all feasible covariance allocations satisfying the power constraint

$$\sum_{j\in S\in E} \int_{\mathcal{H}} \mathrm{tr}(\mathbf{K}_{S,jj}(\nu))\mathrm{d}F(\nu) \leq P_j \qquad j \in [1:K]\,. \tag{4.22}$$

Proof. Achievability follows from an argument mimicking that in [106]. Notably, capacity may be achieved either with a "multiple-codebook variable-rate"' [40] or with a "single-codebook constant-rate" [12] scheme. The converse is provided in Appendix C.2. □

The above theorem states that any improvement in capacity due to channel state information at the transmitter is due solely to the ability to dynamically allocate the power and linear signaling strategies according to the channel state. In the case where all transmitters and receivers have only a single antenna ($t_1 = \cdots = t_K = r = 1$), the above characterization simplifies further, as through Cauchy-Schwartz, it suffices to only consider rank-one covariance allocations as in (4.10), which are

$$\mathbf{K}_S = \mathbf{B}_S \mathbf{B}_S^* \qquad\qquad \mathbf{B}_S = \begin{bmatrix} \rho_{S,j_1}(\nu)\sqrt{p_S(\nu)}/h_{j_1}(\nu) \\ \vdots \\ \rho_{S,j_{|S|}}(\nu)\sqrt{p_S(\nu)}/h_{j_{|S|}}(\nu) \end{bmatrix},$$

where each vector $\rho_S(\nu) = \begin{bmatrix} \rho_{S,j_1}(\nu) & \cdots & \rho_{S,j_{|S|}}(\nu) \end{bmatrix}$, which we call a load balance vector, is on the $|S|$-dimensional simplex for each message index $S \in E$ and each fading state $\nu \in \mathcal{H}$. In this case, the covariance allocation is simply parameterized by the its received power $p_S(\nu) = \mathbf{H}_S(\nu)\mathbf{K}_S(\nu)\mathbf{H}_S^*(\nu)$ and its load-balance vector $\rho_S(\nu)$. Moreover the polytope (4.21) may be simply described as

$$C_f(\{\rho_S(\nu), p_S(\nu)\}) = \left\{ R \in \mathbb{R}_+^E : \sum_{S \in B} R_S \le \int_{\mathcal{H}} \log\left(1 + \sum_{S \in B} p_S(\nu)\right) \mathrm{d}F(\nu) \ \forall B \in \mathcal{F}_{\downarrow} \right\}. \quad (4.23)$$

This leads to the following corollary

Corollary 4.5.2. *The capacity of the ergodic fading SISO MAC with implicit cooperation and with complete channel state information at the transmitters and receiver is given by*

$$C(P) = \bigcup_{\{\rho_S(\nu), p_S(\nu)\} \in \mathcal{F}} C_f(\{\rho_S(\nu), p_S\})$$

where $\mathcal{F}$ is the set of all feasible received powers and load-balance vectors allocations satisfying the power constraint

$$\sum_{j \in S \in E} \int_{\mathcal{H}} \frac{\rho_{S,jj}^2(\nu) p_S(\nu)}{|h_j(\nu)|^2} \mathrm{d}F(\nu) \le P_j \qquad j \in [1:K] \quad (4.24)$$

Proof. A consequence of Cauchy-Schwartz, this may be shown by applying the analogous argument, mutatis mutandis, from the non-fading case. □

4.6 Fading Channel: Optimal Power Allocations

The boundary points of the capacity region for the fading Gaussian MAC are the set of rates such that no component can be increased with the other components remaining fixed, while remaining within the capacity region. In other language, these are known as Pareto-optimal points, and correspond to optimal operating points: any other operating point in the capacity region is dominated component-wise by some boundary point.

For convex regions, every boundary point lies on a supporting hyperplane. In particular, pursuant to the discussion following Theorem 4.5.1, the capacity region $\mathcal{C}(P)$ of the fading MIMO MAC is convex. As such, we may characterize the boundary surface $\mathcal{B}(P)$ of the capacity region $\mathcal{C}(P)$ as the closure of all points R^* such that R^* is a solution to an optimization problem

$$\max_{R \in \mathcal{C}(P)} \ \mu \cdot R \tag{4.25}$$

for some rate-profile vector $\mu \in \mathbb{R}_+^E$. If, for each such rate-profile vector, we denote the set of rate tuples R^* which attain the maximum (4.25) by $\mathcal{R}(\mu)$, then the map $\mathcal{R} : \mathbb{R}_+^E \mapsto \mathcal{B}(P)$ from the space of rate-profile vectors to the boundary surface can be understood as a parameterization of the boundary surface of the capacity boundary surface. Therefore, to obtain a complete description of the boundary surface of the capacity region, it suffices to consider any subset $\mathcal{S} \subset \mathbb{R}_+^E$ such that $\mathcal{R}(\mathcal{S}) = \mathcal{B}(P)$. In fact, this is precisely the case in the multiple access channel with cooperation: rather than needing to consider all rate-profile choices $\mu \in \mathbb{R}_+^E$, it suffices to only consider those in the subset

$$\mathcal{O} = \{\mu \in \mathbb{R}_+^E : \mu_S > \mu_{S'} \text{ only if } S \subset S'\},$$

as formalized in Lemma 4.4.1.

For such a rate-profile vector in $\mathcal{O}$, any enumeration $E = \{S_1, \ldots, S_M\}$ which orders the elements of μ in decreasing order $\mu_{S_1} \geq \cdots \geq \mu_{S_M}$ is necessarily a non-decreasing enumeration of E with respect to the inclusion order: $S_i \subset S_j$ only if $i < j$. So, for each $i \in [1 : M]$, the subset $\{S_1, \ldots, S_i\}$ is a down-set in the down-set lattice $\mathcal{F}_\downarrow$ and is thus a fixed point of the operator $\mathcal{Z}_{\mathcal{F}_\downarrow}$ defined in (2.16): $\mathcal{Z}_{\mathcal{F}_\downarrow}(\{S_1, \ldots, S_i\}) = \{S_1, \ldots, S_i\}$. Thus, a polymatroid corner point

corresponding to this rate-profile vector is simply given by $x_{S_1} = f(\{e_1\})$,

$$x_{e_i} = f(\{S_1, \ldots, S_i\}) - f(\{S_1, \ldots, S_{i-1}\})$$

for $2 \leq i \leq n$.

With this observation, and the knowledge of polymatroid structure, the fading capacity region is expressible as the weighted sum of the capacity regions of parallel time-invariant Gaussian channels per channel state $\nu \in \mathcal{H}$:

Lemma 4.6.1.

$$C_f(\{\mathbf{K}_S\}) = \left\{ \left(R_S = \int_{\mathcal{H}} r_S(\nu) \mathrm{d}F(\nu) : S \in E \right) : r(\nu) \in C_g(\{\mathbf{H}_k(\nu)\}, \{\mathbf{K}_S(\nu)\}) \quad \forall \nu \in \mathcal{H} \right\}. \tag{4.26}$$

Proof. Define $\mathcal{E}$ to bet the right hand side of (4.26). By definition, $\mathcal{E} \subseteq C_f(\{\mathbf{K}_S\})$. But as $C_f(\{\mathbf{K}_S\})$ is a polymatroid over the down-set lattice $\mathcal{F}_\downarrow$, it is the convex hull of its vertices. For some enumeration on E, these vertices are (non-uniquely) specified by all permutations $\pi : [1:M] \mapsto [1:M]$ as

$$\begin{aligned}
R_{S_{\pi(1)}} &= \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \mathbf{H}_{S_{\pi(1)}}(\nu)\mathbf{K}_{S_{\pi(1)}}(\nu)\mathbf{H}^*_{S_{\pi(1)}}(\nu)\right) \mathrm{d}F(\nu) \\
R_{S_{\pi(i)}} &= \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \sum_{S \in B_i} \mathbf{H}_S(\nu)\mathbf{K}_S(\nu)\mathbf{H}^*_S(\nu)\right) \mathrm{d}F(\nu) \\
&\qquad - \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \sum_{S \in B_{i-1}} \mathbf{H}_S(\nu)\mathbf{K}_S(\nu)\mathbf{H}^*_S(\nu)\right) \mathrm{d}F(\nu) \qquad 2 \leq i \leq M
\end{aligned}$$

where $B_i = \mathcal{Z}_{\mathcal{F}_\downarrow}(\{S_{\pi(1)}, \ldots, S_{\pi(i)}\})$. But each such vertex is the the convex combination of the corresponding vertices in the collection of polymatroids $C_g(\{\mathbf{H}_k(\nu)\}, \{\mathbf{K}_S(\nu)\})$, indexed by the states $\nu \in \mathcal{H}$. Hence, $C_f(\{\mathbf{K}_S\}) \subseteq \mathcal{E}$ as well. □

4.6.1 Langrangian Characterization of the Capacity Region

Let's focus on the optimization problem (4.25) for each $\mu \in \mathcal{O}$, which by Lemma 4.6.1 is simply:

Problem 4.6.1

$$\begin{aligned}
&\text{Maximize} && \sum_{S\in E}\mu_S\int_{\mathcal{H}} r_S(\nu)\mathrm{d}F(\nu)\\
&\text{Subject to} && \sum_{j\in S\in E}\int_{\mathcal{H}}\mathrm{tr}(\mathbf{K}_{S,jj}(\nu))\mathrm{d}F(\nu)\leq P_j\ \forall S\in E && \text{(4.27a)}\\
& && \mathbf{K}_S(\nu)\succeq\mathbf{0}\quad\forall S\in E,\forall\nu\in\mathcal{H} && \text{(4.27b)}\\
& && r(\nu)\in\mathcal{C}_g(\{\mathbf{H}_k(\nu)\},\{\mathbf{K}_S(\nu)\}).\ \forall\nu\in\mathcal{H} && \text{(4.27c)}
\end{aligned}$$

The optimization problem is formulated directly in terms of the rate allocations $r(\nu) = (r_S(\nu) : S \in E)$ and covariance allocations. These constraints are convex: this can be seen by noting the inequalities (4.27a)-(4.27b) are linear and that the concavity of $\log\det(\cdot)$ implies that the polymatroid constraint (4.27c) is convex. As the objective is also linear, Problem 4.6.1 is convex and we may employ standard convex optimization techniques to help solve it.

Let $\mathcal{D}$ denote the set of covariance allocations and rate allocations satsifying (4.27b)-(4.27c), which by the prior observations, is convex. Consider problem 4.6.1 as an optimization problem over the domain $\mathcal{D}$ with an explicit constraint of (4.27a). Then its corresponding Lagrangian, with a vector of dual variables $\lambda = [\lambda_1,\ldots,\lambda_K]\in\mathbb{R}_+^K$ associated with the power constraints (4.27a), is defined over domain $\mathcal{D}$ as

$$\begin{aligned}
\mathcal{L}(\{\mathbf{K}_S(\nu)\},\{r_S(\nu)\},\lambda) = &\sum_{S\in E}\mu_S\left(\int_{\mathcal{H}} r_S(\nu)\mathrm{d}F(\nu)\right)\\
&-\sum_{j=1}^{K}\lambda_j\left(\sum_{j\in S\in E}\int_{\mathcal{H}}\mathrm{tr}(\mathbf{K}_{S,jj}(\nu))\mathrm{d}F(\nu)-P_j\right).
\end{aligned}$$

Of interest is the Lagrange dual function, defined to be

$$g(\lambda) = \max_{(\{\mathbf{K}_S(\nu)\},\{r_S(\nu)\})\in\mathcal{D}}\mathcal{L}(\{\mathbf{K}_S(\nu)\},\{r_S(\nu)\},\lambda),$$

as it provides an upper bound to the optimal value, p^*, of Problem 4.6.1: $\min_{\lambda\in\mathbb{R}_+^K} g(\lambda)\geq p^*$. In fact, Problem 4.6.1 is sufficiently nice to guarantee that this upper bound is tight: as the feasible set has non-empty interior[5] , Slater's condition holds and so the duality gap, $p^* - \min_\lambda g(\lambda)$, is zero.

[5] With sufficiently large powers, the polytopes $\mathcal{C}_g(\{\mathbf{K}_S(\nu)\})$ can be made to contain any rate-tuple.

This suggests that a strategy to obtain an optimal solution to Problem 4.6.1 is to first obtain the Lagrange dual function (thereby obtaining candidate optimal covariance and rate allocations, $\{\mathbf{K}^*(\nu)\}$ and $r^*(\nu)$, per choice of dual variable λ) and then to minimize the Lagrange dual function to choose a "best" choice from these candidate covariance and rate allocations. Let λ^* be any such minimizer. Since the problem is convex and the duality gap is zero, the Karush-Kuhn-Tucker (KKT) optimality conditions state that this best choice will solve Problem 3.1; that is, any primal optimal solution set minimizes $\mathcal{L}(\{\mathbf{K}_S^*(\nu)\}, \{r_S^*(\nu)\}, \lambda^*)$ and satisfies the power constraint (4.27a) simultaneously.

To pursue this strategy efficiently, we need an efficient method for both of the optimization steps: that of maximizing the Lagrangian and that of minimizing the Lagrange dual function. The former step may be significantly simplified through the observation that it decomposes into a set of independent optimizations problems per state $\nu \in \mathcal{H}$. To see this, re-arrange the Lagrangian as

$$\mathcal{L}(\{\mathbf{K}_S(\nu)\}, r(\nu), \lambda) = \int_{\mathcal{H}} \left(\sum_{S \in E} \mu_S r_S(\nu) - \sum_{j=1}^{K} \lambda_j \sum_{j \in S \in E} \mathrm{tr}(\mathbf{K}_{S,jj}(\nu)) \right) \mathrm{d}F(\nu) + \lambda \cdot P,$$

and define, for each state $\nu \in \mathcal{H}$, $g'_\nu(\lambda)$ to be the optimal value of the following optimization problem.

Problem 4.6.1.ν

$$\begin{aligned} \text{Maximize} \quad & \sum_{S \in E} \mu_S r_S(\nu) - \sum_{j=1}^{K} \lambda_j \sum_{j \in S \in E} \mathrm{tr}(\mathbf{K}_{S,jj}(\nu)) \\ \text{Subject to} \quad & \mathbf{K}_S(\nu) \succeq \mathbf{0} \quad \forall S \in E, && \text{(4.28a)} \\ & r(\nu) \in \mathcal{C}_g(\{\mathbf{H}_k(\nu)\}, \{\mathbf{K}_S(\nu)\}). && \text{(4.28b)} \end{aligned}$$

Then the Lagrange dual function may be decomposed as

$$g(\lambda) = \int_{\mathcal{H}} g'_\nu(\lambda) \mathrm{d}F(\nu) + \lambda \cdot P.$$

This approach, of decomposing the Lagrange dual function into a series of much smaller and simpler optimization problems, is known as the dual-decomposition method.

For each state $\nu \in \mathcal{H}$, Problem 4.6.1.ν may be simplified further through consideration of the special polymatroidal structure of the constraint (4.28b). In particular, with the explicit formulae for its vertices, Problem 4.6.1.ν is equivalently stated in the following simpler form. Recall that we have restricted attention to a $\mu \in \mathcal{O}$ and so an enumeration $E = \{S_1, \ldots, S_M\}$ which orders the elements of the rate-profile vector in descending order (i.e. $\mu_{S_1} \geq \cdots \geq \mu_{S_M}$) has that each subset $\{S_1, \ldots, S_i\}$ is a down-set of E under the inclusion order $\subseteq$. Thus, an the equivalent formulation of Problem 4.6.1.ν is

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{M} \delta_i \log \left| \mathbf{I} + \sum_{j=1}^{i} \mathbf{H}_{S_j}(\nu) \mathbf{K}_{S_j}(\nu) \mathbf{H}_{S_j}^*(\nu) \right| - \sum_{j=1}^{K} \lambda_j \sum_{j \in S \in E} \operatorname{tr}(\mathbf{K}_{S,jj}(\nu)) \\ \text{Subject to} \quad & \mathbf{K}_S(\nu) \succeq \mathbf{0} \quad \forall S \in E, \end{aligned}$$

where $\delta_1 = \mu_{S_1}$ and $\delta_i = \mu_{S_i} - \mu_{S_{i-1}}$ for $2 \leq i \leq M$. This simplified optimization problem is again concave.

We distinguish between two cases: those for which Problem 4.6.1.ν is solvable exactly and thereby permitting consideration of infinite fading state spaces $\mathcal{H}$ and those for which we require numerical methods to solve Problem 4.6.1.ν. Below, we show that the SISO case falls under the former category, while for the case where there are multiple antennas at at least one terminal, we resort to numerical methods.

4.6.1.1 MIMO case: Numerical Solution

As Problem 4.6.1.ν is convex with a twice differentiable objective we may solve it efficiently via the Interior-Point method for each $\nu \in \mathcal{H}$.

4.6.1.2 SISO case: Analytical Solution

In the SISO case, we use Corollary 4.5.2 to re-state Problem 4.6.1.ν in the following form.

Problem 4.6.1.ν.SISO

$$\text{Maximize} \quad \sum_{i=1}^{M} \delta_i \log \left(1 + \sum_{j=1}^{i} p_{S_j}(\nu) \right) - \sum_{j=1}^{K} \lambda_j \sum_{j \in S \in E} \frac{\rho_{S,jj}^2(\nu) p_S(\nu)}{|h_j(\nu)|^2}$$

$$\text{Subject to} \qquad p_S(\nu) \geq 0 \quad \forall S \in E,$$

$$\rho_{S,j}(\nu) \geq 0 \quad \forall (S,j) \text{ with } j \in S \in E$$

$$\sum_{j \in S} \rho_{S,j} = 1 \quad \forall S \in E$$

To solve this, we may use the result of Tse, but before we can do this, we need make a simplification. Observe that the in the problem above the load-balance allocations $\rho_S(\nu)$ are present only in the negative term in the objective. Hence, to maximize the objective, we are free to minimize this negative term with respect to the load-balance vector while keeping the received powers $p_S(\nu)$ fixed.

This minimum has an explicit solution. Rearranging the summation, the negative term is lower bounded as

$$\begin{aligned} \sum_{j=1}^{K} \frac{\lambda_j}{|h_j(\nu)|^2} \sum_{j \in S \in E} \rho_{S,j}^2(\nu) p_S(\nu) &= \sum_{S \in E} p_S(\nu) \sum_{j \in S} \rho_{S,j}^2(\nu) \frac{\lambda_i}{|h_j(\nu)|^2} \\ &\geq \sum_{S \in E} p_S(\nu) \left(\sum_{j \in S} \frac{|h_j(\nu)|^2}{\lambda_j} \right)^{-1}, \end{aligned} \tag{4.29}$$

as a consequence of Cauchy-Schwartz:

$$\left(\sum_{j \in S} \rho_{S,j}(\nu) \right)^2 \leq \left(\sum_{j \in S} \frac{|h_j(\nu)|^2}{\lambda_j} \right) \left(\sum_{j \in S} \rho_{S,j}^2(\nu) \frac{\lambda_j}{|h_j(\nu)|^2} \right).$$

With the minimizing choice $\rho_{S,j}^*(\nu) = \frac{|h_j(\nu)|^2}{\lambda_j} \eta_S(\nu)$, this lower bound is attained, where for each $S \in E$, we define

$$\eta_S(\nu) = \left(\sum_{j \in S} \frac{|h_j(\nu)|^2}{\lambda_j} \right)^{-1}. \tag{4.30}$$

So, Problem 4.6.1.ν.SISO is equivalently stated as

$$\max_{p_S(\nu) \geq 0 \ \forall S \in E} \sum_{i=1}^{M} \delta_i \log \left(1 + \sum_{j=1}^{i} p_{S_j}(\nu) \right) - \eta_{S_i}(\nu) p_S(\nu).$$

A solution to this optimization can be found through the greedy method developed in Tse and

Hanly [106] for the non-cooperative MAC. Define the marginal utility functions [6]

$$u_S(z) = \frac{\mu_S}{1+z} - \eta_S(\nu) \quad \forall S \in E$$
$$u^*(z) = \left[\max_{S\in E} u_S(z)\right]^+,$$

which provide an upper bound on the achievable rate in Problem 4.6.1.ν.SISO:

$$\sum_{i=1}^{M} \mu_i \left[\log\left(1 + \sum_{j=1}^{i} p_{S_j}(\nu)\right) - \log\left(1 + \sum_{j=1}^{i-1} p_{S_j}(\nu)\right)\right] - \eta_{S_i}(\nu) p_{S_i}(\nu)$$
$$= \sum_{i=1}^{M} \int_{\sum_{j=1}^{i-1} p_{S_j}(\nu)}^{\sum_{j=1}^{i} p_{S_j}(\nu)} u_{S_i}(z)\mathrm{d}z \le \int_0^\infty u^*(z)\mathrm{d}z.$$

Critically, this upper bound can actually be attained. As each utility function $u_S(z)$ is monotonically decreasing in z, they intersect each other in at most one location. Thus, there is a sequence $0 = z_0 < z_1 < \cdots < z_J$ of J terms where z_J is the smallest $z \ge 0$ for which $u^*(z) = 0$ (if there is no z, set $z_J = \infty$) and $u^*(z) = u_{S'_k}(z)$ for $z \in [z_k, z_{k+1}]$ with $E = \{S'_1, \ldots, S'_M\}$ as a (possibly different) enumeration than $E = \{S_1, \ldots, S_M\}$. Define

$$p^*_{S'_k}(\nu) = z_{k+1} - z_k \qquad k = 0, \ldots, J-1$$
$$p^*_{S'_k}(\nu) = 0 \qquad k \ge J$$
$$r^*_{S'_k}(\nu) = \log(1 + z_{k+1}) - \log(1 + z_k) \qquad k = 0, \ldots, J-1$$
$$r^*_{S'_k}(\nu) = 0 \qquad k \ge J$$

Then this rate point attains the maximum,

$$\int_0^\infty u^*(z)\mathrm{d}z = \sum_{k=1}^{M} \mu_{S'_k} r^*_{S'_k}(\nu) - \eta_{S'_k}(\nu) \cdot p^*_{S'_k}(\nu),$$

and we claim that this rate point is lies on the boundary of the the capacity region. To see this note that the rate allocation corresponds to a successive decoding scheme whereby the message indexed by S'_i is decoded after decoding the messages $S'_{i+1}, \ldots, S'_M$. For this rate tuple to lie on the boundary of the capacity region, the enumeration $\{S'_i\}$ on E must satisfy $S'_i \subset S'_j$ only if $i > j$ so

[6] Here, $x^+ = \max(x, 0)$.

that "more common" messages are decoded prior to "more private" messages. To see that this is true, consider a pair $i \neq j$ with $S_i' \subset S_j'$. As $\mu_{S_i'} \geq \mu_{S_j'}$ (by assumption $\mu \in \mathcal{O}$) and $\eta_{S_i'}(\mu) \geq \eta_{S_j'}(\mu)$ (by definition), there must be intersection point $z^* > 0$ such that

$$u_S(z) \geq u_{S'}(z) \qquad z \leq z^*$$
$$u_S(z) \leq u_{S'}(z) \qquad z \geq z^*.$$

Hence, necessarily $i > j$. Thus, $\{r_S^*(\nu) : S \in E\}$ and $\{p_S^*(\nu) : S \in E\}$ are the **optimal** received power and rate allocations for the state $\nu \in \mathcal{H}$.

4.6.2 Optimizing the Lagrange Dual Function

To minimize the dual Langrange function, $g(\lambda)$, we require a method that needn't rely on differentiability. As the the dual Lagrange function is convex, an appropriate method is the ellipsoid method, we converges in polynomial time and only requires knowledge of some subgradient of the objective function, which is provided with the following:

Lemma 4.6.2. *If $\{\mathbf{K}_S(\nu)\}$ and $\{r_S^*(\nu)\}$ maximize the Lagrangian over $\mathcal{D}$ at λ, i.e. $\mathcal{L}(\{\mathbf{K}_S^*(\nu)\}, r^*(\nu), \lambda) = g(\lambda)$, then the vector $\mathbf{u}$ defined as*

$$u_j = P_j - \sum_{j \in S \in E} \int_{\mathcal{H}} \mathbf{K}_{S,jj}^* \mathrm{d}F(\nu) \quad j \in [1:K]\,,$$

is a subgradient of g at λ.

Proof. For any vector $\delta \in \mathbb{R}_+^K$, we have

$$g(\delta) \geq \mathcal{L}(\{\mathbf{K}_S^*(\nu)\}, r^*(\nu), \delta) = g(\lambda) + \sum_{k=1}^{K} (\delta_k - \lambda_k) \left(P_j - \sum_{j \in S \in E} \int_{\mathcal{H}} \mathbf{K}_{S,jj}^* \mathrm{d}F(\nu) \right)$$

so that $g(\delta) - g(\lambda) \geq u \cdot (\delta - \lambda)$. □

This is similar in spirit to the corresponding prescriptions given for the non-cooperative case in [71] and generalizes an analogous result in [65].

4.7 Constant Gap Characterization

4.7.1 Constate State

To clarify what gains sending a common message might provide over sending private messages, we develop a constant gap-to capacity characterization, which in turn leads to a DoF characterization of capacity with general message sets:

Theorem 4.7.1. *For the K user MIMO fading Gaussian MAC with common information, the set of rate tuples $(R_S : S \in E)$ for which*

$$\sum\nolimits_{D\in\downarrow S} R_D \leq I(S) \triangleq \log\det\left(\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j/t_j\right) \tag{4.31}$$

for all non-empty $S \subseteq [K]$ is achievable by a simple single coding strategy that does not require cooperation or channel state information. Moreover, if $(R_S : S \in E)$ is an achievable rate tuple, then it must satisfy

$$\sum\nolimits_{D\in\downarrow S} R_D \leq O(S) \triangleq \log\det\left(\mathbf{I} + |S| \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j\right) \tag{4.32}$$

for all non-empty $S \subseteq [K]$.

Proof. Our inner bound can be arrived at by the general MIMO MAC results with the specific choice of input distribution: all common-message auxiliary random variables are set to zero ($U_S = 0$ for all $S \in E$ with $|S| \geq 2$) while the K transmitters are independent with $\mathbf{X}_j \sim \mathcal{CN}(\mathbf{0}, (P_j/t_j)\mathbf{I})$. The outer bound will follow from from the converse for the fading case in Section 4.7.4 □

Noteworthy is that this inner bound is achieved with non-cooperative coding and with no channel state information at the transmitters (no CSIT). Thus the corollaries in the following section bound the benefit that cooperative coding and CSIT may have.

In the case of the much studied MIMO MAC with private messages only [121, 71] the bounds of Theorem 4.7.1 (specialized by setting $R_S = 0$ for $|S| > 1$) imply that the optimization of transmit covariances [121],[71] to maximize sum-rate results in an expansion of sum rate by no greater than

the constant gap of Corollary 4.7.2 (when $S = [K]$) while requiring global CSIT at all transmitters. In the private message case, we may tighten the outer bound (4.32) to

$$\sum\nolimits_{S\in\uparrow\{j\}} R_S \leq \log\det\left(\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j\right)$$

by not having to consider correlated inputs; this bound follows by mimicking the steps in 4.7.4 while omitting the unnecessary block diagonalization bound step provided by Lemma 4.7.6. Thus, when we only have private messages to send, the gap of Corollary 4.7.2 may be tightened to $\min\{r, t(S)\}\log\left(\hat{t}_S\right)$.

4.7.1.1 Quantifiable gap

A salient feature of this result is that the gap between the inner and outer bounds admits a convenient characterization:

- The gap can be bounded by a constant independent of both the channel state $\mathbf{H}$ and the signal to noise ratios $(P_1, \ldots, P_K)$.
- If the entries of the channel state $\mathbf{H}$ are assumed to have been drawn iid from a distribution with bounded variance, then the gap may be thought of a random variable whose distribution and expectation can be characterized with random matrix theory. In particular, by such results, the gap is with high probability very close to its expected value which in turn has a deterministic characterization.

Corollary 4.7.2 (Constant-Gap)**.** *For each $S \subseteq [K]$,*

$$O(S) - I(S) \leq \min\{r, t(S)\}\log\left(|S|\hat{t}_S\right) \tag{4.33a}$$

$$\leq \min\{r, t([K])\}\log(K\hat{t}_{[K]}), \tag{4.33b}$$

where $t(S) = \sum_{j\in S} t_j$ and $\hat{t}_S = \max_{j\in S} t_j$

Proof. We can relax the above inner and outer bounds as

$$I(S) = \log\det\left(\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j/t_j\right) \geq \log\det\left(\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j\right) - \min\left\{r, t(S)\right\}\log\left(\hat{t}_S\right)$$

$$O(S) = \log\det\left(\mathbf{I} + |S|\sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j\right) \leq \log\det\left(|\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j\mathbf{H}_j^* P_j\right) + \min\left\{r, t(S)\right\}\log(|S|).$$

to provide (4.33a). Relaxing (4.33a) to (4.33b) provides a uniform bound on all such gaps. □

A special case of interest is in space-division multiple-access (SDMA), where $t_1 = \cdots = t_K = 1$ and the uniform gap above is $\min\{r, K\}\log(K)$.

4.7.2 Time-varying state: Fading Channel

The above results can be extended, via a separable coding scheme, to the fading K-user MIMO MAC-CM, provided that both the transmitter and receiver have non-causal knowledge of the channel fading state. As each transmitter need only satisfy the power constraint on average, it is free to choose a different transmit power per channel fading state, described by the power allocations $\rho(\nu) : \mathcal{H} \to \mathbb{R}_+^K$, whose jth entry $\rho_j(\nu)$ is the jth transmitter's power allocation. A power allocation is admissible if $\int_{\mathcal{H}} \rho_j(\nu)\mathrm{d}F(\nu) \leq P_j$ for each $j \in [K]$. An approximate capacity characterization for the fading case is:

Theorem 4.7.3. *For the K user fading MIMO fading Gaussian MAC with common information, the set of rate tuples $(R_S : S \in E)$ for which*

$$\sum_{D\in\downarrow S} R_D \leq I_{\rho(\cdot)}(S) \triangleq \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \sum\nolimits_{j\in S} \mathbf{H}_j(\nu)\mathbf{H}_j^*(\nu)\rho_j(\nu)/t_j\right)\mathrm{d}F(\nu) \tag{4.34}$$

for all non-empty subsets $S \subseteq [K]$ and some admissible power allocation $\rho(\cdot)$ is achievable by a simple coding strategy with a single strategy per power allocation. Moreover, if $(R_S : S \in E)$ is an achievable rate tuple, then it must satisfy

$$\sum_{D\in\downarrow S} R_D \leq O_{\rho(\cdot)}(S) \triangleq \int_{\mathcal{H}} \log\det\left(\mathbf{I} + |S|\sum\nolimits_{j\in S} \mathbf{H}_j(\nu)\mathbf{H}_j^*(\nu)\rho_j(\nu)\right)\mathrm{d}F(\nu) \tag{4.35}$$

*for all non-empty subsets $S \subseteq [K]$ and **some** admissible power allocation $\rho(\cdot)$.*

Proof. The achievable rate region for a constant channel state in Theorem 4.7.1 may be extended, via a separable coding scheme, to the achievable rate region given above for a time-varying channel fade process. As the proof involves standard arguments (e.g. [40], [31]) that are not specific to the problem here, we only sketch the proof. The proof involves the multiplexing over a discretization of the channel fading state space. With a $\hat{\mathbf{h}}$ as a finite discretization of the state space, we rate split as

$$R_S = \sum_{\hat{\mathbf{h}}} p(\hat{\mathbf{h}}) R_{S,\hat{\mathbf{h}}}. \tag{4.36}$$

At the encoder, multiplex the larger block of n contiguous channel uses into a collection of smaller sub-blocks over which the discretized channel fading state is constant. Over each sub-block, encode and decode as in the constant channel state case with the signal-to-noise ratios determined by the power allocation. The receiver, who also demultiplexes the channel transmission block into the constant-state channel transmission subblocks, is able to decode each message split $(m_{S,\hat{\mathbf{h}}} : S \in E)$ per state $\hat{\mathbf{h}}$ provided that that for each $\hat{\mathbf{h}}$,

$$\sum_{D \in \downarrow S} R_{D,\hat{\mathbf{h}}} \leq \log\det\left(\mathbf{I} + \sum\nolimits_{j \in S} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^* \rho_j(\hat{\mathbf{h}})/t_j\right)$$

for all $S \subseteq [K]$. Hence, per the rate split (4.36), we may achieve any rate tuple which satisfies

$$\sum_{D \in \downarrow S} R_D \leq E_{\hat{\mathbf{H}}} \log\det\left(\mathbf{I} + \sum\nolimits_{j \in S} \hat{\mathbf{H}}_j \hat{\mathbf{H}}_j^* \rho_j(\hat{\mathbf{H}})/t_j\right)$$

for all $S \subseteq [K[$. Taking a series of increasingly finer quantizations of the state space provides that (4.34) is achievable. The outer bound is detailed in Section 4.7.4. □

Mimicking the proof of Corollary 4.7.2 provides an analogous corollary to the above theorem. In particular, the gap between the inner and outer bounds is easily bounded with a gap independent of the power allocation policy.

Corollary 4.7.4. *For each $S \subseteq [K]$ and admissible power allocation $\rho(\cdot)$,*

$$O_{\rho(\cdot)}(S) - I_{\rho(\cdot)}(S) \leq \min\{r, t(S)\} \log\left(|S|\hat{t}_S\right).$$

As this gap holds for all power allocations, it holds for those that are optimal in some sense. For example, it would hold for the choice of $\rho(\cdot)$ that maximizes the weighted sum-rate objective $\sum_{S\in E}\mu_S R_S$ subject to the constraints (4.35) for some set of non-negative rate-rewards $(\mu_S : S \in E)$. For this power allocation and rate-reward vector, the gap in Corollary 4.7.4 would represent the gap to capacity.

4.7.3 DoF Region

A simple consequence of the above result is that the Degrees-of-Freedom (DoF) region for MIMO MAC with general message sets is easily characterized. The degrees of freedom (DoF) region is, informally, the capacity region in the high signal to noise ratio regime, characterizing the number of complex symbols which may be allocated towards the transmission of each message. Formally, it is the set of non-negative tuples $(d_S : S \in E)$ satisfying

$$\sum_{S\in E}\mu_S d_S \leq \limsup_{P\to\infty}\left(\sup_{R(P)\in\mathcal{C}_K(P,\ldots,P)}\left[\sum_{S\in E}\mu_S R_S(P)\right]\frac{1}{\log(P)}\right)$$

for all non-negative weight vectors $(\mu_S \geq 0 : S \in E)$ [51]. By considering successively higher power budgets, the signal-to-noise ratio can be made arbitrarily high so that the impact of the additive noise is eliminated and the competition among the various signals to be transmitted is brought to the forefront.

By considering setting $P_1 = \cdots = P_K = P$ in the constant gap capacity characterizations of Theorem 4.7.1 and allowing the power budget P to tend to infinity, we find:

Corollary 4.7.5. *The DoF region for the MIMO MAC with general message sets is given by*

$$\sum_{S\in\cup_{j\in A}\uparrow\{j\}} d_S \leq \operatorname{rank}(\mathbf{H}_A) \tag{4.37}$$

where $\mathbf{H}_A$ *is the horizontal concatenation of the matrices* $(\mathbf{H}_j : j \in A)$.

Of particular importance is the observation signaling with a common message provides no benefit over the private message signaling, whose DoF region is given by

$$\sum_{j\in A} d_{\{j\}} \leq \operatorname{rank}(\mathbf{H}_A). \tag{4.38}$$

Further evidence of this is that one can layer down-set rate-splitting over a private-message only DoF optimal signaling scheme to achieve the DoF region of Corollary 4.7.5. Doing so introduces rate-splits: projecting away those rate-splits can be done efficiently via the procedure introduced in 3.4.1.

4.7.4 Outer Bound

Suppose that for the fading K-user MIMO MAC with common messages there exists a sequence of codes indexed by block length n which communicate at the rate triple $(R_S : S \in E)$ and which achieve a vanishing probability of error $P_e^{(n)}$ as n tends to infinity. Pick the code corresponding to block length n, and for each $S \in E$, let M_S be a uniformly distributed message on $[2^{nR_S}]$. Let $\mathbf{X}_1^n, \ldots, \mathbf{X}_K^n$ and $\mathbf{Y}^n$ be the random variables be induced by the messages, the encoders, and the channel.

Let $C(\mathbf{X}) = \log\det(\mathbf{I} + \mathbf{X})$ and

$$\begin{aligned} M(\mathcal{B}) &= (M_S)_{S\in\mathcal{B}} && \text{for } \mathcal{B} \subseteq E \\ \mathbf{X}_t(S) &= (\mathbf{X}_{j,t})_{j\in S} && \text{for } S \subseteq [K] \\ \mathbf{H}(N) &= (\mathbf{H}(i))_{i\in N} && \text{for } N \subseteq [n]. \end{aligned}$$

By Fano's inequality, there is a vanishing non-negative sequence $\epsilon_n \to 0$ such that for every non-empty subset $S \subseteq [K]$,

$$\begin{aligned} n\left(\sum\nolimits_{D\in\downarrow S} R_D - \epsilon_n\right) &\overset{(i)}{\leq} I(M_{\downarrow S}; \mathbf{Y}^n) \\ &\leq I(M_{\downarrow S}; \mathbf{Y}^n | M_{E\setminus\downarrow S}, \mathbf{H}([n])) \\ &= \sum_{t=1}^{n} h(\mathbf{Y}_t | M_{E\setminus\downarrow S}, \mathbf{H}([n]), \mathbf{Y}^{t-1}) - h(\mathbf{Z}_t) \\ &\leq \sum_{t=1}^{n} h(\mathbf{Y}_t | \mathbf{X}_{[K]\setminus S,t}, \mathbf{H}([n])) - h(\mathbf{Z}_t) \\ &= \sum_{t=1}^{n} E_{\mathbf{H}([n])=\mathbf{h}([n])} h(\mathbf{Y}_t | \mathbf{X}_{[K]\setminus S,t}, \mathbf{H}([n]) = \mathbf{h}([n])) \\ &\qquad - h(\mathbf{Z}_t) \end{aligned}$$

$$
\begin{aligned}
&\overset{(ii)}{\leq} \sum_{t=1}^{n} E_{\mathbf{H}([n])} C\left(\mathbf{H}_S(t)\mathbf{K}_{St}(\mathbf{H}([n]))\mathbf{H}_S^*(t)\right) \\
&\overset{(iii)}{\leq} \sum_{t=1}^{n} E_{\mathbf{H}([n])} C\left(|S| \sum_{j\in S} \mathbf{H}_j(t)\mathbf{K}_{jj,t}(\mathbf{H}([n]))\mathbf{H}_j^*(t)\right) \\
&\overset{(iv)}{\leq} \sum_{t=1}^{n} E_{\mathbf{H}([n])} C\left(|S| \sum_{j\in S} \mathbf{H}_j(t)\mathbf{H}_j^*(t)\rho_{jt}(\mathbf{H}([n]))\right) \\
&= \sum_{t=1}^{n} E_{\mathbf{H}(t)} E_{\mathbf{H}([n]\setminus\{t\})} C\left(|S| \sum_{j\in S} \mathbf{H}_j(t)\mathbf{H}_j^*(t)\rho_{jt}(\mathbf{H}([n]))\right) \\
&\overset{(v)}{\leq} \sum_{t=1}^{n} E_{\mathbf{H}(t)} C\left(|S| \sum_{j\in S} \mathbf{H}_j(t)\mathbf{H}_j^*(t)\rho_{jt}(\mathbf{H}(t))\right) \\
&\overset{(vi)}{=} nE_{\mathbf{H}} \left(\sum_{t=1}^{n} \frac{1}{n} C\left(|S| \sum_{j\in S} \mathbf{H}_j\mathbf{H}_j^*\rho_{jt}(\mathbf{H})\right)\right) \\
&\overset{(vii)}{\leq} nE_{\mathbf{H}} C\left(\textstyle\sum_{j\in S} \mathbf{H}_j\mathbf{H}_j^*\rho_j(\mathbf{H})\right).
\end{aligned}
$$

where $E_{\mathbf{H}}\rho_j(\mathbf{H}) \leq P_j$. The labeled steps above follow by

(i) Fano's inequality.

(ii) the definitions in (4.39),(4.40) below and the fact that circularly symmetric Gaussian distributions maximize conditional entropy subject to a joint covariance constraint.

(iii) Lemma 4.7.6 below and the fact that $C(\mathbf{X}) \leq C(\mathbf{Y})$ if $\mathbf{0} \preceq \mathbf{X} \preceq \mathbf{Y}$.

(iv) (4.41a) and the fact that for any $\mathbf{A} \succeq \mathbf{0}$, $\mathbf{A} \preceq \mathrm{tr}(\mathbf{A})\mathbf{I}$.

(v) the concavity of $C(\mathbf{X})$ over $\mathbf{X} \succeq \mathbf{0}$, Jensen's inequality, and the definitions (4.41).

(vi) the stationarity of the stochastic process $\mathbf{H}(t)$.

(vii) the same reasoning as (v) above.

We now explain the covariance notation used above. For each channel state sequence realization $\mathbf{H}([n]) = \mathbf{h}([n])$, channel use $t \in [n]$, and any subset $S = \{j_1, \ldots, j_{|S|}\} \subseteq [K]$ where the indices

are ordered in increasing order, $j_1 < \cdots < j_{|S|}$, let

$$\mathbf{K}_{St}(\mathbf{h}([n])) = \mathrm{Cov}\left(\begin{bmatrix} \mathbf{X}_{j_1,t} \\ \vdots \\ \mathbf{X}_{j_{|S|},t} \end{bmatrix}, \begin{bmatrix} \mathbf{X}_{j_1,t} \\ \vdots \\ \mathbf{X}_{j_{|S|},t} \end{bmatrix} \middle| \mathbf{H}([n]) = \mathbf{h}([n]) \right)$$
$$= \begin{bmatrix} \mathbf{K}_{j_1 j_1 t}(\mathbf{h}([n])) & \cdots & \mathbf{K}_{j_1 j_{|S|} t}(\mathbf{h}([n])) \\ \vdots & \ddots & \vdots \\ \mathbf{K}^*_{j_1 j_{|S|} t}(\mathbf{h}([n])) & \cdots & \mathbf{K}_{j_{|S|} j_{|S|} t}(\mathbf{h}([n])) \end{bmatrix}. \tag{4.39}$$

For the same subset $S = \{j_1, \ldots, j_{|S|}\}$, denote the portion of the channel fading matrix seen by the inputs listed in S as

$$\mathbf{H}_S(t) = \begin{bmatrix} \mathbf{H}_{j_1}(t) & \cdots & \mathbf{H}_{j_{|S|}}(t) \end{bmatrix}. \tag{4.40}$$

Our development was specifically interested in the empirical power allocation and certain averages of this power allocation,

$$\rho_{jt}(\mathbf{H}([n]) = \mathrm{tr}\Big(\mathbf{K}_{jjt}(\mathbf{h}([n])) \Big) \tag{4.41a}$$
$$\rho_{jt}(\mathbf{H}(t)) = E_{\mathbf{H}([n]\setminus\{t\})} \rho_{jt}(\mathbf{H}([n]))$$
$$\rho_j(\mathbf{H}) = \frac{1}{n} \sum_{t=1}^{n} \rho_{jt}(\mathbf{H}).$$

As the code is achievable, its empirical power allocation must satisfy the power constraint, so that $\frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{H}([n])} \rho_{jt}(\mathbf{H}([n])) \le P_j$. Hence $\int_{\mathcal{H}} \rho_j(\nu) \mathrm{d}F(\nu) \le P_j$ as well.

The following lemma was critical in the above outer bound.

Lemma 4.7.6. *If* $\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{K}^*_{1p} & \cdots & \mathbf{K}_{pp} \end{bmatrix} \succeq \mathbf{0}$ *where* $p \ge 2$, *then*

$$\begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{K}^*_{1p} & \cdots & \mathbf{K}_{pp} \end{bmatrix} \preceq p \begin{bmatrix} \mathbf{K}_{11} & & \\ & \ddots & \\ & & \mathbf{K}_{pp} \end{bmatrix}.$$

Proof. This may be shown by induction. The lemma is trivially true for $p = 1$. Now assume that the lemma is true for any $p-1$ block partition of a positive semi-definite matrix where $p \geq 2$. Consider a $\mathbf{K} \succeq \mathbf{0}$ with a p block partition as stated in the hypothesis. Group the bottom right $(p-1) \times (p-1)$ block of blocks to form the following 2×2 partition of $\mathbf{K}$,

$$\mathbf{K} = \left[\begin{array}{c|ccc} \mathbf{K}_{11} & \mathbf{K}_{12} & \cdots & \mathbf{K}_{1p} \\ \hline \mathbf{K}_{12}^* & \mathbf{K}_{22} & \cdots & \mathbf{K}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{1p}^* & \mathbf{K}_{2p}^* & \cdots & \mathbf{K}_{pp} \end{array}\right].$$

Then by Lemma C.3.1 in the Appendix, three necessary and sufficient conditions for $\mathbf{K}$ to be positive semi-definite are

$$\mathbf{K}_{11} \succeq \mathbf{0} \tag{4.42a}$$

$$\text{Null}(\mathbf{K}_{11}) \subseteq \text{Null}\left(\begin{bmatrix} \mathbf{K}_{12} & \cdots & \mathbf{K}_{1p} \end{bmatrix}^*\right) \tag{4.42b}$$

$$\begin{bmatrix} \mathbf{K}_{22} & \cdots & \mathbf{K}_{2p} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{2p}^* & \cdots & \mathbf{K}_{pp} \end{bmatrix} \succeq \begin{bmatrix} \mathbf{K}_{12}^* \\ \vdots \\ \mathbf{K}_{1p}^* \end{bmatrix} \mathbf{K}_{11}^+ \begin{bmatrix} \mathbf{K}_{12} & \cdots & \mathbf{K}_{1p} \end{bmatrix}. \tag{4.42c}$$

Construct

$$\tilde{\mathbf{K}} = p \begin{bmatrix} \mathbf{K}_{11} & & \\ & \ddots & \\ & & \mathbf{K}_{pp} \end{bmatrix} - \begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{1p}^* & \cdots & \mathbf{K}_{pp} \end{bmatrix} = \left[\begin{array}{c|ccc} (p-1)\mathbf{K}_{11} & -\mathbf{K}_{12} & \cdots & -\mathbf{K}_{1p} \\ \hline -\mathbf{K}_{12}^* & (p-1)\mathbf{K}_{22} & \cdots & -\mathbf{K}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{K}_{1p}^* & -\mathbf{K}_{2p}^* & \cdots & (p-1)\mathbf{K}_{pp} \end{array}\right].$$

We will show that $\tilde{\mathbf{K}}$ meets the three conditions (4.42) for its constituent blocks so that (with Lemma C.3.1) we may conclude $\tilde{\mathbf{K}} \succeq \mathbf{0}$, our desired result. The first two such conditions follow as direct consequences of (4.42a) and (4.42b),

$$(p-1)\mathbf{K}_{11} \succeq \mathbf{0} \tag{4.43a}$$

$$\text{Null}((p-1)\mathbf{K}_{11}) \subseteq \text{Null}\left(-\begin{bmatrix} \mathbf{K}_{12} & \cdots & \mathbf{K}_{1p} \end{bmatrix}^*\right). \tag{4.43b}$$

For the last condition, note that

$$\begin{bmatrix} -\mathbf{K}_{12}^* \\ \vdots \\ -\mathbf{K}_{1p}^* \end{bmatrix} ((p-1)\mathbf{K}_{11})^+ \begin{bmatrix} -\mathbf{K}_{12} & \cdots & -\mathbf{K}_{1p} \end{bmatrix} = \frac{1}{p-1} \begin{bmatrix} \mathbf{K}_{12}^* \\ \vdots \\ \mathbf{K}_{1p}^* \end{bmatrix} \mathbf{K}_{11}^+ \begin{bmatrix} \mathbf{K}_{12} & \cdots & \mathbf{K}_{1p} \end{bmatrix}$$

$$\overset{(i)}{\preceq} \frac{1}{p-1} \begin{bmatrix} \mathbf{K}_{22} & \cdots & \mathbf{K}_{2p} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{2p}^* & \cdots & \mathbf{K}_{pp} \end{bmatrix} \overset{(ii)}{\preceq} \begin{bmatrix} \mathbf{K}_{22} & & \\ & \ddots & \\ & & \mathbf{K}_{pp} \end{bmatrix}$$

$$\overset{(iii)}{\preceq} \begin{bmatrix} (p-1)\mathbf{K}_{22} & \cdots & -\mathbf{K}_{2p} \\ \vdots & \ddots & \vdots \\ -\mathbf{K}_{2p}^* & \cdots & (p-1)\mathbf{K}_{pp} \end{bmatrix}, \tag{4.44}$$

where (i) follows by (4.42c) and $(ii), (iii)$ both follow by the inductive assumption. The result follows by noting (4.43)-(4.44) are sufficient conditions for $\tilde{\mathbf{K}}$ to be positive semi-definite. □

Chapter 5

Broadcast Channel with General Message Sets

5.1 Motivation

In contrast to the Multiple Access Channel, answers for the Broadcast Channel are hard to come by, including capacity for the general two-user two-private message discrete memoryless channel. Even with a focus on only multiple unicast transmission, progress has been made in specific cases. In the discrete memoryless setting, capacity is known in cases where the outputs can be ordered in terms of channel quality. In the MIMO setting, capacity was determined recently [116], and is equal to the Marton's inner bound with inputs selected according to the dirty-paper coding rule.

Similarly, progress for the case with simultaneous unicast and multicast sessions has been limited to special cases [74, 37, 29]:

- The two-user MIMO setting, where capacity and its approximate Degrees-of-freedom characterization require subtle answers.
- A noise-less, but rate-limited broadcast network known as a combination network [75], with one transmitter and several receivers connected through an intermediate layer of nodes and connections (a three-user version is depicted in Figure 5.1). When there are three or more receivers, it is known that network coding is required and for the case where capacity is known, linear network coding suffices.
- For three-user discrete memoryless channels with two degraded message sets, capacity can

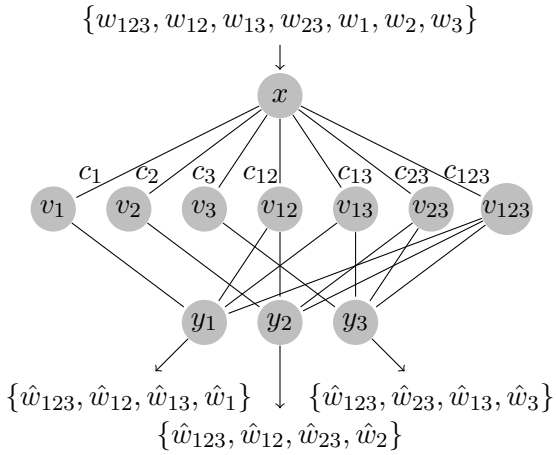


Figure 5.1: Three-user combination network, the links between the first and second layer are rate-limited, while the links between the second and third layer are not.

be found by first constructing a code for a more general message set, and then projecting away the undesired message set [74]. This demonstrates that a more general understanding of general message sets may be required for the understanding of simpler message sets.

In more detail; the two-user MIMO capacity was only recently determined in [37] through the development of novel arguments for the sufficiency of Gaussian inputs in the attainment of capacity. There, capacity is characterized implicitly, and an explicit characterization with a prescription for optimal input covariances appears difficult to obtain. An approximately optimal input covariance is provided in [29], where the choice provided is shown to attain capacity to within a finite gap, where the gap is independent of the signal to noise ratio, but dependent on the condition number of the channel matrix.

For the combination network, some partial answers are known. For the three user case, the capacity region characterized in Grokop and Tse [42]. Tian [105] generalizes this to a symmetric K-user setting, characterizing capacity not as an explicit polyhedral region but implicitly in terms of rate-splits. Despite this, a clever converse argument reliant on the **submodularity** of entropy demonstrates that the rate region achievable by linear network coding is the capacity region. The central role of submodularity in the prior works is brought to the forefront with the recent work

of Salimi **et al.** [91], who develop a notion of **generalized** cut-set bounds. With this viewpoint Salimi **et al.**, simplify the converse arguments of [105, 42] and uncover an explicit polyhedral representation of the symmetric K-user capacity region.

Motivated by its relevance and the difficulty of the general case, we study the MIMO and focus on the approximate notion of capacity known as the degress-of-freedom (DoF) region.This measure characterizes the trade-offs inherent in the transmission of distinct messages across the same channel medium as the trade-offs in the assignment of the available signal spaces to the transmission of different messages. We start by observing that the DoF region for the two-user case can be attained without appeal to dirty paper coding and through the use of simple linear precoding and rate-transfer operations. More precisely, capacity may be attained by concatenating two types of codes together: a zero-forcing inner code, and a rate-transfer outer code. Inspired by this observation, we propose a general inner bound with the same structure: a zero-forcing inner code, and a linear network coding outer code, as depicted in Figure 5.2.

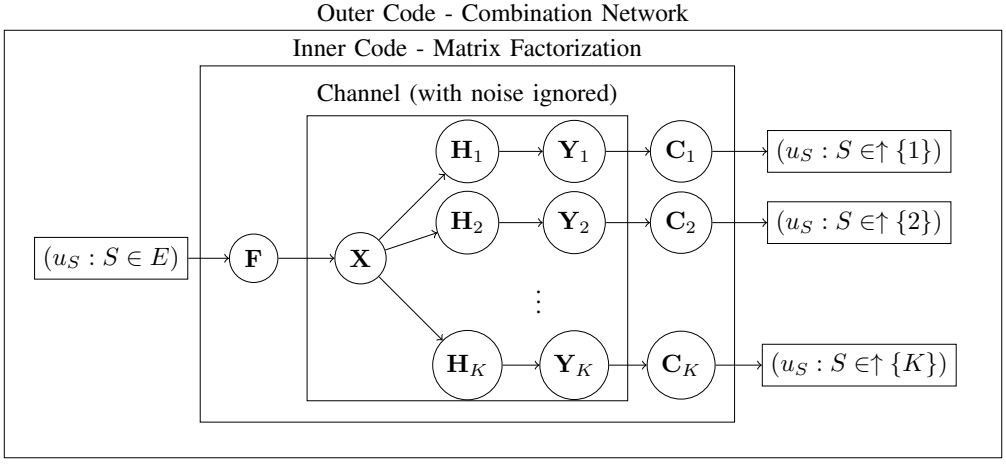


Figure 5.2: Schematic of Degrees-of-Freedom inner Bound for the BC as a concatenation of two codes.

For the zero-forcing inner code, we consider a recursive row-vector selection procedure. The possibilities of the row-vector selection process are described through polymatroid constraints, with the bounds being the dimension of the direct sum over various elements of the intersection lattice of the arrangement of row spaces. The resultant set of achievable DoF can be further enlarged through

consideration of linear network coding. The development of the rates achievable through linear network coding relies on prior work on the finite-field broadcast networks with linear input/output relations [58] to develop a scalable inner bound applicable to all cases. For certain three-user and symmetric cases, we find a precise polyhedral representation of this inner bound; for the remaining cases, the inner bound remains implicitly defined through the introduction of rate-splits.

When the recursive process eventually selects all row vectors, it yields a matrix factorization which enables a converse argument to be applied. In these cases, the proposed DoF region is optimal, and its governing constraints require not only cut-set bounds, but generalized cut-set bounds.

5.2 DoF Inner Bound

5.2.1 Channel Model and Preliminaries

The Gaussian vector broadcast channel (BC) is

$$\begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \mathbf{X} + \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_K \end{bmatrix}, \tag{5.1}$$

where $\mathbf{X}$ is the input of dimension t, the jth output $\mathbf{Y}_j$ is of dimension r_j, and the channel matrix $\mathbf{H}_j$ between the two is of size $r_j \times t$. Each of the additive noises $\mathbf{Z}_j$ are Gaussian with identity covariance. Without loss of generality, we assume that $r_j \leq t$ and that $\text{rank}(\mathbf{H}_j) = r_j$. When we take the channel matrices to be generic, the resultant DoF region will rely only on the transmit and collective receive dimensions, which we refer to with the notation $t \times (r_1, r_2, \ldots, r_K)$.

Our interest is in studying this channel where the transmitter has both private and common messages to send. Again, E is a message index set, containing subsets $S \subseteq [1:K]$, which lists the receivers at which the corresponding message source m_S is desired. In the Gaussian setting, the input must satisfy the power constraint $E\left[\frac{1}{n}\sum_{t=1}^{n} \|\mathbf{X}_t\|^2\right] \leq P$, where the expectation is with respect to a uniform distribution on the messages, is satisfied. A rate tuple $(R_S(P) : S \in E)$ is said

to be achievable if its reconstruction error, as a function of block length n, tends to zero for some sequence of codes satisfying the power constraint with respect to the power budget P. The closure of all such rate tuples is the capacity region $\mathcal{C}(P)$.

The degrees of freedom (DoF) region is, informally, the capacity region in the high signal to noise ratio regime, characterizing the number of complex symbols which may be allocated towards the transmission of each message. Formally, it is the set of non-negative tuples $(d_S : S \in E)$ satisfying

$$\sum_{S\in E} \mu_S d_S \leq \limsup_{P\to\infty} \left(\sup_{R(P)\in\mathcal{C}(P)} \left[\sum_{S\in E} \mu_S R_S(P) \right] \frac{1}{\log(P)} \right)$$

for all non-negative weight vectors $(\mu_S \geq 0 : S \in E)$ [51]. By considering successively higher power budgets, the signal-to-noise ratio can be made arbitrarily high so that the impact of the additive noise is eliminated and the competition among the various signals to be transmitted is brought to the forefront.

5.2.2 Two-user Case: Matrix Factorization

We start with a known result, the DoF region characterized by Ekrem and Ulukus [29], which relies on a matrix factorization known as the generalized singular value decomposition (SVD). For the purposes of the DoF region, a simpler matrix factorization suffices, which forgoes desirable numerical attributes of the GSVD has for the purposes of a simpler analytical treatment. We review the development of Ekrem and Ulukus [29] first, which is reliant on the GSVD first. The generalized singular value decomposition (GSVD) is, for the pair of channel matrices $\mathbf{H}_1, \mathbf{H}_2$, given by

$$\mathbf{\Phi}_1 \mathbf{H}_1 \mathbf{\Phi}_0 = \mathbf{\Sigma}_1 [\mathbf{\Omega}^{-1} \mathbf{0}]$$

$$\mathbf{\Phi}_2 \mathbf{H}_2 \mathbf{\Phi}_0 = \mathbf{\Sigma}_2 [\mathbf{\Omega}^{-1} \mathbf{0}]$$

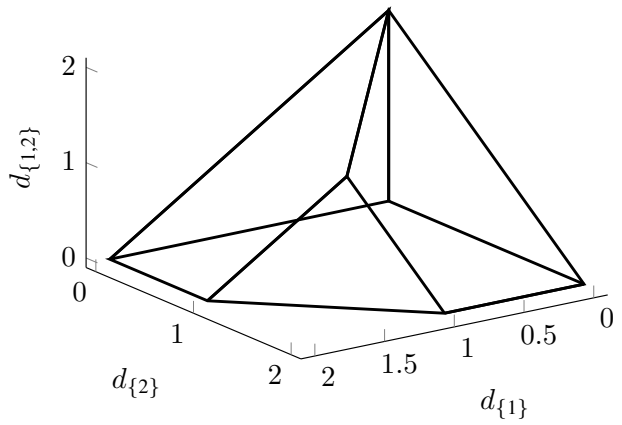


Figure 5.3: Two user DoF region for a generic $3 \times (2,2)$ two-user MIMO BC with multicasting at the physical layer.

where $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_0$ are unitary, $\mathbf{\Omega}$ is non-singular and of size $k = \text{rank}(\mathbf{H})$, and

$$\mathbf{\Sigma}_1 = \begin{bmatrix} \mathbf{I}_{\phi_1 \times \phi_1} & & \\ & \mathbf{D}_{1,\phi_{12} \times \phi_{12}} & \\ & & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{r_1 \times k} \quad \mathbf{\Sigma}_2 = \begin{bmatrix} \mathbf{0} & & \\ & \mathbf{D}_{1,\phi_{12} \times \phi_{12}} & \\ & & \mathbf{I}_{\phi_2 \times \phi_2} \end{bmatrix} \in \mathbb{R}^{r_2 \times k}.$$

Here, ϕ_{12} is the dimension of the common subspace $\text{Null}(\mathbf{H}_1)^\perp \cap \text{Null}(\mathbf{H}_2)^\perp$ [1], while each dimension ϕ_i are the dimensions available for private message broadcasting after those directions which lie in the common subspace have been set aside; that is, $\phi_i = \text{rank}(\mathbf{H}_i) - \phi_{12}$. With this decomposition, the two-user BC can be transformed into a set of parallel scalar BCs, for which a previously known converse [32] provides that any achievable DoF tuple must satisfy

$$\mathbf{E}_2 \begin{bmatrix} d_1 \\ d_2 \\ d_{12} \end{bmatrix} \leq \mathbf{E}_2 \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_{12} \end{bmatrix} \qquad \text{where} \qquad \mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{5.2}$$

Through dirty paper coding and a covariance choice dependent on the GSVD, these conditions can be shown to be sufficient as well as necessary for achievability. An example of such a region, for a $3 \times (2,2)$ BC with generic channel matrices (that is, $\mathbf{H}_1, \mathbf{H}_2$ are both of dimension 2×3 assuring, through genericity, that $\phi_1 = \phi_2 = \phi_{12} = 1$), is depicted in Figure 5.3.

[1] $A^\perp$ denotes the orthogonal complement of A: $A^\perp = \{y : \langle y, x \rangle = 0 \ \forall x \in A\}$, where $\langle y, x \rangle = y^* x$ is the standard inner product.

The prior formulation provides a relatively strong notion of capacity; capacity to within a gap independent of the signal to noise ratio. If interested in directly obtaining the DoF region, simpler arguments suffice. The converse, in fact, is immediate as (5.2) is simply

$$\begin{aligned} d_{12} + d_1 &\leq \text{rank}(\mathbf{H}_1) \\ d_{12} + d_2 &\leq \text{rank}(\mathbf{H}_2) \\ d_{12} + d_1 + d_2 &\leq \text{rank}(\mathbf{H}), \end{aligned} \tag{5.3}$$

which are recognizable as the standard cut-set bounds.

Achievability can also be shown more simply by concatenating two simple codes. The first is a zero-forcing inner code which characterizes all those DoF tuples which are achievable by carrying each message-bearing symbol on at most one linear dimension within the transmit space. The second is a rate-splitting code, which considers those common DoF points requiring that the message-bearing symbol occupy more than one linear dimension in the transmit space. Overlaying the outer code over the inner code achieves all points within the outer bound (5.3).

5.2.2.1 Inner code

The inner code parallels the development of the private message case, where the transmission scheme assigns each message-bearing symbol to a different channel row vector so that all assigned row vectors are linearly independent. As it suffices to have any linearly independent selection, pick from each channel matrix $\mathbf{H}_j$ only $\text{rank}(\mathbf{H}_j)$ of its rows, and redefine the channel matrix to be this subset of rows. In this way, we can assume that $r_j = \text{rank}(\mathbf{H}_j)$, without affecting the row spaces involved.

Another change to the selection of the transmission row vectors is relevant, by applying a nonsingular receive filter $\mathbf{G}_j$, the jth transmitter can change transmission rows that the transmitter has available to an alternative basis than the one provided by $\mathbf{H}_j$. Thus, we focus on the row spaces, rather that the specific row vectors, where the following achievable scheme will provide a means to select the appropriate receive filters. Preferring to work with column vectors over row vectors, we

focus on the equivalent representation of the row space of $\mathbf{H}_j$ as the column space of $\mathbf{H}_j^*$,

$$\mathcal{R}_1 = \text{Range}(\mathbf{H}_2^*) = \text{Null}(\mathbf{H}_2)^\perp$$

$$\mathcal{R}_2 = \text{Range}(\mathbf{H}_1^*) = \text{Null}(\mathbf{H}_1)^\perp,$$

along with their intersection

$$\mathcal{R}_{12} = \mathcal{R}_1 \cap \mathcal{R}_2.$$

This common subspace $\mathcal{R}_{12}$, whose dimension is the previously defined ϕ_{12}, plays a central role for the transmission of common messages to both receivers: a common message-bearing symbol intended to be received at both destinations and is along a single single transmit direction only if that direction is within the common subspace $\mathcal{R}_{12}$.

For concreteness, choose matrices

$$\tilde{\mathbf{H}}_{12} = \begin{bmatrix} \mathbf{x}_1^{12} \\ \vdots \\ \mathbf{x}_{\phi_{12}}^{12} \end{bmatrix} \qquad \tilde{\mathbf{H}}_1 = \begin{bmatrix} \mathbf{x}_1^{1} \\ \vdots \\ \mathbf{x}_{\phi_2}^{1} \end{bmatrix} \qquad \tilde{\mathbf{H}}_1 = \begin{bmatrix} \mathbf{x}_2^{1} \\ \vdots \\ \mathbf{x}_{\phi_2}^{2} \end{bmatrix}.$$

where rows of $\tilde{\mathbf{H}}_{12}$ form a basis for the common subspace $\mathcal{R}_{12}$ and the rows of $\tilde{\mathbf{H}}_1$ contains rows within R_1 but linearly independent of those rows in $\tilde{\mathbf{H}}_{12}$. Similarly, the rows of $\tilde{\mathbf{H}}_2$ contains rows within R_2 but linearly independent of those rows in $\tilde{\mathbf{H}}_{12}$. Then there are a pair of non-singular $\mathbf{G}_1, \mathbf{G}_2$ such that

$$\mathbf{G}_1\mathbf{H}_1 = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \\ \tilde{\mathbf{H}}_{12} \end{bmatrix} \qquad\qquad \mathbf{G}_2\mathbf{H}_2 = \begin{bmatrix} \tilde{\mathbf{H}}_{12} \\ \\ \tilde{\mathbf{H}}_2 \end{bmatrix}.$$

Now, consider the set of non-negative DoF tuples (d_1', d_2', d_{12}') satisfying[2]

$$\begin{aligned} d_{12}' &\le \dim(\mathcal{R}_{12}) = \phi_{12} \\ d_1' + d_{12}' &\le \dim(\mathcal{R}_1 \oplus \mathcal{R}_{12}) = \phi_1 + \phi_{12} \\ d_2' + d_{12}' &\le \dim(\mathcal{R}_2 \oplus \mathcal{R}_{12}) = \phi_2 + \phi_{12} \\ d_1' + d_2' + d_{12}' &\le \dim(\mathcal{R}_1 \oplus \mathcal{R}_2 \oplus \mathcal{R}_{12}) = \phi_1 + \phi_2 + \phi_{12}, \end{aligned} \tag{5.4}$$

[2] $A \oplus B = \{a + b : a \in A, b \in B\}$ denotes the direct sum of two subspaces A and B.

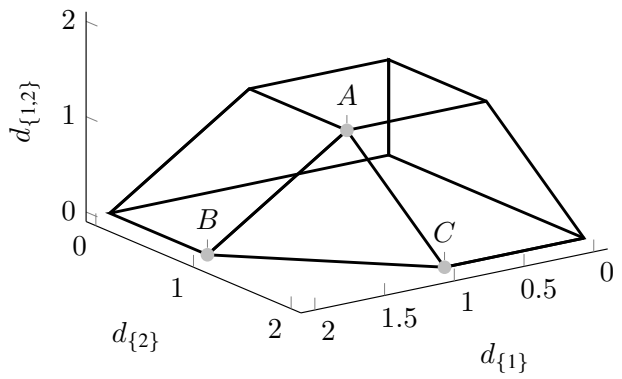


Figure 5.4: The polytope (5.4) for a generic $3 \times (2,2)$ two-user MIMO BC. Each of the three verticies A, B, C can be achieved with zero-forcing and the remaining region can be achieved with time sharing.

where $\dim(V)$ is the dimension of the linear space V. As in the private message case, this polytope has a special structure: it is a polymatroid, though it may not be apparent at first. To see this, let $\mathbf{F}_S$ be a matrix with $\text{Range}(\mathbf{F}_S) = \mathcal{R}_S$ for $S \in \{1, 2, 12\} = E$. Then each of the bounds in (5.4) may be reformulated as

$$\sum\nolimits_{S \in B} d_S \leq f(B) \text{ for } B \in \Big\{ \{12\}, \{1,12\}, \{2,12\}, \{1,2,12\} \Big\} \tag{5.5}$$

where $f(B) = \dim(\oplus_{S \in B} \mathcal{R}_S) = \text{rank}(\mathbf{F}_B)$, and $\mathbf{F}_B$ is the matrix formed by horizontally concatenating the matrices $(\mathbf{F}_S : S \in B)$. Thus we recognize $f(B)$ as a normalized, nonincreasing, submodular function (c.f. (2.13)) over the subsets $B \subseteq E$. Further, as $\mathcal{R}_{12} \subseteq \mathcal{R}_1$ and $\mathcal{R}_{12} \subseteq \mathcal{R}_2$, we observe that all of the defining inequalities of $\mathcal{P}(f)$, given by

$$\sum\nolimits_{S \in B} d_S \leq f(B) \quad \text{ for all } B \subseteq E,$$

are implied by those in (5.5). Thus, we have an explicit description of each vertex of $\mathcal{P}(f)$. If we can achieve each of these vertices, then through time-sharing we can achieve any point within $\mathcal{P}(f)$. By the redundancy of the bounds, $\mathcal{P}(f)$ has three unique verticies:

A $(d_1^*, d_2^*, d_{12}^*) = (\phi_1, \phi_2, \phi_{12})$

B $(d_1^*, d_2^*, d_{12}^*) = (\phi_1 + \phi_{12}, \phi_2, 0)$

C $(d_1^*, d_2^*, d_{12}^*) = (\phi_1, \phi_2 + \phi_{12}, 0)$;

see Figure 5.4 for an example of such a polymatroid.

All of the corner points can be achieved in a similar manner; for illustrative purposes we focus on point A. This corner point attains the maximum of any linear program

$$\text{max.} \sum\nolimits_{S \in E} \mu_S d_S \qquad \text{s.t.} \qquad d \in \mathcal{P}(f)$$

with $\mu_{12} \geq \max\{\mu_1, \mu_2\}$ as the explicit formula for polymatroid vertices provides

$$\begin{aligned} d_{12}^* &= \dim(\mathcal{R}_{12}) = \text{rank}(\tilde{\mathbf{H}}_{12}) = \phi_{12} \\ d_1^* &= \dim(\mathcal{R}_1) - \dim(\mathcal{R}_{12}) = \text{rank}(\tilde{\mathbf{H}}_1) = \phi_1 \\ d_2^* &= \dim(\mathcal{R}_1 \oplus \mathcal{R}_2) - \dim(\mathcal{R}_1) = \text{rank}(\tilde{\mathbf{H}}_2) = \phi_2. \end{aligned}$$

To attain this DoF point, assigning transmit directions to message-bearing symbols as follows.

$i = 1$ Select d_{12}^* linearly independent row of $\tilde{\mathbf{H}}_{12}$ for the transmission of m_{12}, and assemble those rows into the $d_{12}^* \times t$ matrix $\mathbf{F}_{12}$.

$i = 2$ As

$$\text{rank}(\mathbf{F}_{12}) = \text{rank}(\tilde{\mathbf{H}}_{12}) \qquad \text{rank}(\begin{bmatrix} \mathbf{F}_{12} & \tilde{\mathbf{H}}_1 \end{bmatrix}) = \text{rank}(\mathbf{H}_1),$$

the augmentation property assures that we can select d_1^* rows from $\tilde{\mathbf{H}}_1$ and assemble them into a $d_1^* \times t$ matrix $\mathbf{F}_1$ such that $\begin{bmatrix} \mathbf{F}_{12} & \mathbf{F}_1 \end{bmatrix}$ has linearly independent rows

$i = 3$ As

$$\text{rank}(\begin{bmatrix} \mathbf{F}_{12} & \mathbf{F}_1 \end{bmatrix}) = \text{rank}(\mathbf{H}_1) \qquad \text{rank}(\begin{bmatrix} \mathbf{F}_{12} & \mathbf{F}_1 & \tilde{\mathbf{H}}_2 \end{bmatrix}) = \text{rank}(\mathbf{H}_{[1:2]}),$$

the augmentation property assures that we can select d_2^* rows from $\tilde{\mathbf{H}}_2$ and assemble them into a $t \times d_2^*$ matrix $\mathbf{F}_2$ such that $\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{F}_{12} & \mathbf{F}_1 & \mathbf{F}_2 \end{bmatrix}$ has linearly independent rows.

By the linear independence of columns of $\mathbf{F}$, we have

$$\mathbf{G}_1\mathbf{H}_1\tilde{\mathbf{H}}^{-1} = \begin{matrix} \\ \phi_1 \\ \phi_{12} \end{matrix} \begin{matrix} \begin{matrix} \phi_1 & \phi_{12} & \phi_2 \end{matrix} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} \end{matrix} \qquad \mathbf{G}_2\mathbf{H}_2\tilde{\mathbf{H}}^{-1} = \begin{matrix} \\ \phi_{12} \\ \phi_2 \end{matrix} \begin{matrix} \begin{matrix} \phi_1 & \phi_{12} & \phi_2 \end{matrix} \\ \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \end{matrix}.$$

Hence, by sending a message-bearing symbol along a direction in $\tilde{\mathbf{H}}_{12}$, we may recover it at both receivers, and if we send a message-bearing symbol along a direction in $\tilde{\mathbf{H}}_i$, then we may recover it only at the ith receiver. In an analogous fashion, we may attain the corner points B and C.

5.2.2.2 Outer code

To achieve that entire region (5.3), we enlarge the region (5.4) through a **rate-splitting** operation. Let (d_1', d_2', d_{12}') be a DoF tuple within the region (5.4) and thus achievable. If both d_1' and d_2' are positive, we may set aside a portion $0 \leq \Delta_{12} \leq \min(d_1, d_2)$ of each DoF not for the transmission of a private message, but for the transmission of a common message. Doing so yields the achievability of the DoF tuple

$$d_1 = d_1' - \Delta_{12} \qquad d_2 = d_2' - \Delta_{12} \qquad d_{12} = d_{12}' + \Delta_{12}.$$

By projecting away this rate-split, any DoF tuple within (5.3) is achievable.

5.2.3 Towards a K-user extension

With the insight that in the two-user case, the DoF region follows straightforwardly form a matrix factorization, we seek similarly useful matrix factorizations for the K-user case.

Define the individual, and shared row spaces as before, with

$$R_{\{j\}} = \text{Null}(\mathbf{H}_j)^{\perp} \qquad R_S = \bigcap_{j \in S} R_{\{j\}} \ .$$

We call this the row-space intersection lattice, as these linear subspaces are partially ordered by inclusion: $R_S \subset R_{S'}$ if $S' \subset S$. Then a three-user matrix factorization analogous to the two-user

case of K- users would have the first users' matrix would factor as

$$\mathbf{G}_1\mathbf{H}_1 = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_{12} \\ \tilde{\mathbf{H}}_{13} \\ \tilde{\mathbf{H}}_{123} \end{bmatrix} \quad \text{s.t.} \quad \begin{aligned} \text{Range}([\tilde{\mathbf{H}}_1^*, \tilde{\mathbf{H}}_{12}^*, \tilde{\mathbf{H}}_{13}^*, \tilde{\mathbf{H}}_{123}^*]) &= \dim(\text{R}_1) \\ \text{Range}([\tilde{\mathbf{H}}_{12}^*, \tilde{\mathbf{H}}_{123}^*]) &= \dim(\text{R}_{12}) \\ \text{Range}([\tilde{\mathbf{H}}_{13}^*, \tilde{\mathbf{H}}_{123}^*]) &= \dim(\text{R}_{13}) \\ \text{Range}([\tilde{\mathbf{H}}_{123}^*]) &= \dim(\text{R}_{123}) \end{aligned} \quad \text{and } \tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \\ \tilde{\mathbf{H}}_{12} \\ \tilde{\mathbf{H}}_3 \\ \tilde{\mathbf{H}}_{13} \\ \tilde{\mathbf{H}}_{23} \\ \tilde{\mathbf{H}}_{123} \end{bmatrix} \text{ has a right-inverse.}$$

5.2.3.1 Complications

This cannot be done in general, for example, consider a three-user channel where there are three transmit antennas, the first two users have two transmit antennas and the last user has a single antenna (succinctly, a $3 \times (2, 2, 1)$ BC). Suppose the channel matrices are

$$\mathbf{H}_1 = \begin{bmatrix} 0.54 & -1.3 & -1.3 \\ 1.8 & -0.43 & 3 \end{bmatrix} \quad \mathbf{H}_2 = \begin{bmatrix} -2.3 & 0.34 & 0.73 \\ 0.86 & 3.6 & -0.063 \end{bmatrix} \quad \mathbf{H}_3 = \begin{bmatrix} 0.32 & 2.8 & 0.71 \end{bmatrix}$$

The row spaces of the first two users share a single two vector, so that R_{12} is the line parameterized by this vector. The remaining shared subspaces contain only the zero vector. Re-parametrized row spaces by applying a receive filter at the output of the first two users so that one receive direction lies in R_{12} and the other direction does not. Doing so yields

$$\mathbf{G}_1\mathbf{H}_1 = \begin{bmatrix} -0.59 & 0.77 & 0.23 \\ 1.8 & -0.43 & 3 \end{bmatrix} \quad \mathbf{G}_2\mathbf{H}_2 = \begin{bmatrix} -0.59 & 0.77 & 0.23 \\ 0.055 & 1 & 0.04 \end{bmatrix} \quad \mathbf{H}_3 = \begin{bmatrix} 0.11 & 0.96 & 0.25 \end{bmatrix}$$

In this case, there are too many row vectors to select; the matrix which aggregate all possible row vector choices after the re-paramterization into the row-space intersection lattice gives

$$\tilde{\mathbf{H}} = \begin{bmatrix} -0.59 & 0.77 & 0.23 \\ 1.8 & -0.43 & 3 \\ 0.055 & 1 & 0.04 \\ 0.11 & 0.96 & 0.25 \end{bmatrix},$$

which cannot have a right-inverse as it has fewer columns than rows.

5.2.4 Recursive Selection and Polymatroid Inner Boud

For the particular example just mentioned, a resolution is to turn off one of the available transmit antennas, and to select rows recursively. Doing so yields a sub channel matrix $\tilde{H}$ which is square and invertible. Doing so is implicitly prioritizing certain messages over other messages. While the user whose antenna is turned off has access to fewer degrees-of-freedom the remaining users may benefit by now being permitted access to more degrees-of-freedom.

Consider two prioritizations as listed in the following table,

$1st$	$2nd$	$3rd$	$4th$
$\{1,2\}$	$\{1\}$	$\{3\}$	$\{2\}$
$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$

For each prioritization, we greedily assign rows for the purpose of achieving a DoF point. We assign as many rows as possible for the transmission of the message source with the highest priority, then as many rows as are left and linearly independent of the prior selection for the transmission of the source with the second highest priority, and so on.

In the first example above, the highest priority message source is that which corresponds to the message source m_{12}. There is only one option of a row vector suitable for transmission of a common message DoF d_{12}, the row vector corresponding to $\tilde{\mathbf{H}}_{12}$. Continuing as described, we then select $\tilde{\mathbf{H}}_1$ for the transmission of the message source m_1 and $\tilde{\mathbf{H}}_3$ for the message source 2. At this point, we must stop: the aggregate sub-channel matrix

$$\tilde{\mathbf{H}} = \begin{bmatrix} -0.59 & 0.77 & 0.23 \\ 1.8 & -0.43 & 3 \\ 0.11 & 0.96 & 0.25 \end{bmatrix}$$

is full-rank, and adding more vectors will not lead to a matrix factorization. Using the inverse of $\tilde{\mathbf{H}}$ as pre-coding matrix, we achieve the DoF point $(d_1, d_2, d_{12}, d_3, d_{13}, d_{23}, d_{123}) = (1, 0, 1, 1, 0, 0, 0)$.

For the second priority assignment, we place the priority of the private message source m_1 over that of the message source m_{12} it shares knowledge of with the second user. There are two vectors suitable for the transmission of m_1: both $\tilde{\mathbf{H}}_{12}$ and $\tilde{\mathbf{H}}_1$: though the symbol that will be sent along $\tilde{\mathbf{H}}_{12}$ will be receivable at the second recover, the second receiver will simply ignore that output (effectively turning of the antenna there). The next priority assignment is to pick any vector suitable for the transmission of m_2. There is one left, and it is $\tilde{\mathbf{H}}_2$. At this point there are no more row vectors that are linearly independent of the previously chosen vectors and we must stop. The aggregate sub channel matrix is then

$$\tilde{\mathbf{H}} = \begin{bmatrix} -0.59 & 0.77 & 0.23 \\ 1.8 & -0.43 & 3 \\ 0.055 & 1 & 0.04 \end{bmatrix}$$

Using the inverse of $\tilde{\mathbf{H}}$ as pre-coding matrix, we achieve the DoF point $(d_1, d_2, d_{12}, d_3, d_{13}, d_{23}, d_{123}) = (2, 1, 0, 0, 0, 0, 0)$.

The K-user extension of this idea is to prioritize the elements of E in descending preference. Enumerate the elements of E to reflect this, so that the the highest priority message is the one corresponding to S_1, the second highest priority message is the one corresponding to S_2, and so on. Then, at the ith step, greedily choose as many row vectors from R_{S_i} as possible that are linearly independent from the already chosen vectors; i.e. they are linearly independent of the subspace $R_{S_1} \oplus \cdots \oplus R_{S_M}$. The number of rows selected at each step is then given by

$$\begin{aligned} d_{S_1} &= \dim(d_{S_1}) \\ d_{S_2} &= \dim(d_{S_1} \oplus d_{S_2}) - \dim(d_{S_1}) \\ &\vdots \\ d_{S_i} &= \dim(d_{S_i} \oplus \cdots \oplus d_{S_1}) - \dim(d_{S_{i-1}} \oplus d_{S_i} \cdots \oplus d_{S_1}) \end{aligned}$$

This algorithm is correct, by the notion of the augmentation property of matroid theory. Let $\mathbf{B}_{1:i-1}$ be the previously selected rows, which has rank $\text{rank}(\mathbf{B}_{1:i-1}) = \dim(R_{S_{i-1}} \oplus R_{S_i} \cdots \oplus R_{S_1})$.

Now let, $\mathbf{A}_i$ have rows which are a basis for d_{S_i} so that

$$\text{rank}\left(\begin{bmatrix}\mathbf{B}_{1:i-1}\\ \mathbf{A}_i\end{bmatrix}\right)) = \dim(R_{S_i} \oplus R_{S_i} \cdots \oplus R_{S_1})$$

Then, there are d_{S_i} rows in $\mathbf{A}_i$, assembled in $\mathbf{R}_i$, such that

$$\text{rank}\left(\begin{bmatrix}\mathbf{B}_{1:i-1}\\ \mathbf{R}_i\end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix}\mathbf{B}_{1:i-1}\\ \mathbf{A}_i\end{bmatrix}\right)$$

By the development of the polymatroid theory, we recognize the convex hull of all such achievable points as a polymatroid, given by

$$\mathcal{P}(\rho) = \left\{ d \in \mathbb{R}_+^E : \sum_{S \in B} d_S \leq \rho(B) = \dim\left(\bigoplus_{S \in B} d_S\right) \qquad \forall B \subseteq E \right\},$$

By the up-set lattice structure of the row-space intersection lattice, where $R_S \subset R_{S'}$ if $S' \subset S$, only those bounds corresponding to up-sets under the inclusion order are needed. Thus, we can state the above more simply as follows

Lemma 5.2.1 (Polymatroid DoF Inner Bound)**.** *Any Dof tuple in*

$$\mathcal{P}(\rho) = \left\{ d \in \mathbb{R}_+^E : \sum_{S \in B} d_S \leq \rho(B) \qquad \forall B \in \mathcal{F}_\uparrow \right\}. \tag{5.6}$$

is achievable.

5.2.5 Linear Network Coding

The inner code limits attention to those achievable schemes which assign each unique message-bearing symbol to at most one transmit direction. This, however is not optimal, as the simple example of sending the global transmit message suggests: consider the $K \times K$ (or, in our notation, $K \times \{1\}^K$) MISO BC with linearly independent channel matrix rows. Then there is no common message subspace $\mathcal{R}_{[1:K]}$, but yet we may still achieve a single DoF for the common message $m_{[1:K]}$ by replicating a single common message symbol K times, with one replication for each of the channel vectors $\mathbf{H}_j$. Another example is the three-user case $3 \times (2, 2, 2)$ MIMO BC, where there is

no common message subspace available for all three receivers ($\mathcal{R}_{\{1,2,3\}} = \{0\}$) but there are three pair-wise common message subspaces ($\mathcal{R}_S$ with $|S| = 2$), each of dimension one, and each linearly independent from the other two. Then by distributing two triple-wise common message-bearing symbols $x_1^{\{1,2,3\}}, x_2^{\{1,2,3\}}$ as

$$\begin{bmatrix} w_{\{1,2\}} \\ w_{\{1,3\}} \\ w_{\{2,3\}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1^{\{1,2,3\}} \\ x_2^{\{1,2,3\}} \end{bmatrix} \tag{5.7}$$

to produce three symbols $w_{\{1,2\}}, w_{\{1,3\}}, w_{\{2,3\}}$, to be assigned (respectively) to the row vectors $\tilde{\mathbf{h}}_{\{1,2\}}, \tilde{\mathbf{h}}_{\{1,3\}}, \tilde{\mathbf{h}}_{\{2,3\}}$ which (again, respectively) define the subspaces $\mathcal{R}_{\{1,2\}}, \mathcal{R}_{\{1,3\}}, \mathcal{R}_{\{2,3\}}$, we can achieve two DoF's for the triple-wise common message $m_{\{1,2,3\}}$ by using three transmit dimensions.

With linear network coding, this idea of using more than one transmit direction per common message DoF can be generalized. More precisely, the type of data mixing present in (5.7) can be generalized through consideration of random matrices: any 3×2 matrix where each of its 2×2 sub matrices is full rank would suffice, and matrices with entries generated i.i.d. from a continuous distribution satisfy this requirement with probability one. This is the analysis of [58] for a finite-field setting, and mimicking their analysis provides the following.

Lemma 5.2.2. *Any DoF tuple $(d_S : S \in E)$ satisfying the **covering** constraints*

$$d_S \leq \sum_{S' : j \in S'} d_{S \to S'} \quad \text{for all } j \in S \text{ and } S \in E, \tag{5.8}$$

*the **packing** constraints*

$$\sum_{S' \in B} \sum_{S : S \cap S' \neq \emptyset} d_{S \to S'} \leq \rho(B) \qquad \forall B \in \mathcal{F}_{\uparrow}, \tag{5.9}$$

and the non-negativity constraints

$$d_{S \to S'} \geq 0 \text{ for all } S, S' \in E, S \cap S' \neq \emptyset \tag{5.10}$$

is achievable.

Proof. Suppose we are given a set of non-negative rate-splits satsifying the covering and packing

constraints (5.8),(5.9). Define

$$\sum\nolimits_{S:S\cap S'\neq\emptyset} d_{S\to S'} = \phi_{S'} \qquad \forall S' \in E \tag{5.11}$$

and let $\{S_1, \ldots, S_M\}$ be an enumeration of the message index set E. Then by Lemma 5.2.1, we can send the set of message-bearing symbols

$$\{u_1^{S_1}, \ldots, u_{n\phi_{S_1}}^{S_1}, \ldots, u_1^{S_M}, \ldots, u_{n\phi_{S_M}}^{S_M}\}$$

over n channel uses such that each symbol u_i^S is recoverable at all receivers listed in S. To send a different set of message-bearing symbols

$$\{v_1^{S_1}, \ldots, v_{nd_{S_1}}^{S_1}, \ldots, v_1^{S_M}, \ldots, v_{nd_{S_M}}^{S_M}\},$$

where v_i^S is intended to be recovered at each receiver listed in S, but where the tuple $(d_S : S \in E)$ is outside of the polymatroid (5.6), we cannot directly apply Lemma 5.2.1. Rather we distribute information via linear combinations among a new set of symbols which we can recover at the receivers and which have sufficient information to resolve the original symbols.

For each ordered pair (S, S') in the message index set E with non-empty intersection, randomly select a set a matrix $\mathbf{F}_{S\to S'} \in \mathbb{C}^{nd_S \times nd_{S\to S'}}$ by picking its entries independently. For each pair $(S, S') \in E$ with nonempty intersection, linearly combine the set of symbols $\{v_i^S : i \in [1 : nd_S]\}$ into an intermediate set of symbols with

$$\begin{bmatrix} \tilde{u}_1^{S\to S'} \\ \vdots \\ \tilde{u}_{nd_{S\to S'}}^{S\to S'} \end{bmatrix} = \mathbf{F}_{S\to S'} \begin{bmatrix} v_1^S \\ \vdots \\ v_{nd_S}^S \end{bmatrix}.$$

Those new symbols contain some, but not necessarily all, information about the original symbols. For each set $S' \in E$, assemble these new symbols as

$$\{\tilde{u}_i^{S\to S'} : i \in [1 : nd_{S\to S'}], S \cap S' \neq \emptyset\}.$$

Then, by the packing constraint (5.11) and Lemma 5.2.1, we can recover all the symbols $\{\tilde{u}_i^{S\to S'} i \in [1 : nd_{S\to S'}], S \cap S' \neq \emptyset\}$ at each receiver listed in S'. Focus on a single such receiver, $j \in S'$. While

receiver j is aware of all symbols of the form $u_i^{A\to B}$ for some pair $(A,B)\in E$ having $j\in A\cap B$, focus on only those originating from the source S. This select set of symbols are related to the original set of symbols v_i^S by the linear relation

$$\begin{bmatrix}\mathbf{u}_{S\to S_1''}\\ \vdots \\ \mathbf{u}_{S\to S_m''}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{F}_{S\to S_1''}\\ \vdots \\ \mathbf{F}_{S\to S_m''}\end{bmatrix}}_{\mathbf{F}} \begin{bmatrix}v_1^S\\ \vdots \\ v_{nd_S}^S\end{bmatrix} \qquad \mathbf{u}_{S\to S''} = \begin{bmatrix}u_1^{S\to S''}\\ \vdots \\ u_{nd_{S\to S''}}^{S\to S''}\end{bmatrix},$$

where $\{S_1'',\dots,S_m''\}$ is an enumeration of the elements in $\{S'' : j\in S\cap S''\}$. By the packing constraint, the $n\sum_{i=1}^m d_{S\to S_i''}\times nd_S$ matrix $\mathbf{F}$ has more rows than columns. Moreover, as its elements were drawn independently and identically, it is has full column rank. Thus each of the nd_S symbols $\{v_i^S : i\in[1:nd_S]\}$ are recoverable at this receiver. Repeating the argument for all other message indices $S\in E$ with $j\in S$, and all other receivers, yields the desired result. □

5.2.6 Optimality

When the message sources are ordered in a non-increasing manner according to the inclusion order, and when the recursive selection procedure eventually selects all vectors, the above inner bound is optimal. In this case, the procedure yields a matrix factorization, the bound $\rho(B)$ is modular, so that it may be expressed as $\rho(B)=\sum_{S\in B}\phi_S$. To prove the outer bound, we re-write the original channel as degraded combination network. In this case, tools from the combination network can be directly applied. In particular, the notion of generalized cut-set bounds are generally needed. Doing so, yields the following results, the details of which we provide later. We refer to all channels for which the recursive procedure picks the inverse of entire channel matrix, with its row spaces re-parametrized according to the row-space intersection lattice, as bound-modular BCs.

Theorem 5.2.3 (Modular Matrix Dimensions - Three Users)**.** *Suppose the three-user BC is bound-*

modular. Then the three user DoF region is given by

$$\mathbf{E}_3 \begin{bmatrix} d_{\{1\}} \\ d_{\{2\}} \\ d_{\{3\}} \\ d_{\{1,2\}} \\ d_{\{1,3\}} \\ d_{\{2,3\}} \\ d_{\{1,2,3\}} \end{bmatrix} \leq \mathbf{E}_3 \begin{bmatrix} \phi_{\{1\}} \\ \phi_{\{2\}} \\ \phi_{\{3\}} \\ \phi_{\{1,2\}} \\ \phi_{\{1,3\}} \\ \phi_{\{2,3\}} \\ \phi_{\{1,2,3\}} \end{bmatrix} \qquad \mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 2 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 3 \\ 2 & 1 & 2 & 2 & 2 & 2 & 3 \\ 2 & 2 & 1 & 2 & 2 & 2 & 3 \\ 2 & 2 & 2 & 2 & 2 & 2 & 3 \end{bmatrix}. \tag{5.12}$$

Proof. Achievability follows by the inner bound given and the converse follows by Theorem 5.2.6 along with the extremal inequalities given in Proposition A.0.1. □

Theorem 5.2.4 (Modular Matrix Dimensions - Symmetric K Users)**.** *Suppose that the K-user BC is symmetric and so that $\phi_S = \phi_{|S|}$ for every $S \in E$. If we enforce that $d_S = d_{|S|}$, then the DoF region is given by*

$$\sum_{r=1}^{K} \beta_Q(r) \sum_{S:|S|\geq r} d_S \leq \sum_{r=1}^{K} \beta_Q(r) \sum_{S:|S|\geq r} \phi_S \tag{5.13}$$

for all subsets $Q \subseteq [1:K] \setminus \{1\}$ where

$$\beta'_Q(r) = \prod_{q\in Q} (q - [\![q < r]\!]) \qquad \beta_Q(r) = \begin{cases} \beta'_Q(r) & \text{if } r \notin Q \\ 0 & \text{if } r \in Q \end{cases} \tag{5.14}$$

for any $r \in [1:K]$ and the convention taken is that the product over an empty set is one.

Proof. Achievability by the previous section, while the converse follows by Theorem 5.2.6 and the extreme inequalities given in Proposition A.0.2. □

Corollary 5.2.5 (Mostly too many receive antennas)**.** *Suppose that $r_j = K-1$ for all $j \in [1:K]$ and $t = K$. Then the symmetric DoF, where the constraint that $d_S = d_{|S|}$ is enforced, is given by the region in Theorem 5.2.4 with $\phi_S = [\![|S| = K-1]\!]$.*

Proof. By assumption of genericity, each $\mathcal{R}_S = \cap_{j \in S} \text{Null}(\mathbf{H}_j)^{\perp}$ with $|S| = K-1$ is a one-dimensional line within the K-dimensional transmit space. Moreover, each of the K such lines are linearly independent and at each receiver, $\mathbf{H}_j$ can be left multiplied so that each of its $K-1$ rows corresponds to one of the $K-1$ lines $\{\mathcal{R}_S : j \in S \in E\}$. Thus this situation leads to a unique matrix factorization and a modular bound on the achievable rate region - precisely the setting where the outer bound of Theorem 5.2.4 matches the inner bound developed in this paper. □

5.2.6.1 Degraded Combination BC

Assume that the channel is bound-modular which implies that $\rho(B) = \dim\left(\bigoplus_{S \in B} R_S\right) = \sum_{S \in B} \phi_S$ for some set of non-negative integers $(\phi_S : S \in E)$ and matrices $\tilde{H}_S$, with ϕ_S linearly independent rows within R_S that are linearly independent of any elements in $\bigoplus_{S' : S \subset S'} R_{S'}$. Specializing this to the the case where B indexes the elements of the intersection lattice of row spaces available to the jth receiver gives

$$\text{rank}(\mathbf{H}_j) = \rho(\{S : j \in S \subseteq [1:K]\}) = \sum_{S : j \in S \subseteq [1:K]} \phi_S.$$

Let $\mathcal{S}_j = \{S : j \in S \subseteq [1:L], \phi_S > 0\}$. As the recursive row selection procedure eventually selects all row vectors, there is a re-parameterization of each of the receiver channel matrices into block rows $\tilde{\mathbf{H}}_S$ which contain rows only in R_S. For this jth receiver, this is accomplished by applying a non-singular $r_j \times r_j$ matrix $\mathbf{G}_j$ to the jth receiver's output, such that

$$\mathbf{G}_j = \begin{bmatrix} \mathbf{G}_{j,S_1} \\ \vdots \\ \mathbf{G}_{j,S_n} \end{bmatrix} \qquad \mathbf{G}_{j,S} \in \mathbb{C}^{\phi_S \times r_j} \qquad \mathbf{G}_{j,S}\mathbf{H}_j = \tilde{\mathbf{H}}_S,$$

where $\{S_1, \ldots, S_n\}$ is an enumeration of the elements in $\mathcal{S}_j$.

We will define a collection of independent enhanced channels $\{\tilde{Y}_S : S \in E\}$, such that the channel outputs $\mathbf{Y}_j$ are degraded versions of these enhanced channels, with a Markov structure as in the combination network. A depiction of these enhanced channels and their Markov relation to the actually channel outputs is depicted in Figure 5.5 for the three-uer BC.

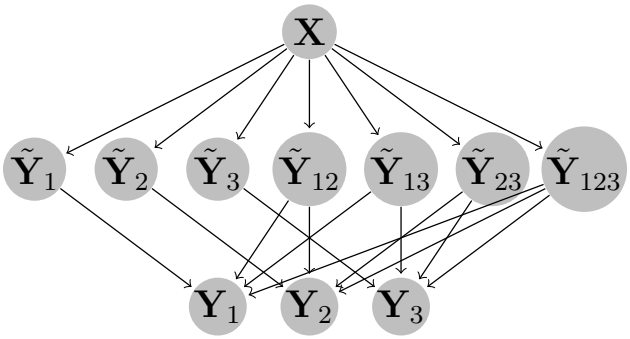


Figure 5.5: Combination Markov structure for Bound-Modular Broadcast Channels.

As $\mathbf{G}_j$ is non-singular, $\mathbf{G}_j\mathbf{G}_j^*$ is positive definite, with minimum eigenvalue $\lambda_j > 0$. Let $\lambda_{min} = \min_j \lambda_j$ so that $\mathbf{G}_j\mathbf{G}_j^* \succeq \lambda_{min}\mathbf{I} \succ \mathbf{0}$ for each j. Define outputs

$$\tilde{\mathbf{Y}}_j = \mathbf{G}_j\mathbf{H}_j\mathbf{X} + \tilde{\mathbf{N}}, \qquad \tilde{\mathbf{N}} \sim \mathcal{CN}(\mathbf{0}, \lambda_{min}\mathbf{I}),$$

which we may write the original channel outputs $\mathbf{Y}_j$ in terms of,

$$\mathbf{Y}_j = \mathbf{G}_j^{-1}(\tilde{\mathbf{Y}}_j + \tilde{\mathbf{Z}}) \qquad \tilde{\mathbf{Z}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{G}_j\mathbf{G}_j^* - \lambda_{min}\mathbf{I}).$$

Further decompose each $\tilde{\mathbf{Y}}_j$ into components $\tilde{\mathbf{Y}}_{j,S}$ indexed by those subsets $S \subseteq [1:K]$ with $\phi_S > 0$. For each $j \in S$, we have $\tilde{\mathbf{Y}}_{j,S} = \tilde{\mathbf{H}}_S\mathbf{X} + \mathbf{N}_{j,S}$ where $\mathbf{N}_{j,S} \sim \mathcal{CN}(\mathbf{0}, \lambda_{min}\mathbf{I}_{\phi_S \times \phi_S})$. Then the enhanced channel outputs of interest will be $\tilde{\mathbf{Y}}_S$, which has as block rows $(\tilde{\mathbf{Y}}_{j,S} : j \in S)$:

$$\tilde{\mathbf{Y}}_S = \begin{bmatrix} \tilde{\mathbf{Y}}_{j_1,S} \\ \vdots \\ \tilde{\mathbf{Y}}_{j_S,S} \end{bmatrix} \qquad \{j_1, \ldots, j_S\} = S.$$

Define an intermediate set of channel outputs $\mathbf{V}_j$, which have as block outputs $(\tilde{\mathbf{Y}}_S : j \in S, \phi_S > 0)$. Then as the original outputs $\mathbf{Y}_j$ are degraded versions of $\tilde{\mathbf{Y}}_j$, and these are degraded versions of

$\mathbf{V}_j$, which are combinations of the enhanced outputs $\mathbf{Y}_S$, we have re-written the network as a degraded combination network, with Markov structure as depicted in Figure 5.5.

To see that the prior framework leads to a matrix factorization, let $\tilde{\mathbf{H}}$ be the matrix with block rows $\{\mathbf{H}_S : S \in E\}$. This matrix contains all of the selected row vectors, and so by assumption, it represents the entire channel matrix without any antennas turned off. Parse the inverse $\tilde{\mathbf{H}}^{-1}$ so that the columns which correspond to the rows $\tilde{\mathbf{H}}_S$ are denoted by the $t \times \phi_S$ matrix $\mathbf{T}_S$. Then, as $\tilde{\mathbf{H}}$ is an inverse, we have that for each tuple (j, S, S') with $j \in S \subseteq [1:K]$ and $S' \subseteq [1:K]$,

$$\mathbf{G}_{j,S}\mathbf{H}_j\mathbf{T}_{S'} = \delta_{j,S,S'} \triangleq \begin{cases} \mathbf{I}_{\phi_S \times \phi_S} & \text{if } S = S' \\ \mathbf{0}_{\phi_S \times \phi_{S'}} & \text{else} \end{cases} \tag{5.15}$$

and the concatenated matrix $\mathbf{F} = [\mathbf{T}_S : S \subseteq [1:K]]$ is a full rank $t \times t$ matrix. In this case,

$$\begin{bmatrix} \mathbf{G}_1 & & \\ & \ddots & \\ & & \mathbf{G}_K \end{bmatrix} \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \mathbf{F} = \begin{bmatrix} \delta_{1,S_1} & \cdots & \delta_{1,S_M} \\ \vdots & \ddots & \vdots \\ \delta_{K,S_1} & \cdots & \delta_{K,S_M} \end{bmatrix}$$

where $\delta_{j,S}$ is non-zero only if $j \in S$, in which case with the enumeration $\{S_1^j, \ldots, S_N^j\} = \{S : j \in S \in E\}$,

$$\delta_{j,S} = \begin{bmatrix} \delta_{j,S,S_1^j} \\ \vdots \\ \delta_{j,S,S_N^j} \end{bmatrix}.$$

By enumerating the non-empty subsets according to their "flipped" binary representation $\chi(S)$, we have that the resultant matrix $[\delta_j, S_i]$ has a block-triangular structure. For example, with the

three-user case we have (5.16).

$$\begin{bmatrix} & & \\ & & \\ \delta_{1,S_1} & \cdots & \delta_{1,S_7} \\ & & \\ \hline & & \\ & & \\ \delta_{2,S_1} & \cdots & \delta_{2,S_7} \\ & & \\ \hline & & \\ & & \\ \delta_{3,S_1} & \cdots & \delta_{3,S_7} \\ & & \end{bmatrix} = \begin{array}{c} \\ \phi_{100} \\ \phi_{110} \\ \phi_{101} \\ \phi_{111} \\ \hline \phi_{010} \\ \phi_{110} \\ \phi_{011} \\ \phi_{111} \\ \hline \phi_{001} \\ \phi_{101} \\ \phi_{011} \\ \phi_{111} \end{array} \begin{array}{c} \begin{array}{cccccccc} \phi_{000} & \phi_{100} & \phi_{010} & \phi_{110} & \phi_{001} & \phi_{101} & \phi_{011} & \phi_{111} \end{array} \\ \begin{bmatrix} & I & & & & & & \\ & & & I & & & & \\ & & & & & I & & \\ & & & & & & & I \\ \hline & & I & & & & & \\ & & & I & & & & \\ & & & & & & I & \\ & & & & & & & I \\ \hline & & & & I & & & \\ & & & & & I & & \\ & & & & & & I & \\ & & & & & & & I \end{bmatrix} \end{array}, \tag{5.16}$$

where the row and column labels indicated the dimensions. Assign $\mathbf{T}_S^+$ to be the left inverse of $\mathbf{T}_S$ and $\mathbf{R}_{j,S}^+$ to be the right inverse of $\mathbf{R}_{j,S}$ when $j \in S$, and otherwise the zero matrix of dimensions $r_j \times \phi_S$. Then by the above we have that

$$\begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{1,S_1}^+ & \ldots & \mathbf{R}_{1,S_M}^+ \\ & & \\ \mathbf{R}_{K,S}^+ & \ldots & \mathbf{R}_{K,S_M}^+ \end{bmatrix} \begin{bmatrix} \mathbf{T}_{S_1}^+ \\ \vdots \\ \mathbf{T}_{S_M}^+ \end{bmatrix},$$

a matrix factorization of the joint channel matrix $\mathbf{H}$.

5.2.7 Generalized DoF Cut-Set Bounds

To the degraded combination BC, we can apply converse arguments which were applied to the noiseless combination networks. They involve extremal inequalities of general submodular functions as in [91]. We borrow much of the framework of [91], and reproduce it below, modifying

their central theorem from a noiseless network coding setting to our DoF MIMO setting. The inequalities which will govern our DoF regions will all be of the form

$$\sum_{i\in\mathcal{I}}\alpha_i \sum_{S\in\Phi_i(\uparrow\{1\},\ldots,\uparrow\{K\})} d_S \leq \sum_{i\in\mathcal{I}}\alpha_i \sum_{S\in\Phi_i(\uparrow\{1\},\ldots,\uparrow\{K\})} \phi_S \tag{5.17}$$

for some non-empty finite set $\mathcal{I}$, a collection of nonnegative reals $(\alpha_i : i \in \mathcal{I})$, and a collection of set operators $(\Phi_i : i \in \mathcal{I})$. Each unique set operator will consist of a pre-defined sequence of unions and intersections to take place on its arguments. These bounds are **generalized** cut-set bounds in that the set operators Φ_i may include intersections, rather than only including unions, as per the typical cut-set bounds of prior literature.

Consider a septuple

$$(\alpha_i : i \in \mathcal{I}), (\beta_j : j \in \mathcal{J}), (\gamma_l : l \in \mathcal{L}), (\Pi_j : j \in \mathcal{J}), (\Gamma_l^+ : l \in \mathcal{L}), (\Gamma_l^- : l \in \mathcal{L}) \tag{5.18}$$

where the first three tuples are collections on nonnegative reals, and the latter three collections are those of set operators. Borrowing the language of [91], this septuple identifies an extremal inequality for submodular functions if

$$\begin{aligned}\sum_{i\in\mathcal{I}}\alpha_i f(\Phi_i(S_1,\ldots,S_K)) \leq \sum_{j\in\mathcal{J}}\beta_j f(\Pi_i(S_1,\ldots,S_K)) \\ + \sum_{l\in\mathcal{L}}\gamma_l\Big(f(\Gamma_i^+(S_1,\ldots,S_K)) - f(\Gamma_i^-(S_1,\ldots,S_K))\Big)\end{aligned} \tag{5.19}$$

holds for any K subsets $(S_k : k \in [1:K])$ of E and any submodular function f over the subsets of E, and holds with equality for any K subsets $(S_k : k \in [1:K])$ of E and any modular function f over subsets of E. For the outer bound, the following theorem mimics Theorem 1 of [91] to the present Broadcast Channel DoF setting.

Theorem 5.2.6. *Let $\mathcal{I}$ be a non-empty finite set, $(\alpha_i : i \in \mathcal{I})$ be a collection of nonnegative reals, and $(\Phi_i : i \in \mathcal{I})$ be a collection of set operators. Then the generalized cut-set bound* (5.17) *holds if there exist*

(1) two nonempty finite sets $\mathcal{J}$ and $\mathcal{L}$;

(2) nonnegative reals $(\beta_j : j \in \mathcal{J})$ *and* $(\gamma_l : l \in \mathcal{L})$*;*

(3) set operators $(\Pi_j : j \in \mathcal{J}), (\Gamma_l^+ : l \in \mathcal{L})$*, and* $(\Gamma_l^- : l \in \mathcal{L})$*;*

(4) a set of numbers $(\phi_S : S \in E)$ *and non-singular matrices* $(\mathbf{G}_j : j \in [1 : K])$*;*

such that:

(1) the septuple (5.18) *identifies an external inequality for submodular functions;*

(2) $(\Pi_j : j \in \mathcal{J})$ *and* $(\Gamma_l^+ : l \in \mathcal{L})$ *are collections of subset unions; and*

(3) for any $l \in \mathcal{L}$ *and any* K *subsets* $(S_k : k \in [1 : K])$ *of* E*,* $\Gamma_l^+(S_1, \ldots, S_k) \supseteq \Gamma_l^+(S_1, \ldots, S_K)$*.*

Proof. Provided in the Appendix. □

5.3 General Bound

While the above bound works for special cases of the MIMO DoF setting, there are several cases where it does not work. To for To guide the development of a more general DoF inner bound, which overcomes the shortcomings of the prior inner bound, we propose a general inner bound for the discrete memoryless channel.

This inner bound contains several key elements

- Superposition Coding
- Up-set Rate-splittting
- Binning

The bounded is given only implicitly, with extra variables which need be projected away to provide the resultant achievable rate region. Using some of just the superposition coding and up-set rate-splitting improves on the prior inner bound for the DoF region and reproduces know results for the combination network.

Recall that E refers to the message index set, containing subsets of the set of users, which indicate which users desire the indexed message source. And, as in the MAC, let $\leq$ be a **superposition** order; that is, $S < S'$ only if $S \subset S'$.

5.3.1 Superposition Coding

Let $(R_S : S \in E)$ be a non-negative rate-tuple and let $(U_S : S \in E)$ be an auxiliary random tuple that factors recursively according to

$$p(U_E) = \prod_{S \in E} p(U_S | U_{\uparrow S \setminus S}).$$

Let the input $X \leftarrow (U_S : S \in E)$ be a deterministic function of the auxiliary random tuple. Consider recursively generating codewords as described in the discrete memoryless MAC Chapter. Specifically, for each $S \in E$, and for each $m_{\uparrow S} \in \prod_{S' \in \uparrow S} \{1, \ldots, 2^{nR_{S'}}\}$, pick a random codeword

$$u_S^n(m_{\uparrow S}) = \prod_{t=1}^{N} p_{U_S | U_{\uparrow S \setminus S}}(u_{St} | U_{\uparrow S \setminus S, t})$$

Define $\mathcal{S}_j = \{S : j \in S \in E\}$ as the set of desired messages for receiver j. By the same argument as used for the DM MAC, joint decoding at each receiver is successful if

$$\sum_{S \in B} R_S \leq I(U_B; Y_j | U_{\mathcal{S}_j \setminus B}) \tag{5.20}$$

for each down-set of $\mathcal{S}_j$. As previously noted—each is of these is a polymatroid with support $\mathcal{S}_j$. The set of all such conditions, then, are an intersection of polymatroids with overlapping, but not identical, supports.

5.3.2 Up-set rate-splitting

Define target rates $(T_S : S \in E)$ and reconstructed $(R_S : S \in E)$ as follows

Target Rates	Reconstructed Rates
$T_S = \sum_{S' \in \uparrow S} r_{S \to S'}$	$R_{S'} = \sum_{S \in \downarrow S'} r_{S \to S'}$

If the reconstructed (R_S) satisfy the conditions (5.20) at all receivers, then the target rates are achievable. This enlarges the region s we may alway just choose $r_{S \to S'}$ non-zero only if $S' = S$, in which case we choose it to be R_S: $r_{S \to S'} = [\![S' = S]\!] R_S$.

5.3.3 Binning

A last element is to consider binning, which allows consideration of **arbitrary** input distributions. The central idea is create an excessively large codebook, with rates $\tilde{R}_S \geq R_S$ where each message has a list of codewords of exponential size $2^{n(\tilde{R}_S - R_S)}$, rather than a single codeword. If the rate excesses $\tilde{R}_S - R_S$ are sufficiently large, then every message can jointly select a set of codewords that appear as though they were jointly generated with respect to an arbitrary joint distribution, rather than according to its recursive marginal distributions.

In particular, the excess rates will be $\tilde{R}_S$, with the excess over the desire rate being $r_S = \tilde{R}_S - R_S$ for each $S \in E$. The key result is a recursive version of the Mutual Covering of El Gamal and Kim's text[31].

Lemma 5.3.1 (Recursive Mutual Covering Lemma)**.** *Let $(U_S : S \in E)$ have arbitrary joint distribution $p(u_S : S \in E)$. Pick an order on E. With respect to this order, recursively generate length-n vectors*

$$u_S^n(m_S) \sim \prod_{t=1}^{n} p(u_{St} | u_{Rt} : R \in \uparrow' S)$$

for each $m_S \in [1 : 2^{nr_S}]$ and each $S \in E$. Then the probability that $(u_S(m_S) : S \in E)$ is jointly ϵ-typical for some $(m_S : S \in E)$ tends to one as $n \to \infty$ if the non-negative rates $(r_S : S \in E)$ satisfy

$$r(G) \geq \sum_{S \in G} H(U_S | U_{\uparrow' S}) - H(U_G) \triangleq \gamma(G), \tag{5.21}$$

for all up-sets $G \subseteq E$.

Proof. See Appendix D.3. □

This unbounded polyhedron is a contra-polymatroid, defined only over the up-set lattice $\mathcal{F}_\uparrow$. This follows as

- $\gamma(G)$ is supermodular: $\gamma(F \cap G) + \gamma(F \cup G) \geq \gamma(F) + \gamma(G)$, a consequence of the submodularity of entropy.

- $\gamma(G)$ is non-increasing, a consequence of the fact that condition reduces entropy: for $F \subseteq G$,

$$\begin{aligned}\gamma(G) - \gamma(F) &= \left(\sum_{S \in G \setminus F} H(U_S|U_{\uparrow' S})\right) - H(U_{G\setminus F}|U_F) \\ &\geq \sum_{S \in G\setminus F} \Big(H(U_S|U_{\uparrow' S}) - H(U_S|U_F)\Big) \\ &\geq \sum_{S \in G\setminus F} \Big(H(U_S|U_F) - H(U_S|U_F)\Big) \\ &= 0\end{aligned}$$

- $\gamma(G)$ is normalized: $\gamma(\emptyset) = 0$ as the sum is vacuous.

This new mutual covering lemma provides a basis for a Marton-type achievable scheme for the K-user Broadcast Channel with common message.

Theorem 5.3.2 (Marton Extension)**.** *For a K-user DM-BC with general message sets, any non-negative rate tuple satisfying the covering constraints*

$$T_S = \sum_{S' \in \uparrow S} r_{S \to S'} \qquad \forall S \in E,$$

the packing constraints,

$$R_{S'} = \sum_{S \in \downarrow S'} r_{S \to S'} \qquad \forall S' \in E,$$

the joint decoding constraints

$$\sum\nolimits_{S \in B} \tilde{R}_S \leq I(U_B; Y_j|U_{\mathcal{S}_j \setminus B}) \tag{5.22}$$

for every down-set within $\mathcal{S}_j$ and every $j \in [1:K]$, and the binning constraints

$$\sum\nolimits_{S \in G} (\tilde{R}_S - R_S) \leq \rho_R(G) \tag{5.23}$$

for every up-set $G \in \mathcal{F}_\uparrow(E; \subseteq)$.

Proof. Follows by the previous subsections. □

5.4 Combination Network

The prior bound is optimal in select cases. To show one application, consider the three-user combination network, where the input consists of $|E|$ parts, where $E = 2^{[1:K]} \backslash \emptyset$ indexes these parts. Let V_S be the Sth part; it can take values in an alphabet $\mathcal{V}_S$ with cardinality C_S. There are K outputs, where the Kth output is the collection of the input parts $(V_S : j \in S \in E)$.

Apply Marton's extension as described in Theorem 5.3.2 with a specific auxiliary choice, with $U_S \sim \text{Uniform}(C_S)$ chosen independently, and with the input V_S set to U_S. As the auxiliary random variables are independent, the binning constraints are vacuous and we may simply eliminate binning from the description of the achievable rate region. The inner bound thus reduces to the set of rate tuples $(R_S : S \in E)$ for which as set of non-negative rate-splits $(r_{S \to S'} : S \subset S')$ exits such that

$$\sum_{S \in B} \tilde{R}_S \leq \sum_{S \in B} C_S \qquad \forall B \in \mathcal{F}_{\downarrow}(\mathcal{S}_j; \leq) \qquad \text{and} \qquad \begin{aligned} \tilde{R}_{S'} &= \sum_{S \in \downarrow S'} r_{S \to S'} \\ R_S &= \sum_{S' \in \uparrow S} r_{S \to S'} \end{aligned} .$$

For small networks, we can project away the rate-splits through Fourier-Motzkin. This yields that any non-negative rate within

$$\mathbf{E}_3 \begin{bmatrix} R_{\{1\}} \\ R_{\{2\}} \\ R_{\{3\}} \\ R_{\{1,2\}} \\ R_{\{1,3\}} \\ R_{\{2,3\}} \\ R_{\{1,2,3\}} \end{bmatrix} \leq \mathbf{E}_3 \begin{bmatrix} C_{\{1\}} \\ C_{\{2\}} \\ C_{\{3\}} \\ C_{\{1,2\}} \\ C_{\{1,3\}} \\ C_{\{2,3\}} \\ C_{\{1,2,3\}} \end{bmatrix} \tag{5.24}$$

where $\mathbf{E}_3$ is as defined in (5.12), is achievable. Through the standard cut-set arguments as well as generalized cut-set arguments, this inner bound is optimal [42, 91]. In contrast to the work of Grokop and Tse [42], this inner bound was shown without linear network coding, but with superposition encoding, joint decoding, and simple up-set rate-splitting.

Chapter 6

Semi-Deterministic Inteference Channel with Common Information

6.1 Introduction

Finding the capacity region of general multi-terminal networks is an elusive goal: even the simplest of networks evade precise capacity characterizations. Recent developments, however, illustrate that searching for capacity **approximations** is far more fruitful. Here, we extend an approximate capacity result for the two-user IC from a setting with only private messages to a setting with both private and common messages.

The canonical IC dates back to Shannon [96], and is an example of a simple network for which the capacity region is yet to be found. The best known inner bound is the Han-Kobayashi scheme (HK), which was originally described in 1981 [47] and more compactly described in 2008 by Chong et al. [17]. To date, attempts to strictly improve or to show optimality have fallen short in the general DM setting. In some special cases, progress has been made. For ICs with strong interference [18], for discrete additive degraded ICs [6], and for a class of deterministic ICs [30], the HK scheme is known to be optimal. For a class of semi-deterministic ICs, the HK scheme is within a quantifiable gap of optimality [103], of which a subclass are the Gaussian ICs, where a single HK strategy suffices to achieve optimality to within a constant gap [33, 57].

Much less work has been done for the IC with common information (ICC). This model consists of the same physical channel as the IC, but expands the set of possible sources to include those that are correlated by a common part. While we do not study sources with arbitrary correlations, which are more difficult to treat[19], sources with common messages can model scenarios with

transmitter cooperation and/or cognition. For example, in the multiple access channel, the capacity with common information [99] can be directly applied to determine the capacity with conferencing encoders [10, 118, 119]. This connection, which suggests that the distinction between models with an implicit encoder conference (i.e. sources with common information) and models with an explicit encoder conference, is minimal, provides hope that models with common information may help shed light on what benefits transmitter cooperation may provide in assisting communication over more general multi terminal channels.

The first results for the ICC were by Tan [101] where both inner and outer bounds are presented for the DM case. Recent results have tightened these bounds by carrying over insights developed in the larger body of work on the IC with only private information. The best inner bound was developed in Jiang-Xin-Garg [53], where a scheme motivated by the HK scheme is presented. Noteworthy aspects of this scheme are that it reduces to the HK scheme when there is no common information sent and that it is optimal for certain subclasses of ICs: a deterministic model [53] and a semi-deterministic model subject to a multi-letter strong interference condition [16]. Focusing on the scalar Gaussian setting, Vaze and Varanasi [109] find that a simple explicit representation of this scheme is within one bit of the capacity region. Moreover, they demonstrate through the generalized degrees of freedom metric that there is a potentially unbounded capacity gain to be had over transmitting with only private information.

While Tan's outer bound has not yet been tightened for all DM ICs, we show that it can be tightened for any in a certain subclass. Specifically, we consider the same semi-deterministic model of Telatar and Tse [103] (which differs from the "strong interference" semi-deterministic model of [16]), but for the case that each transmitter can send both common and private information. We show that this outer bound is within a quantifiable gap of the Jiang-Xin-Garg inner bound. Moreover, this gap result reduces to the Telatar and Tse result when there is no common information to send and predicts the constant gap to capacity region result for the Gaussian case.

6.1.1 Setting

Before describing the details of our capacity characterization, we provide short formal definitions of the channel models we consider. The general memoryless interference channel (IC) consists of two users where each user $k \in \{1,2\}$ has an input $X_k \in \mathcal{X}_k$ and output $Y_k \in \mathcal{Y}_k$ pair. The key complication is that the outputs depend on both inputs, not merely the paired input; that is, the probability transition function is

$$p(y_1^n, y_2^n | x_1^n, x_2^n) = \prod_{i=1}^n p_{Y_1,Y_2|X_1,X_2}(y_{1i}, y_{2i} | x_{2i}, x_{1i}).$$

Such a channel is discrete if the alphabets $\mathcal{Y}_k, \mathcal{X}_k$ are finite.

Regarding how we use this channel, we consider the case where each user k has both a both private and common messages to send; that is the interference channel with common information (ICC). Formally, for $k \in \{1,2\}$, transmitter k desires to send a private message $m_k \in \mathcal{W}_k = [1 : W_k]$ together with common message $m_0 \in \mathcal{W}_0 = [1 : W_0]$ to receiver k. To do so, each transmitter/receiver pair k may code over n transmissions (a block length of n) with an encoder $e_k : \mathcal{W}_0 \times \mathcal{W}_k \mapsto \mathcal{X}_k^n$ and a decoder $d_k : \mathcal{Y}_k^n \mapsto \mathcal{W}_0 \times \mathcal{W}_k$. Denoting $\hat{M}_{t,k}$ as the estimate of the message M_t at a receiver k, let $P_{e,k}^{(n)} = P((\hat{M}_{0,k}, \hat{M}_{k,k}) \neq (M_0, M_k))$ be the probability of error at receiver k. Then we say that a rate triple (R_0, R_1, R_2) is achievable if, for every $\epsilon > 0$, there exists a block length n and corresponding code such that $W_t \geq 2^{n(R_t - \epsilon)}$ for $t = 0, 1, 2$ and $P_{e,k}^{(n)} < \epsilon$.

While we use theory for the general interference channel for our inner bound, for the outer bound we focus on a structured subclass that offers both sufficient structure to be tractable and sufficient generality to be useful. The subclass we consider is the "semi-deterministic" kind proposed in [103], with an underlying deterministic structure originating in [30]. In a channel of this class, if $(k,l) \in \{(1,2),(2,1)\}$, then the output Y_k is a deterministic function of a perfect copy its paired transmitter's signal (X_k) and a noisy copy of the interfering transmitter's signal ($S_l \in \mathcal{S}_l$) obtained by passing the the interfering transmitter's signal X_l through a memoryless channel defined by $p_{S_l|X_l}$. That is,

$$Y_1 = f_1(X_1, S_2) \qquad\qquad X_2 \xrightarrow{p_{S_2|X_2}} S_2$$

$$Y_2 = f_2(X_2, S_1) \qquad\qquad X_1 \xrightarrow{p_{S_1|X_1}} S_1$$

where the f_1 and f_2 are both deterministic functions. Moreover, the maps $f_k(x_k, s_k)$ are invertible in s_k for each $x_k \in \mathcal{X}_k$. This model has two attractive features. The first, as we shall see, is that the invertiblility condition provides enough structure to declare statements analogous to $H(Y_1|X_1) = H(S_2|X_1)$, which in turn allow the development of a "relatively tight" outer bound. The second is that when the alphabets are continuous, this model encapsulates the The first, as we shall see, is that the invertiblility condition allows for the development of a relatively tight outer bound with the ability to declare statements analogous to $H(Y_1|X_1) = H(S_2|X_1)$. The second is that this model encapsulates the Gaussian interference channel, where the additive noise can be confined to the interfering links and the functions f_k are linear and invertible.

6.1.2 Background

For the general discrete memoryless interference channel without common information (i.e. $R_0 = 0$), the largest known region of achievable rate pairs (R_1, R_2) is the Han-Kobayashi region [17, 47]. Recently, Jiang-Xin-Garg [53] extended this result to the discrete memoryless interference channel with common information (i.e. $R_0 \geq 0$) and we reproduce their result here,

Theorem 6.1.1. *Consider a discrete memoryless interference channel described by $p_{Y_1,Y_2|X_1,X_2}$ and a set of random variables U_0, U_1, U_2, X_1, X_2 which satisfy the Markov condition*

$$(U_1, X_1) \multimap U_0 \multimap (U_2, X_2),$$

and where the U_k are defined over finite alphabets $\mathcal{U}_k$. Then any non-negative rate triple $(R_0, R_1, R_2) \in \mathcal{R}_{JXG}(\mathbf{B})$ is achievable where

$$\begin{aligned}
\mathcal{R}_{JXG}(\mathbf{B}) = \{(R_0, R_1, R_2) : \\
R_1 < B_3 \\
R_2 < B_3' \qquad\qquad & R_0 + R_1 + R_2 < B_1 + B_5' \\
R_0 + R_1 < B_5 \qquad\qquad & R_0 + R_1 + R_2 < B_1' + B_5
\end{aligned}$$

$$
\begin{aligned}
R_0 + R_2 &< B_5' & R_1 + 2R_2 &< B_1' + B_2 + B_4' \\
R_1 + R_2 &< B_1 + B_4' & 2R_1 + R_2 &< B_1 + B_2' + B_4 \\
R_1 + R_2 &< B_1' + B_4 & R_0 + R_1 + 2R_2 &< B_1' + B_2 + B_5' \\
R_1 + R_2 &< B_2 + B_2' & R_0 + 2R_1 + R_2 &< B_1 + B_2' + B_5\}.
\end{aligned} \tag{6.1}
$$

and the bounds $\mathbf{B} = (B_1, \ldots, B_5, B_1', \ldots, B_5')$ *are*

$$
\begin{aligned}
B_1 &= I(Y_1; X_1 | U_0, U_1, U_2) & B_1' &= I(Y_2; X_2 | U_0, U_1, U_2) \\
B_2 &= I(Y_1; X_1, U_2 | U_0, U_1) & B_2' &= I(Y_2; X_2, U_1 | U_0, U_2) \\
B_3 &= I(Y_1; X_1 | U_0, U_2) & B_3' &= I(Y_2; X_2 | U_0, U_1) \\
B_4 &= I(Y_1; X_1, U_2 | U_0) & B_4' &= I(Y_2; X_2, U_1 | U_0) \\
B_5 &= I(Y_1; X_1, U_0, U_2) & B_5' &= I(Y_2; X_2, U_0, U_1),
\end{aligned} \tag{6.2}
$$

These results continue to hold when we allow the alphabets to be continuous and the channel is well-behaved, as can be shown with a limiting argument on successively refined discretizations (Chapter 3, pgs. 50-51[31]). In particular, a Gaussian channel is one such well-behaved channel and the above theorem applies. For example, the scalar Gaussian ICC fits in the semi-deterministic mold by setting

$$f_k(X_k, S_l) = \sqrt{\mathrm{SNR}_k} X_k + S_l$$

where $S_l = \sqrt{\mathrm{INR}_l} X_l + Z_l$ ($Z_l \sim \mathcal{CN}(0,1)$) is a noisy copy of the interfering transmitter's signal X_l and the linear combinations $f_k(X_k, \cdot)$ are invertible for each X_k.

In [109], it is shown that for the scalar Gaussian ICC (with $\mathrm{SNR}_k = \Lambda_k$ and $\mathrm{INR}_l = \Gamma_l$) that the single region $\mathcal{R}_{JXG}(B_1, \ldots, B_5, B_1', \ldots, B_5')$ described by the private message power $x_l = \min(1, 1/\Gamma_l)$ at transmitter l and bounds

$$
\begin{aligned}
B_1 &= \log_2(1 + \Lambda_1 x_1 + \Gamma_2 x_2) & B_1' &= \log_2(1 + \Lambda_2 x_2 + \Gamma_1 x_1) \\
B_2 &= \log_2(1 + \Lambda_1 x_1 + \Gamma_2) & B_2' &= \log_2(1 + \Lambda_2 x_2 + \Gamma_1) \\
B_3 &= \log_2(1 + \Lambda_1 + \Gamma_2 x_2) & B_3' &= \log_2(1 + \Lambda_2 + \Gamma_1 x_1)
\end{aligned}
$$

$$B_4 = \log_2(1 + \Lambda_1 + \Gamma_2) \qquad B'_4 = \log_2(1 + \Lambda_2 + \Gamma_1)$$

$$B_5 = B_4 \qquad B'_5 = B'_4$$

is within one bit of the capacity region irrespective of the signal-to-noise ratios Λ_k and interference-to-noise ratios Γ_k.

6.2 Result

Consider a semi-deterministic interference channel, which is characterized by its transition probabilities $p_{S_k|X_k}$, functions f_k, and alphabets $\mathcal{Y}_k, \mathcal{X}_k, \mathcal{S}_k$. Append to the channel variables (Y_1, Y_2, S_1, S_2) two auxiliary variables $(U_1, U_2) \in \mathcal{S}_1 \times \mathcal{S}_2$ which are independently and identically generated in the same manner as (S_1, S_2) are; that is, the following are four parallel memoryless channels:

$$X_1 \xrightarrow{p_{S_1|X_1}} S_1 \qquad X_2 \xrightarrow{p_{S_2|X_2}} S_2$$

$$X_1 \xrightarrow{p_{S_1|X_1}} U_1 \qquad X_2 \xrightarrow{p_{S_2|X_2}} U_2.$$

If $X_1 \multimap U_0 \multimap X_2$, then providing (X_1, X_2) as an input to the appended channel above induces a distribution on $(U_0, U_1, U_2, X_1, X_2)$ for which $U_1 \multimap X_1 \multimap U_0 \multimap X_2 \multimap U_2$ and is hence a valid input distribution for the hypothesis of Theorem 6.1.1. If we restrict the alphabets $\mathcal{Y}_k, \mathcal{X}_k, \mathcal{S}_k$ to be finite, so that our discussion is limited to discrete channels, then Theorem 6.1.1 applies with the mutual information bounds (6.2) simplifying to

$$\begin{aligned}
I_1 &= H(Y_1|U_0, U_1, U_2) - H(S_2|U_0, U_2) & I'_1 &= H(Y_2|U_0, U_1, U_2) - H(S_1|U_0, U_1) \\
I_2 &= H(Y_1|U_0, U_1) - H(S_2|U_0, U_2) & I'_2 &= H(Y_2|U_0, U_2) - H(S_1|U_0, U_1) \\
I_3 &= H(Y_1|U_0, U_2) - H(S_2|U_0, U_2) & I'_3 &= H(Y_2|U_0, U_1) - H(S_1|U_0, U_1) \\
I_4 &= H(Y_1|U_0) - H(S_2|U_0, U_2) & I'_4 &= H(Y_2|U_0) - H(S_1|U_0, U_1) \\
I_5 &= H(Y_1) - H(S_2|U_0, U_2) & I'_5 &= H(Y_2) - H(S_1|U_0, U_1).
\end{aligned} \tag{6.3}$$

That is, any rate triple $(R_0, R_1, R_2) \in \mathcal{R}_i(U_0, X_1, X_2) = \mathcal{R}_{JXG}(I_1, \ldots, I_5, I_1', \ldots, I_5')$ (as in (6.1)) is achievable and

$$\mathcal{R}_i = \bigcup_{X_1 - \circ - U_0 - \circ - X_2} \mathcal{R}_i(U_0, X_1, X_2)$$

serves as an inner bound to the capacity region of the semi-deterministic interference channel with common information.

Define $\mathcal{R}_o(U_0, X_1, X_2) = \mathcal{R}_{JXG}(O_1, \ldots, O_5, O_1', \ldots, O_5')$ where

$$\begin{aligned}
O_1 &= H(Y_1|U_0, U_1, X_2) - H(S_2|X_2) & O_1' &= H(Y_2|U_0, X_1, U_2) - H(S_1|X_1) \\
O_2 &= H(Y_1|U_0, U_1) - H(S_2|X_2) & O_2' &= H(Y_2|U_0, U_2) - H(S_1|X_1) \\
O_3 &= H(Y_1|U_0, X_2) - H(S_2|X_2) & O_3' &= H(Y_2|U_0, X_1) - H(S_1|X_1) \\
O_4 &= H(Y_1|U_0) - H(S_2|X_2) & O_4' &= H(Y_2|U_0) - H(S_1|X_1) \\
O_5 &= H(Y_1) - H(S_2|X_2) & O_5' &= H(Y_2) - H(S_1|X_1).
\end{aligned} \tag{6.4}$$

Then the main result given here is

Theorem 6.2.1. *If (R_0, R_1, R_2) is achievable for the semi-deterministic interference channel with common information, then $(R_0, R_1, R_2) \in \mathcal{R}_o$ where*

$$\mathcal{R}_o = \bigcup_{X_1 - \circ - U_0 - \circ - X_2} \mathcal{R}_o(U_0, X_1, X_2).$$

We relegate the proof of this outer bound to Section 6.4. The salient characteristic of this outer bound is that it of the same form as the inner bound $\mathcal{R}_i$ and that the gap between the two bounds is easily quantifiable.

Theorem 6.2.2. *If $(R_0, R_1, R_2) \in \mathcal{R}_o(U_0, X_1, X_2)$ for some set of random variables $(U_0, X_1, X_2) \in \mathcal{P}$, then*

$$(R_0 - \max(G_1, G_2), R_1 - G_2, R_2 - G_1) \in \mathcal{R}_i(U_0, X_1, X_2)$$

where $G_k = I(X_k; S_k|U_0, U_k)$.

Proof. We can relax all the positive entropy terms in the outer bounds (6.4) by replacing the X_k with U_k, e.g.,

$$H(Y_1|U_0,U_1,U_2) \geq H(Y_1|U_0,U_1,U_2,X_2) \overset{(i)}{=} H(Y_1|U_0,U_1,X_2)$$
$$H(Y_1|U_0,U_2) \geq H(Y_1|U_0,U_2,X_2) \overset{(ii)}{=} H(Y_1|U_0,X_2)$$

where both (i) and (ii) follow as $Y_1 \multimap (W,X_2) \multimap U_2$ for any W other than S_2 (e.g., $W = U_0$ or $W = (U_0,U_1)$[1]). Define these new relaxed outer bounds as $O_j^r, O_j'^r$ and let the region defined by these bounds be $\mathcal{R}_0^r$. Because the channels $X_k \to S_k$ are memoryless, $H(S_k|X_k) = H(S_k|X_k,U_0,U_k)$, and so subtracting the inner bounds (6.3) from these relaxed outer bounds provides

$$O_j^r - I_j = I(X_2;S_2|U_0,U_2) = G_2$$
$$O_j'^r - I_j' = I(X_1;S_1|U_0,U_1) = G_1$$

for $j = 1,2,\ldots,5$. In summary, if $(R_0,R_1,R_2) \in \mathcal{R}_o \subset \mathcal{R}_0^r$, then $(R_0 - \max(G_1,G_2), R_1 - G_2, R_2 - G_1) \in \mathcal{R}_i$. □

6.2.1 Relationship to previous results

The above result incorporates some previous results.

- If $S_k = X_k$, then the channel is the deterministic ICC and, as noted in [53], the region $\mathcal{R}_i$ is optimal. Our result reproduces this: if $S_k = X_k$, then $U_k = S_k$ and

$$G_k = I(X_k,S_k|U_k,U_0) = 0 \qquad \text{for } k = 1,2.$$

  That is, $\mathcal{R}_o = \mathcal{R}_i$.

[1] For (ii), we can factor $p(y_1|u_0,u_2,x_2)$ as

$$\frac{\sum_{x_1,s_2} p(u_0)p(x_1|u_0)p(x_2|u_0)p(u_2|x_2)p(s_2|x_2)p(y_1|x_1,s_2)}{p(u_0)p(x_2|u_0)p(u_2|x_2)},$$

which is independent of u_2 and thereby equal to $p(y_1|u_0,x_2)$. We can factor $p(y_1|u_0,u_1,u_2,x_2)$ analogously to obtain (i).

- We can apply our bounds to the semi-deterministic IC without common information. To do this, set $R_0 = 0$ and $U_0 = Q$, a coded time-sharing variable revealed to both the senders and receivers. Then all bounds that involve R_0 become redundant, $\mathcal{R}_o$ reproduces the outer bound of [103], and $\mathcal{R}_i$ reproduces the Han-Kobayashi region (as pointed out in [53]). The gap between the bounds reproduces the gap in [103],

$$G_k = I(X_k, S_k|U_k, Q).$$

- Our outer bound is a subset of Tan's outer bound in his initial study [101] of the ICC,

$$\begin{aligned}\mathcal{R}_{\text{Han}}(U, X_1, X_2) = \{(R_0, R_1, R_2 :& \\ R_1 \leq H_1 &= I(X_1; Y_1|X_2, U) \\ R_2 \leq H_1' &= I(X_2; Y_2|X_1, U) \\ R_1 + R_2 \leq H_2 &= I(X_1, X_2; Y_1, Y_2|U) \\ R_0 + R_1 + R_2 \leq H_3 &= I(X_1, X_2; Y_1, Y_2)\}\end{aligned}$$

where X_1 and X_2 are independent given U. Specializing these bounds to the semi-deterministic channel,

$$\begin{aligned}H_1 &= H(Y_1|X_2, U) - H(S_2|X_2) \\ H_1' &= H(Y_2|X_1, U) - H(S_1|X_1) \\ H_2 &= H(Y_1|U) + H(Y_2|U, Y_1) - H(S_2|X_2) - H(S_1|X_1) \\ H_3 &= H(Y_1) + H(Y_2|Y_1) - H(S_2|X_2) - H(S_1|X_1).\end{aligned}$$

To see that $\mathcal{R}_o(U_0, X_1, X_2) \subset \mathcal{R}_{\text{Han}}(U_0, X_1, X_2)$, note that $O_3 = H_1$, $O_3' = H_1'$, and

$$H(Y_2|U_0, X_1, U_2) = H(Y_2|U_0, Y_1, X_1, S_2) \leq H(Y_2|U_0, Y_1) \leq H(Y_2|Y_1).$$

Hence, $O_1' + O_4 \leq H_2$ and $O_1' + O_5 \leq H_3$.

6.3 Vector Gaussian ICC

In addition to subsuming previous results, these results provide novel bounds for the vector (multi-input multi-output) Gaussian interference channel with common information. With $(k,l) \in \{(1,2),(2,1)\}$, transmitter k having t_k dimensions, and receiver l having r_l dimensions, the vector Gaussian interference channel and corresponding auxiliary variables U_k are

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{H}_{11}\mathbf{X}_1 + \mathbf{S}_2 & \mathbf{S}_2 &= \mathbf{H}_{12}\mathbf{X}_2 + \mathbf{Z}_1 \\ & & \mathbf{U}_2 &= \mathbf{H}_{12}\mathbf{X}_2 + \mathbf{Z}_1' \\ \mathbf{Y}_2 &= \mathbf{H}_{22}\mathbf{X}_2 + \mathbf{S}_1 & \mathbf{S}_1 &= \mathbf{H}_{21}\mathbf{X}_1 + \mathbf{Z}_2 \\ & & \mathbf{U}_1 &= \mathbf{H}_{21}\mathbf{X}_1 + \mathbf{Z}_2' \end{aligned}$$

where the $\mathbf{Z}_l, \mathbf{Z}_l'$ have distribution $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{r_l \times r_l})$, independently of each other and all other random variables. Practical power limitations require that every coding scheme respect a power constraint $\sum_{t=1}^{n} \mathbf{X}_{kt}^* \mathbf{X}_{kt} \leq nP_k$ at transmitter k.

6.3.1 Constant Gap

The capacity characterization developed for the discrete setting continues to hold in the Gaussian setting, on which we further elaborate in Section 6.3.2. Particularly appealing is that the gap to the capacity region can be crisply stated: for any $(U_0, \mathbf{X}_1, \mathbf{X}_2)$ (not necessarily Gaussian),

$$\begin{aligned} G_k &= I(\mathbf{X}_k; \mathbf{S}_k | U_0, \mathbf{U}_k) \\ &= h(\mathbf{S}_k | U_0, \mathbf{U}_k) - h(\mathbf{S}_k | \mathbf{X}_k, U_0, \mathbf{U}_k) \\ &\leq h(\mathbf{Z}_k - \mathbf{Z}_k' | \mathbf{U}_k) - h(\mathbf{Z}_k) \\ &\overset{(i)}{\leq} h(\mathbf{Z}_k - \mathbf{Z}_k' | \mathbf{H}_{jk}\mathbf{X}_k^G + \mathbf{Z}_j) - h(\mathbf{Z}_k) \\ &= \log\det(2\mathbf{I} - (\mathbf{I} + \mathbf{H}_{lk}\mathrm{Cov}[\mathbf{X}_k, \mathbf{X}_k]\mathbf{H}_{lk}^*)^{-1}) \\ &\overset{(ii)}{=} \sum\nolimits_{i=1}^{n_k} \log(2 - 1/(1+\lambda_i(k))) \\ &\leq n_k \text{ bits}, \end{aligned} \tag{6.5}$$

the rank of $\mathbf{H}_{lk}$. Step (i) follows with $\mathbf{H}_{jk}\mathbf{X}_k^G + \mathbf{Z}_j$ as a Gaussian random variable with the same mean and covariance as $\mathbf{H}_{jk}\mathbf{X}_k + \mathbf{Z}_j$ and by the fact that Gaussian distributions maximize conditional entropy subject to a joint covariance constraint [104]. Step (ii) follows with $\lambda_1(k) \geq \lambda_2(k) \geq \cdots$ as the eigenvalues of $\mathbf{H}_{lk}\text{Cov}[\mathbf{X}_k]\mathbf{H}_{lk}^*$. Moreover, these bounds can be further relaxed to a **constant** gap irrespective of the channel parameters in each $\mathbf{H}_{ij}$: $n_k \leq \min\{t_k, r_l\}$.

In summary, if (R_0, R_1, R_2) is achievable for **some** coding scheme, then $(R_0 - \max(n_1, n_2), R_1 - n_2, R_2 - n_1)$ is achievable with the coding scheme described above. For the scalar case, this gap is no more than one bit and in accordance with previous results [109]. For the vector case, where results are more difficult to attain, this gap is clean and tighter relative to gaps provided in prior literature (e.g.[57] for the case with only private information).

Remark: Our scheme achieves this tighter gap at the expense of complexity of description: our region is a union over an infinite number of input distributions while both [109] and [57] demonstrate their results with a single input distribution. In what follows, we reduce some of our complexity of description by demonstrating that the restriction to Gaussian distributions is without loss of generality.

6.3.2 Capacity characterization

The achievability and converse carry over in a straightforward manner from the discrete to Gaussian setting, with both the inner and outer bounds characterized as a union over distributions in the set

$$\{(U_0, \mathbf{X}_1, \mathbf{X}_2) \ : \ \mathbf{X}_1 \multimap U_0 \multimap \mathbf{X}_2, \ E[\mathbf{X}_k^*\mathbf{X}_k] \leq P_k\}.$$

With a little bit of work one can show that in fact it suffices to restrict the union to the Gaussian subset. The challenge lies in demonstrating that Gaussian distributions maximize certain mutual information terms **subject to** a Markov condition.

Such subtleties were first observed for the two user scalar (single antenna) Gaussian multiple-access channel with conferencing encoders in [10] and for its three-user vector (multiple antenna) generalization in [118]. By mimicking these results, we can state the following.

Theorem 6.3.1. *For the Gaussian ICC, any achievable rate triple (R_0, R_1, R_2) must be in*

$$\mathcal{R}_o^G = \bigcup_{\substack{\mathbf{X}_1^G \multimap U_0^G \multimap \mathbf{X}_2^G \\ \mathrm{E}[(\mathbf{X}_k^G)^* \mathbf{X}_k^G] \leq P_k}} \mathcal{R}_o(\mathbf{U}_0, \mathbf{X}_1, \mathbf{X}_2).$$

Proof. A slight modification of the converse argument in Section 6.4 to account for the power constraints provides that any achievable rate must be in

$$\mathcal{R}_o = \bigcup_{\substack{\mathbf{X}_1 \multimap U_0 \multimap \mathbf{X}_2 \\ \mathrm{E}[\mathbf{X}_k^* \mathbf{X}_k] \leq P_k}} \mathcal{R}_o(U_0, \mathbf{X}_1, \mathbf{X}_2).$$

Let $(U_0^G, \mathbf{X}_1^G, \mathbf{X}_2^G)$ be the Gaussian random triple with the same mean and joint covariance as the random triple $(U_0, \mathbf{X}_1, \mathbf{X}_2)$. Then, as Gaussian random variables maximize conditional entropy subject to a joint covariance constraint [104], we know that $\mathcal{R}_o(U_0, \mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{R}_o(U_0^G, \mathbf{X}_1^G, \mathbf{X}_2^G)$. However, there is no guarantee that $\mathbf{X}_1^G \multimap U_0^G \multimap \mathbf{X}_2^G$ forms a Markov Chain. To circumvent this problem, introduce

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{01} \\ \mathbf{V}_{02} \end{bmatrix} = \begin{bmatrix} E[\mathbf{X}_1 | U_0] \\ E[\mathbf{X}_2 | U_0] \end{bmatrix}.$$

As $\mathbf{V}$ is a function of U_0, we can relax the bounds O_j, O_j' by substituting $\mathbf{V}$ in for U_0 to get

$$\mathcal{R}_o(U_0, \mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{R}_o(\mathbf{V}, \mathbf{X}_1, \mathbf{X}_2).$$

Define, for $k \in \{1, 2\}$, $\tilde{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{V}_{0k}$. By the orthogonality principle for MMSE estimation, the estimation error $\tilde{\mathbf{X}}_k$ is orthogonal to any function of U_0 with finite second moment; hence, $\mathrm{Cov}(\tilde{\mathbf{X}}_k, \mathbf{V}) = \mathbf{0}$ for both $k \in \{1, 2\}$. Essentially, $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}_2$ are also orthogonal:

$$\begin{aligned} \mathrm{Cov}(\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) &\overset{(i)}{=} EE[(\mathbf{X}_1 - E[\mathbf{X}_1|U_0])(\mathbf{X}_2 - E[\mathbf{X}_2|U_0])^* | U_0] \\ &\overset{(ii)}{=} E[(E[\mathbf{X}_1|U_0] - E[\mathbf{X}_1|U_0])(E[\mathbf{X}_2|U_0] - E[\mathbf{X}_2|U_0])^*] = \mathbf{0}, \end{aligned}$$

where (i) follows by the tower property of expectation for random variables with finite second moment[45] and (ii) follows as $\mathbf{X}_1 \multimap U_0 \multimap \mathbf{X}_2$.

Let $(\mathbf{V}^G, \mathbf{X}_1^G, \mathbf{X}_2^G)$ be the Gaussian random triple with the same mean and joint covariance as $(\mathbf{V}, \mathbf{X}_1, \mathbf{X}_2)$. As $\tilde{\mathbf{X}}_1^G, \tilde{\mathbf{X}}_2^G$, and $\mathbf{V}^G$ are orthogonal and jointly Gaussian, they are also all

independent. Hence, by the decomposition

$$\begin{bmatrix} \mathbf{X}_1^G \\ \mathbf{X}_2^G \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}_1^G \\ \tilde{\mathbf{X}}_2^G \end{bmatrix} + \mathbf{V}^G$$

we have $\mathbf{X}_1^G \multimap\!\!\circ\!\!- \mathbf{V}^G \multimap\!\!\circ\!\!- \mathbf{X}_2^G$. As we preserved covariances, $E[(\mathbf{X}_k^G)^* \mathbf{X}_k^G] \leq P_k$ for both k as well. Now we can state

$$\mathcal{R}_o(U_0, \mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{R}_o(\mathbf{V}, \mathbf{X}_1, \mathbf{X}_2) \subset \mathcal{R}_o(\mathbf{V}^G, \mathbf{X}_1^G, \mathbf{X}_2^G),$$

with $(\mathbf{V}^G, \mathbf{X}_1^G, \mathbf{X}_2^G)$ in the desired set of input distributions. That is, $\mathcal{R}_o \subset \mathcal{R}_o^G$. □

On the achievability side, successively refined discretizations (Chapter 3, pgs. 50-51[31]), along with a restriction to Gaussian distributions that satisfy the power constraints to within a vanishing ϵ, extend Theorem 6.1.1 to demonstrate that

$$\mathcal{R}_i^G = \bigcup_{\substack{\mathbf{X}_1^G \multimap\!\!\circ\!\!- \mathbf{U}_0^G \multimap\!\!\circ\!\!- \mathbf{X}_2^G \\ \mathrm{E}[(\mathbf{X}_k^G)^* \mathbf{X}_k^G] \leq P_k}} \mathcal{R}_i(\mathbf{U}_0, \mathbf{X}_1, \mathbf{X}_2)$$

is an inner bound to the Gaussian ICC.

Remark: The method of Theorem 6.3.1 may be applied to determine the capacity of the two user vector (multiple antenna) Gaussian MAC with common information or conferencing encoders. This is a more direct means of determining capacity than would be specializing the results for the more general three user vector Gaussian MAC in [118], and as such we provide details in the Appendix.

6.4 Proof of outer bound

Suppose that for the discrete semi-deterministic interference channel (where all alphabets are finite) there exists a sequence of codes indexed by block length n which communicate at the rate triple (R_0, R_1, R_2) and which achieve a vanishing probability of error $P_e^{(n)}$ as n tends to infinity. Pick the code corresponding to block length n, and let M_0, M_1, and M_2 be uniformly distributed messages on $\left[1 : 2^{nR_0}\right]$, $\left[1 : 2^{nR_1}\right]$, and $\left[1 : 2^{nR_2}\right]$. Let X_1^n, X_2^n, Y_1^n, and Y_2^n be the random variables

be induced by the messages, the encoders, and the channel. As in our description of the achievable scheme, augment this set of channel variables by a sequence of random variables (U_1^n, U_2^n) obtained by passing X_1^n and X_2^n through memoryless side channels defined by $p_{S_1|X_1}$ and $p_{S_2|X_2}$.

6.4.1 Preliminaries

We make note of a few relationships that must be true for this set of random variables. Recall our convention that $(k,l) \in \{(1,2),(2,1)\}$. Because the the maps $f_k(x_k,\cdot)$ are one-to-one,

$$H(Y_k^n|X_k^n, M_0, \tilde{U}_k) = H(S_l^n|X_k^n, M_0, \tilde{U}_k) = H(S_l^n|M_0) \tag{6.6}$$

for any random variable $\tilde{U}_k$ conditionally independent of S_l^n given M_0 (e.g., $\tilde{U}_k = M_k$). Similarly, if $\tilde{W}_k$ is conditionally independent of S_l^n given X_l^n,

$$H(Y_k^n|X_k^n, X_l^n, \tilde{W}_k) = H(S_l^n|X_k^n, X_l^n, \tilde{W}_k) = H(S_l^n|X_l^n). \tag{6.7}$$

As $Y_{1i} = f_1(X_{1i}, S_{2i})$ is a function of S_{2i} but **not** of S_{1i}, $(M_0, X_2^n, Y_1^n) \multimap X_1^n \multimap S_1^n$ forms a Markov chain with distribution, by construction, equivalent to the Markov Chain $(M_0, X_2^n, Y_1^n) \multimap - X_1^n \multimap U_1^n$. Swap the indices one and two to obtain an analogous result. By these two distributional equivalences,

$$H(Y_k^n|M_0, S_k^n) = H(Y_k^n|M_0, U_k^n) \tag{6.8}$$

$$H(Y_k^n|M_0, X_l^n, S_k^n) = H(Y_k^n|M_0, X_l^n, U_k^n). \tag{6.9}$$

6.4.2 Bounds

By Fano's inequality, we have that $nR_k \le I(M_k;Y_k) + n\epsilon_n$ and $n(R_0+R_k) \le I(M_0,M_k;Y_k) + n\epsilon_n$ for $k = 1,2$ and some sequence ϵ_n for which $\epsilon_n \to 0$ as $n \to \infty$. So,

$$\begin{aligned} n(R_1 - \epsilon_n) &\le I(M_1;Y_1^n) \le I(M_1;Y_1^n, M_0, X_2^n) \\ &\overset{(i)}{=} I(M_1;Y_1^n|M_0,X_2^n) \\ &\le H(Y_1^n|M_0,X_2^n) - H(Y_1^n|M_0,M_1,X_1^n,X_2^n) \end{aligned}$$

$$\stackrel{(ii)}{=} H(Y_1^n|M_0, X_2^n) - H(S_2^n|X_2^n)$$
$$\stackrel{(iii)}{\leq} \sum\nolimits_i H(Y_{1i}|M_0, X_2^n) - H(S_{2i}|X_{2i}) \tag{6.10a}$$

where (i) follows by the independence of M_1 and (M_0, M_2), (ii) follows by (6.7) with $\tilde{W}_1 = (M_0, M_1)$, and (iii) follows as $X_2^n \to S_2^n$ is a discrete memoryless channel. Similarly,

$$\begin{aligned}
&n(R_0 + R_1 - \epsilon_n) \\
&\quad \leq I(M_0, M_1; Y_1^n) \leq I(X_1^n, X_2^n; Y_1^n) \\
&\quad \stackrel{(iv)}{=} H(Y_1^n) - H(S_2^n|X_2^n) \\
&\quad \leq \sum\nolimits_i H(Y_{1i}) - H(S_{2i}|X_{2i}) \qquad (6.10b) \\
&n(R_1 + R_2 - 2\epsilon_n) \\
&\quad \leq I(M_1; Y_1^n, M_0) + I(M_2; Y_2^n, M_0, S_2^n, X_1^n) \\
&\quad = I(M_1; Y_1^n|M_0) + I(M_2; Y_2^n, S_2^n|M_0, X_1^n) \\
&\quad \leq I(M_1, X_1^n; Y_1^n|M_0) + I(M_2, X_2^n; Y_2^n, S_2^n|M_0, X_1^n) \\
&\quad = I(M_1, X_1^n; Y_1^n|M_0) + I(M_2, X_2^n; S_2^n|M_0, X_1^n) + I(M_2, X_2^n; Y_2^n|M_0, X_1^n, S_2^n) \\
&\quad \stackrel{(v)}{=} H(Y_1^n|M_0) - H(S_2^n|M_0) + H(S_2^n|M_0) - H(S_2^n|X_2^n) + H(Y_2^n|M_0, X_1^n, U_2^n) - H(S_1^n|X_1^n) \\
&\quad \leq \sum\nolimits_i \left[H(Y_{1i}|M_0) + H(Y_{2i}|M_0, X_{1i}, U_{2i}) - H(S_{2i}|X_{2i}) - H(S_{1i}|X_{1i})\right] \qquad (6.10c)
\end{aligned}$$

where (iv) follows by (6.7) with $\tilde{W}_1 = \emptyset$, (v) follows by applying (6.6) to the second and third terms (with $\tilde{U}_1 = M_1$ and $\tilde{U}_1 = \emptyset$ respectively), applying (6.7) to the fourth and sixth terms (with $\tilde{W}_1 = (M_0, M_2)$ and $\tilde{W}_2 = (M_0, M_2, S_2^n)$ respectively), and applying (6.9) to the fifth term. The following argument parallels precisely the argument preceding (6.10c) with the exception of the first term at each step,

$$\begin{aligned}
&n(R_0 + R_1 + R_2 - 2\epsilon_n) \\
&\quad \leq I(M_0, M_1; Y_1^n) + I(M_2; Y_2^n, M_0, S_2^n, X_1^n) \\
&\quad \leq I(M_0, M_1, X_1^n; Y_1^n) + I(M_2; Y_2^n, S_2^n|M_0, X_1^n) \\
&\quad \leq \sum\nolimits_i \left[H(Y_{1i}) + H(Y_{2i}|M_0, X_{1i}, U_{2i}) - H(S_{2i}|X_{2i}) - H(S_{1i}|X_{1i})\right] \qquad (6.10d)
\end{aligned}$$

Again, in a similar manner as the above arguments,

$$
\begin{aligned}
&n(2R_1 + R_2 - 3\epsilon_n) \\
&\leq I(M_1; Y_1^n, M_0) + I(M_1; Y_1^n, M_0, S_1^n, X_2^n) + I(M_2; Y_2^n, S_2^n, M_0) \\
&= I(M_1; Y_1^n | M_0) + I(M_1; Y_1^n, S_1^n | M_0, X_2^n) + I(M_2; Y_2^n, S_2^n | M_0) \\
&\leq I(M_1, X_1^n; Y_1^n | M_0) + I(M_1, X_1^n; Y_1^n, S_1^n | M_0, X_2^n) + I(M_2, X_2^n; Y_2^n, S_2^n | M_0) \\
&= I(M_1, X_1^n; Y_1^n | M_0) + I(M_1, X_1^n; S_1^n | M_0, X_2^n) + I(M_1, X_1^n; Y_1^n | M_0, X_2^n, S_1^n) \\
&\quad + I(M_2, X_2^n; S_2^n | M_0) + I(M_2, X_2^n; Y_2^n | M_0, S_2^n) \\
&\overset{(vi)}{=} H(Y_1^n | M_0) - H(S_2^n | M_0) + H(S_1^n | M_0) - H(S_1^n | X_1^n) + H(Y_1^n | M_0, X_2^n, U_1^n) \\
&\quad - H(S_2^n | X_2^n) + H(S_2^n | M_0) - H(S_2^n | X_2^n) + H(Y_2^n | M_0, U_2^n) - H(S_1^n | M_0) \\
&\leq \sum\nolimits_i \big[H(Y_{1i} | M_0) + H(Y_{1i} | M_0, X_{2i}, U_{1i}) + H(Y_{2i} | M_0, U_{2i}) - H(S_{1i} | X_{1i}) - 2H(S_{2i} | X_{2i}) \big] \quad (6.10e)
\end{aligned}
$$

where (vi) follows by applying (6.6) to the second, third, and tenth terms (with $\tilde{U}_1 = M_1$, $\tilde{U}_2 = \emptyset$, and $\tilde{U}_2 = (M_2, S_2^n)$ respectively), applying (6.7) to the fourth, sixth, and eighth terms (with $\tilde{W}_2 = (M_0, M_1)$, $\tilde{W}_1 = (M_0, M_1, S_1^n)$, and $\tilde{W}_1 = M_2$ respectively), applying (6.8) to the ninth term, and applying (6.9) to the fourth term. If we parallel the argument preceding (6.10e) exactly except for the first term at each step, we find,

$$
\begin{aligned}
&n(R_0 + 2R_1 + R_2 - 3\epsilon_n) \\
&\quad \leq I(M_0, M_1; Y_1^n) + I(M_1; Y_1^n, M_0, S_1^n, X_2^n) + I(M_2; Y_2^n, S_2^n, M_0) \\
&\quad \leq \sum\nolimits_i \big[H(Y_{1i}) + H(Y_{1i} | M_0, X_{2i}, U_{1i}) + H(Y_{2i} | M_0, U_{2i}) - H(S_{1i} | X_{1i}) - 2H(S_{2i} | X_{2i}) \big]. \quad (6.10f)
\end{aligned}
$$

By symmetry, we can extend the above results to

$$
\begin{aligned}
n(R_2 - \epsilon_n) &\leq \sum\nolimits_i H(Y_{2i} | M_0, X_{1i}) - H(S_{1i} | X_{1i}) && (6.11a) \\
n(R_0 + R_2 - \epsilon_n) &\leq \sum\nolimits_i H(Y_{2i}) - H(S_{1i} | X_{1i}) && (6.11b) \\
n(R_1 + R_2 - 2\epsilon_n) &\leq \sum\nolimits_i \big[H(Y_{2i} | M_0) + H(Y_{1i} | M_0, X_{2i}, U_{1i}) \\
&\qquad - H(S_{1i} | X_{1i}) - H(S_{2i} | X_{2i}) \big] && (6.11c)
\end{aligned}
$$

$$n(R_0+R_1+R_2-2\epsilon_n) \le \sum\nolimits_i \big[H(Y_{2i}) + H(Y_{1i}|M_0,X_{2i},U_{1i}) - H(S_{1i}|X_{1i}) - H(S_{2i}|X_{2i})\big] \tag{6.11d}$$

$$\begin{aligned} n(R_1+2R_2-3\epsilon_n) \le \sum\nolimits_i \big[&H(Y_{2i}|M_0) + H(Y_{2i}|M_0,X_{1i},U_{2i}) + H(Y_{1i}|M_0,U_{1i}) \\ &- H(S_{2i}|X_{2i}) - 2H(S_{1i}|X_{1i})\big] \end{aligned} \tag{6.11e}$$

$$\begin{aligned} n(R_0+R_1+2R_2-3\epsilon_n) \le \sum\nolimits_i \big[&H(Y_{2i}) + H(Y_{2i}|M_0,X_{1i},U_{2i}) + H(Y_{1i}|M_0,U_{1i}) \\ &- H(S_{2i}|X_{2i}) - 2H(S_{1i}|X_{1i})\big]. \end{aligned} \tag{6.11f}$$

Finally, we also have

$$\begin{aligned} &n(R_1+R_2-2\epsilon_n) \\ &\quad\le I(M_1;Y_1^n,S_1^n,M_0) + I(M_2;Y_2^n,S_2^n,M_0) \\ &\quad= I(M_1;Y_1^n,S_1^n|M_0) + I(M_2;Y_2^n,S_2^n|M_0) \\ &\quad\le I(M_1,X_1^n;Y_1^n,S_1^n|M_0) + I(M_2,X_2^n;Y_2^n,S_2^n|M_0) \\ &\quad= I(M_1,X_1^n;S_1^n|M_0) + I(M_1,X_1^n;Y_1^n|M_0,S_1^n) + I(M_2,X_2^n;S_2^n|M_0) + I(M_2,X_2^n;Y_2^n|M_0,S_2^n) \\ &\quad\overset{(vii)}{=} H(S_1^n|M_0) - H(S_1^n|X_1^n) + H(Y_1^n|M_0,U_1^n) - H(S_2^n|M_0) + H(S_2^n|M_0) - H(S_2^n|X_2^n) \\ &\qquad + H(Y_2^n|M_0,U_2^n) - H(S_1^n|M_0) \\ &\quad\le \sum\nolimits_i \big[H(Y_{1i}|M_0,U_{1i}) + H(Y_{2i}|M_0,U_{2i}) - H(S_{2i}|X_{2i}) - H(S_{1i}|X_{1i})\big] \end{aligned} \tag{6.12}$$

where (vii) follows by applying (6.6) to the fourth and eighth terms (with $\tilde{U}_1 = (M_1, S_1^n)$ and $\tilde{U}_2 = (M_2, S_2^n)$ respectively), applying (6.7) to the second and sixth terms (with $\tilde{W}_2 = M_1$ and $\tilde{W}_1 = M_2$ respectively), and applying (6.8) to the third and seventh terms. Setting Q, X_1, X_2 to be random variables with Q uniformly distributed on $[1,\dots,n]$, $X_k = X_{kQ}$, $U_0 = (M_0, Q)$, and noting that

$$H(Y_k|Q) \le H(Y_k) \qquad H(S_k|X_k,Q) = H(S_k|X_k)$$

as the two side channels $X_k \to S_k$ are time-invariant; that is, $p(s_{ki}|x_{ki}) = p_{S_k|X_k}(s_{ki}|x_{ki})$ for each i. Hence, the inequalities (6.10)-(6.12) imply

$$(R_0-\epsilon_n, R_1-\epsilon_n, R_2-\epsilon_n) \in \mathcal{R}_o(U_o, X_1, X_2) \subset \mathcal{R}_o.$$

As $\mathcal{R}_0$ is closed, we see that any achievable rate is in $\mathcal{R}_0$.

Remark: As mentioned in [103], the key to this type of bound is to choose receiver side information such that the mutual information quantities that bound $\sum_k c_k R_k$ will single-letterize. Our choices of side information are analogous to the choices in [103] (we choose S_k^n over U_k^n, though the proof works equally well with either choice), with the caveat that we provide the common message M_0 as additional side information whenever we are **not** bounding $R_0 + R_k$. With our choices of side information, the only multi-letter entropy terms that do not cancel and have a negative coefficient are $H(S_k^n|X_k^n)$, which single-letterize by definition.

6.5 Conclusion

We have shown that for a certain class of interference channels with common information the distance to optimality of the Jiang-Xin-Garg achievable region can be bounded. The most notable aspect of this class of interference channels are the implications it carries for signaling over Gaussian channels with correlated inputs. In particular, our results show that the Jiang-Xin-Garg achievable scheme is within a constant gap of the capacity region for both scalar (single antenna) and vector (multiple antenna) Gaussian channels.

Chapter 7

Summary and Future Directions

7.1 Summary

This thesis generalizes classical results on fundamental limits of network communication from settings with specific message sets to settings with general message sets. Novel exact capacity characterizations are provided for the MAC (many-to-one), as well as novel approximate capacity characterizations for both the BC (one-to-many) and IC (two-to-two) networks. The established fundamentals limits provided here could help inform the design of, and pave the way towards, the development of a richer physical layer interface for next-generation wireless networks.

The central theme underlying all of these contributions is that order, and recursion, is fundamental to communication with general message sets. General message sets have a partial order associated with them: some messages are strictly more fundamental, which can be equated with a notion of order putting the more fundamental messages above those less fundamental, than others. Revisiting the classical random coding technique of superposition coding through this lens of order theory provides new insights:

- Dependencies in codebook design, and in input distribution factorizations, recurse up this partial order while dually, decoding errors propagate down this partial order.
- The achievable rate region of recursive encoding (that is, superposition coding) is of a special structure, that of a polymatroid. The defining bounds needn't be the entire boolean lattice of all subsets, but a sublattice, namely the down-set lattice, of that boolean lattice.

For the MAC, the polymatroid observation allows one to conclude that a recursive decoding procedure attains the capacity boundary. Order can be further elaborated upon without appeal to recursive codebook generation, through rate-splitting and variable-splitting.

As the vertices of polymatroids have an explicit analytic expression, the boundary of the achievable rate regions associated with superposition coding can be simply explored. In the Gaussian MAC, where a sufficient set of optimal input distributions is a convex set, exploiting this explicit characterization of the vertices enables efficient computation of the optimal power allocation allocation through a convex program.

In the BC, recursive encoding strategies can be formulated for both a degrees-of-freedom analysis and for a discrete memoryless analysis. An inner bound for the degrees-of-freedom region of the K-user BC is formulated, based on a recursive row vector selection procedure coupled with a network coding strategy, which attains capacity in select cases. An inner bound is also formulated for the discrete memoryless channel, which combines superposition coding, rate-splitting, and binning.

7.2 Future Directions

This thesis indicates that poylmatroidal structure may be more pervasive—and useful—than previously thought. Specializing the proposed inner bound for the discrete memoryless BC to the MIMO Gaussian BC could plausibly yield the degrees-of-freedom region. An intriguing question for the subject of further study is whether the set of covariance choices which attain this DoF region can be made to be finite, and preferably small and finite. Moreover, it is of interest to know whether the convex hull of the achievable DoF points thus enabled can be expressed simply.

Another direction of study, which may help direct focus towards types of outer bounds which might be helpful for the BC, is to determine a closed-form expression for the general BC inner bound proposed in Chapter 5 which combines superposition, rate-splitting, and binning. Here, both the rate-splits $r_{S\to S'}$ and the excess rates $\tilde{R}_S$ are projected away. Focusing on the case **without** binning is itself on interest: it may lead to a general inner bound for the general asymmetric K-

user combination network, which at the moment is only known for the symmetric setting or for the two- or three-user case.

Polymatroidal structure is known to exist in dual coding problems, such as the Multiple Description and Source Coding problems. It would be of interest to see whether a formulation with K sources, with an arbitrary collection of common parts, has relevance and could be developed to the same degree as the MAC with general message sets. Similarly, if light is shed no the BC through a polymatroid framework, perhaps light can be shed on the problem of multiple description coding through a polymatroid framework.

Lastly, it remains an open question as to whether or not the polymatriod framework might have anything meaningful to contribute to interference networks, or many-to-many communication.

Bibliography

[1] Rudolf Ahlswede. Multi-way communication channels. In 2nd International Symposium on Information Theory, 1973.

[2] Steen A Andersson, David Madigan, Michael D Perlman, and Christopher M Triggs. On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. Journal of Statistical Planning and Inference, 48(1):25–46, 1995.

[3] Steen A Andersson and Michael D Perlman. Lattice models for conditional independence in a multivariate normal distribution. The Annals of Statistics, pages 1318–1358, 1993.

[4] T Ando. Concavity of certain maps on positive definite matrices and applications to Hadamard products. Linear Algebra and its Applications, 26:203–241, 1979.

[5] J Barros and S D Servetto. Network information flow with correlated sources. IEEE Transactions on Information Theory, 52(1):155–170, January 2006.

[6] R Benzel. The capacity region of a class of discrete additive degraded interference channels (Corresp.). IEEE Transactions on Information Theory, 25(2):228–231, 1979.

[7] T. Berger. Multiterminal source coding. In The Information Theory Approach to Communications, CISM Courses and Lectures, pages 171–231. Springer- Verlag, New York, 1978.

[8] E. Biglieri, J. Proakis, and S Shamai. Fading channels: Information-theoretic and communications aspects. IEEE Transactions on Information Theory, 44(6):2619–2692, October 1998.

[9] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, New York, NY, USA, 2004.

[10] S.I Bross, A Lapidoth, and M.A Wigger. The Gaussian MAC with conferencing encoders. In Proceedings of the 2008 International Conference on Information Theory (ISIT)., pages 2702–2706, July 2008.

[11] V.R. Cadambe and S.A Jafar. Sum-capacity and the unique separability of the parallel Gaussian MAC-Z-BC network. In IEEE International Symposium on Information Theory Proceedings (ISIT), 2010, pages 2318–2322. IEEE, June 2010.

[12] G Caire and S Shamai. On the capacity of some channels with channel state information. IEEE Transactions on Information Theory, 45(6):2007–2019, September 1999.

[13] Teena Carroll, Joshua Cooper, and Prasad Tetali. Counting antichains and linear extensions in generalizations of the boolean lattice. `http://people.math.sc.edu/cooper/calegbl.pdf`, 2009.

[14] T Chan and A Grant. On capacity regions of non-multicast networks. In Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, pages 2378–2382. IEEE, 2010.

[15] R S Cheng and Sergio Verdu. Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling. IEEE Transactions on Information Theory, 39(3):773–785, 1993.

[16] H-F. Chong and M. Motani. The capacity region of a class of semideterministic interference channels. IEEE Transactions on Information Theory, 55(2), February 2009.

[17] H F Chong, Mehul Motani, H K Garg, and H El Gamal. On the Han–Kobayashi Region for the Interference Channel. IEEE Transactions on Information Theory, 54(7):3188–3195, 2008.

[18] M Costa and Abbas El-Gamal. The Capacity Region of the Discrete Memoryless Interference Channel with Strong Interference (Corresp.). IEEE Transactions on Information Theory, 33(5):710–711, 1987.

[19] T Cover, A.E. Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. IEEE Transactions on Information Theory, 26(6):648–657, November 1980.

[20] T. Cover and J.A. Thomas. Elements of Information Theory. Wiley-Interscience, july 2006.

[21] Thomas M Cover. Broadcast channels. IEEE Transactions on Information Theory, 18(1):2–14, 1972.

[22] Thomas M Cover. An achievable rate region for the broadcast channel. IEEE Transactions on Information Theory, 21(4):399–404, 1975.

[23] Thomas M Cover and Joy A Thomas. Elements of information theory. John Wiley & Sons, 2nd edition, 2006.

[24] B.A. Davey and H.A. Priestley. Introduction to Lattices and Order. Cambridge University Press, second edition, 2002.

[25] K. De Bruyn, V Prelov, and E van der Meulen. Reliable Transmission of Two Correlated Sources over an Asymmetric Multiple-Access Channel. IEEE Transactions on Information Theory, 33(5):716–718, September 1987.

[26] S N Diggavi, N Al-Dhahir, A Stamoulis, and A.R Calderbank. Great Expectations: The Value of Spatial Diversity in Wireless Networks. Proc. iEEE, 92(2):219–270, February 2004.

[27] R Dougherty, C Freiling, and K Zeger. Networks, Matroids, and Non-Shannon Information Inequalities. IEEE Transactions on Information Theory, 53(6):1949–1969, 2007.

[28] Jack Edmonds. Submodular functions, matroids, and certain polyhedra. Combinatorial structures and their applications, pages 69–87, 1970.

[29] Ersen Ekrem and Sennur Ulukus. An Outer Bound for the Gaussian MIMO Broadcast Channel With Common and Private Messages. IEEE Transactions on Information Theory, 58(11):6766–6772, 2012.

[30] A. El-Gamal and M. Costa. The capacity region of a class of deterministic interference channels (corresp.). IEEE Transactions on Information Theory, 28(2):343–346, 1982.

[31] A. El-Gamal and Y-H. Kim. Network Information Theory. Cambridge University Press, New York, NY, USA, 2012.

[32] Abbas El-Gamal. Capacity of the Product and Sum of two Unmatched Broadcast Channels. Problems of Information Transmission, 1980.

[33] R.H. Etkin, D.N.C Tse, and H. Wang. Gaussian Interference Channel Capacity to Within One Bit. IEEE Transactions on Information Theory, 54(12):5534–5562, December 2008.

[34] Satoru Fujishige. Polymatroidal dependence structure of a set of random variables. Information and Control, 39(1):55–72, 1978.

[35] Robert G Gallager. Information theory and reliable communication, volume 2. Springer, 1968.

[36] S I Gel'fand and M S Pinsker. Capacity of a broadcast channel with one deterministic component. Problemy Peredachi Informatsii, 16(1):24–34, January 1980.

[37] Yanlin Geng and Chandra Nair. The Capacity Region of the Two-Receiver Gaussian Vector Broadcast Channel With Private and Common Messages. IEEE Transactions on Information Theory, 60(4):2087–2104, 2014.

[38] Z Goldfeld, Haim H Permuter, and Benjamin M Zaidel. The Finite State MAC With Cooperative Encoders and Delayed CSI. IEEE Transactions on Information Theory, 60(10):6181–6203, October 2014.

[39] A Goldsmith, S.A Jafar, I Maric, and S Srinivasa. Breaking Spectrum Gridlock With Cognitive Radios: An Information Theoretic Perspective. In Proceedings of the IEEE, pages 894–914, May 2009.

[40] A J Goldsmith and P P Varaiya. Capacity of fading channels with channel side information. IEEE Transactions on Information Theory, 43(6):1986–1992, November 1998.

[41] A Grant, B Rimoldi, R.L Urbanke, and P.A Whiting. Rate-splitting multiple access for discrete memoryless channels. IEEE Transactions on Information Theory, 47(3):873–890, March 2001.

[42] Leonard Grokop and D N Tse. Fundamental Constraints on Multicast Capacity Regions. arXiv.org, September 2008.

[43] D Gunduz and O Simeone. On the capacity region of a multiple access channel with common messages. In Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, pages 470–474, 2010.

[44] A Haghi, R Khosravi-Farsani, M R Aref, and F Marvasti. The Capacity Region of p -Transmitter/ q -Receiver Multiple-Access Channels With Common Information. IEEE Transactions on Information Theory, 57(11):7359–7376, November 2011.

[45] B. Hajek. A Exploration of Random Processes for Engineers. Lecture Notes, `http://www.ifp.illinois.edu/~hajek/Papers/randomprocesses.html`, December 2011.

[46] B Hajek and M Pursley. Evaluation of an achievable rate region for the broadcast channel. IEEE Transactions on Information Theory, 25(1):36–46, January 1979.

[47] T Han and Kingo Kobayashi. A new achievable rate region for the interference channel. IEEE Transactions on Information Theory, 27(1):49–60, January 1981.

[48] T. S. Han. The capacity region of general multiple-access channel with certain correlated sources. Information and Control, 40(1):37–60, 1979.

[49] S V Hanly and D.N.C Tse. Multiaccess fading channels. II. Delay-limited capacities. IEEE Transactions on Information Theory, 44(7):2816–2831, November 1998.

[50] N J A Harvey, R Kleinberg, and A R Lehman. On the capacity of information networks. IEEE Transactions on Information Theory, 52(6):2345–2364, June 2006.

[51] SA Jafar and S Shamai. Degrees of freedom region of the MIMO X channel. IEEE Transactions on Information Theory, 54(1):151–170, 2008.

[52] Syed A Jafar and Andrea J Goldsmith. On the capacity of the vector MAC with feedback. IEEE Transactions on Information Theory, 52(7):3259–3264, 2006.

[53] J Jiang, Y Xin, and HK Garg. Interference channels with common information. IEEE Transactions on Information Theory, 54(1):171–187, 2008.

[54] N Jindal and Zhi-Quan Luo. Capacity Limits of Multiple Antenna Multicast. Information Theory, 2006 IEEE International Symposium on, pages 1841–1845, 2006.

[55] N Jindal, S Vishwanath, and A Goldsmith. On the duality of Gaussian multiple-access and broadcast channels. IEEE Transactions on Information Theory, 50(5):768–783, May 2004.

[56] E Karipidis, N D Sidiropoulos, and Zhi-Quan Luo. Quality of Service and Max-Min Fair Transmit Beamforming to Multiple Cochannel Multicast Groups. IEEE Transactions on Signal Processing, 56(3):1268–1279, 2008.

[57] S. Karmakar and M.K. Varanasi. The Capacity Region of the MIMO Interference Channel and Its Reciprocity to Within a Constant Gap. IEEE Transactions on Information Theory, 59(8):4781–4797, 2013.

[58] Mohammad Khojastepour and A Keshavarz-Haddad. Multicast Achievable Rate Region of Deterministic Broadcast Channel. In Communications (ICC), 2011 IEEE International Conference on, pages 1–6. IEEE, 2011.

[59] B. Korte and J. Vygen. Combinatorial Optimization. Springer-Verlag, Berlin Heidelberg New York, 5th edition, 2012.

[60] R.T. Krishnamachari and M.K. Varanasi. MIMO performance under covariance matrix feedback. In Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, 2011.

[61] A Lapidoth and M Wigger. On the AWGN MAC with imperfect feedback. IEEE Transactions on Information Theory, 56(11):5432–5476, November 2010.

[62] L Li and A Goldsmith. Capacity and optimal resource allocation for fading broadcast channels. I. Ergodic capacity. IEEE Transactions on Information Theory, 47(3):1083–1102, March 2001.

[63] L Li and A Goldsmith. Capacity and optimal resource allocation for fading broadcast channels. II. Outage capacity. IEEE Transactions on Information Theory, 47(3):1103–1127, March 2001.

[64] H. H. J. Liao. Multiple Access Channels. PhD thesis, University of Hawaii, Honolulu, HI, 1972.

[65] N Liu and S. Ulukus. Capacity region and optimum power control strategies for fading Gaussian multiple access channels with common data. IEEE Transactions on Communications, 54(10):1815–1826, October 2006.

[66] László Lovász. Submodular functions and convexity. In Mathematical Programming The State of the Art, pages 235–257. Springer, 1983.

[67] Angel Lozano and Nihar Jindal. Are yesterday's information-theoretic fading models and performance metrics adequate for the analysis of today's wireless systems? Communications Magazine, IEEE, 50(11):210–217, 2012.

[68] Mohammad Ali Maddah-Ali, Amin Mobasher, and Amir K Khandani. Fairness in multiuser systems with polymatroid capacity region. IEEE Transactions on Information Theory, 55(5):2128–2138, May 2009.

[69] K Marton. A coding theorem for the discrete memoryless broadcast channel. IEEE Transactions on Information Theory, 25(3):306–311, 1979.

[70] J Mietzner, R Schober, L Lampe, W Gerstacker, and P Hoeher. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. IEEE Communications Surveys & Tutorials, 11(2):87–105, April 2009.

[71] M Mohseni, R Zhang, and J M Cioffi. Optimized transmission for fading multiple-access and broadcast channels with multiple antennas. IEEE Journal on Selected Areas in Communications, 24(8):1627–1639, August 2006.

[72] C. T. Mullis. Advanced linear systems. Lecture Notes, 2009.

[73] Kazuo Murota. Discrete Convex Analysis. Monographs on Discrete Mathematics and Applications. SIAM, 2003.

[74] Chandra Nair and Abbas El-Gamal. The capacity region of a class of three-receiver broadcast channels with degraded message sets. IEEE Transactions on Information Theory, 55(10):4479–4493, 2009.

[75] Chi Kin Ngai and R W Yeung. Network coding gain of combination networks. Information Theory Workshop, 2004. IEEE, pages 283–287, 2004.

[76] L H Ozarow. The capacity of the white Gaussian multiple access channel with feedback. IEEE Transactions on Information Theory, 30(4):623–629, july 1984.

[77] C C Paige. The general linear model and the generalized singular value decomposition. Linear Algebra and its Applications, 1985.

[78] S C G Periaswamy, D R Thompson, H P Romero, and J Di. Fingerprinting radio frequency identification tags using timing characteristics. In Proceedings Workshop on RFID Security 2010 Asia, pages 73–82, February 2010.

[79] N Pippenger. Entropy and enumeration of Boolean functions. IEEE Transactions on Information Theory, 45(6):2096–2100, August 1999.

[80] V. V. Prelov. Transmission over a multiple-access channel with a special source hierarchy. Problemy Peredachi Informatsii, 20(4):3–10, 1984.

[81] Stefano Rini and Andrea J Goldsmith. A general approach to random coding for multi-terminal networks. Information Theory and Applications Workshop (ITA), 2013, pages 1–9, 2013.

[82] Stefano Rini and Andrea J Goldsmith. On the interference channel with common messages and the role of rate-sharing. Information Theory Workshop (ITW), 2013 IEEE, pages 1–5, 2013.

[83] H. Romero and M. Varanasi. Approximate Capacity of the K-user Fading MIMO MAC with Common Information. In Proc. 51st Annu. Allerton Conf.,, page forthcoming, 2013.

[84] H. Romero and M. Varanasi. Capacity of the K-user Discrete Memoryless Multiple Access Channel with Common Information and with or without State. In Proc. 51st Annu. Allerton Conf.,, page forthcoming, 2013.

[85] H Romero and M. Varanasi. MIMO Multiaccess Channels with Cooperation. in preparation, June 2014.

[86] H Romero and M. Varanasi. Order and Polymatroidal Structure in the Multiple Access Channel with Implicit Cooperation. in preparation, June 2014.

[87] H P Romero, K A Remley, D F Williams, and C-M Wang. Electromagnetic measurements for counterfeit detection of radio frequency identification cards. IEEE Transactions on Microwave Theory and Techniques, 57(5):1383–1387, May 2009.

[88] H P Romero, K A Remley, D F Williams, C-M Wang, and T X Brown. Identifying RF identification cards from measurements of resonance and carrier harmonics. IEEE Transactions on Microwave Theory and Techniques, 58(7):1758–1765, July 2010.

[89] H P Romero and M.K Varanasi. Bounds on the Capacity Region for a Class of Interference Channels With Common Information. IEEE Transactions on Information Theory, 59(8):4811–4818, August 2013.

[90] S.M. Ross. A First Course in Probability. Pearson Prentice Hall, 2010.

[91] A Salimi, T. Liu, and S Cui. Generalized Cut-Set Bounds for Broadcast Networks. arXiv.org, 2013.

[92] L Sankar, Xiaohu Shang, E Erkip, and H.V Poor. Ergodic Fading Interference Channels: Sum-Capacity and Separability. IEEE Transactions on Information Theory, 57(5):2605–2626, May 2011.

[93] R Schaefer and H Boche. Physical Layer Service Integration in Wireless Networks: Signal processing challenges. Signal Processing Magazine, 2014.

[94] Alexander Schrijver. Combinatorial Optimization Polyhedra and Efficiency. Algorithms and Combinatorics 24. Springer-Verlag, Berlin, 2003.

[95] C. E. Shannon. A mathematical model of communication. The Bell System Technical Journal, 27:379–423–623–656, October 1948.

[96] Claude Elwood Shannon. Two-way communication channels. Proc. 4th Berkeley Symp. Math. Stat. Prob, 1:611–644, 1961.

[97] O Simeone, O Somekh, G Kramer, H.V Poor, and S Shamai. Three-user Gaussian multiple access channel with partially cooperating encoders. Asilomar, pages 85–89, 2008.

[98] Osvaldo Simeone, Oren Somekh, Gerhard Kramer, H Vincent Poor, and Shlomo Shitz Shamai. Three-user Gaussian multiple access channel with partially cooperating encoders. Asilomar, pages 85–89, 2008.

[99] D Slepian and J K Wolf. Coding Theorem for Multiple Access Channels With Correlated Sources. Bell System Technical Journal, 52(7):1037–1076, 1973.

[100] Changho Suh and D.N.C Tse. Feedback Capacity of the Gaussian Interference Channel to Within 2 Bits. Information Theory, IEEE Transactions on, 57(5):2667–2685, May 2011.

[101] HH Tan. Two-user interference channels with correlated information sources*. Information and Control, 44(1):77–104, 1980.

[102] R Tandon and S Ulukus. Dependence Balance Based Outer Bounds for Gaussian Networks With Cooperation and Feedback. IEEE Transactions on Information Theory, 57(7):4063–4086, July 2011.

[103] I. Emre Telatar and D N Tse. Bounds on the capacity region of a class of interference channels. Information Theory, 2007. ISIT 2007. IEEE International Symposium on, pages 2871–2874, 2007.

[104] J Thomas. Feedback can at most double Gaussian multiple access channel capacity (Corresp.). IEEE Transactions on Information Theory, 33(5):711–716, 1987.

[105] Chao Tian. Latent Capacity Region: A Case Study on Symmetric Broadcast With Common Messages. IEEE Transactions on Information Theory, 57(6):3273–3285, 2011.

[106] D.N.C Tse and S V Hanly. Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities. IEEE Transactions on Information Theory, 44(7):2796–2815, November 1998.

[107] A.M. Tulino and S Verdu. Random Matrix Theory and Wireless Communications. Now Publishers Inc, Hanover, MA, 2004.

[108] S.Y. Tung. Multiterminal Source Coding. PhD thesis, Cornell University, School of Electrical Engineering, Ithaca, N.Y., 1978.

[109] C S Vaze and M K Varanasi. Independent Signaling Achieves the Capacity Region of the Gaussian Interference Channel With Common Information to Within One Bit. IEEE Transactions on Information Theory, 60(10):6070–6079, October 2014.

[110] S Vishwanath, N Jindal, and A Goldsmith. Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. IEEE Transactions on Information Theory, 49(10):2658–2668, October 2003.

[111] P Viswanath and David N C Tse. Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality. IEEE Transactions on Information Theory, 49(8):1912–1921, August 2003.

[112] H. S. Wang and N Moayeri. Finite-state Markov channel-a useful model for radio communication channels. Vehicular Technology, IEEE Transactions on, 44(1):163–171, 1995.

[113] I-Hsiang Wang and D.N.C Tse. Interference Mitigation Through Limited Receiver Cooperation. IEEE Transactions on Information Theory, 57(5):2913–2940, May 2011.

[114] I-Hsiang Wang and D.N.C Tse. Interference Mitigation Through Limited Transmitter Cooperation. IEEE Transactions on Information Theory, 57(5):2941–2965, May 2011.

[115] I-Hsiang Wang and D.N.C Tse. Interference Mitigation Through Limited Transmitter Cooperation. IEEE Transactions on Information Theory, 57(5):2941–2965, 2011.

[116] Hanan Weingarten, Yossef Steinberg, and Shlomo Shitz Shamai. The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. IEEE Transactions on Information Theory, 52(9):3936–3964, 2006.

[117] J E Wieselthier, Gam D Nguyen, and Anthony Ephremides. On the construction of energy-efficient broadcast and multicast trees in wireless networks. In INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, pages 585–594. IEEE, 2000.

[118] M.A Wigger and G Kramer. Three-user MIMO MACs with cooperation. In 2009 IEEE Information Theory Workshop, pages 221–225, June 2009.

[119] F Willems. The discrete memoryless multiple access channel with partially cooperating encoders (Corresp.). IEEE Transactions on Information Theory, 29(3):441–445, May 1983.

[120] A Wyner. The common information of two dependent random variables. IEEE Transactions on Information Theory, 21(2):163–179, March 1975.

[121] W Yu, W Rhee, S. Boyd, and J M Cioffi. Iterative water-filling for Gaussian vector multiple-access channels. IEEE Transactions on Information Theory, 50(1):145–152, January 2004.

[122] Rui Zhang, Shuguang Cui, and Ying-Chang Liang. On Ergodic Sum Capacity of Fading CognitiveMultiple-Access and Broadcast Channels. IEEE Transactions on Information Theory, 55(11):5161–5178, November 2009.

[123] Xin Zhang, Jun Chen, S B Wicker, and T Berger. Successive Coding in Multiuser Information Theory. IEEE Transactions on Information Theory, 53(6):2246–2254, June 2007.

Appendix A

Submodular Inequalities

The following are reproduced verbatim from [91].

Proposition A.0.1. *Let E be a finite ground set and i, j, k be three distinct integers from $\{1, 2, 3\}$. Then the following inequalities hold for any three subsets S_i, S_j, S_k of E for submodular functions. If the function is modular, then all such inequalities hold with equality.*

$$
\begin{aligned}
&f(S_i \cup S_j \cup S_k + f(S_i \cap S_j) \\
&\qquad \leq f(S_i) + f(S_j) + f(S_k) + (f(S_k) - f(S_k \cap (S_i \cup S_j)), \\
&f(S_i \cup S_j \cup S_k) + f((S_i \cap S_j) \cup (S_j \cap S_k) \cup (S_i \cap S_k)) \\
&\qquad \leq f(S_i) + f(S_j) + f(S_k) + (f(S_k) - f(S_i \cup S_j \cap S_k) \\
&f(S_i \cup S_j \cup S_k) + f(S_i \cap S_j \cap S_k) + f(S_i \cup S_j), \\
&\qquad \leq f(S_i) + f(S_j) + f(S_k) + (f(S_i) - f(S_i \cap (S_j \cap S_k))) + (f(S_j) - f(S_j \cap (S_i \cap S_k)) \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&2f(S_i \cup S_j \cup S_k) + f(S_i \cap S_j \cap S_k) \\
&\qquad \leq f(S_i) + f(S_j) + f(S_k) + (f(S_i) - f(S_i \cap (S_j \cap S_k))) \\
&\qquad\qquad + (f(S_j) - f(S_j \cap (S_i \cap S_k)) + (f(S_k) - f(S_k \cap (S_i \cap S_j)) \ .
\end{aligned}
$$

Proposition A.0.2. *Let E be a nonempty finite grounds set and Q be a subset of $[1 : K] \setminus \{1\}$.*

Define $\gamma_Q = \prod_q \in Q(q-1)$ *and take* $\beta_Q(r)$ *and* $\beta'_Q(r)$ *to be as defined in* (5.14). *Then the inequality*

$$\sum_{r=1}^{K} \beta_Q(r) f(S^{(r)}) \leq \gamma_Q \sum_{k=1}^{K} f(S_k) + \sum_{r \in Q} \sum_{k=1}^{r} \beta'_Q(r)(f(S_k) - f(S_k \cap S^{(r)})),$$

is an extremal inequality for submodular functions over any K *subsets* $(S_k : k \in [1:K])$ *of* E *and where*

$$S^{(r)} = \bigcup_{U \subseteq [1:K]:|U|=r} \left(\bigcap_{k \in U} S_k \right)$$

is the set of elements that belong to at least r*-out-of-*K *subsets from the collection* $(S_k : k \in [1:K])$.

Appendix B

DM-MAC Achievability

B.1 Achievability

To demonstrate achievability of the capacity characterization in Theorem 3.3.1, we use random coding and joint typicality decoding, where our notion of typicality is robust typicality.

B.1.1 Superposition Coding

B.1.1.1 Notion of Typicality

The notion of typicality we use is that of robust typicality [31]: with $X \sim p(x)$ as a discrete random variable with finite support $\mathcal{X}$ and with $\delta \in (0,1)$ as parameter that may be infinitesimal, a $n-$sequence $x^n \in \mathcal{X}^n$ is δ-typical with respect to the probability mass function $p(x)$ if

$$\left| \frac{|\{i : x_i = x\}|}{n} - p(x) \right| \leq \delta p(x) \quad \forall\ x \in \mathcal{X}.$$

Let $\mathcal{T}_\delta^{(n)}(X)$ be the set of all such δ-typical n-sequences. For a tuple of discrete, finite-support random variables $(X_1, \ldots, X_m) \sim p(x_1, \ldots, x_m)$, define the set of $\delta-$typical n-sequences $(x_1^n, \ldots, x_m^n)$ to be $\mathcal{T}_\delta^n(X_1, \ldots, X_m) = \mathcal{T}_\delta^n(X)$; that is, as the typical set for a single random variable $X \equiv (X_1, \ldots, X_m)$.

B.1.1.2 Encoding

Fix a message index set $E \subseteq 2^{[1:K]} \backslash \emptyset$ of cardinality M and assign to it some superposition order. Fix a typicality parameter $\delta > 0$, an auxiliary random tuple $(U_S : S \in E) \in L(E; \leq)$, and a

set of K deterministic Shannon strategies

$$x_j : \prod_{S:j\in S\in E} \mathcal{U}_S \mapsto \mathcal{X}_j \qquad j \in [1:K]\,. \tag{B.1}$$

B.1.1.3 Codebook Generation

Let $S_1, S_2, \ldots, S_M$ be an exhaustive, never-increasing listing of the members of E so that $S_i < S_j$ only if $j < i$. Let $m_B = (m_S : S \in B)$ for each subset $B \subseteq E$. We generate the codebook by successively considering the codewords for the message M_{S_1}, then M_{S_2}, and so on. First, for each $m_{S_1} \in \left[1 : 2^{nR_{S_1}}\right]$, generate an independent codeword with symbols drawn i.i.d. according to $p(u_{S_1})$, i.e. $u^n_{S_i}(m_{\uparrow S_i}) \sim \prod_{t=1}^n p(u_{S_i,t})$. Now, fix an $i > 1$ and suppose we've generated the codewords for all prior messages M_j with $j < i$. Then, for each message tuple

$$m_{\uparrow S_i\setminus S_i} \equiv (m_{S_j} : S_i < S_j) \in \prod_{j:S_i<S_j} \left[1 : 2^{nR_{S_j}}\right],$$

generate an independent codeword for each

$$m_{S_i} \in \left[1 : 2^{nR_{S_i}}\right]$$

with symbols drawn i.i.d. as

$$u^n_{S_i}(m_{\uparrow S_i}) \sim \prod_{t=1}^n p(u_{S_i,t}|u_{\uparrow S_i\setminus S_i,t}(m_{\uparrow S_i\setminus S_i})). \tag{B.2}$$

B.1.1.4 Encoding

To select which input symbol to send during the tth transmission, the jth transmitter uses the Shannon strategies (B.1):

$$x_{j,t}(m_S : j \in S \in E) = x_j\Big(u_{S,t}(m_{\uparrow S}) : j \in S \in E\Big)$$

for all $t \in [1 : n]$.

B.1.1.5 Joint typicality decoding

We decode by joint typicality with respect to the auxiliary codewords. With $m \equiv (m_S : S \in E)$, let $T(m)$ be the event that

$$(u^n_{S_1}(m_{\uparrow S_1}), \ldots, u^n_{S_M}(m_{\uparrow S_M}), Y^n) \tag{B.3}$$

is jointly typical (i.e., an element of $\mathcal{T}^{(n)}_\delta(U_{S_1}, \ldots, U_{S_M}, Y)$). Declare $\hat{m}$ to be the sent message if it is the **unique** message tuple for which $T(m)$ occurs. If no such message tuple exists or more than one such message tuple exists, declare an error.

B.1.1.6 Analysis of Error

By the symmetry of the code construction, the average probability of error over all codebooks and all codewords is equal to the average probability of error over all codebooks that just the message tuple with $m_S = 1$ for all $S \in E$ was sent (which can succinctly denote with the shorthand $m \equiv (m_S : S \in E) = \mathbf{1}$).

Consider a candidate message tuple $(\hat{m}_S : S \in E)$ with

$$\begin{aligned} \hat{m}_S &\neq 1 \quad \text{for } S \in B \\ \hat{m}_S &= 1 \quad \text{for } S \notin B \end{aligned} \tag{B.4}$$

for some $B \subseteq E$. By the encoding dependency (B.2), the output Y^n would be statistically dependent on codewords $u^n_{S_k}(\hat{m}_{\uparrow S_k})$ with $\uparrow S_k \subseteq E \backslash B$, but would **not** be statistically dependent on any codeword $u^n_{S_k}(\hat{m}_{\uparrow S_k})$ with $\uparrow S_k \not\subseteq E \backslash B$. Hence, the distribution that the random tuple in (B.3) obeys for any such candidate message tuple is

$$\left(\prod_{t=1}^{n} p\Big(Y_t \Big| (u_{S_k,t}(\hat{m}_{\uparrow S_k}) : S_k \in \mathcal{Z}^c_{\mathcal{F}_\downarrow}(B)\Big)\right) p\Big(u_{S_k,t}(\hat{m}_{\uparrow S_k}) : S_k \in E\Big)$$

where

$$\mathcal{Z}^c_{\mathcal{F}_\downarrow}(B) = \bigcup \{\uparrow S_k : \uparrow S_k \subseteq E \backslash B)\} = E \backslash \bigcup \{\downarrow S_k : S_k \in B\} = E \backslash \mathcal{Z}_{\mathcal{F}_\downarrow}(B)$$

is the **largest** element of the up-set lattice $\mathcal{F}_\uparrow$ contained in B and $\mathcal{Z}_{\mathcal{F}_\downarrow}(B)$ as the **smallest** element of the down-set lattice family $\mathcal{F}_\downarrow$ containing B. Hence, by the joint typicality lemma [31], there is an ϵ_δ (with $\epsilon_\delta \to 0$ as $\delta \to 0$) for which

$$P(T(\hat{m})) \leq 2^{-n(I(U_{\mathcal{Z}_{\mathcal{F}_\downarrow}(B)};Y|U_{\mathcal{Z}^c_{\mathcal{F}_\downarrow}(B)})-\epsilon_\delta)}$$

By contrast, the law of large numbers assures that the codewords corresponding to the true message will be jointly typical with high probability with the received vector; more precisely, $P(T(\mathbf{1})) \to 1$ as $n \to \infty$. With the union bound,

$$\begin{aligned}
P(\text{Error} \cap T(\mathbf{1})) &= P(\cup_{\hat{m}\neq\mathbf{1}} T(\hat{m})) \\
&\leq \sum_{\hat{m}\neq\mathbf{1}} P(T(\hat{m})) \\
&= \sum_{B\subseteq E} \sum_{\hat{m}:\text{(B.4)}} P(T(\hat{m})) \\
&\leq \sum_{B\subseteq E} \sum_{\hat{m}:\text{(B.4)}} 2^{-n(I(U_{\mathcal{Z}_{\mathcal{F}_\downarrow}(B)};Y|U_{\mathcal{Z}^c_{\mathcal{F}_\downarrow}(B)})-\epsilon_\delta)} \\
&\leq \sum_{B\subseteq E} 2^{-n(I(U_{\mathcal{Z}_{\mathcal{F}_\downarrow}(B)};Y|U_{\mathcal{Z}^c_{\mathcal{F}_\downarrow}(B)})-\sum_{S\in B} R_S-\epsilon_\delta)}
\end{aligned}$$

Hence the described encoding scheme achieves a vanishing probability of error if

$$\sum\nolimits_{S\in B} R_S \leq I(U_{\mathcal{Z}_{\mathcal{F}_\downarrow}(B)};Y|U_{\mathcal{Z}^c_{\mathcal{F}_\downarrow}(B)}) \quad \text{for all } B \subseteq E.$$

If $B \notin \mathcal{F}_\downarrow$, then B is a **strict** subset of $B' = \mathcal{Z}_{\mathcal{F}_\downarrow}(B)$ and hence its corresponding inequality is redundant (as $R_S \geq 0$ for all $S \in E$) given the corresponding inequality for B', which is an element of the down-set lattice family $\mathcal{F}_\downarrow$. Hence there is still a vanishing probability of error if

$$\sum\nolimits_{S\in B} R_S \leq I(U_B;Y|U_{E\setminus B}) \qquad \text{for all } B \in \mathcal{F}_\downarrow.$$

B.2 Achievability via Rate-Delegation

By Theorem 3.3.1, we know that for a fixed message index set E and superposition order $\leq$ on E, coding recursively along the principal up-sets of $\mathcal{F}$ provides that any rate tuple in $\mathcal{P}_{\mathcal{F}_\downarrow(E;\leq)}(\rho_{x,U})$

is achievable for a fixed auxiliary tuple $U \in L(E;\leq)$ and set of K Shannon strategies $x_1, \ldots, x_K$. However, as indicated in the two-user case, it may be possible to achieve a larger polytope by concatenating the achievability step of Theorem (3.3.1) with rate load delegation of "more common" message rates onto "less common" message rates as in (3.23) and (3.24).

More precisely, by introducing the non-negative **rate-splits**

$$r \equiv (r_{(S',S)} : S', S \in E \text{ and } S' \subseteq S), \tag{B.5}$$

we may achieve any target rate

$$R_S = \sum\nolimits_{S' \in E : S' \subseteq S} r_{(S',S)} \quad S \in E, \tag{B.6}$$

where common messages delegate part of their rate load to less common messages, if the reconstructed rates

$$\tilde{R}_{S'} = \sum\nolimits_{S \in E : S' \subseteq S} r_{(S',S)} \quad S' \in E \tag{B.7}$$

satisfy

$$\sum_{S' \in B} \tilde{R}_{S'} \leq \rho_U(B) \qquad \text{for } B \in \mathcal{F}_{\downarrow}(E :\leq), \tag{B.8}$$

as the resultant rates $\tilde{R}_{S'}$ may then be achieved by coding recursively along the principal up-sets elements of $\mathcal{F}_{\downarrow}(E :\leq)$ per the results of Theorem 3.3.1.

To determine what target rates $R \equiv (R_S : S \in E)$ are achievable in this manner, we must the project polytope constraint (B.8) defined in the high-dimensional space of the rate-splits r onto the much lower-dimensional space in which the target rates lie. A general procedure for such a projection is the iterative Fourier-Motkzin elimination procedure. However, such a procedure is generally **not** scalable to arbitrary dimensions: each step of Fourier-Motkzin may exponentially increase the number of resultant inequalities to keep track of. In our setting, where there are combinatorially many rate splits to deal with, such an approach is unwieldy. Rather than attempt to use Fourier-Motkzin, we instead find that an alternative approach leveraging polymatroidal properties which is tractable and scalable to the general K-user case.

Our approach relies on the dual description of convex sets (i.e. the characterization of a convex set's boundary) and the observation that the convex constraint (B.8) is polymatroidal not only in the reconstructed rates, but also directly in the rate splits. With this framework, we demonstrate the following

Lemma B.2.1. *Let $\mathcal{R} \subset \mathbb{R}^E$ be the region consisting of all non-negative target rates for which there exists a non-negative rate split as in* (B.5) *satisfying* (B.6)-(B.8)*. Then $\mathcal{R} = \mathcal{P}_{\mathcal{F}_\downarrow(E;\subseteq)}(\rho_U)$, the polymatroid defined over E and constrained by the down-sets of E under the inclusion order $\subseteq$.*

Proof. For the sake of brevity, let $\mathcal{F} = \mathcal{F}_\downarrow(E;\leq)$ and $\mathcal{F}_\downarrow = \mathcal{F}_\downarrow(E;\subseteq)$. As the constraints for $\mathcal{R}$ are linear, it must be convex. Hence, by convexity, it is fully characterized by the maximal values of the collection of linear programs

$$\text{maximize } \mu^T R \qquad\qquad \text{subject to } \tilde{R} \in \mathcal{P}_{\mathcal{F}}(\rho_U) \tag{B.9}$$

corresponding to each choice of $\mu \in \mathbb{R}^E$.

To solve each such linear program, we first express the optimization problem directly in terms of the rate splits. To describe this, we first introduce convenient notation. Let

$$E_r = \{(S', S) : S', S \in E, S' \subseteq S\}$$

be the index set of all rate-splits (B.5). Enumerate the message index set as $E = \{S_1, \ldots, S_M\}$ so that $\mu_{S_1} \geq \cdots \mu_{S_k} > 0 \geq \mu_{S_{k+1}} \geq \cdots \geq \mu_M$ and the rate-split message index set E_r as

$$\{(S', S)_1, \ldots, (S', S)_{n_1}, \ldots, (S', S)_{n_{M-1}+1}, \ldots, (S', S)_{n_M}\}$$

where sequence of numbers $0 = n_0 \leq n_1 \leq \ldots \leq n_M$ are defined implicitly so that

$$\{(S', S)_{n_{i-1}+1}, \ldots, (S', S)_{n_i}\} = \{(S', S_i) : S' \in \downarrow S_i\}, \tag{B.10}$$

for each $i \in [1 : M]$, where we remind the user of the down-set notation (see Table 2.4), taken with respect to the message index set E with the inclusion order $\subseteq$. Then the index $j \in [1 : n_M]$

enumerates all of the rate-splits. For each $j \in [1 : n_M]$, let $\nu_j = \mu_i$ when $n_i < j \leq n_{i+1}$ so that we may write

$$\sum_{i=1}^{M} \mu_{S_i} R_{S_i} = \sum_{i=1}^{M} \mu_{S_i} \left(\sum_{j=n_i}^{n_{i+1}-1} R_{(S',S)_j} \right) = \sum_{j=1}^{n_M} \nu_j R_{(S',S)_j}.$$

Define the one-to-one map $\mathcal{Z} : \mathcal{F} \mapsto \mathcal{F}_r$ via

$$\mathcal{Z}(B) = \{(S, S') : S' \in \uparrow S \text{ for some } S \in B\}$$

between subsets of the index set of rate splits E_r and the index set of messages E. Notably, $\mathcal{Z}$ is a lattice homomorphism (recall Definition 4) and hence its image $\mathcal{F}_r$ is also a lattice set family (but over the rate-split index set E_r rather than the message index set E). For each $B_r \in \mathcal{F}_r$, define the map $\rho'_U : \mathcal{F}_r \mapsto \mathbb{R}_+$ with

$$\begin{aligned} \rho'_U(B_r) &= \rho_U \circ \mathcal{Z}^{-1}(B_r) \\ \mathcal{Z}^{-1}(B_r) &= \left\{ S' : \text{there is some } S \in E \text{ with } (S', S) \in Z \right\}. \end{aligned}$$

In particular, as the inverse of a one-to-one lattice homomorphism is again a lattice homomorphism, we know that ρ'_U is again a polymatroid function (but over $\mathcal{F}_r$ rather than $\mathcal{F}$).

Hence, when the polytope (B.8) is written in terms of the rate splits directly, we recognize it as a polymatroid over the lattice set family $\mathcal{F}_r$. So, the initial linear program (B.9) is equivalent to

$$\text{maximize } \nu^T r \qquad \text{subject to } r \in \mathcal{P}_{\mathcal{F}_r}(\rho'_U). \tag{B.11}$$

By polymatroidal properties, a maximizing rate point is, with S'_j as the first index of the pair $(S', S)_j$,

$$\begin{aligned} r_{(S',S)_1} &= \rho'_U \circ \mathcal{Z}_{\mathcal{F}_r}(\{(S', S)_1\}) = \rho_U(\{S'_1\}) \\ r_{(S',S)_j} &= \rho'_U \circ \mathcal{Z}_{\mathcal{F}_r}(\{(S', S)_1, \ldots, (S', S)_j\}) - \rho'_U \circ \mathcal{Z}_{\mathcal{F}_r}(\{(S', S)_1, \ldots, (S', S)_{j-1}\}) \\ &= \rho_U \left(\bigcup_{k=1}^{j} \{S'_k\} \right) - \rho_U \left(\bigcup_{k=1}^{j-1} \{S'_k\} \right) \qquad 2 \leq j \leq n_k \\ r_{(S',S)_j} &= 0 \qquad n_{k+1} \leq j \leq n_M. \end{aligned}$$

where $\mathcal{Z}_{\mathcal{F}_r}(B_r)$ is the smallest element of lattice set family $\mathcal{F}_r$ containing $B_r \subseteq E_r$ and $\mathcal{Z}_{\mathcal{F}}(B)$ is the smallest element of $\mathcal{F}$ containing $B \subseteq E$. The simplifications above follow by the relation

$$\mathcal{Z}^{-1} \circ \mathcal{Z}_{\mathcal{F}_r}(B_r) = \left\{ S' : \exists S \in E \text{ with } (S', S) \in B_r \right\}.$$

Then by the rate delegation formula (B.6) and the enumeration (B.10) of the rate-split index set E_r, we have that for $m > k$, $R_{S_m} = 0$, while for $m \leq k$

$$R_{S_1} + \cdots + R_{S_m} = \sum\nolimits_{j=1}^{n_m} r_{(S',S)_j} = \rho_U \left(\cup_{i=1}^{m} \downarrow S_i \right) = \rho_U \circ \mathcal{Z}_{\mathcal{F}_\downarrow}(\{S_1, \ldots, S_m\}),$$

as $\bigcup_{i=1}^{m} \downarrow S_i \in \mathcal{F}_\downarrow$. Abbreviating $\mathcal{Z}_{\mathcal{F}_\downarrow}(B)$ to $\mathcal{Z}_\downarrow(B)$, the above is simply that

$$\begin{aligned}
R_{S_1} &= \rho_U \circ \mathcal{Z}_\downarrow(\{S_1\}) \\
R_{S_m} &= \rho_U \circ \mathcal{Z}_\downarrow(\{S_1, \ldots, S_m\}) - \rho_U \circ \mathcal{Z}_\downarrow(\{S_1, \ldots, S_{m-1}\}) && 2 \leq m \leq k \\
R_{S_m} &= 0 && k+1 \leq m \leq M.
\end{aligned}$$

Hence the rate region attainable by first rate-delegating and then coding with respect to $\mathcal{F}$ has exactly the same supporting hyperplanes as the polymatroid $\mathcal{P}_{\mathcal{F}_\downarrow}(\rho_U)$. As two convex regions with the same set of supporting hyperplanes are equal, we conclude that $\mathcal{R} = \mathcal{P}_{\mathcal{F}_\downarrow}(\rho_U)$. □

Remark Notably, we may delegate common message rates to less common message rates and subsequently code without superposition to achieve the same capacity region as would be achieved had we simply coded with superposition. This equivalence has also been noted in other contexts; for example, in the interference channel with the recent result of [82]. These results suggest that in general, rate delegation followed by with coding without superposition achieves the same rate region as coding with superposition.

B.3 Converse

Suppose that there exists a sequence of codes indexed by block length n which communicate at a rate tuple $(R_S : S \in E)$ and which achieve a vanishing probability of error as n tends to infinity. Pick the code corresponding to block length n and and let M_S be uniformly distributed on

$\left[1 : 2^{nR_S}\right]$ for each $S \in E$. Let $X_1^n, \ldots, X_K^n, Y^n$ be the random variables induced by the messages, the encoders, and the channel.

Let $\leq$ be any superposition order on E and let $\mathcal{F}_\downarrow$ be the associated down-set lattice family. By Fano's inequality we may conclude that for any $B \in \mathcal{F}_\downarrow$,

$$n\left(\sum_{S\in B} R_S\right) \leq I(M_B; Y) + n\epsilon_n$$

for some non-negative vanishing sequence $\epsilon_n \to 0$. Hence,

$$\begin{aligned}
n\left(\sum_{S\in B} R_S - \epsilon_n\right) &\leq I(M_B; Y^n | M_{E\setminus B}) \\
&= H(Y^n | M_{E\setminus B}) - H(Y^n | M_E) \\
&= \sum_{t=1}^n H(Y_t | M_{E\setminus B}, Y^{t-1}) - \sum_{t=1}^n H(Y_t | X_{1t}, \ldots, X_{Kt}) \\
&\leq \sum_{t=1}^n H(Y_t | M_{E\setminus B}) - \sum_{t=1}^n H(Y_t | X_{1t}, \ldots, X_{Kt}) \\
&= n\left(H(Y_Q | U_{E\setminus B}, Q) - H(Y_Q | X_{1Q}, \ldots, X_{KQ})\right) \\
&\leq n\left(H(Y_Q | U_{E\setminus B}) - H(Y_Q | X_{1Q}, \ldots, X_{KQ})\right) \\
&= nI(U_B, Y | U_{E\setminus B}).
\end{aligned}$$

where $U_S = M_S$ for each $S \subseteq [1 : K]$, $Q \sim \text{Uniform}([1 : n])$, $Y = Y_Q$, and $X = X_Q$. Hence, $(R_S - \epsilon_n)_{S\in E} \in \mathcal{P}_{\mathcal{F}_\downarrow}(\rho_U)$. Moreover, by the joint independence of $(U_S : S \in E)$, $U \equiv (U_S : S \in E) \in L(E; =) \subseteq L(E; \leq)$. By definition of the set of feasible encoders (see Table 2.2), there are K deterministic functions $x_j(\cdot)$ (one for each $j \in [1 : K]$) such that

$$X_j = x_j((U_S : j \in S \in E)).$$

Appendix C

Gaussian MAC

C.1 Successive Group Decoding Proof

Recall Lemma 3.4.2, and the discussion immediately following. Define, per collection of $\mathbf{K}_E = (\mathbf{K}_S : S \in E)$, $\rho(\mathbf{K}_E; \cdot) : 2^E \mapsto \mathbb{R}_+$ with

$$\rho(\mathbf{K}_E; B) = \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S\in B} \mathbf{P}_S \mathbf{K}_S \mathbf{P}_S^*\right)\mathbf{H}^*\right), \tag{C.1}$$

for each $B \subseteq E$, which is normalized, nondecreasing, and submodular as a result of Lemma 3.2.1. It suffices to show the following for the converse.

Lemma C.1.1. *The K-user MIMO MAC with message index set E is dominated by rate points that satisfy*

$$\sum_{S\in B_i} R_S = \rho(\mathbf{K}_E; B_i) \text{ for all } i \in [1:k] \tag{C.2a}$$

$$\sum_{S\in B'} R_S \leq \rho(\mathbf{K}_E; B') \text{ for all } B' \in \mathcal{F}_{\downarrow}(E; \subseteq) \text{ satisfying}$$

$$B_i \subseteq B' \subseteq B_{i+1} \text{ for some } i \in [1:k-1]\,. \tag{C.2b}$$

for some admissible set of covariances and some Hierarchical Decoding chain $\{B_1, \ldots, B_k\}$.

Proof. By Theorem 4.3, a rate-tuple is achievable only if it is in a polymatroid

$$\left\{R \in \mathbb{R}_+^E : \sum_{S\in B} R_S \leq \rho(\mathbf{K}_E; B) \;\forall\; B \in \mathcal{F}_{\downarrow}(E; \leq)\right\}$$

for some admissible set of covariance matrices $\mathbf{K}_e$. By the greedy algorithm for maximizing the weighted sum-rate over polymatroids, all rate points in the polymatroid above are dominated by those that satisfy $\sum_{S\in E} R_S = \rho(\mathbf{K}_E; E)$. Pick one such dominating rate tuple R'. We proceed by induction.

Root Case By assumption, (C.2) is satisfied with respect to the length-2 chain $\emptyset = B_1 \subset B_2 = E$ and the family of down-sets $\mathcal{F}_\downarrow(E;\leq)$.

Inductive Step Assume that (C.2) holds with respect to a length-k chain $\emptyset = B_1 \subset B_2 \subset \cdots \subset B_k = E$ of the down-set lattice $\mathcal{F}_\downarrow(E;\leq)$ that is **not** a hierarchical decoding chain.

Fix i to be the smallest index where $G_i = B_{i+1}\backslash B_i$ violates (3.25). Then there is a pair $(S, S') \in G_i$ with $S \subset S'$ and no $S'' \in G_i$ with $S'' \subset S$. For this pair, set $B'_1 = B_i \cup \{S\} \in \mathcal{F}_\downarrow(E;\leq)$ and let $B'_2, \ldots, B'_m$ be the list of any remaining subsets $B \in \mathcal{F}_\downarrow(E;\leq)$ containing S, but not containing S', and satisfying $B_i \subset B \subset B_{i+1}$. There are two cases

(1) If $R(B'_l) = \rho(\mathbf{K}_E; B'_l)$ for some B'_l, continue to the next inductive step with the chain $B_1, \ldots, B_i, B'_l, B_{i+1}, \ldots, B_k$ in place of $\{B_1, \ldots, B_k\}$.

(2) Suppose $R(B'_l) < \rho(\mathbf{K}_E; B'_l)$ for all $l \in [1:m]$. For each $\epsilon_{S,S'} \in [0,1]$, consider the **covariance** split

$$\begin{aligned}
\mathbf{K}'_S &= \epsilon_{S,S'}\mathbf{K}_S \\
\mathbf{K}'_{S'} &= (1-\epsilon_{S,S'})\mathbf{P}_{S'}^T\mathbf{P}_S\mathbf{K}_S\mathbf{P}_S^T\mathbf{P}_{S'} + \mathbf{K}_{S'} \\
\mathbf{K}'_{S''} &= \mathbf{K}_{S''} \quad \forall S'' \in E\backslash\{S, S'\}.
\end{aligned}$$

Observe that for each $k \in [1:m]$, each

$$\begin{aligned}
\rho(\mathbf{K}'_E; B'_l) &= \log\det\left(\mathbf{I} + \mathbf{H}\left(\sum_{S''\in B'_l} \mathbf{P}_{S''}\mathbf{K}_{S''}\mathbf{P}_{S''}^*\right)\mathbf{H}^*\right), \\
&= \log\det\left(\mathbf{I} + \mathbf{H}\left(\epsilon_{S,S'}\mathbf{P}_S\mathbf{K}_S\mathbf{P}_S^* + \sum_{S''\in B'_l\backslash\{S\}} \mathbf{P}_{S''}\mathbf{K}_{S''}\mathbf{P}_{S''}^*\right)\mathbf{H}^*\right),
\end{aligned}$$

is a continuous function of $\epsilon_{S,S'}$ onto the interval $[\rho(\mathbf{K}_E; B'_l \backslash \{S\}), \rho(\mathbf{K}_E; B'_l)]$. Moreover, as

$$\begin{aligned} R_S &= \sum_{S \in B'_1} R_S - \sum_{S \in B_i} R_S \\ &< \rho(\mathbf{K}'_E; B'_1) - \rho(\mathbf{K}'_E; B_i) \\ &= \log\det\left(\mathbf{I} + \epsilon_{S,S'}\mathbf{H}\Big(\mathbf{P}_S\mathbf{K}_S\mathbf{P}_S^*\Big)\mathbf{H}^*\left(\mathbf{I} + \mathbf{H}\Big(\sum_{S'' \in B_i} \mathbf{P}_{S''}\mathbf{K}_{S''}\mathbf{P}_{S''}^*\Big)\mathbf{H}^*\right)^{-1}\right) \end{aligned} \tag{C.3}$$

where the right hand side is a continuous function of $\epsilon_{S,S'}$ onto the interval $[0, \rho(\mathbf{K}_E; B_i \cup \{S\}) - \rho(\mathbf{K}_E; B_i)]$, we know by the intermediate value theorem that there exists a $\epsilon_{S,S'} > 0$ such that for some $l \in [1:m]$,

$$\sum_{S \in B'_l} R_S = \rho(\mathbf{K}'_E; B'_l)$$

$$\sum_{S \in B'_j} R_S \leq \rho(\mathbf{K}'_E; B'_j) \qquad l \neq j \in [1:m]$$

Continue to the next inductive step with $\mathbf{K}'_E$ in place of $\mathbf{K}_E$ and $\{B_1, \ldots, B_i, B'_l, B_{i+1}, \ldots, B_k\}$ in place of $\{B_1, \ldots, B_k\}$.

□

Corollary C.1.2. *The K-user MIMO MAC with message index set E is dominated by rate points that are contained in a polytope*

$$\left\{R \in \mathbb{R}_+^E : \sum_{S \in B} R_S \leq \rho(\mathbf{K}'_E; B) \ \forall \ B \subseteq E\right\}.$$

for some admissible set of covariance matrices.

Proof. Follows by the Lemmas 3.4.2 and C.1.1. □

C.2 Fading MAC Converse

Let $E \subseteq 2^{[1:K]}$ be a family of subsets closed under intersection which index the message sources to be transmitted by the K-users over the channel (4.18). In our converse, recall the

following result from Section 4.3.3.1:

Lemma C.2.1 (Entropy Maximization subject to Markov and Covariance constraints)**.** *Suppose* $(U_S : S \in E)$ *is a tuple of independent random variables such that*[1] $\mathbf{X}_j \leftarrow (U_S : j \in S \in E)$*, where* $\mathbf{X}_j \in \mathbb{C}^{t_j \times 1}$*, for each* $j \in [1:K]$*. Suppose further that* $\mathbf{Y} = \sum_{j=1}^{K} \mathbf{H}_j \mathbf{X}_j + \mathbf{Z} \in \mathbb{C}^{r \times 1}$ *with conformable matrices* $\{\mathbf{H}_j\}$ *and with* $\mathbf{Z}$ *as a circularly symmetric Gaussian vector with identity covariance.*

Then there exists a set of independent, jointly Gaussian random vectors $\{\mathbf{V}_S^G\}, \{\mathbf{X}_j^G\}$ *satsifying*

- *For each* $S \in E$*,* $\mathbf{V}_S^G \in \mathbb{C}^{(\sum_{j \in S} t_j) \times 1}$ *with* $|S|$ *partitions* $\mathbf{V}_{S,j}^G \in \mathbb{C}^{t_j \times 1}$*, indexed by* $j \in S$*.*
- $\mathbf{X}_j^G = \sum_{j \in S \in E} \mathbf{V}_{S,j}^G$ *for each* $j \in [1:K]$*.*
- $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_j) = \mathrm{Cov}(\mathbf{X}_j^G, \mathbf{X}_j^G)$ *for each* $j \in [1:K]$*.*

such that with $\mathbf{Y}^G = \sum_{j=1}^{K} \mathbf{H}_j \mathbf{X}_j^G + \mathbf{Z}$*,*

$$h(\mathbf{Y}|U_S : S \in E \backslash B) \leq h(\mathbf{Y}^G|\mathbf{V}_S^G : S \in E \backslash B) \leq \log(2\pi)^r \left| \mathbf{I} + \sum_{S \in B} \mathbf{H}_S \mathbf{Q}_S \mathbf{H}_S^* \right|$$

for each down-set $B \in \mathcal{F}_{\downarrow}$*. Here, for each* $S \in E$*,* $\mathbf{Q}_S$ *is the covariance of* $\mathbf{V}_S^G$ *and* $\mathbf{H}_S$ *is the conformable concatenation of the channel matrices* $\{\mathbf{H}_j : j \in S\}$ *such that* $\mathbf{Y}^G = \sum_{S \in E} \mathbf{H}_S \mathbf{V}_S^G + \mathbf{Z}$*.*

Suppose there exists a sequence of codes indexed by block length n which communicate at a rate tuple $(R_S : S \in E)$ and which achieve a vanishing probability of error as n tends to infinity. Fix the block length to be n and choose the code from this sequence corresponding to this block length. Let M_S uniformly distributed on $\left[1 : 2^{nR_S}\right]$, independently of the other messages M_S' for each $S \in E$.

With Fano's inequality, we know that for any $B \in \mathcal{F}_{\downarrow}$,

$$\sum_{S \in B} R_S \leq H(M_S : S \in E; \mathbf{Y}^n) \leq n\epsilon_n$$

[1] Recall that $A \leftarrow B$ denotes that A is a deterministic function of B.

for some vanishing sequence ϵ_n. Then, by standard manipulations we have that for each $B \in \mathcal{F}_{\downarrow}$,

$$\begin{aligned} n\left(\sum_{S\in B} R_S - \epsilon_n\right) &\leq I((M_S : S \in B); \mathbf{Y}^n) \\ &\leq I((M_S : S \in B); \mathbf{Y}^n, (M_S : S \in E\backslash B), \mathbf{H}^n) \\ &= I((M_S : S \in B); \mathbf{Y}^n | (M_S : S \in E\backslash B), \mathbf{H}^n) \\ &= \sum_{t=1}^{n} I((M_S : S \in B); \mathbf{Y}_t | (M_S : S \in E\backslash B), \mathbf{Y}^{t-1}, \mathbf{H}^n) \\ &\leq \sum_{t=1}^{n} h(\mathbf{Y}_t | (M_S : S \in E\backslash B), \mathbf{H}^n) - h(\mathbf{Z}_t) \\ &= \sum_{t=1}^{n} \int_{\mathcal{H}} E\left[h(\mathbf{Y}_t | M_{\{E\backslash B\}}, \mathbf{H}^n, \mathbf{H}(t) = \mathbf{H}(\nu))\right] \mathrm{d}F(\nu) - r\log(2\pi). \end{aligned} \tag{C.4}$$

where the expectation is over all channel state sequences $\mathbf{H}^n$ such that $\mathbf{H}(t) = \mathbf{H}(\nu)$ and we adopted the shorthand $M_{\{E\backslash B\}} = (M_S : S \in E\backslash B)$. We now apply Lemma C.2.1 for each instantiation of the channel sequence $\mathbf{H}^n$, where we denote the corresponding covariance matrices by $\mathbf{Q}_S(t, \mathbf{H}^n, \nu)$. Then for the tth channel use and each possible instantiation the possible channel state $\mathbf{H}(t) = \mathbf{H}(\nu)$,

$$\begin{aligned} &E\left[h(\mathbf{Y}_t | (M_S : S \in E\backslash B), \mathbf{H}^n, \mathbf{H}(t) = \mathbf{H}(\nu))\right] - r\log(2\pi) \\ &\leq E\left[\log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{Q}_S(t, \mathbf{H}^n, \nu)\mathbf{H}_S^*(\nu)\right)\right] \\ &\overset{(i)}{\leq} \log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)E[\mathbf{Q}_S(t, \mathbf{H}^n, \nu)]\mathbf{H}_S^*(\nu)\right) \\ &\overset{(ii)}{\leq} \log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{Q}_S(t, \nu)\mathbf{H}_S^*(\nu)\right) \end{aligned}$$

where (i) follows by the concavity of $\log\det(\mathbf{I} + \mathbf{X})$ over $\mathbf{X} \succeq \mathbf{0}$ and (ii) follows by defining $\mathbf{Q}_S(t, \nu) = E[\mathbf{Q}_S(t, \mathbf{H}^n, \nu)]$. Hence, continuing from (C.4), we have

$$\begin{aligned} n\left(\sum\nolimits_{S\in B} R_S - \epsilon_n\right) &\leq \sum_{t=1}^{n} \int_{\mathcal{H}} \log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{Q}_S(t, \nu)\mathbf{H}_S^*(\nu)\right) \mathrm{d}F(\nu) \\ &\overset{(iii)}{=} \int_{\mathcal{H}} \sum_{t=1}^{n} \log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{Q}_S(t, \nu)\mathbf{H}_S^*(\nu)\right) \mathrm{d}F(\nu) \\ &\overset{(iv)}{\leq} \int_{\mathcal{H}} n\log\det\left(\mathbf{I} + \sum_{S\in B} \mathbf{H}_S(\nu)\mathbf{Q}_S(\nu)\mathbf{H}_S^*(\nu)\right) \mathrm{d}F(\nu), \end{aligned}$$

where (iii) follows by the stationarity of the channel state sequence and (iv) follows by the concavity of $\log\det(\mathbf{I}+\mathbf{X})$ and by the definition $\mathbf{Q}_S(\nu)=\frac{1}{n}\sum_{t=1}^{n}E[\mathbf{Q}_S(t,\mathbf{H}^n,\nu)]$.

By assumption, the achievable code respects the average power constraint and hence

$$\int_{\mathcal{H}}\sum_{j\in S\in E}\mathrm{tr}(\mathbf{Q}_{S,jj}(\nu))\le P_j \qquad \forall j\in[1:K]\,.$$

In summary, any achievable rate is no more than an ϵ_n outside of $\mathcal{C}_f(\{\mathbf{H}_k(\nu)\},\{\mathbf{Q}_S(\nu)\})$ for a collection of covariance allocations $\{\mathbf{Q}_S(\nu)\}$ which satisfy the power constraint (4.19). As this ϵ_n can be taken to be arbitrarily small, we conclude that the region in Theorem 4.3 is an outer bound to the capacity region of the fading three-user vector Gaussian MAC with common information.

C.3 Positivity Condition

We quote a result in [72].

Lemma C.3.1. *Let* $\mathbf{Q}=\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^* & \mathbf{D}\end{bmatrix}$. *Then* $\mathbf{Q}\succeq\mathbf{0}$ *iff*[2]

$$\{\mathbf{A}\succeq\mathbf{0},\ \mathbf{D}\succeq\mathbf{B}^*\mathbf{A}^{+}\mathbf{B},\ \mathrm{Null}(\mathbf{A})\subseteq\mathrm{Null}(\mathbf{B}^*)\}.$$

Proof. • Let $\mathbf{Q}\succeq\mathbf{0}$. If $\mathbf{x}\in\mathrm{Null}(\mathbf{A})$, then

$$\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix}^*\mathbf{Q}\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix}=\mathbf{x}^*\mathbf{A}\mathbf{x}=\mathbf{0}\implies\mathbf{0}=\mathbf{Q}\begin{bmatrix}\mathbf{x}\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{A}\mathbf{x}\\ \mathbf{B}^*\mathbf{x}\end{bmatrix}$$

and so $\mathbf{x}\in\mathrm{Null}(\mathbf{B}^*)$. Moreover

$$\mathbf{0}\preceq\begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}^*\mathbf{Q}\begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}=\mathbf{A}$$

$$\mathbf{0}\preceq\begin{bmatrix}-\mathbf{A}^{+}\mathbf{B}\\ \mathbf{I}\end{bmatrix}^*\mathbf{Q}\begin{bmatrix}-\mathbf{A}^{+}\mathbf{B}\\ \mathbf{I}\end{bmatrix}=\mathbf{D}-\mathbf{B}^*\mathbf{A}^{+}\mathbf{B}.$$

[2] $\mathbf{A}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{A}$.

- If $\text{Null}(\mathbf{A}) \subseteq \text{Null}(\mathbf{B}^*)$, then

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{B}^*\mathbf{A}^+ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{B}^*\mathbf{A}^+\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^+\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \mathbf{Q}.$$

  If $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{D} \succeq \mathbf{B}^*\mathbf{A}^+\mathbf{B}$ as well, then $\mathbf{Q} \succeq \mathbf{0}$.

□

Appendix D

DM Broadcast Cannel

D.1 Generalized cut-set Bound Framework

We only outline the proof, as the details left out mimic precisely the proof of Theorem 1 in [91]. The adaptation to our Gaussian setting depends critically the properties (P1)-(P3), Notably, the factorization can be thought of as providing a set of parallel channels $\tilde{\mathbf{Y}}_{j,S}$ indexed by the message index set E. By the rank condition that

$$\text{rank}(\tilde{\mathbf{Y}}_S) = \text{rank}(\tilde{\mathbf{Y}}_{j,S}) \qquad \forall j \in S,$$

we may think of the channel provided by $\tilde{\mathbf{Y}}_S$ a corruption-free symbol-pipe that has the same DoF-rate to all of the $|S|$ receivers listed in S. In proving our outer bounds, we will provide not only the receiver output

$$\tilde{\mathbf{Y}}_j = (\tilde{\mathbf{Y}}_{j,S} : S \in E \text{ with } j \in S),$$

but also the side information

$$(\tilde{\mathbf{Y}}_{i,S} : i \neq j, S \in E, \{i,j\} \subseteq S).$$

Proof. Suppose that $(R_S : S \in E)$ is an achievable rate tuple. Then, necessarily, by Fano's inequality we may conclude that

$$H(M_{\cup_{j\in A}\uparrow\{j\}}|\tilde{\mathbf{Y}}^n_{\cup_{j\in A}\uparrow\{j\}}) \leq n\epsilon_n$$

for all subsets $A \subseteq [1:K]$. For brevity, define the set functions

$$H_{M,\tilde{\mathbf{Y}}^n}(A) = H(M_A, \tilde{\mathbf{Y}}_A^n) \qquad H_M(A) = H(M_A)$$
$$H_{\tilde{\mathbf{Y}}^n}(A) = H(\tilde{\mathbf{Y}}_A^n) \qquad C(A) = \sum_{S \in A} \nu_S.$$

It is well known that entropy, in general is submodular. Moreover, by the independence of the messages, we know that $H_M(A) \sum_{S \in A} H(M_S)$, thus providing that H_M is not only submodular, but is also modular. By definition, $C(A)$ is modular. By the independence bound on the joint entropy of random variables, we have that independence bound on the joint entropy of a collection of random variables,

$$\begin{aligned} H_{\tilde{\mathbf{Y}}^n}(A) &\leq \sum_{S \in A} H(\tilde{\mathbf{Y}}_S^n) \\ &\leq \sum_{S \in A} n \log\det(\pi)(\lambda_m \mathbf{I} + \tilde{\mathbf{H}}_S \tilde{\mathbf{H}}_S^* P) \\ &\leq \sum_{S \in A} n(\nu_S \log(P) + o(\log(P))), \end{aligned} \tag{D.1}$$

where the second inequality follows as the power constraint requires that the average codebook covariance

$$\mathbf{K} = \frac{1}{n} \sum_{t=1}^{n} \mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_t)$$

must have trace less than or equal to P and hence all its eigenvalues are bounded above by P. Thus, $\mathbf{K} \preceq P\mathbf{I}$ and the desired bound follows.

With this in mind, we may use the extremal inequalities (5.19) to write, with $(\uparrow \{1\}, \ldots, \uparrow \{K\}) = \mathcal{A}_{[1:K]}$,

$$\begin{aligned} &\sum_{i \in \mathcal{I}} \alpha_i H_M(\Phi_i(\mathcal{A}_{[1:K]})) \\ &\quad \leq \sum_{i \in \mathcal{I}} \alpha_i H_{M,\tilde{\mathbf{Y}}^n}(\Phi_i(\mathcal{A}_{[1:K]})) \\ &\quad \leq \sum_{j \in \mathcal{J}} \beta_j H_{M,\tilde{\mathbf{Y}}^n}(\Pi_i(\mathcal{A}_{[1:K]})) + \sum_{l \in \mathcal{L}} \gamma_l \big(H_{M,\tilde{\mathbf{Y}}^n}(\Gamma_i^+(\mathcal{A}_{[1:K]})) - H_{M,\tilde{\mathbf{Y}}^n}(\Gamma_i^-(\mathcal{A}_{[1:K]})) \big) \\ &\quad \overset{(i)}{\leq} \sum_{j \in \mathcal{J}} \beta_j H_{M,\tilde{\mathbf{Y}}^n}(\Pi_i(\mathcal{A}_{[1:K]})) + \sum_{l \in \mathcal{L}} \gamma_l \left(H_{M,\tilde{\mathbf{Y}}^n}(\Gamma_i^+(\mathcal{A}_{[1:K]}) \backslash \Gamma_i^-(\mathcal{A}_{[1:K]})) \right) \end{aligned}$$

$$\overset{(ii)}{\leq} \sum_{j\in\mathcal{J}} \beta_j n C(\Phi_i(\mathcal{A}_{[1:K]}))\log(P) + \sum_{l\in\mathcal{L}} n\gamma_l \big(C(\Gamma_i^+(\mathcal{A}_{[1:K]})) - C(\Gamma_i^-(\mathcal{A}_{[1:K]}))\big)\log(P)$$

$$+ no(\log(P)) + n\epsilon_n$$

$$\overset{(iii)}{\leq} n\sum_{i\in\mathcal{I}} \alpha_i (C(\Phi_i(\mathcal{A}_{[1:K]})) + o(\log(P)) + \epsilon_n),$$

where (i) follows as conditioning reduces entropy and (ii) follows by (D.1) and (iii) follows by the fact that equality holds in the extremal inequality (5.19) for modular functions. Then the desired result follows by taking $P, n \to \infty$ by noting that

$$n\sum_{i\in\mathcal{I}}\alpha_i \sum_{S\in\Phi_i(\mathcal{A}_{[1:K]})} R_S = \sum_{i\in\mathcal{I}} \alpha_i H_M(\Phi_i(\mathcal{A}_{[1:K]}))$$

and similarly

$$\sum_{i\in\mathcal{I}} \alpha_i C(\Phi_i(\mathcal{A}_{[1:K]}) = \sum_{i\in\mathcal{I}} \alpha_i \sum_{S\in\Phi_i(\mathcal{A}_{[1:K]})} \phi_S.$$

□

D.2 On Linear Subspaces

Proposition D.2.1. *For any two linear subspaces A, B of $\mathbb{C}^t$, $(A \cap B^\perp) \cap B = \{0\}$.*

Proof. If $x \in (A \cap B^\perp) \cap B$ then $x \in B^\perp \cap B$. Thus $\langle x, y\rangle = 0$ for any $y \in B$, including x itself. Thus $\|x\|^2 = 0$ and necessarily $x = 0$. □

Proposition D.2.2. *Suppose that A, B are linear subspaces of $\mathbb{C}^t$ and $B \subseteq A$. Then $\{(A\cap B^\perp), B\}$ is a direct sum decomposition of A.*

Proof. By Proposition D.2.1, $(A \cap B^\perp) \cap B = \{0\}$. By Gram-Schmidt, there exists a matrix $\mathbf{Q}_A = \begin{bmatrix}\mathbf{Q}_B & \mathbf{Q}_C\end{bmatrix}$ with orthogonal columns such that $\text{Range}(\mathbf{Q}_B) = B$ and $\text{Range}(\mathbf{Q}_A) = A$. Take any $x \in A \cap B^\perp$, which must have a representation $x = \mathbf{Q}_B y_B + \mathbf{Q}_C y_C$. As $x \in B^\perp$, $\mathbf{Q}_B^* x = 0$ and $x = \mathbf{Q}_c y_C$. Thus $\text{Range}(\mathbf{Q}_C) = (A \cap B^\perp)$, which implies that $A = (A \cap B^\perp) \oplus B$. □

Proposition D.2.3. *Let $A_1, \ldots, A_M$ be linear subspaces of $\mathcal{C}^t$. Then*

$$\left(\bigoplus_{i=1}^M A_i\right)^\perp = \bigcap_{i=1}^M A_i^\perp.$$

Proof. Let $\mathbf{A}_i$ be a matrix whose columns span A_i. As $\text{Null}(\mathbf{X}^*) = \text{Range}(\mathbf{X})^\perp$ for any matrix $\mathbf{X}$, when the superscript $*$ denotes the Hermitian transpose, we have

$$\left(\bigoplus_{i=1}^{M} A_i\right)^\perp = \text{Range}([\mathbf{A}_1, \ldots, \mathbf{A}_M])^\perp = \text{Null}\left(\begin{bmatrix} \mathbf{A}_1^* \\ \vdots \\ \mathbf{A}_M^* \end{bmatrix}\right) = \bigcap_{i=1}^{M} A_i^\perp.$$

□

D.3 Recursive Mutual Covering Lemma Proof

Let $\mathcal{T}_\epsilon^{(n)}(U_E)$ be the set of jointly ϵ-typical length-n sequences $(u_S^n : S \in E)$ with respect to the joint distribution of U_E. Let

$$\mathcal{A} = \{(m_S : S \in E) : (u_S^n(m_S) : S \in E) \in \mathcal{T}_\epsilon^{(n)}(U_E)\}$$

be the the set of all independently randomly generated vectors which appear as though they were jointly generated (by being jointly typical). As before, Chebyshev's inequality supplies $P(|\mathcal{A}| = 0) \leq \text{Var}(|\mathcal{A}|)/E(|\mathcal{A}|)^2$. The probability mass function governing the distribution of a codeword tuple $(u_S^n(m_s) : S \in E)$ is independent of the message tuple $m \equiv (m_S : S \in E)$. Thus each codeword tuple has the same probability of being jointly typical, which we define to be

$$P(U_E^n(m) \in \mathcal{T}_\epsilon(U_E)) = P(U_E^n(\mathbf{1}) \in \mathcal{T}_\epsilon(U_E)) \triangleq p.$$

By linearity of expectation, $E[|\mathcal{A}|] = 2^{nr(E)}p$. Introduce

$$B(m(E), m'(E)) = [\![U_E^n(m_E) \in \mathcal{T}_\epsilon(U_E)]\!][\![U_E^n(m'_E) \in \mathcal{T}_\epsilon(U_E)]\!] \ .$$

As before,

$$\text{Var}[|\mathcal{A}|] \leq \sum_{D \subset S} 2^{n(2r(D)+r(D^c))} p_D,$$

where $p_D = E[B(m(E), m'(E))]$ with $(m_S = 1 : S \in E)$ and $(m'_S = 1 : S \notin D)$ but $(m'_S = 2 : S \in D)$, more succinctly stated as $m = \mathbf{1}$ and $m' = \mathbf{1} + \mathbf{1}(D)$ where $\mathbf{1}(D) = ([\![S \in D]\!] : S \in E)$ is the indicator vector for a subset D of E. By the recursive generating procedure, if $m_S \neq m_{S'}$ for some

$S \in E$, then all the vectors corresponding to $R \in\downarrow S$ appear as though they were independently generated, even if the indices m_R and m'_R match. In other words,

$$p\Big(U^n(\mathbf{1}) = u^n, U^n(\mathbf{1}+\mathbf{1}(D)) = v^n\Big) = \prod_{S\in\downarrow D} p(u_S|u_{\uparrow' S})p(v_S|v_{\uparrow' S}) \times \prod_{S\in E\setminus\downarrow D} p(u_S|u_{\uparrow' S})$$

for all potential sequences u^n, v^n with $u_S = v_S$ when $S \notin\downarrow D$. If these potential sequences are in addition also jointly ϵ-typical (denote the set of such sequences with $\mathcal{T}(\epsilon, D)$), then

$$p\left(U^n(\mathbf{1}) = u^n, U^n(\mathbf{1}+\mathbf{1}(D)) = v^n\right) \leq 2^{-n(d_D - \delta_\epsilon/6)}$$

where

$$d_D = 2\sum_{S\in\downarrow D} H(U_S|U_{\uparrow' S}) + \sum_{S\in E\setminus\downarrow D} H(U_S|U_{\uparrow' S}).$$

Consider the set of sequences v^n that are jointly $\epsilon-$typical with respect to the **joint** distribution on U_E and have components in $E\setminus D$ fixed to $v_S = u_S$ for $S \in E\setminus D$, which we denote by $\mathcal{S}(\epsilon, u_{E\setminus D})$. Then by standard arguments, $\log|\mathcal{S}(\epsilon, u_{E\setminus D})| \leq nH(U_D|U_{E\setminus D})$. Combining the above observations yields a bound on the probability that the pair of sequences $U^n(\mathbf{1}) = u^n, U^n(\mathbf{1}+\mathbf{1}(D)) = v^n$ are jointly typical:

$$\begin{aligned}
p_D &= \sum_{u_D^n, v_D^n \in \mathcal{T}(\epsilon, D)} p\Big(U^n(\mathbf{1}) = u^n, U^n(\mathbf{1}+\mathbf{1}(D)) = v^n\Big) \\
&= \sum_{u_D^n, v_D^n \in \mathcal{T}(\epsilon, D)} 2^{-nd_D} \\
&= \sum_{u_D^n : u^n \ \epsilon-\text{typical}} \sum_{v^n \in \mathcal{S}(\epsilon, u_{E\setminus D})} 2^{-nd_D} \\
&\leq 2^{nH(U_E) + nH(U_D|U_{E\setminus D}) - nd_D + \delta_\epsilon/2},
\end{aligned}$$

This bound, combined with the bound

$$\frac{1}{n}\log p \geq H(U_E) - \sum_{S\in E} H(U_S|U_{\uparrow' S}) + \delta_\epsilon/4$$

yields

$$\frac{\mathrm{Var}(|\mathcal{A}|)}{E[|\mathcal{A}|]^2} \leq \sum_{D\subset E} 2^{-n\triangle(D)}$$

where

$$\triangle(D) = r(E \backslash D) - \left(\sum_{S \in E \backslash \downarrow D} H(U_S | U_{\uparrow' S}) - H(U_{E \backslash D}) \right) - \delta_\epsilon.$$

Thus the error tends to zero if the conditions

$$r(G) \geq \sum_{S \in \mathcal{X}_\uparrow(G)} H(U_S | U_{\uparrow' S}) - H(U_G),$$

where $\mathcal{X}_\uparrow(G)$ is the largest up-set contained within G, hold for all subsets $G \subseteq E$. But not all of these inequalities are necessary. Observe that as the rates are non-negative, and $\mathcal{X}_\uparrow(G)$ is a subset of G,

$$r(G) \geq r(\mathcal{X}_\uparrow(G)) \geq \sum_{S \in \mathcal{X}_\uparrow(G)} H(U_S | U_{\uparrow' S}) - H(U_{\mathcal{X}_\uparrow(G)}) \geq \sum_{S \in \mathcal{X}_\uparrow(G)} H(U_S | U_{\uparrow' S}) - H(U_G).$$

Thus, the inequality corresponding to $\mathcal{X}_\uparrow(G)$ implies the inequality corresponding to G. As such, it suffices to only enforce those inequalities corresponding to the up-sets of E.