Title

Capacity Bounds and Coding Schemes for Cloud Radio Access Networks

Author
Ganguly, Shouvik

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Capacity Bounds and Coding Schemes for Cloud Radio Access Networks

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

in

Electrical Engineering
(Communication Theory and Systems)

by

Shouvik Ganguly

Committee in charge:

Professor Young-Han Kim, Chair
Professor Bhaskar D. Rao
Professor Dinesh Bharadia
Professor Jason Ross Schweinsberg
Professor Paul H. Siegel

2020

The Dissertation of Shouvik Ganguly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

To my parents and my sister

without whom I would not be who I am

EPIGRAPH

Every difficulty slurred over will be a ghost

to disturb your repose later on.

—Frederic Chopin

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The PhD journey, while it can be called long and arduous, has been nonetheless extremely rewarding and fulfilling for me, changing me even as a person in many ways. The credit should go, first and foremost, to my advisor, Prof. Young-Han Kim. When I first contacted him all those years back, I did not know how working with him as an advisor would be; it was, after all, his interests and his profile which drew me to strive for an opportunity to be his student. It worked out better than even the most optimistic me could have hoped for; not only did his working style and his approach to research teach me a lot, but also, and more importantly, his compassion, his understanding of the mindset of a student, his life advice, and overall, his paternalistic attitude, helped me navigate these years and, I hope, made me a better researcher. It would take far too many pages to adequately convey all the ways in which he taught and inspired me; therefore, I would like to mention a few key aspects. First and foremost, his passion and thirst for learning and research can hardly fail to inspire anyone. The weekly group meetings, where participants are free to talk about any "interesting" topic, including brain-twisting math puzzles and sometimes obscure but elegant results, exemplify his philosophy, namely, that research should be fun above everything else. He motivated me on many occasions to change my thinking style and taught me to be more rigorous in my approach to problems. His style of attacking problems by breaking them down into simpler cases and his guiding principle that the best metric of time well-spent on research is my own perception of how valuable the contribution is and how much fulfillment it brought me, gave me a holistic view of the journey and broadened my horizon indescribably. Finally, his attention to detail and his incessant criticism of my writing has improved my technical communication skills beyond measure; this is another tangible gift I take with me in my future journey.

I would like to thank my doctoral committee members for graciously agreeing to serve on the committee, and for their valuable encouragement and suggestions along the way. I would like to thank Prof. Jason Schweinsberg for being a meticulous teacher; I owe most of my knowledge in probability theory to his extremely well-planned lectures over the course of three quarters.

His teaching style has forever spoiled me and redefined my expectations as to how a math class should be. I thank Prof. Bhaskar Rao for being an exemplary teacher as well, always being able to convey his enjoyment and his appreciation of the beauty of the subject matter to the audience; even without a few of the recommended pre-requisites under my belt, his teaching style made it more than easy to internalize the material and feel excited about the applications. I thank Prof. Paul Siegel for his brilliant treatment of probabilistic codes in his class and for his pace so as to be able to present both state-of-the-art and classical results in an easily accessible fashion. I thank Prof. Dinesh Bharadia for challenging me to think critically about the engineering questions that may arise from my research, and for making me ponder over how to sell my theoretical results to real-world practitioners.

I wish to thank my fellow students in the group for being exceptional peers in this shared journey. I thank Pinar for her friendship as well as constant encouragement and help in navigating the PhD milestones, many of which we crossed at the same time. Her willingness to help and limitless patience for my series of questions made the logistic hurdles much easier to clear. I thank Jongha for being an amazing friend and collaborator; the long discussions, technical and otherwise, enriched me and opened my eyes. I will cherish the "intellectual resonance" that we shared, and I am especially grateful to him for pushing me to give my one hundred percent at every step. I thank Nadim, Alankrita, and Jiun-Ting for their friendships and for fostering a group dynamic that turned the workplace into a comfortable, nurturing environment. In particular, I also thank Nadim for our endless foodie adventures and for helping keep the mood light even at tough times through friendly banter. I thank Lele for being a perfect role model as a group senior and for consistently setting an example as a devoted and accomplished researcher. I have benefited immensely from the opportunity to collaborate with her, through her constant encouragement and her philosophy of never giving up until the summit is conquered. I thank the other seniors in the group – Yu Xiang, Chiao-Yi, and Fatemeh – for making me feel welcome in the group from day one. I thank Dheeraj, Moeen, Chaitanya, Vaishakh, Hanwen, and Tugcan for light-hearted conversation and good company on the fourth floor.

I thank all the graduate division and ECE advisors – in particular, Teresa Ta – for their constant help and support with logistics. I thank all companies and organizations whose funds made this journey possible; in particular, I thank Samsung Research America, Dallas, and ETRI, Korea. I thank UCSD for the opportunity of being a student here and for providing me with a valuable nurturing ground.

My journey of these six years would have been incomplete without my friends, who became family and turned San Diego into a home away from home. Sridip, Ahanjit, and Brato have been with me every step of the way and shared the good times and the bad. They have provided emotional support, been on shared journeys of exploration – be it trying out a new restaurant or joining me on my first road-trip after getting a driver's license, been critical when I needed someone to spell out the harsh truth rather than be nice, encouraged me to follow my heart without fear, in work as in life, and occasionally, unknowingly provided much-needed peer pressure to get me to work harder and get stuff done. Sumit, Siddhartha Da, Madhukar, Rashmi, Sasank, Varun, Athira, Bodhi, Diptarka Da, and Uri blessed my life and made the moments not spent working that much more special. I thank Nasir for being a permanent host in LA and for being an always-ready companion for crazy road-trips and other experiences. I thank Netai Mama and Anindita Mami for being de facto "guardians" in a new city and for making me feel that I never left my culture behind while moving here. Among my IIT friends, I thank Manish and Soumyajit in particular, for the more-than-a-decade-old friendships that carried over to the other side of the world. I thank Pratik, Anindya, and Simanti for making my two summers in Dallas enjoyable and helping make the experience a tiny bit closer to San Diego. I thank Abhijeet and Akshit, who welcomed me to Ames, IA with open arms and became the first friends I made in the US. I thank my cousin Avi for being my "friend, philosopher, and misguide" throughout my life and especially during my time in the US. He taught me many aspects of life here and thanks to him, I got to meet my amazing sister-in-law Melissa, as well as Pratik, Soham, Royal, and Shannon, all of whom I'm indebted to for making the journey worthwhile.

I wish to separately thank La Jolla (San Diego), which has been my home for the past

six years. As far as life decisions go, moving here was one of the best I ever made. It forever changed my personality and my outlook on life, making me more open-minded, outgoing, and, I hope, less socially awkward than the nerdy young adult who first stepped on these shores six years back. It has been a place of many firsts for me, and the town and its people will always hold a special place in my heart.

I thank my early educators for fostering my lifelong love for math. This section would be incomplete without the mention of Udaykaka, my constant friend and mentor for the first eighteen years of my life and an inexhaustible source of inspiration even now. I owe a lot of my knack for logical and abstract thinking, as well as my grasp of written English, to his guidance during my formative years. I thank my undergraduate advisor, Prof. R. K. Bansal of IIT Kanpur, for introducing me to the wonderful subject of information theory and for motivating me to pursue higher education in this field.

Finally, my family – immediate as well as extended – has been with me through thick and thin every step of the way. None of what I did would be possible without my parents or my sister. They have always cared for me and been worried about me, sometimes too much, but nonetheless, encouraged me to always aim high and never compromise on following my passion, no matter how rough the road got. I love and cherish them, and consider myself fortunate to have them. This dissertation is dedicated to my family.

Chapter 2 is, in part, a reprint of the material in the paper: Shouvik Ganguly and Young-Han Kim, “On the Capacity of Cloud Radio Access Networks”, *Proceedings of the IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017, as well as parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, “On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul”, *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*. The dissertation author was the primary researcher and author of this material.

Chapter 3 is, in part, a reprint of the material in the paper: Shouvik Ganguly and Young-Han Kim, “Capacity Scaling for Cloud Radio Access Networks with Limited Orthogonal Fronthaul”, *Proceedings of the IEEE International Symposium on Information Theory*, Paris, France, July 2019, as well as parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, “On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul”, *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*. The dissertation author was the primary researcher and author of this material.

Chapter 4 is, in part, a reprint of parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, “On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul”, *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*. The dissertation author was the primary researcher and author of this material.

Chapter 5 is, in part, a reprint of the material in the paper: Shouvik Ganguly, Lele Wang, and Young-Han Kim, “A Functional Construction of Codes for Multiple Access and Broadcast Channels”, accepted for publication at the *IEEE International Symposium on Information Theory*, Los Angeles, USA, June 2020. The dissertation author was the primary researcher and author of this material.

VITA

2013 B.Tech. in Electrical Engineering, Indian Institute of Technology, Kanpur.

2020 Ph.D. in Electrical Engineering (Communication Theory and Systems), University of California San Diego.

PUBLICATIONS

List of publications

Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, "On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*.

Shouvik Ganguly, Lele Wang, and Young-Han Kim, "A Functional Construction of Codes for Multiple Access and Broadcast Channels", accepted for publication at the *IEEE International Symposium on Information Theory*, Los Angeles, USA, June 2020.

Shouvik Ganguly and Young-Han Kim, "Capacity Scaling for Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *Proceedings of the IEEE International Symposium on Information Theory*, Paris, France, July 2019.

Shouvik Ganguly and Young-Han Kim, "On the Capacity of Cloud Radio Access Networks", *Proceedings of the IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017.

J. Jon Ryu, Shouvik Ganguly, Young-Han Kim, Yung-Kyun Noh, and Daniel D. Lee, "Nearest Neighbor Density Functional Estimation based on Inverse Laplace Transform", *arXiv:1805.08342 [math.ST], May 2018*, submitted to *Annals of Statistics*.

Shouvik Ganguly, K. Sahasranand, and Vinod Sharma, "A New Algorithm for Distributed Nonparametric Sequential Detection", *Proceedings of the IEEE International Conference on Communications*, Sydney, Australia, June 2014.

Shouvik Ganguly, K. Sahasranand, and Vinod Sharma, "A New Algorithm for Nonparametric Sequential Detection", *Proceedings of the National Conference on Communications*, Kanpur, India, February 2014.

Shouvik Ganguly and Sridip Pal, "Bounds on the Density of States and the Spectral Gap in CFT-2", *Phys. Rev. D*, vol. 101, no. 10, pp. (106022) 1–14, May 2020.

ABSTRACT OF THE DISSERTATION

Capacity Bounds and Coding Schemes for Cloud Radio Access Networks

by

Shouvik Ganguly

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California San Diego, 2020

Professor Young-Han Kim, Chair

Cloud radio access networks (C-RANs) have been touted as a viable alternative to current communication network architectures for handling larger data volumes and higher throughput requirements, in order to serve a growing number of data-hungry devices. In this dissertation, we study information theoretic models of uplink and downlink C-RANs and explore questions on optimal data throughputs and large-network size asymptotics. In addition, as a first step on the path to achieving these tradeoffs in practice, we demonstrate how one can build codes for multiuser networks and provide finite-blocklength performance guarantees by starting from single-user codes.

Chapter 1

Introduction

With ever-increasing demands for higher data rates, better coverage, and more reliable connectivity for a large number of devices, new network architectures and protocols are expected to play an important role in future communication systems. The cloud radio access network (C-RAN) architecture [42, 43] is one of the candidates, being implemented extensive in the upcoming 5G systems. In this architecture, communication over a group of cells is coordinated by a cloud-based central processor. Fig. 1.1 depicts uplink and downlink C-RAN systems schematically.

Base stations in a C-RAN, unlike in conventional cellular networks, do not perform all network functionalities locally, but instead delegate most of them to a central processor by communicating with it over wired or wireless fronthaul links. If these links have unbounded capacities, the base stations act as spatially distributed antennas of a conventional multiple-input multiple-output (MIMO) system, that use beamforming to coordinate transmission and mitigate interference among multiple cells. For the more realistic situation of limited fronthaul link capacities, beamforming in a downlink C-RAN is typically performed at the central processor assuming no capacity constraints, and the corresponding baseband signals are digitized individually for each base-station and transmitted through the fronthaul links. For uplink, the received signal at each base-station is similarly digitized individually according to the corresponding link capacity, and transmitted to the base-station. These approaches often lead to high fronthaul capacity requirements.

As an alternative to this greedy beamforming–digitization approach, we investigate near-optimal coding schemes for the C-RAN architecture and their achievable throughput tradeoffs by modeling the entire system as a two-hop relay network. In this model, which was studied, for example, in [43, 68, 69], the base stations act as relays that summarize the received signals from user devices to the central processor (uplink) and transmit the prescribed signals from the central processor to user devices (downlink). Communication-theoretic results on this model were presented in a recent volume edited by Quek, Peng, Simeone, and Yu [43].

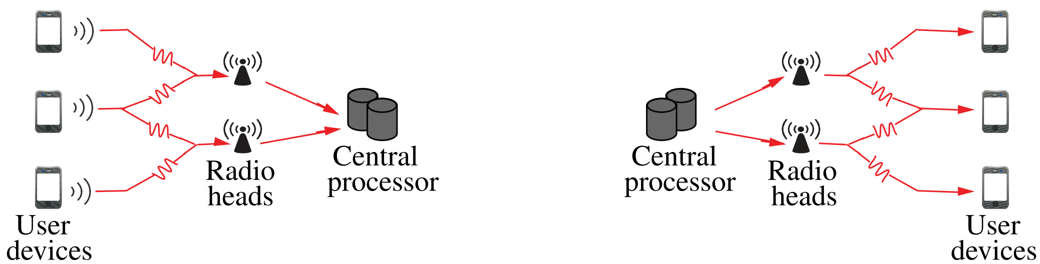


Figure 1.1. (a) Uplink and (b) downlink cloud radio access networks.

Several coding schemes have been proposed in the literature for the uplink C-RAN with K users (senders) and L relays. Zhou and Yu [69] applied the network compress–forward relaying scheme [26] to this model and showed, by optimizing over quantizers, that under some symmetry assumptions, this scheme achieves the optimal sum-rate within $L/2$ bits per real dimension uniformly over all K and all channel parameters. Sanderovich, Someskh, Poor, and Shamai [47] used the same scheme and analyzed the large-user asymptotics (i.e., the scaling law) of symmetric achievable rates when all fronthaul links have equal capacities. Zhou, Xu, Yu, and Chen [68] subsequently showed that under a sum-capacity constraint on the fronthaul links, the coding scheme in [69] and [47] can be simplified through successive cancellation decoding, generalizing an earlier result for the single-sender multiple relay network [46]. Aguerri and Zaidi [2] proposed a hybrid coding scheme of network compress–forward and compute–forward [35], and demonstrated that it outperforms the better of the two in general. Aguerri, Zaidi, Caire, and Shamai [3] specialized the noisy network coding scheme [29] to the uplink C-RAN, the achievable rate region of which coincides with that of network compress–forward [47, 69]. Park et al. [38] studied joint decompression and decoding for the uplink.

The most general outer bound on the capacity region of the uplink C-RAN can be obtained by specializing the cutset bound [18]; see, for example, [69] and the references therein, as well as Proposition 2.2.2 in this thesis. The cutset bound has been further tightened under additional assumptions. Aguerri, Zaidi, Caire, and Shamai [3] studied the uplink C-RAN in which the relays are oblivious of the codebooks of the senders, and demonstrated that network compress–forward (or noisy network coding) achieves the capacity region. Simeone, Somekh, Erkip, Poor, and Shamai [52] studied the uplink C-RAN with one sender, L oblivious relays, and unreliable fronthaul links, and derived an upper bound on the capacity, which was numerically shown to be close to the network compress–forward lower bound under certain network parameters.

For the downlink C-RAN with L relays and K receivers, a variety of coding schemes have been proposed. Hong and Caire [23] studied a low-complexity reverse compute–forward scheme for symmetric rates. Liu and Yu [31] applied network coding and beamforming to the downlink model with a noiseless multi-hop fronthaul. Motivated by the MAC–BC duality, Liu, Patil, and Yu [30] proposed compression-based schemes and established a duality between achievable rate regions for the uplink and downlink C-RANs. El Bakouri and Nazer [14, 15] applied integer-forcing based joint beamforming and compression strategies and demonstrated a duality between uplink and downlink C-RANs under this framework. Bidokhti and Kramer [7] studied the 2-relay, single-user downlink C-RAN and used rate-splitting across relays and Marton coding with common message to derive capacity lower bounds. Bidokhti, Kramer, and Shamai [8] studied the L-relay, single-user downlink C-RAN and used Marton coding and rate-splitting across relays, but this time with no common message (due to the complexity for $L > 2$). Wang, Wigger, and Zaidi [59] studied the three-hop, 2-relay, 2-user downlink C-RAN with relay cooperation, where the relays communicate with each other once, simultaneously, per network use. They applied the *generalized data-sharing (G-DS)* and distributed decode–forward (DDF) [28] coding schemes to this network, numerically showing that G-DS outperforms DDF in the low-power regime with a Gaussian second hop. More recently, Patil and Yu [41] have shown that under fronthaul sum-capacity constraints, a successive encoding scheme achieves the same rate region as DDF. Shamai and Zaidel [48] introduced a combined linear pre-processing and encoding scheme for the Gaussian downlink C-RAN based on factorizing the channel gain matrix and using writing on dirty paper [10, 11], which was shown to enhance the performance. Jing et al. [24] studied base station cooperation for downlink transmission in a multicell model for the soft handoff scenario [65] and analyzed the performance of linear precoding schemes under this framework. Simeone et al. [53] studied multicell processing for the downlink and investigated transmission schemes requiring partial or complete channel knowledge at the base stations. Liu and Kang [32] developed an achievability scheme for the $L = 2$ relays and $K = 2$ users case by combining Marton coding [17, Section 8.3] with using correlated codewords for the multiple access channel. In a dual approach to [38], Park et al. [39] studied joint precoding and compression for the downlink.

The most general outer bound on the capacity region of the downlink C-RAN can be obtained by specializing the cutset bound [18]; see, for example, [8] and the references therein, as well as Proposition 2.3.2 in this thesis. The cutset bound has been further tightened for specific network configurations. Bidokhti and Kramer [7] derived capacity upper bounds for the 2-relay, single-user downlink C-RAN by tightening the cutset bound through channel enhancement techniques [25, 36]. These bounds are tight for the single-user symmetric Gaussian C-RAN under certain parameters, establishing the capacity for those cases. These upper bounds were further generalized to $L \geq 3$ relays and a single user by Bidokhti, Kramer, and Shamai [8]. In related work, Yang et al. [66] developed outer bounds for the downlink multicell processing model with $L = 2$ relays and $K = 2$ users.

For a more comprehensive review of both uplink and downlink C-RANs, the reader is referred to [40, 42, 51, 54] and the references therein.

In Chapter 2, we specialize network compress–forward (or equivalently, noisy network coding) to the uplink C-RAN and approximate the capacity region to within a constant gap per user, independent of the channel gain matrix and the power constraint. A similar approximation is achieved for the downlink C-RAN by simplifying the distributed decode–forward coding scheme [28].

In Chapter 3, we characterize the scaling behavior of the C-RAN sum-capacity for large network size under various channel models and compare it to the capacities of currently-used network architectures.

Chapter 4 extends the results of Chapters 2 and 3 to MIMO C-RANs, in which users and relays have multiple local antennas.

While it is useful to theoretically quantify the best possible throughput tradeoffs across multiple users in a communication network, the aforementioned coding schemes are often used more as proof techniques than as a prescription on how to efficiently transmit information over real networks in practice. The practical problem of coding for point-to-point channels has seen enormous advances in recent years, with the advent and extensive studies of several low-complexity coding schemes that approach or achieve the Shannon capacity. Of particular note among these are the turbo codes [6], low-density parity-check (LDPC) codes [19, 27, 45], and polar codes [4]. While a multitude of information-theoretic results on achievable rate tradeoffs for multi-user channels exist in the literature (see, for example, [17, Chapter 1] and [16] for comprehensive reviews), we are far from achieving known inner bounds with low complexity coding schemes due to the high computational complexity in implementing some sophisticated multi-user coding schemes, such as Marton coding [17, Theorem 8.3 and Proposition 8.1] for broadcast channels and simultaneous decoding [17, Chapter 4.5.1] for multiple access channels.

At such a juncture, Chapter 5 starts out with the long term goal of turning theoretical coding schemes into efficient approaches in practice. In this chapter, we start out with Gelfand–Pinsker (GP) codes [21] for binary-input, binary-state channels and construct codes for binary-input multiple access channels (MAC) and finite-alphabet broadcast channels (BC). Coding for these single-hop, multiuser networks are an important first step towards building practical codes for C-RAN and even more general multihop networks.

Chapter 6 concludes the thesis, along with comments on possible future directions of investigation.

Throughout the dissertation, we follow the notation in [17]. In addition, we use the following. In Chapters 2 and 4, we use $||A||_F := \sqrt{\text{tr}(AA^T)} = \sqrt{\text{tr}(A^TA)}$ to denote the Frobenius norm of a matrix A. For a natural number n, we denote by $[n]$ the set $\{1,\ldots,n\}$. We denote a finite tuple of objects $(x_l, l \in \mathcal{S})$ by the shorthand notation $x(\mathcal{S})$, for $\mathcal{S} \subseteq \mathbb{N}$. For example, $x([n]) = x^n = (x_1,\ldots,x_n)$. For a tuple of random variables $X(\mathcal{S}) := (X_l, l \in \mathcal{S})$ and a random variable Y, we define the *total correlation*

$$I^*(X(\mathcal{S})\,|\,Y) := \sum_{x(\mathcal{S}),y} p(x(\mathcal{S}),y) \log \frac{p(x(\mathcal{S})\,|\,y)}{\prod_{l\in\mathcal{S}} p(x_l\,|\,y)} = \sum_{l\in\mathcal{S}} I(X_l; X([l-1]\cap\mathcal{S})\,|\,Y)$$

as a multivariate generalization of conditional mutual information [62]. For functions f and g from $\mathbb{N}$ to $\mathbb{R}$, we say $f \sim g$ if $f(n)/g(n) \to 1$ as $n \to \infty$. Further, $\log(\cdot)$ and $\ln(\cdot)$ denote logarithms to base 2 and base e, respectively. All information measures are in bits.

Chapter 2

Approximate Capacities of C-RANs

Uplink and downlink cloud radio access networks are modeled as two-hop K-user L-relay networks, whereby small base-stations act as relays for end-to-end communications and are connected to a central processor via orthogonal fronthaul links of finite capacities. Simplified versions of network compress–forward (or noisy network coding) and distributed decode–forward are presented to establish inner bounds on the capacity region for uplink and downlink communications, that match the respective cutset bounds to within a finite gap independent of the channel gains and signal to noise ratios.

2.1 Introduction

In this chapter, we apply network compress–forward (or equivalently, noisy network coding) to the uplink C-RAN and show that the scheme achieves the capacity region approximately within $(1/2)\log(eL)$ bits per user per real dimension, regardless of the channel gain matrix, power constraint, and the number of users K. When the fronthaul link capacities are unbounded, the approximation is precise and the network compress–forward inner bound (as well as the cutset outer bound) coincides with the fronthaul-unlimited uplink capacity region.

Likewise, we specialize and simplify the distributed decode–forward coding scheme [28] to the downlink C-RAN with capacity-limited single-hop fronthaul. In this scheme, multicoding at the encoder (as in Marton coding for broadcast channels [17]) is coupled with coding for fronthaul links, which allows more efficient coordination among the transmitted signals at the base-stations. We show that our rate region achieves a per-user gap of $(1/2)\log(eKL)$ bits per real dimension from the cutset bound. This result refines the best-known linear gap from capacity for this model (implicit in [30]).

2.2 Uplink C-RANs

2.2.1 General Model

We model the uplink C-RAN as a two-hop relay network in Fig. 2.1, where the first hop, namely, the (wireless) channel from the user devices (senders) to the radio heads (relays), is modeled as a discrete memoryless network $p(y^L|x^K)$, and the second hop, namely, the channel from the radio heads to the central processor, consists of orthogonal links of capacities $C_1,\ldots,C_L$ bits per real dimension, decoupled from the first hop. To be more precise, the channel output at the central processor (receiver) is $(W_1,\ldots,W_L)$, where $W_l \in [2^{nC_l}]$ is a reliable estimate of what relay l communicates to the receiver over n transmissions. We assume without loss of generality that these communication links are noiseless.

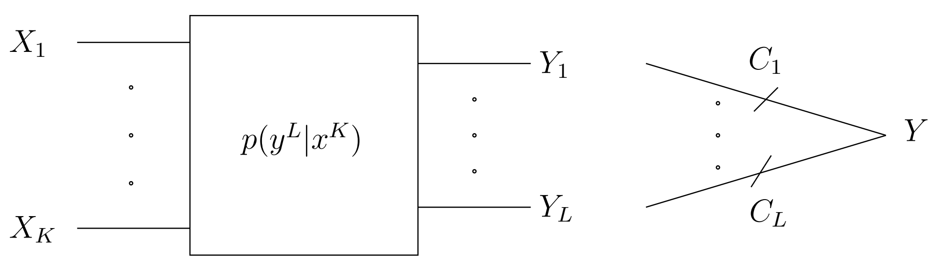


Figure 2.1. Uplink network model.

A $(2^{nR_1},\ldots,2^{nR_K},n)$ code for this network consists of K message sets $[2^{nR_1}],\ldots,[2^{nR_K}]$; K encoders, where encoder $k \in [K]$ assigns a codeword x_k^n to each $m_k \in [2^{nR_k}]$; L relay encoders, where relay encoder $l \in [L]$ assigns an index $w_l \in [2^{nC_l}]$ to each received sequence y_l^n; and a decoder that assigns message estimates $(\hat{m}_1,\ldots,\hat{m}_K)$ to each index tuple $w^L := (w_1,\ldots,w_L)$. We assume that the messages $M_1,\ldots,M_K$ are uniformly distributed and independent of each other. The average probability of error is defined as $P_e^{(n)} = \mathsf{P}(\cup_{k=1}^K \{\hat{M}_k \neq M_k\})$. A rate tuple $(R_1, \ldots, R_K)$ is achievable if there is a sequence of $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity region is defined as the closure of the set of all achievable rate tuples.

The noisy network coding scheme [29] can be specialized [3] to the uplink C-RAN model as follows.

Codebook generation. Fix a pmf $p(q)\prod_{k=1}^K p(x_k \,|\, q)\prod_{l=1}^L p(\hat{y}_l \,|\, y_l, q)$. Randomly generate a time-sharing sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$. For each message $m_k \in [2^{nR_k}]$, generate $x_k^n(m_k) \sim \prod_{i=1}^n p_{X_k|Q}(x_{ki|q_i})$ conditionally independently. Define auxiliary indices $t_l \in [2^{n\hat{R}_l}]$, $l \in [L]$, for some auxiliary rates $\{\hat{R}_l, l \in [L]\}$. For each
$(w_l, t_l) \in [2^{nC_l}] \times [2^{n\hat{R}_l}]$ and $l \in [L]$, generate $\hat{y}_l^n(w_l, t_l) \sim \prod_{i=1}^n p_{\hat{Y}_l|Q}(\hat{y}_{li} \,|\, q_i)$.

Encoding. For $k \in [K]$, to send message m_k, encoder k transmits $x_k^n(m_k)$.

Relaying. On receiving y_l^n, relay l finds (w_l, t_l) such that $(q^n, y_l^n, \hat{y}_l^n(w_l, t_l)) \in \mathcal{T}_{\varepsilon'}^{(n)}$ and transmits w_l to the central processor via the noiseless fronthaul. The compression at relay l succeeds w.h.p. if

$$C_l + \hat{R}_l > I(Y_l; \hat{Y}_l). \tag{2.1}$$

Decoding. Let $\varepsilon > \varepsilon'$. Upon receiving the index tuple w^L, the central processor finds message estimates $\hat{m}_1, \ldots, \hat{m}_K$ such that

$$(q^n, x_1^n(\hat{m}_1), \ldots, x_K^n(\hat{m}_K), \hat{y}_1^n(w_1, t_1), \ldots, \hat{y}_L^n(w_L, t_L)) \in \mathcal{T}_\varepsilon^{(n)}$$

for some $t_1, \ldots, t_L$. The decoding succeeds w.h.p. if

$$\begin{aligned}\sum_{k\in\mathcal{S}_1} R_k + \sum_{l\in\mathcal{S}_2} \hat{R}_l < I(X(\mathcal{S}_1); \hat{Y}(\mathcal{S}_2^c) \,|\, X(\mathcal{S}_1^c)) + \sum_{l\in\mathcal{S}_2} I(Y_l; \hat{Y}_l) \\ - \sum_{l\in\mathcal{S}_2} I\big(Y_l; \hat{Y}_l \,|\, \hat{Y}([l-1]\cap\mathcal{S}_2), \hat{Y}(\mathcal{S}_2^c), X^K\big)\end{aligned} \tag{2.2}$$

for every $\mathcal{S}_1 \subseteq [K], \mathcal{S}_2 \subseteq [L]$ such that $\mathcal{S}_1 \neq \emptyset$. Combining (2.1) and (2.2) to eliminate the auxiliary rates $\hat{R}_l$ and introducing a time-sharing random variable Q leads to the following inner bound on the capacity region of this network. (See Section 2.2.3 for a complete proof.)

Proposition 2.2.1 (Network compress–forward inner bound for the uplink C-RAN)**.** *A rate tuple $(R_1, \ldots, R_K)$ is achievable if*

$$\sum_{k \in \mathcal{S}_1} R_k \leq I(X(\mathcal{S}_1); \hat{Y}(\mathcal{S}_2^c) \,|\, X(\mathcal{S}_1^c), Q) + \sum_{l \in \mathcal{S}_2} C_l - \sum_{l \in \mathcal{S}_2} I\left(Y_l; \hat{Y}_l \,|\, \hat{Y}\left([l-1] \cap \mathcal{S}_2\right), \hat{Y}(\mathcal{S}_2^c), X^K, Q\right) \tag{2.3}$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ for some pmf $p(q)\prod_{k=1}^{K} p(x_k \,|\, q) \prod_{l=1}^{L} p(\hat{y}_l \,|\, y_l, q)$.

Specializing the cutset bound [18] to the uplink C-RAN model leads to the following.

Proposition 2.2.2 (Cutset outer bound for the uplink C-RAN)**.** *If a rate tuple $(R_1, \ldots, R_K)$ is achievable for the uplink C-RAN, then*

$$\sum_{k \in \mathcal{S}_1} R_k \leq I(X(\mathcal{S}_1); Y(\mathcal{S}_2^c) \,|\, X(\mathcal{S}_1^c), Q) + \sum_{l \in \mathcal{S}_2} C_l \tag{2.4}$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ for some pmf $p(q)\prod_{k=1}^{K} p(x_k \,|\, q)$.

For completeness, we provide a proof of Proposition 2.2.2 in Section 2.2.3.

Remark 2.2.1. As the fronthaul capacities $C_1, \ldots, C_L$ tend to infinity, this uplink C-RAN channel model becomes identical to the "fronthaul-unlimited" uplink channel from K senders to a single receiver with L receive antennas, i.e., the multiple access channel $p(y^L | x^K)$ with K senders $X_1, \ldots, X_K$ and one receiver Y^L. In this regime, both the inner and outer bounds can be shown to converge to the capacity region of the fronthaul-unlimited uplink channel, characterized by rate tuples $(R_1,\dots,R_K)$ satisfying

$$\sum_{k\in\mathcal{S}_1} R_k \le I(X(\mathcal{S}_1);Y^L\,|\,X(\mathcal{S}_1^c),Q)$$

for every $\mathcal{S}_1 \subseteq [K]$ for some pmf $p(q)\prod_{k\in[K]} p(x_k\,|\,q)$. In contrast, for *finite* fronthaul link capacities $C_1,\dots,C_L$, no matter how large, we can always find networks for which the capacity region of the uplink C-RAN is strictly smaller than the fronthaul-unlimited uplink capacity region, as demonstrated in Chapter 3.3.1.

2.2.2 Gaussian Model

We now assume that the first hop of the network is Gaussian, i.e.,

$$Y^L = GX^K + Z^L,$$

where $G \in \mathbb{R}^{L\times K}$ is a (deterministic) channel gain matrix and Z^L is a vector of independent $\mathrm{N}(0,1)$ noise components. We also assume the average power constraint P on each sender, i.e.,

$$\sum_{i=1}^{n} x_{ki}^2(m_k) \le nP, \quad m_k \in [2^{nR_k}], \quad k \in [K].$$

The network compress–forward inner bound in Proposition 2.2.1 can be specialized to this Gaussian network model to show the achievability of all rate tuples $(R_1,\dots,R_K)$ such that

$$\sum_{k\in\mathcal{S}_1} R_k \le \frac{1}{2}\log\left|\frac{P}{\sigma^2+1} G_{\mathcal{S}_2^c,\mathcal{S}_1} G_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_2} C_l - \frac{|\mathcal{S}_2|}{2}\log\left(1+\frac{1}{\sigma^2}\right) =: f_{\mathrm{in}}(\mathcal{S}_1,\mathcal{S}_2) \quad (2.5)$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ for some $\sigma^2 > 0$. Here, $G_{\mathcal{S}_2^c,\mathcal{S}_1}$ is the submatrix of G formed by the rows with indices in $\mathcal{S}_2^c$ and the columns with indices in $\mathcal{S}_1$. This follows by considering X^K to be a vector of i.i.d. $\mathrm{N}(0,P)$ random variables, and setting $\hat{Y}_l = Y_l + \hat{Z}_l$, $l \in [L]$, where $\hat{Z}_l \sim \mathrm{N}(0,\sigma^2)$. For convenience, for every $\sigma^2 > 0$, we denote the set of tuples $(R_1,\dots,R_K)$ satisfying (2.5) and hence, achievable by network compress–forward (NCF), by $\mathscr{R}_{\text{up}}^{\text{NCF}}(\sigma^2)$. We also denote the achievable sum-rate for each $\sigma^2 > 0$ by

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) := \sup_{(R_1,\ldots,R_K)} \{R_1 + \cdots + R_K : (R_1,\ldots,R_K) \in \mathscr{R}_{\text{up}}^{\text{NCF}}(\sigma^2)\} \tag{2.6}$$

$$= \min_{\mathcal{S}_2 \subseteq [L]} \Big(\frac{1}{2}\log\Big|\frac{P}{\sigma^2+1} G_{\mathcal{S}_2^c,[K]} G_{\mathcal{S}_2^c,[K]}^T + I\Big| + \sum_{l\in\mathcal{S}_2} C_l - \frac{|\mathcal{S}_2|}{2}\log\big(1+\frac{1}{\sigma^2}\big)\Big). \tag{2.7}$$

We establish the following useful property of the inner bound (2.5), which will be useful in developing some insight into the nature of the achievable region, as well as in approximating the capacity region.

Lemma 2.2.1. *For any* $\mathcal{S}_2 \subseteq [L]$ *and* $\mathcal{S}_1' \subseteq \mathcal{S}_1 \subseteq [K]$,

$$f_{\text{in}}(\mathcal{S}_1', \mathcal{S}_2) \le f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2).$$

Proof. Letting $G_{\mathcal{S}_2^c,k}$ denote the column vector consisting of the elements of G with row index in $\mathcal{S}_2^c$ and column index k, we have

$$G_{\mathcal{S}_2^c,\mathcal{S}_1} G_{\mathcal{S}_2^c,\mathcal{S}_1}^T = \sum_{k\in\mathcal{S}_1} G_{\mathcal{S}_2^c,k} G_{\mathcal{S}_2^c,k}^T \succeq \sum_{k'\in\mathcal{S}_1'} G_{\mathcal{S}_2^c,k'} G_{\mathcal{S}_2^c,k'}^T = G_{\mathcal{S}_2^c,\mathcal{S}_1'} G_{\mathcal{S}_2^c,\mathcal{S}_1'}^T,$$

which implies that

$$\frac{1}{2}\log\Big|\frac{P}{\sigma^2+1} G_{\mathcal{S}_2^c,\mathcal{S}_1} G_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\Big| \ge \frac{1}{2}\log\Big|\frac{P}{\sigma^2+1} G_{\mathcal{S}_2^c,\mathcal{S}_1'} G_{\mathcal{S}_2^c,\mathcal{S}_1'}^T + I\Big|$$

and hence, that $f_{\text{in}}(\mathcal{S}_1', \mathcal{S}_2) \le f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2)$, since the other terms remain the same. □

Lemma 2.2.1 immediately implies that

$$\min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1', \mathcal{S}_2) \le \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2). \tag{2.8}$$

Remark 2.2.2. We can establish (2.8) directly from Proposition 2.2.1, which implies that it continues to hold for the general inner bound. Moreover, we can show that the inner bound (2.5) is a polymatroid for each fixed $\sigma^2 > 0$.

We now specialize the cutset bound in Proposition 2.2.2 to the Gaussian uplink C-RAN model.

Lemma 2.2.2. *The cutset bound* (2.4) *can be simplified and relaxed for the Gaussian model as*

$$\begin{aligned}\sum_{k\in\mathcal{S}_1} R_k &\le \frac{1}{2}\log\left|PG_{\mathcal{S}_2^c,\mathcal{S}_1}G^T_{\mathcal{S}_2^c,\mathcal{S}_1}+I\right|+\sum_{l\in\mathcal{S}_2} C_l \\ &=: f_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2). \end{aligned} \tag{2.9}$$

Proof. Continuing from (2.4), we have

$$\begin{aligned}
\sum_{k\in\mathcal{S}_1} R_k &\le I(X(\mathcal{S}_1);Y(\mathcal{S}_2^c)\,|\,X(\mathcal{S}_1^c),Q)+\sum_{l\in\mathcal{S}_2} C_l \\
&\overset{(a)}{=} h(Y(\mathcal{S}_2^c)\,|\,X(\mathcal{S}_1^c),Q)-\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&= h\left(G_{\mathcal{S}_2^c,\mathcal{S}_1}X(\mathcal{S}_1)+G_{\mathcal{S}_2^c,\mathcal{S}_1^c}X(\mathcal{S}_1^c)+Z(\mathcal{S}_2^c)\,|\,X(\mathcal{S}_1^c),Q\right)-\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&= h\left(G_{\mathcal{S}_2^c,\mathcal{S}_1}X(\mathcal{S}_1)+Z(\mathcal{S}_2^c)\,|\,X(\mathcal{S}_1^c),Q\right)-\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&\overset{(b)}{=} h(G_{\mathcal{S}_2^c,\mathcal{S}_1}X(\mathcal{S}_1)+Z(\mathcal{S}_2^c)\,|\,Q)-\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&= \sum_q h(G_{\mathcal{S}_2^c,\mathcal{S}_1}X(S_1)+Z(\mathcal{S}_2^c)\,|\,Q=q)p(q)-\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&\overset{(c)}{\le} \sum_q \frac{1}{2}\log\left((2\pi e)^{|\mathcal{S}_2^c|}\left|G_{\mathcal{S}_2^c,\mathcal{S}_1}\,\mathsf{E}\left[X(\mathcal{S}_1)X(\mathcal{S}_1)^T\,|\,Q=q\right]G^T_{\mathcal{S}_2^c,\mathcal{S}_1}+I\right|\right)p(q) \\
&\qquad\qquad -\frac{|\mathcal{S}_2^c|}{2}\log(2\pi e)+\sum_{l\in\mathcal{S}_2} C_l \\
&\overset{(d)}{=} \sum_q \frac{1}{2}\log\left|G_{\mathcal{S}_2^c,\mathcal{S}_1}K'_{\mathcal{S}_1}(q)G^T_{\mathcal{S}_2^c,\mathcal{S}_1}+I\right|p(q)+\sum_{l\in\mathcal{S}_2} C_l
\end{aligned}$$

$$\overset{(e)}{\leq} \frac{1}{2}\log\left|G_{\mathcal{S}_2^c,\mathcal{S}_1}K_{\mathcal{S}_1}G_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_2} C_l$$
$$\leq \frac{1}{2}\log\left|PG_{\mathcal{S}_2^c,\mathcal{S}_1}G_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_2} C_l.$$

Here, (a) follows since $Y(\mathcal{S}_2^c)$ is an i.i.d. Gaussian vector given X^K, (b) follows since $X(\mathcal{S}_1)$ and $X(\mathcal{S}_1^c)$ are conditionally independent given Q, (c) follows from the (vector) maximum entropy principle, and in (d), $K'_{\mathcal{S}_1}(q)$ is a diagonal matrix consisting of $\{\mathsf{E}[X_k^2\,|\,Q=q], k\in\mathcal{S}_1\}$. In (e), $K_{\mathcal{S}_1}$ is a diagonal matrix consisting of $\{\mathsf{E}[X_k^2], k\in\mathcal{S}_1\}$, and (e) follows from the concavity of the log–determinant function of a symmetric matrix. Finally, the last inequality follows since each diagonal entry of $K_{\mathcal{S}_1}$ is upper-bounded by P. □

Our main goal of this section is to quantify how well network compress–forward performs for the Gaussian network, by comparing its achievable rates in (2.5) with the cutset bound in (2.9). In particular, we establish the following result.

Theorem 2.2.1. *For every $G\in\mathbb{R}^{L\times K}$ and every $P\in\mathbb{R}^+$, if a rate tuple $(R_1,\ldots,R_K)$ is in the cutset bound* (2.9)*, then the rate tuple $((R_1-\Delta)^+,\ldots,(R_K-\Delta)^+)$ is achievable, where*

$$\Delta \leq \frac{1}{2}\log(eL) \approx \frac{1}{2}\log L + 0.722.$$

Moreover, the sum-rate gap between the cutset bound and the network compress–forward inner bound is upper-bounded as

$$\Delta_{\text{sum}} := R_{\text{sum}}^{CS} - \sup_{\sigma^2>0} R_{\text{sum}}^{NCF}(\sigma^2) \leq \begin{cases} \frac{L}{2}H(K/L) \leq \frac{K}{2}\log(eL/K), & L\geq 2K, \\ \frac{L}{2}, & L<2K, \end{cases}$$

irrespective of P and G, where $H(\cdot)$ is the binary entropy function.

Remark 2.2.3. Theorem 2.2.1 tightens the existing sum-rate gap of $L/2$ bits per real dimension [69].

Proof of Theorem 2.2.1. Let

$$\Delta := \max_{\substack{\mathcal{S}_1 \subseteq [K] \\ \mathcal{S}_1 \neq \emptyset}} \frac{\min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2) - \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2)}{|\mathcal{S}_1|}. \tag{2.10}$$

Suppose that $(R_1, \ldots, R_K)$ lies in the cutset bound, and let $\mathcal{A} = \{k : R_k > \Delta\}$. Then, for every nonempty $\mathcal{S}_1 \subseteq [K]$,

$$\begin{aligned}
\sum_{k \in \mathcal{S}_1} (R_k - \Delta)^+ &= \sum_{k \in \mathcal{S}_1 \cap \mathcal{A}} (R_k - \Delta) \\
&= \sum_{k \in \mathcal{S}_1 \cap \mathcal{A}} R_k - |\mathcal{S}_1 \cap \mathcal{A}| \Delta \\
&\overset{(a)}{\leq} \min_{\mathcal{S}_2} \left[f_{\text{out}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2) \right] - \left(\min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2) - \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2) \right) \\
&= \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2) \\
&\overset{(b)}{\leq} \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2),
\end{aligned}$$

where (a) follows from the cutset bound (2.9) and the fact that

$$\begin{aligned}
\Delta &= \max_{\mathcal{S}_1} \frac{\min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2) - \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2)}{|\mathcal{S}_1|} \\
&\geq \frac{\min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2) - \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1 \cap \mathcal{A}, \mathcal{S}_2)}{|\mathcal{S}_1 \cap \mathcal{A}|},
\end{aligned}$$

and (b) follows from (2.8). Hence, Δ, as defined in (2.10), satisfies the requirements of Theorem 2.2.1. Now, for every $\sigma^2 > 0$,

$$\begin{aligned}
\Delta &= \max_{\mathcal{S}_1} \frac{\min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2) - \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2)}{|\mathcal{S}_1|} \\
&\overset{(a)}{\leq} \max_{\mathcal{S}_1, \mathcal{S}_2} \frac{f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2) - f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2)}{|\mathcal{S}_1|}
\end{aligned}$$

$$
\begin{aligned}
&\stackrel{(b)}{=} \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{1}{2|\mathcal{S}_1|} \log \frac{\left| PG_{\mathcal{S}_2^c,\mathcal{S}_1} G^T_{\mathcal{S}_2^c,\mathcal{S}_1} + I \right|}{\left| \frac{P}{\sigma^2+1} G_{\mathcal{S}_2^c,\mathcal{S}_1} G^T_{\mathcal{S}_2^c,\mathcal{S}_1} + I \right|} \right] \\
&\stackrel{(c)}{=} \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{1}{2|\mathcal{S}_1|} \sum_{i=1}^{\operatorname{rank}(G_{\mathcal{S}_2^c,\mathcal{S}_1})} \log \frac{P\beta_i + 1}{\frac{P}{\sigma^2+1}\beta_i + 1} + |\mathcal{S}_2| \log \left(1 + \frac{1}{\sigma^2}\right) \right] \\
&\stackrel{(d)}{\leq} \max_{\substack{k\in[K] \\ l\in\{0,\ldots,L\}}} \left[\frac{\min\{L-l,k\}}{2k} \log(1+\sigma^2) + \frac{l}{2k} \log\left(1+\frac{1}{\sigma^2}\right) \right]. \qquad (2.11)
\end{aligned}
$$

Here, (a) follows from the fact that for functions f and g defined over a finite set $\mathcal{X}$, such that $g \geq f$ everywhere on $\mathcal{X}$, $\min_{x\in\mathcal{X}} g(x) - \min_{x\in\mathcal{X}} f(x) \leq \max_{x\in\mathcal{X}}[g(x) - f(x)]$, (b) follows from (2.5) and (2.9), and in (c), $\beta_1, \beta_2, \ldots$ are the (nonnegative) eigenvalues of $G_{\mathcal{S}_2^c,\mathcal{S}_1} G^T_{\mathcal{S}_2^c,\mathcal{S}_1}$. Finally, in (d), we take $|\mathcal{S}_1| = k$, $|\mathcal{S}_2| = l$, and upper-bound $\operatorname{rank}(G_{\mathcal{S}_2^c,\mathcal{S}_1})$ by $\min\{L-l,k\}$. The maximization in (2.11) yields

$$
\Delta \leq \begin{cases} \frac{1}{2}\log(\sigma^2+1) + \frac{L-1}{2}\log(1+\frac{1}{\sigma^2}), & \sigma^2 \geq 1, \\ \frac{L}{2}\log(1+\frac{1}{\sigma^2}), & \sigma^2 \leq 1. \end{cases}
$$

Since this holds for every $\sigma^2 > 0$, we set $\sigma^2 = L-1$ for $L \geq 2$ to obtain

$$
\begin{aligned}
\Delta &\leq \frac{1}{2}\log L + \frac{L-1}{2}\log\left(1+\frac{1}{L-1}\right) \\
&\stackrel{(a)}{\leq} \frac{1}{2}\log L + \frac{L-1}{2} \cdot \frac{1}{L-1}\log e \\
&\leq \frac{1}{2}\log(eL). \qquad (2.12)
\end{aligned}
$$

Here, (a) follows since from elementary calculus, we know that for $x > 0$, $\log(1+x) \leq x \log e$.

For $L = 1$, we can choose $\sigma^2 = 1$ to obtain $\Delta \leq 1$. This, together with (2.12), establishes the first part of Theorem 2.2.1. For the sum-rate gap, we simply consider

$$
\Delta_{\text{sum}} \leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left(f_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2) - f_{\text{in}}(\mathcal{S}_1,\mathcal{S}_2) \right)
$$

$$\leq \max_{\substack{k\in[K]\\ l\in\{0,\ldots,L\}}} \left[\frac{\min\{L-l,k\}}{2}\log(1+\sigma^2)+\frac{l}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right]. \tag{2.13}$$

Maximization of (2.13) over l and k yields, for $\sigma^2 \geq 1$,

$$\Delta_{\text{sum}} \leq \begin{cases} \frac{K}{2}\log(1+\sigma^2)+\frac{L-K}{2}\log(1+\frac{1}{\sigma^2}), & L \geq K, \\ \frac{L}{2}\log(1+\sigma^2), & L < K. \end{cases}$$

For $L \leq 2K$, we can then choose $\sigma^2 = 1$ to obtain an upper bound $\Delta_{\text{sum}} \leq L/2$. For $L > 2K$, we can choose $\sigma^2 = L/K - 1 \geq 1$ to obtain

$$\begin{aligned} \Delta_{\text{sum}} &\leq \frac{K}{2}\log\left(\frac{L}{K}\right)+\frac{L-K}{2}\log\left(1+\frac{K}{L-K}\right) \\ &= \frac{L}{2}\left(\frac{K}{L}\log\left(\frac{L}{K}\right)+\left(1-\frac{K}{L}\right)\log\frac{1}{1-\frac{K}{L}}\right) \\ &= \frac{L}{2}H(K/L), \end{aligned} \tag{2.14}$$

completing the proof. □

2.2.3 Proofs of General Bounds on the Capacity Region

Proof of Proposition 2.2.1. Our analysis of the coding scheme follows that in [29] but is considerably simpler because of the relative simplicity of our network model. We omit the time-sharing sequence q^n for simplicity of notation.

Without loss of generality, let $m^K = (1,\ldots,1)$ be the messages sent. Then the error events are:

$$\mathcal{E}_0 = \left\{ \left(Y_l^n, \hat{Y}_l^n(w_l,t_l)\right) \notin \mathcal{T}_{\varepsilon'}^{(n)} \text{ for all } (w_l,\ t_l,) \text{ for some } l \right\}.$$

$$\mathcal{E}_1 = \left\{ \left(X_1^n(1),\ldots,X_K^n(1),\hat{Y}_1^n(W_1,t_1),\ldots,\hat{Y}_L^n(W_L,t_L)\right) \notin \mathcal{T}_{\varepsilon}^{(n)} \text{ for all } t^L \right\}.$$

$$\mathcal{E}_2 = \left\{ \left(X_1^n(m_1),\ldots,X_K^n(m_K),\hat{Y}_1^n(W_1,t_1),\ldots,\hat{Y}_L^n(W_L,t_L)\right) \in \mathcal{T}_{\varepsilon}^{(n)} \text{ for some } t^L \text{ and some } m^K \neq (1,\ldots,1) \right\}.$$

Here, $(W_1,\ldots,W_L)$ represent the indices transmitted by the relays. By the packing lemma and union of events, $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$ if

$$C_l + \hat{R}_l > I(Y_l;\hat{Y}_l) \tag{2.15}$$

for all $l \in [L]$.

By the Markov lemma [17, Lemma 12.1] and union of events bound ($\hat{Y}_l \to Y_l \to X^K$ form a Markov chain), $\mathsf{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c) \to 0$ as $n \to \infty$.

To analyze $\mathsf{P}(\mathcal{E}_2)$, let $t^L = (1,\ldots,1)$ be the t-indices chosen at the relays. Then, by the union of events bound,

$$\begin{aligned} \mathsf{P}(\mathcal{E}_2) &\leq \sum_{\substack{m^K,t^L \\ m^K \neq (1,\ldots,1)}} \mathsf{P}\left(\left(X_1^n(m_1),\ldots,X_K^n(m_K),\hat{Y}_1^n(W_1,t_1),\ldots,\hat{Y}_L^n(W_L,t_L)\right) \in \mathcal{T}_{\varepsilon}^{(n)} \right) \\ &=: \sum_{\substack{m^K,t^L: \\ m^K \neq (1,\ldots,1)}} p_{m^K,t^L}. \end{aligned} \tag{2.16}$$

In order to bound each term on the right-hand side of (2.16), we need the following generalization of the joint typicality Lemma.

Lemma 2.2.3. *[29, Lemma 2] Let $(X^N,Y^N,Z) \sim p(x^N,y^N,z)$. Let the n-length random vector*

$\hat{\mathbf{Z}}$ be distributed according to some arbitrary pmf $p(\hat{\mathbf{z}})$ and let

$$(\hat{X}_1^n, \ldots, \hat{X}_N^n, \hat{Y}_1^n, \ldots, \hat{Y}_N^n) \sim \prod_{i=1}^{n} p_{X^N}(\hat{x}_{1i}, \ldots, \hat{x}_{Ni}) \prod_{k=1}^{N} \prod_{i=1}^{n} p_{Y_k|X_k}(\hat{y}_{ki}|\hat{x}_{ki})$$

be distributed independently of $\hat{\mathbf{Z}}$. Then, there exists $\delta(\varepsilon)$ that tends to zero as $\varepsilon \to 0$, such that

$$P\Big((\hat{\mathbf{Z}}, \hat{X}_1^n, \ldots, \hat{X}_N^n, \hat{Y}_1^n, \ldots, \hat{Y}_N^n) \in \mathcal{T}_\varepsilon^{(n)}\Big) \leq 2^{-n[I(Z;X^N) + \sum_{k=1}^{N} I(Y_k;X^N,Y^{k-1},Z|X_k) - \delta(\varepsilon)]}.$$

For a given t^L and $m^K \neq (1, \ldots, 1)$, let $\mathcal{S}_2(t^L) = \{l \in [L] : t_l \neq 1\}$ and $\mathcal{S}_1(m^K) = \{k \in [K] : m_k \neq 1\}$. Then, $(X^n(\mathcal{S}_1(m^K)), \hat{Y}^n(\mathcal{S}_2(t^L)))$ is independent of $(X^n(\mathcal{S}_1^c(m^K)), \hat{Y}^n(\mathcal{S}_2^c(t^L)))$. Then, using $(X^n(\mathcal{S}_1^c(m^K)), \hat{Y}^n(\mathcal{S}_2^c(t^L)))$ as $\hat{\mathbf{Z}}^n$ in Lemma 2.2.3, we obtain

$$\begin{aligned} & p_{m^K,t^L} \\ & \leq 2^{-n\left[I(X(\mathcal{S}_1(m^K)); X(\mathcal{S}_1^c(m^K)), \hat{Y}(\mathcal{S}_2^c(t^L)))\right]} \times \\ & \qquad 2^{-n\left[\sum_{l \in \mathcal{S}_2(t^L)} I(\hat{Y}_l; X(\mathcal{S}_1(m^K)), \hat{Y}([l-1] \cap \mathcal{S}_2(t^L)), X(\mathcal{S}_1^c(m^K)), \hat{Y}(\mathcal{S}_2^c(t^L))) - \delta(\varepsilon)\right]}. \end{aligned}$$

The terms in the exponent of the right-hand expression (excluding the factor of $-n$) are given by

$$
\begin{aligned}
& I(X(\mathcal{S}_1(m^K));X(\mathcal{S}_1^c(\mathbf{m})),\hat{Y}(\mathcal{S}_2^c(t^L))) \\
&\qquad + \sum_{l\in\mathcal{S}_2(t^L)} I(\hat{Y}_l;X(\mathcal{S}_1(m^K)),\hat{Y}([l-1]\cap\mathcal{S}_2(t^L)),X(\mathcal{S}_1^c(m^K)),\hat{Y}(\mathcal{S}_2^c(t^L))) - \delta(\varepsilon) \\
&\overset{(a)}{=} I(X(\mathcal{S}_1(m^K));\hat{Y}(\mathcal{S}_2^c(t^L))\,|\,X(\mathcal{S}_1^c(m^K))) + \sum_{l\in\mathcal{S}_2(t^L)} I(Y_l;\hat{Y}_l) \\
&\qquad - \sum_{l\in\mathcal{S}_2(t^L)} \big(I(Y_l;\hat{Y}_l) - I(\hat{Y}_l;\hat{Y}([l-1]\cap\mathcal{S}_2(t^L)),\hat{Y}(\mathcal{S}_2^c(t^L)),X^K)\big) - \delta(\varepsilon) \\
&\overset{(b)}{=} I(X(\mathcal{S}_1(m^K));\hat{Y}(\mathcal{S}_2^c(t^L))\,|\,X(\mathcal{S}_1^c(m^K))) + \sum_{l\in\mathcal{S}_2(t^L)} I(Y_l;\hat{Y}_l) \\
&\qquad - \sum_{l\in\mathcal{S}_2(t^L)} I\big(Y_l;\hat{Y}_l\,|\,\hat{Y}([l-1]\cap\mathcal{S}_2(t^L)),\hat{Y}(\mathcal{S}_2^c(t^L)),X^K\big) - \delta(\varepsilon).
\end{aligned}
$$

Here, (b) follows from the fact that $\big(\hat{Y}([l-1]\cap\mathcal{S}_2(t^L)),\hat{Y}(\mathcal{S}_2^c(t^L)),X^K\big) \to Y_l \to \hat{Y}_l$ form a Markov chain and (a) follows from the independence of $X(S_1(m^K))$ and $X(S_1^c(m^K))$.

Defining

$$
\begin{aligned}
& J(\mathcal{S}_1,\mathcal{S}_2) \\
& := I(X(\mathcal{S}_1);\hat{Y}(\mathcal{S}_2^c)|X(\mathcal{S}_1^c)) + \sum_{l\in\mathcal{S}_2} I(Y_l;\hat{Y}_l) - \sum_{l\in\mathcal{S}_2} I\left(Y_l;\hat{Y}_l\middle|\hat{Y}([l-1]\cap\mathcal{S}_2),\hat{Y}(\mathcal{S}_2^c),X^K\right)
\end{aligned}
$$

and continuing (2.16), we have

$$
\begin{aligned}
\mathsf{P}(\mathcal{E}_2) &\le \sum_{\substack{m^K,t^L:\\ m^K\neq(1,\dots,1)}} 2^{-n[J(\mathcal{S}_1(m^K),\mathcal{S}_2(t^L))-\delta(\varepsilon)]} \\
&\le \sum_{\substack{\mathcal{S}_1\subseteq[K]\\ \mathcal{S}_2\subseteq[L]\\ \mathcal{S}_1\neq\emptyset}} 2^{-n[J(\mathcal{S}_1,\mathcal{S}_2)-\sum_{k\in\mathcal{S}_1}R_k-\sum_{l\in\mathcal{S}_2}\hat{R}_l-\delta(\varepsilon)]}.
\end{aligned}
$$

Therefore, $\mathsf{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if

$$\sum_{k\in\mathcal{S}_1} R_k + \sum_{l\in\mathcal{S}_2} \hat{R}_l < J(\mathcal{S}_1,\mathcal{S}_2) \tag{2.17}$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ such that $\mathcal{S}_1 \neq \emptyset$. Combining (2.17) with (2.15) to eliminate the auxiliary rates $(\hat{R}_1,\dots,\hat{R}_L)$, we obtain the inequalities

$$\sum_{k\in\mathcal{S}_1} R_k < I(X(\mathcal{S}_1);\hat{Y}(\mathcal{S}_2^c)|X(\mathcal{S}_1^c)) + \sum_{l\in\mathcal{S}_2} C_l - \sum_{l\in\mathcal{S}_2} I\left(Y_l;\hat{Y}_l \middle| \hat{Y}\left([l-1]\cap\mathcal{S}_2\right),\hat{Y}(\mathcal{S}_2^c),X^K\right) \tag{2.18}$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ such that $\mathcal{S}_1 \neq \emptyset$. □

Proof of Proposition 2.2.2. For $k \in [K]$, let M_k denote the message communicated by sender k and let W_l denote the index sent by relay l to the central processor. Also, for $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$, denote by $X_i(\mathcal{S}_1)$ the tuple $(X_{ki}, k \in \mathcal{S}_1)$ and by $Y_i(\mathcal{S}_2)$ the tuple $(Y_{li}, l \in \mathcal{S}_2)$. Similarly, $X^n(\mathcal{S}_1)$ stands for $(X_{ki}, k \in \mathcal{S}_1, i \in [n])$ and $Y^n(\mathcal{S}_2)$ stands for $(Y_{li}, l \in \mathcal{S}_2, i \in [n])$. Then, for every $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$, $X^n(\mathcal{S}_1)$ is a function of $M(\mathcal{S}_1)$ and $W(\mathcal{S}_2)$ is a function of $Y^n(\mathcal{S}_2)$. For every $\mathcal{S}_1 \subseteq [K], \mathcal{S}_1 \neq \emptyset$ and $\mathcal{S}_2 \subseteq [L]$, we must have, by Fano's inequality,

$$H(M(\mathcal{S}_1)\,|\,M(\mathcal{S}_1^c),Y^n(\mathcal{S}_2^c),W^L) \le H(M(\mathcal{S}_1)\,|\,W^L) \le n\varepsilon_n,$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Therefore, since $H(M(\mathcal{S}_1)\,|\,M(\mathcal{S}_1^c)) = n\sum_{k\in\mathcal{S}_1} R_k$, we have

$$
\begin{aligned}
& n \sum_{k\in\mathcal{S}_1} R_k \\
& \le I(M(\mathcal{S}_1); Y^n(\mathcal{S}_2^c), W^L \,|\, M(\mathcal{S}_1^c)) + n\varepsilon_n \\
& \overset{(a)}{=} I(M(\mathcal{S}_1); Y^n(\mathcal{S}_2^c), W(\mathcal{S}_2) \,|\, M(\mathcal{S}_1^c)) + n\varepsilon_n \\
& = I(M(\mathcal{S}_1); Y^n(\mathcal{S}_2^c) \,|\, M(\mathcal{S}_1^c)) + I(M(\mathcal{S}_1); W(\mathcal{S}_2) \,|\, M(\mathcal{S}_1^c), Y^n(\mathcal{S}_2^c)) + n\varepsilon_n \\
& \le \sum_{i=1}^{n} I(M(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, M(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c)) + H(W(\mathcal{S}_2)) + n\varepsilon_n \\
& \overset{(b)}{=} \sum_{i=1}^{n} I(M(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, M(\mathcal{S}_1^c), X_i(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c)) + H(W(\mathcal{S}_2)) + n\varepsilon_n \\
& \le \sum_{i=1}^{n} I(M(\mathcal{S}_1), M(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_2)) + n\varepsilon_n \\
& \overset{(c)}{=} \sum_{i=1}^{n} I(M(\mathcal{S}_1), M(\mathcal{S}_1^c), X_i(\mathcal{S}_1), Y^{i-1}(\mathcal{S}_2^c); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_2)) + n\varepsilon_n \\
& \overset{(d)}{=} \sum_{i=1}^{n} I(X_i(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_2)) + n\varepsilon_n \\
& \overset{(e)}{\le} \sum_{i=1}^{n} I(X_i(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + n \sum_{l\in\mathcal{S}_2} C_l + n\varepsilon_n.
\end{aligned}
$$

Here, (a) follows since $W(\mathcal{S}_2^c)$ is a function of $Y^n(\mathcal{S}_2^c)$, (b) follows since $X_i(\mathcal{S}_1^c)$ is a function of $M(\mathcal{S}_1^c)$, (c) follows since $X_i(\mathcal{S}_1)$ is a function of $M(\mathcal{S}_1)$, (d) follows since $(M(\mathcal{S}_1), M(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c)) \to (X_i(\mathcal{S}_1), X_i(\mathcal{S}_1^c)) \to (Y_i(\mathcal{S}_2^c))$ form a Markov chain (by the memorylessness of the first hop), and (e) follows since $W(\mathcal{S}_2)$ is supported on a set of size $\prod_{l\in\mathcal{S}_2} 2^{nC_l}$. Defining a random variable $Q \sim \mathrm{Unif}([n])$ independent of all the other random variables, writing $X(\mathcal{S}_1) := X_Q(\mathcal{S}_1)$ and $Y(\mathcal{S}_2) := Y_Q(\mathcal{S}_2)$, and letting n tend to infinity leads to

$$
\sum_{k\in\mathcal{S}_1} R_k \le I(X(\mathcal{S}_1); Y(\mathcal{S}_2^c) \,|\, X(\mathcal{S}_1^c), Q) + \sum_{l\in\mathcal{S}_2} C_l \tag{2.19}
$$

for all $\mathcal{S}_1 \subseteq [K], \mathcal{S}_1 \neq \emptyset, \mathcal{S}_2 \subseteq [L]$ for some pmf $p(q)\prod_{k\in[K]} p(x_k\,|\,q)$, and thus completes the proof. □

2.3 Downlink C-RANs

2.3.1 General Model

Similar to the uplink case, we model the downlink C-RAN as a two-hop relay network in Fig. 2.2, where the first hop (central processor to radio heads) consists of orthogonal noiseless links of capacities $C_1, \ldots, C_L$ bits per real dimension and the second hop (radio heads to user devices or receivers) is modeled as a discrete memoryless network $p(y^K \mid x^L)$.

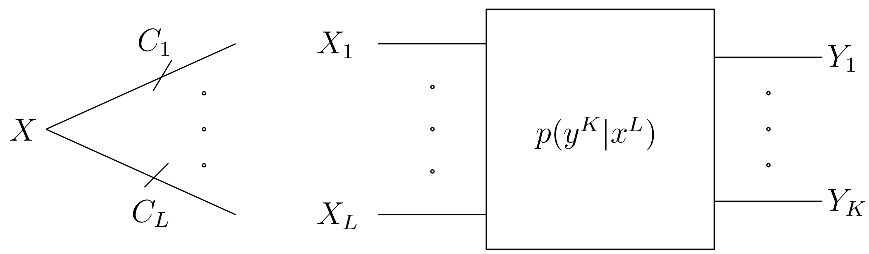


Figure 2.2. Downlink network model.

A $(2^{nR_1}, \ldots, 2^{nR_K}, n)$ code for this network consists of K message sets $[2^{nR_1}], \ldots, [2^{nR_k}]$; an encoder $w^L(m_1, \ldots, m_K) \in \prod_{l=1}^{L} [2^{nC_l}]$; relay encoders $x_l^n(w_l)$, $l \in [L]$; and decoders $\hat{m}_k(y_k^n) \in [2^{nR_k}]$, $k \in [K]$. The average probability of error, achievability of a rate tuple, and the capacity region are defined similar to Section 2.2.1.

The distributed decode–forward coding scheme [28] can be specialized to the downlink C-RAN model as follows.

Codebook generation. Fix a pmf $p(x^L, u^K)$. For each $w_l \in [2^{nC_l}], l \in [L]$, generate $x_l^n(w_l) \sim \prod_{i=1}^{n} p_{X_l}(x_{li})$. Define auxiliary indices $s_k \in [2^{n\tilde{R}_k}], k \in [K]$, for some auxiliary rates $(\tilde{R}_k, k \in [k])$. For each $(m_k, s_k) \in [2^{nR_k}] \times [2^{n\tilde{R}_k}]$ and $k \in [K]$, generate $u_k^n(m_k, s_k) \sim \prod_{i=1}^{n} p_{U_k}(u_{ki})$.

Encoding. To transmit messages $m^K = (m_1, \ldots, m_K)$, the encoder transmits $w^L = (w_1, \ldots, w_L)$, such that

$$(x_1^n(w_1), \ldots, x_L^n(w_L), u_1^n(m_1, s_1), \ldots, u_K^n(m_K, s_K)) \in \mathcal{T}_{\epsilon'}^{(n)}$$

for some $s^K \in [2^{n\tilde{R}_1}] \times \cdots \times [2^{n\tilde{R}_K}]$. The encoding succeeds w.h.p. if

$$\sum_{l \in \mathcal{S}_1^c} C_l + \sum_{k \in \mathcal{S}_2^c} \tilde{R}_k > I^*(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \tag{2.20}$$

for every $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$.

Relaying. On receiving the index w_l, relay l transmits $x_l^n(w_l)$.

Decoding. Let $\varepsilon > \varepsilon'$. Upon receiving y_k^n, receiver k finds a message estimate $\hat{m}_k$ such that

$$(u_k^n(\hat{m}_k, s_k), y_k^n) \in \mathcal{T}_\varepsilon^{(n)}$$

for some s_k. The decoding at receiver k succeeds w.h.p. if

$$R_k + \tilde{R}_k < I(U_k; Y_k). \tag{2.21}$$

Combining (2.20) and (2.21) to eliminate the auxiliary rates $\tilde{R}_1, \ldots, \tilde{R}_K$ leads to the following inner bound on the capacity region of this network. (See Section 2.3.3 for a complete proof.)

Proposition 2.3.1 (Distributed decode–forward inner bound for the downlink C-RAN)**.** *A rate tuple $(R_1, \ldots, R_K)$ is achievable for the downlink C-RAN if*

$$\begin{aligned}\sum_{k \in \mathcal{S}_2^c} R_k < I(X(\mathcal{S}_1); U(\mathcal{S}_2^c) | X(\mathcal{S}_1^c)) + \sum_{l \in \mathcal{S}_1^c} C_l - \sum_{k \in \mathcal{S}_2^c} I(U_k; X^L | Y_k) \\ - I^*(U(\mathcal{S}_2^c) \,|\, X^L) - I^*(X(\mathcal{S}_1^c))\end{aligned} \tag{2.22}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$ for some pmf $p(x^L, u^K)$, such that

$$I^*(X(\mathcal{S}_1)) \leq \sum_{l \in \mathcal{S}_1} C_l$$

for all $\mathcal{S}_1 \subseteq [L]$.

Remark 2.3.1. As the fronthaul capacities $C_1, \ldots, C_L$ tend to infinity, this downlink C-RAN channel model becomes identical to the "fronthaul-unlimited" downlink channel from a single sender with L transmit antennas to K receivers, i.e., the broadcast channel $p(y^K \,|\, x^L)$ with one sender X^L and K receivers $Y_1, \ldots, Y_K$. In this regime, the distributed decode–forward inner bound converges to the Marton coding inner bound with no common messages [17, Theorem 8.3], characterized by rate tuples $(R_1, \ldots, R_K)$ satisfying

$$\sum_{k \in \mathcal{S}_2} R_k \leq \sum_{k \in \mathcal{S}_2} I(U_k; Y_k) - I^*(U(\mathcal{S}_2))$$

for every $\mathcal{S}_2 \subseteq [K]$ for some pmf $p(u^K)$ and some function $x^L(u^K)$.

Specializing the cutset bound [18] to the downlink C-RAN model leads to the following.

Proposition 2.3.2 (Cutset outer bound for the downlink C-RAN)**.** *If a rate tuple $(R_1, \ldots, R_K)$ is achievable for the downlink C-RAN, then*

$$\sum_{k \in \mathcal{S}_2^c} R_k \leq I(X(\mathcal{S}_1); Y(\mathcal{S}_2^c) | X(\mathcal{S}_1^c)) + \sum_{l \in \mathcal{S}_1^c} C_l \tag{2.23}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$ for some pmf $p(x^L)$.

We provide a proof of Proposition 2.3.2 in Section 2.3.3.

2.3.2 Gaussian Model

We now assume that the second hop of the network is Gaussian, i.e., $Y^K = HX^L + Z^K$, where $H \in \mathbb{R}^{K \times L}$ is a channel gain matrix and Z^K is a vector of i.i.d. $\mathrm{N}(0,1)$ noise components. We also impose the average power constraint P on each relay. For this Gaussian network model, the distributed decode–forward inner bound in Proposition 2.3.1 can be specialized to establish the achievability of all rate tuples $(R_1,\ldots,R_K)$ such that

$$\sum_{k\in\mathcal{S}_2^c} R_k \le \frac{1}{2}\log\left|\frac{P}{\sigma^2}H_{\mathcal{S}_2^c,\mathcal{S}_1}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_1^c} C_l - \frac{|\mathcal{S}_2^c|}{2}\log\left(1+\frac{1}{\sigma^2}\right) =: F_{\text{in}}(\mathcal{S}_1,\mathcal{S}_2) \tag{2.24}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$ for some $\sigma^2 > 0$. This follows by setting X^L to be a vector of i.i.d. $\mathrm{N}(0,P)$ random variables and defining $U^K = GX^L + \hat{Z}^K$, where $\hat{Z}^K \sim \mathrm{N}(0,\sigma^2 I)$ is independent of Z^K. For every $\sigma^2 > 0$, we denote the set of rate tuples $(R_1,\ldots,R_K)$ satisfying (2.24) by $\mathscr{R}_{\text{down}}^{\text{DDF}}(\sigma^2)$. We also denote the achievable sum-rate for each $\sigma^2 > 0$ by

$$R_{\text{sum}}^{\text{DDF}}(\sigma^2) := \sup_{(R_1,\ldots,R_K)} \{R_1+\cdots+R_K : (R_1,\ldots,R_K) \in \mathscr{R}_{\text{down}}^{\text{DDF}}(\sigma^2)\} \tag{2.25}$$

$$= \min_{\mathcal{S}_1\subseteq[L]} \Big(\frac{1}{2}\log\left|\frac{P}{\sigma^2}H_{[K],\mathcal{S}_1}H_{[K],\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_1^c} C_l\Big) - \frac{K}{2}\log\big(1+\frac{1}{\sigma^2}\big). \tag{2.26}$$

Similar to Section 2.2.2, the cutset bound in Proposition 2.3.2 can be specialized to the rate region characterized by

$$\sum_{k\in\mathcal{S}_2^c} R_k \le \frac{1}{2}\log\left|H_{\mathcal{S}_2^c,\mathcal{S}_1}\Gamma_{\mathcal{S}_1|\mathcal{S}_1^c}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right| + \sum_{l\in\mathcal{S}_1^c} C_l$$

$$=: F_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2) \tag{2.27}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$ for some covariance matrix $\Gamma \succeq 0$ satisfying $\Gamma_{ll} \le P$ for all $l \in [L]$. Here, $\Gamma_{\mathcal{S}_1|\mathcal{S}_1^c}$ is the conditional covariance matrix given by

$$\Gamma_{\mathcal{S}_1|\mathcal{S}_1^c} = \Gamma_{\mathcal{S}_1,\mathcal{S}_1} - \Gamma_{\mathcal{S}_1,\mathcal{S}_1^c}\Gamma_{\mathcal{S}_1^c,\mathcal{S}_1^c}^{-1}\Gamma_{\mathcal{S}_1^c,\mathcal{S}_1}.$$

For each covariance matrix Γ, we denote the set of rate tuples $(R_1,\ldots,R_K)$ satisfying (2.27) by

$\mathscr{R}_{\text{down}}^{\text{CS}}(\Gamma)$. We denote the sum-rate upper bound by

$$R_{\text{sum}}^{\text{CS}} := \sup_{(R_1,\ldots,R_K),\Gamma} \{R_1 + \cdots + R_K : (R_1,\ldots,R_K) \in \mathscr{R}_{\text{down}}^{\text{CS}}(\Gamma) \text{ for some } \Gamma\}. \tag{2.28}$$

The achievable per-user rate gap Δ, as well as the sum-rate gap Δ_{sum} between the cutset bound and the distributed decode–forward inner bound (2.24), can be bounded as in the following result.

Theorem 2.3.1. *For every $H \in \mathbb{R}^{K\times L}$ and $P \in \mathbb{R}^+$, if a rate tuple $(R_1,\ldots,R_K)$ is in the cutset bound* (2.27)*, then the rate tuple $((R_1-\Delta)^+,\ldots,(R_K-\Delta)^+)$ is achievable, where*

$$\Delta \le \frac{1}{2}\log(eKL) \approx \frac{1}{2}\log(KL) + 0.722.$$

Moreover, the sum-rate gap between the cutset bound and the distributed decode–forward inner bound is upper-bounded as

$$\Delta_{\text{sum}} := R_{\text{sum}}^{CS} - \sup_{\sigma^2>0} R_{\text{sum}}^{DDF}(\sigma^2) \le \frac{K}{2} + \frac{\min\{L,K\}}{2}\log L$$

irrespective of P and H.

To prove Theorem 2.3.1, we need the following lemma, which is immediate from elementary calculus.

Lemma 2.3.1. *For $x > 1$, $x\log x - (x-1)\log(x-1) \le \log(ex)$.*

Proof. Let $f(x) = x\log x$ for $x > 0$. We then have $f'(x) = \log x + (1/\ln 2) = \log(ex)$, which is increasing in x. Therefore, for $x > 1$,

$$\begin{aligned} f(x) - f(x-1) &\le f'(x)\,(x-(x-1)) \\ &= \log(ex). \end{aligned}$$

□

Proof of Theorem 2.3.1. Note that unlike (2.8) in Section 2.2.2, F_{in} is not necessarily monotonic. We overcome this difficulty by rephrasing the inner bound (2.24) as

$$\sum_{k\in\mathcal{S}_2^c} R_k \le \min_{\mathcal{T}_2\subseteq\mathcal{S}_2} F_{\text{in}}(\mathcal{S}_1,\mathcal{T}_2). \tag{2.29}$$

We observe that the right-hand side of (2.29) is increasing with $\mathcal{S}_2^c$ for a fixed $\mathcal{S}_1$, so we can apply the technique developed in the proof of Theorem 2.2.1 to compute an upper bound on Δ. We thus write

$$\begin{aligned}
\Delta &= \max_{\mathcal{S}_2\subsetneq[K]}\left[\frac{\min_{\mathcal{S}_1} F_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2)}{|\mathcal{S}_2^c|} - \frac{\min_{\mathcal{S}_1}\min_{\mathcal{T}_2\subseteq\mathcal{S}_2} F_{\text{in}}(\mathcal{S}_1,\mathcal{T}_2)}{|\mathcal{S}_2^c|}\right] \\
&\le \max_{\substack{\mathcal{S}_1\subseteq[L]\\ \mathcal{S}_2\subsetneq[K]\\ \mathcal{T}_2\subseteq\mathcal{S}_2}} \frac{F_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2)-F_{\text{in}}(\mathcal{S}_1,\mathcal{T}_2)}{|\mathcal{S}_2^c|} \\
&= \max_{\substack{\mathcal{S}_1\subseteq[L]\\ \mathcal{S}_2\subsetneq[K]\\ \mathcal{T}_2\subseteq\mathcal{S}_2}} \frac{1}{2|\mathcal{S}_2^c|}\left[\log\frac{\left|H_{\mathcal{S}_2^c,\mathcal{S}_1}\Gamma_{\mathcal{S}_1|\mathcal{S}_1^c}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T+I\right|}{\left|\frac{P}{\sigma^2}H_{\mathcal{T}_2^c,\mathcal{S}_1}H_{\mathcal{T}_2^c,\mathcal{S}_1}^T+I\right|}+|\mathcal{T}_2^c|\log\left(1+\frac{1}{\sigma^2}\right)\right] \\
&\overset{(a)}{\le} \max_{\substack{\mathcal{S}_1\subseteq[L]\\ \mathcal{S}_2\subsetneq[K]\\ \mathcal{T}_2\subseteq\mathcal{S}_2}} \frac{1}{2|\mathcal{S}_2^c|}\left[\log\frac{\left|H_{\mathcal{S}_2^c,\mathcal{S}_1}\Gamma_{\mathcal{S}_1}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T+I\right|}{\left|\frac{P}{\sigma^2}H_{\mathcal{S}_2^c,\mathcal{S}_1}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T+I\right|}+|\mathcal{T}_2^c|\log\left(1+\frac{1}{\sigma^2}\right)\right],
\end{aligned} \tag{2.30}$$

where (a) follows since $\Gamma_{\mathcal{S}_1}\succeq\Gamma_{\mathcal{S}_1|\mathcal{S}_1^c}$ and for any matrix A and $\alpha>0$, $|I+\alpha AA^T|$ increases when we add more rows to A. Writing $\Gamma_{\mathcal{S}_1}=U\Lambda U^T$, where U is orthogonal and Λ is diagonal, and letting $H_{\mathcal{S}_2^c,\mathcal{S}_1}U=[b_1\quad b_2\quad\cdots\quad b_{|\mathcal{S}_1|}]$, where $b_1,\ldots,b_{|\mathcal{S}_1|}$ are $|\mathcal{S}_2^c|\times 1$ vectors satisfying

$\sum_{l=1}^{|\mathcal{S}_1|} \|b_l\|^2 = \left\| H_{\mathcal{S}_2^c,\mathcal{S}_1} \right\|_F^2$, we have

$$\begin{aligned}
\log \frac{|H_{\mathcal{S}_2^c,\mathcal{S}_1} \Gamma_{\mathcal{S}_1} H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I|}{|\frac{P}{\sigma^2} H_{\mathcal{S}_2^c,\mathcal{S}_1} H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I|} &= \log \frac{\left| I + \sum_{l=1}^{|\mathcal{S}_1|} \lambda_l b_l b_l^T \right|}{\left| I + \frac{P}{\sigma^2} \sum_{l=1}^{|\mathcal{S}_1|} b_l b_l^T \right|} \\
&\overset{(a)}{\leq} \log \frac{\left| I + P|\mathcal{S}_1| \sum_{l=1}^{|\mathcal{S}_1|} b_l b_l^T \right|}{\left| I + \frac{P}{\sigma^2} \sum_{l=1}^{|\mathcal{S}_1|} b_l b_l^T \right|} \\
&\overset{(b)}{=} \sum_{k=1}^{|\mathcal{S}_2^c|} \log \frac{1 + P|\mathcal{S}_1|\mu_k}{1 + \frac{P}{\sigma^2}\mu_k} \\
&\leq |\mathcal{S}_2^c| \log\left(\sigma^2 |\mathcal{S}_1|\right),
\end{aligned}$$

provided $\sigma^2 \geq \frac{1}{|\mathcal{S}_1|}$. Here, (a) follows since the trace of $\Gamma_{\mathcal{S}_1}$ is upper bounded by $P|\mathcal{S}_1|$ and in (b), $\mu_1, \ldots, \mu_{|\mathcal{S}_2^c|}$ are the (non-negative) eigenvalues of $\sum_{l=1}^{|\mathcal{S}_1|} b_l b_l^T$. Continuing from (2.30), we thus have

$$\begin{aligned}
\Delta &\leq \max_{\substack{\mathcal{S}_1 \subseteq [L] \\ \mathcal{S}_2 \subsetneq [K] \\ \mathcal{T}_2 \subseteq \mathcal{S}_2}} \left[\frac{|\mathcal{T}_2^c| \log\left(1 + \frac{1}{\sigma^2}\right)}{2|\mathcal{S}_2^c|} + \frac{1}{2}\log\left(\sigma^2 |\mathcal{S}_1|\right) \right] \\
&= \frac{K}{2} \log\left(1 + \frac{1}{\sigma^2}\right) + \frac{1}{2}\log(\sigma^2 L). \qquad (2.31)
\end{aligned}$$

This holds for every $\sigma^2 \geq 1$, so we set $\sigma^2 = K - 1$ (for $K \geq 2$) to obtain

$$\begin{aligned}
\Delta &\leq \frac{1}{2}\log L + \frac{1}{2}\left(K \log K - (K-1)\log(K-1)\right) \\
&\overset{(a)}{\leq} \frac{1}{2}\left(\log L + \log K + \frac{1}{\ln 2}\right) \\
&= \frac{1}{2}\log(eKL).
\end{aligned}$$

Here, (a) follows from lemma 2.3.1. For $K=1$, we can set $\sigma^2=1$ in (2.31) to obtain

$$\Delta \leq \frac{1}{2}\log(2L) \leq \frac{1}{2}\log(eL).$$

This establishes the first part of the theorem.

For the sum-rate gap, consider

$$\begin{aligned}
\Delta_{\text{sum}} &\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left(F_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2) - F_{\text{in}}(\mathcal{S}_1,\mathcal{S}_2)\right) \\
&\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{1}{2}\log\frac{\left|H_{\mathcal{S}_2^c,\mathcal{S}_1}\Gamma_{\mathcal{S}_1}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right|}{\left|\frac{P}{\sigma^2}H_{\mathcal{S}_2^c,\mathcal{S}_1}H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I\right|} + frac|\mathcal{S}_2^c|2\log\left(1+\frac{1}{\sigma^2}\right)\right] \\
&\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{1}{2}\log\frac{\left|I + \sum_{l=1}^{|\mathcal{S}_1|}\lambda_l b_l b_l^T\right|}{\left|I + \frac{P}{\sigma^2}\sum_{l=1}^{|\mathcal{S}_1|} b_l b_l^T\right|} + frac|\mathcal{S}_2^c|2\log\left(1+\frac{1}{\sigma^2}\right)\right] \\
&\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{\min\{|\mathcal{S}_2^c|,|\mathcal{S}_1|\}}{2}\log(\sigma^2|\mathcal{S}_1|) + \frac{|\mathcal{S}_2^c|}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right] \qquad (2.32)
\end{aligned}$$

if $\sigma^2 \geq 1/|\mathcal{S}_1|$ for each $\mathcal{S}_1 \neq \emptyset$. Maximization of (2.32) over $|\mathcal{S}_1|$ and $|\mathcal{S}_2^c|$ yields, for $\sigma^2 \geq 1$,

$$\Delta_{\text{sum}} \leq \frac{\min\{L,K\}}{2}\log(L\sigma^2) + \frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right). \qquad (2.33)$$

We can then choose $\sigma^2 = 1$ in (2.33) to obtain

$$\Delta_{\text{sum}} \leq \frac{\min\{L,K\}}{2}\log L + \frac{K}{2},$$

completing the proof. □

2.3.3 Proofs of General Bounds on the Capacity Region

Throughout this section, we use the following additional notation. For a function $f:\mathbb{N}\to[0,\infty)$ and a real number $r\neq 0$, we say

$$f(n)\doteq 2^{nr}$$

if

$$r=\lim_{n\to\infty}\frac{\log f(n)}{n}.$$

Proof of Proposition 2.3.1. For analyzing the coding scheme and proving Proposition 2.3.1, we will need the Markov lemma [17, Lemma 12.1] and the following additional elementary result.

Lemma 2.3.2. *Let $\mathcal{C}$, $\mathcal{T}_1$ and $\mathcal{T}_2$ be disjoint and finite index sets and fix a pmf $p(x(\mathcal{C}\cup\mathcal{T}_1\cup\mathcal{T}_2))$. For each $k\in\mathcal{C}\cup\mathcal{T}_1\cup\mathcal{T}_2$, we independently generate X_k^n according to the marginals $\prod_{i=1}^n p_{X_k}(x_{ki})$. Then, as $n\to\infty$,*

$$P\left(X^n(\mathcal{C}\cup\mathcal{T}_1\cup\mathcal{T}_2)\in\mathcal{T}_\varepsilon^{(n)}\right)\doteq 2^{-nI^*(X(\mathcal{C}\cup\mathcal{T}_1\cup\mathcal{T}_2))}, \tag{2.34}$$

and

$$P\left(X^n(\mathcal{C}\cup\mathcal{T}_1)\in\mathcal{T}_\varepsilon^{(n)},X^n(\mathcal{C}\cup\mathcal{T}_2)\in\mathcal{T}_\varepsilon^{(n)}\right)\doteq 2^{-n[I^*(X(\mathcal{C}\cup\mathcal{T}_1))+I^*(X(\mathcal{T}_2))+I(X(\mathcal{C});X(\mathcal{T}_2))]}. \tag{2.35}$$

Our analysis of the coding scheme follows that in [28] but is considerably simpler because of the relative simplicity of our network model.

Without loss of generality, let $m^K=(1,\dots,1)$ be the messages sent. Then the error events are:

$$\mathcal{E}_0 = \Big\{ \Big(X_1^n(w_1), \ldots, X_L^n(w_L), U_1^n(1,s_1), \ldots, U_K^n(1,s_K)\Big) \notin \mathcal{T}_{\varepsilon'}^{(n)} \text{ for all } w^L \text{ and } s^K \Big\}.$$

$$\mathcal{E}_1 = \Big\{ (U_k^n(1,s_k), Y_k^n) \notin \mathcal{T}_{\varepsilon}^{(n)} \text{ for all } s_k, \text{ for some } k \Big\}.$$

$$\mathcal{E}_2 = \Big\{ (U_k^n(m_k,s_k), Y_k^n) \in \mathcal{T}_{\varepsilon}^{(n)} \text{ for some } k, \text{ some } s_k, \text{ and some } m_k \neq 1 \Big\}.$$

By the packing lemma and union of events, $\mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_0^c) \to 0$ as $n \to \infty$ if

$$R_k + \tilde{R}_k < I(U_k; Y_k) \tag{2.36}$$

for all $k \in [K]$. By the Markov lemma and union of events bound ($U_k \to X^L \to Y_k$ form a Markov chain), $\mathsf{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c) \to 0$ as $n \to \infty$.

To analyze the error event $\mathcal{E}_0$, we observe that by the manner in which the codebook is generated, $\mathsf{P}(\mathcal{E}_0)$ remains the same if we index the U_ks only by the indices s_k and drop the m_k. In the following analysis, we do this to simplify notation.

Let

$$\mathcal{A} := \Big\{ (w^L, s^K) : (X_1^n(w_1), \ldots, X_L^n(w_L), U_1^n(s_1), \ldots, U_K^n(s_K)) \in \mathcal{T}_{\varepsilon'}^{(n)} \Big\}.$$

Then, $\mathsf{P}(\mathcal{E}_0) = \mathsf{P}(|\mathcal{A}| = 0)$. We can write

$$|\mathcal{A}| = \sum_{w^L, s^K} Z(w^L, s^K),$$

where

$$Z(w^L, s^K) := \mathbb{1}_{\big\{ \big(X_1^n(w_1), \ldots, X_L^n(w_L), U_1^n(s_1), \ldots, U_K^n(s_K)\big) \in \mathcal{T}_{\varepsilon'}^{(n)} \big\}}.$$

We have

$$\mathsf{E}[Z(w^L,s^K)] = \mathsf{P}\left((X_1^n(w_1),\dots,X_L^n(w_L),U_1^n(s_1),\dots,U_K^n(s_K)) \in \mathcal{T}_{\varepsilon'}^{(n)}\right) =: p_1$$

By (2.34),

$$p_1 \doteq 2^{-nI^*(X^L,U^K)}. \tag{2.37}$$

For $w^L, w'^L \in [2^{nC_1}] \times \cdots \times [2^{nC_l}]$ and $s^K, s'^K \in [2^{n\tilde{R}_1}] \times \cdots \times [2^{n\tilde{R}_K}]$, define

$$\mathcal{S}_1(w^L,w'^L) := \left\{l \in [L] : w_l \neq w'_l\right\},$$
$$\mathcal{S}_2(s^K,s'^K) := \left\{k \in [K] : s_k \neq s'_k\right\}.$$

Then, using (2.35) with index sets

$$\mathcal{C} := \left\{w_l : l \in S_1(w^L,w'^L)\right\} \cup \left\{s_k : k \in S_2(s^K,s'^K)\right\},$$
$$\mathcal{T}_1 := \left\{w_l : l \in S_1^c(w^L,w'^L)\right\} \cup \left\{s_k : k \in S_2^c(s^K,s'^K)\right\}, \text{ and}$$
$$\mathcal{T}_2 := \left\{w'_l : l \in S_1^c(w^L,w'^L)\right\} \cup \left\{s'_k : k \in S_2^c(s^K,s'^K)\right\},$$

we have

$$\begin{aligned}\mathsf{E}[Z(w^L,s^K)Z(w'^L,s'^K)] &\doteq 2^{-n[I^*(X^L,U^K)+I^*(X(\mathcal{S}_1),U(\mathcal{S}_2))+I(X(\mathcal{S}_1^c),U(\mathcal{S}_2^c);X(\mathcal{S}_1),U(\mathcal{S}_2))]}\\ &=: p_2(\mathcal{S}_1,\mathcal{S}_2),\end{aligned} \tag{2.38}$$

where in the definition of p_2, we hide the dependence on w^L, w'^L, s^K, s'^K. We then have

$$\begin{aligned}
&\mathsf{E}\left[|\mathcal{A}|^2\right] \\
&= \sum_{w^L, s^K} \mathsf{E}[Z(w^L, s^K)] + \sum_{\substack{w^L, w'^L, s^K, s'^K: \\ w^L \neq w'^L \text{ or } s^K \neq s'^K}} \mathsf{E}[Z(w^L, s^K) Z(w'^L, s'^K)] \\
&= p_1 \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k\right)} \\
&\quad + \sum_{\substack{\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subseteq [K], \\ \mathcal{S}_1 \neq \emptyset \text{ or } \mathcal{S}_2 \neq \emptyset}} \Bigg(p_2(\mathcal{S}_1, \mathcal{S}_2) \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k\right)} \cdot \left(2^{n \sum_{l \in \mathcal{S}_1} C_l} - 1\right) \times \\
&\qquad\qquad \left(2^{n \sum_{k \in \mathcal{S}_2} \tilde{R}_k} - 1\right) \Bigg) \\
&\leq p_1 \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k\right)} \\
&\quad + \sum_{\substack{\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subseteq [K], \\ \mathcal{S}_1 \neq \emptyset \text{ or } \mathcal{S}_2 \neq \emptyset}} p_2(\mathcal{S}_1, \mathcal{S}_2) \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k + \sum_{l \in \mathcal{S}_1} C_l + \sum_{k \in \mathcal{S}_2} \tilde{R}_k\right)}.
\end{aligned}$$

Noting that $p_2([L],[K]) = p_1^2$, we then have

$$\begin{aligned}
\mathrm{Var}(|\mathcal{A}|) \leq\ & p_1 \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k\right)} \\
& + \sum_{\substack{\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subseteq [K], \\ \mathcal{S}_1 \neq \emptyset \text{ or } \mathcal{S}_2 \neq \emptyset, \\ \mathcal{S}_1 \neq [L] \text{ or } \mathcal{S}_2 \neq [K]}} p_2(\mathcal{S}_1, \mathcal{S}_2) \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k + \sum_{l \in \mathcal{S}_1} C_l + \sum_{k \in \mathcal{S}_2} \tilde{R}_k\right)}.
\end{aligned} \tag{2.39}$$

We also have

$$\mathsf{E}[|\mathcal{A}|] = p_1 \cdot 2^{n\left(\sum_{l=1}^L C_l + \sum_{k=1}^K \tilde{R}_k\right)}. \tag{2.40}$$

Using (2.37), (2.38), (2.39), and (2.40) and manipulating exponents, we finally have, for some $\delta(\varepsilon')$ that goes to zero as $\varepsilon' \to 0$,

$$\frac{\mathrm{Var}(|\mathcal{A}|)}{\mathsf{E}[|\mathcal{A}|]^2} \leq \sum_{\substack{\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subseteq [K], \\ \mathcal{S}_1 \neq [L] \text{ or } \mathcal{S}_2 \neq [K]}} 2^{-n\left(\sum_{l \in \mathcal{S}_1^c} C_l + \sum_{k \in \mathcal{S}_2^c} \tilde{R}_k - I(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) - \delta(\varepsilon')\right)}.$$

Thus, using the inequality $\mathsf{P}(|\mathcal{A}|=0) \leq \mathrm{Var}(|\mathcal{A}|)/\,\mathsf{E}[|\mathcal{A}|^2]$, we conclude that $\mathsf{P}(\mathcal{E}_0) \to 0$ as $n \to \infty$ if

$$\sum_{l\in\mathcal{S}_1^c} C_l + \sum_{k\in\mathcal{S}_2^c} \tilde{R}_k > I^*(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \tag{2.41}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$.

Combining this with (2.36) to eliminate the auxiliary rates, the rates R_k satisfy, for every $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$,

$$\begin{aligned}
&\sum_{k\in\mathcal{S}_2^c} R_k \\
&< \sum_{l\in\mathcal{S}_1^c} C_l + \sum_{k\in\mathcal{S}_2^c} I(U_k;Y_k) - I^*(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \\
&= \sum_{l\in\mathcal{S}_1^c} C_l + \sum_{k\in\mathcal{S}_2^c} \big(I(U_k;X^L,Y_k) - I(U_k;X^L|Y_k)\big) - I^*(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \\
&\overset{(a)}{=} \sum_{l\in\mathcal{S}_1^c} C_l - \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L|Y_k) + \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L) - I^*(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \\
&= \sum_{l\in\mathcal{S}_1^c} C_l - \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L|Y_k) - \sum_{k\in\mathcal{S}_2^c} H(U_k|X^L) - \sum_{l\in\mathcal{S}_1^c} H(X_l) + H(X(\mathcal{S}_1^c), U(\mathcal{S}_2^c)) \\
&= \sum_{l\in\mathcal{S}_1^c} C_l - \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L|Y_k) - \sum_{k\in\mathcal{S}_2^c} H(U_k|X^L) - I^*(X(\mathcal{S}_1^c)) + H(U(\mathcal{S}_2^c)|X(\mathcal{S}_1^c)) \\
&= \sum_{l\in\mathcal{S}_1^c} C_l - \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L|Y_k) - H(U(\mathcal{S}_2^c)|X^L) - I^*(U(\mathcal{S}_2^c)|X^L) \\
&\qquad\qquad - I^*(X(\mathcal{S}_1^c)) + H(U(\mathcal{S}_2^c)|X(\mathcal{S}_1^c)) \\
&= \sum_{l\in\mathcal{S}_1^c} C_l - \sum_{k\in\mathcal{S}_2^c} I(U_k;X^L|Y_k) + I(X(\mathcal{S}_1);U(\mathcal{S}_2^c)|X(\mathcal{S}_1^c)) - I^*(X(\mathcal{S}_1^c)) - I^*(U(\mathcal{S}_2^c)|X^L). \quad (2.42)
\end{aligned}$$

Here, (a) follows from the fact that $U_k \to X^L \to Y_k$ form a Markov chain. In addition, if $\mathcal{S}_2 = [K]$

in (2.41), we obtain the additional conditions

$$\sum_{l\in\mathcal{S}_1^c} C_l > I^*(X(\mathcal{S}_1^c)) \tag{2.43}$$

for every $\mathcal{S}_1 \subsetneq [L]$. This completes the proof. □

Remark 2.3.2. The constraints (2.43) can be shown to be inactive using techniques similar to [28, Appendix E].

Proof of Proposition 2.3.2. We use $X_i(\mathcal{S}_1), X^n(\mathcal{S}_1), Y_i(\mathcal{S}_2), Y^n(\mathcal{S}_2)$ to convey similar meanings as in the proof of Proposition 2.2.2 in Section 2.2.3. For $k \in [K]$, let M_k denote the message intended for receiver k and let W_l denote the index communicated by the central processor to relay l. Then, for every $\mathcal{S}_1 \subseteq [L]$, $X^n(\mathcal{S}_1)$ is a function of $W(\mathcal{S}_1)$, which is itself a function of M^K. For every $\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subsetneq [K]$, we must have, by Fano's inequality,

$$H(M(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), Y^n(\mathcal{S}_2^c)) \leq H(M(\mathcal{S}_2^c)\,|\,Y^n(\mathcal{S}_2^c)) \leq n\varepsilon_n,$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Therefore, since $H(M(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2)) = n\sum_{k\in\mathcal{S}_2^c} R_k$, we have

$$\begin{aligned}
& n\sum_{k\in\mathcal{S}_2^c} R_k \\
&\leq I(M(\mathcal{S}_2^c); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2)) + n\varepsilon_n \\
&\leq I(M(\mathcal{S}_2^c), W(\mathcal{S}_1^c); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2)) + n\varepsilon_n \\
&= I(M(\mathcal{S}_2^c); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), W(\mathcal{S}_1^c)) + I(W(\mathcal{S}_1^c); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2)) + n\varepsilon_n \\
&\leq I(M(\mathcal{S}_2^c); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), W(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n \\
&\overset{(a)}{=} I(M(\mathcal{S}_2^c), W(\mathcal{S}_1); Y^n(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), W(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n \\
&= \sum_{i=1}^n I(M(\mathcal{S}_2^c), W(\mathcal{S}_1); Y_i(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), W(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^n I(M(\mathcal{S}_2^c), W(\mathcal{S}_1); Y_i(\mathcal{S}_2^c)\,|\,M(\mathcal{S}_2), W(\mathcal{S}_1^c), Y^{i-1}(\mathcal{S}_2^c), X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n
\end{aligned}$$

$$\leq \sum_{i=1}^{n} I(M^K, W^L, Y^{i-1}(\mathcal{S}_2^c); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n$$
$$\stackrel{(c)}{=} \sum_{i=1}^{n} I(M^K, W^L, Y^{i-1}(\mathcal{S}_2^c), X_i(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n$$
$$\stackrel{(d)}{=} \sum_{i=1}^{n} I(X_i(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + H(W(\mathcal{S}_1^c)) + n\varepsilon_n$$
$$\stackrel{(e)}{\leq} \sum_{i=1}^{n} I(X_i(\mathcal{S}_1); Y_i(\mathcal{S}_2^c) \,|\, X_i(\mathcal{S}_1^c)) + n \sum_{l \in \mathcal{S}_1^c} C_l + n\varepsilon_n.$$

Here, (a) follows since conditioned on $M(\mathcal{S}_2)$, $W(\mathcal{S}_1)$ is a function of $M(\mathcal{S}_2^c)$; (b) follows since $X_i(\mathcal{S}_1^c)$ is a function of $W(\mathcal{S}_1^c)$; (c) follows since $X_i(\mathcal{S}_1)$ is a function of W^L; (d) follows since $(M^K, W^L, Y^{i-1}(\mathcal{S}_2^c)) \to (X_i(\mathcal{S}_1), X_i(\mathcal{S}_1^c)) \to (Y_i(\mathcal{S}_2^c))$ form a Markov chain (by the memorylessness of the second hop), and (e) follows since $W(\mathcal{S}_1^c)$ is supported on a set of size $\prod_{l \in \mathcal{S}_1^c} 2^{nC_l}$. Defining a random variable $Q \sim \text{Unif}([n])$ independent of all other random variables, writing $X(\mathcal{S}_1) := X_Q(\mathcal{S}_1)$ and $Y(\mathcal{S}_2) := Y_Q(\mathcal{S}_2)$, noting that $Q \to X^L \to Y^K$ form a Markov chain, and letting n tend to infinity leads to

$$\sum_{k \in \mathcal{S}_1} R_k \leq I(X(\mathcal{S}_1); Y(\mathcal{S}_2^c) \,|\, X(\mathcal{S}_1^c)) + \sum_{l \in \mathcal{S}_1^c} C_l \tag{2.44}$$

for all $\mathcal{S}_1 \subseteq [L], \mathcal{S}_2 \subsetneq [K]$ for some pmf $p(x^L)$, and thus completes the proof. □

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Shouvik Ganguly and Young-Han Kim, "On the Capacity of Cloud Radio Access Networks", *Proceedings of the IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017, as well as parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, "On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*. The dissertation author was the primary researcher and author of this material.

Chapter 3

Capacity Scaling of C-RANs

The approximate capacity regions of the uplink and downlink C-RAN are compared with the capacity regions for networks with no capacity limit on the fronthaul. Although it takes infinite fronthaul link capacities to achieve these “fronthaul-unlimited” capacity regions exactly, these capacity regions can be approached *approximately* with finite-capacity fronthaul. The total fronthaul link capacities required to approach the fronthaul-unlimited sum-rates (for uplink and downlink) are characterized. Based on these results, the capacity scaling law in the large network size limit is established under certain uplink and downlink network models, both theoretically and via simulations.

3.1 Introduction

In this chapter, we quantify the minimum fronthaul capacity required to achieve the fronthaul-unlimited uplink and downlink capacity regions approximately. We then use these results to characterize the scaling behavior of the uplink and downlink C-RAN sum-capacities for large network size under various channel models and demonstrate that the C-RAN sum-rates exhibit similar large-user asymptotics as the fronthaul-unlimited sum-capacities for a range of channel models.

3.2 Fronthaul-unlimited Networks

Large-network asymptotics and other types of scaling behavior for multi-terminal systems have been explored ever since the introduction of MIMO. Telatar [56] examined single-user MIMO systems with multiple transmit and receive antennas and quantified the gains over single-antenna systems in terms of capacities and error exponents. It was also shown that the ergodic capacity of such systems scales linearly with the number of antennas in the large-antenna limit under *rich scattering* (see Section 3.3.2). Under a similar channel model, Tse et al. [57] studied the tradeoffs between throughput (multiplexing gain) and error performance (diversity gain) for multiple access networks. Verdú [58] evaluated transmit energy thresholds for multi-antenna systems in the wideband (low spectral efficiency) regime under Gaussian as well as Laplacian noise distributions. For a comprehensive review of literature on the asymptotic behavior of MIMO systems, we refer the reader to [37, 56, 58] and the references therein.

3.3 Uplink C-RANs

3.3.1 Comparisons with Fronthaul-Unlimited Uplink

In this section, we examine the effect of the capacities C_l of the fronthaul links (in particular, their sum $C_\Sigma := C_1 + \cdots + C_L$) on the capacity region of the uplink C-RAN. Recall from Remark 2.2.1 that as the fronthaul link capacities approach infinity, the uplink C-RAN capacity region becomes the same as the fronthaul-unlimited uplink capacity region in the limit. However, as shown by the following example, this limit is in general unattainable when the link capacities are finite.

Example 3.3.1. Consider the single-sender, 2-relay Gaussian uplink C-RAN with first hop given by

$$Y_1 = gX + Z_1,$$

$$Y_2 = gX + Z_2,$$

where $g \in \mathbb{R} \setminus \{0\}$ and Z_1, Z_2 are i.i.d. $\mathrm{N}(0,1)$ noise components. Let us denote the fronthaul link capacities of this network by C_1 and C_2, and let there be an average power constraint $P > 0$ on the sender. Then the first hop has conditionally i.i.d. outputs Y_1, Y_2 given X. If $C_2 = \infty$, this network is equivalent to the relay channel model studied in the Gaussian version of Cover's problem [12] and by the results of Wu, Barnes, and Özgür [63], the capacity for any finite C_1 is strictly less than the capacity for $C_1 = \infty$. Thus, even for this simple network, the fronthaul-unlimited uplink capacity $R^{\infty}_{\text{sum}} := (1/2)\log(1+2g^2P)$ is unattainable unless both the fronthaul link capacities are infinite. On the positive side, it is possible to approximately achieve the fronthaul-unlimited

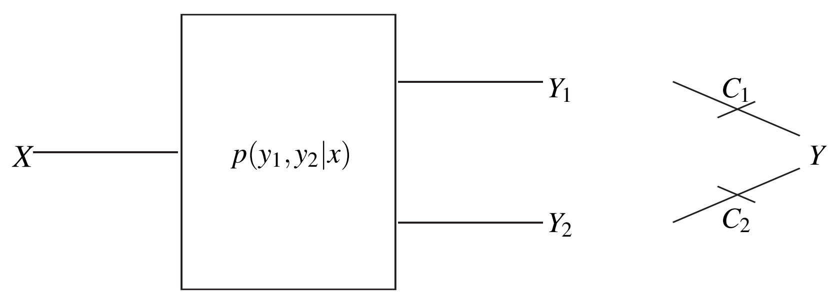


Figure 3.1. A single sender 2-relay uplink C-RAN.

uplink sum-rate for finite fronthaul capacities, provided we spend a sufficient amount of extra capacity on the fronthaul. Suppose that we have a certain amount of total capacity C_Σ to spend on the fronthaul links, which we are free to allocate in any way among the two links. The cutset bound implies that we cannot hope to achieve the capacity $(1/2)\log(1+2g^2P)$ unless C_Σ is at least equal to this amount. However, if we set

$$C_1 = C_2 = \frac{1}{2} + \frac{1}{4}\log(1+2g^2P).$$

and thus spend the fronthaul sum-capacity of

$$C_\Sigma = \frac{1}{2}\log(1+2g^2P)+1,$$

then it can be shown, by taking $\sigma^2 = 1$ in (2.7), that the uplink C-RAN sum-rate is

$$\begin{aligned}
&\min\left\{\frac{1}{2}\log(1+g^2P), \frac{1}{2}\log\left(1+\frac{g^2P}{2}\right)+C_1-\frac{1}{2},\right.\\
&\qquad\qquad\left.\frac{1}{2}\log\left(1+\frac{g^2P}{2}\right)+C_2-\frac{1}{2}, C_1+C_2-1\right\}\\
&=\min\left\{\frac{1}{2}\log(1+g^2P), \frac{1}{2}\log\left(1+\frac{g^2P}{2}\right)+\frac{1}{4}\log\left(1+2g^2P\right), \frac{1}{2}\log(1+2g^2P)\right\}\\
&=\frac{1}{2}\log(1+g^2P)\\
&\overset{(b)}{\geq}\frac{1}{2}\log(1+2g^2P)-\frac{1}{2},
\end{aligned} \tag{3.1}$$

where (b) follows since $(1+2g^2P)\leq 2(1+g^2P)$. Thus, using a total fronthaul link capacity only 1 bit higher than the fronthaul-unlimited uplink capacity, we can achieve the fronthaul-unlimited uplink capacity within half a bit, irrespective of P and g. Thus we can achieve the fronthaul-unlimited uplink sum-capacity within a finite additive gap using a total fronthaul link capacity which is also finitely larger than the fronthaul-unlimited sum-capacity in the additive sense. This statement is formalized and generalized in Corollary 3.3.1 to Theorem 3.3.1.

The result (3.1) holds for every P and therefore, with this fronthaul allocation strategy, we can achieve $R_{\text{sum}}^{\text{NCF}}/R_{\text{sum}}^{\infty}$ approaching 1 as $P\to\infty$, with a total fronthaul link capacity C_Σ satisfying $C_\Sigma/R_{\text{sum}}^{\infty}\to 1$ as $P\to\infty$. In fact, we can go one step further and show, letting P go to infinity in (3.1), that at high SNR, using a C_Σ whose ratio to R_{sum}^{∞} is 1 within $O(1/\log P)$, network compress–forward can achieve a sum-rate whose ratio to R_{sum}^{∞} is also 1 within $O(1/\log P)$. Thus, in a multiplicative sense as well, only a slightly larger fronthaul capacity is sufficient to approximate the fronthaul-unlimited capacity for this network. This statement is formalized and explored in Corollary 3.3.2.

We first quantify the fronthaul requirement for network compress–forward to approximate the fronthaul-unlimited uplink sum-capacity in Theorem 3.3.1, from which the additive and multiplicative gap results follow as corollaries.

Theorem 3.3.1. *If*

$$C_\Sigma \geq \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right| + \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right) =: C^*(\sigma^2)$$

for some $\sigma^2 > 0$, *then there exist fronthaul link capacities* $C_1, C_2, \ldots, C_L \geq 0$ *with* $\sum_{l\in[L]} C_l = C_\Sigma$ *at which network compress–forward can achieve a sum-rate*

$$R_{\text{sum}}^{NCF}(\sigma^2) = \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right|.$$

Conversely, to achieve a sum-rate of $(1/2)\log|I + PGG^T|$, *we must have a total fronthaul capacity*

$$C_\Sigma \geq \frac{1}{2}\log|I + PGG^T|.$$

Proof. The achievable sum-rate can be written as

$$\begin{aligned} R_{\text{sum}}^{\text{NCF}}(\sigma^2) &= \min_{\mathcal{S}_2\subseteq[L]}\left(\frac{1}{2}\log\left|\frac{P}{\sigma^2+1}G_{\mathcal{S}_2^c,[K]}G_{\mathcal{S}_2^c,[K]}^T + I\right| + \sum_{l\in\mathcal{S}_2} C_l - \frac{|\mathcal{S}_2|}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right) \\ &= \min_{\mathcal{S}_2\subseteq[L]}\Big(\phi(\mathcal{S}_2^c) + \psi(\mathcal{S}_2)\Big), \end{aligned} \tag{3.2}$$

where

$$\phi(\mathcal{S}_2^c) := \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}G_{\mathcal{S}_2^c,[K]}G_{\mathcal{S}_2^c,[K]}^T + I\right|$$

and

$$\psi(\mathcal{S}_2) := \sum_{l\in\mathcal{S}_2}\left(C_l - \frac{1}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right).$$

If $C_l \geq (1/2)\log(1+1/\sigma^2)$ for all $l \in [L]$, the set functions ϕ and ψ are zero on the null set, monotonically increasing (with respect to the partial ordering defined by set inclusion), and are *submodular*, i.e., we have $\phi(\emptyset) = 0$, $\phi(\mathcal{S}) \leq \phi(\mathcal{T})$ if $\mathcal{S} \subseteq \mathcal{T}$, and $\phi(\mathcal{S} \cup \mathcal{T}) + \phi(\mathcal{S} \cap \mathcal{T}) \leq \phi(\mathcal{S}) + \phi(\mathcal{T})$. The sets

$$\mathscr{P}(\phi) := \left\{(x_1,\ldots,x_L) \subseteq \mathbb{R}_+^L : \sum_{l\in\mathcal{S}} x_l \leq \phi(\mathcal{S}), \mathcal{S} \subseteq [L]\right\}$$

and $\mathscr{P}(\psi)$, defined in a similar manner, are referred to as *polymatroids* [13], which generalize two-dimensional pentagonal regions to L dimensions. The following celebrated result can rewrite (3.2) in an alternative form.

Lemma 3.3.1 (Edmonds's polymatroid intersection theorem [13])**.** *If $\mathscr{P}(\phi)$ and $\mathscr{P}(\psi)$ are two polymatroids, then*

$$\max\left\{\sum_{l\in[L]} x_l : (x_1,\ldots,x_L) \in \mathscr{P}(\phi) \cap \mathscr{P}(\psi)\right\} = \min_{\mathcal{S}\subseteq[L]} \left(\phi(\mathcal{S}) + \psi(\mathcal{S}^c)\right).$$

Using Lemma 3.3.1, we can rewrite (3.2) as

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) = \max_{y^L}\left\{\sum_{l\in[L]} y_l : y_l \leq \psi(\{l\}), l \in [L], \sum_{l\in\mathcal{S}_2} y_l \leq \phi(\mathcal{S}_2), \mathcal{S}_2 \subseteq [L]\right\}.$$

Now, let us fix

$$C_\Sigma \geq \phi([L]) + \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right) = \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right| + \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right) \tag{3.3}$$

such that $C_1,\ldots,C_L$ are constrained to satisfy $C_1+\ldots+C_L = C_\Sigma$. Choose a point $\mathbf{y}^* \equiv (y_1^*,\ldots,y_L^*) \in \mathscr{P}(\phi)$ such that $y_1^* + \ldots + y_L^* = \phi([L])$. Such a point always exists since

$\mathscr{P}(\phi)$ is a polymatroid. The point $\tilde{\mathbf{y}} \equiv (\tilde{y}_1, \ldots, \tilde{y}_L)$ defined by

$$\tilde{y}_l = \frac{C_\Sigma - \frac{L}{2}\log\left(1 + \frac{1}{\sigma^2}\right)}{\phi([L])} y_l^*, \quad l \in [L],$$

satisfies $\tilde{y}_1 + \cdots + \tilde{y}_L = C_\Sigma - (L/2)\log(1+1/\sigma^2)$. Therefore, choosing $C_l = \tilde{y}_l + (1/2)\log(1 + 1/\sigma^2)$ for each l, $\mathscr{P}(\phi)$ becomes the cuboid

$$\{(y_1, \ldots, y_L) : y_l \leq \tilde{y}_l, l \in [L]\}$$

with corner point $\tilde{\mathbf{y}}$. Moreover, this cuboid includes the point $\mathbf{y}^*$, since $\tilde{y}_l \geq y_l^*$ for each l by (3.3). Thus, the point $\mathbf{y}^*$ lies in the intersection $\mathscr{P}(\phi) \cap \mathscr{P}(\psi)$ and therefore, network compress–forward, with this choice of $C_1, \ldots, C_L$, achieves the sum-rate

$$y_1^* + \cdots + y_L^* = \phi([L]) = \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right|,$$

establishing the result. The converse follows immediately from the cutset bound. □

Remark 3.3.1. Given σ^2, P, and G, coming up with a specific allocation $(C_1, \ldots, C_L)$ satisfying the sum fronthaul constraint is equivalent to finding a point $\mathbf{y}^*$, as seen from the proof. Such a point can be found, moreover, by solving a linear feasibility problem

$$\begin{aligned} \text{find} \quad & (y_1^*, \ldots, y_L^*) \\ \text{subject to} \quad & \sum_{l \in \mathcal{S}} y_l^* \leq \phi(\mathcal{S}), \quad \mathcal{S} \subsetneq [L], \\ & \sum_{l \in [L]} y_l^* = \phi([L]). \end{aligned}$$

Thus, the fronthaul allocation problem is equivalent to checking the feasibility of a linear program with $2^L - 2$ inequalities and one equality.

Remark 3.3.2. As an immediate consequence of the polymatroid representation, the best sum-rate achievable for a given total fronthaul capacity $C_\Sigma > 0$ can be expressed as

$$R_{\text{sum}}^{\max}(C_\Sigma) = \sup_{\sigma^2>0} \min\left\{C_\Sigma - \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right), \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right|\right\}.$$

The first term in the minimum increases monotonically from 0 to ∞ as σ^2 increases from 0 to ∞, while for $G \neq 0$, the second term decreases monotonically from ∞ to 0. Therefore, there is a unique $\sigma_*^2(C_\Sigma)$ at which the supremum is attained and the two terms in the minimum are equal for $\sigma^2 = \sigma_*^2(C_\Sigma)$. This also shows that

$$\lim_{C_\Sigma\to\infty} \sigma_*^2(C_\Sigma) = 0$$

and hence, that

$$\lim_{C_\Sigma\to\infty} R_{\text{sum}}^{\max}(C_\Sigma) = \frac{1}{2}\log|PGG^T + I| = R_{\text{sum}}^{\infty}.$$

Thus, our coding scheme is asymptotically optimal in the limit of large fronthaul sum-capacity.

Theorem 3.3.1 leads to a formalization of the achievable additive and multiplicative gaps from the fronthaul-unlimited uplink sum-capacity, that were briefly explored in Example 3.3.1.

Corollary 3.3.1 (Additive gap from fronthaul-unlimited uplink sum-capacity)**.** *Denote by R_{sum}^{∞} the fronthaul-unlimited uplink sum-capacity, which is given by $(1/2)\log|I+PGG^T|$. Then, for every P and G for some $\sigma^2 > 0$, if $C_\Sigma = R_{\text{sum}}^{\infty} + \Delta_1(\sigma^2)$, then there exist $C_1,\dots,C_L$ with $\sum_{l\in[L]} C_l = C_\Sigma$, at which $R_{\text{sum}}^{NCF}(\sigma^2) \geq R_{\text{sum}}^{\infty} - \Delta_2(\sigma^2)$, where*

$$\Delta_1(\sigma^2) = \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right)$$

and

$$\Delta_2(\sigma^2) = \frac{\min\{K,L\}}{2}\log\left(1+\sigma^2\right).$$

Proof. We have, from Theorem 3.3.1, that if

$$C_{\Sigma} = \frac{1}{2}\log\left|\frac{P}{\sigma^2+1}GG^T + I\right| + \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right),$$

then

$$C_{\Sigma} - R_{\text{sum}}^{\infty} = \frac{1}{2}\log\frac{\left|\frac{P}{\sigma^2+1}GG^T+I\right|}{|PGG^T+I|} + \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right) \leq \frac{L}{2}\log\left(1+\frac{1}{\sigma^2}\right), \tag{3.5}$$

and

$$\begin{aligned} R_{\text{sum}}^{\infty} - R_{\text{sum}}^{\text{NCF}}(\sigma^2) &= \frac{1}{2}\log\frac{|PGG^T+I|}{\left|\frac{P}{\sigma^2+1}GG^T+I\right|} \\ &\overset{(a)}{\leq} \frac{\text{rank}(G)}{2}\log(1+\sigma^2) \\ &\leq \frac{\min\{K,L\}}{2}\log(1+\sigma^2), \end{aligned} \tag{3.6}$$

where (a) follows from the inequality $(1+\alpha)/(1+\alpha/x) \leq x$ for $x > 1, \alpha > 0$. Equations (3.5) and (3.6) establish the result. □

As a concrete illustration of the gaps established by Corollary 3.3.1, taking $\sigma^2 = L$ in (3.5) and (3.6) yields

$$\begin{aligned} \Delta_1(\sigma^2) &= \frac{L}{2}\log\left(1+\frac{1}{L}\right) \\ &\overset{(a)}{\leq} \frac{L}{2}\cdot\frac{\log e}{L} \\ &= \frac{\log e}{2} \end{aligned}$$

and

$$\Delta_2(\sigma^2) = \frac{\min\{K,L\}}{2}\log(1+L).$$

Here, (a) follows since $\log(1+x) \leq x \log e$, $x > 0$. Similarly, setting $\sigma^2 = 1/L$ in (3.5) and (3.6) yields

$$\Delta_1(\sigma^2) = \frac{L}{2} \log(1+L)$$

and

$$\Delta_2(\sigma^2) = \frac{\min\{K,L\}}{2} \log\left(1 + \frac{1}{L}\right) \leq \frac{\min\{K,L\}}{2} \cdot \frac{\log e}{L} \leq \frac{\log e}{2}.$$

Various choices of σ^2, as well as the corresponding tradeoffs between Δ_1 and Δ_2, are summarized in Table 3.1. As noted before, Corollary 3.3.1, being a channel- and SNR-independent result,

Table 3.1. Additive gap from fronthaul-unlimited uplink sum-capacity.

σ^2	$\Delta_1(\sigma^2)$	$\Delta_2(\sigma^2)$
L	$\frac{\log e}{2}$	$\frac{\min\{K,L\}}{2} \log(1+L)$
1	$\frac{L}{2}$	$\frac{\min\{K,L\}}{2}$
$\frac{1}{L}$	$\frac{L}{2} \log(1+L)$	$\frac{\log e}{2}$

implies that both $R_{\text{sum}}^{\text{NCF}}/R_{\text{sum}}^{\infty}$ and $C^*/R_{\text{sum}}^{\infty}$ approach 1 at high SNR. The next result is a further refinement of this statement.

Corollary 3.3.2 (Multiplicative gap from fronthaul-unlimited uplink sum-capacity at high SNR)**.** *For a fixed channel gain matrix G, let $P \to \infty$ and let σ^2 be chosen as $\sigma^2 = \sigma^2(P)$ such that*

$$\lim_{P\to\infty} P\sigma^2(P) = \infty$$

and

$$\lim_{P\to\infty} \sigma^2(P)/P = 0.$$

Then,

$$1-\frac{R_{\text{sum}}^{NCF}}{R_{\text{sum}}^{\infty}}\sim\frac{\log\left(1+\sigma^2(P)\right)}{\log P}$$

and

$$\frac{C^*}{R_{\text{sum}}^{\infty}}-1\sim\begin{cases}\dfrac{\frac{L\log e}{\sigma^2(P)}-\text{rank}(G)\log(\sigma^2(P))}{\text{rank}(G)\log P}, & \sigma^2(P)\xrightarrow{P\to\infty}\infty,\\ \dfrac{L\log(1/\sigma^2(P))}{\text{rank}(G)\log P}, & \sigma^2(P)\xrightarrow{P\to\infty}0,\\ \dfrac{\text{rank}(G)\log\left(\frac{1}{1+\sigma^2}\right)+L\log\left(1+\frac{1}{\sigma^2}\right)}{\text{rank}(G)\log P}, & \sigma^2>0 \textit{ is fixed,}\end{cases}$$

where $R_{\text{sum}}^{NCF}, R_{\text{sum}}^{\infty}$, and C^* depend on P and G (as well as $\sigma^2(\cdot)$ for C^* and R_{sum}^{NCF}).

Proof. Let $\beta_1,\ldots,\beta_{\text{rank}(G)}$ be the non-zero eigenvalues of GG^T. Then, we have

$$\begin{aligned}1-\frac{R_{\text{sum}}^{\text{NCF}}(P)}{R_{\text{sum}}^{\infty}(P)}&=1-\frac{\sum_{l=1}^{\text{rank}(G)}\log\left(1+\frac{P\beta_l}{1+\sigma^2(P)}\right)}{\sum_{l=1}^{\text{rank}(G)}\log(1+P\beta_l)}\\&=\frac{\sum_{l=1}^{\text{rank}(G)}\log\left(\frac{1+P\beta_l}{1+\frac{P\beta_l}{1+\sigma^2(P)}}\right)}{\sum_{l=1}^{\text{rank}(G)}\log(1+P\beta_l)}\\&=\frac{\sum_{l=1}^{\text{rank}(G)}\log\left(1+\frac{P\sigma^2(P)\beta_l}{1+\frac{P\beta_l}{1+\sigma^2(P)}}\right)}{\sum_{l=1}^{\text{rank}(G)}\log(1+P\beta_l)}\\&\sim\frac{\text{rank}(G)\log(1+\sigma^2(P))}{\text{rank}(G)\log P}\\&=\frac{\log(1+\sigma^2(P))}{\log P},\end{aligned}\tag{3.7}$$

and

$$\frac{C^*(P)}{R_{\text{sum}}^{\infty}(P)}-1=\frac{\sum_{l=1}^{\text{rank}(G)}\log\left(\frac{1+\frac{P\beta_l}{1+\sigma^2(P)}}{1+P\beta_l}\right)+L\log\left(1+\frac{1}{\sigma^2(P)}\right)}{\sum_{l=1}^{\text{rank}(G)}\log(1+P\beta_l)}.\tag{3.8}$$

If $\sigma^2(P) \to \infty$ as $P \to \infty$, (3.8) leads to

$$\frac{C^*(P)}{R_{\text{sum}}^{\infty}(P)} - 1 \sim \frac{\text{rank}(G)\log\left(1/\sigma^2(P)\right) + \frac{L\log e}{\sigma^2(P)}}{\text{rank}(G)\log P} \sim \frac{\log\left(1/\sigma^2(P)\right)}{\log P}, \tag{3.9}$$

and if $\sigma^2(P) \to 0$ as $P \to \infty$, (3.8) leads to

$$\begin{aligned}
\frac{C^*(P)}{R_{\text{sum}}^{\infty}(P)} - 1 &= \frac{\sum_{l=1}^{\text{rank}(G)} \log\left(1 - \frac{P\beta_l\sigma^2(P)}{(1+\sigma^2(P))(1+P\beta_l)}\right) + L\log\left(1 + \frac{1}{\sigma^2(P)}\right)}{\sum_{l=1}^{\text{rank}(G)} \log(1+P\beta_l)} \\
&\sim \frac{\text{rank}(G)\sigma^2(P) + L\log(1/\sigma^2(P))}{\text{rank}(G)\log P} \\
&\sim \frac{L\log(1/\sigma^2(P))}{\text{rank}(G)\log P}.
\end{aligned} \tag{3.10}$$

Similarly, if $\sigma^2 > 0$ is fixed, (3.8) leads to

$$\frac{C^*(P)}{R_{\text{sum}}^{\infty}(P)} - 1 \sim \frac{\text{rank}(G)\log\left(\frac{1}{1+\sigma^2}\right) + L\log\left(1 + \frac{1}{\sigma^2}\right)}{\text{rank}(G)\log P}. \tag{3.11}$$

□

For various choices of $\sigma^2(P)$, (3.7), (3.9), (3.10), and (3.11) enable us to make several statements about the behaviors of the ratios $R_{\text{sum}}^{\text{NCF}}/R_{\text{sum}}^{\infty}$ and $C^*/R_{\text{sum}}^{\infty}$ at high SNR. These are summarized in Table 3.2. As another result that demonstrates the asymptotically optimal fronthaul link capacity, we examine how $R_{\text{sum}}^{\text{NCF}}$, R_{sum}^{∞}, and C^* scale with network size for specific network models, in the next section.

3.3.2 Capacity Scaling

In this section, as opposed to keeping the network size fixed and varying the SNR and the channel coefficients, we let the network size grow and examine how the sum-rates and the fronthaul capacity requirement behave under certain network models. In Section 3.3.2, we consider a channel model, referred to as the *rich scattering model*, where the entries of the

Table 3.2. Multiplicative gap from fronthaul-unlimited uplink sum-capacity.

$\sigma^2(P)$	$\frac{C^*(P)}{R^\infty_{\text{sum}}(P)} - 1$	$1 - \frac{R^{\text{NCF}}_{\text{sum}}(P)}{R^\infty_{\text{sum}}(P)}$
1	$O\left(\frac{1}{\log P}\right)$	$O\left(\frac{1}{\log P}\right)$
$\log P$	$O\left(\frac{\log\log P}{\log P}\right)$	$O\left(\frac{\log\log P}{\log P}\right)$
$(\log P)^{-\varepsilon}$, $\varepsilon \in (0,1)$	$O\left(\frac{\log\log P}{\log P}\right)$	$O\left(\frac{1}{(\log P)^{1+\varepsilon}}\right)$

channel gain matrix G are i.i.d. $\mathrm{N}(0,1)$ random variables. In Section 3.3.2, we study a *stochastic geometry* model through simulations, where users and relays are physically distributed over a two-dimensional area at random, and the channel coefficient between a particular user–relay pair is determined by the Euclidean distance between the two.

In contrast to the current treatment, large network size asymptotics for achievable symmetric rates was considered in [47] for $L = K$ and equal fronthaul link capacities, under various localized interference models such as the Wyner model [65] and the soft-handoff model [55]. Specifically, under these models, the limit of the symmetric achievable rate was computed as the network size grows to infinity. The high- and low-SNR behaviors of this limit were then studied. Fading was incorporated into the same localized interference model and similar studies were made on the limit of the ergodic capacity.

Rich scattering

We consider a *rich scattering* network model with slow fading, where the entries of the channel gain matrix G are i.i.d. random variables with variance 1 and are assumed fixed for the duration of transmission. Moreover, the average power constraint P is kept fixed. We recall the following.

Lemma 3.3.2 (Telatar [56], Silverstein [50])**.** *Let W be an $m \times n$ random matrix with i.i.d. entries*

W_{ij}, each of which has unit variance. Then, as $n \to \infty$ such that $n/m \to \rho \in [1, \infty)$, the limiting density of the empirical distribution of the eigenvalues of WW^T/m is given, almost surely, by

$$f_\Lambda(\lambda) = \frac{\sqrt{(\lambda - \alpha(\rho))(\beta(\rho) - \lambda)}}{2\pi\lambda} \mathbf{1}_{[\alpha(\rho), \beta(\rho)]},$$

where $\alpha(\rho) := (\sqrt{\rho} - 1)^2$ and $\beta(\rho) := (\sqrt{\rho} + 1)^2$. On the other hand, if $n/m \to \infty$, all the eigenvalues of WW^T/n approach 1 a.s.

Using Lemma 3.3.2, we can establish the following result on the large network size behavior of $R_{\text{sum}}^{\text{NCF}}(\sigma^2)$, $C^*(\sigma^2)$, and R_{sum}^{∞}.

Theorem 3.3.2. *Let the entries of the $L \times K$ channel gain matrix G be distributed as i.i.d. random variables with unit variance, and let $\sigma^2 = \sigma^2(L, K) > 0$. If $L \to \infty$ such that $L/K \to \rho \in (1, \infty]$ and $L/\sigma^2 \to \infty$, then*

$$R_{\text{sum}}^{\infty} \sim \frac{K}{2} \log L$$

and

$$R_{\text{sum}}^{NCF} = C^* - \frac{L}{2} \log\left(1 + \frac{1}{\sigma^2}\right) \sim \frac{K}{2} \log(L/\sigma^2),$$

a.s. in G. Similarly, if $K \to \infty$ such that $L/K \to \rho \in [0, 1)$ and $K/\sigma^2 \to \infty$, then

$$R_{\text{sum}}^{\infty} \sim \frac{L}{2} \log K$$

and

$$R_{\text{sum}}^{NCF} = C^* - \frac{L}{2} \log\left(1 + \frac{1}{\sigma^2}\right) \sim \frac{L}{2} \log(K/\sigma^2),$$

a.s. in G.

Theorem 3.3.2 acts as a powerful tool to examine the large network size asymptotics for $R_{\text{sum}}^{\text{NCF}}, R_{\text{sum}}^{\infty}$, and the fronthaul link capacity requirement C^*. We consider various scaling regimes of L and K, namely, K fixed and L growing, $L = \gamma K$ with $\gamma \notin \{0, 1\}$, and $L = K^\gamma$ with $\gamma \notin \{0, 1\}$.

For each case, we choose $\sigma^2 = \sigma^2(L,K)$ appropriately and use Theorem 3.3.2 to establish the scaling laws for this rich scattering model. The results are summarized in Table 3.3.

Table 3.3. Sum-rate scaling for fronthaul-limited and fronthaul-unlimited uplink C-RAN; $\gamma > 0$, $\gamma \neq 1$, $0 < \delta < \gamma - 1$, $0 < \varepsilon < 1$.

L vs. K		σ^2	C^*	$R_{\text{sum}}^{\text{NCF}}$	R_{sum}^{∞}
$L = \gamma K$	$(\gamma > 1)$	1	$\frac{K}{2}\log L$	$\frac{K}{2}\log L$	$\frac{K}{2}\log L$
	$(\gamma < 1)$	1	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$
$L = K^{\gamma}$	$(\gamma > 1)$	$K^{\gamma-1}$	$\frac{K}{2}\log K$	$\frac{K}{2}\log K$	$\frac{K}{2}\log L$
		$K^{\gamma-1-\delta}$	$\frac{K^{1+\delta}}{2}\log e$	$(1+\delta)\frac{K}{2}\log K$	
	$(\gamma < 1)$	1	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$
K fixed		L^{ε}	$\frac{L^{1-\varepsilon}}{2}\log e$	$(1-\varepsilon)\frac{K}{2}\log L$	$\frac{K}{2}\log L$
L fixed		1	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$	$\frac{L}{2}\log K$

Derivation of Table 3.3. First note that for all cases considered,

$$R_{\text{sum}}^{\text{MIMO}} \sim \frac{\min\{K,L\}}{2}\log\left(P\max\{K,L\}\right) \sim \frac{\min\{K,L\}}{2}\log\left(\max\{K,L\}\right).$$

For $L = \gamma K$ with $\gamma > 1$, we can choose $\sigma^2 = 1$ to obtain, from Theorem 3.3.1 and Lemma 3.3.2,

$$C^*(\sigma^2) \sim \frac{K}{2}\log(L/2) + \frac{L}{2} \sim \frac{K}{2}\log L,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{K}{2}\log(L/2) \sim \frac{K}{2}\log L. \tag{3.12}$$

For $L = \gamma K$ with $\gamma \in (0,1)$, we can similarly choose $\sigma^2 = 1$ to obtain, from Theorem 3.3.1 and Lemma 3.3.2,

$$C^*(\sigma^2) \sim \frac{L}{2}\log(K/2) + \frac{L}{2} \sim \frac{L}{2}\log K,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{L}{2}\log(K/2) \sim \frac{L}{2}\log K. \tag{3.13}$$

For $L = K^\gamma$ with $\gamma > 1$, we can choose $\sigma^2 = K^{\gamma-1}$ to obtain

$$C^*(\sigma^2) \sim \frac{K}{2}\log(L/K^{\gamma-1}) + \frac{L\log e}{2K^{\gamma-1}} \sim \frac{K}{2}\log K,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2)) \sim \frac{K}{2}\log(L/K^{\gamma-1}) \sim \frac{K}{2}\log K.$$

Alternatively, for the same scaling regime, we can choose $\sigma^2 = K^{\gamma-1-\delta}$ for some $\delta \in (0, \gamma-1)$ to obtain

$$C^*(\sigma^2) \sim \frac{K}{2}\log(L/K^{\gamma-1-\delta}) + \frac{L\log e}{2K^{\gamma-1-\delta}} \sim \frac{K^{1+\delta}}{2}\log e,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{K}{2}\log(L/K^{\gamma-1-\delta}) \sim (1+\delta)\frac{K}{2}\log K.$$

For $L = K^\gamma$ with $\gamma \in (0,1)$, we can choose $\sigma^2 = 1$ to obtain

$$C^*(\sigma^2) \sim \frac{L}{2}\log(K/2) + \frac{L}{2} \sim \frac{L}{2}\log K,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{L}{2}\log(K/2) \sim \frac{L}{2}\log K.$$

For fixed K with L growing, we can choose $\sigma^2 = L^{\varepsilon}$ for some $\varepsilon \in (0,1)$ to obtain

$$C^*(\sigma^2) \sim \frac{K}{2}\log(L/L^{\varepsilon}) + \frac{L\log e}{2L^{\varepsilon}} \sim \frac{L^{1-\varepsilon}}{2}\log e,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{K}{2}\log(L/L^{\varepsilon}) \sim (1-\varepsilon)\frac{K}{2}\log L.$$

Finally, for fixed L with K growing, we can choose $\sigma^2 = 1$ to obtain

$$C^*(\sigma^2) \sim \frac{L}{2}\log(K/2) + \frac{L}{2} \sim \frac{L}{2}\log K,$$

and

$$R_{\text{sum}}^{\text{NCF}}(\sigma^2) \sim \frac{L}{2}\log(K/2) \sim \frac{L}{2}\log K.$$

□

Remark 3.3.3. When L is fixed and K is growing, or $L = K^{\gamma}$ with $\gamma < 1$, the sum-rates scale sublinearly in K and therefore, the per-user rate is asymptotically zero if one attempts to serve all users fairly.

Remark 3.3.4. We note that most of the classical scaling results in the literature, as reviewed in Section 3.2, consider ergodic capacities and their limits. In contrast, our results focus on *global* interference network models with certain known statistical properties of the channel and make high-probability predictions on achievable rate regions.

Remark 3.3.5. The theory developed here does not lead to the same a.s. statements for the case $L = K \to \infty$. As a workaround, for every $\varepsilon > 0$, however small, one can choose to only serve

$(1-\varepsilon)K$ of the users, thereby leading to a sum-rate scaling of

$$(1-\varepsilon)K\log K/2$$

for this case, in accordance with Table 3.3.

Proof of Theorem 3.3.2. We will prove the first part, i.e., the case when $L \to \infty$. The second part will follow from this by exchanging the roles of K and L. Note that

$$\log\left|I+\frac{P}{\sigma^2+1}GG^T\right| = \log\left|I+\frac{P}{\sigma^2+1}G^TG\right|.$$

In the regime considered, $L \geq K$ eventually, therefore we can use G^T in place of W in Lemma 3.3.2. Assume first that $\rho < \infty$. We can conclude from Lemma 3.3.2 that w.p. 1, for every $\delta > 0$, there exists $L^*(\delta)$ such that for all $L \geq L^*(\delta)$, all eigenvlaues $\Lambda_1,\ldots,\Lambda_K$ of G^TG/K lie in

$$\left[(\sqrt{\rho}-1)^2-\delta,(\sqrt{\rho}+1)^2+\delta\right].$$

Therefore, w.p. 1, for every $L \geq L^*(\delta)$, we have

$$\begin{aligned}
&\log\left|I+\frac{P}{\sigma^2+1}GG^T\right| \\
&=\sum_{k=1}^{K}\log\left(1+\frac{PK}{\sigma^2+1}\Lambda_k\right) \\
&\in\left[K\log\left(1+\frac{PK}{\sigma^2+1}((\sqrt{\rho}-1)^2-\delta)\right),K\log\left(1+\frac{PK}{\sigma^2+1}((\sqrt{\rho}+1)^2+\delta)\right)\right]. \qquad (3.14)
\end{aligned}$$

Further, we have

$$\begin{aligned}
&\lim_{L\to\infty} \frac{K\log\left(1+\frac{PK}{\sigma^2+1}((\sqrt{\rho}\pm 1)^2\pm\delta)\right)}{K\log\left(\frac{PL}{\sigma^2}\right)} \\
&= \lim_{L\to\infty} \frac{\log\left(1+\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{L/K}\right)}{\log\left(\frac{PL}{\sigma^2+1}\right)} \\
&= \lim_{L\to\infty} \left[\frac{\log\left(1+\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{L/K}\right)}{\log\left(1+\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{\rho}\right)} \times \frac{\log\left(1+\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{\rho}\right)}{\log\left(\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{\rho}\right)}\right. \\
&\qquad\qquad \left.\times \frac{\log\left(\frac{PL}{\sigma^2+1}\cdot\frac{((\sqrt{\rho}\pm 1)^2\pm\delta)}{\rho}\right)}{\log\left(\frac{PL}{\sigma^2}\right)}\right] \\
&= 1,
\end{aligned}$$

since each factor approaches 1. Therefore, from (3.14), we obtain the limiting behavior

$$\log\left|I+\frac{P}{\sigma^2+1}GG^T\right| \sim K\log\left(\frac{PL}{\sigma^2}\right) \sim K\log(L/\sigma^2) \quad \text{w.p. } 1.$$

For the case $L/K\to\infty$, since all K eigenvalues of G^TG/L approach 1, we have

$$\log\left|I+\frac{P}{\sigma^2+1}GG^T\right| \sim K\log\left(\frac{PL}{\sigma^2}\right) \sim K\log(L/\sigma^2) \quad \text{w.p. } 1.$$

A similar line of reasoning with $P/(\sigma^2+1)$ replaced by P yields the result for R^{∞}_{sum}. □

Stochastic geometry

We now consider an alternative network model based on stochastic geometry [5]. In this model, users and relays are distributed over a 100m × 100m area according to independent Poisson point processes with intensities λ_u and λ_r (per 10^4 square meters) respectively. All channel gains are assumed to be real and unchanged for the duration of transmission.

As an initial simple model, the gain G_{lk} from sender k to relay l, separated by Euclidean distance r_{lk}, is modeled by $\max\{r_0, r_{lk}\}^{-\beta}$, where β is the *path loss exponent* and r_0 (set to 1 meter) is a minimum link distance to prohibit the singularity of the path loss for $r_{lk} \to 0$. In contrast to the rich scattering model, the channel randomness now comes exclusively from the placement of user and relay nodes, and once the nodes are fixed, the channel coefficients become deterministic. Therefore, this simple model approximates line-of-sight (LOS) propagation with no multipath component.

Following [33], we also study a more practical channel model based on stochastic geometry, where multipath effects such as blockage, shadowing, and fast fading are considered. More specifically, the multipath channel gain is given by

$$G_{lk} = \begin{cases} G_{lk}^{(\mathrm{LOS})} & \text{w.p. } p_{\mathrm{LOS}}(r_{lk}), \\ G_{lk}^{(\mathrm{NLOS})} & \text{w.p. } 1 - p_{\mathrm{LOS}}(r_{lk}), \end{cases}$$

where

$$G_{lk}^{(\mathrm{LOS})} = \frac{A_{lk}^{(\mathrm{LOS})} \Theta_{lk}^{(\mathrm{LOS})}}{\kappa \left(\max\{r_0, r_{lk}\}\right)^{\beta^{(\mathrm{LOS})}}}$$

and

$$G_{lk}^{(\mathrm{NLOS})} = \frac{A_{lk}^{(\mathrm{NLOS})} \Theta_{lk}^{(\mathrm{NLOS})}}{\kappa \left(\max\{r_0, r_{lk}\}\right)^{\beta^{(\mathrm{NLOS})}}}.$$

Here, NLOS stands for non-LOS. The random variable A_{lk} represents the fast fading component for modeling small-scale fluctuations in the envelope of the links in LOS and in NLOS. $A_{lk}^{(\mathrm{LOS})}$ and $A_{lk}^{(\mathrm{NLOS})}$ follow a *Nakagami-m* distribution with $m = 2$ and scale parameter $\Omega = 1$, and a *Rayleigh* distribution with scale parameter $\Omega = 1$, respectively. The factor Θ_{lk} models the shadowing effect due to changes in the surrounding environment. We consider a typical log-normal shadowing and set $\Theta_{lk}^{(\mathrm{LOS})}$ and $\Theta_{lk}^{(\mathrm{NLOS})}$ as log-normal random variables with means and standard deviations as specified in [33]. We also assume that A_{lk} and Θ_{lk} are independently distributed. The parameter κ is the free-space path loss at a distance of 1 meter from the sender at the center frequency f_c (which is set to 2.1 GHz here), and $\beta^{(\mathrm{LOS})}$ and $\beta^{(\mathrm{NLOS})}$ denote the path loss exponent for LOS and NLOS scenarios, respectively. We take $\beta^{(\mathrm{LOS})} = 2.5$ and $\beta^{(\mathrm{NLOS})} = 3.5$. Finally, $p_{\mathrm{LOS}}(r_{lk})$ represents the probability that the link is in LOS and is modeled according to the 3GPP urban micro (UMi) channel model [1] as

$$p_{\mathrm{LOS}}(r_{lk}) = \min\{18/r_{lk}, 1\}\left(1 - e^{-r_{lk}/36}\right) + e^{-r_{lk}/36},$$

where r_{lk} is measured in meters.

For simulating the large network asymptotics for all the aforementioned channel models, we examine the cases $\lambda_r = 2\lambda_u$, $\lambda_r = \lambda_u^2$, and λ_u fixed. The corresponding median sum-rates for fronthaul-limited and fronthaul-unlimited C-RAN uplink, as well as the corresponding C^* required, are plotted as functions of λ_u in Fig. 3.2 for different values of β. The median values are taken over 1000 runs of the simulations. For each simulation run, σ^2 is chosen so as to (numerically) minimize

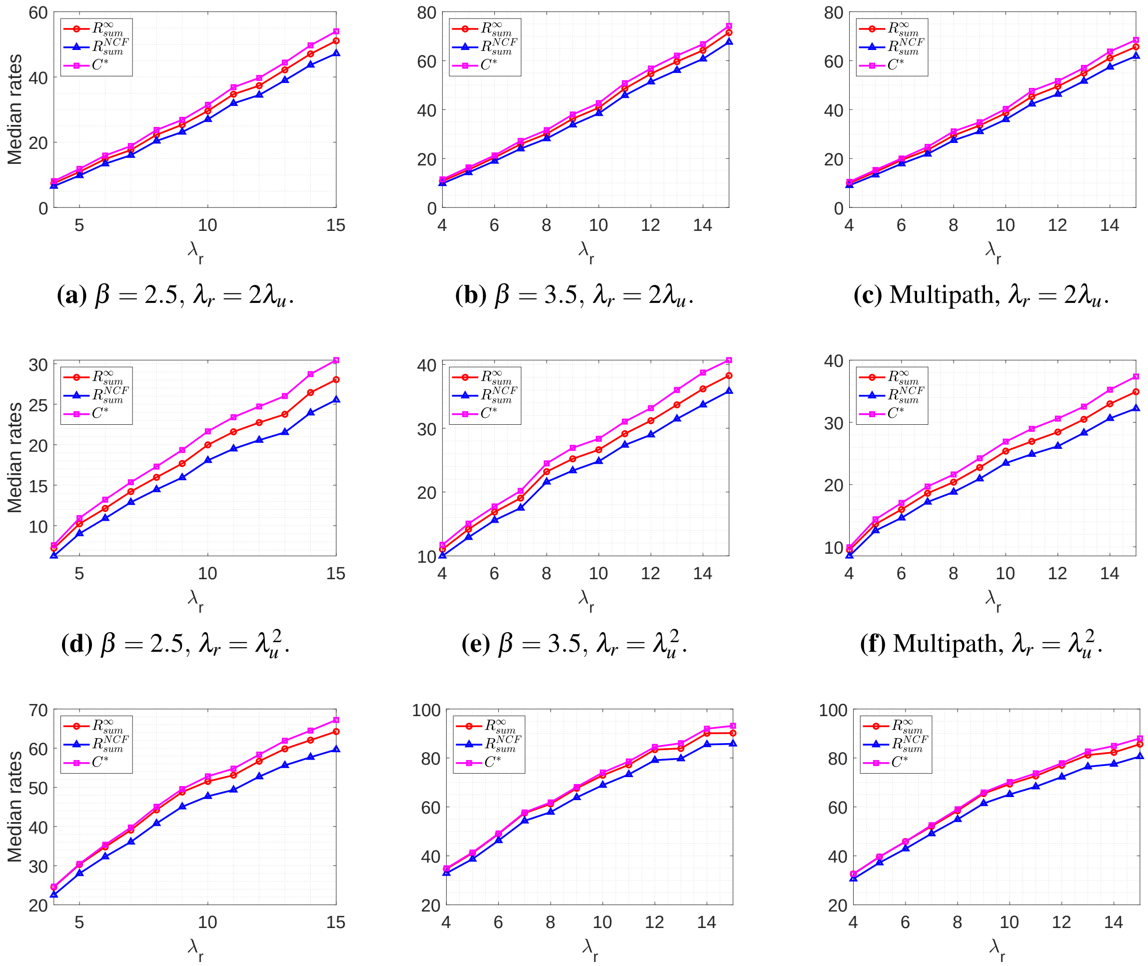
$$\max\left\{C^*(\sigma^2) - R_{\mathrm{sum}}^{\infty}, R_{\mathrm{sum}}^{\infty} - R_{\mathrm{sum}}^{\mathrm{NCF}}(\sigma^2)\right\}.$$

From the plots, we observe that $R_{\mathrm{sum}}^{\mathrm{NCF}}$ scales in a similar fashion as $R_{\mathrm{sum}}^{\infty}$, remaining only slightly lower, provided we have a slightly larger amount to spend on the fronthaul. Moreover, for $\lambda_r = 2\lambda_u$ as well as for $\lambda_r = \lambda_u^2$, the sum-rates show an approximately linear scaling with λ_r (and hence with L), unlike the $K \log L$ scaling observed for the rich scattering case. This loss seems to be caused by the dependence among the channel coefficients, which are still identically distributed but not independent of each other.

3.4 Downlink C-RANs

3.4.1 Comparisons with Fronthaul-Unlimited Downlink

Similar to Section 3.3.1, we can use Edmonds's polymatroid intersection theorem to quantify the total fronthaul $C_{\Sigma} := C_1 + \cdots + C_L$ required to approximate the fronthaul-unlimited



(a) $\beta = 2.5$, $\lambda_r = 2\lambda_u$. **(b)** $\beta = 3.5$, $\lambda_r = 2\lambda_u$. **(c)** Multipath, $\lambda_r = 2\lambda_u$.

(d) $\beta = 2.5$, $\lambda_r = \lambda_u^2$. **(e)** $\beta = 3.5$, $\lambda_r = \lambda_u^2$. **(f)** Multipath, $\lambda_r = \lambda_u^2$.

(g) $\beta = 2.5$, $\lambda_u = 10$. **(h)** $\beta = 3.5$, $\lambda_u = 10$. **(i)** Multipath, $\lambda_u = 10$.

Figure 3.2. Uplink capacity scaling under stochastic geometry.

downlink sum-capacity.

Theorem 3.4.1. *If*

$$C_\Sigma \geq \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right| =: C^*(\sigma^2)$$

for some $\sigma^2 > 0$, *then there exist* $C_1, C_2, \ldots, C_L \geq 0$ *with* $\sum_{l\in[L]} C_l = C_\Sigma$ *at which distributed decode–forward can achieve a sum-rate*

$$R_{\text{sum}}^{DDF}(\sigma^2) = \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right| - \frac{K}{2}\log\left(1 + \frac{1}{\sigma^2}\right).$$

Conversely, to achieve a sum-rate of $(1/2)\log|I + PHH^T|$, *we must have a total fronthaul* *capacity*

$$C_\Sigma \geq \frac{1}{2}\log|I + PHH^T|.$$

Proof. We assume that

$$\frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right| \geq \frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right), \tag{3.15}$$

since a negative sum-rate has no physical meaning. Define $r_k := R_k + (1/2)\log(1+1/\sigma^2)$, $k \in [K]$. We will work with the tuple $(r_1, \dots, r_K)$ instead of $(R_1, \dots, R_K)$. The maximum sum $r_1 + \cdots + r_K$ corresponding to $\mathscr{R}_{\text{down}}^{\text{DDF}}(\sigma^2)$ can be written as

$$\begin{aligned} r_{\max} &= \min_{\mathcal{S}_1 \subseteq [L]} \Big(\frac{1}{2}\log\Big|\frac{P}{\sigma^2}H_{[K],\mathcal{S}_1}H_{[K],\mathcal{S}_1}^T + I\Big| + \sum_{l \in \mathcal{S}_1^c} C_l\Big) \\ &= \min_{\mathcal{S}_1 \subseteq [L]} \Big(\phi(\mathcal{S}_1) + \psi(\mathcal{S}_1^c)\Big), \end{aligned} \tag{3.16}$$

where

$$\phi(\mathcal{S}_1) := \frac{1}{2}\log\Big|\frac{P}{\sigma^2}H_{[K],\mathcal{S}_1}H_{[K],\mathcal{S}_1}^T + I\Big|$$

and

$$\psi(\mathcal{S}_1^c) := \sum_{l \in \mathcal{S}_1^c} C_l$$

are such that $\mathscr{P}(\phi)$ and $\mathscr{P}(\psi)$ are both polymatroids. Therefore, by Edmonds's polymatroid intersection theorem,

$$r_{\max} = \max_{y^L}\left\{\sum_{l \in [L]} y_l : y_l \leq \psi(\{l\}), l \in [L], \sum_{l \in \mathcal{S}_1} y_l \leq \phi(\mathcal{S}_1), \mathcal{S}_1 \subseteq [L]\right\}.$$

Now, let us fix

$$C_\Sigma \geq \phi([L]) = \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right| \tag{3.17}$$

such that $C_1, \dots, C_L$ are constrained to satisfy $C_1 + \cdots + C_L = C_\Sigma$. Choose a point

$\mathbf{y}^* \equiv (y_1^*, \cdots, y_L^*) \in \mathscr{P}(\phi)$ such that $y_1^* + \ldots + y_L^* = \phi([L])$ and $y_l^* \geq (1/2)\log(1+1/\sigma^2)$ for each l. Such a point always exists since $\mathscr{P}(\phi)$ is a polymatroid and since (3.15) holds. The point $\tilde{\mathbf{y}} \equiv (\tilde{y}_1, \ldots, \tilde{y}_L)$ defined by

$$\tilde{y}_l = \frac{C_\Sigma}{\phi([L])} y_l^*, \quad l \in [L],$$

satisfies $\tilde{y}_1 + \cdots + \tilde{y}_L = C_\Sigma$. Therefore, choosing $C_l = \tilde{y}_l$ for each l, $\mathscr{P}(\phi)$ becomes the cuboid with corner point $\tilde{\mathbf{y}}$. Moreover, this cuboid includes the point $\mathbf{y}^*$, since $\tilde{y}_l \geq y_l^*$ for each l by (3.17). Thus, the point $\mathbf{y}^*$ lies in the intersection $\mathscr{P}(\phi) \cap \mathscr{P}(\psi)$ and therefore,

$$r_{\max} \geq y_1^* + \cdots + y_L^* = \phi([L]) = \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right|,$$

which implies that distributed decode–forward with the same fronthaul link capacities $(C_1, \ldots, C_L)$ can achieve

$$R_{\text{sum}}^{\text{DDF}} = \phi([L]) - \frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right) = \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right| - \frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right),$$

establishing the result. The converse follows immediately from the cutset bound. □

Remark 3.4.1. The best sum-rate achievable by our coding scheme for a given total fronthaul capacity $C_\Sigma > 0$ can be expressed as

$$R_{\text{sum}}^{\max}(C_\Sigma) = \sup_{\sigma^2>0}\left(\min\left\{C_\Sigma, \frac{1}{2}\log\left|\frac{P}{\sigma^2}HH^T + I\right|\right\} - \frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right).$$

Remark 3.4.2. As demonstrated in [67], one can write the sum-capacity of the fronthaul-unlimited downlink with channel gain matrix $H \in \mathbb{R}^{K\times L}$ as the solution of the optimization problem

$$\begin{aligned}
&\min_{Q}\max_{\Sigma}\frac{1}{2}\log\frac{|H^T\Sigma H+Q|}{|Q|}\\
\text{subject to}\quad &\Sigma\succeq 0 \quad \text{diagonal}, \quad \text{tr}(\Sigma)\le 1,\\
&Q\succeq 0 \quad \text{diagonal}, \quad \text{tr}(Q)\le 1/P.
\end{aligned}$$

Taking $Q=(1/PL)I$ and $\Sigma=I$ therefore yields an upper bound

$$R_{\text{sum}}^{\infty}\le\frac{1}{2}\log|PLHH^T+I|.$$

3.4.2 Capacity Scaling

Similar to Section 3.3.2, we first consider a rich scattering model. We can use Lemma 3.3.2 to establish the following theorem on the large network size behavior of $R_{\text{sum}}^{\text{DDF}}(\sigma^2)$, $C^*(\sigma^2)$, and R_{sum}^{∞}. The proof is similar to that of Theorem 3.3.2 and is omitted.

Theorem 3.4.2. *Let the entries of the $K\times L$ channel gain matrix H be distributed as i.i.d. random variables with variance* 1*, and let $\sigma^2=\sigma^2(L,K)>0$. If $L\to\infty$ such that $L/K\to\rho\in(1,\infty]$ and $L/\sigma^2\to\infty$, then*

$$\frac{1}{2}\le\liminf\frac{R_{\text{sum}}^{\infty}}{K\log L}\le\limsup\frac{R_{\text{sum}}^{\infty}}{K\log L}\le 1$$

and

$$R_{\text{sum}}^{DDF}=C^*-\frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right)\sim\frac{K}{2}\log(L/\sigma^2)-\frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right),$$

a.s. in H. Similarly, if $K\to\infty$ such that $L/K\to\rho\in[0,1)$ and $K/\sigma^2\to\infty$, then

$$\frac{1}{2}\le\liminf\frac{R_{\text{sum}}^{\infty}}{L\log K}\le\limsup\frac{R_{\text{sum}}^{\infty}}{L\log K}\le 1$$

and

$$R_{\text{sum}}^{DDF}=C^*-\frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right)\sim\frac{L}{2}\log(K/\sigma^2)-\frac{K}{2}\log\left(1+\frac{1}{\sigma^2}\right),$$

a.s. in H.

Using Theorem 3.4.2 and choosing $\sigma^2 = \sigma^2(L,K)$ appropriately, we summarize the scaling laws for $R_{\text{sum}}^{\text{DDF}}, R_{\text{sum}}^{\infty}$, and the fronthaul link capacity requirement C^* in Table 3.4.

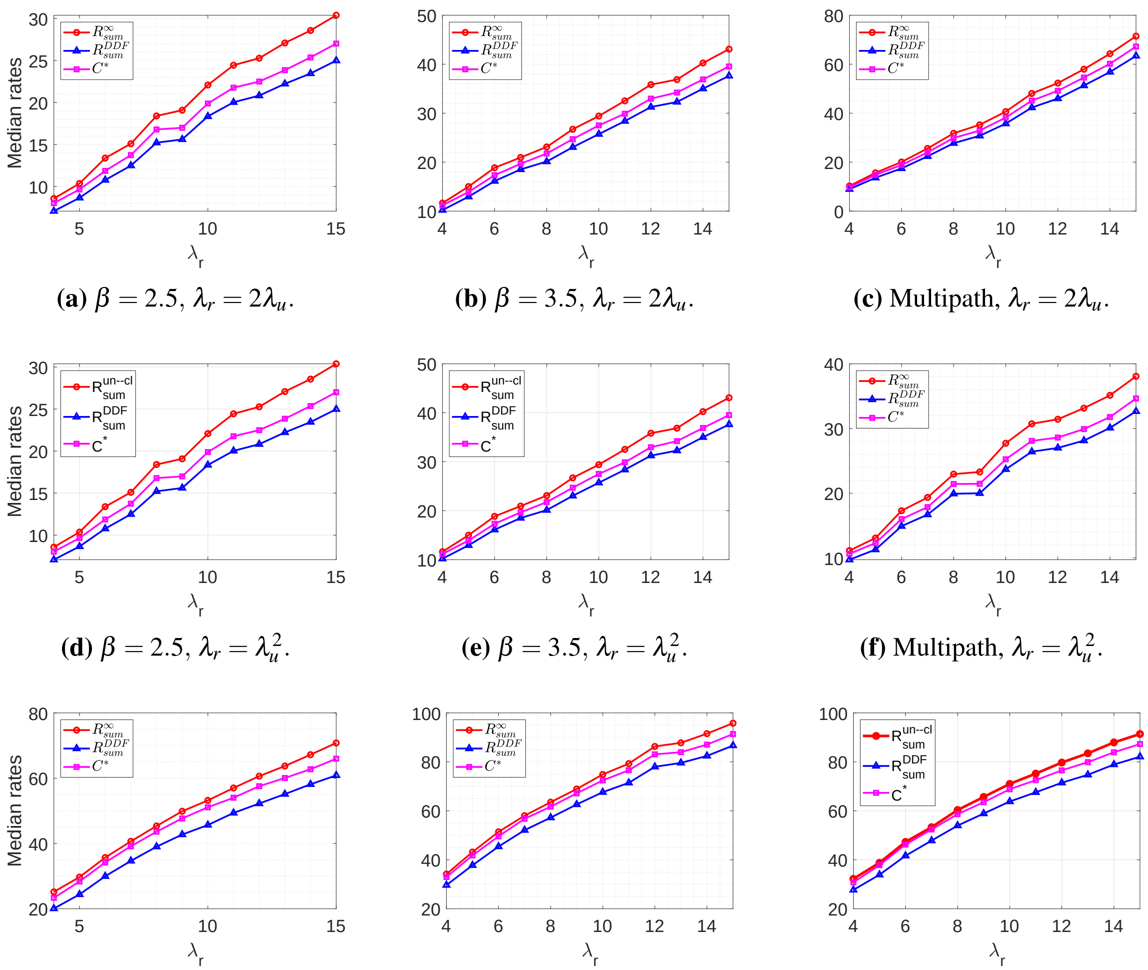
Remark 3.4.3. Unlike Table 3.3, Table 3.4 does not have an exact coefficient in the scaling law for R_{sum}^{∞} for downlink. The upper bound in Remark 3.4.2 scales as $L \log K$ or $K \log L$, while $R_{\text{sum}}^{\text{DDF}}$ serves as a lower bound on R_{sum}^{∞}.

For a stochastic geometry model similar to that in Section 3.3.2, Fig. 3.3 plots the median sum-rates obtained experimentally over 1000 simulation runs each, for different scaling regimes and different path loss exponents. The power constraint P at each relay is kept fixed. For each

Table 3.4. Sum-rate scaling for fronthaul-limited and fronthaul-unlimited downlink C-RAN; $\alpha \in [1/2, 1]$, $\gamma > 0$, $\gamma \neq 1$, $0 < \delta < \gamma - 1$, $0 < \varepsilon < 1$.

L vs. K		σ^2	C^*	$R_{\text{sum}}^{\text{DDF}}$	R_{sum}^{∞}
$L = \gamma K$	$(\gamma > 1)$	1	$\frac{K}{2} \log L$	$\frac{K}{2} \log L$	$\alpha K \log L$
	$(\gamma < 1)$	1	$\frac{L}{2} \log K$	$\frac{L}{2} \log K$	$\alpha L \log K$
$L = K^\gamma$	$(\gamma < 1)$	$L^{1/\gamma - 1}$	$\frac{L}{2} \log L$	$\frac{L}{2} \log L$	$\alpha L \log K$
		$L^{1/\gamma - 1 + \delta}$	$\frac{(1-\delta) L \log L}{2}$	$\frac{(1-\delta) L \log L}{2}$	
	$(\gamma > 1)$	1	$\frac{K}{2} \log L$	$\frac{K}{2} \log L$	$\alpha K \log L$
K fixed		1	$\frac{K}{2} \log L$	$\frac{K}{2} \log L$	$\alpha K \log L$

realization of the channel gain matrix H, σ^2 is chosen to (numerically) maximize $R_{\text{sum}}^{\text{DDF}}$, and then C^* is calculated using this value of σ^2. As before, the C-RAN downlink sum-rate closely tracks the fronthaul-unlimited downlink sum-capacity using a similar amount of fronthaul capacity. We note here that the plots show an *upper* bound on R_{sum}^{∞}, corresponding to choosing randomized values of the entries of the matrix Q in the dual characterization mentioned in Remark 3.4.2 and maximizing over the input covariance matrix Σ using the singular value decomposition of the channel gain matrix and water-filling power allocation (see, for example, [17, Section 9.1]).



(a) $\beta = 2.5$, $\lambda_r = 2\lambda_u$. **(b)** $\beta = 3.5$, $\lambda_r = 2\lambda_u$. **(c)** Multipath, $\lambda_r = 2\lambda_u$.

(d) $\beta = 2.5$, $\lambda_r = \lambda_u^2$. **(e)** $\beta = 3.5$, $\lambda_r = \lambda_u^2$. **(f)** Multipath, $\lambda_r = \lambda_u^2$.

(g) $\beta = 2.5$, $\lambda_u = 10$. **(h)** $\beta = 3.5$, $\lambda_u = 10$. **(i)** Multipath, $\lambda_u = 10$.

Figure 3.3. Downlink capacity scaling under stochastic geometry.

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Shouvik Ganguly and Young-Han Kim, "Capacity Scaling for Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *Proceedings of the IEEE International Symposium on Information Theory*, Paris, France, July 2019, as well as parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, "On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE Transactions on Information Theory*. The dissertation author was the primary researcher and author of this material.

Chapter 4

MIMO C-RANs

4.1 Introduction

In this chapter, we extend the results of Chapters 2 and 3 to the situation in which each user has N_u local antennas and each relay has N_r local antennas. The apparently more general situation in which users and/or relays have different numbers of antennas can be handled by setting the channel gain matrix appropriately. We assume a total average transmit power constraint P at each node and consider channel matrices $G \in \mathbb{R}^{N_r L \times N_u K}$ and $H \in \mathbb{R}^{N_u K \times N_r L}$ for uplink and downlink C-RANs, respectively. The objects $\mathscr{R}_{\text{up}}^{\text{NCF}}$, $R_{\text{sum}}^{\text{NCF}}$, $\mathscr{R}_{\text{up}}^{\text{CS}}$, $R_{\text{sum}}^{\text{CS}}$, $\mathscr{R}_{\text{down}}^{\text{DDF}}$, $R_{\text{sum}}^{\text{DDF}}$, and $\mathscr{R}_{\text{down}}^{\text{CS}}$ are defined as in Chapter 2.

4.2 Approximate Capacity

Similar to (2.5) in Section 2.2.2, the network compress–forward inner bound $\mathscr{R}_{\text{up}}^{\text{NCF}}$ for the uplink Gaussian MIMO C-RAN is characterized by the rate tuples $(R_1, \ldots, R_K)$ satisfying

$$\begin{aligned}\sum_{k \in \mathcal{S}_1} R_k &\le \frac{1}{2} \log \left| \frac{\sum_{k \in \mathcal{S}_1} G_{\mathcal{S}_2^c, k} \Gamma_k G_{\mathcal{S}_2^c, k}^T}{\sigma^2 + 1} + I \right| + \sum_{l \in \mathcal{S}_2} C_l - \frac{N_r |\mathcal{S}_2|}{2} \log \left(1 + \frac{1}{\sigma^2}\right) \\ &=: f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K) \qquad (4.1)\end{aligned}$$

for all $\mathcal{S}_1 \subseteq [K]$ and $\mathcal{S}_2 \subseteq [L]$ for some $\sigma^2 > 0$ and for some covariance matrices $\Gamma_1, \ldots, \Gamma_K \succeq 0$ such that $\text{tr}(\Gamma_k) = P$, $k \in [K]$. Here, for $\mathcal{S}_2 \subseteq [L]$ and $k \in [K]$, $G_{\mathcal{S}_2^c, k}$ denotes the $N_r |\mathcal{S}_2^c| \times N_u$

channel gain matrix between user k and the relays in $\mathcal{S}_2^c$. For each fixed $\sigma^2, \Gamma_1, \ldots, \Gamma_K$, we denote this region by $\mathscr{R}_{\text{up}}^{\text{NCF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_K)$. The cutset bound $\mathscr{R}_{\text{up}}^{\text{CS}}$ is characterized by rate tuples $(R_1, \ldots, R_K)$ satisfying

$$\begin{aligned}\sum_{k \in \mathcal{S}_1} R_k &\le \frac{1}{2} \log \Big| \sum_{k \in \mathcal{S}_1} G_{\mathcal{S}_2^c,k} \Gamma_k G_{\mathcal{S}_2^c,k}^T + I \Big| + \sum_{l \in \mathcal{S}_2} C_l \\ &=: f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K). \end{aligned} \tag{4.2}$$

Similar to (2.24) in Section 2.3.2, the distributed decode–forward inner bound $\mathscr{R}_{\text{down}}^{\text{DDF}}$ for the downlink Gaussian MIMO C-RAN is characterized by rate tuples $(R_1, \ldots, R_K)$ satisfying

$$\begin{aligned}\sum_{k \in \mathcal{S}_2^c} R_k &\le \frac{1}{2} \log \Big| \frac{\sum_{l \in \mathcal{S}_1} H_{\mathcal{S}_2^c,l} \Gamma_l H_{\mathcal{S}_2^c,l}^T}{\sigma^2} + I \Big| + \sum_{l \in \mathcal{S}_1^c} C_l - \frac{N_u |\mathcal{S}_2^c|}{2} \log \left(1 + \frac{1}{\sigma^2}\right) \\ &=: F_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_L) \end{aligned} \tag{4.3}$$

for all $\mathcal{S}_1 \subseteq [L]$ and $\mathcal{S}_2 \subseteq [K]$ for some $\sigma^2 > 0$ and for some covariance matrices $\Gamma_1, \ldots, \Gamma_L \succeq 0$ satisfying $\text{tr}(\Gamma_l) = P$, $l \in [L]$. For each fixed $\sigma^2, \Gamma_1, \ldots, \Gamma_L$, we denote this region by $\mathscr{R}_{\text{down}}^{\text{DDF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_L)$. The cutset bound $\mathscr{R}_{\text{down}}^{\text{CS}}$ is characterized by rate tuples $(R_1, \ldots, R_K)$ satisfying

$$\begin{aligned}\sum_{k \in \mathcal{S}_2^c} R_k &\le \frac{1}{2} \log \Big| H_{\mathcal{S}_2^c, \mathcal{S}_1} \tilde{\Gamma}_{\mathcal{S}_1 | \mathcal{S}_1^c} H_{\mathcal{S}_2^c, \mathcal{S}_1}^T + I \Big| + \sum_{l \in \mathcal{S}_1^c} C_l \\ &=: F_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2, \tilde{\Gamma}), \end{aligned} \tag{4.4}$$

where $\tilde{\Gamma}$ is a general $N_r L \times N_r L$ input covariance matrix satisfying the block trace constraints. Here, $H_{\mathcal{S}_2^c,l}$ denotes the $N_u |\mathcal{S}_2^c| \times N_r$ channel gain matrix between relay l and the users in $\mathcal{S}_2^c$, and $H_{\mathcal{S}_2^c, \mathcal{S}_1}$ denotes the $N_u |\mathcal{S}_2^c| \times N_r |\mathcal{S}_1|$ channel gain matrix between the relays in $\mathcal{S}_1$ and the users in $\mathcal{S}_2^c$.

We have the following result on the achievable per-user gaps from the capacity of the

MIMO C-RAN.

Proposition 4.2.1. *For every* $G \in \mathbb{R}^{N_r L \times N_u K}$ *and every* $P \in \mathbb{R}^+$*, if a rate tuple* $(R_1, \ldots, R_K)$ *is in* $\mathscr{R}_{\text{up}}^{CS}$*, then the rate tuple* $((R_1 - \Delta^{\text{up}})^+, \ldots, (R_K - \Delta^{\text{up}})^+)$ *is achievable, where*

$$\Delta^{\text{up}} \leq \frac{N_u}{2} \log\left(\frac{eN_r L}{N_u}\right).$$

Moreover,

$$\Delta_{\text{sum}}^{\text{up}} := R_{\text{sum}}^{CS} - \sup_{\sigma^2, \Gamma_1, \ldots, \Gamma_K} R_{\text{sum}}^{NCF}(\sigma^2, \Gamma_1, \ldots, \Gamma_K) \leq \begin{cases} \frac{N_r L}{2} H\left(\frac{N_u K}{N_r L}\right), & N_r L \geq 2N_u K, \\ \frac{N_r L}{2}, & N_r L < 2N_u K. \end{cases}$$

Similarly, for every $H \in \mathbb{R}^{N_u K \times N_r L}$ *and* $P \in \mathbb{R}^+$*, if a rate tuple* $(R_1, \ldots, R_K)$ *is in* $\mathscr{R}_{\text{down}}^{CS}$*, then the rate tuple* $((R_1 - \Delta^{\text{down}})^+, \ldots, (R_K - \Delta^{\text{down}})^+)$ *is achievable, where*

$$\Delta^{\text{down}} \leq \begin{cases} \frac{N_u}{2} \log(eN_r LK), & N_u < N_r L, \\ \frac{N_r L}{2} \log(eN_u K), & N_u \geq N_r L, \quad N_u K \geq 2N_r L, \\ \frac{N_u}{2} + \frac{N_r L}{2} \log(N_r L), & K = 1, \quad N_r L \leq N_u < 2N_r L. \end{cases}$$

Moreover,

$$\Delta_{\text{sum}}^{\text{down}} := R_{\text{sum}}^{CS} - \sup_{\sigma^2, \Gamma_1, \ldots, \Gamma_L} R_{\text{sum}}^{DDF}(\sigma^2, \Gamma_1, \ldots, \Gamma_L) \leq \frac{N_u K}{2} + \frac{\min\{N_r L, N_u K\}}{2} \log(N_r L).$$

Remark 4.2.1. Proposition 4.2.1 recovers the results of Theorems 2.2.1 and 2.3.1 when we set $N_u = N_r = 1$. From the expressions for Δ^{up} and Δ^{down}, we observe that the capacity gaps are the same as if there were $N_r \times L$ single-antenna relays. The sum-rate gaps are the same as if there were $N_u \times K$ single-antenna users, while Δ^{up} and Δ^{down} are in general larger than the gaps obtained with $N_u \times K$ single-antenna users.

Proof sketch of Proposition 4.2.1. The proof is an extension of the proofs of Theorems 2.2.1 and 2.3.1. Let us focus on the uplink first. Similar to Lemma 2.2.1, $f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K)$ satisfies the monotonicity property for each fixed $\mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K$. Therefore, similar to the line of argument in Section 2.2.2, the per-user rate gap from the cutset bound can be upper-bounded as

$$\begin{aligned}
\Delta^{\text{up}} &\leq \max_{\substack{\mathcal{S}_1 \subseteq [K] \\ \mathcal{S}_1 \neq \emptyset}} \Bigg[\frac{\max_{\Gamma_1,\ldots,\Gamma_K} \min_{\mathcal{S}_2} f_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K)}{|\mathcal{S}_1|} \\
&\qquad\qquad - \frac{\max_{\Gamma_1,\ldots,\Gamma_K} \min_{\mathcal{S}_2} f_{\text{in}}(\mathcal{S}_1, \mathcal{S}_2, \Gamma_1, \ldots, \Gamma_K)}{|\mathcal{S}_1|} \Bigg] \\
&\leq \max_{\substack{\mathcal{S}_1 \subseteq [K] \\ \mathcal{S}_1 \neq \emptyset}} \max_{\Gamma_1,\ldots,\Gamma_K} \max_{\mathcal{S}_2} \frac{1}{2|\mathcal{S}_1|} \log \frac{\left| \sum_{k \in \mathcal{S}_1} G_{\mathcal{S}_2^c,k} \Gamma_k G_{\mathcal{S}_2^c,k}^T + I \right|}{\left| \frac{\sum_{k \in \mathcal{S}_1} G_{\mathcal{S}_2^c,k} \Gamma_k G_{\mathcal{S}_2^c,k}^T}{\sigma^2+1} + I \right|} + \frac{N_r |\mathcal{S}_2|}{2} \log\left(1 + \frac{1}{\sigma^2}\right) \\
&\overset{(a)}{\leq} \max_{\substack{k \in [K] \\ l \in \{0,\ldots,l\}}} \left[\frac{\min\{N_r(L-l), N_u k\}}{2k} \log(1+\sigma^2) + \frac{N_r l}{2} \log\left(1 + \frac{1}{\sigma^2}\right) \right],
\end{aligned}$$

where in (a), we set $|\mathcal{S}_1| = k$, $|\mathcal{S}_2| = l$, and upper-bound rank $\left(\sum_{k \in \mathcal{S}_1} G_{\mathcal{S}_2^c,k} \Gamma_k G_{\mathcal{S}_2^c,k}^T\right)$ by $\min\{N_r(L-l), N_u k\}$. The maximization yields

$$\Delta^{\text{up}} \leq \begin{cases} \frac{N_u}{2} \log(1+\sigma^2) + \frac{N_r L - N_u}{2} \log(1 + \frac{1}{\sigma^2}), & \sigma^2 \geq 1, N_r L \geq N_u, \\ \frac{N_r L}{2} \log(1+\sigma^2), & \sigma^2 \geq 1, N_r L < N_u, \\ \frac{N_r L}{2} \log(1 + \frac{1}{\sigma^2}), & \sigma^2 \leq 1. \end{cases}$$

Since this holds for every $\sigma^2 > 0$, we set

$$\sigma^2 = \frac{N_r L}{N_u} - 1$$

for $L \geq 2N_u/N_r$ to obtain

$$\begin{aligned}\Delta^{\text{up}} &\leq \frac{N_u}{2}\log\frac{N_rL}{N_u} + \frac{N_rL - N_u}{2}\log\left(1 + \frac{N_u}{N_rL - N_u}\right) \\ &\stackrel{(a)}{\leq} \frac{N_u}{2}\log\frac{eN_rL}{N_u}.\end{aligned} \tag{4.5}$$

Here, (a) follows since from elementary calculus, we know that for $x > 0$, $\log(1+x) \leq x\log e$. For $L < 2N_u/N_r$, we can choose $\sigma^2 = 1$ to obtain

$$\Delta^{\text{up}} \leq \frac{N_rL}{2} \stackrel{(a)}{\leq} \frac{N_u}{2}\log\frac{eLN_r}{N_u},$$

where (a) follows from the inequality $x \leq \log(ex)$ for $x < 1/2$. This, together with (4.5), establishes the per-user rate gap for the uplink MIMO C-RAN. For the sum-rate gap, we consider

$$\begin{aligned}\Delta^{\text{up}}_{\text{sum}} &\leq \max_{\mathcal{S}_1,\mathcal{S}_2,\Gamma_1,\ldots,\Gamma_K} \left(f_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2,\Gamma_1,\ldots,\Gamma_K) - f_{\text{in}}(\mathcal{S}_1,\mathcal{S}_2,\Gamma_1,\ldots,\Gamma_K)\right) \\ &\leq \max_{\substack{k\in[K] \\ l\in\{0,\ldots,L\}}} \left[\frac{\min\{N_r(L-l), N_uk\}}{2}\log(1+\sigma^2) + \frac{N_rl}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right].\end{aligned} \tag{4.6}$$

Maximization of (4.6) over l and k yields, for $\sigma^2 \geq 1$,

$$\Delta^{\text{up}}_{\text{sum}} \leq \begin{cases} \frac{N_uK}{2}\log(1+\sigma^2) + \frac{N_rL - N_uK}{2}\log(1+\frac{1}{\sigma^2}), & N_rL \geq N_uK, \\ \frac{N_rL}{2}\log(1+\sigma^2), & N_rL < N_uK. \end{cases}$$

For $N_rL \leq 2N_uK$, we can then choose $\sigma^2 = 1$ to obtain an upper bound $\Delta^{\text{up}}_{\text{sum}} \leq N_rL/2$. For $N_rL > 2N_uK$, we can choose $\sigma^2 = N_rL/N_uK - 1 \geq 1$ to obtain

$$\begin{aligned}\Delta^{\text{up}}_{\text{sum}} &\leq \frac{N_uK}{2}\log\left(\frac{N_rL}{N_uK}\right) + \frac{N_rL - N_uK}{2}\log\left(1 + \frac{N_uK}{N_rL - N_uK}\right) \\ &= \frac{N_rL}{2}H(N_uK/N_rL).\end{aligned} \tag{4.7}$$

For the downlink MIMO C-RAN with channel gain matrix $H \in \mathbb{R}^{N_u K \times N_r L}$, the per-user rate gap from the cutset bound can be upper-bounded, similar to Section 2.3.2, as

$$\begin{aligned}
\Delta^{\text{down}} &\leq \max_{\mathcal{S}_2 \subsetneq [K]} \Bigg[\frac{1}{|\mathcal{S}_2^c|} \Bigg(\max_{\tilde{\Gamma}} \min_{\mathcal{S}_1} F_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2, \tilde{\Gamma}) \\
&\qquad - \max_{\Gamma_1,\ldots,\Gamma_L} \min_{\mathcal{S}_1} \min_{\mathcal{T}_2 \subseteq \mathcal{S}_2} F_{\text{in}}(\mathcal{S}_1, \mathcal{T}_2, \Gamma_1, \ldots, \Gamma_L) \Bigg) \Bigg] \\
&\leq \max_{\mathcal{S}_1, \mathcal{S}_2, \mathcal{T}_2 \subseteq \mathcal{S}_2} \max_{\tilde{\Gamma}} \min_{\Gamma_1,\ldots,\Gamma_L} \frac{F_{\text{out}}(\mathcal{S}_1, \mathcal{S}_2, \tilde{\Gamma}) - F_{\text{in}}(\mathcal{S}_1, \mathcal{T}_2, \Gamma_1, \ldots, \Gamma_L)}{|\mathcal{S}_2^c|} \\
&\leq \max_{\mathcal{S}_1, \mathcal{S}_2, \mathcal{T}_2 \subseteq \mathcal{S}_2} \max_{\tilde{\Gamma}} \min_{\Gamma_1,\ldots,\Gamma_L} \frac{1}{2|\mathcal{S}_2^c|} \Bigg[\log \frac{\left| H_{\mathcal{S}_2^c,\mathcal{S}_1} \tilde{\Gamma}_{\mathcal{S}_1} H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I \right|}{\left| \frac{\sum_{l \in \mathcal{S}_1} H_{\mathcal{S}_2^c,l} \Gamma_l H_{\mathcal{S}_2^c,l}^T}{\sigma^2} + I \right|} \\
&\qquad + N_u |\mathcal{T}_2^c| \log \left(1 + \frac{1}{\sigma^2} \right) \Bigg]. \qquad (4.8)
\end{aligned}$$

Writing $\tilde{\Gamma}_{\mathcal{S}_1} = U \Lambda U^T$ where U is orthogonal and Λ is diagonal, letting $H_{\mathcal{S}_2^c,\mathcal{S}_1} U =: [B_1 \quad B_2 \quad \cdots \quad B_{|\mathcal{S}_1|}]$, where $B_1, \ldots, B_{|\mathcal{S}_1|}$ are $N_u |S_2^c| \times N_r$ matrices satisfying $\sum_{l=1}^{|\mathcal{S}_1|} \text{tr}(B_l^T B_l) = \left\| H_{\mathcal{S}_2^c,\mathcal{S}_1} \right\|_F^2$, and taking $\Gamma_l = (P/N_r) I$ for each l, we have

$$\begin{aligned}
\log \frac{\left| H_{\mathcal{S}_2^c,\mathcal{S}_1} \tilde{\Gamma}_{\mathcal{S}_1} H_{\mathcal{S}_2^c,\mathcal{S}_1}^T + I \right|}{\left| \frac{\sum_{l \in \mathcal{S}_1} H_{\mathcal{S}_2^c,l} \Gamma_l H_{\mathcal{S}_2^c,l}^T}{\sigma^2} + I \right|} &= \log \frac{\left| I + \sum_{l=1}^{|\mathcal{S}_1|} B_l \Lambda_l B_l^T \right|}{\left| I + \frac{P}{N_r \sigma^2} \sum_{l=1}^{|\mathcal{S}_1|} B_l B_l^T \right|} \\
&\overset{(a)}{\leq} \log \frac{\left| I + P|\mathcal{S}_1| \sum_{l=1}^{|\mathcal{S}_1|} B_l B_l^T \right|}{\left| I + \frac{P}{N_r \sigma^2} \sum_{l=1}^{|\mathcal{S}_1|} B_l B_l^T \right|} \\
&\leq \min\{N_u |\mathcal{S}_2^c|, N_r |\mathcal{S}_1|\} \log \left(\sigma^2 N_r |\mathcal{S}_1| \right),
\end{aligned}$$

provided $\sigma^2 \geq \frac{1}{N_r |\mathcal{S}_1|}$. Here, (a) follows since the trace of $\tilde{\Gamma}_{\mathcal{S}_1}$ is upper bounded by $P|\mathcal{S}_1|$.

Continuing from (4.8), we thus have

$$\begin{aligned}\Delta^{\text{down}} &\leq \max_{\mathcal{S}_1,\mathcal{S}_2,\mathcal{T}_2\subseteq\mathcal{S}_2}\left[\frac{N_u|\mathcal{T}_2^c|\log\left(1+\frac{1}{\sigma^2}\right)}{2|\mathcal{S}_2^c|}+\frac{\min\{N_u|\mathcal{S}_2^c|,N_r|\mathcal{S}_1|\}}{2|\mathcal{S}_2^c|}\log\left(\sigma^2 N_r|\mathcal{S}_1|\right)\right]\\ &=\frac{N_uK}{2}\log\left(1+\frac{1}{\sigma^2}\right)+\frac{\min\{N_u,N_rL\}}{2}\log(\sigma^2N_rL).\end{aligned}\tag{4.9}$$

This holds for every $\sigma^2\geq 1$, so we set $\sigma^2=K-1$ for $K\geq 2$ and $N_rL\geq N_u$ to obtain

$$\begin{aligned}\Delta^{\text{down}} &\leq \frac{N_u}{2}\log L+\frac{N_u}{2}\left(\log N_r+K\log K-(K-1)\log(K-1)\right)\\ &\leq \frac{N_u}{2}\log(eN_rLK).\end{aligned}\tag{4.10}$$

For $K=1$ and $N_rL\geq N_u$, we can set $\sigma^2=1$ in (4.9) to obtain

$$\Delta^{\text{down}}\leq\frac{N_u}{2}\log(2N_rL)\leq\frac{N_u}{2}\log(eN_rL).\tag{4.11}$$

For $N_rL<N_u$ and $N_uK\geq 2N_rL$, set

$$\sigma^2=\frac{N_uK}{N_rL}-1$$

to obtain

$$\Delta^{\text{down}}\leq\frac{N_uK}{2}\log(N_uK)-\frac{N_uK-N_rL}{2}\log(N_uK-N_rL)\leq\frac{N_rL}{2}\log(eN_uK),\tag{4.12}$$

and for $N_rL<N_u<2N_rL$ and $K=1$, set $\sigma^2=1$ to obtain

$$\begin{aligned}\Delta^{\text{down}} &\leq \frac{N_u}{2}+\frac{N_rL}{2}\log(N_rL)\\ &\leq\frac{N_u}{2}\log(eN_rL).\end{aligned}\tag{4.13}$$

The results (4.10)–(4.13) establish the per-user gap results for the downlink MIMO C-RAN.

For the sum-rate gap, we similarly obtain

$$\begin{aligned}\Delta_{\text{sum}}^{\text{down}} &\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \max_{\tilde{\Gamma}} \min_{\Gamma_1,\ldots,\Gamma_L} \left(F_{\text{out}}(\mathcal{S}_1,\mathcal{S}_2,\tilde{\Gamma}) - F_{\text{in}}(\mathcal{S}_1,\mathcal{S}_2,\Gamma_1,\ldots,\Gamma_L)\right) \\ &\leq \max_{\mathcal{S}_1,\mathcal{S}_2} \left[\frac{\min\{N_u|\mathcal{S}_2^c|, N_r|\mathcal{S}_1|\}}{2}\log(\sigma^2 N_r|\mathcal{S}_1|) + \frac{N_u|\mathcal{S}_2^c|}{2}\log\left(1+\frac{1}{\sigma^2}\right)\right]\end{aligned} \tag{4.14}$$

if $\sigma^2 \geq 1/N_r|\mathcal{S}_1|$ for each $\mathcal{S}_1 \neq \emptyset$. Maximization of (4.14) over $|\mathcal{S}_1|$ and $|\mathcal{S}_2^c|$ yields, for $\sigma^2 \geq 1$,

$$\Delta_{\text{sum}}^{\text{down}} \leq \frac{\min\{N_r L, N_u K\}}{2}\log(N_r L\sigma^2) + \frac{N_u K}{2}\log\left(1+\frac{1}{\sigma^2}\right). \tag{4.15}$$

We can then choose $\sigma^2 = 1$ in (4.15) to obtain

$$\Delta_{\text{sum}}^{\text{down}} \leq \frac{\min\{N_r L, N_u K\}}{2}\log(N_r L) + \frac{N_u K}{2},$$

completing the proof. □

4.3 Fronthaul Requirement and Capacity Scaling

In this section, we quantify the fronthaul requirements for the uplink and downlink MIMO C-RAN sum-capacities to approximate the fronthaul-unlimited capacities. To this end, we first note that similar to Sections 2.2.2 and 2.3.2, one can characterize the achievable sum-rates

$R_{\text{sum}}^{\text{NCF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_K)$ and $R_{\text{sum}}^{\text{DDF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_L)$ as

$$\begin{aligned}
&R_{\text{sum}}^{\text{NCF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_K) \\
&= \min_{\mathcal{S}_2 \subseteq [L]} \left(\frac{1}{2} \log \left| \frac{1}{\sigma^2 + 1} \sum_{k \in [K]} G_{\mathcal{S}_2^c, k} \Gamma_k G_{\mathcal{S}_2^c, k}^T + I \right| + \sum_{l \in \mathcal{S}_2} C_l - \frac{N_r |\mathcal{S}_2|}{2} \log \left(1 + \frac{1}{\sigma^2} \right) \right) \\
&= \max_{y^L} \Bigg\{ \sum_{l \in [L]} y_l : y_l \leq C_l - \frac{N_r}{2} \log \left(1 + \frac{1}{\sigma^2} \right), l \in [L], \\
&\qquad\qquad \sum_{l \in \mathcal{S}_2} y_l \leq \log \left| \frac{1}{\sigma^2 + 1} \sum_{k \in [K]} G_{\mathcal{S}_2^c, k} \Gamma_k G_{\mathcal{S}_2^c, k}^T + I \right|, \mathcal{S}_2 \subseteq [L] \Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
&R_{\text{sum}}^{\text{DDF}}(\sigma^2, \Gamma_1, \ldots, \Gamma_L) + \frac{N_u K}{2} \log \left(1 + \frac{1}{\sigma^2} \right) \\
&= \min_{\mathcal{S}_1 \subseteq [L]} \left(\frac{1}{2} \log \left| \frac{1}{\sigma^2} \sum_{l \in \mathcal{S}_1} H_{[K], l} \Gamma_l H_{[K], l}^T + I \right| + \sum_{l \in \mathcal{S}_1^c} C_l \right) \\
&= \max_{y^L} \left\{ \sum_{l \in [L]} y_l : y_l \leq C_l, l \in [L], \sum_{l \in \mathcal{S}_1} y_l \leq \frac{1}{2} \log \left| \frac{1}{\sigma^2} \sum_{l \in \mathcal{S}_1} H_{[K], l} \Gamma_l H_{[K], l}^T + I \right|, \mathcal{S}_1 \subseteq [L] \right\},
\end{aligned}$$

leading to the following extension of Theorems 3.3.1 and 3.4.1.

Proposition 4.3.1. *If*

$$C_\Sigma \geq \frac{1}{2} \log \left| \frac{1}{\sigma^2 + 1} \sum_{k \in [K]} G_{[L], k} \Gamma_k G_{[L], k}^T + I \right| + \frac{N_r L}{2} \log \left(1 + \frac{1}{\sigma^2} \right) =: C^*(\sigma^2, \Gamma_1, \ldots, \Gamma_K)$$

for some $\sigma^2, \Gamma_1, \ldots, \Gamma_K$, then there exist fronthaul link capacities $C_1, C_2, \ldots, C_L \geq 0$ with $\sum_{l \in [L]} C_l = C_\Sigma$ at which network compress–forward can achieve a sum-rate

$$R_{\text{sum}}^{NCF}(\sigma^2, \Gamma_1, \ldots, \Gamma_K) = \frac{1}{2} \log \left| \frac{1}{\sigma^2 + 1} \sum_{k \in [K]} G_{[L], k} \Gamma_k G_{[L], k}^T + I \right|.$$

If

$$C_\Sigma \geq \frac{1}{2}\log\left|\frac{1}{\sigma^2}\sum_{l\in[L]} H_{[K],l}\Gamma_l H_{[K],l}^T + I\right| =: C^*(\sigma^2,\Gamma_1,\ldots,\Gamma_L)$$

for some $\sigma^2,\Gamma_1,\ldots,\Gamma_L$*, then there exist* $C_1,C_2,\ldots,C_L \geq 0$ *with* $\sum_{l\in[L]} C_l = C_\Sigma$ *at which distributed decode–forward can achieve a sum-rate*

$$R_{\text{sum}}^{DDF}(\sigma^2,\Gamma_1,\ldots,\Gamma_L) = \frac{1}{2}\log\left|\frac{1}{\sigma^2}\sum_{l\in[L]} H_{[K],l}\Gamma_l H_{[K],l}^T + I\right| - \frac{N_u K}{2}\log\left(1+\frac{1}{\sigma^2}\right).$$

Remark 4.3.1. Following a similar line of reasoning as in [67, Section IV-B], one can write the sum-capacity of the fronthaul-unlimited MIMO downlink with channel gain matrix $H \in \mathbb{R}^{N_u K\times N_r L}$ as the solution of the optimization problem

$$\begin{aligned}
&\min_Q \max_\Sigma \frac{1}{2}\log\frac{|H^T\Sigma H + Q|}{|Q|}\\
\text{subject to}\quad & Q \quad \text{non-negative diagonal},\\
&\Sigma = \begin{bmatrix}\Sigma_1 & & & \\ & \Sigma_2 & & \\ & & \ddots & \\ & & & \Sigma_K\end{bmatrix},\\
&\text{submatrix}\quad Q([(l-1)N_r+1:lN_r],[(l-1)N_r+1:lN_r])\\
&\qquad\qquad\text{having equal diagonals for}\quad l=1,\ldots,L,\\
&\operatorname{tr}(Q)\leq N_r/P,\\
&\operatorname{tr}(\Sigma)\leq 1.
\end{aligned}$$

Here, $\Sigma_k \succeq 0$ is of size $N_u\times N_u$ for every $k\in[K]$.

Remark 4.3.2. Generalizing Remarks 3.3.2 and 3.4.1, the best sum-rates achievable for a total fronthaul capacity $C_\Sigma > 0$ can be expressed as

$$R_{\text{sum}}^{\max}(C_\Sigma) = \sup_{\sigma^2>0} \min\left\{ C_\Sigma - \frac{N_r L}{2} \log\left(1+\frac{1}{\sigma^2}\right), \right.$$
$$\left. \max_{\Gamma_1,\ldots,\Gamma_K} \frac{1}{2} \log\left| \frac{1}{\sigma^2+1} \sum_{k\in[K]} G_{[L],k} \Gamma_k G_{[L],k}^T + I \right| \right\}$$

for the uplink, and

$$R_{\text{sum}}^{\max}(C_\Sigma) = \sup_{\sigma^2>0} \left(\min\left\{ C_\Sigma, \max_{\Gamma_1,\ldots,\Gamma_L} \frac{1}{2} \log\left| \frac{1}{\sigma^2} \sum_{l\in[l]} H_{[K],l} \Gamma_l H_{[K],l}^T + I \right| \right\} \right.$$
$$\left. - \frac{N_u K}{2} \log\left(1+\frac{1}{\sigma^2}\right) \right)$$

for the downlink.

We have the following result on the large-network asymptotics of the MIMO C-RAN.

Proposition 4.3.2. *Let the entries of the $N_r L \times N_u K$ channel gain matrix G be distributed as i.i.d.* N$(0,1)$, *and let $\sigma^2 = \sigma^2(N_r, L, N_u, K) > 0$. If $N_r L \to \infty$ such that $N_r L / N_u K \to \rho \in (1, \infty]$ and $N_r L/\sigma^2 \to \infty$, and if N_u is kept fixed, then*

$$R_{\text{sum}}^{\infty} \sim \frac{N_u K}{2} \log(N_r L)$$

and for every choice of $\Gamma_1, \ldots, \Gamma_K$,

$$R_{\text{sum}}^{NCF} = C^* - \frac{N_r L}{2} \log\left(1+\frac{1}{\sigma^2}\right) \sim \frac{N_u K}{2} \log(N_r L/\sigma^2),$$

a.s. in G. Similarly, let the entries of the $N_u K \times N_r L$ channel gain matrix H be distributed as i.i.d. N$(0,1)$, *and let $\sigma^2 = \sigma^2(N_r, L, N_u, K) > 0$. If $L \to \infty$ such that $N_r L / N_u K \to \rho \in (1, \infty]$ and*

$L/(N_u\sigma^2) \to \infty$, *and if* N_r *is kept fixed, then for every choice of* $\Gamma_1,\ldots,\Gamma_L$,

$$R_{\text{sum}}^{DDF} = C^* - \frac{N_u K}{2}\log\left(1+\frac{1}{\sigma^2}\right) \sim \frac{N_u K}{2}\log(L/\sigma^2) - \frac{N_u K}{2}\log\left(1+\frac{1}{\sigma^2}\right),$$

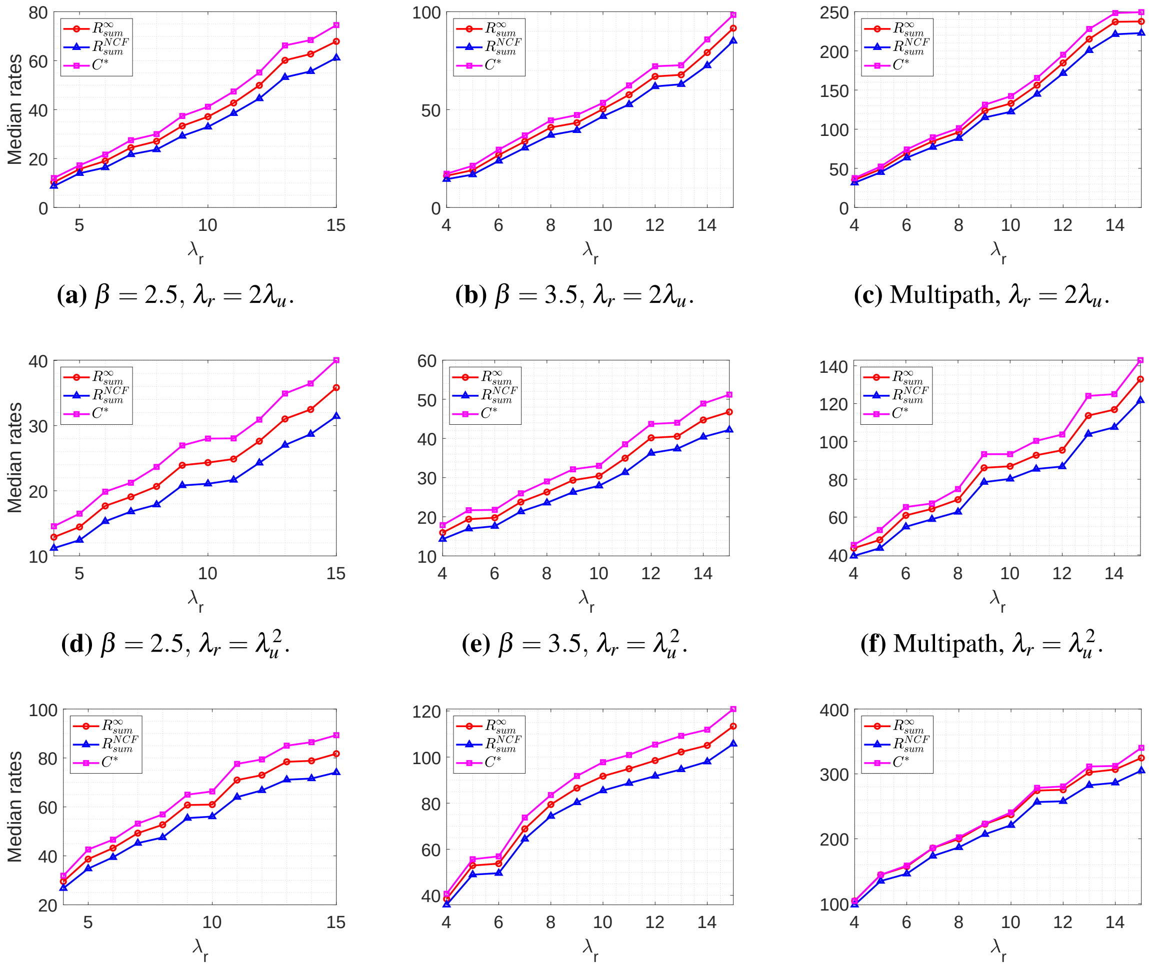
a.s. in H.

Remark 4.3.3. Comparing Proposition 4.3.2 with Theorems 3.3.2 and 3.4.2 shows that under the rich scattering model, the large network asymptotics of the MIMO C-RAN is the same as if there were N_uK users and N_rL relays. Thus, Tables 3.3 and 3.4 can be easily generalized to the MIMO case through appropriate choices of $\sigma^2(N_r, L, N_u, K)$.

Remark 4.3.4. In Proposition 4.3.2, N_u and N_r are held fixed for uplink and downlink, respectively, so that the scaling results remain invariant to the power allocation across the local antennas at each user and at each relay, respectively.

Similar to Sections 3.3.2 and 3.4.2, Figs. 4.1 and 4.2 plot $R_{\text{sum}}^{\text{NCF}}$, $R_{\text{sum}}^{\text{DDF}}$, C^*, and R_{sum}^{∞} under a stochastic geometry model. Both N_r and N_u are kept fixed ($N_r = N_u = 4$) for these simulations. For the downlink, as before, we plot an upper bound on R_{sum}^{∞}, obtained by a grid search over eligible values of Q in Remark 4.3.1. At each node (user or relay), the shadowing effect is considered to be the same across all local antennas, while the small-scale fading is taken as i.i.d.

To quantify the advantages of having multiple local antennas at each user and each relay (i.e., the advantage of “using MIMO”), Figs. 4.3 and 4.4 plot the best sum-rates achievable for a given C_Σ as the number of local antennas grows, in accordance with Remark 4.3.2. For these simulations, we take K users and L relays distributed uniformly over a 100m $\times$ 100m area, where $K = 4$ and $L = 6$. We consider the cases $C_\Sigma = 20$, 40, 60, and 80 bits per transmission. We take $N_u = N_r$ in all these simulations, for simplicity. The gains from MIMO are more significant at higher C_Σ and under multipath models. Intuitively, a larger fronthaul provides a pipeline for the flow of the extra information available through the use of a larger number of antennas in the



(a) $\beta = 2.5$, $\lambda_r = 2\lambda_u$. **(b)** $\beta = 3.5$, $\lambda_r = 2\lambda_u$. **(c)** Multipath, $\lambda_r = 2\lambda_u$.

(d) $\beta = 2.5$, $\lambda_r = \lambda_u^2$. **(e)** $\beta = 3.5$, $\lambda_r = \lambda_u^2$. **(f)** Multipath, $\lambda_r = \lambda_u^2$.

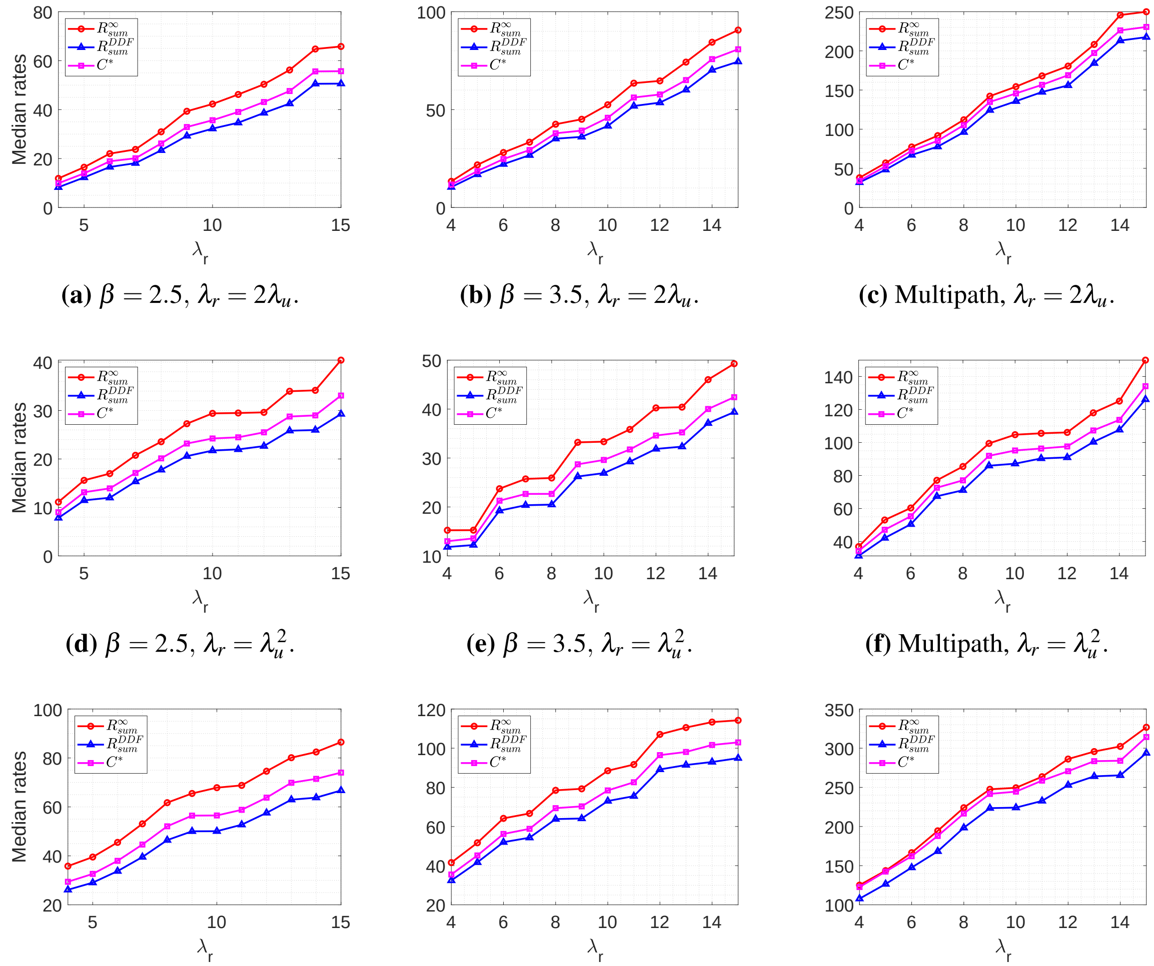
(g) $\beta = 2.5$, $\lambda_u = 10$. **(h)** $\beta = 3.5$, $\lambda_u = 10$. **(i)** Multipath, $\lambda_u = 10$.

Figure 4.1. MIMO uplink capacity scaling under stochastic geometry.

wireless hop of the network, while a smaller fronthaul is a bottleneck to achieving the full MIMO gains. In addition, for multipath models, i.i.d. small-scale fading across different antennas at the local nodes leads to MIMO gains at higher C_{Σ}, while for the simple LOS models, the channel gains across different antennas at each node are almost identical and provide little diversity gain.

Acknowledgment

This chapter is, in part, a reprint of parts of the paper: Shouvik Ganguly, Seung-Eun Hong, and Young-Han Kim, "On the Capacity Regions of Cloud Radio Access Networks with Limited Orthogonal Fronthaul", *arXiv:1912.04483 [cs.IT], December 2019*, submitted to *IEEE*



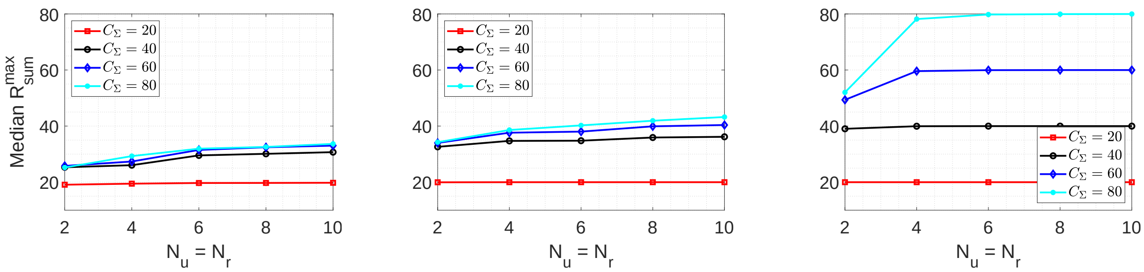
(a) $\beta = 2.5$, $\lambda_r = 2\lambda_u$.
(b) $\beta = 3.5$, $\lambda_r = 2\lambda_u$.
(c) Multipath, $\lambda_r = 2\lambda_u$.

(d) $\beta = 2.5$, $\lambda_r = \lambda_u^2$.
(e) $\beta = 3.5$, $\lambda_r = \lambda_u^2$.
(f) Multipath, $\lambda_r = \lambda_u^2$.

(g) $\beta = 2.5$, $\lambda_u = 10$.
(h) $\beta = 3.5$, $\lambda_u = 10$.
(i) Multipath, $\lambda_u = 10$.

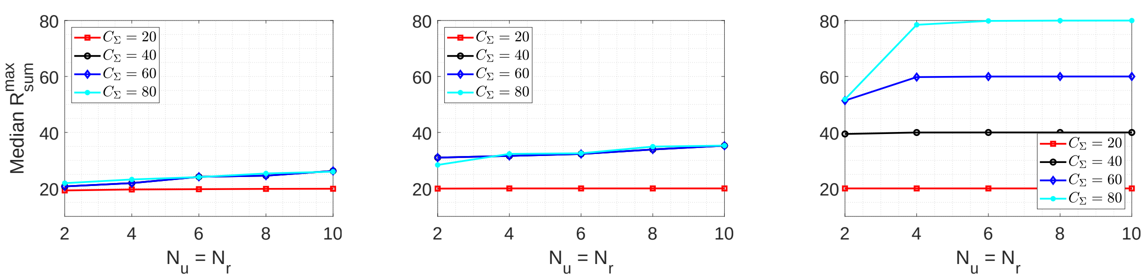
Figure 4.2. MIMO downlink capacity scaling under stochastic geometry.

Transactions on Information Theory. The dissertation author was the primary researcher and author of this material.



(a) $\beta = 2.5$, $K = 4, L = 6$. **(b)** $\beta = 3.5$, $K = 4, L = 6$. **(c)** Multipath, $K = 4, L = 6$.

Figure 4.3. MIMO uplink capacity scaling with antenna number under fixed sum-fronthaul.



(a) $\beta = 2.5$, $K = 4, L = 6$. **(b)** $\beta = 3.5$, $K = 4, L = 6$. **(c)** Multipath, $K = 4, L = 6$.

Figure 4.4. MIMO downlink capacity scaling with antenna number under fixed sum-fronthaul.

Chapter 5

Towards Practical Codes: Lego-brick Approach

In this chapter, codes are developed for two-user multiple access and broadcast channels starting from Gelfand–Pinsker codes with known block lengths, rates, and error performances. Guarantees are provided on the block error rates of the MAC and BC codes in terms of the parameters of the constituent Gelfand–Pinsker codes. These guarantees hold as long as the constituent codes satisfy the assumed properties on rate, codeword weights, and performances, irrespective of the basic structure and other properties.

5.1 Introduction

The channel coding problem has been studied extensively ever since Shannon, in his seminal 1948 paper [49], modeled a point-to-point communication channel as a collection of conditional probability distributions. The point-to-point channel coding theorem was established, among others, by Shannon [49] and Gallager [20]. In a parallel direction of research, the practical problem of coding for point-to-point channels has seen enormous advances in recent years, as alluded to in Chapter 1.

This chapter attempts to answer the following question: what happens when a code, whose performance is known in some setting through simulations or theoretical studies, is used for a different problem? More specifically, given point-to-point channel codes with certain known parameters such as block length n, rate R, and block error probability ε, we attempt to come up with coding schemes for multi-user channels whose performance can be directly obtained as a simple function of parameters of the original code, without redoing extensive studies in the new setting. In essence, we treat encoders and decoders of known codes as "black boxes" (or "Lego-bricks") satisfying some primitive properties and assemble them (potentially with other simple "bricks" such as interleavers or dithers) to build a bigger "box" for a different, and potentially more complicated, scenario, with performance guarantees. Such a theory enables one to leverage commercial off-the-shelf codes (such as those studied in [4, 6, 19, 27, 45]) for single user channels, or even hypothetical codes to be invented in future, to build codes for multi-user communication.

What are the minimum primitive properties these "Lego-bricks" should satisfy while being versatile in building various network communication codes? Given such "Lego-bricks", how do we assemble them in different network communication scenarios? How does the performance guarantee translate between different communication settings? These questions were studied between channel coding and Slepian–Wolf coding first by Wyner for binary symmetric channels and doubly symmetric binary source [64] and later for general binary-input channels and general Slepian–Wolf problems [60, 61]. In this chapter, we propose another "Lego-brick", which can build Gelfand–Pinsker codes for channels with state, channel codes for (asymmetric) point-to-point channels, and Marton coding for broadcast channels, among others. The focus of the chapter is on how the performance of one code in a certain communication setting can be translated into the performance of another code in a different setting.

We start out with primitive Gelfand–Pinsker (GP) codes [21] for binary-input, binary-state channels and construct codes for binary-input multiple access channels (MAC) and finite-alphabet broadcast channels (BC). In addition to primitive GP codes, we use a random interleaver that applies to a length-n sequence, a permutation chosen uniformly at random from the $n!$ possible permutations, as well as shared random bits between transmitters and receivers.

The rest of the chapter is organized as follows. Section 5.2.1 introduces the primitive GP encoding and decoding blocks we use throughout the paper and establishes the performance of this code when a random interleaver is used on the output of the encoder and additional random bits are shared between the encoder and the decoder. Section 5.2.2 shows how to construct an ordinary point-to-point channel code from the primitive GP code. Section 5.3 develops a coding scheme for the 2-user binary-input MAC using two channel codes (ultimately derived from primitive GP codes). Section 5.4 develops a coding scheme for the 2-user BC using a primitive GP code and a channel code. Throughout the chapter, we follow the notation in [17], with the exception that for a natural number n, we use $[n]$ to denote the set $\{1,\ldots,n\}$. In addition, $|x^n| := |\{i \in [n] : x_i = 1\}|$ denotes the Hamming weight of a binary sequence $x^n \in \{0,1\}^n$ and for two binary sequences x^n, y^n, we denote by $x^n \oplus y^n := \{z^n : z_i = x_i \oplus y_i, i \in [n]\}$ the bitwise XOR operation or equivalently, binary addition without carry.

5.2 Gelfand–Pinsker codes to channel codes

5.2.1 Gelfand–Pinsker coding

A Gelfand–Pinsker problem $p(y\,|\,u,s)p(s)$ consists of finite alphabets $\mathcal{U} = \mathcal{S} = \{0,1\}$ and $\mathcal{Y}$, a collection of conditional probability mass functions $p(y,s\,|\,u)$ on $\mathcal{Y} \times \mathcal{S}$ for $u \in \mathcal{U}$ (referred to as the "channel" with input u and state s), and a probability mass function $p(s)$ on $\mathcal{S}$.

A $(R,n,\alpha,\varepsilon,\delta)$ code (g,ψ) depicted in Fig. 5.1 for the Gelfand–Pinsker problem $p(y\,|\,u,s)p(s)$ consists of

- an encoder $g : [2^{nR}] \times \mathcal{S}^n \to \mathcal{U}^n$ that maps each message m and each state sequence s^n to a codeword $u^n = g(m,s^n)$ such that $|g(m,s^n) \oplus s^n| = n\alpha$ for every $m \in [2^{nR}]$, $s^n \in \mathcal{S}^n$,

- a decoder $\psi : \mathcal{Y}^n \to [2^{nR}]$ that assigns a message estimate $\hat{m} = \psi(y^n)$ to each received sequence y^n.

The average probability of error of the code with a *perturbed input* is defined as

$$P_e^{(n)}(z^n) := \sum_m \sum_{s^n} \left(2^{-nR} \prod_{i=1}^{n} p_S(s_i) \, \mathsf{P}(\hat{M} \neq m \,|\, U^n = g(m, s^n) \oplus z^n, S^n = s^n) \right)$$

for $z^n \in \{0,1\}^n$. The *maximal average probability of error under sublinear perturbation* is

$$\max_{\substack{z^n \in \{0,1\}^n: \\ |z^n| \le n^{1/2+\delta}}} P_e^{(n)}(z^n) = \varepsilon. \tag{5.1}$$

The condition (5.1) states that the message is decoded correctly w.h.p. as long as the Hamming weight of the perturbation z^n is not larger than $n^{1/2+\delta}$. This condition is motivated by the existence of practical codes with low decoding complexity and large block lengths n, such as Reed–Muller codes [34, 44] and BCH codes [9], for which the minimum distance can be made to grow as $n^{1/2+\delta}$ or faster by choosing code parameters appropriately.

Remark 5.2.1. For the rest of this chapter, we assume that such a code satisfying (5.1) exists for every channel $p(y \,|\, u, s)$, every $\alpha \in (0,1)$, and every $\varepsilon > 0$, however small, for some large-enough block length n and some $\delta \in (0, 1/2)$.

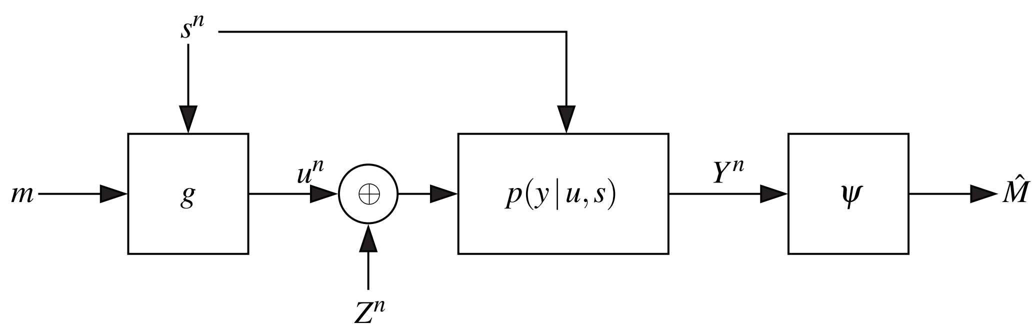


Figure 5.1. Primitive GP code.

We now adapt the primitive Gelfand–Pinsker code to a form that is more useful in coding for multi-user channels. Specifically, we add a random dither $W^n \sim$ i.i.d. Bern$(1/2)$ to the codeword as well as to the observed state sequence, and apply a uniform interleaver Γ_n (i.e., a permutation of n objects chosen uniformly at random) to both of them. To compensate for the interleaver, we apply the operation Γ_n^{-1} to the channel output Y^n and try to decode the message M assuming W^n is also available at the decoder. This arrangement is shown in Fig. 5.2. The following result connects the probability of error of this scheme to that of a slightly different Gelfand–Pinsker problem.

Lemma 5.2.1. *Consider a Gelfand–Pinsker problem $q(y,w\,|\,u,s)p(s)$, where the channel q is obtained by adding a* random dither $W^n \sim$ i.i.d. Bern$(1/2)$ *to the input U^n as well as the state S^n of the channel $p(y\,|\,u,s)$, and taking (Y^n,W^n) as the output, as shown in Fig. 5.2. Suppose that we have a $(R,n,\alpha,\varepsilon,\delta)$ code (g,ψ) for the problem $q(y,w\,|\,u,s)p(s)$. Then,*

$$\sum_{m,s^n}\left(\prod_{i=1}^{n} p_S(s_i) 2^{-nR}\cdot P\Big(\psi(\Gamma_n^{-1}(Y^n),W^n)\neq m \Big| U^n=\Gamma_n(g(m,s^n)),S^n=\Gamma_n(s^n)\Big)\right)\leq\varepsilon.$$

Lemma 5.2.1 illustrates that the arrangement in Fig. 5.2 can achieve the same rate and probability of error as a usual primitive Gelfand–Pinsker code adapted to the channel $q(y,w\,|\,u,s)\equiv p(w)p_{Y|U,S}(y\,|\,u\oplus w,s\oplus w)$.

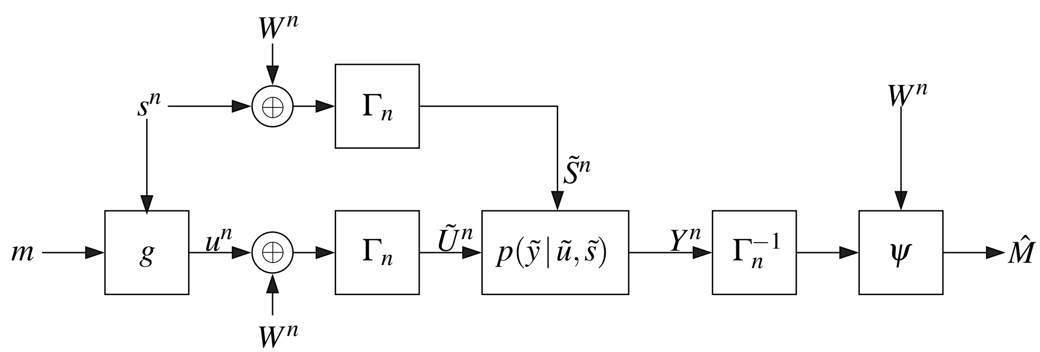


Figure 5.2. GP code with interleavers and dithers.

5.2.2 Point-to-point channel coding

The Gelfand–Pinsker code described in Section 5.2.1 can be easily converted into a point-to-point channel code, as described in this section. Similar to [61], we define a binary-input discrete memoryless channel $p(y\,|\,u)$ as consisting of an input alphabet $\mathcal{U}=\{0,1\}$, a finite output alphabet $\mathcal{Y}$, and a collection of conditional probability mass functions $p(y|u)$ on $\mathcal{Y}$ for $u \in \mathcal{U}$. A (R,n,ε) code (f,ϕ) for the channel $p(y|u)$ consists of

- an encoder $f : [2^{nR}] \to \mathcal{U}^n$ that maps each message m to a codeword $u^n = f(m)$,
- a decoder $\phi : \mathcal{Y}^n \to [2^{nR}]$ that assigns a message estimate $\hat{m} = \phi(y^n)$ to each received sequence y^n.

The average probability of error of this code is

$$\sum_m 2^{-nR} \cdot \mathsf{P}(\phi(Y^n) \neq m | U^n = f(m)) = \varepsilon.$$

Now, suppose we have a $(R,n,\alpha,\varepsilon,\delta)$ code (g,ψ) for the Gelfand–Pinsker problem $p(y|u,s)p(s)$, where $p(y|u,s) \equiv p(y|u)$ (i.e., the channel output is independent of the state given the channel input) and $p_S(0) = 1$. Define $f : [2^{nR}] \to \mathcal{U}^n$ by $f(m) = g(m,\mathbf{0})$, where $\mathbf{0}$ is the all-zero sequence. Then, (f,ψ) forms a code for the channel $p(y|u)$ with length n and rate R, and has average probability of error

$$\sum_m 2^{-nR} \cdot \mathsf{P}(\psi(Y^n) \neq m | U^n = f(m)) = \sum_m 2^{-nR} \cdot \mathsf{P}(\psi(Y^n) \neq m | U^n = g(m,\mathbf{0})).$$

We write

$$
\begin{aligned}
&\mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,\mathbf{0})) \\
&= \mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,\mathbf{0}), S^n = \mathbf{0}) \prod_{i=1}^{n} p_S(0) \\
&\qquad\qquad + \sum_{s^n \neq \mathbf{0}} \mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,\mathbf{0}), S^n = s^n) \prod_{i=1}^{n} p_S(s_i) \\
&\overset{(a)}{=} \mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,\mathbf{0}), S^n = \mathbf{0}) \\
&\overset{(b)}{=} \mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,\mathbf{0}), S^n = \mathbf{0}) \prod_{i=1}^{n} p_S(0) \\
&\qquad\qquad + \sum_{s^n \neq \mathbf{0}} \mathsf{P}(\psi(Y^n) \neq m \,|\, U^n = g(m,s^n), S^n = s^n) \prod_{i=1}^{n} p_S(s_i) \\
&\leq \varepsilon,
\end{aligned}
$$

where (a) and (b) follow since $\prod_{i=1}^{n} p_S(s_i) = 0$ for $s^n \neq \mathbf{0}$, and the last step follows from condition (5.1). Thus, we have obtained a (R,n,ε') channel code from a $(R,n,\alpha,\varepsilon,\delta)$ Gelfand–Pinsker code, where $\varepsilon' \leq \varepsilon$. Similar to Lemma 5.2.1, the following result adapts the Gelfand–Pinsker code to a channel with dithered and permuted input.

Lemma 5.2.2. *Consider a Gelfand–Pinsker problem $q(y,w\,|\,u,s)p(s)$, where $p_S(0) = 1$ and the channel $q(y,w\,|\,u,s) \equiv q(y,w\,|\,u)$ is obtained by adding a* random dither $W^n \sim$ i.i.d. Bern$(1/2)$ *to the input U^n to the channel $p(y\,|\,u)$ and taking (Y^n,W^n) as the output, as shown in Fig. Consider a $(R,n,\alpha,\varepsilon,\delta)$ code (g,ψ) for the problem $q(y,w\,|\,u,s)p(s)$. Then,*

$$
\sum_{m} \left(2^{-nR} \cdot P\Big(\psi(\Gamma_n^{-1}(Y^n), W^n) \neq m \Big| U^n = \Gamma_n(g(m,\mathbf{0}))\Big) \right) \leq \varepsilon.
$$

Remark 5.2.2. While this approach of coding for a point-to-point channel by using a Gelfand–Pinsker code may seem redundant since a multitude of good codes already exist for several different classes of channels, this enables us to code for multi-user channels by using only Gelfand–Pinsker blocks, rather than adding separate channel coding blocks as and when we need

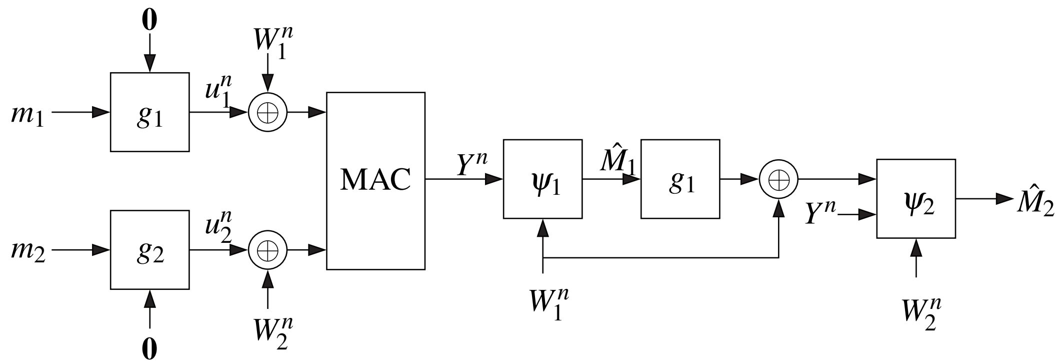


Figure 5.3. Coding for MAC using primitive GP codes.

it. This aligns with the spirit of this work, which aims to eventually code for complex channels, starting out with a minimal number of elementary blocks.

5.3 Coding for two-user MAC

In this section, we put together two channel codes, derived, as described in Section 5.2.2, from Gelfand–Pinsker codes, to build a code for a binary-input discrete memoryless multiple-access channel (DM-MAC) $p(y\,|\,u_1,u_2)$, defined as consisting of input alphabets $\mathcal{U}_1 = \mathcal{U}_2 = \{0,1\}$, a finite output alphabet $\mathcal{Y}$, and a collection of conditional probability mass functions $p(y\,|\,u_1,u_2)$ on $\mathcal{Y}$ for $(u_1,u_2) \in \mathcal{U}_1 \times \mathcal{U}_2$. A (R_1,R_2,n,ε) code (f_1,f_2,ϕ) for the channel $p(y\,|\,u_1,u_2)$ consists of

- encoders $f_j : [2^{nR_j}] \to \mathcal{U}_j^n$ for $j = 1,2$, that map each message $m_1 \in [2^{nR_1}]$ to a codeword $u_1^n = f_1(m_1)$ and each message $m_2 \in [2^{nR_2}]$ to a codeword $u_2^n = f_2(m_2)$,
- a decoder $\phi : \mathcal{Y}^n \to [2^{nR_1}] \times [2^{nR_2}]$ that assigns message estimates $(\hat{m}_1,\hat{m}_2) = \phi(y^n)$ to each received sequence y^n.

The average probability of error of this code is

$$2^{-n(R_1+R_2)} \cdot \sum_{m_1,m_2} \left(\mathsf{P}\left(\phi(Y^n) \neq (m_1,m_2) \middle| (U_1^n,U_2^n) = (f_1(m_1), f_2(m_2)) \right) \right) = \varepsilon.$$

Now, let us add independent dithering sequences $W_1^n, W_2^n \sim$ i.i.d. Bern$(1/2)$ to the codewords U_1^n and U_2^n and make W_1^n, W_2^n available at the decoder. We will construct a code for this modified MAC

$$q(y, w_1, w_2 \,|\, u_1, u_2) := p(w_1)p(w_2)p_{Y|U_1,U_2}(y \,|\, u_1 \oplus w_1, u_2 \oplus w_2)$$

using codes for the point-to-point channels

$$\begin{aligned} q(y, w_1 \,|\, u_1) &= \sum_{u_2 \in \mathcal{U}_2} q(y, w_1 \,|\, u_1, u_2)p(u_2) \\ &= \sum_{\substack{u_2 \in \mathcal{U}_2 \\ w_2 \in \mathcal{W}_2}} \Big(p(w_1)p(w_2)p(u_2) \times p_{Y|U_1,U_2}(y \,|\, u_1 \oplus w_1, u_2 \oplus w_2) \Big) \end{aligned}$$

and

$$q(y, u_1 \oplus w_1, w_2 \,|\, u_2) = p(w_2)p(u_1 \oplus w_1)p_{Y|U_1,U_2}(y \,|\, u_1 \oplus w_1, u_2 \oplus w_2).$$

Suppose that we have a $(R_1, n, \alpha, \varepsilon_1, \delta)$ code (g_1, ψ_1) for the Gelfand–Pinsker problem

$$q(y, w_1 \,|\, u_1, s_1)p(s_1) \equiv q(y, w_1 \,|\, u_1)p(s_1),$$

where $p_{S_1}(0) = 1$. Similarly, let us consider a $(R_2, n, \alpha, \varepsilon_2, \delta)$ code (g_2, ψ_2) for the Gelfand–Pinsker problem

$$q(y, u_1 \oplus w_1, w_2 \,|\, u_2, s_2)p(s_2) \equiv q(y, u_1 \oplus w_1, w_2 \,|\, u_2)p(s_2),$$

where $p_{S_2}(0) = 1$. The following result demonstrates that combining these two codes in the manner shown in Fig. 5.3 yields a $(R_1, R_2, n, \varepsilon')$ code for the MAC $\tilde{p}(y, w_1, w_2 \,|\, u_1, u_2)$, where $\varepsilon' \le \varepsilon_1 + \varepsilon_2$.

Proposition 5.3.1. *Define* $f_1(m_1) := g_1(m_1, \mathbf{0})$, $f_2(m_2) := g_2(m_2, \mathbf{0})$, *and*

$$\phi(y^n, w_1^n, w_2^n) := (\psi_1(y^n, w_1^n), \psi_2(y^n, g_1(\psi_1((y^n, w_1^n))) \oplus w_1^n, w_2^n)),$$

i.e., the decoding function ϕ yields message estimates $\hat{m}_1 = \psi_1(y^n, w_1^n)$ and $\hat{m}_2 = \psi_2(y^n, f_1(\hat{m}_1) \oplus w_1^n, w_2^n)$. Then, the average probability of error is bounded as

$$
\begin{aligned}
\varepsilon' &:= P((\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)) \\
&= 2^{-n(R_1+R_2)} \sum_{m_1,m_2} P\Big(\phi(Y^n, W_1^n, W_2^n) \neq (m_1, m_2) \Big| U_1^n = f_1(m_1), U_2^n = f_2(m_2)\Big) \\
&\leq \varepsilon_1 + \varepsilon_2.
\end{aligned}
$$

Proof sketch. One can show that addition of the dithering sequences ensures that the channel $U_1^n \to (Y^n, W_1^n)$ obtained by averaging over $M_2 \sim \text{Unif}([2^{nR_2}])$ and W_2^n is discrete memoryless and charaterized by the conditional probability distribution

$$\prod_{i=1}^{n} p_{W_1}(w_{1i}) p_{Y|U_1}(y_i \,|\, u_{1i} \oplus w_{1i}). \tag{5.2}$$

This implies that the probability of incorrectly decoding the first message is bounded as

$$\mathsf{P}(\psi_1(Y^n, W_1^n) \neq M_1) \leq \varepsilon_1 \tag{5.3}$$

by our assumption on the $(R_1, n, \alpha, \varepsilon_1, \delta)$ Gelfand–Pinsker code. Similarly, the channel $U_2^n \to (Y^n, U_1^n \oplus W_1^n, W_2^n)$ is discrete memoryless and therefore, if the first message is decoded correctly, the probability of incorrectly decoding the second message is bounded as

$$\mathsf{P}(\psi_2(Y^n, U_1^n \oplus W_1^n, W_2^n) \neq M_2) \leq \varepsilon_2. \tag{5.4}$$

We can now combine (5.3) and (5.4) to bound the average probability of error $\varepsilon' = \tilde{\mathsf{P}}((\hat{M}_1, \hat{M}_2) \neq$

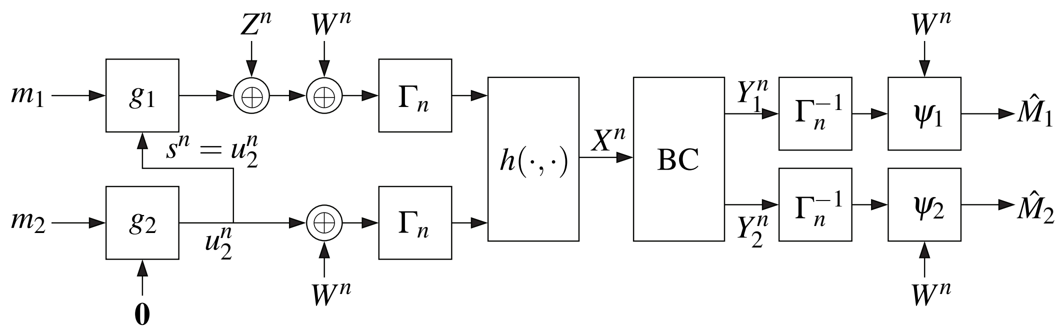


Figure 5.4. Coding for BC using primitive GP codes.

(M_1, M_2)). We write

$$
\begin{aligned}
&\mathsf{P}\left(\{\hat{M}_1 \neq M_1\} \cup \{\hat{M}_2 \neq M_2\}\right) \\
&= \mathsf{P}\left(\{\hat{M}_1 \neq M_1\} \cup \{\hat{M}_1 = M_1, \hat{M}_2 \neq M_2\}\right) \\
&\leq \mathsf{P}(\hat{M}_1 \neq M_1) + \mathsf{P}(\hat{M}_1 = M_1, \hat{M}_2 \neq M_2) \\
&= \mathsf{P}(\psi_1(Y^n, W_1^n) \neq M_1) + \mathsf{P}(\psi_1(Y^n, W_1^n) = M_1, \psi_2(Y^n, f_1(M_1) \oplus W_1^n, W_2^n) \neq M_2) \\
&\leq \mathsf{P}(\psi_1(Y^n, W_1^n) \neq M_1) + \mathsf{P}(\psi_2(Y^n, U_1^n \oplus W_1^n, W_2^n) \neq M_2) \\
&\leq \varepsilon_1 + \varepsilon_2.
\end{aligned}
$$

□

Remark 5.3.1. The coding scheme used here corresponds to the well-known *successive cancellation decoding* [17, Chapter 4.5.1], where the message of user 1 is decoded first and the corresponding message estimate is used to decode the message of user 2. One can also implement the decoding order $2 \to 1$ to achieve a different rate pair for the same MAC.

5.4 Coding for two-user BC

In this section, we put together a Gelfand–Pinsker code and another channel code derived from a Gelfand–Pinsker code, to build a code for a discrete memoryless broadcast channel (DM-BC) $p(y_1, y_2 | x)$, defined as consisting of a finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}_1$, $\mathcal{Y}_2$, and a collection of conditional probability mass functions $p(y_1, y_2 | x)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$ for

$x \in \mathcal{X}$. A $(R_1, R_2, n, \varepsilon)$ code (f, ξ_1, ξ_2) for the channel $p(y_1, y_2 | x)$ consists of

- an encoder $f : [2^{nR_1}] \times [2^{nR_2}] \to \mathcal{X}$ that maps each message pair (m_1, m_2) to a codeword $x^n = f(m_1, m_2)$,
- decoders $\xi_j : \mathcal{Y}_j^n \to [2^{nR_j}]$ for $j = 1, 2$, that assign message estimates $\hat{m}_1 = \xi_1(y_1^n)$ and $\hat{m}_2 = \xi_2(y_2^n)$ to received sequences y_1^n and y_2^n, respectively.

The average probability of error of this code is

$$2^{-n(R_1+R_2)} \sum_{m_1, m_2} \mathsf{P}\left(\{\xi_1(Y_1^n) \neq m_1\} \cup \{\xi_2(Y_2^n) \neq m_2\} \middle| X^n = f(m_1, m_2) \right) = \varepsilon.$$

Now, let us implement the Marton coding scheme for the broadcast channel using primitive GP codes. Take $\mathcal{U}_1 = \mathcal{U}_2 = \{0, 1\}$ and fix a mapping $h : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$. We will use $u_1^n \in \mathcal{U}_1^n$ and $u_2^n \in \mathcal{U}_2^n$ to carry the messages m_1 and m_2, respectively. We will use a (R_2, n, ε_2) channel code (f_2, ϕ_2) for carrying the message m_2 and a $(R_1, n, \alpha, \varepsilon_1, \delta)$ Gelfand–Pinsker code (g_1, ψ_1) for carrying the message m_1, using the codeword carrying m_2 as the state sequence known to the encoder, as in Marton coding [17, Chapter 8.3]. Analogous to Remark 5.3.1, we could also have flipped the *encoding order* and used a Gelfand–Pinsker code for m_2 and a channel code for m_1. In this section, we will finally add a perturbative noise Z^n to the codeword u_1^n and make use of condition (5.1). Z^n is generated as follows. Let $\mathcal{I}_0 \subseteq [n]$ and $\mathcal{I}_1 \subseteq [n]$ be the indices of the positions where 0 and 1, respectively, occur in $u_1^n \oplus u_2^n$, i.e., $\mathcal{I}_0 = \{i \in [n] : u_{1i} \oplus u_{2i} = 0\}$ and $\mathcal{I}_1 = \{i \in [n] : u_{1i} \oplus u_{2i} = 1\}$. Denote the (sorted) indices in $\mathcal{I}_0$ by $j_1, j_2, \ldots, j_{n(1-\alpha)}$, and the sorted indices in $\mathcal{I}_1$ by $l_1, \ldots, l_{n\alpha}$. Let Q be a $\mathrm{Binom}(n, \alpha)$ random variable. If $Q = k$ for $n\alpha < k \leq n$, we choose Z^n to have 1s at the positions $j_1, \ldots, j_{k-n\alpha}$, and 0s everywhere else. If $Q = k$ for $0 \leq k < n\alpha$, we choose Z^n to have 1s at the positions $l_1, \ldots, l_{n\alpha-k}$, and 0s everywhere else. Finally, if $Q = n\alpha$, we take $Z^n = \mathbf{0}$. It can be shown that with this choice of Z^n, $|Z^n \oplus u^n(m, \tilde{s}^n) \oplus \tilde{s}^n|$ is distributed as $\mathrm{Binom}(n, \alpha)$. The perturbative noise Z^n is crucial to introducing correlation among the codewords carrying the two messages. We will also add the same dithering sequence $W^n \sim$ i.i.d. Bern$(1/2)$ to each of $(u_1^n \oplus Z^n)$ and u_2^n and make W^n available to both decoders as common randomness. We will then apply a random permutation $\Gamma_n(\cdot)$ on the sequences $(u_1^n \oplus Z^n \oplus W^n)$ and $u_2^n \oplus W^n$, similar to Lemma 5.2.1, and finally, generate the transmitted codeword X^n as

$$X_i = h(\Gamma_n(u_{1i} \oplus Z_i \oplus W_i), \Gamma_n(u_{2i} \oplus W_i)), \quad i \in [n].$$

The arrangement is put together as shown in Fig. 5.4. We note here that the Gelfand–Pinsker code used is for the effective channel $q_1(y_1, w \mid u_1, u_2)$, obtained by adding a random dither $W \sim$ Bern$(1/2)$ to the input U_1 and state U_2 of the channel $p(y_1 \mid u_1, u_2)$ defined using the BC $p(y_1, y_2 \mid x)$ and the map $x = h(u_1, u_2)$. Similarly, the channel code used is for the channel $q_2(y_2, w \mid u_2)$ obtained by adding W to the input U_2 of the channel $p(y_2 \mid u_2)$.

Lemma 5.4.1. *For the arrangement shown in Fig. 5.4,*

$$(\Gamma_n(g_1(m_1, f_2(m_2)) \oplus Z^n \oplus W^n), \Gamma_n(f_2(m_2) \oplus W^n)) \sim \text{i.i.d. DSBS}(\alpha)$$

for every $(m_1, m_2) \in [2^{nR_1}] \times [2^{nR_2}]$.

The following result demonstrates that combining the Gelfand–Pinsker code and the point-to-point channel code as shown in Fig. 5.4 yields a $(R_1, R_2, n, \varepsilon')$ code for the BC $p(y_1, y_2, w \mid x)$, where $\varepsilon' \leq \varepsilon_1 + \varepsilon_2 + 2e^{-2n^{2\delta}}$.

Proposition 5.4.1. *For the coding scheme depicted in Fig. 5.4, the average probability of error* *is bounded as*

$$\begin{aligned}
\varepsilon' &:= P((\hat{M}_1,\hat{M}_2) \neq (M_1,M_2)) \\
&= 2^{-n(R_1+R_2)} \sum_{m_1,m_2} P\Big(\{\psi_1(\Gamma_n^{-1}(Y_1^n),W^n) \neq m_1\} \cup \{\phi_2(\Gamma_n^{-1}(Y_2^n),W^n) \neq m_2\} \Big| \\
&\qquad\qquad U_1^n = g_1(m_1,f_2(m_2)), U_2^n = f_2(m_2)\Big) \\
&\leq \varepsilon_1 + \varepsilon_2 + 2e^{-2n^{2\delta}}.
\end{aligned}$$

Proof sketch. We first note that similar to Section 5.3, the channel $U_2^n \to (Y_2^n, W^n)$ obtained by averaging over $M_1 \sim \text{Unif}([2^{nR_1}])$ and Z^n, as well as the channel $U_1^n \to (Y_1^n, W^n)$ obtained by averaging over Z^n, is discrete memoryless. Therefore, we have

$$\begin{aligned}
&\mathsf{P}(\psi_1(\Gamma_n^{-1}(Y_1^n),W^n) \neq M_1) \\
&= \sum_{|z^n| \leq n^{1/2+\delta}} p(z^n)\, \mathsf{P}\left(\psi_1(\Gamma_n^{-1}(Y_1^n),W^n) \neq M_1 \middle| Z^n = z^n\right) \\
&+ \sum_{|z^n| > n^{1/2+\delta}} p(z^n)\, \mathsf{P}\left(\psi_1(\Gamma_n^{-1}(Y_1^n),W^n) \neq M_1 \middle| Z^n = z^n\right) \\
&\overset{(a)}{\leq} \varepsilon_1 \mathsf{P}\left(|Z^n| \leq n^{1/2+\delta}\right) + \mathsf{P}\left(|Z^n| > n^{1/2+\delta}\right) \\
&\leq \varepsilon_1 + \mathsf{P}(|Q - n\alpha| > n \cdot n^{-1/2+\delta}) \\
&\overset{(b)}{\leq} \varepsilon_1 + 2e^{-2n\cdot(n^{-1/2+\delta})^2} = \varepsilon_1 + 2e^{-2n^{2\delta}},
\end{aligned}$$

where in (a), we use the error probability bound (5.1) for the $(R_1, n, \alpha, \varepsilon_1, \delta)$ Gelfand–Pinsker code and in (b), we use the Hoeffding bound [22] $\mathsf{P}(|Q - n\alpha| > n\beta) \leq 2\exp(-2n\beta^2)$ for $Q \sim \text{Binom}(n,\alpha)$. Similarly, by the memorylessness of $U_2^n \to (Y_2^n, W^n)$ and the assumed property of the (R_2, n, ε_2) channel code, we conclude that $\mathsf{P}(\phi_2(\Gamma_n^{-1}(Y_2^n),W^n) \neq M_2) \leq \varepsilon_2$. The result is then established by the union bound. □

5.5 Discussions

In this chapter, we have looked at how the performance of a code in a point-to-point setting translates to other scenarios, and how multiple point-to-point codes with certain performance guarantees can be combined to code for multi-user problems. The main focus and guiding principle of this chapter has been the abstraction of complex coding problems into a small number of basic blocks which can be easily realized and in fact, already exist in coding theory and communication theory literature. Combining these practical blocks according to our prescriptions would enable one to code for more complex networks with finite-blocklength performance guarantees.

The next step in this line of work is to generalize these results to multi-hop scenarios including C-RANs. The C-RAN model studied in Chapters 2–4 is one of the simplest 2-hop network models, where the main additional step is compressing the received signals (for uplink) and conveying the precoded signals to the relays (for downlink). One way to handle this is by mapping a K-user L-relay C-RAN problem into a $(K+L)$-user single-hop (MAC or BC) problem. We are currently exploring this line of research.

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Shouvik Ganguly, Lele Wang, and Young-Han Kim, "A Functional Construction of Codes for Multiple Access and Broadcast Channels", accepted for publication at the *IEEE International Symposium on Information Theory*, Los Angeles, USA, June 2020. The dissertation author was the primary researcher and author of this material.

Chapter 6

Concluding Remarks

We conclude this dissertation with comments for future research directions.

In Chapter 2, we developed coding schemes as well as outer bounds for C-RANs and thereby approximated the capacities of uplink and downlink C-RANs. Chapter 3 looks at C-RAN capacities from a slightly different direction and examines the large-network asymptotics under sum-fronthaul constraints, comparing the scaling laws with those of the infinite-fronthaul case. While these results are interesting in that they address the fundamental limits of information flow in C-RANs, a more relevant question for practicing engineers is how these limits can be approached in a real-world communication scenario. To this end, Chapter 5 studies approaches to combine point-to-point channel codes with other basic blocks and develop finite-blocklength, low complexity codes for multiuser networks. This is an ongoing process; I wish to take this forward and come up with tractable codes for C-RANs as well as more complex networks, with performance guarantees approaching those predicted by information theory.

Bibliography

[1] G. T. 36.873, "Study on 3D channel model for LTE (Release 12)," 3rd Generation Partnership Project (3GPP), Tech. Rep. 36.873 (V12.7.0), Dec. 2017.

[2] I. E. Aguerri and A. Zaidi, "Lossy compression for compute-and-forward in limited backhaul uplink multicell processing," *IEEE Trans. Communications*, vol. 64, no. 12, pp. 5227–5238, Dec 2016.

[3] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai, "On the capacity of cloud radio access networks with oblivious relaying," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4575–4596, Jul 2019.

[4] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[5] F. Baccelli, A. El Gamal, and D. N. C. Tse, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, May 2011.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *IEEE International Conference on Communications*, May 1993, pp. 1064–1070.

[7] S. S. Bidokhti and G. Kramer, "Capacity bounds for diamond networks with an orthogonal broadcast channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7103–7122, Dec 2016.

[8] S. S. Bidokhti, G. Kramer, and S. Shamai, "Capacity bounds on the downlink of symmetric, multi-relay, single receiver c-ran networks," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017, pp. 2058–2062.

[9] R. Bose and D. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68 – 79, 1960.

[10] G. Caire and S. Shamai, "On achievable rates in a multi-antenna broadcast downlink," in *Proc. 38th Ann. Allerton Conf. Comm. Control Comput.*, Oct. 2000.

[11] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[12] T. M. Cover and B. Gopinath, Eds., *Open Problems in Communication and Computation*. Springer-Verlag, 1987.

[13] J. Edmonds, *Submodular Functions, Matroids, and Certain Polyhedra*. New York, NY, USA: Gordon and Breach, 1970.

[14] I. El Bakouri and B. Nazer, "Integer-forcing architectures for uplink cloud radio access networks," in *Proc. 55th Ann. Allerton Conf. Comm. Control Comput.*, Oct. 2017, pp. 67–75.

[15] ——, "Uplink–downlink duality for integer-forcing in cloud radio access networks," in *Proc. 56th Ann. Allerton Conf. Comm. Control Comput.*, Oct. 2018, pp. 39–47.

[16] A. El Gamal and T. M. Cover, "Multiple user information theory," *Proc. IEEE*, vol. 68, no. 12, pp. 1466–1483, Dec. 1980.

[17] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[18] A. El Gamal, "On information flow in relay networks," in *IEEE Nat. Telecom Conf.*, vol. 2, November 1981, pp. D4.1.1–D4.1.4.

[19] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[20] ——, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[21] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf.Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[22] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Am. Stat. Assoc*, vol. 58, no. 301, pp. 13–30, 1963.

[23] S. N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sept 2013.

[24] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Downlink macro-diversity in cellular networks," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2007.

[25] W. Kang and N. Liu, "The Gaussian multiple access diamond channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 1499–1503.

[26] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept 2005.

[27] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[28] S. H. Lim, K. T. Kim, and Y.-H. Kim, “Distributed decode–forward for relay networks,” *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4103–4118, Jul 2017.

[29] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, “Noisy network coding,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[30] L. Liu, P. Patil, and W. Yu, “An uplink–downlink duality for cloud radio access network,” in *Proc. IEEE Int. Symp. Inf. Theory*, July 2016, pp. 1606–1610.

[31] L. Liu and W. Yu, “Joint sparse beamforming and network coding for downlink multi-hop cloud radio access networks,” December 2016.

[32] N. Liu and W. Kang, “A new achievability scheme for downlink multicell processing with finite backhaul capacity,” in *Proc. IEEE Int. Symp. Inf. Theory*, June 2014, pp. 1006–1010.

[33] W. Lu and M. Di Renzo, “Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation,” in *18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* , Nov. 2015, pp. 179–188.

[34] D. E. Muller, “Application of boolean algebra to switching circuit design and to error detection,” *Transactions of the I.R.E. Professional Group on Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, Sep. 1954.

[35] B. Nazer and M. Gastpar, “Compute-and-forward: Harnessing interference through structured codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct 2011.

[36] L. H. Ozarow, “On a source-coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[37] L. H. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[38] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, “Joint decompression and decoding for cloud radio access networks,” *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 503–506, May 2013.

[39] ——, “Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks,” *IEEE Trans. Signal Processing*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.

[40] ——, “Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory,” *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[41] P. Patil and W. Yu, “Generalized compression strategy for the downlink cloud radio access network,” Jan 2018.

[42] M. Peng, C. Wang, V. Lau, and H. V. Poor, “Fronthaul-constrained cloud radio access networks: Insights and challenges,” *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[43] T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, Eds., *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge Univ. Press, March 2017.

[44] I. Reed, “A class of multiple-error-correcting codes and the decoding scheme,” *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 38–49, Sep. 1954.

[45] T. J. Richardson and R. L. Urbanke, “The capacity of low-density parity-check codes under message-passing decoding,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[46] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, “Communication via decentralized processing,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, July 2008.

[47] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai, “Uplink macro diversity of limited backhaul cellular network,” *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457–3478, Aug 2009.

[48] S. Shamai and B. M. Zaidel, “Enhancing the cellular downlink capacity via co-processing at the transmitting end,” in *Proc. IEEE VTS 53rd Vehicular Technology Conf.*, May 2001, pp. 1745–1749.

[49] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[50] J. W. Silverstein, “Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices,” *J. Multivar. Anal.*, vol. 55, pp. 331–339, 1995.

[51] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, “Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems,” *Journal of Communications and Networks*, vol. 18, no. 2, pp. 135–149, April 2016.

[52] O. Simeone, O. Somekh, E. Erkip, H. V. Poor, and S. Shamai, “Robust communication via decentralized processing with unreliable backhaul links,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4187–4201, Jul 2011.

[53] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, “Downlink multicell processing with limited-backhaul capacity,” *EURASIP Journal on Advances in Signal Processing*, June 2009.

[54] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor, and S. Shamai, *Cooperative Wireless Cellular Systems: An Information-Theoretic View*, Aug. 2012, vol. 8, no. 1–2.

[55] O. Somekh, B. M. Zeidel, and S. Shamai, "Spectral efficiency of joint multiple cell-site processors for randomly spread DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2625–2637, Jul. 2007.

[56] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov 1999.

[57] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity–multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.

[58] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1329–1343, Jun. 2002.

[59] C.-Y. Wang, M. Wigger, and A. Zaidi, "On achievability for downlink cloud radio access networks with base station cooperation," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5726–5742, Aug 2018.

[60] L. Wang, "Channel coding techniques for network communication," Ph.D. dissertation, University of California, San Diego, 2015.

[61] L. Wang and Y.-H. Kim, "Linear code duality between channel coding and Slepian–Wolf coding," in *Proc. 53rd Ann. Allerton Conf. Comm. Control Comput.*, Sep.–Oct. 2015, pp. 147–152.

[62] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM Journal of Research and Development*, vol. 4, no. 1, pp. 66–82, Jan. 1960.

[63] X. Wu, L. P. Barnes, and A. Özgür, "The capacity of the relay channel: Solution to Cover's problem in the Gaussian case," *arXiv:1701.02043v4 [cs.IT]*, Oct 2018.

[64] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.

[65] ——, "Shannon theoretic approach to a Gaussian cellular multiple access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.

[66] T. Yang, N. Liu, W. Kang, and S. Shamai, "Converse results for the downlink multicell processing with finite backhaul capacity," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 368–379, Sep. 2018.

[67] W. Yu, "Uplink–downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[68] Y. Zhou, Y. Xu, W. Yu, and J. Chen, "On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7402–7418, Dec 2016.

[69] Y. Zhou and W. Yu, "Optimized backhaul compression for uplink cloud radio access network," *IEEE Journ. Sel. Are. Comm.*, vol. 32, no. 6, pp. 1295–1307, June 2014.