Downlink Transmission Strategies for Cloud Radio-Access Networks

by

Pratik Patil

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

Abstract

Downlink Transmission Strategies for Cloud Radio-Access Networks

Pratik Patil

Master of Applied Science

Graduate Department of Electrical and Computer Engineering

University of Toronto

2015

This thesis studies transmission strategies for the downlink of a cloud radio-access network, in which the base stations are connected to a centralized cloud-computing based processor with digital backhaul links. We provide a system-level performance comparison of two fundamentally different strategies, namely the data-sharing strategy and the compression strategy, that differ in the way the backhaul is utilized. It is observed that the performance of both strategies depends crucially on the available backhaul capacity. When the backhaul capacity is low, the data-sharing strategy performs better, while the compression strategy is superior under moderate-to-high backhaul capacity. Using insights from such a comparison, we propose a novel hybrid strategy, combining the data-sharing and compression strategies, that allows for better control over the backhaul capacity utilization. An optimization framework for the hybrid strategy is proposed. Numerical evidence demonstrates the performance gain of the hybrid strategy.

Acknowledgements

It is my great pleasure to thank the many people who have contributed, both directly and indirectly, to this work. First and foremost, I thank my advisor, Professor Wei Yu, for his advice and encouragement throughout these years. His remarkable intuition and insights have often led me to consider problems from different perspectives. I have benefited a great deal from his curiosity, enthusiasm, and efforts to understand the essence of things. The passion and dedication he puts towards research is always inspiring.

I thank Professor Ravi Adve, Professor Ashish Khisti, and Professor Raymond Kwong for taking the time to read the thesis and provide useful comments and suggestions. I am particularly thankful to Professor Khisti for his valuable guidance and support throughout my studies at the University of Toronto.

I owe many thanks to my friends and colleagues who made these past years an enjoyable learning experience. In particular, I thank Binbin Dai for always being available for cheerful and stimulating discussions while collaborating on various topics. The countless interesting dinner conversations on philosophy and mathematics with Siyu Liu have been a great source of fun and recreation. I am grateful for all the advice and help I have received from Gokul Sridharan, Soroush Tabatabaei, and Yuhan Zhou on numerous occasions. I thank the friends in the Communications Group, Arvin Ayoughi, Ahmed Badr, David Ding, Chen Feng, Siddarth Hari, Kianoush Hosseini, Yicheng Lin, Kaveh Mahdaviani, Rafid Mahmood, Peter Sam Raj, Kaiming Shen, Foad Sohrabi, Louis Tan, Wanyao Zhao, Caiyi Zhu, and the staff for creating a wonderful working environment.

The financial support of MITACS Globalink Graduate Fellowship, and additional support from Huawei Technologies, and Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged.

Finally, I thank my parents for their unwavering support and encouragement through all these years.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The next generation (5G) wireless networks are envisioned with an ultra-dense cellular deployment in order to support the ever increasing demand for high-speed data [1]. As a consequence, intercell interference is the main physical layer bottleneck in future cellular networks. Multicell cooperation is a technique that allows neighboring base stations (BSs) to cooperate with each other for joint precoding and joint processing of user data for intercell interference mitigation [2]. This thesis considers a promising future cellular architecture, the Cloud Radio-Access Network (C-RAN), as an enabling platform on which BSs can cooperate for interference mitigation purposes

In the C-RAN architecture, the BSs are connected to centralized cloud-computing based servers via high-speed digital backhaul (wireline or wireless) links. One of the benefits of the C-RAN architecture is that it provides an ability for flexible allocation of radio and computing resources across all the BSs managed by the same central processor and a cost-effective path for upgrading the existing wireless infrastructure for mobile service delivery [3]. But more importantly, it facilitates coordinated and cooperative signal processing across the multiple BSs connected to the same central processor. Joint encoding of user messages in the downlink and joint decoding of user signals in uplink can be performed at the central processor. By enabling the implementation of network

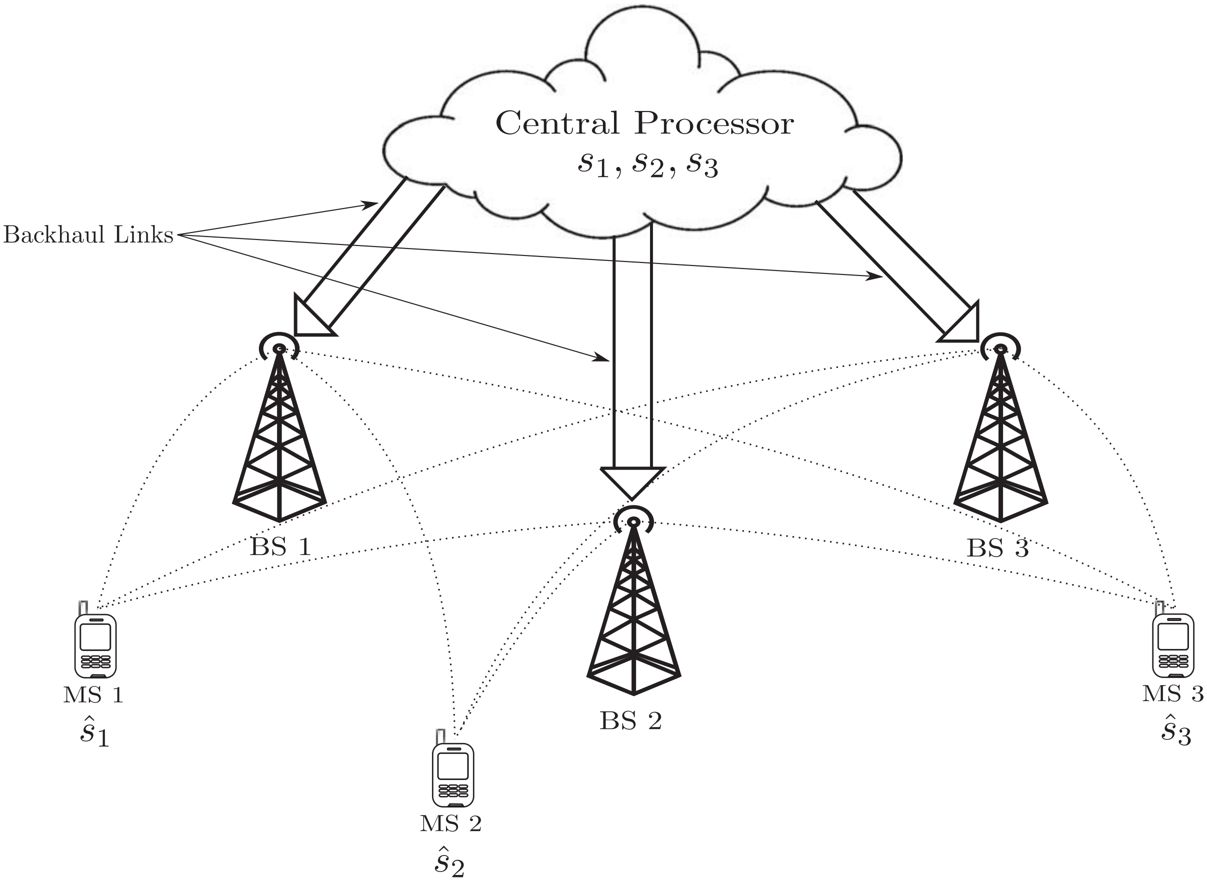

Figure 1.1: Illustration of the downlink transmission in C-RAN.

multiple-input multiple-output (MIMO) or coordinated multi-point (CoMP) concepts [4, 5], C-RAN has the potential to significantly improve the overall throughput of the cellular network.

This thesis studies the downlink transmission in a C-RAN setting. In the downlink C-RAN, as shown in Fig. 1.1, the user data originate from the centralized cloud server and are destined for the mobile devices distributed throughout a geographical area, while the BSs act as *relays* between the user terminals and the cloud. In this sense, the downlink C-RAN can be modeled as a broadcast-relay channel. If the backhaul links between the cloud processor and the BSs have infinite capacities, the capacity analysis for this setting is straightforward, as the downlink C-RAN becomes a vector broadcast channel and the standard network information theoretic results apply [6]. But the limitations on the practical implementation of the C-RAN architecture constrains the backhaul links to have *finite* capacities. In this more realistic case, both the theoretical analysis and the practical system design become much more involved. This thesis studies transmission strategies for the downlink C-RAN with finite backhaul capacities.

There are two fundamentally different existing transmission strategies for the downlink C-RAN, depending on whether the joint precoding operation is performed at the central processor or at the individual BSs. First, this thesis asks the question of how the limited backhaul capacities influence the achievable rates in each strategy, and compares their system-level performance under practical network settings. Second, this thesis proposes a novel hybrid transmission scheme, which allows for better utilization of the finite-capacity backhaul, by combining these two strategies.

The interference mitigation capability of CRAN stems from its ability to jointly encode the user messages across multiple BSs. One way to enable such joint precoding is to simply share each user's message with multiple BSs over the backhaul links. This backhaul transmission strategy, called the *data-sharing* strategy in this thesis, is analogous to the decode-and-forward relaying strategy. As sharing of each user's message across the entire network would require excessively large amount of backhaul capacity, the practical implementation of data-sharing strategy often involves clustering, where each user selects a subset of cooperating BSs and only those BSs in its cooperation cluster receive its message.

As an alternative strategy, the joint precoding of user messages can also be performed at the cloud server rather than at the individual BSs. In fact, one of the the original motivations for C-RAN is to entirely shift the baseband processing from the BSs to the central processor making BS units as simple as possible for easy deployments, upgrades, and maintenance [3]. In this case, the precoded analog signals are compressed and forwarded to the corresponding BSs over the finite-capacity backhaul links for direct transmission by the BS antennas. This approach, called the *compression* strategy in this thesis, is akin to the compress-and-forward relaying strategy. The practical implementation of the compression strategy involves setting the appropriate quantization noise levels under the limited backhaul capacity.

One of the key questions then is, if the functionalities of the BSs should be entirely moved to the central processor, or if there is some benefit to having a functional split between the central processor and the BSs. The main contribution of the thesis is to answer this question.

In the data-sharing strategy the BSs receive clean copies of the user messages. Thus the precise beamformed signals can be computed at the BSs. However, carrying raw user data multiple times consumes high backhaul capacity, hence the cooperation cluster size needs to be small under limited backhaul capacity. On the other hand, in the compression strategy, since the beamformed signals are computed at the central processor, all the available user data can be used to compute the beamformed signals. This allows the possibility of a large cooperation cluster. But the final beamformed signals then need to be compressed, which introduces quantization noises that limit the system performance. Note that the feasibility of the joint cooperative signal processing depends crucially on the availability of the channel state information (CSI) at the BSs and the central processor. In terms of CSI, the data-sharing strategy requires less CSI than the compression strategy due to smaller cluster size in the former.

Individually, both the data-sharing and compression strategies have been studied in the context of C-RAN. However, a fair system-level comparison between the two strategies under practical network settings has not yet been carried out in the literature due to the challenges in solving the corresponding network optimization problems involving user scheduling, beamforming, power control, along with the optimization of clusters for the data-sharing strategy and the optimization of quantization noise levels for the compression strategy. This thesis tackles such a system-level performance evaluation and tries to find the conditions under which one strategy outperforms the other.

The thesis then demonstrates that in a practical C-RAN setting with finite backhaul capacity, instead of individual data-sharing or compression strategies, a hybrid strategy that combines the two can improve the overall system performance. We propose an approach where the central processor directly sends messages for some of the users to the BSs, along with the compressed version of the precoded signals for rest of the users. The intuition behind such an approach is that it is beneficial, in terms of backhaul capacity utilization, to send clean messages for strong users while compressing rest of the interference canceling signals. To quantify the benefit of this hybrid strategy, this thesis proposes an optimization framework to select users for either direct data-sharing or for compression, along with the network-wide beamforming design and the optimization of quantization noise levels for the compressed signals.

1.1 Contributions

The overall contributions of this thesis are as follows.

First, the thesis provides a system-level performance comparison of the data-sharing and compression strategies under finite backhaul capacity and practical network settings. We consider the network-wide optimization frameworks for both strategies to maximize the network utility. We assume explicit per-antenna power constraints and per-BS backhaul constraints. The optimization methodology is based on an equivalence between the weighted sum rate (WSR) maximization and the weighted minimization of sum mean squared error (WMMSE). Specifically,

- We take into account loss due to practical modulation schemes in terms of gap to capacity for both strategies. In addition, for the compression strategy, we introduce a similar notion of gap to rate-distortion limit to account for quantization losses due to non-ideal quantizers used in practice;
- We propose a novel algorithm for the joint optimization of the beamformers and

quantization noise levels for the compression strategy based on the equivalence between the WSR maximization and the WMMSE problem;

- We extend the existing algorithm for the joint optimization of the beamformers and BS cooperation clusters for the data-sharing strategy in [7] to account for per-antenna power constraints and the effect of practical modulation in terms of the gap to capacity factor.

Second, we propose a novel hybrid transmission strategy that combines the data-sharing and compression strategies that allows for better utilization of the limited backhaul capacity. We propose an optimization framework to quantify the performance gains due to the hybrid strategy. Specifically,

- We develop a unified optimization framework that jointly optimizes the network-wide beamformers, user selection for either data-sharing or compression, and the quantization noise levels for the compressed signals. This framework generalizes the frameworks for both the data-sharing and compression strategies.

1.2 Related Work

As pointed before, information theoretically, the downlink of C-RAN is an instance of a broadcast-relay network, where the BSs can be considered as relays. The capacity of such network is unknown. A general coding strategy for the broadcast-relay network is proposed in [8] based on a combination of Marton coding for the general broadcast channel [9] and a coding scheme for deterministic linear relay networks [10]. However, unlike in the uplink of the C-RAN, which is an instance of a multiple-access-relay channel, where compress-and-forward strategies are known to be approximately optimal (in the sense of constant gap to the cutset outer bound), such as quantize-map-forward scheme of [10], or more generally noisy network coding [11], there are no approximate results known on the capacity region for the downlink C-RAN setup. The main difficulty lies in the need for careful coordination among codewords for multiple user messages at the central processor. In the uplink, there was no such need as the central processor decodes all the compressed signals and the original user messages jointly. In the downlink, the central processor can induce coordination among different codewords and the relays potentially need to decode carefully chosen parts of the messages. Recently a new coding scheme that combines Marton coding for single-hop broadcast channels [9] and partial decode-forward for relay channels [12], called distributed decode-forward, is proposed for broadcasting multiple messages over a general relay network in [13]. It is not yet clear how to specialize the proposed scheme to the downlink C-RAN setup as the achievable rate region in the proposed scheme involves auxiliary random variables which are difficult to set properly.

If the backhaul capacity is infinite, downlink C-RAN with a Gaussian channel model reduces to the well-known vector Gaussian broadcast channel, for which dirty paper coding (DPC) achieves the capacity region. For the finite backhaul capacity, however, DPC and other linear precoding schemes cannot be applied directly. In [14], inner bounds for the downlink transmission schemes with different levels of BS cooperation (infinite, limited or no BS cooperation) are studied. The effect of imperfect channel state information (CSI) at the BSs and users is also taken into account.

For compression based strategies, a compressed version of DPC (CDPC) is introduced in [15]. Different transmission strategies that require varying degrees of codebook information (the encoding function information needed to employ DPC) at the BSs are investigated for a simple Wyner type model. We get CDPC, when the BSs are oblivious of any codebook information, where the central processor performs joint DPC, independently compresses the codeword for each BS, and then sends the quantized codeword to the corresponding BS. If some degree of codebook information is available at the BSs, then data-sharing becomes possible. The conclusion of [15] is that oblivious BSs are sufficient in the regime of sufficiently large backhaul capacity for the Wyner model. Recently [16] proposed a multivariate compression strategy across the signals of all the BSs, instead of independent compression for each BS, to better control the effect of resulting total quantization noises at the users by correlating the quantization noises for signals of different BSs. An iterative algorithm achieving a stationary point for the problem of maximizing sum rate with respect to the precoding matrix and the quantization noise covariance matrix is proposed. Our work differs from [16], in that [16] optimizes the covariance matrices of transmit beamformers along with the quantization noise covariance matrix using a rank approximation. In our optimization framework for the compression strategy, we make a novel use of the equivalence between the WSR maximization and the WMMSE problem instead, which does not require any approximation.

For data-sharing based strategies, various ways to selectively share the user messages have been investigated in the literature [17, 18]. Information theoretic results for the downlink network MIMO model using the data-sharing strategy have been reported in [15, 19, 20], but most of these works are limited to certain simplified models. A modified linear Wyner cellular model is studied in [15], and a two-BS, two-user setup is studied in [19]. Our optimization framework for the data-sharing strategy is based on previous work on sparse beamforming in [7]. We extend the algorithm in [7] to account for per-antenna power constraints and the gap to capacity factor.

It is worth pointing out that a third transmission strategy, based on the compute and forward (CoF) strategy for relay networks [21], is proposed in [22] nicknamed reverse compute and forward (RCoF). The roles of BSs and users are reversed in RCoF compared to CoF. Since users do not cooperate, an appropriate invertible precoding is performed to the messages to be sent by the BSs at the central processor such that the effect of linear combination can be undone at the user terminal so that each user obtains just its desired message in the end. But, as with CoF, the performance of RCoF is quite sensitive to the channel coefficients due to the non-integer penalty, since channel coefficients are not exactly matched to the computed integer linear combination. To eliminate this penalty, based on the integer-forcing receiver idea [23], in [22], the effective channel matrix is forced to be an integer matrix by a beamforming strategy named integer-forcing beamforming (IFBF). In IFBF, the precoding matrix is chosen such that the effective channel matrix is an integer matrix, then RCoF is applied to the effective channel matrix with no non-integer penalty. While IFBF removes the non-integer penalty of RCoF, it introduces a signal-to-noise-ratio (SNR) penalty due to the non-unitary precoding matrix. To send the precoded symbols through the limited capacity backhaul links, the central processor forwards the quantized versions to the BSs. The overall scheme is termed compressed IFBF (CIFBF). The main challenge with such coding strategies based on lattice-coding is that the underlying optimization problems often involve integer matrices, which are very difficult to solve in practical networks.

1.3 Organization

The rest of the thesis is organized as follows. Chapter 2 looks at the network-wide optimization for the data-sharing and compression strategies. The optimization frameworks for the two strategies are provided in separate sections and then the numerical system-level performance comparison between the two is made. Chapter 3 proposes the hybrid strategy that combines the data-sharing and compression strategies. We provide a unifying optimization framework for the hybrid strategy that generalizes the data-sharing and compression strategies. Joint optimization of network-wide beamformers, user selection for data-sharing component, and quantization noise optimization for the compressed signal is performed. We then provide system-level numerical evaluation of the hybrid strategy to quantify its performance gains, over the individual data-sharing and compression strategies. Finally, Chapter 4 concludes the thesis outlining the major findings of our study. We also provide some directions for future work.

1.4 Notation

The notation used in this thesis is as follows. *Plain* lower or upper case letters are used to denote scalars, e.g., w, C. Bold face lower letters are used to denote vectors, e.g., $\mathbf{w}$. Bold face upper letters are used to denote matrices, e.g., $\mathbf{H}$. An n-dimensional identity matrix is denoted by either $\mathbf{I}_{n\times n}$, or $\mathbf{I}$ when the dimension is clear from the context. For a scalar, $\mathrm{Re}\{\cdot\}$ denotes its real part and $|\cdot|$ denotes its magnitude. For a vector, $(\cdot)^T$ denotes its transpose, $||\cdot||_p$ denotes its ℓ_p norm. For a matrix, $(\cdot)^{-1}$ denotes its inverse, $(\cdot)^H$ denotes its conjugate transpose (or just conjugate, in case of a scalar). For a random variable, $\mathbb{E}\left[\cdot\right]$ denotes its expected value. Calligraphy letter are used to denote sets, e.g., $\mathcal{L}$. Letters $\mathbb{C}$ and $\mathbb{R}$ are used to denote the set of real and complex numbers respectively.

Chapter 2

Data-sharing versus Compression Strategies

This chapter provides a system-level performance comparison of two fundamentally different transmission strategies, the data-sharing strategy and the compression strategy, for the downlink of a C-RAN. The two strategies differ in the way the limited backhaul is utilized. On one hand, in the data-sharing strategy, the central processor shares the data of each user to a cluster of BSs which then computes the beamformed signals to be transmitted. The backhaul is used to carry raw user data. On the other hand, in the compression strategy, the central processor itself computes the beamformed signals to be transmitted by each BS through capacity-limited backhaul links. The backhaul is used to carry compressed beamformed signals.

Although these strategies have been individually studied in the literature, a fair comparison of the two schemes under practical network settings is challenging because of the complexity in jointly optimizing user scheduling, beamforming, and power control for system-level performance evaluation, along with the need to optimize cooperation clusters for the data-sharing strategy and quantization noise levels for the compression strategy. This chapter presents optimization frameworks to maximize the network utility for both strategies, while taking into account losses due to practical modulation in terms of gap to capacity and due to practical quantization in terms of gap to rate-distortion limit.

Among the family of network utility functions, we adopt the WSR utility. Both optimization frameworks for the data-sharing strategy and the compression strategy exploit the equivalence between the WSR maximization and the WMMSE problem. We point out that, although we consider the WSR as the utility function, the proposed frameworks can easily be extended to any utility function that holds an equivalence with the WMMSE problem. A sufficient condition for any function to hold such equivalence is provided in [24].

Our optimization framework for the data-sharing strategy builds upon [7]. We extend the framework in [7] to include the gap factor for practical modulation and consider more general per-antenna power constraints. For the compression strategy, we consider two models for quantization, depending on whether the codebooks used for compression are kept fixed or are allowed to adapted, and propose a novel algorithm that uses the equivalence between the WSR maximization and the WMMSE problem. We make appropriate distinctions between the cases with single-antenna BSs and multi-antenna BSs, depending on the feasibility of the framework in each of the cases with fixed or adaptive codebooks.

The main conclusion of this chapter is that the compression-based strategy, even with a simple fixed-rate uniform quantizer, outperforms the data-sharing strategy under medium-to-high capacity backhauls. However, the data-sharing strategy outperforms the compression strategy under low capacity backhauls primarily because of the large quantization loss at low backhaul capacity with compression.

This chapter restricts attention to linear precoding strategies, but as mentioned before, possibilities exist for performing nonlinear precoding based on dirty-paper coding [15], and for using the lattice-coding strategy based on compute-and-forward [25] for

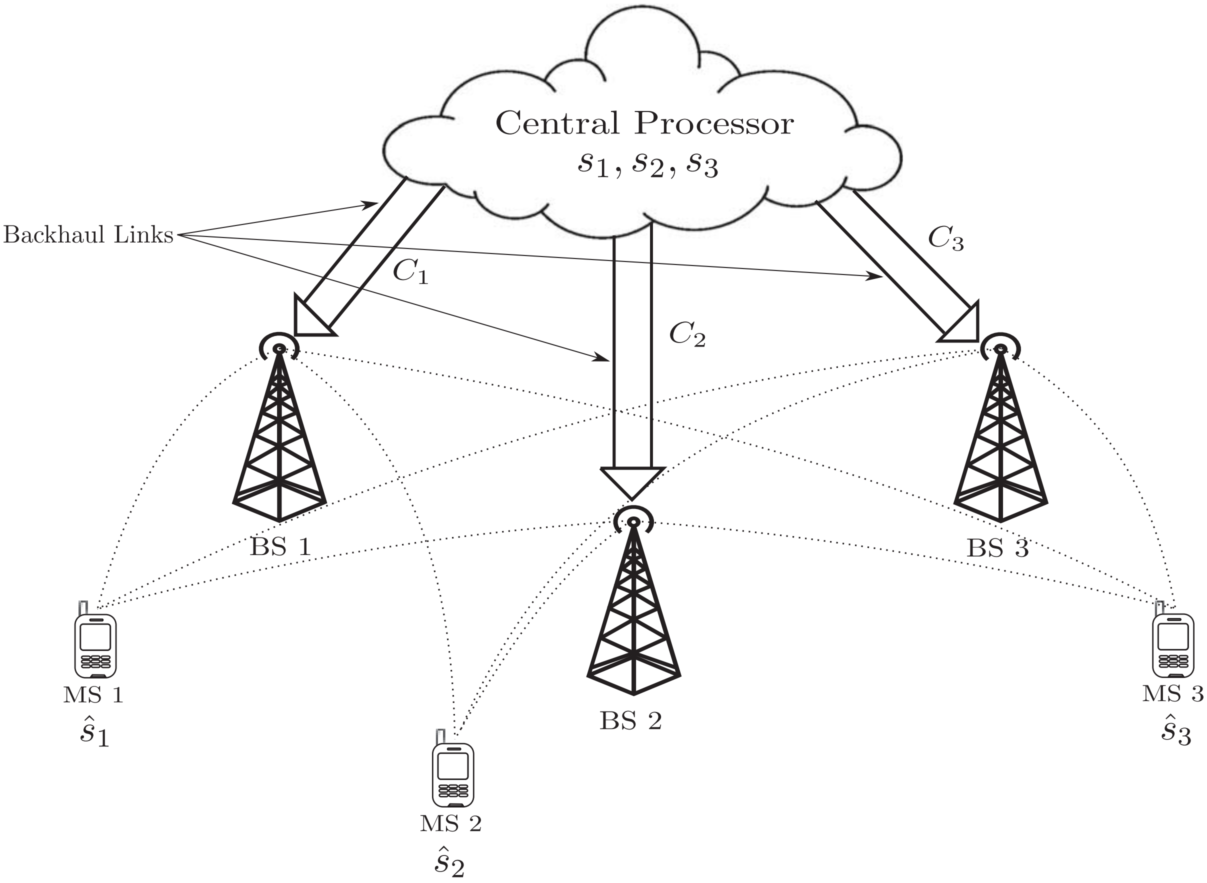

Figure 2.1: Illustration of the C-RAN downlink system.

the downlink C-RAN.

2.1 System Model

Consider the downlink of a C-RAN, as shown in Fig. 2.1, comprising of L BSs equipped with M antennas serving K users equipped with N antennas. All the BSs are connected to a central processor with capacity-limited backhaul links[1]. The capacity of the backhaul link connecting BS l to the central processor is denoted by C_l, $l \in \mathcal{L} = \{1, \ldots, L\}$. We transmit a single independent data stream from the central processor to each user. The user k's information signal is denoted by s_k, $k \in \mathcal{K} = \{1, \ldots, K\}$ and it is assumed to

[1] We use the term backhaul, because the links carry digital data. These links are sometimes referred to as fronthaul links in the C-RAN literature, especially when they carry compressed analog signals.

be chosen independently from a complex Gaussian distribution with zero-mean and unit variance. We assume that the central processor has access to the data and perfect CSI for all the users in the network. The complex signal transmitted by antenna m at BS l is denoted by x_l^m, $m \in \mathcal{M} = \{1, \dots, M\}$, $l \in \mathcal{L}$. We assume a per-antenna transmit power constraint with maximum power budget denoted by P_l^m, i.e.,

$$\mathbb{E}[|x_l^m|^2] \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M}. \tag{2.1}$$

A flat-fading channel model is assumed. Let $\mathbf{x}_l \in \mathbb{C}^{M\times 1} = [x_l^1, \dots, x_l^m]^T$ denote the vector signal transmitted by BS l and $\mathbf{x} \in \mathbb{C}^{LM\times 1} = [\mathbf{x}_1^T, \dots, \mathbf{x}_L^T]^T$ be the aggregate signal from all the BSs. The received signal at user k, $\mathbf{y}_k \in \mathbb{C}^{N\times 1}$, is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{z}_k, \tag{2.2}$$

where $\mathbf{H}_k \in \mathbb{C}^{N\times LM} = [\mathbf{H}_{1,k}, \dots, \mathbf{H}_{L,k}]$ is the channel to user k from all the BSs, $\mathbf{H}_{l,k} \in \mathbb{C}^{N\times M}$ being the channel response from M transmit antennas of BS l to N receive antennas of user k, and $\mathbf{z}_k$ is the additive complex Gaussian noise with zero-mean and variance σ^2 on all of its diagonals.

2.2 Data-sharing Strategy

In the data-sharing strategy, as shown in Fig. 2.2, a cluster of BSs locally form beamformers to cooperatively serve each user. The data for that user is replicated at all the participating BSs in the cluster via the backhaul links. A crucial decision is to select an appropriate cluster of BSs for each user for interference mitigation, while staying under the limited backhaul capacity.

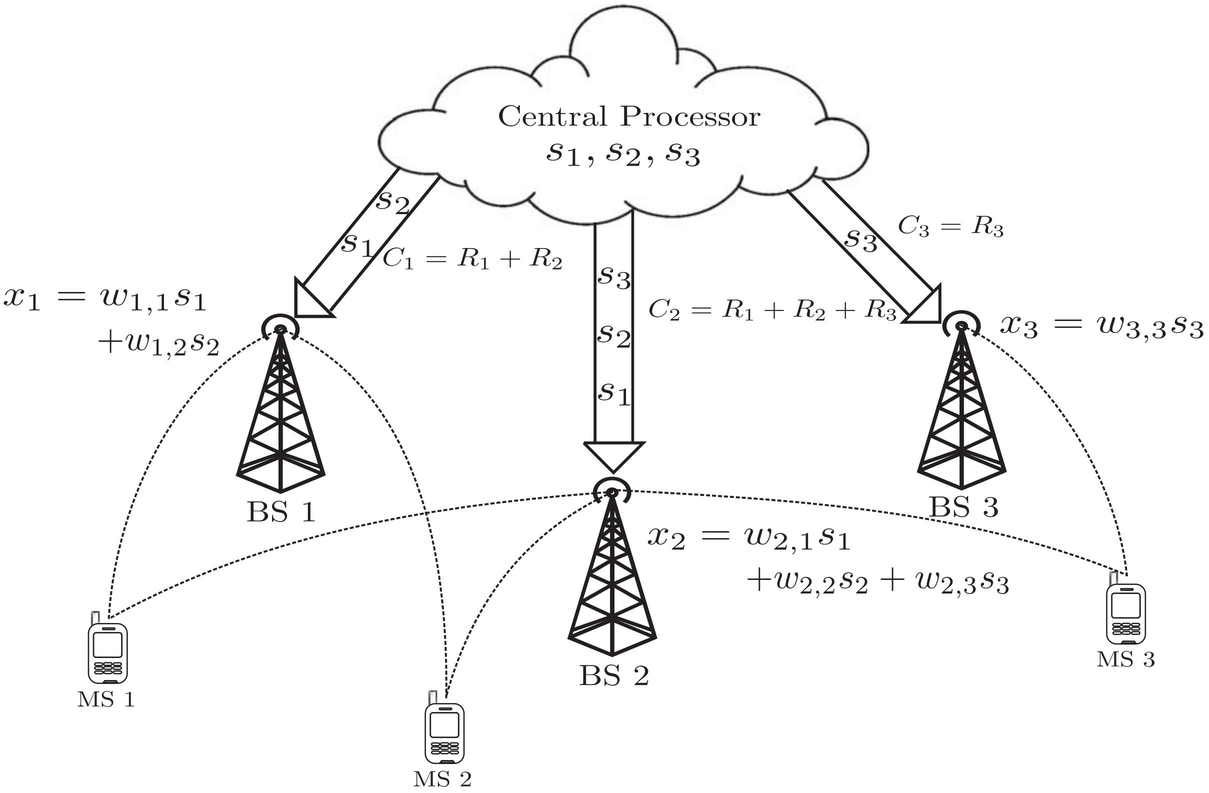

Figure 2.2: Example of the data-sharing strategy for the downlink C-RAN.

2.2.1 Optimization Framework

Let $\mathbf{w}_{l,k} \in \mathbb{C}^{M\times 1} = [w_{l,k}^1, \ldots, w_{l,k}^M]^T$ be the beamforming vector from BS l to user k with $w_{l,k}^m$ denoting the beamforming coefficient from mth antenna of BS l to user k and $\mathbf{w}_k \in \mathbb{C}^{LM\times 1} = [\mathbf{w}_{1,k}^T, \ldots, \mathbf{w}_{L,k}^T]^T$ be the aggregate network-wide beamformer to user k from all the BSs. If user k is not cooperatively served by BS l, then $\mathbf{w}_{l,k} = \mathbf{0}$. This can be equivalently represented by saying that $\|\mathbf{w}_{l,k}\|_2^2 = 0$, if BS l does not participate in serving user k. The beamformed signal $\mathbf{x}$ to be transmitted by all the BSs can be written as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k. \tag{2.3}$$

At user k, the signal-to-interference-plus-noise ratio (SINR) is

$$\text{SINR}_k = \mathbf{w}_k^H \mathbf{H}_k^H \left(\sum_{j\neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{H}_k \mathbf{w}_k. \tag{2.4}$$

The information theoretical achievable rate for user k is related to SINR as $R_k = \log(1 + \text{SINR}_k)$. However, this rate expression assumes Gaussian signaling, while in practice QAM constellations are typically used for the Gaussian channel in the moderate and high SINR regime. To achieve a given data rate, at a certain probability of error, we need an SINR higher than what is suggested above. This extra amount of power is usually captured by a so-called SNR gap, denoted here by Γ_m. Its value is approximately independent of the size of the constellation for square QAM, and can be easily computed as a function of the target probability of error [26]. For example, at $P_e = 10^{-6}$, $\Gamma_m = 9$ dB. The use of error correcting codes may lower the value of Γ_m. Now with the SNR gap taking into account, we can rewrite the achievable rate for user k as

$$R_k = \log\left(1 + \frac{\mathbf{w}_k^H \mathbf{H}_k^H \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_k \mathbf{w}_k}{\Gamma_m}\right). \tag{2.5}$$

The optimization problem of finding the optimal set of BS clusters and beamformers for the data-sharing scheme can now be formulated as a WSR maximization problem under per-antenna power constraints and per-BS backhaul constraints as follows:

$$\begin{aligned} \underset{\{\mathbf{w}_{l,k}\}}{\text{maximize}} \quad & \sum_{k=1}^{K} \alpha_k R_k && \text{(2.6a)} \\ \text{subject to} \quad & \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} && \text{(2.6b)} \\ & \sum_{k=1}^{K} \mathbb{1}\left\{\|\mathbf{w}_{l,k}\|_2^2\right\} R_k \leq C_l, \quad l \in \mathcal{L}, && \text{(2.6c)} \end{aligned}$$

where α_k denotes the priority weight associated with user k at the current user scheduling time slot which can be updated according to proportional fairness criterion, for example. The indicator function $\mathbb{1}\left\{\|\mathbf{w}_{l,k}\|_2^2\right\}$ in the constraint (2.6c) denotes if BS l participates in beamforming to user k, and if so, the user rate R_k is included in the backhaul constraint C_l. The constraint (2.6b) accounts for the per-antenna power constraint at antenna m

of BS l. The beamforming coefficients are computed at the central processor, and are assumed to be transmitted to the BSs without any error. We neglect the backhaul consumption for transmitting the beamformers as the beamformers need to be transmitted only once during each user scheduling time slot and compared with the backhaul needed to send the data, it is a very small fraction. The above formulation considers joint design of BS clustering, beamforming, and power control. Note that it also implicitly does joint user scheduling. This can be seen from the fact that a user k is scheduled, i.e., R_k is non-zero, if and only if its beamformer vector $\mathbf{w}_k$ is non-zero. Thus the user scheduling is implicitly jointly done along with BS clustering and beamforming optimization to satisfy the per-antenna and per-BS backhaul constraints. The optimization problem is solved repeatedly and the BS clusters are dynamically optimized in each time slot as the priority weights are updated.

2.2.2 Optimization Methodology

The presence of the backhaul constraint (2.6c) makes the optimization problem challenging. In this paper, we follow the approximation suggested in [7] to first write the indicator function as a l_0 norm which is then approximated as a weighted l_1 norm as

$$\mathbb{1}\left\{\|\mathbf{w}_{l,k}\|_2^2\right\} = \left\|\|\mathbf{w}_{l,k}\|_2^2\right\|_0 \approx \beta_{l,k}\,\|\mathbf{w}_{l,k}\|_2^2\,, \tag{2.7}$$

where $\beta_{l,k}$ is a constant weight associated with BS l and user k and is updated iteratively according to

$$\beta_{l,k} = \frac{1}{\|\mathbf{w}_{l,k}\|_2^2 + \tau}, \tag{2.8}$$

for some regularization constant $\tau > 0$ and $\|\mathbf{w}_{l,k}\|_2^2$ from the previous iteration. This simplifies the constraint (2.6c) to

$$\sum_{k=1}^{K} \beta_{l,k} \|\mathbf{w}_{l,k}\|_2^2 R_k \leq C_l, \quad l \in \mathcal{L}, \tag{2.9}$$

which is equivalent to a generalized power constraint, if R_k is assumed fixed and heuristically chosen from the previous iteration in an iterative manner. The resulting optimization problem then becomes:

$$\begin{aligned} \underset{\{\mathbf{w}_{l,k}\}}{\text{maximize}} \quad & \sum_{k=1}^{K} \alpha_k R_k && (2.10a) \\ \text{subject to} \quad & \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} && (2.10b) \\ & \sum_{k=1}^{K} \beta_{l,k} \hat{R}_k \|\mathbf{w}_{l,k}\|^2 \leq C_l, \quad l \in \mathcal{L}, && (2.10c) \end{aligned}$$

where $\hat{R}_k$ is the rate from the previous iteration.

Even though the approximated problem (2.10) is still non-convex, it can formulated as an equivalent WMMSE problem using the equivalence between the WSR maximization and the WMMSE problem. The advantage of working with the WMMSE problem is that the optimization variables can be split into groups such that with respect to each group of variables, the optimization problem is convex, if other variables are fixed. Thus we can use the block coordinate descent method to reach a stationary point of (2.10). [27] first established the relationship between the WSR maximization and the WMMSE problem for the MIMO broadcast channel. It is generalized to the MIMO interference channel in [24] and the MIMO interference channel with partial cooperation in [28]. In the context of C-RAN, the equivalence is used in [7]. The difference between the formulation (2.6) and that in [7] is the gap factor Γ_m and per-antenna power constraints, instead of per-BS power constraint. It is not difficult to verify that the equivalence between WSR optimization and WMMSE extends even for (2.10). We state the equivalence explicitly below.

Proposition 2.2.1. *Let* $\mathbf{u}_k \in \mathbb{C}^{N\times 1}$ *denote the receive beamformer at user* k *and the corresponding MSE defined as*

$$\begin{aligned} e_k = &\ \mathsf{E}\left[\left\|\mathbf{u}_k^H\mathbf{y}_k - s_k\right\|_2^2\right] &(2.11)\\ = &\ \mathbf{u}_k^H\left(\Gamma_m\left(\sum_{j\neq k}\mathbf{H}_k\mathbf{w}_j\mathbf{w}_j^H\mathbf{H}_k^H + \sigma^2\mathbf{I}\right) + \mathbf{H}_k\mathbf{w}_k\mathbf{w}_k^H\mathbf{H}_k^H\right)\mathbf{u}_k - 2Re\left\{\mathbf{u}_k^H\mathbf{H}_k\mathbf{w}_k\right\} + 1. & (2.12)\end{aligned}$$

Then the WSR maximization problem (2.10) is equivalent to the following WMMSE problem

$$\underset{\{\mathbf{w}_{l,k}\},\{\mathbf{u}_k\},\{\rho_k\}}{\text{minimize}} \quad \sum_{k=1}^{K}\alpha_k\left(\rho_k e_k - \log\rho_k\right) \qquad (2.13\text{a})$$

$$\text{subject to} \quad \sum_{k=1}^{K}|w_{l,k}^m|^2 \le P_l^m, \quad l\in\mathcal{L}, m\in\mathcal{M} \qquad (2.13\text{b})$$

$$\sum_{k=1}^{K}\beta_{l,k}\hat{R}_k\left\|\mathbf{w}_{l,k}\right\|^2 \le C_l, \quad l\in\mathcal{L} \qquad (2.13\text{c})$$

in the sense that for any stationary point $(\{\mathbf{w}_{l,k}^*\},\{\mathbf{u}_k^*\},\{\rho_k^*\})$ *of the WMMSE problem,* $(\{\mathbf{w}_{l,k}^*\})$ *is a stationary point of (2.10) and vice versa, where* ρ_k *denotes the MSE weight for user* k*.*

Proof. We provide the proof in Appendix A. □

It can be easily verified that the WMMSE problem (2.13) is convex with respect to each of the individual optimization variables $\{\mathbf{w}_{l,k}\},\{\mathbf{u}_k\},\{\rho_k\}$. This allows the block coordinate descent method to be applied iteratively over these variables. Specifically,

- The optimal MSE weight ρ_k under fixed $\{\mathbf{w}_k\}$ and $\{\mathbf{u}_k\}$ is given by

$$\rho_k = e_k^{-1}. \tag{2.14}$$

- The optimal receive beamformer $\mathbf{u}_k$ under fixed $\{\mathbf{w}_k\}$ and $\{\rho_k\}$ is given by

$$\mathbf{u}_k = \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}\right) + \mathbf{H}_k \mathbf{w}_k \mathbf{w}_k^H \mathbf{H}_k^H\right)^{-1} \mathbf{H}_k \mathbf{w}_k. \tag{2.15}$$

- The optimal transmit beamformers $\{\mathbf{w}_k\}$ under fixed $\{\mathbf{u}_k\}$, $\{\rho_k\}$ and fixed $\{\hat{R}_k\}$ can be obtained by solving following quadratically constrained quadratic programming (QCQP):

$$\underset{\{\mathbf{w}_{l,k}\}}{\text{minimize}} \quad \sum_{k=1}^{K} \mathbf{w}_k^H \mathbf{A}_k \mathbf{w}_k - \ \mathrm{Re}\{\mathbf{b}_k^H \mathbf{w}_k\} \tag{2.16a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} \tag{2.16b}$$

$$\sum_{k=1}^{K} \beta_{l,k} \hat{R}_k \left\|\mathbf{w}_{l,k}\right\|_2^2 \leq C_l, \quad l \in \mathcal{L}, \tag{2.16c}$$

where $\{\mathbf{A}_k\} \in \mathbb{C}^{LM \times LM}$ and $\{\mathbf{b}_k\} \in \mathbb{C}^{LM \times 1}$ are defined to be

$$\mathbf{A}_k = \sum_{j \neq k} \alpha_j \rho_j \Gamma_m \mathbf{H}_j^H \mathbf{u}_j \mathbf{u}_j^H \mathbf{H}_j + \alpha_k \rho_k \mathbf{H}_k^H \mathbf{u}_k \mathbf{u}_k^H \mathbf{H}_k, \tag{2.17}$$

$$\mathbf{b}_k = 2\alpha_k \rho_k \mathbf{H}_k^H \mathbf{u}_k. \tag{2.18}$$

We summarize the overall algorithm for the optimization of the data-sharing strategy in Algorithm 1.

Algorithm 1 WSR maximization for the data-sharing strategy

```
Initialization: $\{\beta_{l,k}\}, \{\mathbf{w}_k\}, \{\hat{R}_k\}$;
Repeat:
    1. For fixed $\{\mathbf{w}_k\}$, compute the MMSE receivers $\{\mathbf{u}_k\}$ and the corresponding MSE $\{e_k\}$ according to (2.15) and (2.12);
    2. Update the MSE weights $\{\rho_k\}$ according to (2.14);
    3. For fixed $\{u_k\}$, $\{\rho_k\}$, and $\{\hat{R}_k\}$ in (2.16c), find the optimal transmit beamformers $\{w_{l,k}\}$ by solving (2.16);
    4. Update $\{\beta_{l,k}\}$ as in (2.8);
    5. Compute the achievable rates $\{R_k\}$ according to (2.5). Update $\hat{R}_k = R_k$, $k \in \mathcal{K}$.
Until convergence
```

2.3 Compression Strategy

In the compression strategy, as shown in Fig. 2.3, the central processor computes the beamformed analog signals to be transmitted by the BSs. These signals have to be compressed before they can be forwarded to the corresponding BSs through the finite-capacity backhaul links. The process of compression introduces quantization noises; the quantization noise levels depend on the backhaul capacities.

2.3.1 Optimization Framework

In the data-sharing strategy, the beamformed signal is computed at the BSs. In the compression strategy, the beamformed signal is computed at the central processor, then compressed, sent over the backhaul links, and reproduced by the BSs. Let $\hat{\mathbf{x}}_l \in \mathbb{C}^{M\times 1} = [\hat{x}_l^1, \ldots, \hat{x}_l^M]^T$ denote precoded signal computed at the central processor intended for BS l and $\hat{\mathbf{x}} \in \mathbb{C}^{LM\times 1} = [\hat{\mathbf{x}}_1^T, \ldots, \hat{\mathbf{x}}_L^T]^T$ be the aggregate signal intended for all the BSs. Again let the beamforming vector from BS l to user k by $\mathbf{w}_{l,k} \in \mathbb{C}^{M\times 1} = [w_{l,k}^1, \ldots, w_{l,k}^M]^T$ with $w_{l,k}^m$ being the beamforming coefficient from mth antenna of BS l to user k and $\mathbf{w}_k \in \mathbb{C}^{LM\times 1} = [\mathbf{w}_{1,k}^T, \ldots, \mathbf{w}_{L,k}^T]^T$ be the aggregate network-wide beamformer to user k

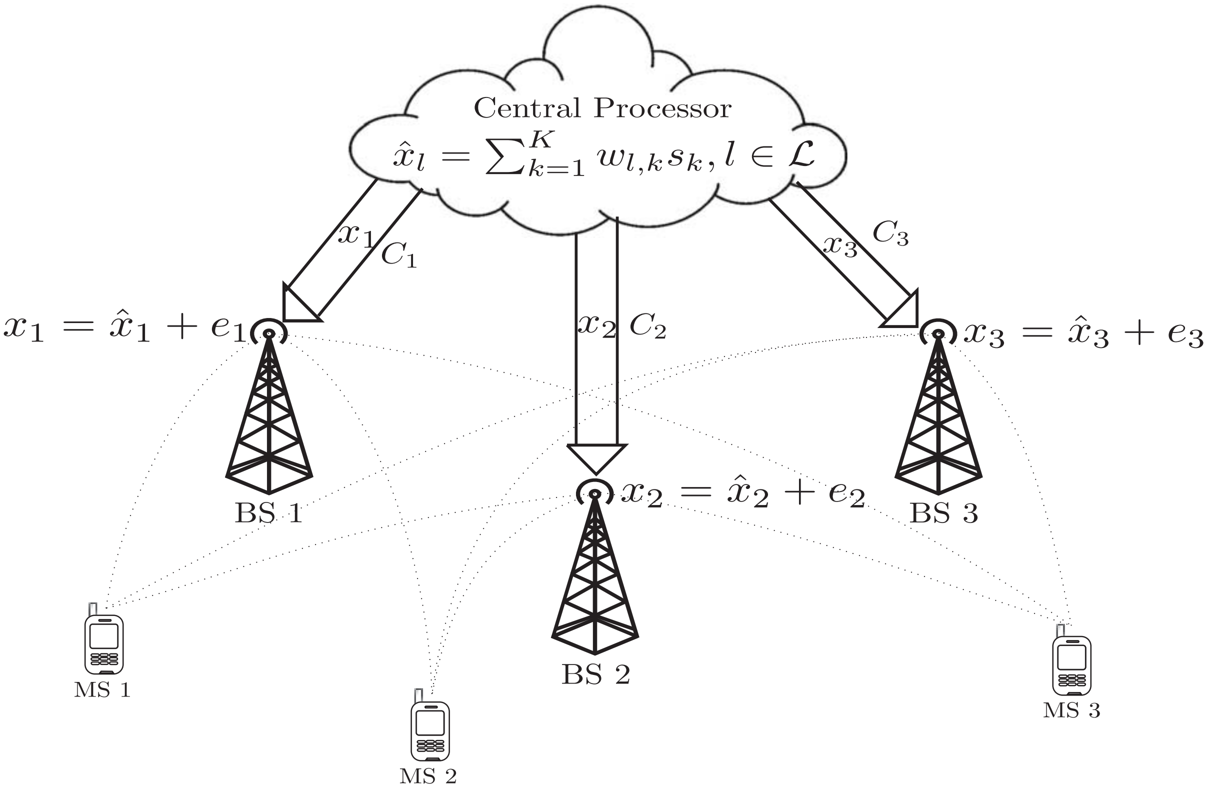

Figure 2.3: Example of the compression strategy for the downlink CRAN.

from all the BSs. We can then write $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \sum_{k=1}^{K} \mathbf{w}_k s_k. \tag{2.19}$$

The analog signals $\hat{\mathbf{x}}$ are then compressed and forwarded to BSs. We model the quantization process for $\hat{\mathbf{x}}$ as

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}, \tag{2.20}$$

where the quantization noise $\mathbf{e}$ is assumed to be complex Gaussian with covariance matrix $\mathbf{Q} \in \mathbf{C}^{LM \times LM}$ and independent of $\hat{\mathbf{x}}$. Under this model, the achievable rate for user k, accounting for the SNR gap, is given by

$$R_k = \log\left(1 + \frac{\mathbf{w}_k^H \mathbf{H}_k^H \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} + \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H\right)^{-1} \mathbf{H}_k \mathbf{w}_k}{\Gamma_m}\right). \tag{2.21}$$

We consider independent quantization at each antenna at all the BSs, in which case $\mathbf{Q}$ is a diagonal matrix with diagonal entries q_l^m. (Multivariate compression is also possible and has been studied in [16].) Assuming an ideal vector quantizer, the quantization noise level q_l^m and the backhaul capacity C_l^m allocated to each antenna at each BS are related as (from rate-distortion theory [29])

$$\log\left(1+\frac{\sum_{k=1}^{K}|w_{l,k}^m|^2}{q_l^m}\right)\leq C_l^m. \tag{2.22}$$

However, the quantizers used in practice for compression can be far from ideal. In order to capture these losses, we introduce a notion of gap to rate-distortion limit. Following [30], we note that the operational distortion, $\delta(R)$, achieved by virtually all practical quantizers at high resolution follow the relation

$$\delta(R)=\Gamma_q \text{var}(X)2^{-R}, \tag{2.23}$$

where $\text{var}(X)$ is the variance of the signal being quantized, R is the rate of quantizer, and Γ_q is a constant that depends on the particular choice of quantizer. For example, for a fixed-rate (uncoded) uniform scalar quantizer, $\Gamma_q=\frac{\sqrt{3}\pi}{2}$, which is approximately 2.72. For a uniform scalar quantizer followed by variable-rate entropy coding, $\Gamma_q=\frac{\pi e}{6}$, which is approximately 1.42. Note that $\Gamma_q=1$ corresponds to the distortion achievable by the best possible vector quantization scheme. Accounting for this, we can rewrite the relation above as

$$\log\left(1+\frac{\Gamma_q\sum_{k=1}^{K}|w_{l,k}^m|^2}{q_l}\right)\leq C_l^m. \tag{2.24}$$

The quantization noise relation described by (2.24) assumes that individual BSs have access to the quantization codebooks used at the central processor for compressing the signals intended for all of their antennas. The quantization codebooks depend on the rate of the quantizer, C_l^m, and the variance of the signal being compressed, $\sum_{k=1}^{K}|w_{l,k}^m|^2$.

Since we are also designing the beamforming coefficients $\{w_{l,k}\}$ at each user scheduling time slot, the variance of the signal being compressed can change at each user scheduling iteration. Also the rate of the quantizer, C_l^m, used for compressing the signal of antenna m at BS l, depends on the backhaul capacity allocated to antenna m of BS l. This allocation can also potentially be changed. Thus, to achieve (2.24), the information about the quantizantion codebooks used at the central procesor for all antennas of a BS needs to be sent to that BS at the start of each user scheduling iteration.

In practice, however, it may not be feasible to convey all such relevant codebook information from the central processor to each individual BS at each user scheduling time slot. We consider below two optimization formulations, one that allows for adaptive quantization codebooks, and other with fixed quantization codebooks, and the algorithms to solve them.

2.3.2 Optimization Methodology

Adaptive Quantization

We refer to the situation when the quantization codebooks are allowed to be changed at the central processor at each user scheduling time slot as adapative quantization. It is adaptive in the sense that, depending on the active users and their priority weights, the quantization codebooks are allowed to adapted. Recall that the quantization codebooks depend on the rate of the quantizer and the variance of the signal being compressed. In the case of single-antenna terminals, since the backhaul capacity per-antenna is fixed (which is same the per-BS backhaul capacity), the rate of the quantizer for that BS is fixed. In the case of multiple antennas, the rate of the quantizer used for an antenna at a BS is not fixed, as we do not have per-antenna backhaul constraints. We can allocate the quantization rates for different quantizers of different antennas so long as we meet the per-BS backhaul constraint. If we consider such flexible allocation, this makes the underlying optimization more difficult, as we have one more set of variables to optimize over. We first consider the case with single-antenna BSs with adaptive quantization, where the rate of the quantizer is fixed (as each BS only has a single antenna), but the variance of the signal to be compressed can change. For the case of multiple antennas considered in the next section, we look at the case of fixed quantization where we fix the rate of the quantizer for each of the antennas as well as the variance of the signals compressed.

In this section, we assume a single antenna at the BSs and the user terminals. For notational clarity, we drop the superscript m denoting the antenna index. The channel vector to user k from all BSs in this case is denoted by a column vector $\mathbf{h}_k \in \mathbb{C}^{L\times 1} = [h_{1,k}, \ldots, h_{L,k}]^T$. The achievable rate R_k is thus given by (2.21) with $\mathbf{H}_k = \mathbf{h}_k^H$, while the quantization relation with adaptive coding is given by (2.24).

The design of the compression strategy can now be stated as a WSR maximization problem over the transmit beamformers and the quantization noise levels as follows:

$$\begin{aligned}
\underset{\{w_{l,k}\},\{q_l\}}{\text{maximize}} \quad & \sum_{k=1}^{K} \alpha_k R_k && \text{(2.25a)}\\
\text{subject to} \quad & \sum_{k=1}^{K} |w_{l,k}|^2 + q_l \le P_l, \quad l \in \mathcal{L} && \text{(2.25b)}\\
& \sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\, q_l \le 0, \quad l \in \mathcal{L}. && \text{(2.25c)}
\end{aligned}$$

The constraint (2.25c) is just a reformulation of (2.24), while the constraint (2.25b) is the per-antenna power constraint at BS l.

Finding the globally optimal solution to (2.25) is challenging. An iterative approach based on the majorize-minimization (MM) algorithm has been suggested in [16]. The algorithm in [16] transforms $\mathbf{w}_k\mathbf{w}_k^H$ into a non-negative definite matrix variable $\mathbf{R}_k$ and ignores the rank constraint on $\mathbf{R}_k$ in the optimization. In this thesis, we propose a novel way to solve (2.25) by reformulating it as an equivalent WMMSE problem and then using the block coordinate descent method between the groups of variables of the transmit beamformers $\{\mathbf{w}_k\}$ and the quantization noise levels $\{q_l\}$, the receive beamformers $\{u_k\}$, and the MSE weights $\{\rho_k\}$. The algorithm can be shown to reach a stationary point of (2.25). Below we state the explicit equivalence.

For a receive beamformer u_k for user k, the corresponding MSE e_k, as defined in (2.11), is

$$e_k = |u_k|^2 \left(\Gamma_m \left(\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k \right) + |\mathbf{h}_k^H \mathbf{w}_k|^2 \right) - 2\text{Re}\{u_k^H \mathbf{h}_k^H \mathbf{w}_k\} + 1. \quad (2.26)$$

Now consider the following WMMSE optimization problem:

$$\underset{\substack{\{w_{l,k}\},\{q_l\},\\ \{u_k\},\{\rho_k\}}}{\text{minimize}} \quad \sum_{k=1}^{K} \alpha_k \left(\rho_k e_k - \log \rho_k\right) \quad (2.27a)$$

$$\text{subject to} \quad \sum_{k=1}^{K} |w_{l,k}|^2 + q_l \leq P_l, \quad l \in \mathcal{L} \quad (2.27b)$$

$$\sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l} - 1}{\Gamma_q} q_l \leq 0, \quad l \in \mathcal{L}, \quad (2.27c)$$

where ρ_k is the WSE weight for user k. The following holds true.

Proposition 2.3.1. *The WSR maximization problem (2.25) is equivalent to the WMMSE problem (2.27) in that for any stationary point* $(\{w_{l,k}^*\}, \{u_k^*\}, \{\rho_k^*\}, \{q_l^*\})$ *of the WMMSE problem,* $(\{w_{l,k}^*\}, \{q_l^*\})$ *is a stationary point of (2.25) and vice versa.*

Proof. We relegate the proof to appendix B. □

It is an easy exercise to verify that the WMMSE problem (2.27) is convex with respect to each of the individual optimization variables $\{w_{l,k}\}, \{u_k\}, \{\rho_k\}, \{q_l\}$. The individual block coordinate descent iterations then are as done as follows.

- The optimal MSE weight ρ_k under fixed $\{\mathbf{w}_k\}$ and $\{u_k\}$ is as given by (2.14)

- The optimal receive beamformer u_k under fixed $\{\mathbf{w}_k\}$ and $\{\rho_k\}$ is given by

$$u_k = \left(\Gamma_m\left(\sum_{j\neq k}|\mathbf{h}_k^H\mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H\mathbf{Q}\mathbf{h}_k\right) + |\mathbf{h}_k^H\mathbf{w}_k|^2\right)^{-1}\mathbf{h}_k^H\mathbf{w}_k. \tag{2.28}$$

- The optimization of the transmit beamformers $\{\mathbf{w}_k\}$ and the quantization noise levels $\{q_l\}$ under fixed $\{u_k\}$ and $\{\rho_k\}$ is solved via the following convex program:

$$\begin{aligned}
\underset{\{w_{l,k}\},\{q_l\}}{\text{minimize}} \quad & \sum_{k=1}^{K}\mathbf{w}_k^H\mathbf{A}_k\mathbf{w}_k - \ \mathrm{Re}\{\mathbf{b}_k^H\mathbf{w}_k\} + \Gamma_m\alpha_k\rho_k|u_k|^2\mathbf{h}_k^H\mathbf{Q}\mathbf{h}_k && (2.29\text{a})\\
\text{subject to} \quad & \sum_{k=1}^{K}|w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\,q_l \leq 0, \quad l\in\mathcal{L} && (2.29\text{b})\\
& \sum_{k=1}^{K}|w_{l,k}|^2 + q_l \leq P_l, \quad l\in\mathcal{L}, && (2.29\text{c})
\end{aligned}$$

where $\{\mathbf{A}_k\}$ and $\{\mathbf{b_k}\}$ are defined to be

$$\mathbf{A}_k = \sum_{j\neq k}\Gamma_m\alpha_j\rho_j|u_j|^2\mathbf{h}_j\mathbf{h}_j^H + \alpha_k\rho_k|u_k|^2\mathbf{h}_k\mathbf{h}_k^H, \tag{2.30}$$

$$\mathbf{b}_k = 2\alpha_k\rho_k u_k\mathbf{h}_k. \tag{2.31}$$

We further observe that the convex optimization problem (2.29) has a particular structure that can be exploited. Observe that the two constraints (2.29b) and (2.29c) provide a lower and an upper bound on $\{q_l\}$, respectively. Since the objective (2.29a) is monotonically decreasing in $\{q_l\}$, we can replace the inequality with equality in the constraint (2.29b) and substitute $\{q_l\}$ from (2.29b) into the objective (2.29a) and the constraint (2.29c). This results in a QCQP problem in only a single set of variables $\{\mathbf{w}_k\}$, which can be solved efficiently.

We summarize the overall algorithm to solve (2.25) in Algorithm 2.

As pointed before, for multiple antennas at the BSs, if we were to consider adaptive

Algorithm 2 WSR maximization for the compression strategy with adapative quantization

```
Initialization: $\{\mathbf{w}_k\}, \{q_l\}$;
Repeat:
    1. For fixed $\{\mathbf{w}_k\}, \{q_l\}$, compute the MMSE receivers $\{u_k\}$ and the corresponding MSE $\{e_k\}$ according to (2.28) and (2.26);
    2. Update the MSE weights $\{\rho_k\}$ according to (2.14);
    3. For fixed $\{u_k\}$ and $\{\rho_k\}$, find the optimal transmit beamformers $\{\mathbf{w}_k\}$ and quantization noise levels $\{q_l\}$ by solving the convex optimization problem (2.29);
Until convergence
```

quantization, the formulation above also needs to tackle the allocation of the available backhaul capacity for different quantizers of different antennas at the same BS while maintaining the per-BS backhaul constraint, since we do not have a per-antenna backhaul constraint. Such an optimization problem needs to deal with the additional set of optimization variables for the allocation of the backhaul capacities, which makes the optimization problem more challenging. For the case of multi-antenna BSs, we focus on the practical case of fixed quantization as discussed in the next section.

Fixed Quantization

In this section, we consider the quantization model when the quantization codebooks are fixed at the central processor and at the BSs. The achievable rate is as given by (2.21). To fix the codebook for the quantizer for antenna m of BS l, we assume that the range of the quantizer is constrained within the power constraint for the antenna, P_l^m. Further, since we have a backhaul constraint on each BS and not on each antenna, we allocate uniform backhaul capacity for each antenna of a BS such that $C_l^m = \frac{C_l}{M}$. With these assumptions, the quantization relation (2.24) becomes

$$\log\left(1 + \frac{\Gamma_q P_l^m}{q_l^m}\right) \leq \frac{C_l}{M}. \tag{2.32}$$

We can now formulate the WSR maximization problem over the transmit beamformers and the quantization noise levels as:

$$\begin{aligned}
&\underset{\{\mathbf{w}_{l,k}\},\{q_l^m\}}{\text{maximize}} && \sum_{k=1}^{K} \alpha_k R_k && \text{(2.33a)}\\
&\text{subject to} && \sum_{k=1}^{K} |w_{l,k}^m|^2 + q_l^m \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} && \text{(2.33b)}\\
& && q_l^m \geq \frac{\Gamma_q P_l^m}{2^{\frac{C_l}{M}} - 1}, \quad l \in \mathcal{L}, && \text{(2.33c)}
\end{aligned}$$

where the constraint (2.33c) is a reformulation of (2.32), and the constraint (2.33b) is the per-antenna power constraint on antenna m at BS l.

Note that the above formulation extends easily, if one allows adaptive quantization for the range of the quantizer by fixing the backhaul allocation, for example with uniform backhaul allocation. The algorithm developed for the case of adapative quantization, Algorithm 2, can be easily applied to the multi-antenna BSs in that case.

In order to solve the optimization problem (2.33), we first observe that the objective (2.33a) is a decreasing function of q_l^m. The constraint (2.33c) provides a lower bound on q_l^m, while the constraint (2.33b) provides an upper bound. Hence the constraint (2.33c) will always be met with equality at a stationary point. Thus we can substitute the value of q_l^m from (2.33c) into the objective (2.33a) as well as the constraint (2.33b) and eliminate the variables q_l^m. This modifies the constraint (2.33b) into

$$\sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m \left(1 - \frac{\Gamma_q}{2^{C_l^m} - 1}\right), \quad l \in \mathcal{L}, m \in \mathcal{M}. \tag{2.34}$$

We then end up with a WSR maximization problem with modified per-antenna power constraints, which is tackled by solving its equivalent WMMSE problem. The MSE for user k as defined in (2.11) is

$$e_k = \mathbf{u}_k^H \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} + \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H \right) + \mathbf{H}_k \mathbf{w}_k \mathbf{w}_k^H \mathbf{H}_k^H \right) \mathbf{u}_k - 2\mathrm{Re}\left\{ \mathbf{u}_k^H \mathbf{H}_k \mathbf{w}_k \right\} + 1. \tag{2.35}$$

under the receiver $\mathbf{u}_k$. We solve the following equivalent WMMSE problem for (2.33) with the modified constraint (2.34).

$$\underset{\{\mathbf{w}_{l,k}\},\{\mathbf{u}_k\},\{\rho_k\}}{\text{minimize}} \quad \sum_{k=1}^{K} \alpha_k \left(\rho_k e_k - \log \rho_k \right) \tag{2.36a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m \left(1 - \frac{\Gamma_q}{2^{C_l^m} - 1} \right), \quad l \in \mathcal{L}, m \in \mathcal{M}. \tag{2.36b}$$

We skip the formal statement of the equivalence and its proof as it can be seen as a special case of the Proposition 2.2.1, by ignoring the constraint (2.10c). The WMMSE problem (2.36a) is solved with block coordinate descent between $\{\mathbf{w_{l,k}}\}$, $\{\mathbf{u}_k\}$, and $\{\rho_k\}$ by solving the following individual optimization problems.

- The optimal MSE weight ρ_k under fixed $\{\mathbf{w}_k\}$ and $\{\mathbf{u}_k\}$ is given by (2.14).
- The optimal receive beamformer $\mathbf{u}_k$ under fixed $\{\mathbf{w}_k\}$ and $\{\rho_k\}$ is given by

$$\mathbf{u}_k = \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} + \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H \right) + \mathbf{H}_k \mathbf{w}_k \mathbf{w}_k^H \right)^{-1} \mathbf{H}_k^H \mathbf{w}_k. \tag{2.37}$$

- The optimal transmit beamformers $\{\mathbf{w}_k\}$ under fixed $\{\mathbf{u}_k\}$ and $\{\rho_k\}$ can be obtained by solving the following QCQP problem:

$$\underset{\{\mathbf{w}_{l,k}\}}{\text{minimize}} \quad \sum_{k=1}^{K} \mathbf{w}_k^H \mathbf{A}_k \mathbf{w}_k - \ \mathrm{Re}\{\mathbf{b}_k^H \mathbf{w}_k\} \tag{2.38a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m \left(1 - \frac{\Gamma_q}{2^{C_l^m} - 1} \right), \quad l \in \mathcal{L}, m \in \mathcal{M}, \tag{2.38b}$$

where $\{\mathbf{A}_k\}$ and $\{\mathbf{b}_k\}$ are as defined in (2.17) and (2.18) respectively.

Cellular Layout	Hexagonal 7-cell wrapped-around
Channel bandwidth	10 MHz
Distance between cells	0.8 km
Number of users/cell	30
Number of macro-BSs/cell	1
Number of pico-BSs/cell	3
Max. Tx power at antenna/macro-BS	43 dBm
Max. Tx Power at antenna/pico-BS	30 dBm
Antenna gain	15 dBi
Background noise	-169 dBm/Hz
Path loss from macro-BS to user	$128.1 + 37.6 \log_{10}(d)$
Path loss from pico-BS to user	$140.7 + 36.7 \log_{10}(d)$
Log-normal shadowing	8 dB
Rayleigh small scale fading	0 dB
SNR gap (Γ_m)	9 dB
Rate-distortion gap (Γ_q)	4.3 dB

Table 2.1: Simulation parameters for 7-cell wrapped-around two-tier heterogeneous network.

The overall algorithm for solving (2.33) is summarized in Algorithm 3.

Algorithm 3 WSR maximization for the compression strategy with fixed quantization

```
Initialization: $\{\mathbf{w}_k\}$;
Repeat:
    1. For fixed $\{\mathbf{w}_k\}$, compute the MMSE receivers $\{\mathbf{u}_k\}$ and the corresponding MSE $\{e_k\}$ according to (2.37) and (2.35);
    2. Update the MSE weights $\{\rho_k\}$ according to (2.14);
    3. For fixed $\{\mathbf{u}_k\}$, $\{\rho_k\}$, find the optimal transmit beamformers $\{\mathbf{w}_{l,k}\}$ by solving (2.38);
Until convergence
```

2.4 Performance Comparison

In this section, we numerically compare the performance of the data-sharing and compression strategies. We consider a 7-cell wrapped-around two-tier heterogeneous network with simulation parameters as listed in Table 2.1. Each cell is a regular hexagon with 1

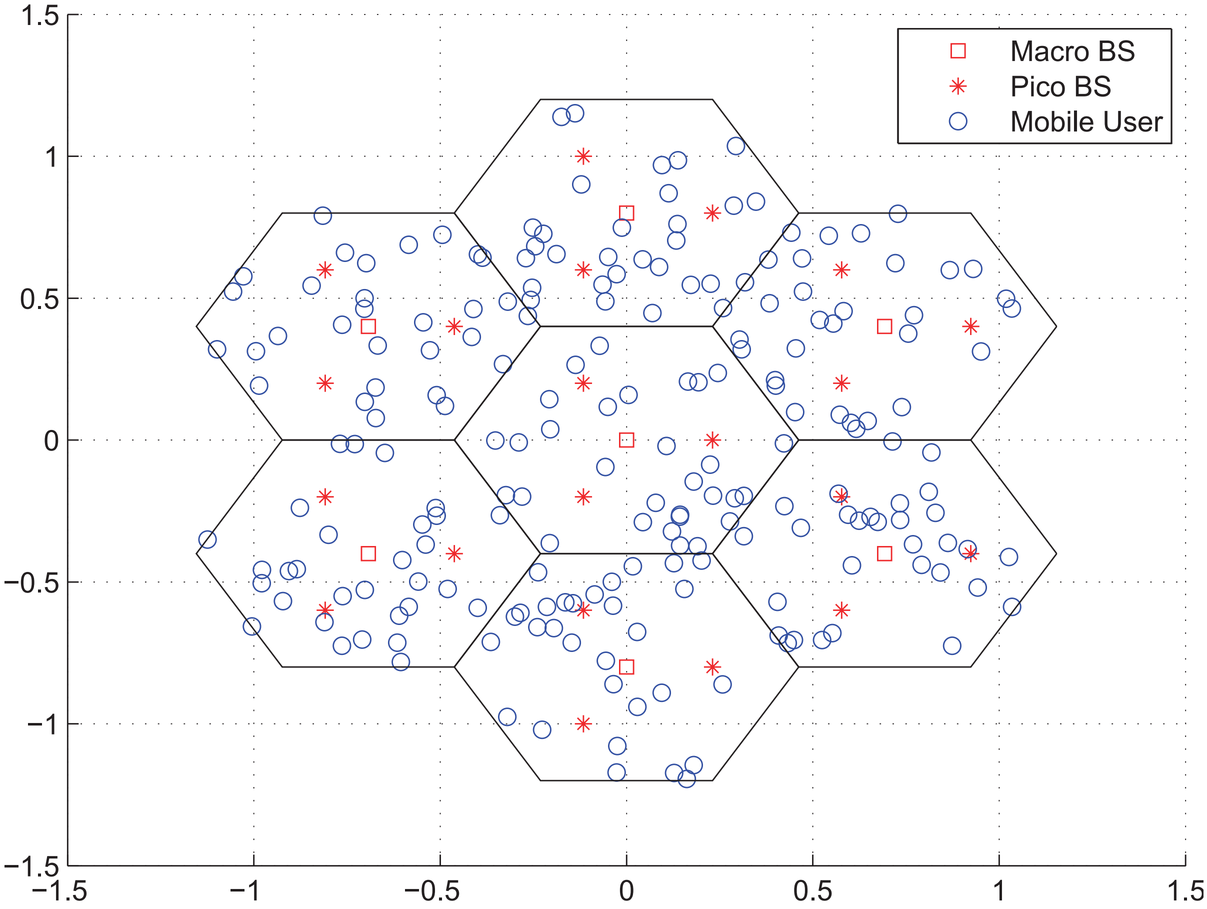

Figure 2.4: Illustration of the 7-cell wrapped around two-tier heterogeneous network with 30 users randomly placed in each cell.

macro-BS at the center and 3 pico-BSs equally separated in space. There are 30 users randomly placed in each cell. A sample of BS and user locations within the network in illustrated in Fig. 2.4. All the macro-BSs and pico-BSs are connected to a centralized processor by capacity-limited backhaul links. We compare the performance of the two strategies under varying backhaul capacities. The combined background noise and interference caused by two tiers of cells outside the 7-cells is estimated to be -150 dBm/Hz. We assume an SNR gap of $\Gamma_m = 9$ dB (corresponding to uncoded QAM transmission) and a gap to rate-distortion limit of $\Gamma_q = 4.3$ dB (corresponding to uncoded fixed-rate uniform scalar quantizer). At each time slot, we solve the respective network optimization problems and update the weights in the WSR maximization according to the proportional

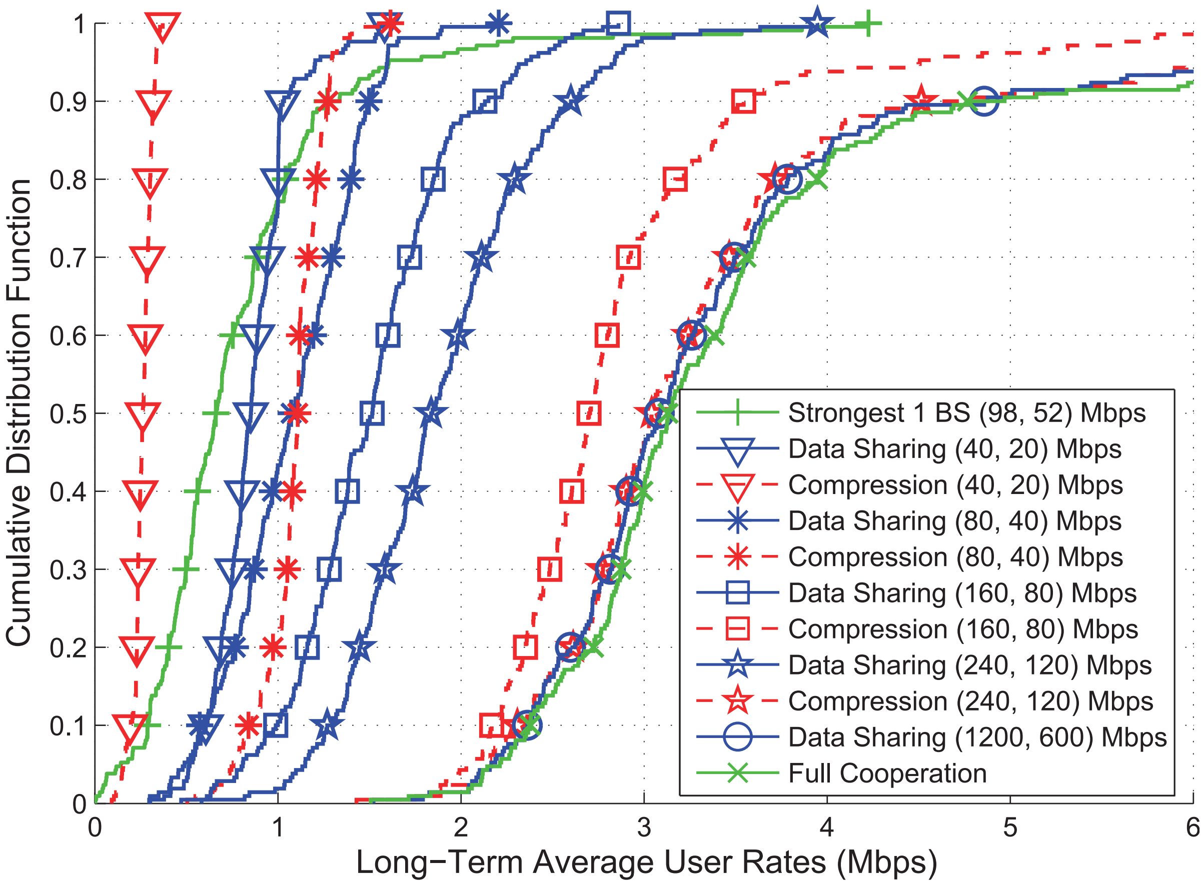

Figure 2.5: Cumulative distribution function of user rates for the data-sharing and compression strategies with single-antenna terminals and adaptive quantization.

fair criterion.

In the first set of simulations, we compare the performance of the data-sharing strategy and the compression strategy with adaptive quantization and single transmit antenna at both the macro-BSs and pico-BSs, and single receive antenna at the users. Fig. 2.5 shows the cumulative distribution of user rates under varying backhaul capacities for both strategies. Plots for the compression strategy are shown in red color, while those for data sharing strategy are shown in blue color. For reference, we also include the full cooperation case with infinite backhaul capacity and the baseline scheme of no cooperation with each user connected to the strongest BS.

When the backhaul capacity is low at 40 Mbps/macro-BS and 20 Mbps/pico-BS, the data-sharing strategy outperforms the compression strategy. The 50-percentile rate for the data-sharing strategy is about 3 times that of the compression strategy. If we double the backhaul capacity to 80 Mbps/macro-BS and 40 Mbps/pico-BS, the compression strategy becomes comparable to the data-sharing strategy and both have about the same 50-percentile user rates. At this operating point, the sum backhaul capacity is about 6 times that of the average sum rate per cell. We also observe that the compression strategy favours low rate users while the data-sharing strategy favours high rate users. A reason for this is that the compression strategy under low backhaul capacity is limited by the quantization noises which are about the same for all the BS signals resulting in more uniform user rates.

We observe that with moderate-to-high backhaul capacity of 160 Mbps/macro-BS and 80 Mbps/pico-BS, the compression strategy outperforms the data-sharing strategy with the 50-percentile rate for the compression strategy more than 2.5 times than that of data-sharing. Increasing the backhaul in this regime improves the compression strategy drastically, while the data-sharing strategy sees only a moderate increase. This is because, at low backhaul capacity, the performance of the compression strategy is limited by the quantization noises. An increase in backhaul capacity reduces the quantization noise levels exponentially, while a similar increase in the backhaul capacity does not buy as much for the data-sharing strategy. Finally with a backhaul of 240 Mbps/macro-BS and 120 Mbps/pico-BS, the compression strategy performs close to the full cooperation limit, while for the data-sharing strategy, backhaul capacities of 1200 Mbps/macro-BS and 600 Mbps/pico-BS are needed to get as close. This is because to match the full cooperation limit, the data-sharing strategy needs large cluster size, leading to significantly higher backhaul capacity.

In the second set of simulations, we compare the performance of the data-sharing strategy and the compression strategy with fixed quantization when the terminals have multiple antennas. We assume 4 antennas per macro-BS, 2 antennas per pico-BS, and

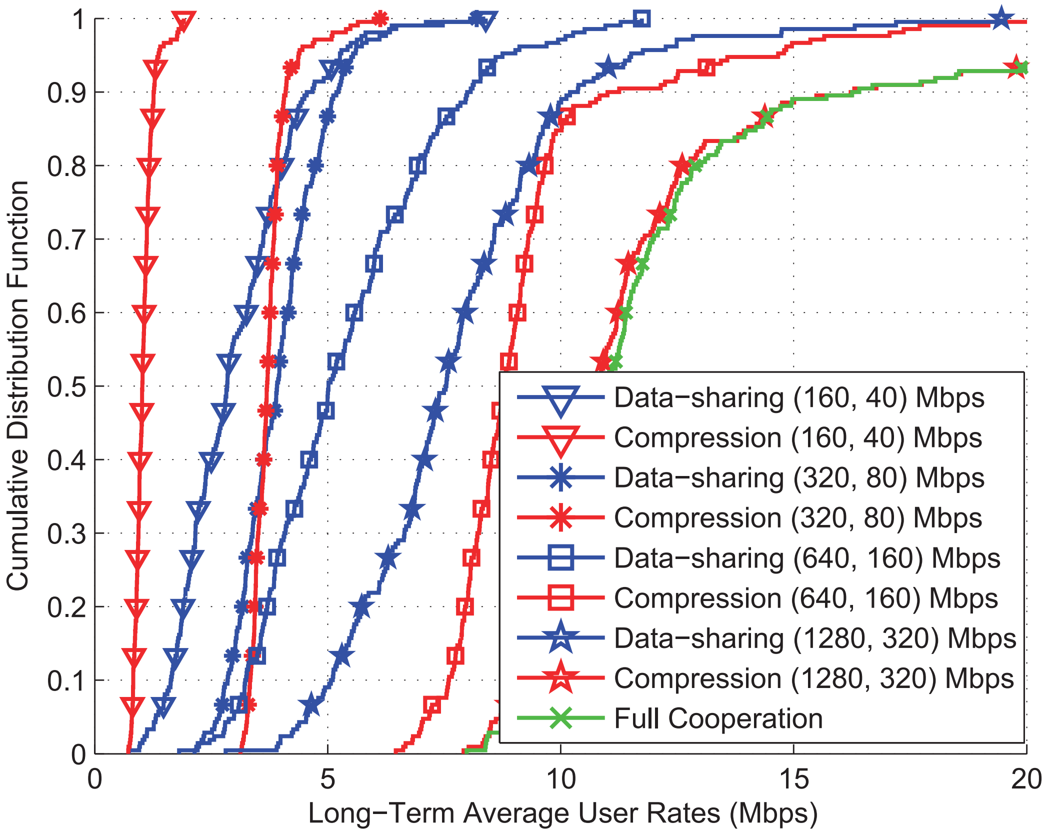

Figure 2.6: Cumulative distribution functions of user rates for the data-sharing and compression strategies with multi-antenna terminals and fixed quantization.

2 receive antennas for each user. Fig. 2.6 shows the cumulative distribution of user rates with varying backhaul capacities for both strategies. For reference, the plot for full cooperation with infinite backhaul capacity is also included. We observe similar trends as in the case of single-antenna terminals and adaptive quantization. When the backahul capacity is low at 160 Mbps/macro-BS and 40 Mbps/pico-BS (note that on average per-antenna backhaul capacities are maintained at 40 Mbps/macro-BS antenna and 20 Mbps/pico-BS antenna), the data-sharing strategy outperforms the compression compression strategy. The 50-percentile rate for the data-sharing strategy is about 2.5 times that of the compression strategy.

If we double the backhaul capacity to 320 Mbps/macro-BS and 80 Mbps/pico-BS, we see that the compression strategy becomes comparable to the data-sharing strategy and both have about the same 40-percentile user rates. In this regime the sum backhaul capacity is about 5 times that of the average sum rate per cell. This is in the similar range to what we observe in the single-antenna case. As the backhaul capacity is increased further at 640 Mbps/macro-BS and 160 Mbps/pico-BS, the compression strategy starts to significantly outperform the data-sharing strategy with the 50-percentile user rate is about 80% more than that of the data-sharing strategy. With the backhaul capacity of 1280 Mbps/macro-BS and 320 Mbps/pico-BS, the compression strategy already achieves the maximum achievable rates of the full cooperation. At this backhaul capacity the quantization noises are small enough that they do not affect the user rates. At the same backhaul capacity, the data-sharing strategy is still far behind that of full cooperation. This is because the backhaul capacity is not high enough to allow for the backhaul exchange required to maintain full cooperation.

It is important to note that the benefits from the compression strategy come at a cost of high CSI requirements at the central processor. To understand the impact of CSI on the data-sharing and compression strategies, we limit the amount of CSI available at the central processor by only allowing CSI of the few strongest BSs for each user. We call such a restriction as clustered CSI, when CSI of only a cluster of BSs around any user is available. All the algorithms can be adapted when such clustered CSI is available for each user.

Fig. 2.7 shows the cumulative distribution of user rates for both strategies when the CSI is limited to only 7 strongest BSs for each user. We observe that the general trend seen in the above two cases remain the same. At low backhaul capacity of 160 Mbps/maco-BS and 40 Mbps/pico-BS, the data-sharing strategy outperforms the compression strategy, while at the high backhaul capacity of 640 Mbps/macro-BS and 160 Mbps/pico-BS, the compression strategy outperforms the data-sharing strategy. However, notice that, in this case the compression strategy is not as significantly better than

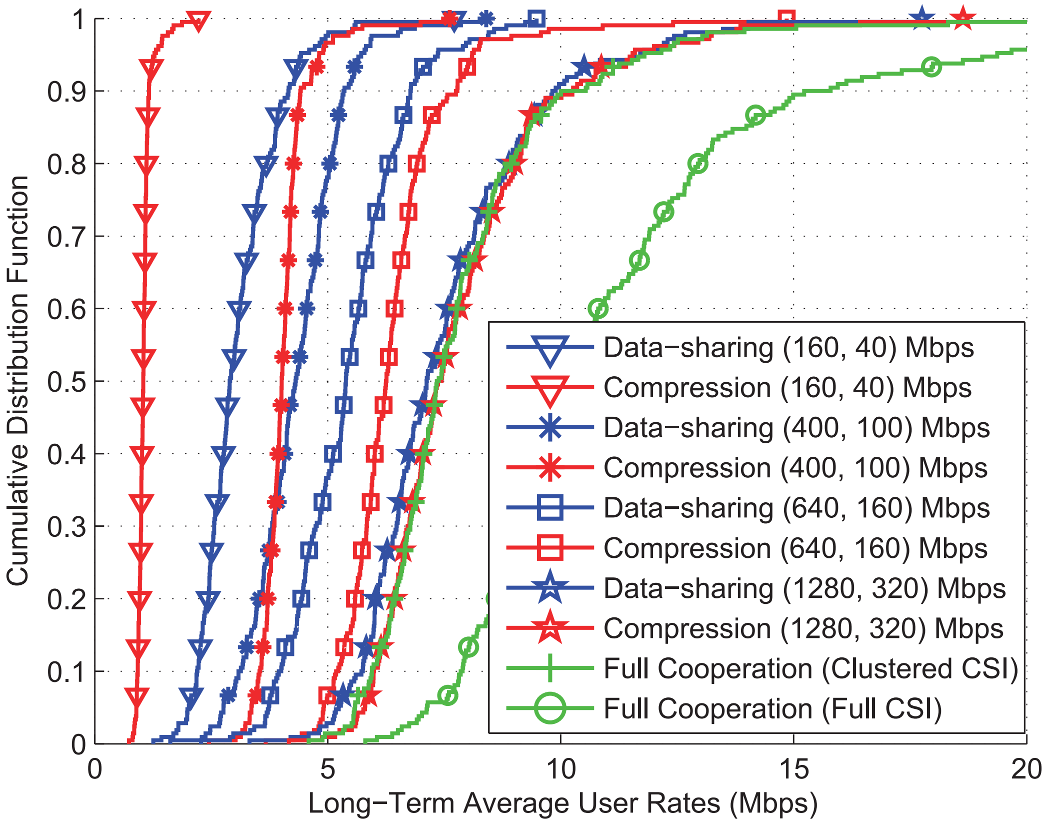

Figure 2.7: Comparison of cumulative distribution functions of user rates for the data-sharing and compression strategies with clustered CSI.

the data-sharing strategy. The 50-percentile user rate of the compression strategy is only 20% better than that of the data-sharing strategy, as compared with the case with full CSI when the it was almost 80% better. We illustrate this point further in the next plot. Finally, when the backhaul capacity is high at 1280 Mbps/macro-BS and 320 Mbps/pico-BS, both the data-sharing and compression strategies saturate to the full cooperation user rates with infinite backhaul capacity under limited CSI. For reference, the plot with full cooperation with infinite backhaul capacity and full CSI is also included to highlight the performance loss that is attributed to the lack of CSI.

In order to closely look at the how the lack of complete CSI affects the data-sharing and compression strategies, we fix the backhaul capacity at 320 Mbps/macro-BS and 80 Mbps/pico-BS. This is the regime where the two strategies are comparable in the case

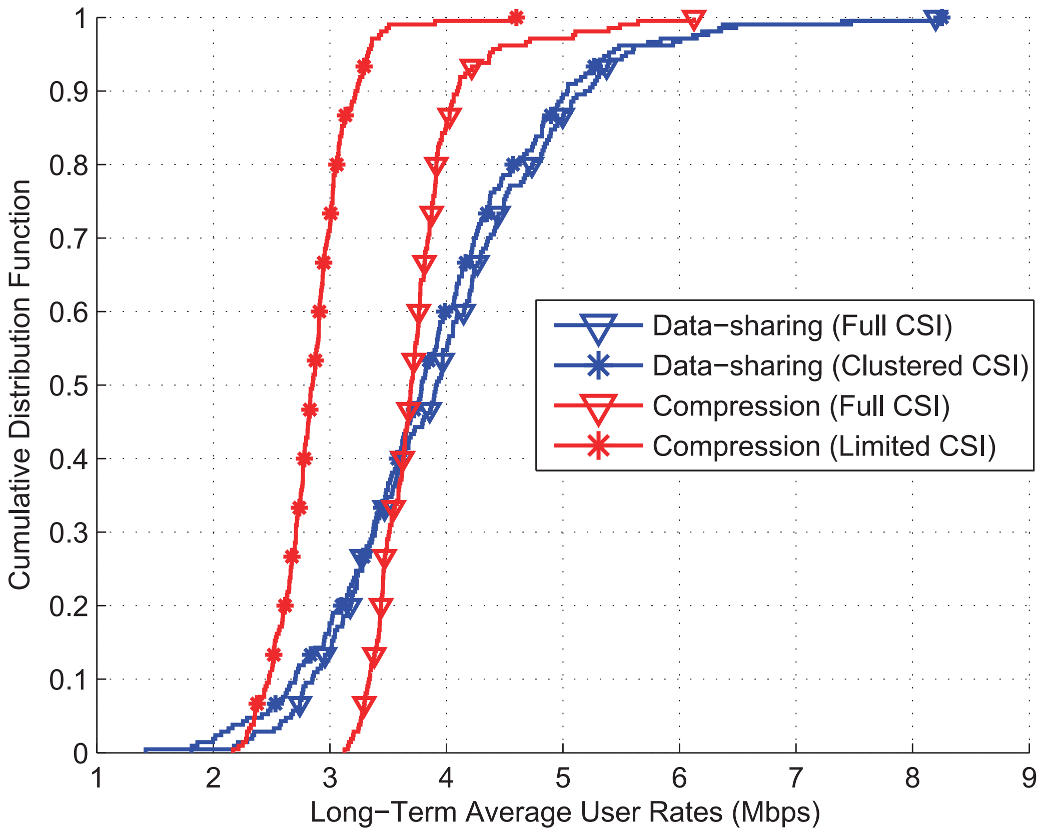

Figure 2.8: Comparison of the cumulative distribution functions of user rates for the data-sharing and compression strategies with full CSI and clustered CSI at the backahul capacity of 320 Mbps/macro-BS and 80 Mbps/pico-BS.

with full CSI. Fig. 2.8 shows the cumulative distribution of user rates for both strategies with full CSI and with clustered CSI at this fixed backhaul capacity. As is evident from the plot, the compression strategy suffers more than the data-sharing strategy when only partial CSI is available. The reason for this behavior is that the compression strategy benefits from having the ability to fully cooperate at the central processor, but when clustered CSI is available at the central processor, the cooperation cluster size at the central processor becomes limited. The data-sharing strategy on the other hand does not pay as much penalty because the cooperation cluster for the data-sharing strategy is already small due to the backhaul capacity limitations. As a result, we also see that the backhaul capacity at which the two strategies are comparable is higher when CSI is restricted, where it is at 400 Mbps/macro-BS and 100 Mbps/pico-BS, than the case with full CSI, where it is at 320 Mbps/macro-BS and 80 Mbps/pico-BS.

2.5 Summary

This chapter compares two fundamentally different strategies, the data-sharing and the compression strategy, for the downlink C-RAN under realistic network settings considering various practical aspects. We provide optimization frameworks for both strategies by exploiting the equivalence between the WSR maximization and the WMMSE problem. We then compare the performance of both strategies under varying backhaul capacities. Our main conclusion is that the backhaul capacity constraint is crucial in deciding which strategy to adopt. The compression strategy offers better user rates for moderate-to-high backhaul capacities, due to its ability to have full cooperation before quantization. But it suffers from high quantization loss at low backhaul capacity in which case it is better to do data-sharing with limited cooperation cluster. Further, the compression strategy is more sensitive to the availability of CSI than the data-sharing strategy, as in the former the benefits stem from the ability to fully cooperate at the central processor, which is affected adversely by the lack of CSI.

Chapter 3

Hybrid Strategy

In the data-sharing based cooperation scheme, the backhaul links are exclusively used to carry user messages. The advantage of such an approach is that BSs get clean messages which they can use for joint encoding. However, the backhaul capacity constraint limits the cooperation cluster size for each user. In the compression based scheme, the precoding operation is exclusively performed at the central processor. The main advantage of such an approach is that, since the central processor has access to all the user data, it can form a joint precoding vector using all the user messages, thus achieving full BS cooperation. Additionally, the BSs can now be completely oblivious of the user codebooks as the burden of preprocessing is shifted from the BSs to the central processor. However, since the precoded signals are compressed, we pay a price in the form of quantization noises.

This paper proposes a hybrid compression and message-sharing strategy in which the precoding operation is split between the central processor and the BSs. The rationale is that as the desired precoded signal typically consists of both strong and weak users, it may be beneficial to send clean messages for the strong users, rather than including them as a part of the signal to be compressed. In so doing, the amplitude of the signal that needs to be compressed can be lowered, and the required number of compression bits reduced.

Building on this intuition, this paper proposes an approach where a part of backhaul capacity is used to send direct messages for some users (for whom the BSs are better off receiving messages directly, instead of their contributions in the compressed precoded signals) and the remaining backhaul capacity is used to carry the compressed signal that combines the contributions from the rest of the users. Typically, each BS receives direct messages for the strong users and compressed precoded signals combining messages of the rest of the weak users in the network. Each BS then combines the direct messages with the decompressed signal, and transmits the resulting precoded signal on its antenna. Note that the appropriate beamforming coefficients are assumed to be available at both the cloud processor and at the BSs.

We point out that a dirty-paper coding based scheme proposed in [15] also makes use of the backhaul links to carry a combination of user message and the compressed version of interfering signal from the neighboring BS in a simplified linear array model. But the scheme of [15] is limited to the simplified linear array model; it also does not provide a method to decide if and what user messages should be shared among the BSs and what signals should be compressed.

3.1 Optimization Framework

In the hybrid strategy, as shown in Fig. 3.1, the central processor computes a part of the beamformed analog signals to be transmitted by BSs. These signals are compressed and sent over to BSs using a part of the backhaul capacity. For rest of the beamformed signal, the central processor sends digital data of selected users to the BSs using the remaining backhaul capacity. To simplify the description of the hybrid strategy, we assume single-antenna at the BSs and the user terminals.

The idea is to introduce separate beamforming coefficients for the data-sharing and compression parts. Let $\mathbf{w}_k^c \in \mathbb{C}^{L\times 1} = [w_{1,k}^c, \ldots, w_{L,k}^c]^T$ be the beamformers for user k

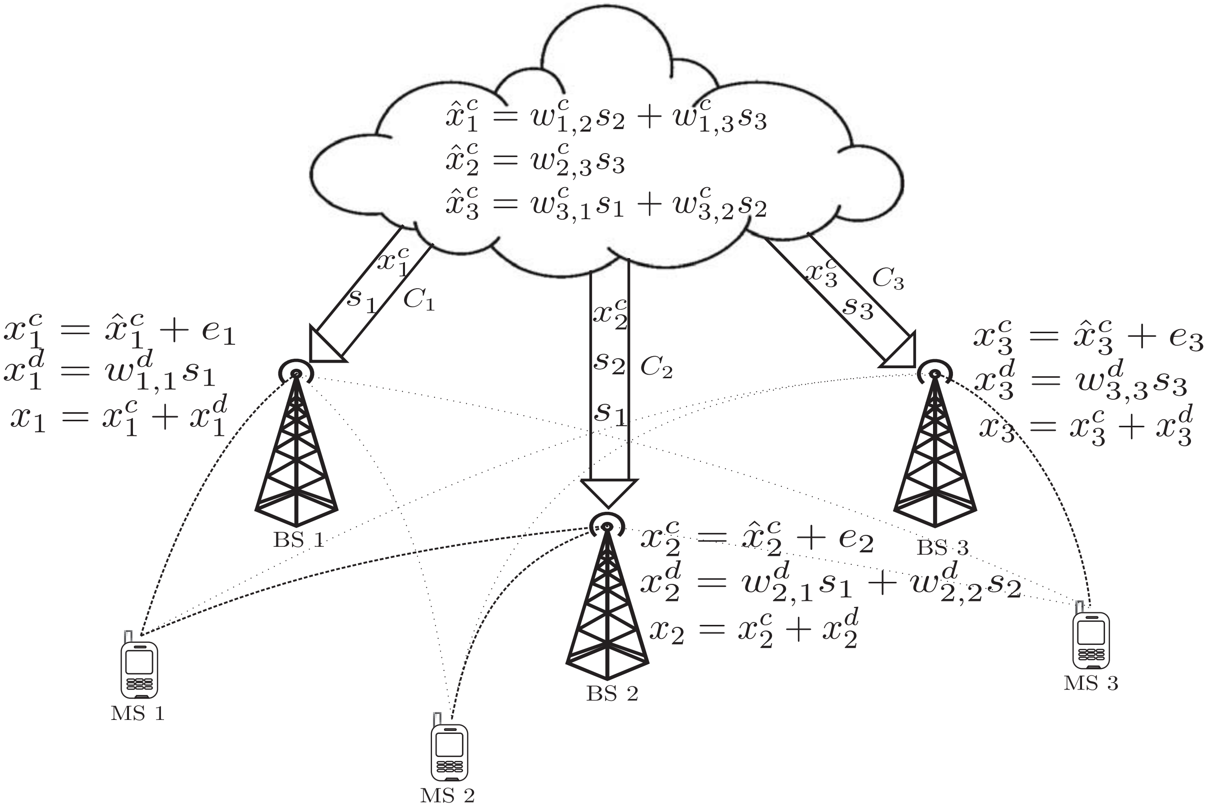
$\hat{x}_1^c = w_{1,2}^c s_2 + w_{1,3}^c s_3$
$\hat{x}_2^c = w_{2,3}^c s_3$
$\hat{x}_3^c = w_{3,1}^c s_1 + w_{3,2}^c s_2$
x_1^c
s_1
C_1
x_2^c
s_2
s_1
C_2
x_3^c
s_3
C_3
$x_1^c = \hat{x}_1^c + e_1$
$x_1^d = w_{1,1}^d s_1$
$x_1 = x_1^c + x_1^d$
$x_3^c = \hat{x}_3^c + e_3$
$x_3^d = w_{3,3}^d s_3$
$x_3 = x_3^c + x_3^d$
$x_2^c = \hat{x}_2^c + e_2$
$x_2^d = w_{2,1}^d s_1 + w_{2,2}^d s_2$
$x_2 = x_2^c + x_2^d$
BS 1
BS 2
BS 3
MS 1
MS 2
MS 3

Figure 3.1: Example of the hybrid data-sharing and compression strategy for the downlink C-RAN.

used to compute the beamformed signal that is going to be compressed at the central processor. Let $\hat{\mathbf{x}}^c \in \mathbb{C}^{L\times 1} = [\hat{x}_1^c, \ldots, \hat{x}_L^c]^T$ denote the beamformed signals intended for all the BSs computed at the central processor. These are given by

$$\hat{\mathbf{x}}^c = \sum_{k=1}^{K} \mathbf{w}_k^c s_k. \tag{3.1}$$

The quantization process for $\hat{\mathbf{x}^c}$ is again modeled as

$$\mathbf{x}^c = \hat{\mathbf{x}}^c + \mathbf{e}, \tag{3.2}$$

where $\mathbf{e}$ is the quantization noise with covariance $\mathbf{Q} \in \mathbb{C}^{L\times L}$ assumed to be Gaussian and independent of $\hat{\mathbf{x}}^c$. Assuming independent quantization at each BS, in which case $\mathbf{Q}$ is a diagonal matrix with diagonal entries q_l, the amount of backhaul capacity consumed by

BS l, C_l^c, for the compression part of its total beamformed signal is given by

$$\log\left(1+\frac{\Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2}{q_l}\right) \leq C_l^c. \tag{3.3}$$

Similarly let $\mathbf{w}_k^d \in \mathbb{C}^{L\times 1} = [w_{1,k}^d, \ldots, w_{L,k}^d]^T$ be the beamformers that are used for data-sharing at the BSs and $\mathbf{x}^d \in \mathbb{C}^{L\times 1} = [x_1^d, \ldots, x_L^d]^T$ denote the beamformed signals computed at the BSs using the direct data given by

$$\mathbf{x}^{\mathbf{d}} = \sum_{k=1}^{K} \mathbf{w}_k^d s_k. \tag{3.4}$$

If BS l does not receive direct data for user k, then $w_{l,k}^d = 0$. The amount of backhaul capacity, C_l^d, consumed by BS l for the data-sharing part is then given by

$$\mathbb{1}\left\{|w_{l,k}^d|^2\right\} R_k \leq C_l^d, \tag{3.5}$$

where the indicator function is used to indicate whether BS l participates in computing the beamformed signal using the direct data for user k. If so, the backhaul needs to support the user rate R_k. Note that we neglect the portion of the backhaul capacity that would be needed to be communicate the beamforming coefficients at the start of each user scheduling iteration as the it is negligible compared to the direct data communicated within that user scheduling iteration. The final beamformed signal transmitted by BSs to the users, $\mathbf{x}$, is then the sum of compressed beamformed signals, $\mathbf{x}^c$, communicated through the backhaul link and the direct beamformed signal, $\mathbf{x}^d$, computed at the BSs, i.e.,

$$\mathbf{x} = \mathbf{x}^c + \mathbf{x}^d. \tag{3.6}$$

The achievable rate for user k is then

$$R_k = \log\left(1 + \frac{|\mathbf{h}_k^H\left(\mathbf{w}_k^c + \mathbf{w}_k^d\right)|^2}{\Gamma_m\left(\sum_{j\neq k}|\mathbf{h}_k^H\left(\mathbf{w}_j^c + \mathbf{w}_j^d\right)|^2 + \sigma^2 + \mathbf{h}_k^H\mathbf{Q}\mathbf{h}_k\right)}\right). \tag{3.7}$$

If we let $\mathbf{w}_k^c + \mathbf{w}_k^d = \mathbf{w}_k$, $k \in \mathcal{K}$, the rate R_k can be simplified to

$$R_k = \log\left(1 + \frac{|\mathbf{h}_k^H\mathbf{w}_k|^2}{\Gamma_m\left(\sum_{j\neq k}|\mathbf{h}_k^H\mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H\mathbf{Q}\mathbf{h}_k\right)}\right), \tag{3.8}$$

where $\mathbf{w}_k \in \mathcal{C}^{L\times 1}$ can be thought of as the final combined beamformer for user k.

Now the WSR maximization problem for the hybrid strategy can then be formulated as follows:

$$\underset{\substack{\{w_{l,k}^d\},\{w_{l,k}^c\},\\ \{w_{l,k}\},\{q_l\}}}{\text{maximize}} \quad \sum_{k=1}^K \alpha_k R_k \tag{3.9a}$$

$$\text{subject to} \quad \sum_{k=1}^K |w_{l,k}|^2 + q_l \leq P_l, \quad l \in \mathcal{L} \tag{3.9b}$$

$$\sum_{k=1}^K \mathbb{1}\left\{|w_{l,k}^d|^2\right\} R_k + \log\left(1 + \frac{\Gamma_q \sum_{k=1}^K |w_{l,k}^c|^2}{q_l}\right) \leq C_l, \quad l \in \mathcal{L} \tag{3.9c}$$

$$w_{l,k}^d + w_{l,k}^c = w_{l,k}, \quad l \in \mathcal{L}, k \in \mathcal{K}. \tag{3.9d}$$

Note that in the problem formulation (3.9) above, it may seem at first that, we allow a more general hybrid strategy where a user k can *both* participate in direct data-sharing to a BS l as well as be part of the signal compressed by that BS, if both the beamforming coefficients $w_{l,k}^c$ and $w_{l,k}^d$ are non-zero. However, it can shown that, if R_k in the constraint (3.9c) is fixed, indeed at most one of the two can be non-zero, i.e., a user may only participate in data-sharing or compression, but not both. Intuitively this is due to the fact that if a user's data is shared at a particular BS, it is always better to put all the beamforming power in the data-sharing beamformer, rather than splitting it with the compression beamformer, to avoid the quantization noise penalty associated with the compression process. A more precise statement is given below.

Proposition 3.1.1. *Any stationary point* $(\{w_{l,k}\}, \{w^c_{l,k}\}, \{w^d_{l,k}\})$ *to the optimization problem (3.9) with fixed* R_k *in the constraint (3.9c) has either* $w^c_{l,k} = 0$*, or* $w^d_{l,k} = 0$*, or both for all* $l \in \mathcal{L}, k \in \mathcal{K}$*.*

Proof. The proof is relegated to Appendix C. □

3.2 Optimization Methodology

The problem (3.9) involves joint optimization of beamforming vector for compression and data-sharing signals $(\{w^c_{l,k}, w^d_{l,k}\})$ (and as a consequence the combined beamformers $\{w_{l,k}\}$), the quantization noise levels $\{q_l\}$ for the compression signal, and the BS clustering for data-sharing (and thus compression), i.e., the decision of which users should data-shared and which users should be compressed for which BSs. In general, the problem is hard as it combines the difficulties with both the individual data-sharing and compression strategies bundled together.

Before we give the joint optimization procedure to solve (3.9), we consider a heuristic procedure that separates the above optimization variables to illustrate where the benefit in the hybrid strategy can come from. To this end, we first obtain the combined network-wide beamformers $\{w_{l,k}\}$ without any backhaul constraints. Then assuming only compression strategy, i.e., $\{w^c_{l,k} = w_{l,k}, w^d_{l,k} = 0\}$, the quantization noise levels $\{q_l\}$ are optimized with these fixed beamformers with the backhaul constraints taken into account. Next, in an iterative manner, we strategically explicitly select the most suitable user for direct data-sharing with some BS, i.e. for some (BS, user) pair (l, k), we make $w^c_{l,k} = 0, w^d_{l,k} = w_{l,k}$, and then re-optimize the quantization noise levels for the remaining compressed part using the modified backhaul capacity. We continue this procedure until no additional users can benefit from data-sharing, instead of being included in the compressed signal. The overall procedure is summarized in Algorithm 4. We describe the components in more detail below.

Algorithm 4 Heuristic design for the hybrid strategy

```
1. Design combined network-wide beamformers $\{w_{l,k}\}$, ignoring the backhaul constraints, using, for example, the WMMSE approach or regularized zero-forcing;

2. Assuming only compression strategy, set $w^c_{l,k} = w_{l,k}, w^d_{l,k} = 0$, $l \in \mathcal{L}, k \in \mathcal{K}$ and optimize the quantization noise levels $\{q_l\}$ taking into account the backhaul constraints and obtain the user rates $\{R_k\}$;

3. Use Algorithm 5 to select users for direct data-sharing.
```

The optimization of the combined network-wide beamformers can be done using the WMMSE approach as discussed in Section 2.2 by ignoring the backhaul constraints. We assume full per-antenna power could be utilized to design these beamformers. Then assuming compression only strategy, optimization of the quantization noise levels with fixed beamforming vectors and the given backhaul capacity constraints can be done by utilizing the relation (3.2) and using the approach discussed in Section 2.3. Note that since we started with beamformers with maximum per-antenna budget allowed, after adding the quantization noise levels, some power constraints might be violated. Thus the initial beamformers may have to be solved again by reducing the maximum power allowed by the amount of the quantization noises. This process may be need to be iterated until a feasible power allocation with the quantization noises is found after Step 2 of Algorithm 4. Finally we now improve upon the initial user rates obtained with compression only strategy, by allowing the data for a subset of users to be sent to the BSs directly through the backhaul links.

To select users for direct data transfer, we compare, for each user, the backhaul capacity required for sending its message directly, with the reduction in backhaul in compressing the rest of the signal if that user is dropped from compression. To illustrate this more precisely, recall the amount of backhaul capacity needed to compress the precoded signal $\hat{x}_l^c$ for BS l to within quantization noise level q_l is approximately $\log\left(\frac{\hat{P}_l}{q_l}\right)$, where $\hat{P}_l = \Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2$. Let $\hat{P}_{l,k} = \Gamma_q |w_{l,k}^c|^2$. If we instead send the data for, say user k, directly to BS l, the signal that needs to be compressed now has smaller power $\hat{P}_l - \hat{P}_{l,k}$. So to compress it to within the same quantization noise level q_l, approximately $\log\left(\frac{\hat{P}_l - \hat{P}_{l,k}}{q_l}\right)$ bits are needed instead. Now, the backhaul capacity required to send the data of user k to BS l is just its achievable rate, namely, R_k. Thus, data-sharing is beneficial for user k on BS l whenever R_k is less than the saving in the quantization bits, or equivalently

$$\log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k > 0. \tag{3.10}$$

This criterion is used to select users for data-sharing. Note that we keep the combined beamforming coefficient for the pair the same, and just move the value from compression beamformer to data-sharing beamformer, i.e., we make $w_{l,k}^d = w_{l,k}, w_{l,k}^c = 0$. Once a user is selected for message sharing, the quantization noise levels for the compressed part are re-optimized with the modified backhaul capacity. Note that the modified backhaul capacity constraint depends on the rate of the selected user, which is a function of the quantization noise levels to be optimized. Hence, we need to iteratively optimize the quantization noise levels assuming fixed rate for that user from the previous iteration, then update the rates, and continue until the rates converge. Note also that the new quantization noise levels obtained also affect the power constraints. However, such effects are small and can be neglected. Algorithm 5 summarizes the user selection procedure for data-sharing based on the criterion (3.10). We use a greedy approach to look for the user which can provide the best improvement in backhaul utilization, then continue the process until no more users would result in further improvement.

Joint Optimization

In this section, we describe the joint optimization methodology to solve the problem (3.9). The main source of difficulty is the constraint (3.9c). The first term is the indicator

Algorithm 5 User selection for data-sharing for the heuristic design

```
Set $n_k = 0, k \in \mathcal{K}$; set $C_{\text{temp}} = C$;
Set $g_{l,k} = \log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k, \quad l \in \mathcal{L}, k \in \mathcal{K}$;
Set $g = \max_{l,k}\{g_{l,k}\}$;
**while** $g > 0$ **do**
    Set $(\hat{l}, \hat{k}) = \arg\max g_{l,k}$ for message sharing;
    Set $\hat{P}_{\hat{l}} = \hat{P}_{\hat{l}} - \hat{P}_{\hat{l},\hat{k}}$; $\hat{P}_{\hat{l},\hat{k}} = 0$; $n_k = n_k + 1$;
    **repeat**
        Set $C = C_{\text{temp}} - \sum_{k=1}^{K} n_k R_k$, and optimize quantization noise levels $\{q_l\}$;
        Update user rates $R_k$;
    **until** user rates converge
    Set $g_{l,k} = \log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k, \quad l \in \mathcal{L}, k \in \mathcal{K}$;
    Set $g = \max_{l,k}\{g_{l,k}\}$;
**end while**
```

function accounting for backhaul consumption due to direct data-sharing, along with the user rate R_k that is also part of the objective function. The second term with the log function in the compression part is a non-convex function of the variables $(\{w^c_{l,k}\}, \{q_l\})$. For the indicator function, as before we approximate it as a weighted l_1 norm as

$$\mathbb{1}\left\{|w^d_{l,k}|^2\right\} = \left\||w^d_{l,k}|^2\right\|_0 \approx \beta^d_{l,k}|w^d_{l,k}|^2, \tag{3.11}$$

where $\beta^d_{l,k}$ is a constant weight associated with BS l and user k and is updated iteratively in an outer loop according to

$$\beta^d_{l,k} = \frac{1}{|w^d_{l,k}|^2 + \tau}, \tag{3.12}$$

for some regularization constant $\tau > 0$ and $|w^d_{l,k}|^2$ from the previous iteration. Similarly, R_k in the constraint (3.9c) is kept fixed from the previous iteration, denoted by $\hat{R}_k$, and is updated in the same outer loop. This simplifies the constraint (3.9c) to

$$\sum_{k=1}^{K} \beta^d_{l,k}\hat{R}_k|w^d_{l,k}|^2 + \log\left(1 + \frac{\Gamma_q \sum_{k=1}^{K}|w^c_{l,k}|^2}{q_l}\right) \le C_l, \quad l \in \mathcal{L}. \tag{3.13}$$

Next, we rewrite the log function in the above constraint (3.13) into sum of two terms as follows:

$$\sum_{k=1}^{K} \beta_{l,k}^{d} \hat{R}_k |w_{l,k}^{d}|^2 + \log\left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^{c}|^2 + q_l\right) - \log(q_l) \leq C_l, \quad l \in \mathcal{L}. \tag{3.14}$$

Thus we need to solve the optimization problem (3.9c) with the constraint (3.9c) modified as (3.14). In this formulation, in the constraint (3.14), $-\log(q_l)$ is a convex function of $\{q_l\}$, but $\log\left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^{c}|^2 + q_l\right)$ is a non-convex function of $(\{w_{l,k}\}, \{q_l\})$. Additionally the objective function (3.9a) is a non-convex function of $(\{w_{l,k}\}, \{q_l\})$. In order to solve the optimization problem (3.9c) with the modified constraint (3.14), we use the iterative successive convex approximation method by linearizing the non-convex part in both the objective and the constraint in an inner loop. First, we transform the objective into a suitable form, by utilizing the relationship between the achievable rate and the MSE. The MSE for user k is defined as

$$e_k = |u_k|^2 \left(\Gamma_m \left(\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k\right) + |\mathbf{h}_k^H \mathbf{w}_k|^2\right) - 2\ \mathrm{Re}\{u_k^H \mathbf{h}_k^H \mathbf{w}_k\} + 1. \tag{3.15}$$

under a receive beamformer u_k. The rate R_k can then be written as

$$R_k = \max_{u_k} \log\left(e_k^{-1}\right). \tag{3.16}$$

Second, to deal with the non-convexity of the log function in the transformed objective function (3.16) and the modified constraint (3.14), we find the appropriate tight convex upper bounds and successively update them. We make use of the following result.

Lemma 3.2.1. *For any positive $x, x_0 \in \mathbb{R}$, $\log x \leq \log x_0 + \frac{1}{x_0} x - 1$, with equality if and only if $x = x_0$.*

We make successive convex approximations to $\log(e_k)$ and $\log\left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^{c}|^2 + q_l\right)$

as follows.

$$\log\left(e_k\right) \leq -\log\left(\rho_k\right) + \rho_k e_k - 1, \tag{3.17}$$

where

$$\rho_k = e_k^{-1}, \tag{3.18}$$

with e_k as defined in (3.15), and $(\{w_{l,k}\}, \{u_k\})$ taken from the previous iteration in an iterative manner in the inner loop. Similarly

$$\log\left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2 + q_l\right) \leq -\log\left(\gamma_l\right) + \gamma_l \left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2 + q_l\right) - 1, \tag{3.19}$$

where

$$\gamma_l = \left(\Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2 + q_l\right)^{-1}, \tag{3.20}$$

with $(\{w_{l,k}^c\}, \{q_l\})$ taken from previous iteration in the inner loop.

Note the similarity of the update (3.18), in the convex upper bound (3.17) for the objective function, to the MSE weight update in Chapter 2 (e.g., (2.14)) used in the iterative algorithm used to solve the equivalent WMMSE problem. The two are in fact related. Another way of looking at the the iterative algorithm for the equivalent WMMSE problem is exactly what we have done above for the objective function. We successively upper bound the log function in the rate R_k after writing it as a function of the transmit and receive beamformers as in (3.16), and then update the convex upper bound in successive block updates in the transmit and receive beamformers. The MSE weights are the multiplying factors in the convex approximations at each step.

Thus, in the end, we iteratively solve the following programs with alternating block updates in the inner loop for fixed $\beta_{l,k}$ and $\hat{R}_k$, and then update $\beta_{l,k}$ according to (3.12) and $\hat{R}_k$ as the modified R_k in the outer loop, as discussed when simplifying the original constraint (3.9c) to (3.13).

- The optimal receive beamformer u_k under fixed $\{\mathbf{w}_k, \mathbf{w}_k^c, \mathbf{w}_k^d\}$ and $\{q_l\}$ is given by

$$u_k = \left(\Gamma_m \left(\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k \right) + |\mathbf{h}_k^H \mathbf{w}_k|^2 \right)^{-1} \mathbf{h}_k^H \mathbf{w}_k. \tag{3.21}$$

- Under fixed $\{u_k\}$, the optimal transmit beamformers $\{\mathbf{w}_k, \mathbf{w}_k^c, \mathbf{w}_k^d\}$ and the optimal quantization noise levels $\{q_l\}$ are obtained by solving the following convex program:

$$\underset{\substack{\{w_{l,k}^d\},\{w_{l,k}^c\},\\ \{w_{l,k}\},\{q_l\}}}{\text{minimize}} \quad \sum_{k=1}^{K} -\alpha_k \rho_k e_k \tag{3.22a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} |w_{l,k}|^2 + q_l \leq P_l, \quad l \in \mathcal{L} \tag{3.22b}$$

$$\sum_{k=1}^{K} \beta_{l,k}^c \hat{R}_k |w_{l,k}^d|^2 + \gamma_l \Gamma_q \sum_{k=1}^{K} |w_{l,k}^c|^2 + \gamma_l q_l - \log(q_l) \leq C_l', \quad l \in \mathcal{L} \tag{3.22c}$$

$$w_{l,k}^d + w_{l,k}^c = w_{l,k}, \quad l \in \mathcal{L}, k \in \mathcal{K}. \tag{3.22d}$$

where $C_l' = C_l + \log(\gamma_l) + 1$.

The overall algorithm for the joint optimization of the problem (3.9) for the hybrid strategy is summarized in Algorithm 6.

Note that the optimization framework and methodology developed in the previous sections can be easily extended to the case with multiple antennas at the BSs and the user terminals, but the main challenge is the computational complexity of the resulting algorithm. We leave the work on developing a low complexity algorithm for the case of multi-antenna terminals for future.

Algorithm 6 WSR maximization for the hybrid strategy

```
Initialization: $\{\mathbf{w}_k, \mathbf{w}_k^c, \mathbf{w}_k^d\}, \{q_l\}, \{\beta_{l,k}^d\}, \{\hat{R}_k\}$;
Repeat:
    1. Repeat:
        (a) For fixed $\{\mathbf{w}_k, \mathbf{w}_k^c, \mathbf{w}_k^d\}, \{q_l\}$, compute the optimal receivers $\{u_k\}$ according to (3.21) and the corresponding MSE $\{e_k\}$ according to (3.15);
        (b) Update the weights $\{\rho_k\}$ according to (3.18);
        (c) Update the weights $\{\gamma_l\}$ according (3.20), for fixed $\{\mathbf{w}_k^c\}, \{q_l\}$;
        (d) For fixed $\{u_k\}$, $\{\rho_k\}$, and $\{\hat{R}_k\}$ in (3.22c), find the optimal transmit beamformers $\{\mathbf{w}_k, \mathbf{w}_k^c, \mathbf{w}_k^d\}$ by solving (3.22);
       Until convergence
    2. Update $\{\beta_{l,k}\}$ as in (3.12);
    3. Compute the achievable rates $\{R_k\}$ according to (3.8). Update $\hat{R}_k = R_K$, $k \in \mathcal{K}$.
Until convergence
```

3.3 Numerical Evaluation

We first consider the a 7-cell wrapped around two-tier heterogeneous network considered in Section 2.4. Each of the terminals is equipped with a single antenna. We compare the hybrid strategy designed with the joint optimization done by Algorithm 6, with the data-sharing and compression strategies, optimized with explicit per-antenna and per-BS backhaul constraints using Algorithm 1 and Algorithm 2 respectively.

The Fig. 3.2 shows the average sum rate as a function of total backhaul capacities across the 7-cell network for the three strategies. For low backhaul capacity, we observe that the data-sharing performs better than the compression strategy. In this case, the hybrid performs just as good. We observe that almost all the users in the final beamformer for the hybrid strategy are data-shared. The reason why hybrid is slightly worse than the data-sharing, even though the data-sharing optimization framework is a special case of the hybrid optimization framework, is because of the log factor in the constraint (3.9c). The quantization noises do not exactly go down to zero because of the numerical

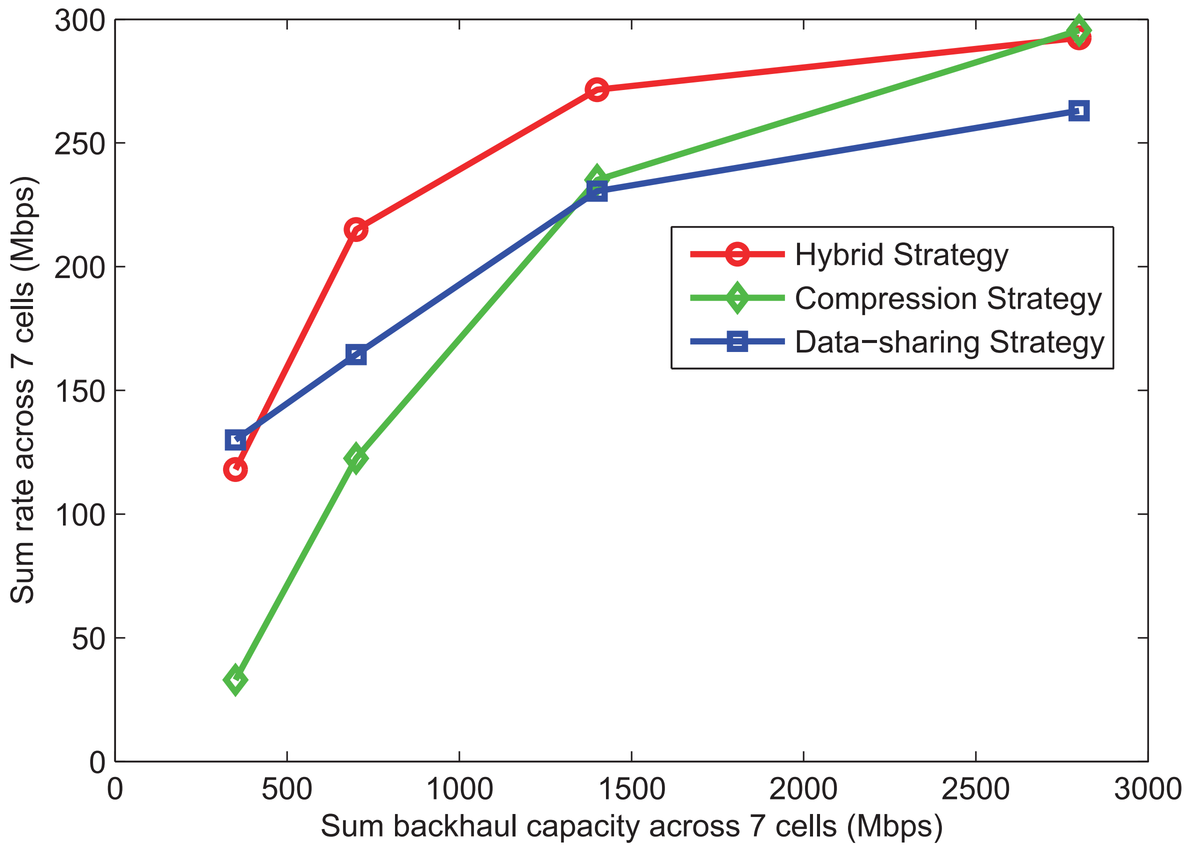

Figure 3.2: Comparison of the hybrid strategy with the data-sharing and compression strategies.

issues. For moderate backhaul capacity, the data-sharing and compression strategies are comparable. This is the regime where hybrid strategy has some potential to provide benefits by having some users participate in data-sharing and rest in the compression. When the backhaual capacity is high, the compression strategy starts to outperform the data-sharing strategy. The hybrid strategy shows some improvement in this regime and the gains tend diminish as we increase the backhaul even further as the rates saturate to the maximum sum rate of the system. Thus overall we see that the hybrid strategy achieves the best of the two strategies under low and high backhaul capacities, and when the backhaul capacities are moderate, there is some benefit from the hybrid design.

For second set of simulations, we present the simulation results showing benefits of

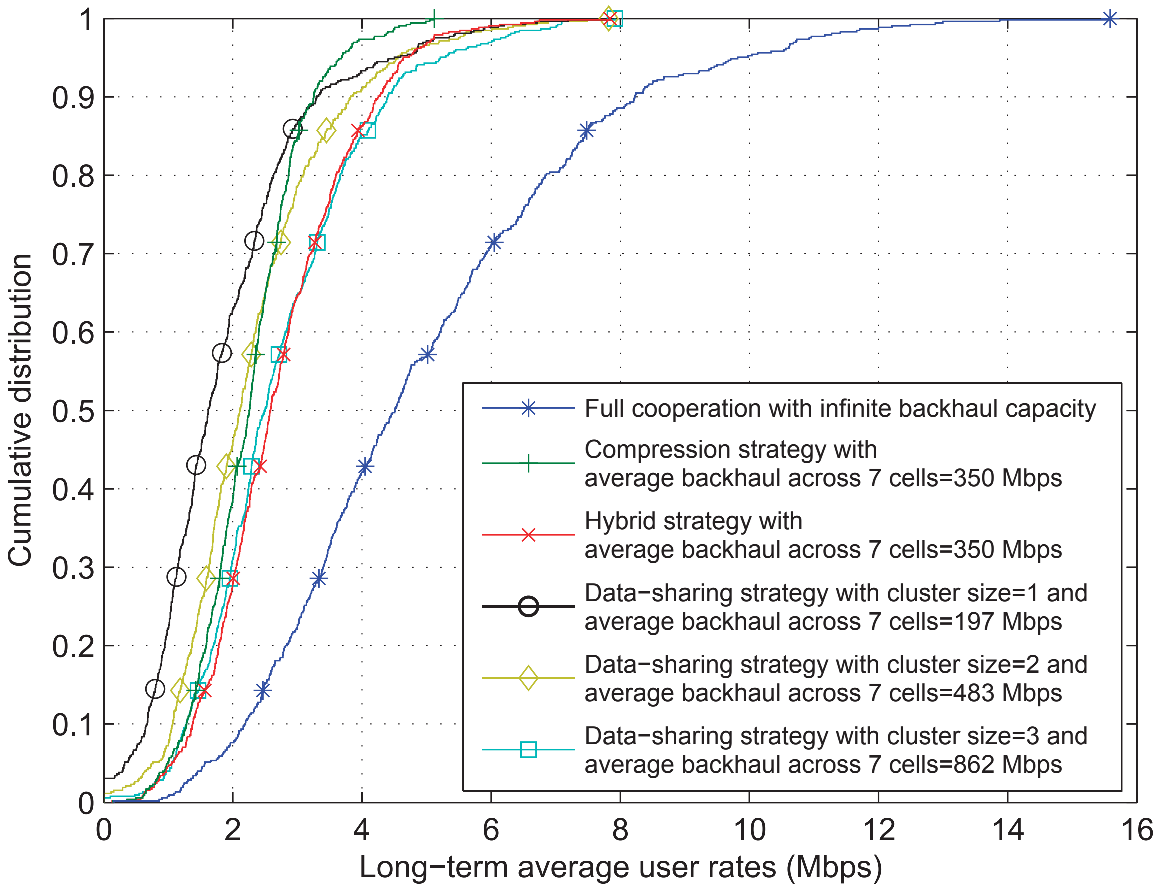

Figure 3.3: Comparison of cumulative distribution functions of user rates for the data-sharing, compression, and hybrid strategies.

the hybrid strategy done with the heuristic design. For simplicity, we first consider a homogenous 7-cell network with 15 users randomly located in each cell. Users are scheduled in a round-robin fashion with one active user scheduled per cell at any given time. The BS-to-BS distance is set at 0.8km, and the noise power spectral density is -162dBm/Hz. The channels from the BSs to the users are generated according to a distance-dependent path-loss model $\text{PL(dB)} = 128.1 + 37.1 \log 10(d)$ with 8dB log-normal shadowing and a Rayleigh fading component, where d is the distance between the BS to the user in km. Perfect channel estimation is assumed, and the CSI is made available to all the BSs and to the centralized processor. A total bandwidth of 10 MHz is assumed.

For algorithmic tractability for designing beamformers for data-sharing part and optimization of quantization noise levels for compression part, a sum power constraint and a sum backhaul constraint over 7 BSs is adopted for this comparison. The reason for such network simplification is that in the heuristic strategy, once a user is selected for data-sharing, its rate is subtracted from the backhaul capacity based on the current user rates and the quantization noises are re-optimized until the rates converge, it is difficult maintain the constraints satisfied in this iterative heuristic process. The average power spectral density at each BS antenna is maintained at -27dBm/Hz. For comparison purposes, for the data-sharing strategy, we fix the cooperation cluster size for each user, picking the strongest BSs according to channel strength, and use the WMMSE approach of Section 2.2 for designing the beamformers. The backhaul capacity consumed is calculated once the user rates are determined. For the compression strategy, the first two steps of Algorithm 4 are performed. The hybrid strategy is done using the heuristic design according to Algorithm 4.

Fig. 3.3 shows the cumulative distribution function of the user rates for the three schemes. In the simulation, weighted sum rate maximization is used as the optimization objective with weights updated according to proportional fairness criterion. It can be seen that both the compression strategy and the hybrid strategy significantly outperform the data-sharing scheme at this backhaul capacity. This is partly also because, for the data-sharing strategy, we fix the cooperation cluster. In particular, the hybrid scheme with 350Mbps backhaul achieves about the same user rates as the data-sharing scheme with 862Mbps, which represents a saving in backhaul capacity by about 60%. Further, the hybrid scheme is also seen to outperform the compression strategy, improving the rate of the 50th percentile user by about 10% at the same backhaul capacity. Thus we see that the hybrid strategy has some potential to provide gains over the individual strategies under moderate backhaul capacity.

3.4 Summary

In this chapter, we propose a hybrid strategy that combines compression-based signaling and data-sharing. Such an approach gives better control the utilization of the backhaul capacity. We propose an optimization framework that generalizes the individual data-sharing and compression strategies, and jointly optimizes the beamformers, user selection for the data-sharing and compression components, and the quantization noise levels for the compressed signals. We then numerically compare the performance of the hybrid strategy with the individual data-sharing and compression strategies. Our main conclusion is that when the backhaul capacity is low, it is better to only do pure data-sharing, and when the backhaul capacity is high, it is preferable to do pure compression. But when the backhaul capacity is moderate, there is some scope for doing the hybrid combination of the two.

Chapter 4

Conclusion

C-RAN has emerged as a promising solution for the next generation wireless cellular networks due to its potential to mitigate intercell interference by means of joint cooperative signal processing at the central processor. The main challenge to the realization of the gains promised by the C-RAN architecture, however, hinges on the effective use of the backhaul links, which in practice are often capacity-limited. This thesis studies different transmission strategies for the downlink C-RAN with limited backhaul capacity, and investigates how the limited backhaul capacity affects the design and system-level performance of these strategies.

First, we compare two fundamentally different strategies, the compression strategy, which is the standard solution for C-RAN, and the data-sharing strategy, which is the traditional implementation in most cellular systems. Our main conclusion is that backhaul capacity constraint is crucial in deciding which strategy to adopt for the downlink C-RAN. If the available backhaul capacity is medium-to-high, the compression strategy outperforms the data-sharing strategy, even with a simple fixed-rate uniform scalar quantizer, due to the possibiilty to have large cooperation cluster at the central processor, whereas using data-sharing, the cluster size is limited by the backhaul capacity. However, if the available backhaul capacity is low, the data sharing strategy outperforms the compression strategy. Under low backhaul capacity the quantization noises introduced in the compression strategy dominate the interference, in which case it is better to just share the data directly with a limited set of BSs rather than to compress. When we also take into account the CSI bottleneck in the network, the performance of the compression strategy suffers more than the data-sharing strategy. This is because, the gain in the compression strategy stems from the possibility to form large cooperation clusters at the central processor, which is affected more compared to the data-sharing strategy, which already has a smaller cluster size due to the limited backhaul capacity.

Next, we propose to combine the data-sharing and compression strategies into a hybrid scheme that can benefit from the advantages of both strategies. Such hybrid combination results in flexibility in terms of backhaul utilization. The optimization framework proposed for the hybrid strategy generalizes both individual strategies. When the backhaul capacity is low, the hybrid strategy reduces to primarily all data-sharing and when the backhaul capacity is high, it reduces to almost all compression. But when the backhaul capacity is moderate, we observe that the system performance can be improved by having the data for some of the users transmitted directly to the BSs and rest of the users compressed using the remaining backhaul capacity. Having the flexibility to switch between data-sharing and compression depending on the available backhaul capacity at different BSs is especially useful in the future dense cellular networks with different tiers of BSs, with different levels of backhaul capacities.

4.1 Future Work

In the compression strategy considered in this thesis, we study the quantization model where the central processor performs independent quantization of the signals intended for each antenna at each BS. There is also a possibility of doing joint compression for the signals intended for different antennas of the *same* BSs. Further, as proposed in [16], it is also possible to do multivariate compression to introduce correlation among the quantization noises of *different* BSs. Such correlation helps by potentially lowering the total quantization noises at the end user after going through the channel between the BSs and the user. It would be interesting to see how the comparison between data-sharing and compression is affected by these quantization models. The main challenge in designing algorithm with such models is the computational complexity. For example, for the multivariate compression in [16], mere evaluation of the quantization rate region grows exponentially with the number of BSs. It would be beneficial to look for ways to reduce this complexity by an alternate rate region characterization.

Although we looked at the effect of partial CSI, in terms of clustered CSI, on the performance of different strategies, it would be interesting to look at the tradeoff between conveying good partial CSI or approximate full CSI. One way to consider such a tradeoff would be account for CSI transfer in the backhaul capacity usage. Apart from CSI, the feasibility of joint cooperative signal processing also depends crucially upon the ability of the BSs to precisely synchronize with each other. How imperfect synchronization affects the performance is also a question of practical importance for the implementation of different cooperative strategies considered in this thesis.

Finally, on a theoretical front, it is of interest to provide a foundation to the transmission strategies considered so far for the downlink C-RAN. In that regard, the first step would be to attempt to characterize the approximate capacity of the downlink C-RAN setup. A future direction would be to try to specialize the achievable rate region of the recently proposed distributed decode-forward scheme of [13] to the downlink C-RAN setup with a good choice of auxiliary random variables.

Appendices

Appendix A

Proof of Proposition 2.2.1

First we show that the WSR maximization problem (2.10) and the WMMSE problem (2.13) have the same global solutions. To show this, first observe that in the WMMSE problem, the optimization variables $\{\mathbf{u}_k\}$ and $\{\rho_k\}$ are unconstrained. Thus by first order optimality conditions, we can easily show that

$$\mathbf{u}_k^\star = \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I}\right) + \mathbf{H}_k \mathbf{w}_k \mathbf{w}_k^H \mathbf{H}_k^H\right)^{-1} \mathbf{H}_k^H \mathbf{w}_k, \quad k \in \mathcal{K} \tag{A.1}$$

$$\rho_k^\star = e_k^{-1}, \quad k \in \mathcal{K} \tag{A.2}$$

as fixing other variables, the objective is a convex function of the variable $\{u_k\}$ and $\{\rho_k\}$. Thus substituting $\{\mathbf{u}_k^\star\}$ and $\{\rho_k^\star\}$ in the optimization problem (2.13), we get the following equivalent problem:

$$\begin{aligned} \underset{\mathbf{w}_{l,k}}{\text{maximize}} \quad & \sum_{k=1}^{K} \alpha_k \log\left(e_k^{-1}\right) && \text{(A.3a)} \\ \text{subject to} \quad & \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} && \text{(A.3b)} \\ & \sum_{k=1}^{K} \beta_{l,k} \hat{R}_k \left\|\mathbf{w}_{l,k}\right\|^2 \leq C_l, \quad l \in \mathcal{L}. && \text{(A.3c)} \end{aligned}$$

By comparing the optimization problem (A.3) with the WSR maximization problem (2.10), all it remains is to show that $\log\left(e_k^{-1}\right)$ is same as R_k, $k \in \mathcal{K}$, which we show below.

By substituting the value of $\mathbf{u}_k^\star$ into e_k and simplifying, we get

$$e_k = 1 - \mathbf{w}_k^H \mathbf{H}_k^H \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) + \mathbf{H}_k \mathbf{w}_k \mathbf{w}_k^H \mathbf{H}_k^H \right)^{-1} \mathbf{H}_k \mathbf{w}_k, \quad k \in \mathcal{K}. \tag{A.4}$$

Now we apply the Woodbury matrix identity to get

$$e_k^{-1} = 1 + \mathbf{w}_k^H \mathbf{H}_k^H \left(\Gamma_m \left(\sum_{j \neq k} \mathbf{H}_k \mathbf{w}_j \mathbf{w}_j^H \mathbf{H}_k^H + \sigma^2 \mathbf{I} \right) \right)^{-1} \mathbf{H}_k \mathbf{w}_k, \quad k \in \mathcal{K}, \tag{A.5}$$

which is exactly the $(1 + \text{SINR}_k)$ from (2.4). Hence we see that the two optimization problems (2.10) and (2.13) have the same global solution.

Now we show that the two optimization problems also have the same set of stationary points by equating the KKT conditions for the two problems.

Note that, since the optimization problem (2.13) is unconstrained in the variables $\{\mathbf{u}_k\}$ and $\{\rho_k\}$, and the objective function is convex in each of these variables when other variables are fixed, for any stationary point $(\{\mathbf{w}_k^\star\}, \{\mathbf{u}_k^\star\}, \{\rho_k^\star\})$ of problem (2.13), the values of stationary points $(\{\mathbf{u}_k^\star\}, \{\rho_k^\star\})$ are as given by (A.1) and (A.2) respectively. Observe that the constraint set for both problems are the same. Hence, it is sufficient to only check the first order optimality conditions. Now writing the respective Lagrangian functions of both the problems and taking their partial derivatives with respect to $\{\mathbf{w}_k\}$, we get

$$\frac{\partial \mathcal{L}_{\text{WSR}}}{\partial \mathbf{w}_k} = -\sum_{k=1}^{K} \alpha_k \frac{\partial R_k}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \lambda_l^m \left(\sum_{k=1}^{K} |w_{l,k}^m|^2 - P_l^m \right) \right)}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \mu_l \left(\sum_k \beta_k^l \hat{R}_k \|\mathbf{w}_k^l\|_2^2 - C_l \right) \right)}{\partial \mathbf{w}_k} = 0, \tag{A.6}$$

$$\frac{\partial \mathcal{L}_{\text{WWMSE}}}{\partial \mathbf{w}_k} = \sum_{k=1}^{K} \rho_k \frac{\partial e_k}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \lambda_l^m \left(\sum_{k=1}^K |w_{l,k}^m|^2 - P_l^m\right)\right)}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \mu_l \left(\sum_k \beta_k^l \hat{R}_k \|\mathbf{w}_k^l\|_2^2 - C_l\right)\right)}{\partial \mathbf{w}_k} = 0, \tag{A.7}$$

where $\lambda_l^m \in \mathbb{R}$, $\mu_l \in \mathbb{R}$, $l \in \mathcal{L}, m \in \mathcal{M}$ denote the dual variables associated with the per-antenna power constraints and per-BS backhaul constraints respectively. The minus sign in the first term of (A.6) results after writing the original problem (2.10) as a minimization problem. Using the fact that $\{\mathbf{u}_k^\star\}$ is given by (A.1) and thus R_k is $\log(e_k^{-1})$, as proved above, we can re-write the partial derivative (A.6) as

$$\frac{\partial \mathcal{L}_{\text{WSR}}}{\partial \mathbf{w}_k} = \sum_{k=1}^{K} \frac{\alpha_k}{e_k} \frac{\partial e_k}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \lambda_l^m \left(\sum_{k=1}^K |w_{l,k}^m|^2 - P_l^m\right)\right)}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \mu_l \left(\sum_k \beta_k^l \hat{R}_k \|\mathbf{w}_k^l\|_2^2 - C_l\right)\right)}{\partial \mathbf{w}_k} = 0. \tag{A.8}$$

Since $\rho_k^\star = e_k^{-1}$, $k \in \mathcal{K}$, we see that for any stationary point $(\{\mathbf{w}_k^\star\}\{\mathbf{u}_k^\star\}, \{\rho_k^\star\})$ of (2.13) which is a solution to (A.7), the point $(\{\mathbf{w}_k^\star\})$ also satisfies (A.6), and vice versa. Hence the problems (2.13) and (2.10) are equivalent in the sense of stationary points.

Appendix B

Proof of Proposition 2.3.1

The proof follows along the same lines as that of the proof for the Proposition 2.2.1 in Appendix A. The main difference is the presence of the new set of variables ($\{q_l\}$), corresponding to quantization noise levels. As we see below, all relations hold even with these new variables. First we show that the WSR maximization problem (2.25) and the WMMSE problem (2.27) have the same global solutions. By first order optimality conditions, we can easily show that

$$u_k^\star = \left(\Gamma_m \left(\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k \right) + |\mathbf{h}_k^H \mathbf{w}_k|^2 \right)^{-1} \mathbf{h}_k^H \mathbf{w}_k \tag{B.1}$$

$$\rho_k^\star = e_k^{-1}, \quad k \in \mathcal{K}. \tag{B.2}$$

Substituting $\{u_k^\star\}$ and $\{\rho_k^\star\}$ in the optimization problem (2.27), we get the following equivalent problem:

$$\begin{aligned}
\underset{\mathbf{w}_{l,k}}{\text{maximize}} \quad & \sum_{k=1}^{K} \alpha_k \log\left(e_k^{-1}\right) && \text{(B.3a)}\\
\text{subject to} \quad & \sum_{k=1}^{K} |w_{l,k}^m|^2 \leq P_l^m, \quad l \in \mathcal{L}, m \in \mathcal{M} && \text{(B.3b)}\\
& \sum_{k=1}^{K} \beta_{l,k} \hat{R}_k \left\|\mathbf{w}_{l,k}\right\|^2 \leq C_l, \quad l \in \mathcal{L}. && \text{(B.3c)}
\end{aligned}$$

Below we verify that $\log\left(e_k^{-1}\right)$ is the same as R_k, $k \in \mathcal{K}$.

By substituting the value of $u_k^\star$ into e_k and simplifying, we get

$$e_k = 1 - \left(\Gamma_m \left(\sum_{j\neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k\right) + |\mathbf{h}_k^H \mathbf{w}_k|^2\right)^{-1} |\mathbf{h}_k^H \mathbf{w}_k|^2, \quad k \in \mathcal{K}. \quad \text{(B.4)}$$

Applying the Woodbury matrix identity, after simplification we get

$$e_k^{-1} = 1 + \frac{|\mathbf{h}_k^H \mathbf{w}_k|^2}{\Gamma_m \left(\sum_{j\neq k} |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k\right)}, \quad k \in \mathcal{K}, \qquad \text{(B.5)}$$

which is exactly the term inside the log expression for rate R_k in (2.21). Hence we see that the two optimization problems (2.25) and (2.27) have the same global solution.

Now we show that the two problems also have the same set of stationary points by equating their KKT conditions. Since the constraints set is same for both problems, we only need to check the first order optimality conditions.

Partial derivatives of the Lagrangian functions of both problems with respect to $\{\mathbf{w}_k\}$ and $\{q_l\}$ are as follows:

$$\frac{\partial \mathcal{L}_{\text{WSR}}}{\partial \mathbf{w}_k} = -\sum_{k=1}^{K} \alpha_k \frac{\partial R_k}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \lambda_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - P_l\right)\right)}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \mu_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\, q_l\right)\right)}{\partial \mathbf{w}_k} = 0. \quad \text{(B.6)}$$

$$\frac{\partial \mathcal{L}_{\text{WSR}}}{\partial q_l} = -\sum_{k=1}^{K} \alpha_k \frac{\partial R_k}{\partial q_l} + \frac{\partial \left(\sum_l \lambda_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - P_l \right) \right)}{\partial q_l} + \frac{\partial \left(\sum_l \mu_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\, q_l \right) \right)}{\partial q_l} = 0. \tag{B.7}$$

$$\frac{\partial \mathcal{L}_{\text{WWMSE}}}{\partial \mathbf{w}_k} = \sum_{k=1}^{K} \rho_k \frac{\partial e_k}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \lambda_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - P_l \right) \right)}{\partial \mathbf{w}_k} + \frac{\partial \left(\sum_l \mu_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\, q_l \right) \right)}{\partial \mathbf{w}_k} = 0. \tag{B.8}$$

$$\frac{\partial \mathcal{L}_{\text{WWMSE}}}{\partial q_l} = \sum_{k=1}^{K} \rho_k \frac{\partial e_k}{\partial q_l} + \frac{\partial \left(\sum_l \lambda_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - P_l \right) \right)}{\partial q_l} + \frac{\partial \left(\sum_l \mu_l \left(\sum_{k=1}^{K} |w_{l,k}|^2 - \frac{2^{C_l}-1}{\Gamma_q}\, q_l \right) \right)}{\partial q_l} = 0. \tag{B.9}$$

Here $\lambda_l \in \mathbb{R}$, $\mu_l \in \mathbb{R}$, $l \in \mathcal{L}$ denote the dual variables associated with the per-antenna power constraints and per-BS backhaul constraints respectively. As done in Appendix A, it is easy to verify that the above equations for the problem (2.25) are same as that for the problem (2.27), by noticing that R_k is $\log(e_k^{-1})$ under $\{\mathbf{u}_k^\star\}$ and $\rho_k^\star = e_k^{-1}$, $k \in \mathcal{K}$. Thus for any stationary point $(\{\mathbf{w}_k^\star\}, \{q_l\}, \{\mathbf{u}_k^\star\}, \{\rho_k^\star\})$ of (2.13) which is a solution to (B.8) and (B.9), the point $(\{\mathbf{w}_k^\star\}, \{q_l\})$ also satisfies (B.6) and (B.7), and vice versa. Hence the problems (2.25) and (2.27) are equivalent in the sense of stationary points.

Appendix C

Proof of Proposition 3.1.1

We prove the proposition using proof by contradiction.

We assume that, contrary to the claim, there is at least one pair of BS l and user k such that the corresponding beamforming coeffient for both the data-sharing and the compression are non-zero, i.e., there is a pair (l, k) of BS and user with compression beamforming coefficient $w_{l,k}^c = c$ and data-sharing beamforming coefficient $w_{l,k}^d$ such that $|c|^2 \neq 0$ and $|d|^2 \neq 0$. Note that both c and d are complex numbers. We now prove that we can always produce another feasible point that has strictly better objective value and thus the assumed solution can not be globally optimal. Additionally, we can also produce a feasible point very close to the assumed solution that gives a strictly better objective value and thus the assumed solution can not be a stationary point of the optimization problem as well. Below we give the new feasible points that contradict the assumption.

To contradict the global optimality, the new point that has strictly better objective value has $w_{l,k}^c = 0$ and $w_{l,k}^d = c + d$. First, with the new point, the total beamforming coefficient for the BS l and user k is still the same, $w_{l,k} = w_{l,k}^c + w_{l,k}^d = c + d$. Thus the total beamforming constraints (3.22d) and the per-antenna power constraints (3.22b) remain unchanged. For the constraint (3.9c) for BS l, since the indicator in first term that accounts for the data-sharing backhaul was non-zero before (since $|d|^2 \neq 0$) and R_k

is assumed to be fixed, the backhaul contribution from that term can not increase. While in the second term in the constraint (3.9c) that accounts for compression backhaul, the the value of $\sum_{k=1}^{K} |w_{l,k}^{c}|^2$ is now decreased by $|c|^2$. Thus the new point along with the old $\{q_l\}$ is feasible. Further, in the rate R_k given by (3.7) the signal power (the numerator) and the interference (first term in the denominator) remain the same. Observe that the rate is a strictly decreasing function of $\{q_l\}$. With the previous q_l and new $(w_{l,k}^{c}, w_{l,k}^{d})$, there is still some slack in the constraint (3.9c) because of the reduction in the numerator of the second term accounting for compression backhaul capacity. Hence we can further decrease q_l, still satisfy the constraint, and get strictly better value for the objective function. Note that the increased rates do not change other constraints as the rate R_k in the constraints (3.9c) is assumed to be fixed. Now to produce a new point that contradicts the stationarity, we take $w_{l,k}^{c} = c - \epsilon$ and $w_{l,k}^{d} = d + \epsilon$, where ϵ is a small complex number such that $|c - \epsilon|^2 < |c|^2$. It is easy to show that such ϵ exists. The rest of the argument follows along the same lines as above for the global optimality, by first showing that new point is feasible, and then showing that it can result in a strictly better objective value by lowering the value of q_l.

Bibliography

[1] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Aurelie Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and Wei Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[3] Aleksandra Checko, Henrik L Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, Michael S Berger, and Lars Dittmann, "Cloud RAN for mobile networks - A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 2014.

[4] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2001.

[5] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming inter-cell interference in indoor wireless systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2007, pp. 83–87.

[6] Abbas El Gamal and Young-Han Kim, *Network information theory*, Cambridge university press, 2011.

[7] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[8] S. Kannan, A. Raja, and P. Viswanath, "Approximately optimal wireless broadcasting," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7154–7167, 2012.

[9] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[10] A.S. Avestimehr, S.N. Diggavi, and D.N.C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.

[11] Sung Hoon Lim, Young-Han Kim, Abbas El Gamal, and Sae-Young Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, 2011.

[12] Thomas M Cover and Abbas El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[13] Sung Hoon Lim, Kwang Taik Kim, and Young-Han Kim, "Distributed decode-forward for broadcast," in *Inf. Theory Workshop (ITW)*. IEEE, 2014, pp. 556–560.

[14] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Nov. 2009.

[15] O. Simeone, O. Somekh, H. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Advances Singal Process.*, Feb. 2009.

[16] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.

[17] Boon Loong Ng, J.S. Evans, S.V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.

[18] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6102–6111, Dec. 2011.

[19] P. Marsch and G. Fettweis, "On base station cooperation schemes for downlink network MIMO under a constrained backhaul," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Nov. 2008.

[20] Sheng Jing, David NC Tse, Joseph B Soriaga, Jilei Hou, John E Smee, and Roberto Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wireless Commun. and Netw.*, vol. 2008, pp. 18, 2008.

[21] Bobak Nazer and Michael Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[22] Song-Nam Hong and Giuseppe Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, 2013.

[23] Jiening Zhan, Bobak Nazer, Uri Erez, and Michael Gastpar, "Integer-forcing linear receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7661–7685, 2014.

[24] Qingjiang Shi, M. Razaviyayn, Zhi-Quan Luo, and Chen He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331 –4340, Sept. 2011.

[25] Song-Nam 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.

[26] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.

[27] S.S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792 –4799, Dec. 2008.

[28] S. Kaviani, O. Simeone, W.A. Krzymien, and S. Shamai, "Linear precoding and equalization for network MIMO with partial cooperation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2083–2096, Jun. 2012.

[29] Toby Berger, "Rate-distortion theory," *Encyclopedia of Telecommunications*, 1971.

[30] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.