

Hybrid Compression and Message-Sharing Strategy for the Downlink Cloud Radio-Access Network

Pratik Patil and Wei Yu

Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario M5S 3G4, Canada
Emails: {ppatil, weiyu}@comm.utoronto.ca

Abstract—Cloud radio-access network (CRAN) is a wireless cellular network architecture in which the base-stations are connected to a cloud-computing based central processor with digital backhaul links. In this setting, the base-stations can be thought of as relays between the central processor and the user terminals. This paper considers the downlink of a cloud radio-access network with finite-capacity backhaul links. We model the overall network as a broadcast-relay channel, and propose a hybrid compression and message-sharing strategy for the downlink transmission. This paper further proposes numerical techniques for optimizing the quantization noise level in the compression part of the hybrid precoding scheme. We compare the system-level performance of the proposed scheme with the pure compression and the pure message sharing schemes to show the benefit of the hybrid strategy for the downlink CRAN.

I. INTRODUCTION

As modern wireless cellular networks move toward full spectrum reuse with progressively smaller cell sizes, the mitigation and management of intercell interference have become increasingly important tasks. Cloud radio-access network (CRAN) is a promising future cellular architecture in which the BSs are connected via high-speed digital backhaul links to centralized cloud computing servers, where joint encoding and decoding of user signals can take place. By allowing coordination and joint signal processing across multiple base-stations (BSs) in the network, the CRAN architecture enables the implementation of network multiple-input multiple-output (MIMO) or coordinated multi-point (CoMP) concepts [1]–[3]; it has the potential to bring significant improvement to future wireless cellular services.

This paper studies the downlink transmission in a CRAN setting. In the CRAN architecture, the user data originate from the centralized cloud server and are destined for the mobile devices distributed throughout a geographical area, whereas the BSs act as *relays* between the user terminals and the cloud. In this sense, the downlink CRAN can be modeled as a broadcast-relay channel. Although the capacity analysis for this setting is straightforward if the backhaul links between the cloud processor and the BSs have infinite capacity, the more realistic case, where the backhaul is constrained to have *finite* capacity, is much more involved. The main contribution of this paper is a novel transmission strategy for the downlink CRAN that allows better utilization of the finite-capacity backhaul.

The interference mitigation capability of CRAN stems from its ability to jointly encode the user messages from 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 *message-sharing* in this paper, can be thought of as analogous to the decode-and-forward relaying strategy. As the sharing of each user's message across the entire network would require excessively large amount of backhaul capacity, practical implementation of message-sharing often involves clustering, where each user selects a subset of cooperating BSs. Various ways to selectively share the user messages have been investigated in the literature [4], [5]. Information theoretic results for the downlink network MIMO model using the message-sharing strategy have also been reported in [6] for a modified linear Wyner cellular model, and in [7] for a two-BS, two-user setup.

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 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 *pure compression* in this paper, is akin to the compress-and-forward relaying strategy. This approach has been investigated in [6], also in [8] where the effect of imperfect channel state information (CSI) is taken into account, and more recently in [9], where the benefit of joint compression across multiple BSs is explored.

The main objective of this paper is to show that in a practical CRAN setting with finite backhaul capacity, instead of pure compression or pure message-sharing, a hybrid scheme that combines the two can bring significant benefit to the overall system performance. This paper proposes an approach where the central processor directly sends messages for some of the users to the BS along with the compressed version of rest of the precoded signal. The intuition is that it is beneficial, in terms of backhaul capacity utilization, to send clean message for strong users while compressing rest of the interference canceling signals. To quantify the benefit of this hybrid strategy, this paper also proposes convex optimization based method for optimizing the quantization noise levels for the compressed part of the precoded signal.

This paper restricts attention to linear precoding strategies, but we mention here, as related work, that possibilities exist for performing nonlinear precoding based on dirty-paper coding [6], and for using lattice-coding based strategy based on compute-and-forward [10] for the downlink CRAN.

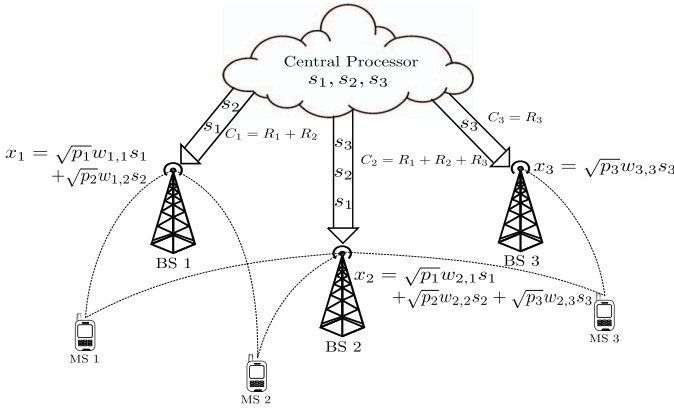


Fig. 1. Example of message-sharing cooperation scheme in downlink CRAN.

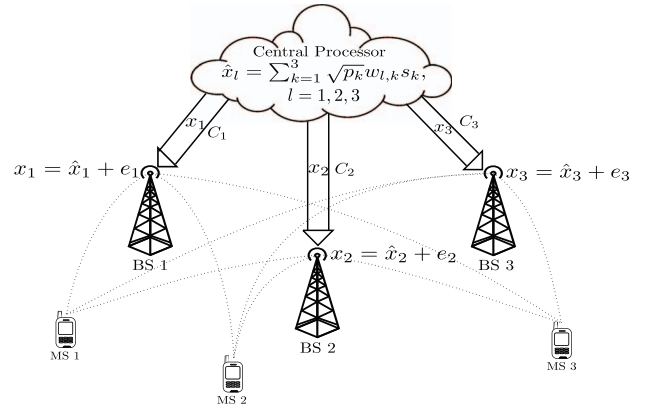


Fig. 2. Compression-based cooperation scheme for the downlink CRAN.

II. SYSTEM MODEL

This paper considers a downlink CRAN with L single-antenna BSs serving K single-antenna remote users. Each of the L BSs is connected to a central processor with a capacity-limited digital backhaul link. A sum capacity constraint is imposed so that the total capacity over the L backhaul links is limited to C bits per channel use. The sum-capacity backhaul constraint is adopted here for convenience. It can model the scenario where the backhaul is implemented in a shared (e.g., wireless) medium.

An independent data stream is transmitted from the central processor to each user. Let x_l be the signal transmitted by BS l . The received signal at user k can be written as

$$y_k = \mathbf{h}_k^H \mathbf{x} + z_k, \quad k = 1, 2, \dots, K \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^{L \times 1} = [x_1, \dots, x_L]^T$ is the aggregate signal from the L BSs, $\mathbf{h}_k \in \mathbb{C}^{L \times 1} = [h_{1,k}, \dots, h_{L,k}]^T$ is the channel from the L BSs to the user k , and z_k is the additive zero-mean Gaussian noise with variance σ^2 . In addition, each BS l has a power constraint P_l so that

$$\mathbb{E}|X_l|^2 \leq P_l, \quad l = 1, 2, \dots, L. \quad (2)$$

Given the setup, this paper seeks to find the optimal encoding and transmission schemes at the central processor and at the BSs that maximize the weighted sum rate of the overall network. For convenience, fixed user scheduling is assumed in this paper. In addition, perfect CSI is assumed to be available both at the central processor and at all the BSs.

III. EXISTING APPROACHES

A. Message Sharing Scheme

Message sharing refers to the cooperation scheme in which the central processor distributes the actual message of each user to its cooperating BSs through the backhaul links. Each BS then forms a precoded signal based on all the user messages available to it, as shown in Fig. 1.

Let s_k be the message signal for user k , assumed to be complex Gaussian with zero-mean and unit variance. Let the normalized beamforming vector from all the BSs to user k be

$\mathbf{w}_k \in \mathbb{C}^{L \times 1} = [w_{1,k}, w_{2,k}, \dots, w_{L,k}]^T$, where $w_{l,k}$ denotes the component of the beamformer at BS l . (If BS l does not participate in cooperatively transmitting to user k , then $w_{l,k} = 0$.) The transmitted vector signal $\mathbf{x}$ from all the BSs can be written as

$$\mathbf{x} = \sum_{k=1}^K \sqrt{p_k} \mathbf{w}_k s_k, \quad (3)$$

where p_k is the power of beam $\mathbf{w}_k$. At the receiver, the signal-to-noise-interference-ratio (SINR) for user k can be expressed as

$$\text{SINR}_k = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2}. \quad (4)$$

The achievable rate for user k is then

$$R_k = \log(1 + \text{SINR}_k). \quad (5)$$

The question of which subset of BSs should serve each user is in general nontrivial. For comparison purpose, this paper uses the following common heuristics for evaluating the achievable rates using the message-sharing scheme, wherein each user forms a cooperating cluster consisting of S BSs with the strongest channels. Under a fixed BS cooperation structure, locally optimal beamformers for maximizing the weighted sum rate subject to BS power constraints can be found using the weighted minimum mean square error (WMMSE) approach [11]. The total amount of backhaul required to support this message-sharing scheme can be calculated based on the achieved user rates multiplied by the number of BSs serving each user.

B. Pure Compression Scheme

In the compression based scheme, the functionality of precoding is completely migrated to the central processor, as shown in Fig. 2. The central processor performs joint encoding of the user messages and forms the analog signals intended to be transmitted by the BSs' antennas. As the precoded signals are analog, they need to be compressed before they can be forwarded to the corresponding BSs through the finite-capacity backhaul links. Compression introduces quantization noises.

The quantization noise level is a function of the backhaul capacity.

Let $\hat{\mathbf{x}} \in \mathbb{C}^{L \times 1} = [\hat{x}_1, \dots, \hat{x}_L]^T$ denote precoded signals intended for BSs 1 to L , which is formed using the beamformers for users 1 to K , i.e., $\mathbf{w}_k \in \mathbb{C}^{L \times 1} = [w_{1,k}, w_{2,k}, \dots, w_{L,k}]^T$ with power p_k :

$$\hat{\mathbf{x}} = \sum_{k=1}^K \sqrt{p_k} \mathbf{w}_k s_k. \quad (6)$$

We denote the power of $\hat{x}_l$ as $\hat{P}_l$. The quantization process for $\hat{\mathbf{x}}$ can be modeled as

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}, \quad (7)$$

where $\mathbf{e}$ is the quantization noise with covariance $\mathbf{Q} \in \mathbb{C}^{L \times L}$ modelled as a Gaussian process and assumed to be independent of $\hat{\mathbf{x}}$. In this case, the received SINR for user k is

$$\text{SINR}_k = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + |\mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k|}. \quad (8)$$

The achievable rate for user k is again $R_k = \log(1 + \text{SINR}_k)$.

For simplicity, this paper assumes independent quantization at each BS, in which case $\mathbf{Q}$ is a diagonal matrix with diagonal entries q_l . (The more general joint compression strategy has been studied in [9].) Assuming ideal quantizer, the quantization noise level q_l and the backhaul capacity C_l are related as

$$\log \left(\frac{\hat{P}_l + q_l}{q_l} \right) \leq C_l. \quad (9)$$

The design of the pure-compression strategy can now be stated as a weighted sum rate maximization problem over the transmit beamformers and the quantization noise levels as follows:

$$\begin{aligned} & \underset{\mathbf{w}_k, p_k, q_l, C_l}{\text{maximize}} && \sum_{k=1}^K \mu_k R_k \\ & \text{subject to} && \log \left(\frac{\hat{P}_l + q_l}{q_l} \right) \leq C_l, \quad 1 \leq l \leq L \\ & && \sum_{l=1}^L C_l \leq C \\ & && \hat{P}_l + q_l \leq P_l, \quad 1 \leq l \leq L \\ & && C_l \geq 0, \quad 1 \leq l \leq L. \end{aligned} \quad (\text{P1})$$

where R_k and $\hat{P}_l$ are both functions of the underlying variables $\mathbf{w}_k, p_k$. Although finding the globally optimal solution to the above problem is a challenging task, an iterative approach based on majorization minimization has been suggested in [9]. In later section of this paper, we also suggest a different method for solving one component of (P1) based on convex optimization.

IV. HYBRID COMPRESSION AND MESSAGE SHARING

In the message-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 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 noise.

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 power 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. This hybrid scheme is illustrated in Fig. 3.

We point out that a dirty-paper coding based scheme proposed in [6] 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 [6] 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.

V. DESIGN METHODOLOGY FOR THE HYBRID SCHEME

We now describe a design methodology for the hybrid compression and message-sharing strategy. The optimization of the hybrid strategy involves the choice of beamforming vector $\mathbf{w}_k$, power p_k , the quantization noise levels q_l , and more importantly the decision of which users should participate in message sharing and which users in compression. To make the

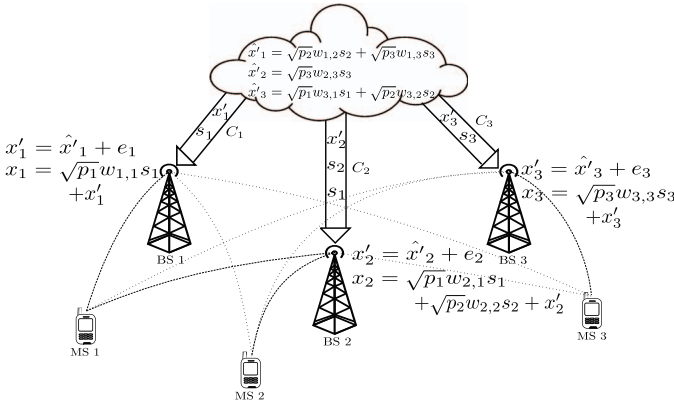


Fig. 3. Example of hybrid compression and message-sharing scheme.

overall problem tractable, in this paper, we fix the network-wide beamformers throughout, and begin the design process with an optimized pure compression scheme. At each iteration of the algorithm, we strategically select the most suitable user for message sharing, then re-optimize the quantization noise levels for the remaining compressed part. We continue this procedure until no additional users can benefit from message sharing instead of being included in the compressed signal.

The overall algorithm is described in Algorithm 1 below. Each of its components are described in more detail in subsequent sections.

Algorithm 1 Hybrid Compression and Message Sharing

- 1: Design fixed network-wide beamformers using, for example, the regularized zero-forcing approach;
 - 2: Assuming pure compression, optimize the quantization noise level in each backhaul link, obtain the user rates;
 - 3: Use Algorithm 2 to select users for message sharing.
-

A. Choose Fixed Network Beamformers

For simplicity, we fix the network beamformers for precoding the user signals over the multiple BSs. We describe here an approach based on regularized zero-forcing beamforming. The beamformers can also be chosen in different ways, for example using the zero-forcing or the WMMSE approach.

The direction for the beamformer of user k , $\mathbf{w}_k$, is chosen to be $\frac{\mathbf{t}_k}{\|\mathbf{t}_k\|}$ for $\mathbf{t}_k \in \mathbb{C}^{L \times 1}$ where

$$[\mathbf{t}_1, \dots, \mathbf{t}_K] = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + \alpha \mathbf{I})^{-1}, \quad (10)$$

$\mathbf{H} \in \mathbb{C}^{K \times L} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^H$, $\mathbf{I}$ is a $K \times K$ identity matrix, and α is a regularization factor. The regularization factor α and the powers p_k associated with each beam are chosen as follows. First, we approximate the SINR for each user by ignoring the residual interference from the other users. Then for a fixed α , the powers p_k associated with each beam can be chosen to maximize the weighted sum rate by solving the following convex optimization problem subject to the per-BS

power constraints:

$$\begin{aligned} & \underset{p_k}{\text{maximize}} && \sum_{k=1}^K \mu_k \log \left(1 + \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sigma^2} \right) \\ & \text{subject to} && \sum_{k=1}^K p_k |\mathbf{w}_{l,k}|^2 \leq P_l, \quad 1 \leq l \leq L \\ & && p_k \geq 0, \quad 1 \leq k \leq K. \end{aligned} \quad (\text{P2})$$

To find the appropriate regularization constant α , we can set it heuristically depending on SNR, or solve (P2) for different α 's and pick the one that maximizes the weighted sum rate.

B. Optimize Pure Compression Scheme

We start with the pure compression strategy, and use the following method for finding the optimal quantization noise level and the resulting achievable user rates with pure compression. This is akin to solving the optimization problem (P1). But instead of using the method in [9], we assume here that the beamformers $\mathbf{w}_k$ and the powers p_k are fixed, and optimize over the quantization noise levels at each BS q_l , or equivalently C_l , as follows:

$$\begin{aligned} & \underset{C_l}{\text{maximize}} && \sum_{k=1}^K \mu_k \log(1 + \text{SINR}'_k) \\ & \text{subject to} && \sum_{l=1}^L C_l \leq C \end{aligned} \quad (\text{P3})$$

where

$$\text{SINR}'_k = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \sum_{l=1}^L \frac{\hat{P}_l |\mathbf{h}_{l,k}|^2}{2^{C_l} - 1}}. \quad (11)$$

The main idea here is to reformulate the problem in terms of C_l by the substitution $q_l = \frac{\hat{P}_l}{2^{C_l} - 1}$. A key observation is that the resulting optimization problem (P3) becomes convex in C_l (assuming fixed p_k and $\mathbf{w}_k$), which allows efficient numerical solution. The proof of concavity is omitted here for brevity.

The variable $\hat{P}_l$ above denotes the power of $\hat{x}_l$ to be compressed, and is assumed to be a constant in the SINR expression (11). Ideally, $\hat{P}_l$ should be set as close to the BS power constraint P_l as possible. But if we set it exactly equal to P_l , after adding quantization noise, the resulting power of the signal transmitted by BS l would exceed the power constraint. For simulation purpose, we start with $\hat{P}_l = P_l$ and decrement $\hat{P}_l$ by the quantization noise level q_l after the optimization. This process may need to be iterated until a feasible power allocation satisfying $\hat{P}_l + q_l \leq P_l$ is found.

C. Greedy User Selection for Message Sharing

We now improve upon the initial user rates obtained with pure compression by allowing the messages for a subset of users to be sent to 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 that we compress the precoded signal $\hat{x}_l = \sqrt{p_1}w_{l,1}s_1 + \sqrt{p_2}w_{l,2}s_2 + \dots + \sqrt{p_K}w_{l,K}s_K$ for BS l . The amount of backhaul needed to compress $\hat{x}_l$ to within quantization noise level q_l is approximately $\log\left(\frac{\hat{P}_l}{q_l}\right)$, where $\hat{P}_l = p_1|w_{l,1}|^2 + p_2|w_{l,2}|^2 + \dots + p_K|w_{l,K}|^2$. Let $\hat{P}_{i,j} = p_j|w_{i,j}|^2$. If we instead send the message for, say user k , directly, 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 message of user k to BS l is just its achievable rate, namely, R_k . Thus, message sharing is beneficial for user k on BS l whenever R_k is less than the saving in the quantization bits, or equivalently

$$\log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k > 0. \quad (12)$$

This criterion is used to select users for message sharing.

Once a user is selected for message sharing, we re-optimize the quantization noise levels for the compressed part of the signals for each BS again by solving (P3) with a modified total backhaul constraint and modified power constraint. 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 solve (P3) assuming fixed rate for that user from the previous iteration, then update the rate and continue the process until the rate converges. Note also that the new quantization noise levels obtained from re-solving (P3) also affect the power constraint. However, such effects are small and can be neglected.

Algorithm 2 summarizes the user selection process for message sharing based on the criterion (12). 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.

Algorithm 2 Select Users for Message Sharing

```

Set  $n_k = 0, \forall k$ ; set  $C_{\text{temp}} = C$ ;
Set  $g_{l,k} = \log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k, \quad \forall(l, k)$ 
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 solve (P3)
        Update user rates  $R_k$ ;
    until user rates converge
    Set  $g_{l,k} = \log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k, \quad \forall(l, k)$ 
    Set  $g = \max_{l,k} \{g_{l,k}\}$ ;
end while

```

VI. SIMULATION RESULTS

We first present simulation results comparing pure compression, pure message-sharing, and the hybrid schemes for

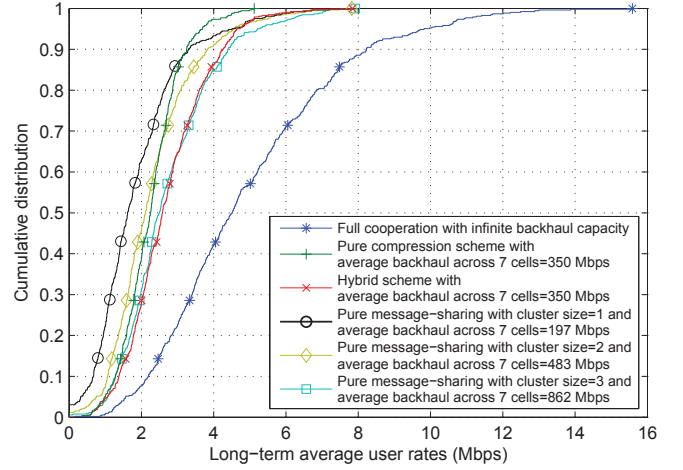


Fig. 4. Comparison of cumulative distribution functions of user rates for the message sharing, pure compression, and hybrid schemes in a 7-cell network.

a 7-cell CRAN 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 message sharing, a sum power constraint over 7 BSs is adopted so that the average power spectral density at each BS antenna is -27dBm/Hz . For the pure message sharing scheme, we adopt the design methodology of Section III-A, namely fixing cooperation cluster size for each user, picking the BSs according to channel strength, and using the WMMSE approach [11] for designing beamformers. The backhaul capacity is calculated once the user rates are determined. For compression and hybrid schemes, we adopt the design methodology of Section V, except the initial network-wide beamformers are chosen using the WMMSE approach with full cooperation over 7 cells.

Fig. 4 shows the cumulative distribution function (CDF) 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 clearly seen that both the pure compression and the hybrid schemes significantly outperform the message-sharing scheme. In particular, the hybrid scheme with 350Mbps backhaul achieves about the same user rates as the message-sharing scheme with 862Mbps, which represents a saving in backhaul capacity by about 60%. Further, the hybrid scheme is also seen to outperform the pure compression-based scheme, improving the rate of the 50th percentile user

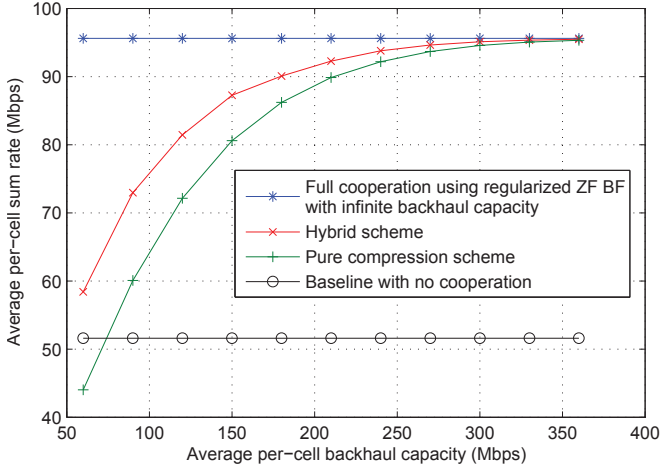


Fig. 5. Per-cell sum rate vs. average per-cell backhaul capacity for the hybrid scheme as compared to the pure compression scheme in a 19-cell topology with center 7 cells forming a cooperating cluster.

by about 10% at the same backhaul.

In a second set of simulations, we consider a larger network with 19 cells wrapper around, 3 sectors per cell, and 10 users randomly located in each sector. The central 7 BSs (i.e., the central 21 sectors) form a cooperation cluster; the out-of-cluster interference produced by the rest of BSs is taken into account. Here, we impose a more realistic per-BS power constraint equivalent to -27dBm/Hz over 10MHz, and use regularized zero-forcing with per-BS power constraint to find the initial beamformers in compression and hybrid designs.

Fig. 5 shows the average per-cell sum rate of the proposed hybrid scheme as compared to the compression based scheme as a function of average per-cell backhaul capacity. As can be seen from the figure, the hybrid scheme improves backhaul utilization as compared to the compression scheme. The improvement is prominent when the backhaul capacity is small and the gap decreases as the backhaul capacity increases. The maximum achievable rate with infinite backhaul capacity using regularized zero-forcing beamforming and the no-cooperation baseline are also plotted for reference. It can be seen that at an operating point of 85Mbps per-cell sum rate, which is about 90% of the full cooperation rate, the hybrid scheme requires a backhaul of 150Mbps, while the pure compression scheme requires 180Mbps. Thus, the hybrid scheme achieves a saving of about 20% in backhaul capacity requirement.

In order to visualize the improvement in network utility, we fix the total backhaul capacity to be 150Mbps and 90Mbps and plot the CDF of user rates of the compression and the hybrid schemes in Fig. 6. The hybrid scheme is seen to improve over the pure compression scheme mostly for high-rate users.

VII. CONCLUSION

This paper studies the downlink transmission in CRAN where the BSs are connected to a cloud processor with finite capacity backhaul links. We propose a hybrid strategy that combines compression-based signaling and message sharing,

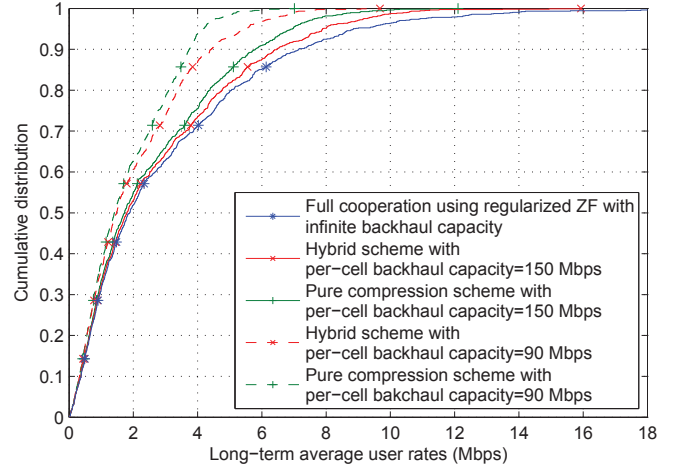


Fig. 6. Comparison of cumulative distribution functions of user rates for the hybrid scheme and the pure compression scheme in a 19-cell topology with center 7 cells forming a cooperating cluster.

which results in better utilization of the backhaul. We propose a design procedure for selecting the appropriate users for message sharing, and an efficient method for the optimization of the quantization noise levels for the compressed part of the signals. Numerical simulations confirm the performance gains of the proposed strategy over the existing approaches.

REFERENCES

- [1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. 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.
- [2] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, May 2001.
- [3] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2007, pp. 83–87.
- [4] B. L. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.
- [5] 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.
- [6] O. Simeone, O. Somekh, H. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Advances Singal Process.*, Feb. 2009.
- [7] 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.
- [8] —, "On downlink network mimo under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE Global Commun. Conf. (GlobeCom)*, Nov. 2009.
- [9] 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.
- [10] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sept. 2013.
- [11] S. Kaviani, O. Simeone, W. 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.