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

ITA 2014

Cloud Radio-Access Network (CRAN)

- CRAN architecture: Base-stations (BSs) are connected to a centralized cloud-computing based processor via backhaul links.
- Motivation: To enable joint multi-cell processing [Gesbert et al., 2010] for effective interference management.
- Uplink: A virtual multiple access channel with BSs as relays. Central processor can perform *joint processing* of user signals.
- Downlink: A broadcast relay channel. Central processor can perform *joint encoding* of user messages.
- This talk is about transmission strategies in the downlink CRAN.

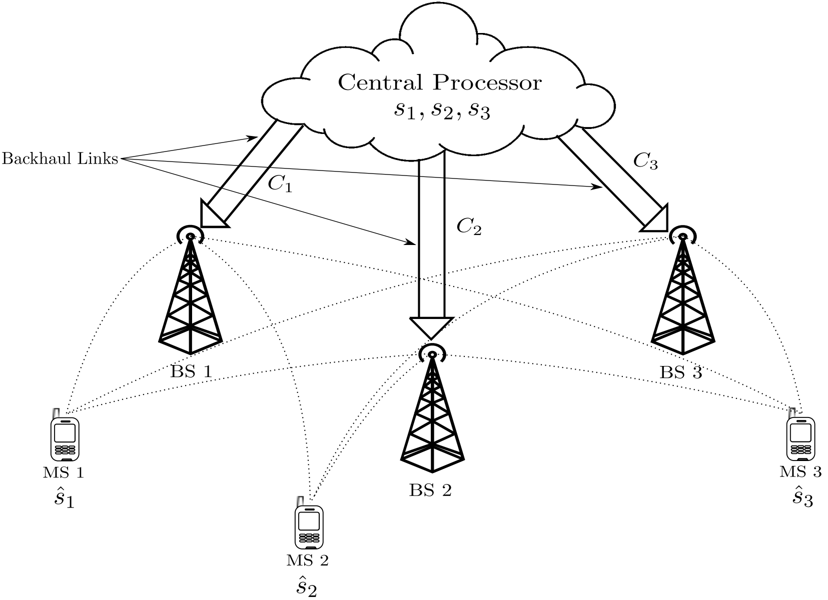

Figure: Illustration of the CRAN downlink

Information Theoretic Problem Setup

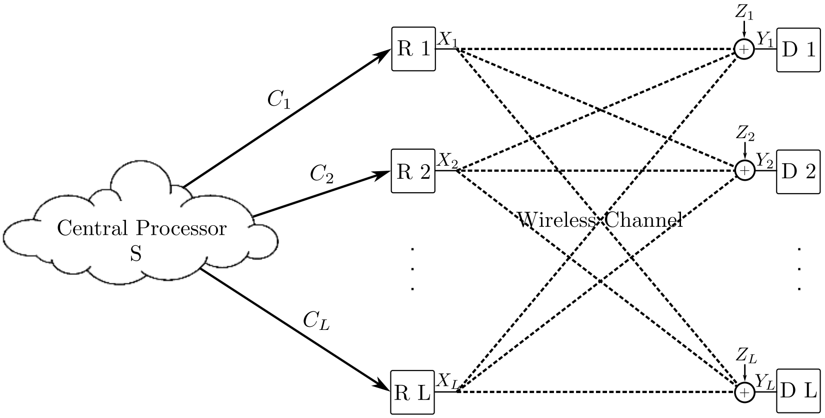

Figure: CRAN downlink

Downlink CRAN

- Data originate from the central processor and is designed for mobile users throughout the geographical area via BSs acting as relays.
- If the capacity of backhaul links is infinite, the downlink trivially reduces to a broadcast channel. Capacity known for Gaussian case.
- But, realistically we have finite capacity of the backhaul links. Capacity analysis challenging for this practical case.
- This talk proposes on a novel transmission strategy for the downlink CRAN with *limited backhaul*.

Previous Work

- User messages (and precoding information) is communicated to BSs via backhaul links.
  - Called as (pure) message-sharing in this talk. Analogous to 'decode-and-forward'.
  - Efficient ways to limit the data transfer studied in, e.g., [Ng et al., 2008, Zakhour and Gesbert, 2011, Zhao et al., 2013].
  - Information theoretic results reported in [Marsch and Fettweis, 2009] for a simplified two-BS, two-user setup.
- Joint precoding performed at the central unit and compressed signals are communicated to BSs (oblivious of codebooks used for users).
  - Called as (pure) compression scheme in this talk. Analogous to 'compress-and-forward'.
  - Dirty paper coding (DPC) performed at the central unit, followed by independent compression investigated in [Simeone et al., 2009].
  - Similar quantized signal based cooperation scheme considered in [Marsch and Fettweis, 2008].
  - Benefits of multivariate compression (instead of independent compression) studied in [Park et al., 2013].

Previous Work

- Downlink counter-part of compute-and-forward (CoF) scheme of [Nazer et al., 2009].
  - Reverse-compute-and-forward strategy is proposed in [Hong and Caire, 2013] where the roles of BSs and users are reversed in CoF. Procoded messages (over finite fields) are transmitted through backhaul links. High sensitivity to channel coefficients, non-integer penalty.
  - Generalization of zero-forcing beamforming based on integer-forcing ([Zhan et al., 2010]) and RCoF is also studied in [Hong and Caire, 2013] to avoid the non-integer penalty at the cost of SNR penalty (incurred due to non-unitary precoding to force effective channel with integers). Broadly, falls under compression-based scheme under finite backhaul (as the resultant precoded signals are real signals as opposed to RCoF).

Problem Statement

- Consider a network-MIMO system with L single-antenna BSs serving K single-antenna users.
- BSs are connected to the central processor via limited backhaul links with total sum capacity limit C bits per channel use.
- All user messages available at the central processor.
- Received signal at user k is $y_k = \mathbf{h}_k^H \mathbf{x} + z_k$ where
  - $\mathbf{x} = [x_1, \cdots, x_L]^T$ is the aggregate signal from the L BSs
  - $\mathbf{h}_k = [h_{1,k}, \cdots, h_{L,k}]^T$ is the channel from the L BSs to the user k
  - z_k is the additive zero-mean Gaussian noise with variance σ^2
- BS l has a power constraint P_l.
- Fixed user scheduling. CSI known to the central processor and all the BSs.
- Objective: What is the optimal encoding and transmission schemes at the central processor and at the BSs that maximize the weighted sum rate of the overall network?

Existing Approach I: Message-Sharing

- One possible way for cooperation is to directly share user messages to BSs through backhaul links. BSs then encode the messages to form the signals to be transmitted. We refer to this as 'message-sharing'.
- Advantage: BSs receive clean copies of user messages.
- Limitation: Due to limited backhaul available, each BS get messages for only a subset of users, resulting in partial cooperation.
- Various ways to select user clusters for limited cooperation have been suggested in literature, e.g,
  [Ng et al., 2008, Zakhour and Gesbert, 2011].

Message-Sharing Basic Setting

- Transmitted signal $\mathbf{x}$ from all BSs is $\mathbf{x} = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{w}_k s_k$ where
  - s_k: zero-mean unit-variance Gaussian signal for user k.
  - $\mathbf{w}_k = [w_{1,k}, \ldots, w_{L,k}]^T$: beamforming direction for user k from L BSs. If BS l does not participate in cooperatively transmitting to user k, $w_{l,k} = 0$.
  - p_k: power of beam $\mathbf{w}_k$.
- Signal-to-noise-interference-ratio at user k, $\text{SINR}_k = \frac{p_k|\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2}$.
- Achievable rate for user k is $R_k = \log(1 + \text{SINR}_k)$.

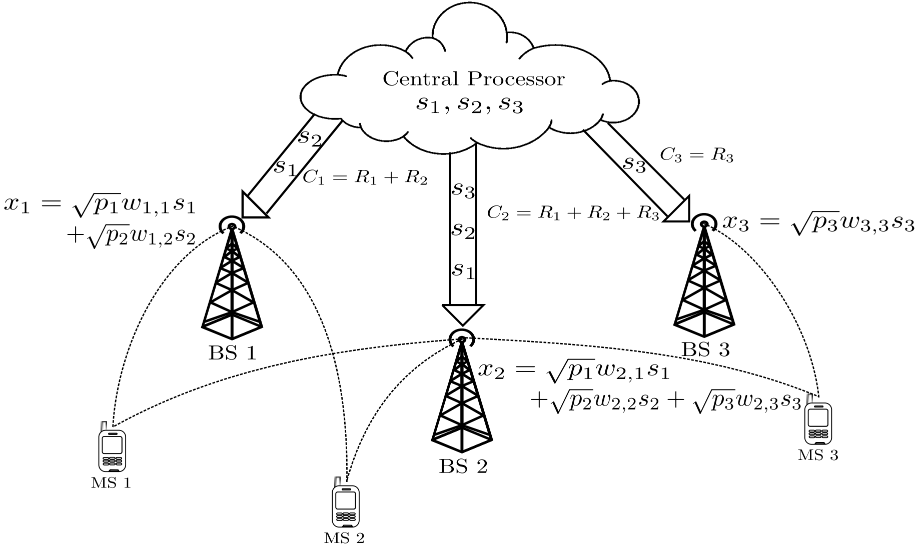

Figure: Example of message-sharing cooperation scheme in downlink CRAN.

Message-Sharing Optimization

- Deciding which subset of BSs should serve each user is nontrivial. For comparison, we consider a cooperating cluster consisting of S BSs for each user with the strongest channels.
- For 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 [Kaviani et al., 2012].
- Total backhaul required can be calculated based on the achieved user rates multiplied by the number of BSs serving each user.

Existing Approach II: Compression-Based Scheme

- The central processor performs joint encoding of the user messages and forms signals intended to be transmitted by the BSs' antennas.
- As the precoded signals are analog, they are compressed and forwarded to the corresponding BSs via finite-capacity backhaul links.
- Advantage: 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. Also, BSs can be completely oblivious of the user codebooks as the burden of preprocessing is shifted from the BSs to the central processor.
- Limitation: Compression introduces quantization noises which can be large for small backhaul capacity.
- This approach has been investigated in [Simeone et al., 2009], [Marsch and Fettweis, 2009], and more recently in [Park et al., 2013].

Compression-Based Scheme Basic Setting

- Precoded signals intended for BSs formed at central processor, $\hat{\mathbf{x}} = [\hat{x}_1, \cdots, \hat{x}_L]^T = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{w}_k s_k$
  - $\mathbf{w}_k$ is normalized beamforming vector for user k
  - p_k is the beampower of $\mathbf{w}_k$
  - Let power of $\hat{x}_l$ be $\hat{P}_l$
- Quantization for $\hat{\mathbf{x}}$ can be modeled as $\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}$, where $\mathbf{e}$ is the quantization noise with covariance $\mathbf{Q}$ and assumed to be independent of $\hat{\mathbf{x}}$.
- Achievable rate for user k is again $R_k = \log(1 + \text{SINR}_k)$ where $\text{SINR}_k = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + |\mathbf{h}_k^H \mathbf{Q} \mathbf{h}_k|}$.
- Assuming ideal quantizer and independent quantization of BSs signals, the quantization noise level q_l and the backhaul capacity C_l must satisfy $\log\left(\frac{\hat{P}_l + q_l}{q_l}\right) \leq C_l$.

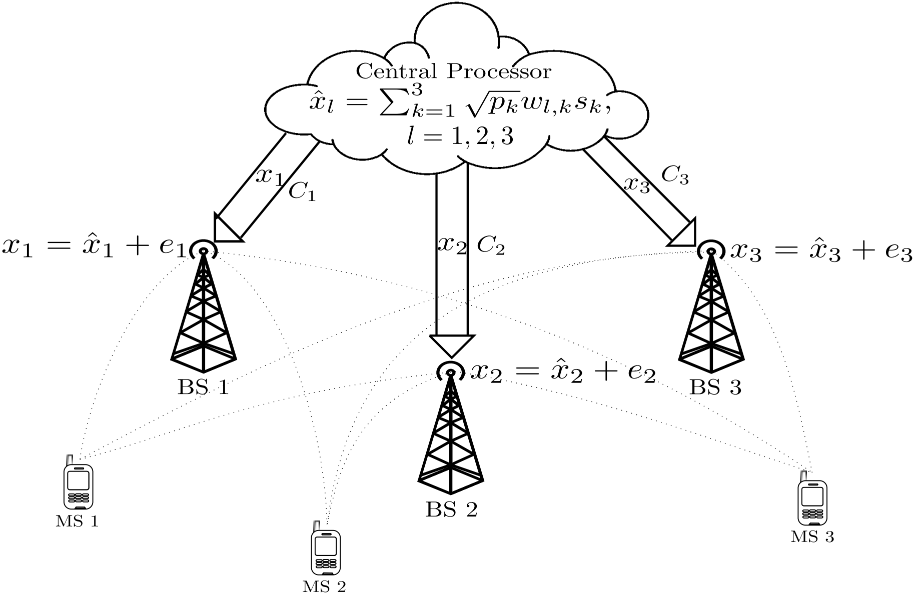

Figure: Compression-based cooperation scheme for the downlink CRAN.

Compression-Based Scheme Optimization

- Overall, weighted sum rate maximization problem becomes

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

- An iterative approach based on majorization minimization is suggested in [Park et al., 2013].
- More general joint compression strategy that leverages correlation between quantization noise levels has also been studied in [Park et al., 2013].

Proposed Hybrid Scheme

- In message-sharing scheme, backhaul links carry user messages.
- In compression-based scheme, backhaul links carry compressed signals.
- In the proposed hybrid scheme, the precoding operation is split between the central processor and the BSs.
- A part of backhaul is used to send direct messages for some users and remaining backhaul is used to carry the compressed signal that combines the contributions from the rest of the users
- Rationale: Since 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.

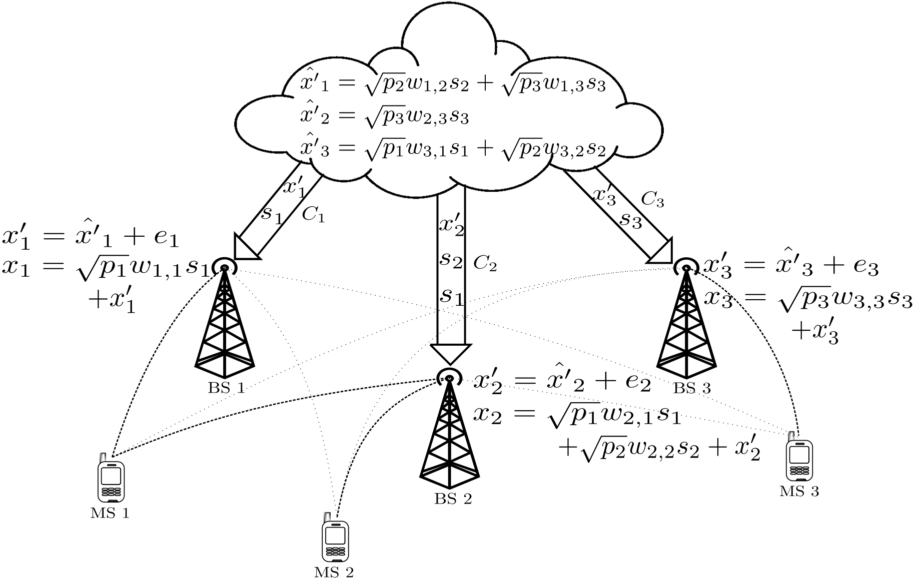

Figure: Example of hybrid compression and message-sharing scheme.

Design Methodology for Proposed Hybrid Scheme

- Design parameters for the hybrid scheme:
  - Which users should do message-sharing and to which BSs?
  - How to design beamforming vectors?
  - How to select quantization noise levels for rest of the compressed signals?
- We propose following design methodology:
  1. Design fixed network-wide beamformers using, for example, the regularized zero-forcing approach or WMMSE approach;
  2. Assuming pure compression, optimize the quantization noise level in each backhaul link, obtain the user rates;
  3. Appropriately select users for message sharing.

1. Choose network-wide beamformers

- Fix the network beamformers for precoding the user signals over the multiple BSs.
- For example, for regularized ZF: direction for the beamformer of user k, $\mathbf{w}_k$, is chosen along $\frac{\mathbf{t}_k}{\|\mathbf{t}_k\|}$ where $[\mathbf{t}_1, \cdots, \mathbf{t}_K] = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \alpha \mathbf{I})^{-1}$ with α as a regularization factor.
- For a fixed α, powers p_k associated with each beam are solved by approximating the SINR for each user (neglecting interference):

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

- Appropriate regularization constant α is set heuristically depending on SNR, or solving (P2) for different α's and picking the one that maximizes the weighted sum rate.

2. Start with pure compression scheme

- We use the following method for finding the optimal quantization noise level and resulting achievable user rates for pure compression.
- Same as solving (P1), but instead of using the MM method as in [Park et al., 2013], 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:

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

where $\mathsf{SINR}'_k = \frac{p_k |\mathbf{h}_k^H \mathbf{w}_k|^2}{\sum_{j \neq k} p_j |\mathbf{h}_k^H \mathbf{w}_j|^2 + \sigma^2 + \sum_{l=1}^{L} \frac{\hat{P}_l |h_{l,k}|^2}{2^{C_l} - 1}}$.

- Advantage: (P3) becomes convex in C_l (assuming fixed p_k and $\mathbf{w}_k$), which allows efficient numerical solution.

3. Select users for message-sharing

- For each user, we compare 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.
- Key idea:
  - Recall we compress the precoded signal $\hat{x}_l = \sum_{k=1}^{K} \sqrt{p_k} w_{l,k} s_k$ for BS l.
  - The amount of backhaul needed to compress x_l to within quantization noise level q_l is approximately $\log\left(\frac{\hat{P}_l}{q_l}\right)$, where $\hat{P}_l = \sum_{k=1}^{K} p_k |w_{l,k}|^2$.
  - 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 $\hat{P}_l - \hat{P}_{l,k}$. 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.
  - 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 $\log\left(\frac{\hat{P}_l}{\hat{P}_l - \hat{P}_{l,k}}\right) - R_k > 0$.

Greedy user selection algorithm

Algorithm 1 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

Simulation Setup I

Cellular Layout	Hexagonal, 7-cell
Channel Bandwidth	10 MHz
Frequency Reuse	1
BS-to-BS distance	0.8km
Number of Tx antennas/BS	1
Number of Rx antennas/user	1
Number of users/cell	15
Background Noise	-162 dBm/Hz
Distance-dependent path loss	$128.1 + 37.6\log_{10}(d)$
Log-normal shadowing	8dB
Rayleigh small scale fading	0dB
Average power per-BS	-27 dBm/Hz
Scheduling	Round Robin

Table: Simulation Parameters

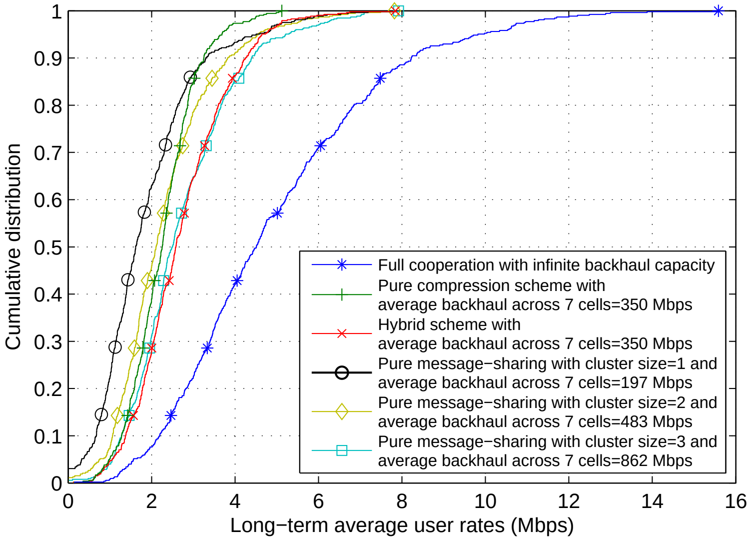

Figure: Comparison of cumulative distribution functions of user rates for the message sharing, pure compression, and hybrid schemes in a 7-cell network.

Simulation Setup II

Cellular Layout	Hexagonal, 19-cell, 3 sectors/cell
Channel Bandwidth	10 MHz
Frequency Reuse	1
BS-to-BS distance	0.8km
Number of Tx antennas/BS	1
Number of Rx antennas/user	1
Number of users/sector	10
Background Noise	-162 dBm/Hz
Distance-dependent path loss	$128.1 + 37.6 \log_{10}(d)$
Log-normal shadowing	8dB
Rayleigh small scale fading	0dB
Maximum BS transmit power	-27 dBm/Hz
Scheduling	Round Robin

Table: Simulation Parameters

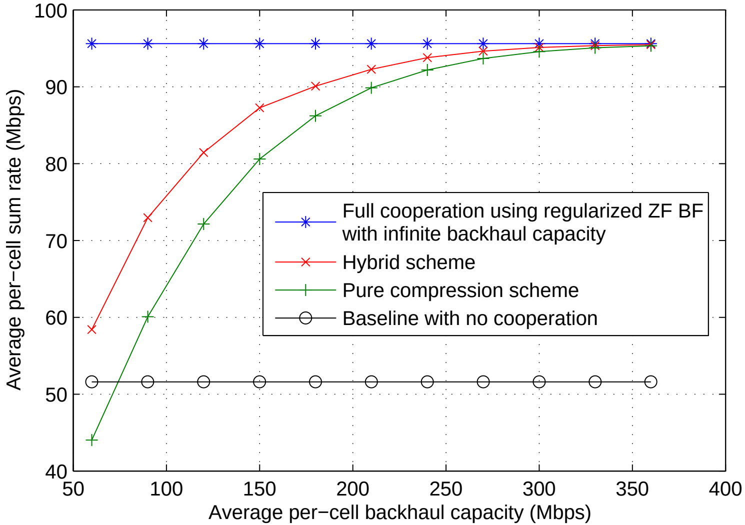

Figure: Per-cell sum rate vs. average per-cell backhaul capacity for the hybrid scheme as compared to the pure compression scheme in a 19-cell topology with center 7 cells forming a cooperating cluster.

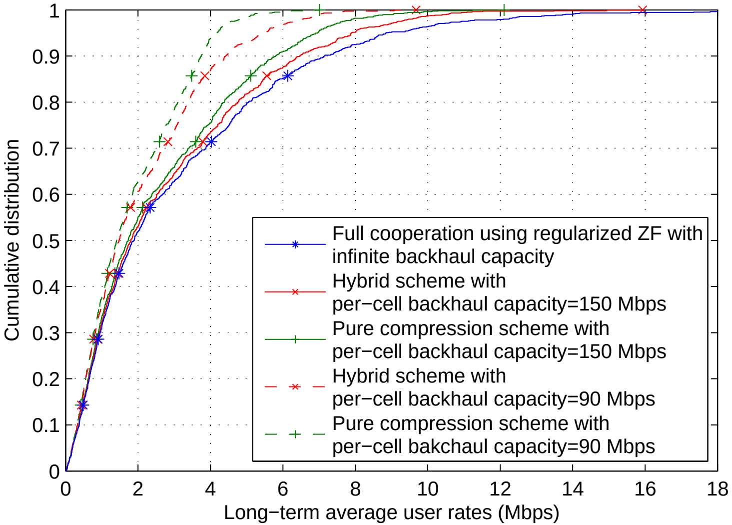

Figure: Comparison of cumulative distribution functions of user rates for the hybrid scheme and the pure compression scheme in a 19-cell topology with center 7 cells forming a cooperating cluster.

Gesbert, D., Hanly, S., Huang, H., Shamai Shitz, S., Simeone, O., and Yu, W. (2010).
Multi-cell MIMO cooperative networks: A new look at interference.
IEEE J. Sel. Areas Commun., 28(9):1380–1408.

Hong, S.-N. and Caire, G. (2013).
Compute-and-forward strategies for cooperative distributed antenna systems.
IEEE Trans. Inf. Theory, 59(9):5227–5243.

Kaviani, S., Simeone, O., Krzymien, W., and Shamai, S. (2012).
Linear precoding and equalization for network MIMO with partial cooperation.
IEEE Trans. Veh. Technol., 61(5):2083–2096.

References II

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

Marsch, P. and Fettweis, G. (2009).
On downlink network mimo under a constrained backhaul and imperfect channel knowledge.
In *Proc. IEEE Global Commun. Conf. (Globecom)*.

Nazer, B., Sanderovich, A., Gastpar, M., and Shamai, S. (2009).
Structured superposition for backhaul constrained cellular uplink.
In *"IEEE Int. Symp. Inf. Theory"*, pages 1530–1534.

Ng, B. L., Evans, J., Hanly, S., and Aktas, D. (2008).
Distributed downlink beamforming with cooperative base stations.
IEEE Trans. Inf. Theory, 54(12):5491–5499.

References III

Park, S.-H., Simeone, O., Sahin, O., and Shamai, S. (2013).
Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks.
IEEE Trans. Signal Process., 61(22):5646–5658.

Simeone, O., Somekh, O., Poor, H., and Shamai (Shitz), S. (2009).
Downlink multicell processing with limited-backhaul capacity.
EURASIP J. Advances Singal Process.

Zakhour, R. and Gesbert, D. (2011).
Optimized data sharing in multicell MIMO with finite backhaul capacity.
IEEE Trans. Signal Process., 59(12):6102–6111.

References IV

Zhan, J., Nazer, B., Erez, U., and Gastpar, M. (2010).
Integer-forcing linear receivers.
In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1022–1026.

Zhao, J., Quek, T., and Lei, Z. (2013).
Coordinated multipoint transmission with limited backhaul data transfer.
"IEEE Trans. Wireless Commun.", 12(6):2762–2775.

Thanks for listening!

Questions/comments/thoughts?