Cloud Radio-Access Networks:
Capacity, Coding Strategies, and Optimization

Wei Yu

Joint work with Binbin Dai, Pratik Patil, Yuhan Zhou, Liang Liu

Department of Electrical and Computer Engineering
University of Toronto, Canada
Future 5G wireless cellular network:

- Requirements: Gbps capacity, 1ms latency, $10^5$ connectivity
- Bottleneck: Path-loss, fading, and interference

Emerging trends:

- Dense
  - Heterogeneous network; Small cell
- Massive
  - Massive MIMO at each BS
- Cooperative
  - Signal processing for interference cancellation

This talk: Capacity and optimization of cooperative networks.
Cooperating BSs in the Cloud

Cloud Radio-Access Network (C-RAN)
Benefits of C-RAN:

- Allows a cost-effective way to deploy and upgrade wireless platform;
- Opens up new possibilities for the optimization of air-interface;
- Enables cooperative communication for interference mitigation;
- Provides an implementation of coordinated multi-point (CoMP).

This talk: Information theoretical analysis of C-RAN

- Multicell Joint Processing for Uplink C-RAN
- Multicell Beamforming for Downlink C-RAN
Wireless Access via the Cloud

Central Processor
$s_1, s_2, s_3$

Fronthaul Links

$C_1, C_2, C_3$

RRH 1
User 1

RRH 2
User 2

RRH 3
User 3
\[ X_1, X_2, \ldots, X_K \text{ are user terminals}; \ Y_1, Y_2, \ldots, Y_L \text{ are RRHs.} \]

- Practical constraint: Fronthaul capacity limited to \( C_l \).
- Goal: To maximize the overall capacities for all users.
What should each RRH do? Local detection vs. compression...

What should the cloud do? Successive vs. joint decoding...

How should we design transmit signaling?
Successive Interference Cancellation in the Cloud

Equivalent channel of user $k$ in the $k^{th}$ decoding stage:

\[
Y_k : \hat{Y}_k = Z_k + \sum_{j \neq k} h_{jk} X_j
\]

- The quantized observation at RRH $k$ is sent to the centralized processor via the fronthaul link of rate $C_k$.
- Previously decoded $X_1$ to $X_{k-1}$ serve as side information for Wyner-Ziv compression and for decoding of $X_k$, achieving:

\[
R_k = \frac{1}{2} \log \frac{1 + \bar{\text{SINR}}_k}{1 + 2^{-2C_k \text{SINR}_k}}
\]

where \(\bar{\text{SINR}}_k = (h_{kk}^2 P_k)/(N_0 + \sum_{j > k} h_{jk}^2 P_j)\)

Wei Yu
C-RAN 8 / 37
Better Strategy: Decoding Based on Cluster of RRHs

- Per-RRH decoding with SIC:

\[ R_k = I(X_k; \hat{Y}_k | X_1, \cdots, X_{k-1}) \]

subject to \( I(Y_k; \hat{Y}_k | X_1, \cdots, X_{k-1}) \leq C_k \).

- Joint-RRH decoding can do better:

\[ R_k = I(X_k; \hat{Y}_1, \cdots, \hat{Y}_L | X_1, \cdots, X_{k-1}), \]

subject to \( I(Y_k; \hat{Y}_k | \hat{Y}_1, \cdots, \hat{Y}_{k-1}) \leq C_k \).
Each RRH compresses $Y_i$ into $\hat{Y}_i$

- Compression can be done with Wyner-Ziv or single-user coding.

The cloud decodes the quantized received signals $\{\hat{Y}_1, \cdots, \hat{Y}_L\}$, then the transmit messages $X_1, X_2, \ldots, X_K$, successively or jointly.

Information theoretical justification:
- Joint decoding proposed by Sanderovich-Somekh-Poor-Shamai ('09) and Sanderovich-Shamai-Steinberg-Kramer ('08)
- Avestimehr-Diggavi-Tse ('09): “Wireless Network Info Flow”
Fact: Assuming Gaussian quantization, optimal input is Gaussian.
Theorem: Assuming Gaussian input, optimal quantizer is Gaussian.
However, joint Gaussian signal/quantization may not be optimal
   • Binary counterexample: Sanderovich-Shamai-Steinberg-Kramer’08
Theorem (Achievable rate region)

Achievable rate under sum fronthaul constraint $C$:

$$\sum_{i \in S} R_i \leq \log \frac{|H_S K_X(S) H_S^H + \Lambda_q + \sigma^2 I|}{|\Lambda_q + \sigma^2 I|}$$

either subject to (for Wyner-Ziv coding, V-MAC-WZ):

$$\log \frac{|H K_X H^H + \Lambda_q + \sigma^2 I|}{|\Lambda_q|} \leq C$$

or subject to (for single-user compression, V-MAC-SU):

$$\log \frac{|\text{diag}(H K_X H^H) + \Lambda_q + \sigma^2 I|}{|\Lambda_q|} \leq C$$

where $\Lambda_q = \text{diag}(q_1, q_2, \ldots, q_L)$ is the quantization noise level.
Noisy Network Coding (Lim-Kim-El Gamal-Chung’11)

- **Cut-set Bound:** $R(S) = \sum_{k \in S} R_k \leq I(x_{ul}(S); y_{ul}(S^c)|x_{ul}(S^c))$
- **Achievable rate using noisy network coding:** $R(S) \leq I(x_{ul}(S); \hat{y}_{ul}(S^c), y_{ul}|x_{ul}(S^c)) - I(y_{ul}(S); \hat{y}_{ul}(S)|x_{ul}, \hat{y}_{ul}(S^c), y_{ul})$
- **Set quantization noise at background noise level:** $\hat{y}_{ul} \approx y_{ul}$. 

Wei Yu
Approximate Optimality of Compress-and-Forward

Successive-decoding region for MAC

\[ R_1 < I(X_1^{ul}, \hat{Y}_1^{ul}, \hat{Y}_2^{ul} | X_2^{ul}); \]
\[ R_2 < I(X_2^{ul}, \hat{Y}_1^{ul}, \hat{Y}_2^{ul} | X_2^{ul}); \]
\[ R_1 + R_2 < I(X_1^{ul}, X_2^{ul}, \hat{Y}_1^{ul}, \hat{Y}_2^{ul}) \]

Wyner-Ziv Compression

\[ C_1 > I(Y_1^{ul}; \hat{Y}_1^{ul} | \hat{Y}_2^{ul}); \]
\[ C_2 > I(Y_2^{ul}; \hat{Y}_2^{ul} | \hat{Y}_1^{ul}); \]
\[ C_1 + C_2 > I(Y_1^{ul}, Y_2^{ul}; \hat{Y}_1^{ul}, \hat{Y}_2^{ul}) \]
Comparing with Noisy Network Coding

\[ R_1 < I(X_{1ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{2ul}^u); \]

\[ R_1 < I(X_{1ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{2ul}^u) + C_1 - I(Y_{1ul}^u; \hat{Y}_1^u | \hat{Y}_2^u, X_{2ul}^u); \]

\[ R_1 < I(X_{1ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{2ul}^u) + C_2 - I(Y_{2ul}^u; \hat{Y}_2^u | \hat{Y}_1^u, X_{2ul}^u); \]

\[ R_1 < I(X_{1ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{2ul}^u) + C_1 + C_2 - I(Y_{1ul}^u, Y_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{2ul}^u); \]

\[ R_2 < I(X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{1ul}^u); \]

\[ R_2 < I(X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{1ul}^u) + C_1 - I(Y_{1ul}^u; \hat{Y}_1^u | \hat{Y}_2^u, X_{1ul}^u); \]

\[ R_2 < I(X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{1ul}^u) + C_2 - I(Y_{2ul}^u; \hat{Y}_2^u | \hat{Y}_1^u, X_{1ul}^u); \]

\[ R_2 < I(X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{1ul}^u) + C_1 + C_2 - I(Y_{1ul}^u, Y_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u | X_{1ul}^u); \]

\[ R_1 + R_2 < I(X_{1ul}^u, X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u); \]

\[ R_1 + R_2 < I(X_{1ul}^u, X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u) + C_1 - I(Y_{1ul}^u; \hat{Y}_1^u | \hat{Y}_2^u); \]

\[ R_1 + R_2 < I(X_{1ul}^u, X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u) + C_2 - I(Y_{2ul}^u; \hat{Y}_2^u | \hat{Y}_1^u); \]

\[ R_1 + R_2 < I(X_{1ul}^u, X_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u) + C_1 + C_2 - I(Y_{1ul}^u, Y_{2ul}^u; \hat{Y}_1^u, \hat{Y}_2^u) \]
Uniform quantization noise level is optimal only at high SQNR.

In general: Jointly optimize transmit and quantization covariances.

Solution: Successive convex approximation with WMMSE.

WMMSE-SCA: Optimal Tx/Rx beamforming then compression.
**Simulation Result: V-MAC-WZ**

**Figure:** CDF of user rates in a 7-cell cluster: VMAC-WZ vs. Per-BS SIC.
Figure: Per-cell sum rate vs. average per-cell fronthaul capacity.
Figure: 12-antenna RRH serving 2 users: Compress vs. Beamform-Compress.
How to enable cooperation across clusters of RRHs?

- Message-sharing with a cluster of RRHs for joint beamforming.
- Precode at the cloud. Compress-forward precoded signals to RRHs.
  - Multivariate compression [Park-Simeone-Sahin-Shamai '13].
Two fundamental coding strategies for downlink C-RAN:

- **Data-Sharing**: CP distributes each user’s data to a cluster of RRHs. Each RRH has access to multiple data streams then precode.
- **Compression**: CP computes the beamformer, then compresses and distributes the precoded signal to the RRHs.

How to best utilize the limited fronthaul?

- In **Data-Sharing**, limit the cluster size;
- In **Compression**, control quantization level.
Optimizing Clustering in Message-Sharing

Cloud Processor

\[ C_1 \rightarrow C_2 \rightarrow \ldots \rightarrow C_{L-1} \rightarrow C_L \]

“Personalized” cloud

© The Walt Disney Company
Sparse Beamforming for the Downlink C-RAN

- Weighted sum-rate maximization under per-RRH power constraints and per-RRH fronthaul constraints assuming single-stream per user:

\[
\begin{align*}
\text{maximize} & \quad \sum_k \alpha_k R_k \\
\text{subject to} & \quad \sum_k \|w^l_k\|_2^2 \leq P_l, \quad \forall l \\
& \quad \sum_k \left\|\|w^l_k\|_2\right\|_0 R_k \leq C_l, \quad \forall l
\end{align*}
\]

- Use $\ell_1$ re-weighting and compressed sensing [Candès-Wakin-Boyd’08]
- The WMMSE approach can be used to find a local optimum. [Christensen-Agarwal-Carvalho-Cioffi ’08], [Shi-Razaviyayn-Luo-He ’11], [Kaviani-Simeone-Krzymien-Shamai ’12]
- Related work: Zhao-Quek-Lei (’13), Luo-Zhang-Lim (’14), Zhuang-Lau (’14),
Better Strategy: Compression for Multicell Beamforming

- Full cooperation possible, but compression introduces quantization noises.
- Optimizing by majorization-minimization: [Park-Simeone-Sahin-Shamai ’13]
Precoded signals intended for RRHs formed at central processor:

\[ \hat{x} = [\hat{x}_1, \cdots, \hat{x}_L]^T = \sum_{k=1}^{K} w_k s_k \]

Quantization for \( \hat{x} \) modeled as \( x = \hat{x} + e \), where \( e \) is the quantization noise with covariance \( Q \), independent of \( \hat{x} \).

Achievable rate for user \( k \) is

\[ R_k = \log \left( 1 + \frac{|h_k^H w_k|^2}{\sum_{j \neq k} |h_k^H w_j|^2 + \sigma^2 + |h_k^H Q h_k|} \right) \]

The fronthaul capacity constraint must satisfy

\[ \log \left( 1 + \frac{\sum_{k=1}^{K} |w_{l,k}|^2}{q_l} \right) \leq C_l \]

Here, \( Q \) is assumed diagonal; multivariate \( Q \) also possible.
Figure: 4-antenna RRH with Independent Compression.
Cut-Set: $R(S) \leq I(x_{dl}^1(S); y_{dl}^1(S^c)|x_{dl}^1(S^c))$

Distributed Decode-Forward: $R(S) \leq I(x_{dl}^1(S); u(S^c)|x_{ul}^1(S^c) - \sum_{k \in S_c} [I(u_{dl}^k; u(S^c_k), x_{dl}^N|y_{dl}^k) + I(x_{dl}^k; x_{dl}^1(S^c_k))]$

To achieve constant gap: Choose $u_k$ close to $y_{dl}^k$. 
Approximate Optimality of Compression-like Strategy

Marton’s Region for Broadcast

\[ R_1 < I(U_1; Y_1^{dl}) \]
\[ R_2 < I(U_2; Y_2^{dl}) \]
\[ R_1 + R_2 < I(U_1; Y_1^{dl}) + I(U_2; Y_2^{dl}) - I(U_1; U_2); \]

Correlated Compression

\[ C_1 > I(X_1^{dl}; U_1, U_2); \]
\[ C_2 > I(X_2^{dl}; U_1, U_2); \]
\[ C_1 + C_2 > I(X_1^{dl}, X_2^{dl}, U_1, U_2) + I(X_1^{dl}; X_2^{dl}); \]
Comparing with Distributed Decode-Forward

\[ R_1 < I(U_1, Y_{dl}^1); \]
\[ R_1 < I(U_1, Y_{dl}^1) + C_1 - I(U_1; X_{dl}^1); \]
\[ R_1 < I(U_1, Y_{dl}^1) + C_2 - I(U_1; X_{dl}^2); \]
\[ R_1 < I(U_1, Y_{dl}^1) + C_1 + C_2 - I(U_1; X_{dl}^1, X_{dl}^2); \]
\[ R_2 < I(U_2, Y_{dl}^2); \]
\[ R_2 < I(U_2, Y_{dl}^2) + C_1 - I(U_2; X_{dl}^1); \]
\[ R_2 < I(U_2, Y_{dl}^2) + C_2 - I(U_2; X_{dl}^2); \]
\[ R_2 < I(U_2, Y_{dl}^2) + C_1 + C_2 - I(U_2; X_{dl}^1, X_{dl}^2); \]
\[ R_1 + R_2 < I(U_1, Y_{dl}^1) + I(U_2, Y_{dl}^2) - I(U_1; U_2); \]
\[ R_1 + R_2 < I(U_1, Y_{dl}^1) + I(U_2, Y_{dl}^2) - I(U_1; U_2) + C_1 - I(U_1, U_2; X_{dl}^1); \]
\[ R_1 + R_2 < I(U_1, Y_{dl}^1) + I(U_2, Y_{dl}^2) - I(U_1; U_2) + C_2 - I(U_1, U_2; X_{dl}^2); \]
\[ R_1 + R_2 < I(U_1, Y_{dl}^1) + I(U_2, Y_{dl}^2) - I(U_1; U_2) + C_1 + C_2 - I(U_1, U_2; X_{dl}^1, X_{dl}^2) \]
\[ - I(X_{dl}^1, X_{dl}^2); \]
Uplink versus Downlink C-RAN

**Uplink**
- Multiple-access-relay channel
- Simple encoders, complex cloud decoder
- Compress-forward with independent or Wyner-Ziv compression
- Noisy network coding within constant gap

**Downlink**
- Broadcast-relay channel
- Simple decoders, complex cloud encoder
- Compression strategy with independent or multivariate compression covering
- Distributed decode-forward within constant gap
Uplink-Downlink Duality in C-RAN

(a) Uplink

(b) Downlink

- Uplink-downlink duality for compression-based beamforming
  - Under same sum-power and individual fronthaul constraints.
  - Achievable rates of the uplink and downlink are the same.

- Generalization of uplink-downlink duality to MAC-BC with relays.
Uplink: Fixed-point method

\[
\begin{align*}
\text{minimize} & \quad P^{\text{ul}}(\{p_i^{\text{ul}}\}) \\
\text{subject to} & \quad R_k^{\text{ul}}(\{p_i^{\text{ul}}, w_i\}, \{q_l^{\text{ul}}\}) \geq R_k, \quad \forall k, \\
& \quad C_i^{\text{ul}}(\{p_i^{\text{ul}}\}, q_l^{\text{ul}}) \leq C_l, \quad \forall l.
\end{align*}
\]

Downlink: Based on uplink solution

\[
\begin{align*}
\text{minimize} & \quad P^{\text{dl}}(\{p_i^{\text{dl}}\}, \{q_i^{\text{dl}}\}) \\
\text{subject to} & \quad R_k^{\text{dl}}(\{p_i^{\text{dl}}, v_i\}, \{q_l^{\text{dl}}\}) \geq R_k, \quad \forall k, \\
& \quad C_i^{\text{dl}}(\{p_i^{\text{dl}}, v_i\}, q_l^{\text{dl}}) \leq C_l, \quad \forall l.
\end{align*}
\]
Achievable rates in C-RAN are significantly influenced by:
- Distances between transmitters and receivers.
- Random channel fading realizations.
- Stochastic geometry provides analytic tool [Andrews-Baccelli-Ganti’11]
Obtaining signal and interference distributions is the main challenge!

- Model distance-dependent channel characterization:
  \[ g_{ilmj} = \sqrt{\beta_{ilmj}} h_{ilmj} \quad \text{with} \quad h_{ilmj} \sim \mathcal{CN}(0, I_M), \quad \beta_{ilmj} = \left(1 + \frac{r_{ilmj}}{d_0}\right)^{-\alpha} \]

- Approximate signal and interference distributions as Gamma distributions with modified parameters [Heath-Wu-Kwon-Soong’11]

  \[ g_{il}^H g_{il} = \sum_{b=1}^{B_l} \beta_{ilbl} h_{ilbl}^H h_{ilbl} \sim \Gamma(k_{il}, \theta_{il}) \]

  where \( k_{il} = M \left( \sum_{b=1}^{B_l} \beta_{ilbl} \right)^2 \), \( \theta_{il} = \frac{\sum_{b=1}^{B_l} \beta_{ilbl}^2}{\sum_{b=1}^{B_l} \beta_{ilbl}} \)

- Key fact:

  \[ \ln(1 + x) = \int_0^\infty \frac{e^{-z}}{z} (1 - e^{-xz}) \, dz \]

  Ergodic rate can be characterized in terms of Laplace transforms!
How Large Should the Cluster Size Be?

Cluster sizes are limited by the fronthaul and by CSI acquisition.

Wei Yu
C-RAN 35 / 37
Conclusions

Cloud radio-access network is an enabling architecture that allows
- Joint signal processing across the RRHs;
- Advanced network optimization.

Network-wide optimization is likely to be done in the cloud.

Summary of results in this talk:
- Uplink: Compression with optimized quantization levels.
- Downlink: Message-sharing and compression are viable strategies.
- Design: Duality, WMMSE, $\ell_1$ reweighting, Succ. Convex Approx.
- Analysis: Information theory, Optimization, Stochastic geometry.

Future wireless cellular architecture:
- Dense, massive, and cooperative.
Further Information


