

An Information Theoretic View of Cloud Radio Access Networks

Shlomo Shamai

The Andrew and Erna Viterbi Department of Electrical Engineering
Technion–Israel Institute of Technology

Supported by the European Union's Horizon 2020,
Research And Innovation Program: ERC 694630

Advanced Communications Center, Tel Aviv University
Annual Workshop & Feder Family Award Ceremony
21 January 2018

Joint work with O. Simeone, S.-H. Park and O. Sahin



Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology
- D. Inter-connected radio units

III. Downlink

- A. Point-to-point compression
- B. Multivariate fronthaul compression
- C. Structured coding
- D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology
- D. Inter-connected radio units

III. Downlink

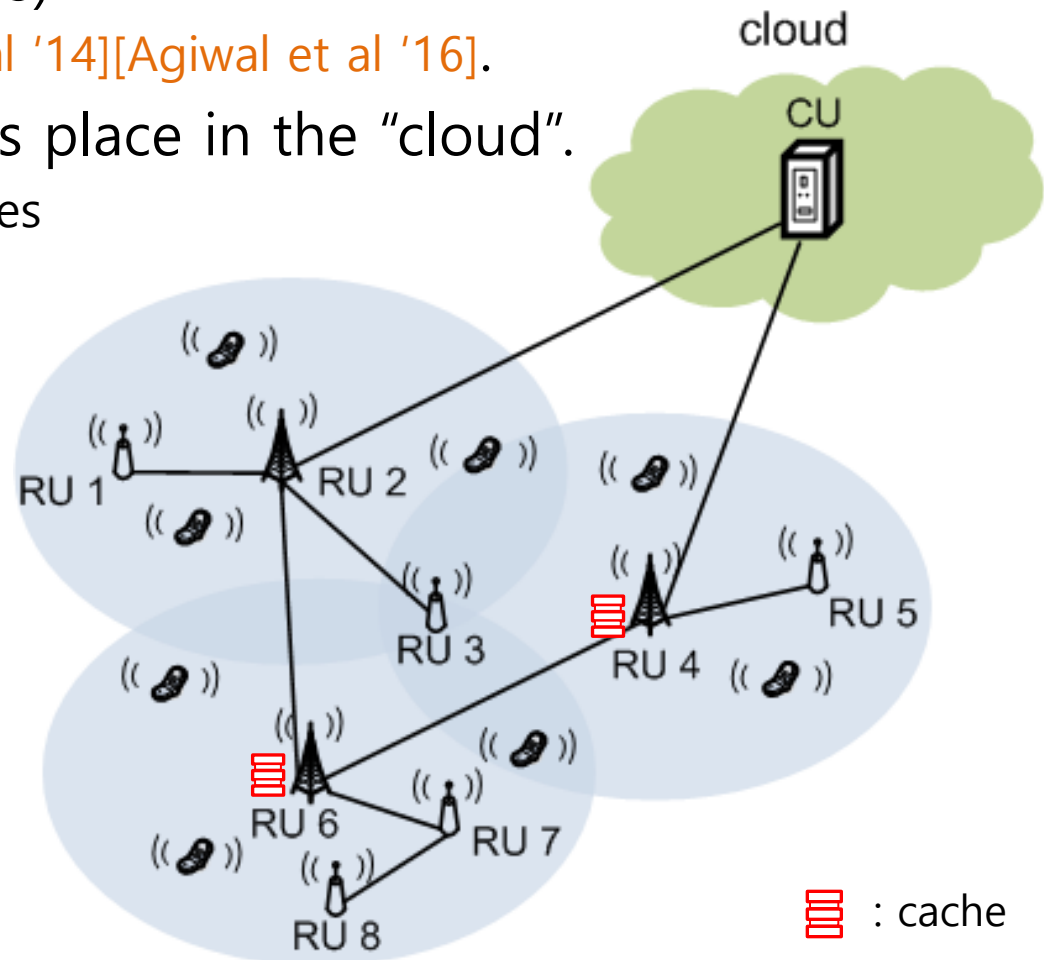
- A. Point-to-point compression
- B. Multivariate fronthaul compression
- C. Structured coding
- D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Cloud/Fog Radio Access Networks

- Base Stations (BSs), macro/pico, operate as radio units (RUs)
[Alcatel-Lucent][China][Rost et al '14][Agiwal et al '16].
- Baseband processing takes place in the “cloud”.
 - Baseband processing includes encoding/decoding of the messages of Mobile Stations (MSs), (i.e., User Equipment (UEs)).
- Fronthaul links carry complex (IQ) baseband signals.
- Network utilization of low data traffic instances for caching.



Cloud Radio Access Networks

Advantages:

- Low-cost deployment of BSs
- Effective interference mitigation via joint baseband processing

Key challenge: Effective transfer of the IQ signals on the fronthaul links [Andrews et al JSAC'14]

Cloud Radio Access Networks

- Common public radio interface (CPRI) standard based on analog-to-digital (ADC)/digital-to-analog converter (DAC)
 [CPRI][IDC]

Table 1. An example link rate calculation for a 3 sector cell with LTE-Advanced.

Parameters	Settings	Units
Sectors	3	
LTE Carriers	5	
Bandwidth	100	MHz
MIMO	2x2	Tx-Rx
Bits-per-I/Q	15	Bits
Protocol	LTE-A	
Throughput	13.8	Gbps

... Need for fronthaul compression

- “Death by Starvation?: backhaul and 5G,” [Lundqvist, CTN-Sep. 2015]

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

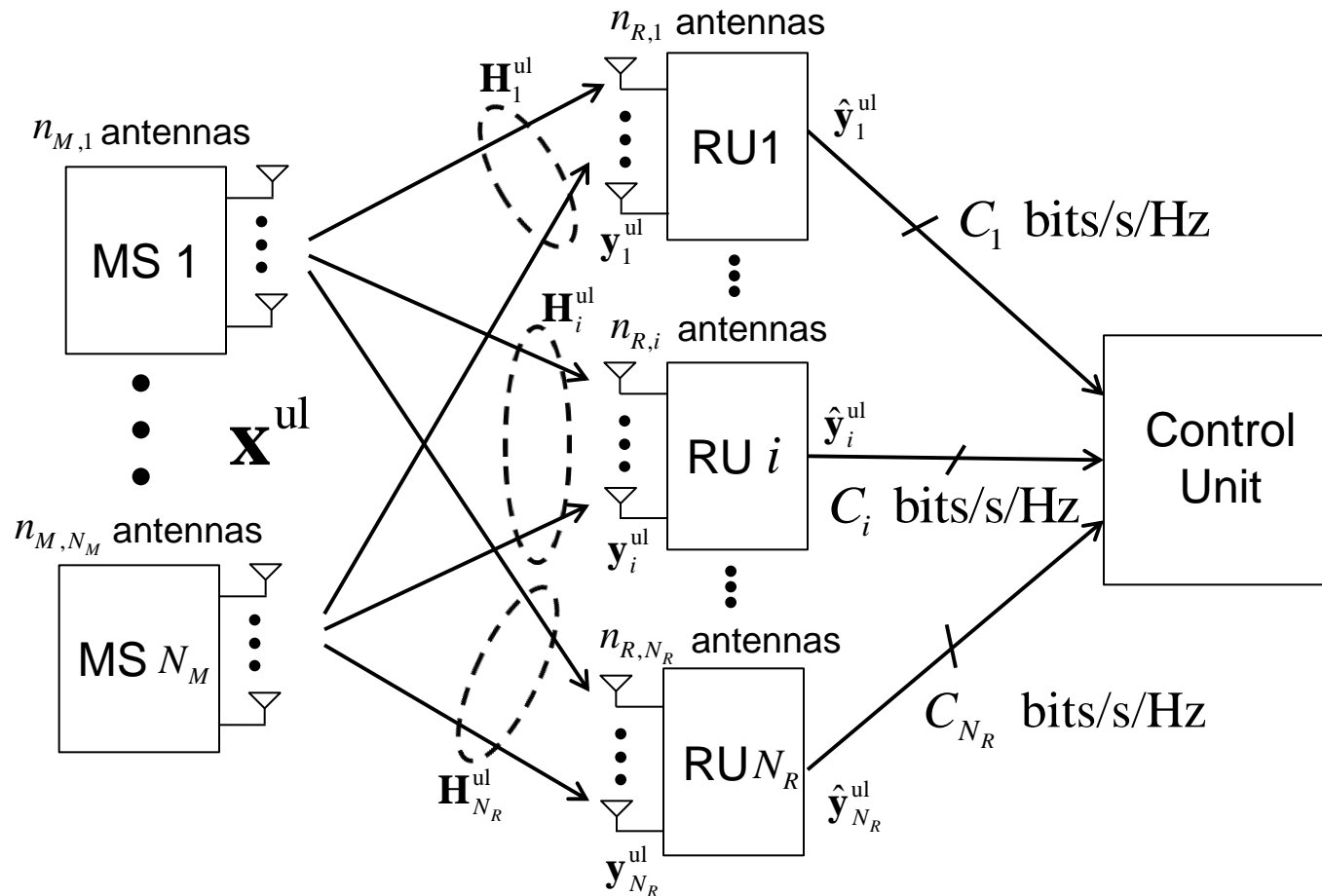
C. Structured coding

D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Basic Settings



Basic Settings

- Assuming flat-fading channel, the received signal at RU i is given by

$$\mathbf{y}_i^{\text{ul}} = \mathbf{H}_i^{\text{ul}} \mathbf{x}^{\text{ul}} + \mathbf{z}_i^{\text{ul}}, \quad i \in \mathcal{N}_R$$

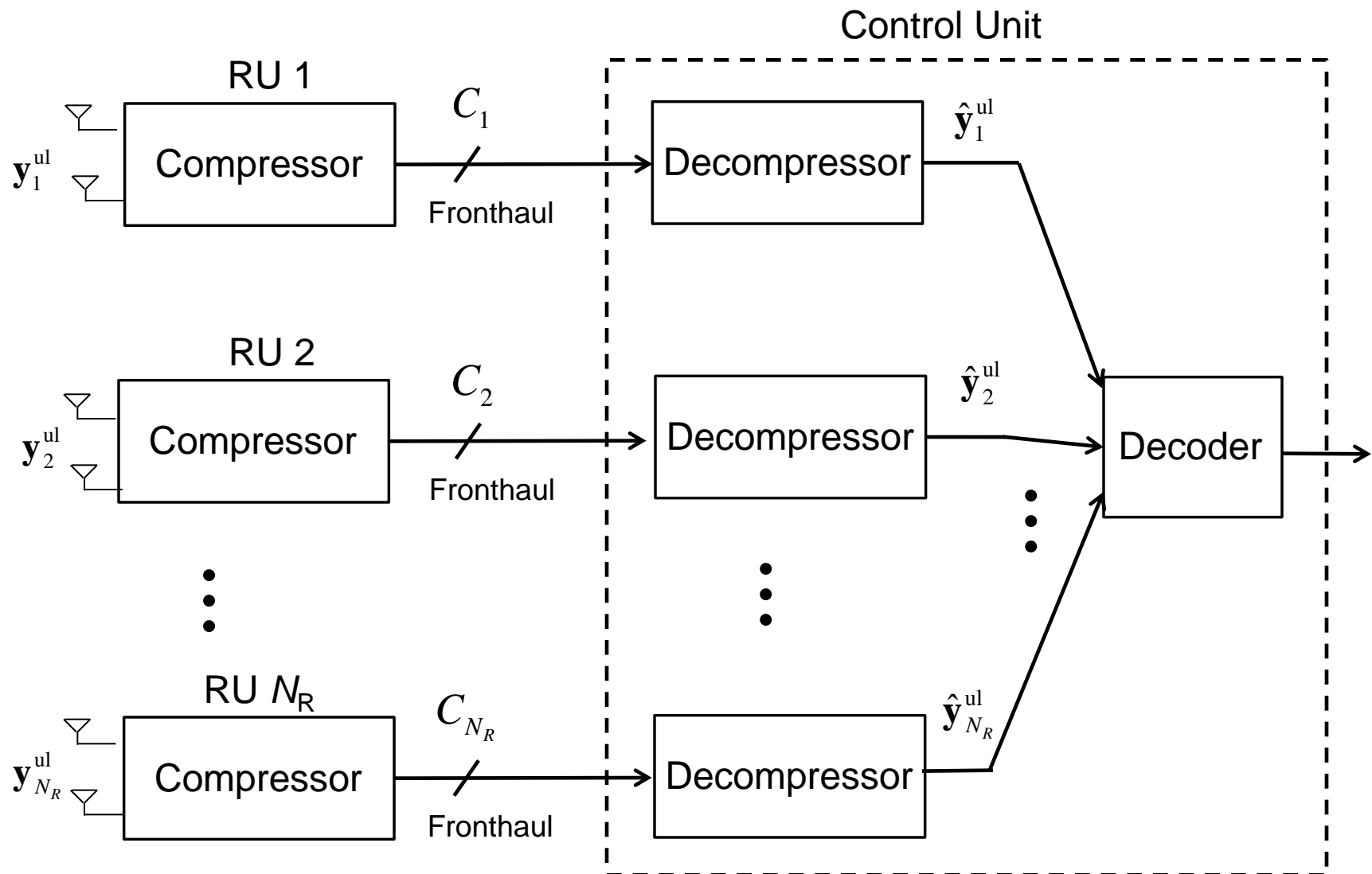
where $\mathbf{H}_i^{\text{ul}} = [\mathbf{H}_{i,1}^{\text{ul}}, \dots, \mathbf{H}_{i,N_M}^{\text{ul}}]$: channel matrix toward to RU i ;

$\mathbf{x}^{\text{ul}} = [\mathbf{x}_1^{\text{ul}H}, \dots, \mathbf{x}_{N_R}^{\text{ul}H}]^H$: vector of symbols transmitted by all MSs;

$\mathbf{z}_i^{\text{ul}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{\mathbf{z}_i^{\text{ul}}})$: noise vector at RU i .

- The fronthaul capacity C_i is normalized to the bandwidth of the uplink channel.
 - For any coding block of n symbols, nC_i bits can be transmitted on the i th fronthaul link.

Point-to-Point Compression



Point-to-Point Compression

- A standard way of modeling the compression at RU i is to adopt the Gaussian direct “test channel” [ElGamal-Kim '11, Ch. 3]

$$\hat{\mathbf{y}}_i^{\text{ul}} = \mathbf{y}_i^{\text{ul}} + \mathbf{q}_i^{\text{ul}},$$

where $\mathbf{q}_i^{\text{ul}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_i^{\text{ul}})$ represents the quantization noise.

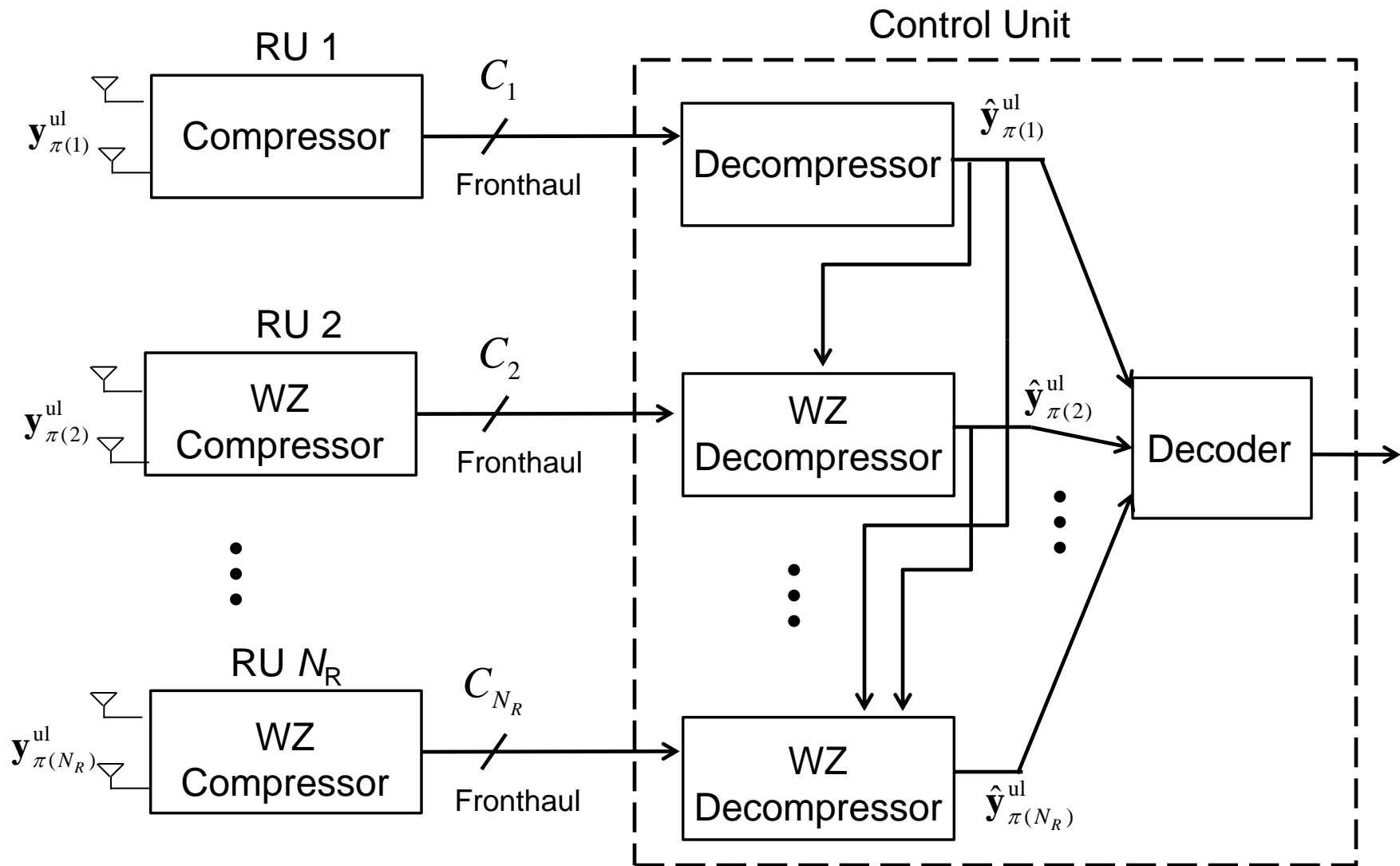
- If the fronthaul capacity C_i satisfies

$$I(\mathbf{y}_i^{\text{ul}}; \hat{\mathbf{y}}_i^{\text{ul}}) \leq C_i,$$

it is possible to design a compression strategy that realizes the given quantization error covariance $\mathbf{\Omega}_i^{\text{ul}}$.

Distributed Fronthaul Compression

[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]



$\pi : \mathcal{N}_R \rightarrow \mathcal{N}_R$ denotes a permutation of RUs' indexes where $\mathcal{N}_R = \{1, \dots, N_R\}$.

Distributed Fronthaul Compression

[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]

- Using Wyner-Ziv compression, a given quantization error matrix $\mathbf{\Omega}_{\pi(i)}^{\text{ul}}$ is attainable if the fronthaul capacity $C_{\pi(i)}$ satisfies

$$I(\mathbf{y}_{\pi(i)}^{\text{ul}}; \hat{\mathbf{y}}_{\pi(i)}^{\text{ul}} | \hat{\mathbf{y}}_{\pi(1)}^{\text{ul}}, \hat{\mathbf{y}}_{\pi(2)}^{\text{ul}}, \dots, \hat{\mathbf{y}}_{\pi(i-1)}^{\text{ul}}) \leq C_{\pi(i)}.$$

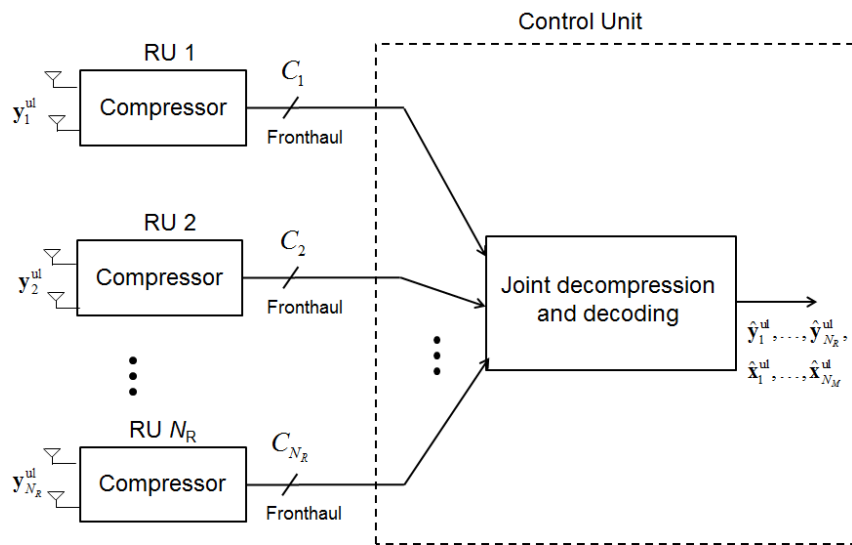
- After the quantized IQ signals $\hat{\mathbf{y}}_1^{\text{ul}}, \dots, \hat{\mathbf{y}}_{N_R}^{\text{ul}}$ are recovered, the CU then performs joint decoding of the signal \mathbf{x}^{ul} sent by all MSs.
 - The uplink sum-rate is given by

$$R_{\text{sum}}^{\text{ul}} = I(\mathbf{x}^{\text{ul}}; \hat{\mathbf{y}}_1^{\text{ul}}, \dots, \hat{\mathbf{y}}_{N_R}^{\text{ul}}).$$

Distributed Fronthaul Compression

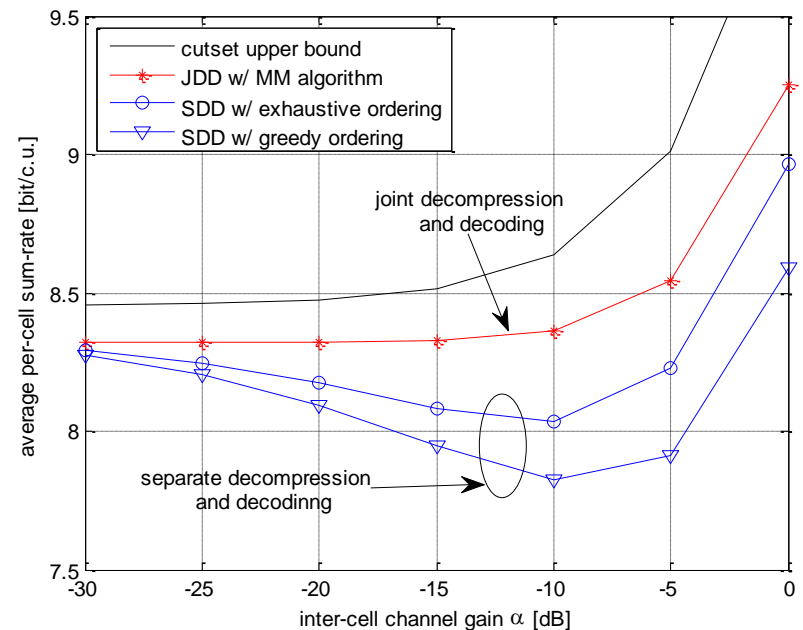
[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]

- Joint decompression and decoding
 - [Sanderovich et al '09][Lim et al '11][Yassaee-Aref '11]
 - Potentially larger rates can be achieved with joint decompression and decoding (JDD) at the central unit [Sanderovich et al '08][Sanderovich et al '09].
 - Now often seen as an instance of noisy network coding [Lim et al '11].
 - Optimal oblivious processing [Aguerri et al '17].



Achievable rate [Sanderovich et al '09][Hong-Caire '15]:

$$R_{\text{sum}} = \min_{S \subseteq \mathcal{N}_B} \left\{ \sum_{j \in S} (C_j - I(\mathbf{y}_j; \hat{\mathbf{y}}_j | \mathbf{x})) + I(\mathbf{x}; \hat{\mathbf{y}}_S) \right\}.$$



Numerical results in 3-cell uplink [Park et al SPL '13]
(SDD: Separate decompression and decoding)

Distributed Fronthaul Compression

[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]

- Sum-rate maximization problem with fronthaul capacity constraints is generally challenging.
- In [Park et al TVT'13], a block-coordinate optimization approach was proposed for successive WZ decompression case.
 - One optimizes the covariance matrices $\mathbf{\Omega}_{\pi(1)}^{\text{ul}}, \dots, \mathbf{\Omega}_{\pi(N_R)}^{\text{ul}}$ following the same order π employed for decompression.
 - At the i th step, for fixed (already optimized) covariances $\mathbf{\Omega}_{\pi(1)}^{\text{ul}}, \dots, \mathbf{\Omega}_{\pi(i-1)}^{\text{ul}}$, the covariance $\mathbf{\Omega}_{\pi(i)}^{\text{ul}}$ is obtained by solving

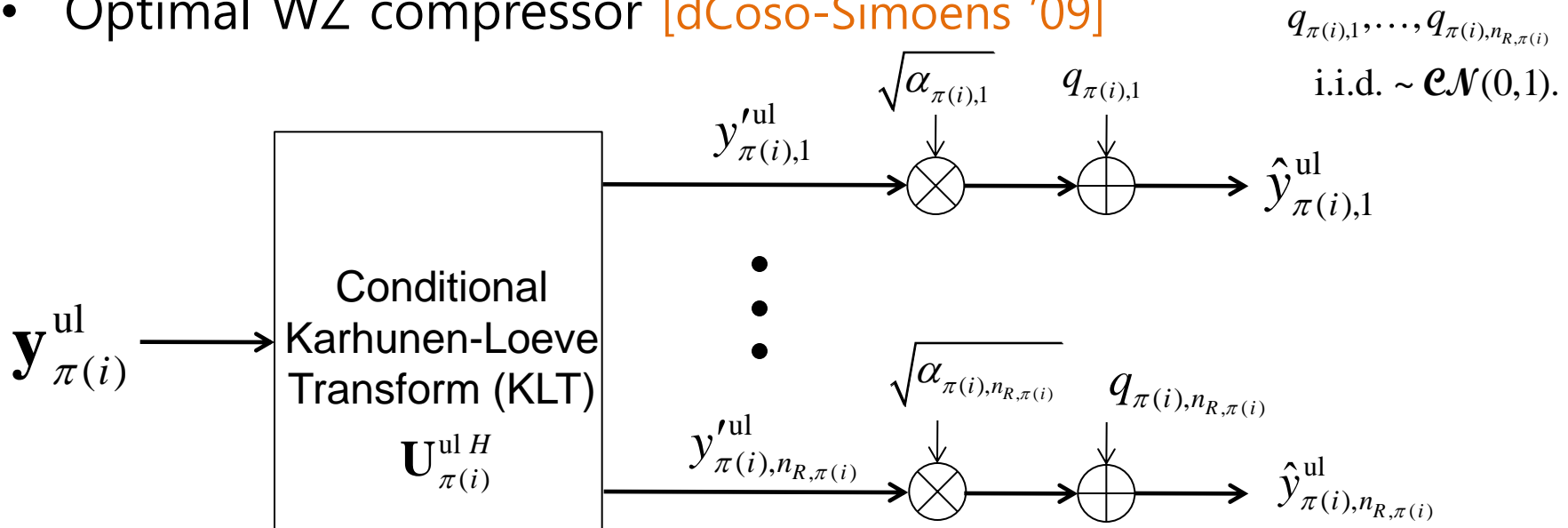
$$\underset{\mathbf{\Omega}_{\pi(i)}^{\text{ul}} \geq \mathbf{0}}{\text{maximize}} \quad I(\mathbf{x}^{\text{ul}}; \hat{\mathbf{y}}_{\pi(i)}^{\text{ul}} \mid \hat{\mathbf{y}}_{\pi(1)}^{\text{ul}}, \hat{\mathbf{y}}_{\pi(2)}^{\text{ul}}, \dots, \hat{\mathbf{y}}_{\pi(i-1)}^{\text{ul}})$$

$$\text{s.t.} \quad I(\mathbf{y}_{\pi(i)}^{\text{ul}}; \hat{\mathbf{y}}_{\pi(i)}^{\text{ul}} \mid \hat{\mathbf{y}}_{\pi(1)}^{\text{ul}}, \hat{\mathbf{y}}_{\pi(2)}^{\text{ul}}, \dots, \hat{\mathbf{y}}_{\pi(i-1)}^{\text{ul}}) \leq C_{\pi(i)}.$$

Distributed Fronthaul Compression

[Sanderovich et al '09] [dCoso-Simoens '09] [Zhou-Yu '11]

- Optimal WZ compressor [dCoso-Simoens '09]



- Unitary transform $\mathbf{U}_{\pi(i)}^{ul H}$ decorrelates the received signal streams when conditioned on the side information signals $\hat{\mathbf{y}}_{\pi(1)}^{ul}, \hat{\mathbf{y}}_{\pi(2)}^{ul}, \dots, \hat{\mathbf{y}}_{\pi(i-1)}^{ul}$.
- Stream-wise multiplication by $\sqrt{\alpha_{\pi(i),1}}, \dots, \sqrt{\alpha_{\pi(i),n_{R,\pi(i)}}}$ represents the compression rate allocation among the streams.
- Statistical independence among quantization noises $q_{\pi(i),1}, \dots, q_{\pi(i),n_{R,\pi(i)}}$ implies that the signals are compressed separately.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

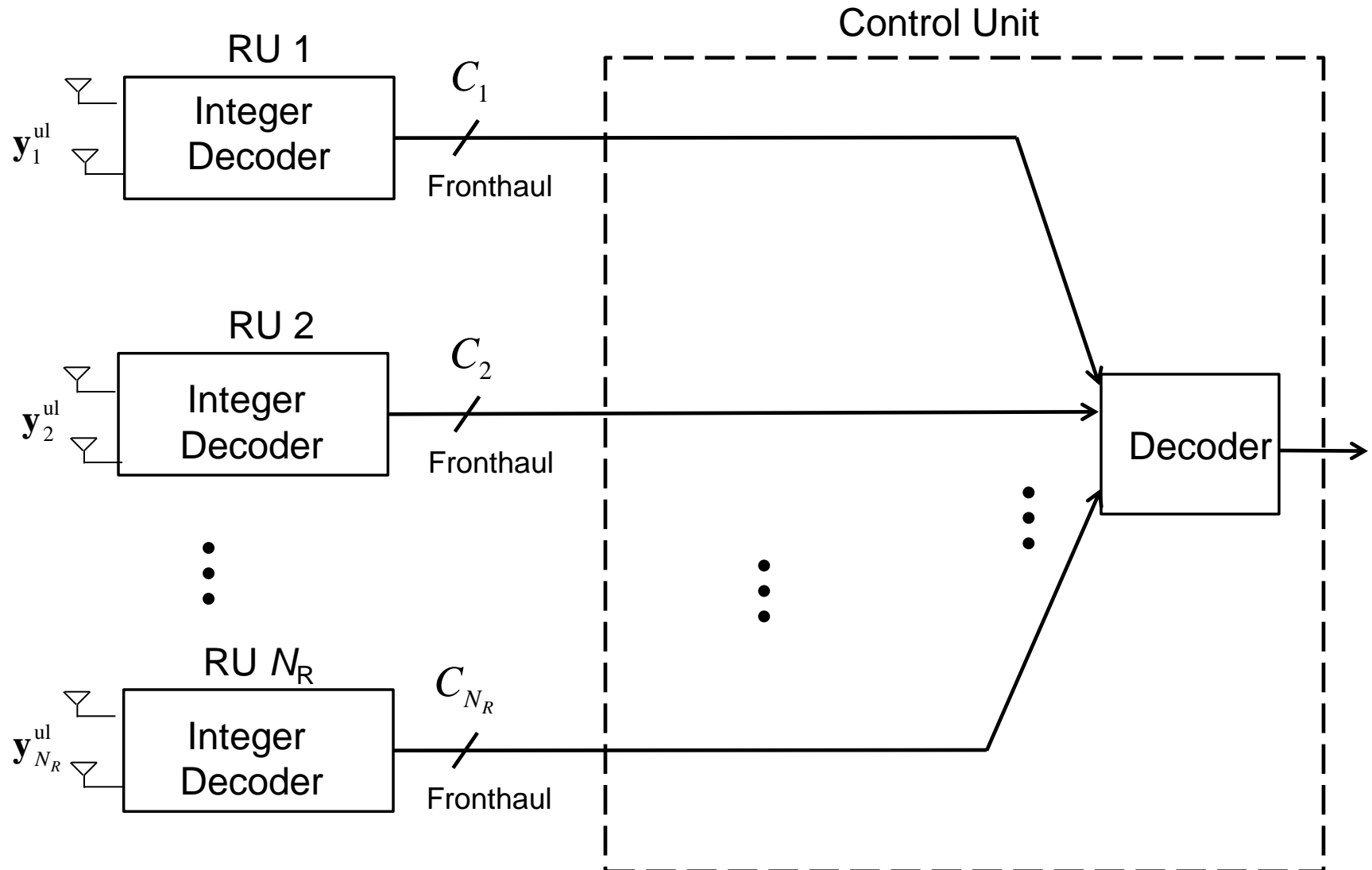
D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Compute-and-Forward

[Nazer et al '09] [Hong and Caire '11]



Compute-and-Forward

[Nazer et al '09] [Hong and Caire '11]

- Compute-and-forward (CoF) [Nazer et al '09]
 - The same codebook is used both for channel encoding at all MSs and for quantization at RUs.
 - Each RU decodes an appropriate (modulo-)sum, with integer weights, of the codewords transmitted by MSs.
 - And then sends a bit stream on the fronthaul link that identifies the decoded codeword within the lattice code.
 - Upon receiving a sufficient number of linear combinations, the CU can invert the resulting linear system and recover the transmitted codewords.
 - For single-antenna uplink system with $N_M = N_R$ and $C_1 = \dots = C_{N_R} = C$, achievable rate per MS is given by

$$R_{\text{per-MS}} = \min \left\{ C, \min_{l \in \mathcal{L}} R(\mathbf{h}_l, \mathbf{a}_l, \text{SNR}) \right\} \quad \text{where} \quad R(\mathbf{h}, \mathbf{a}, \text{SNR}) = \max \left\{ \log \left(\frac{\text{SNR}}{\mathbf{a}^H (\text{SNR}^{-1} \mathbf{I} + \mathbf{h}\mathbf{h}^H)^{-1} \mathbf{a}} \right), 0 \right\};$$

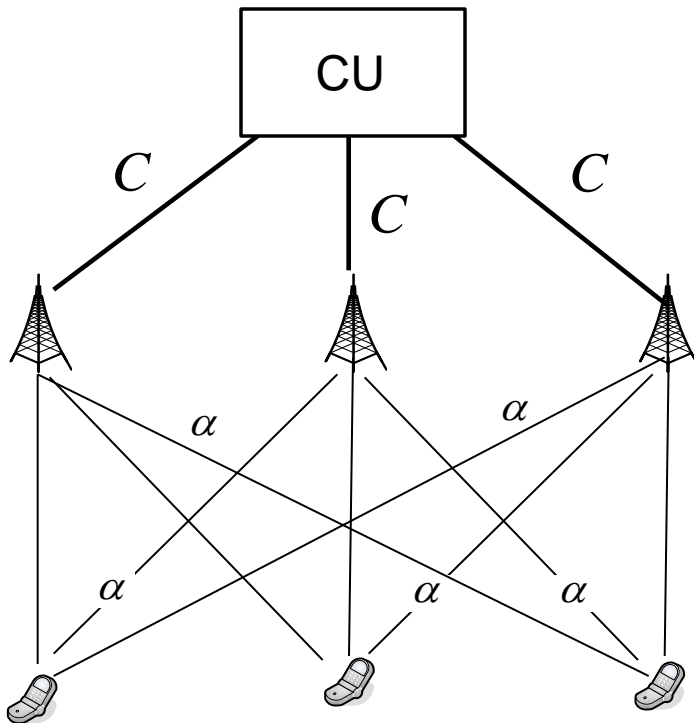
Integer penalty: The signal received at each RU is sum with *non-integer weights* of the codewords transmitted by MSs.

\mathbf{h}_l : channel vector toward RU l ;

\mathbf{a}_l : target integer vector for RU l .

Numerical Example

- Three-cell SISO circular Wyner model



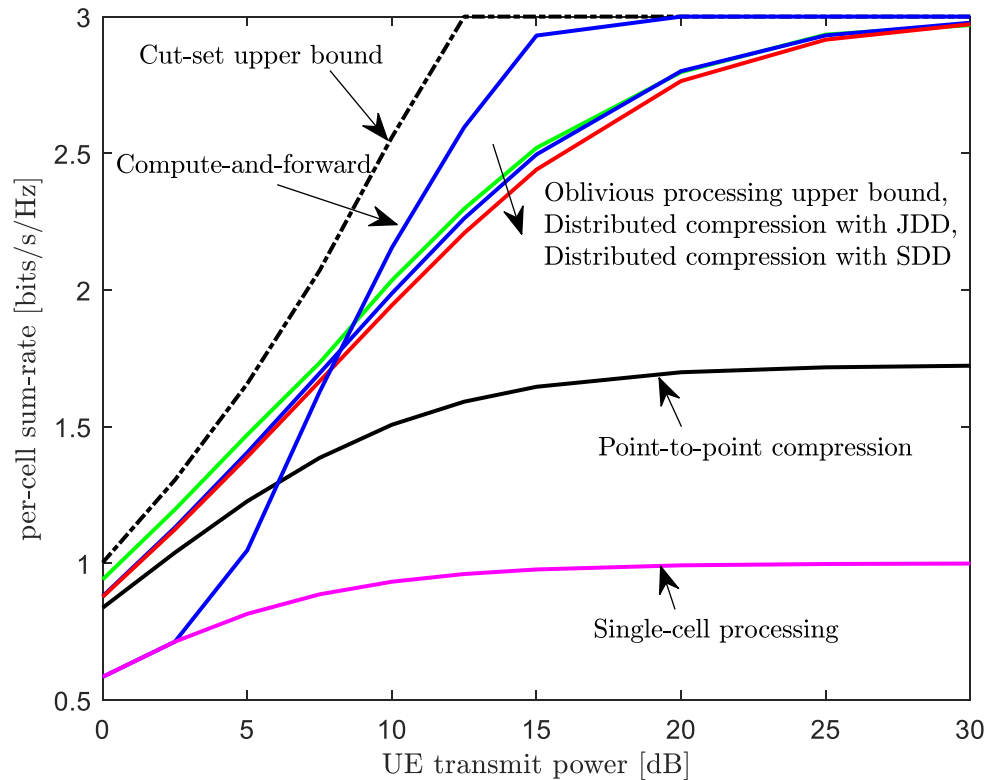
- Each cell contains a single-antenna and a single-antenna RU.
- Inter-cell interference takes place only between adjacent cells.
- The intra-cell and inter-cell channel gains are given by 1 and α , respectively.
- All RUs have a fronthaul capacity of C .

Numerical Example

- Compare the following schemes
 - Single-cell processing
 - Each RU decodes the signal of the in-cell MS by treating all other MSs' signals as noise.
 - Point-to-point fronthaul compression
 - Each RU compresses the received baseband signal and the quantized signals are decompressed in parallel at the control unit.
 - Distributed fronthaul compression [dCoso-Simoens '09]
 - Each RU performs Wyner-Ziv coding on the received baseband signal and the quantized signals are successively recovered at the control unit.
 - Joint Decompression and Decoding (noisy network coding [Sanderovich et al '08])
 - Compute-and-forward [Hong-Caire '11]
 - Each RU performs structured coding.
 - Oblivious processing upper bound
 - RUs cooperate and optimal compression is done over $3C$ fronthaul link.
 - Cutset upper bound [Simeone et al '12]

Numerical Example

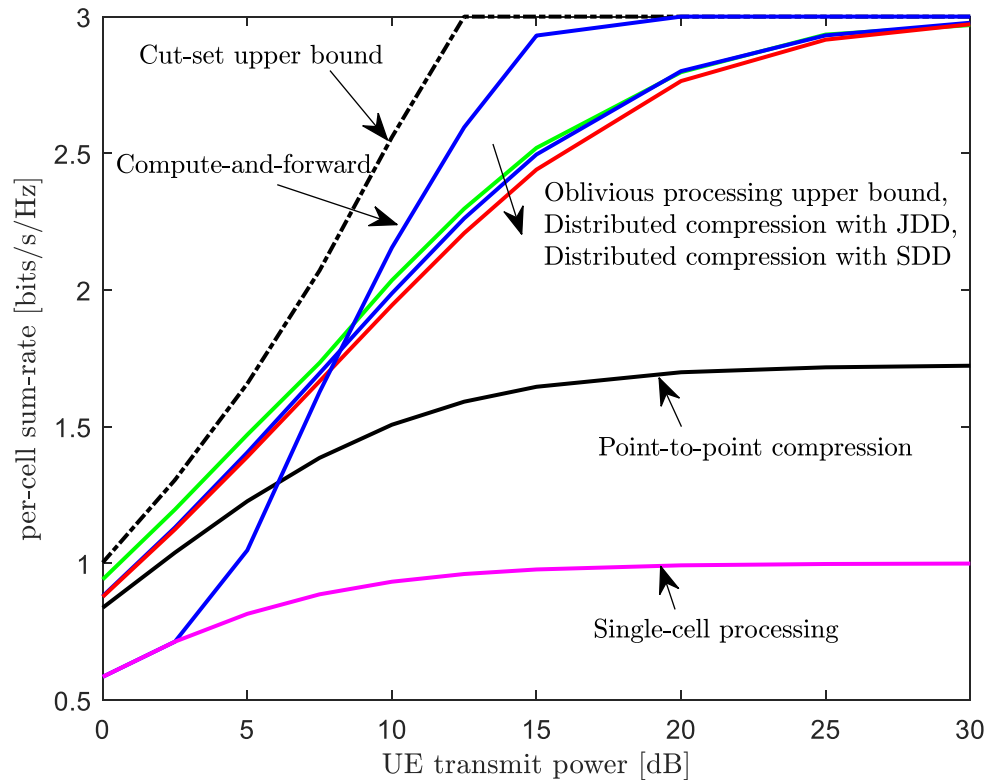
$$\alpha = 1/\sqrt{2} \text{ and } C = 3 \text{ bit/s/Hz}$$



- The performance advantage of distributed compression over point-to-point compression increases as SNR grows larger.
 - At high SNR, the correlation of the received signals at RUs becomes more pronounced.
- Compute-and-Forward
 - At low SNR, its performance coincides with single-cell processing.
 - RUs tend to decode trivial combinations.
 - At high SNR, the fronthaul capacity is the main performance bottleneck, so CoF shows the best performance.

Numerical Example

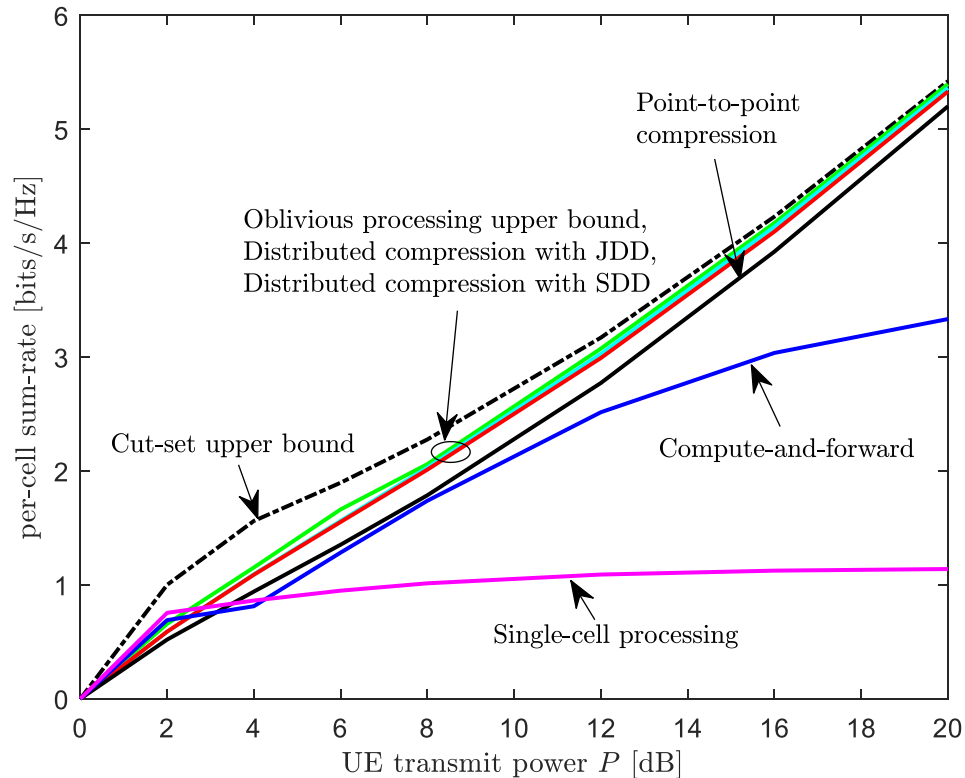
$$\alpha = 1/\sqrt{2} \text{ and } C = 3 \text{ bit/s/Hz}$$



- Distributed compression
 - Joint decompression and decoding does not provide much gain compared to separate decompression and decoding.
 - Optimality of joint decompression and decoding in symmetric case [Aguerri et al '17].

Numerical Example

$$\alpha = 1/\sqrt{2} \text{ and } C = 5 \log_{10} P \text{ bit/s/Hz}$$



- When C increases as $\log(\text{snr})$, CoF is not the best for high SNR.
 - i.e., if C does not limit the performance, the oblivious compression technique will be advantageous than CoF.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

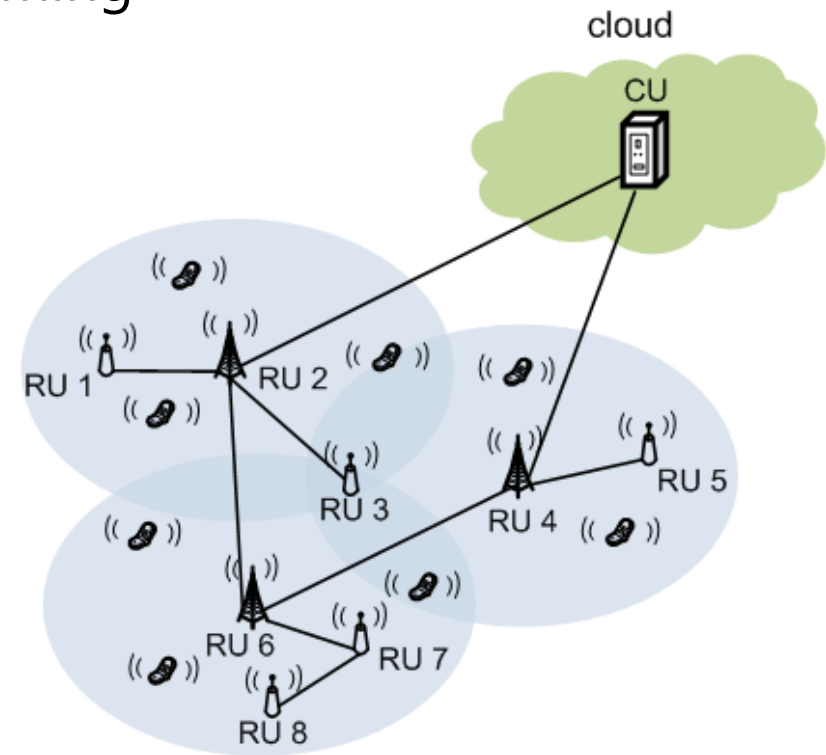
D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

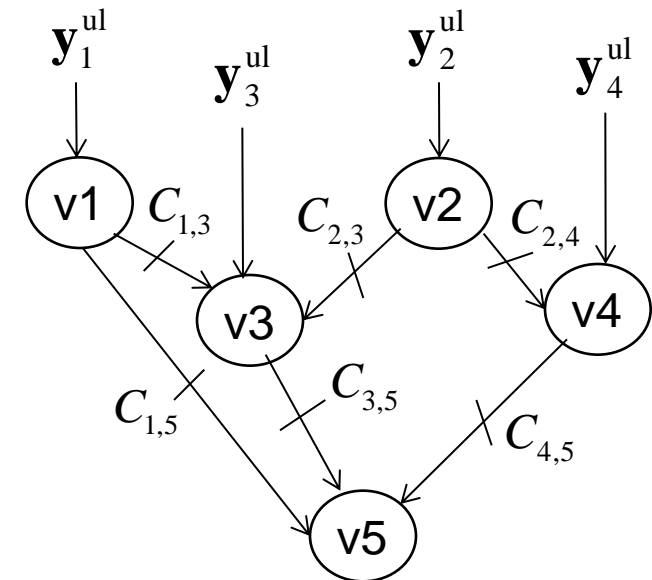
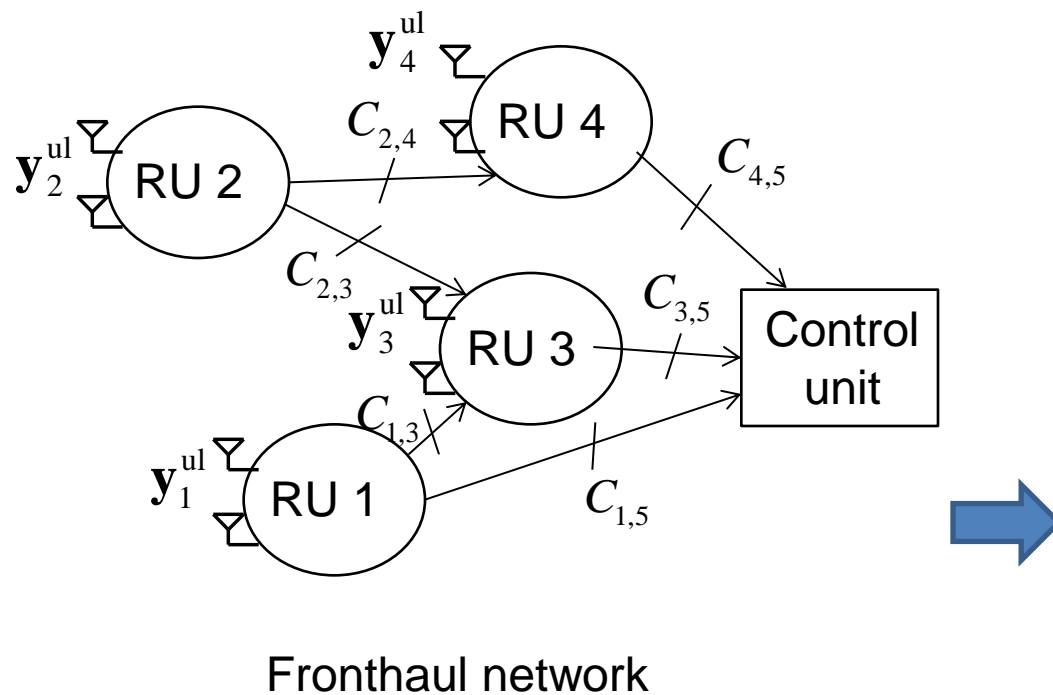
Cloud Radio Access Networks

- In multihop fronthaul networks, each RU may have multiple incoming and outgoing fronthaul links.
- For example, RU 6 in the figure has two incoming and single outgoing links.
- Two different operations, *routing* and *in-network processing*, were compared in [Park et al TVT'15].



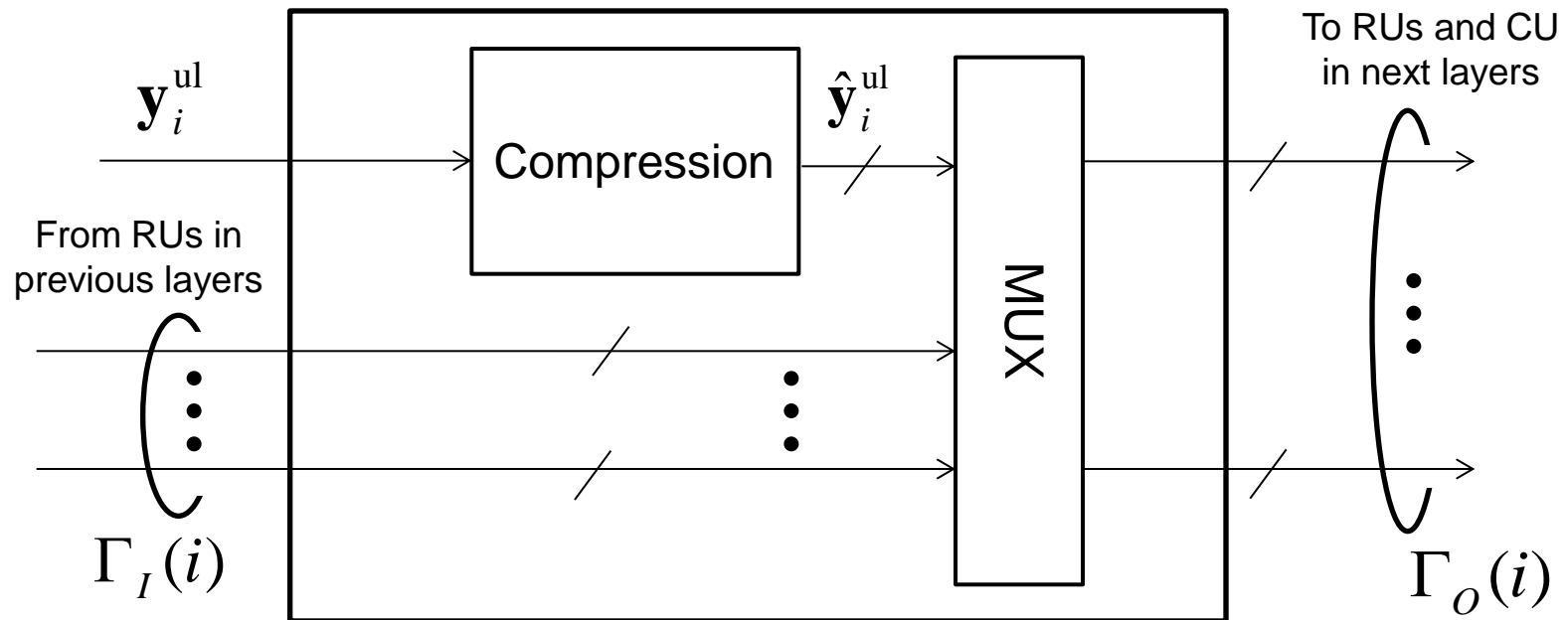
Directed Acyclic Graph

- Multihop fronthaul network modeled as a directed acyclic graph (DAG) [Koetter-Medard '03]



Routing

RU i

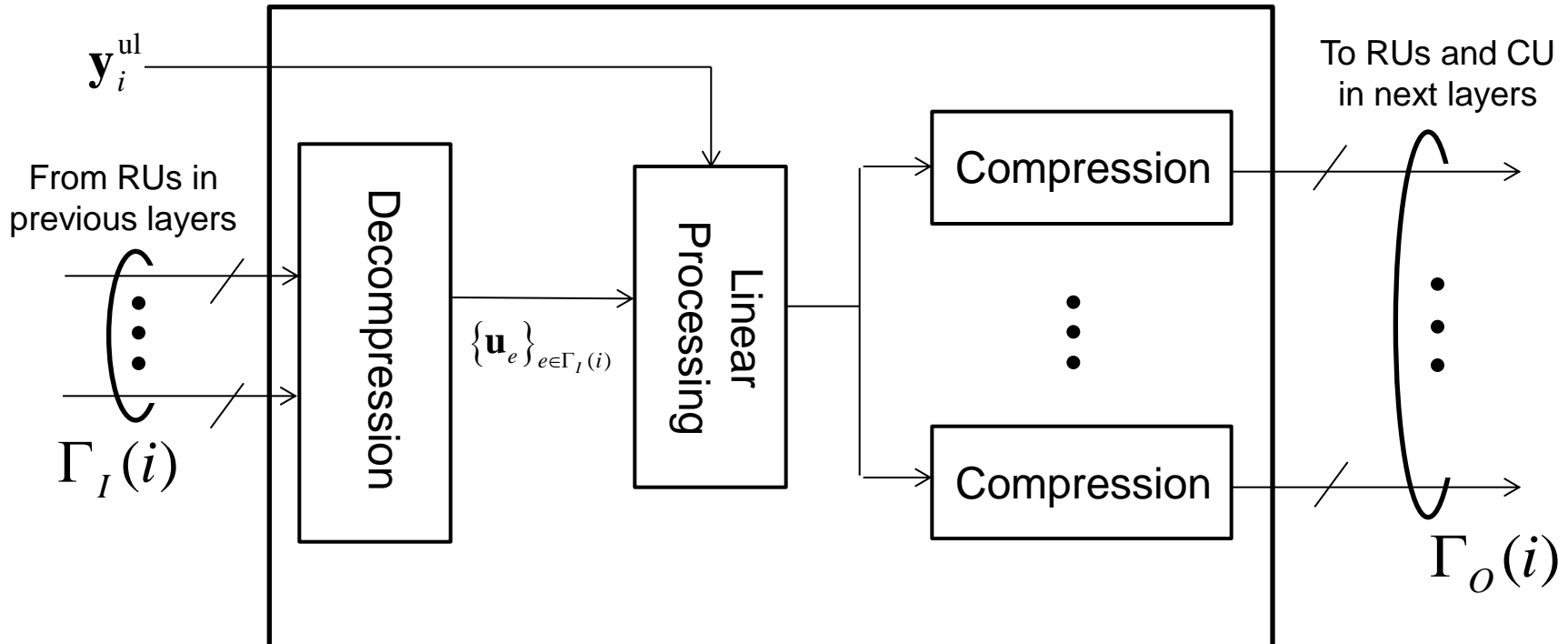


- The bits received on the incoming links are simply **forwarded** on the outgoing links **without any addition processing**.
- This approach requires the optimization of standard flow variables that define the allocation of fronthaul capacity to different bit streams.
 - In [Park et al TVT'15], the problem was addressed via the Majorization Minimization (MM) algorithm [Beck-Teboulle '11].

In-Network Processing

[Park et al TVT'15]

RU i



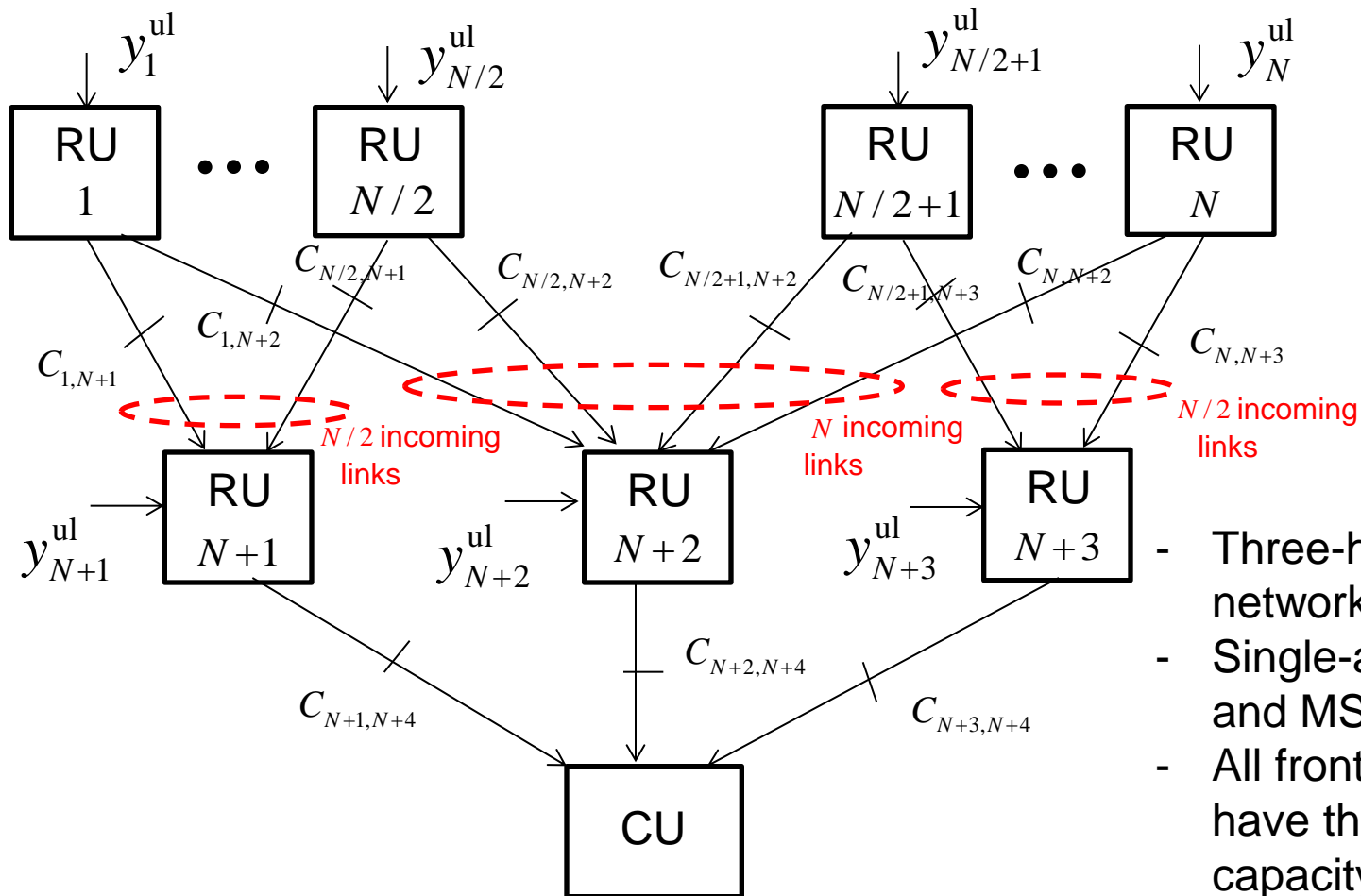
- In a **dense deployment of RUs**, an RU may be connected to a large number of nearby RUs, all of which receive **correlated baseband signals**.
- It is possible to **combine the correlated baseband signals** at the RU in order to reduce redundancy.

In-Network Processing

[Park et al TVT'15]

- In in-network processing, the RU must first **decompress** the received bit streams.
- The decompressed baseband signals are **linearly processed**, along with the IQ signal received locally by the RU.
- The in-network processed signal must be **recompressed** before being sent on the outgoing fronthaul links.
 - The effect of the resulting quantization noise must be counterbalanced by the advantage of in-network processing.
- The optimization of both routing and in-network processing was addressed in [Park et al TVT'15].

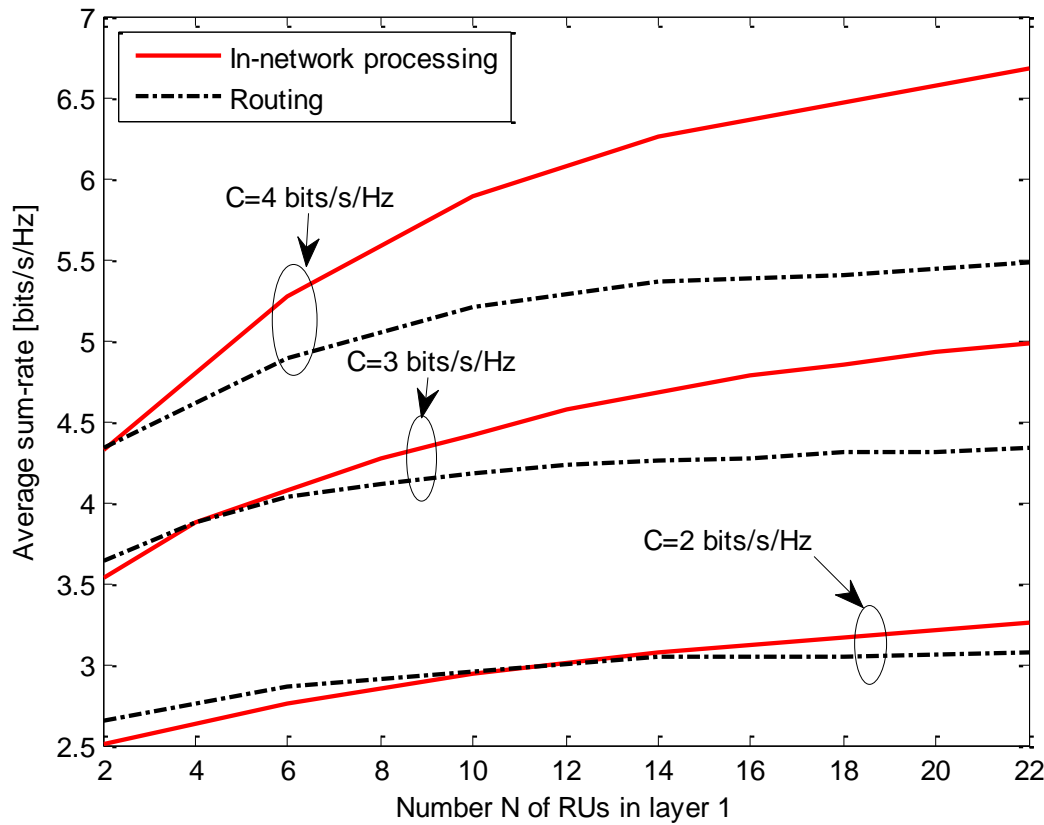
Numerical Example



- Three-hop fronthaul network
- Single-antenna RUs and MSs
- All fronthaul links have the same capacity C .
- i.i.d. Rayleigh fading channels with unit power

Numerical Results

4 MSs, average received per-antenna SNR of 20 dB



- The performance gain of in-network processing over routing becomes more pronounced as the number N of RUs in the first layer increases.
- As the density of the RUs' deployment increases, it is desirable for each RU in layer 2 perform in-network processing.
- In-network processing is more advantageous when the fronthaul links have larger capacity, as the distortion introduced by the recompression step becomes smaller.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

System Model

- C-RAN with inter-connected RUs, Wyner-type example

[Shama et al ITA '17]

- N pairs of RU-UE ($\mathcal{N} = \{1, 2, \dots, N\}$)

- Fronthaul connections

- C bit/symbol between RU-CU
- B bit/symbol between RU-RU

- Uplink channel

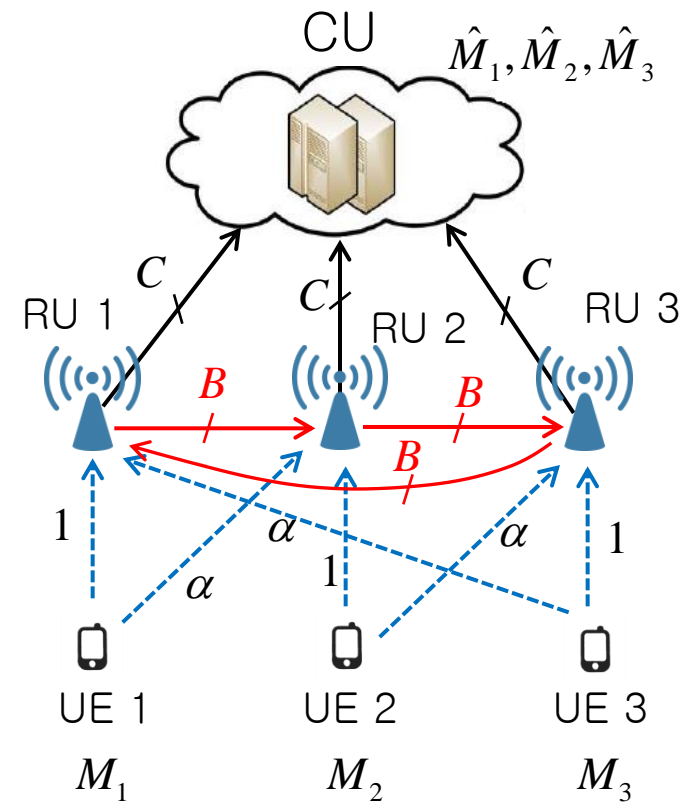
$$Y_i = X_i + \alpha X_{[i-1]} + Z_i,$$

where Y_i : Rx signal RU i ,

X_i : Tx signal of UE i ,

Z_i : Noise at RU i with $Z_i \sim N(0, \sigma^2)$,

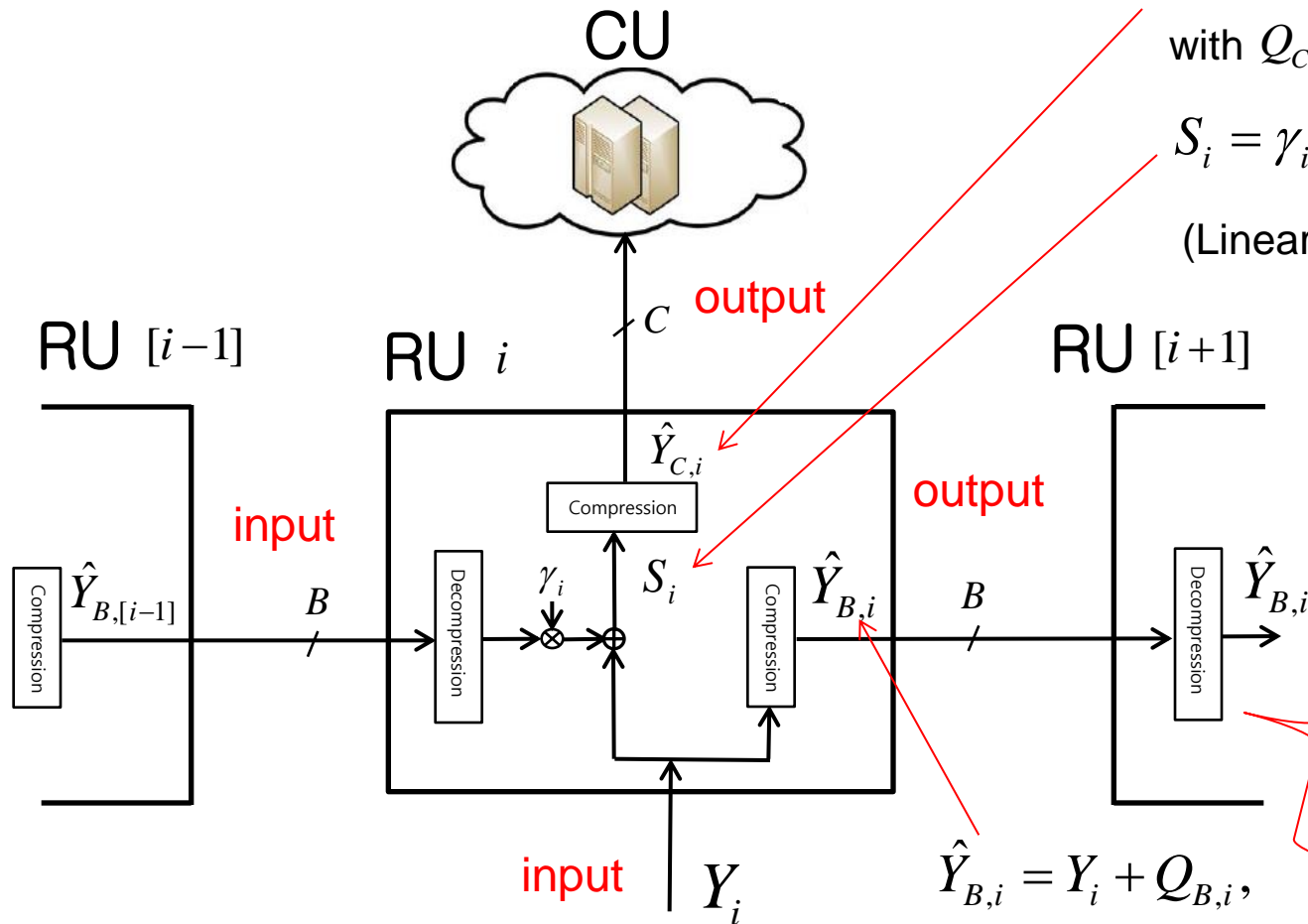
α : Inter-cell channel gain with $\alpha \in [0, 1]$.



<Example for $N = 3$ >

Oblivious Processing at RUs

- In-network processing (INP) at **RU i**



$$\hat{Y}_{C,i} = S_i + Q_{C,i}$$

with $Q_{C,i} \sim N(0, \omega_{C,i})$

$$S_i = \gamma_i \hat{Y}_{B,[i-1]} + Y_i$$

(Linear is optimal.)

$$\hat{Y}_{B,i} = Y_i + Q_{B,i},$$

with quantization noise $Q_{B,i} \sim N(0, \omega_{B,i})$

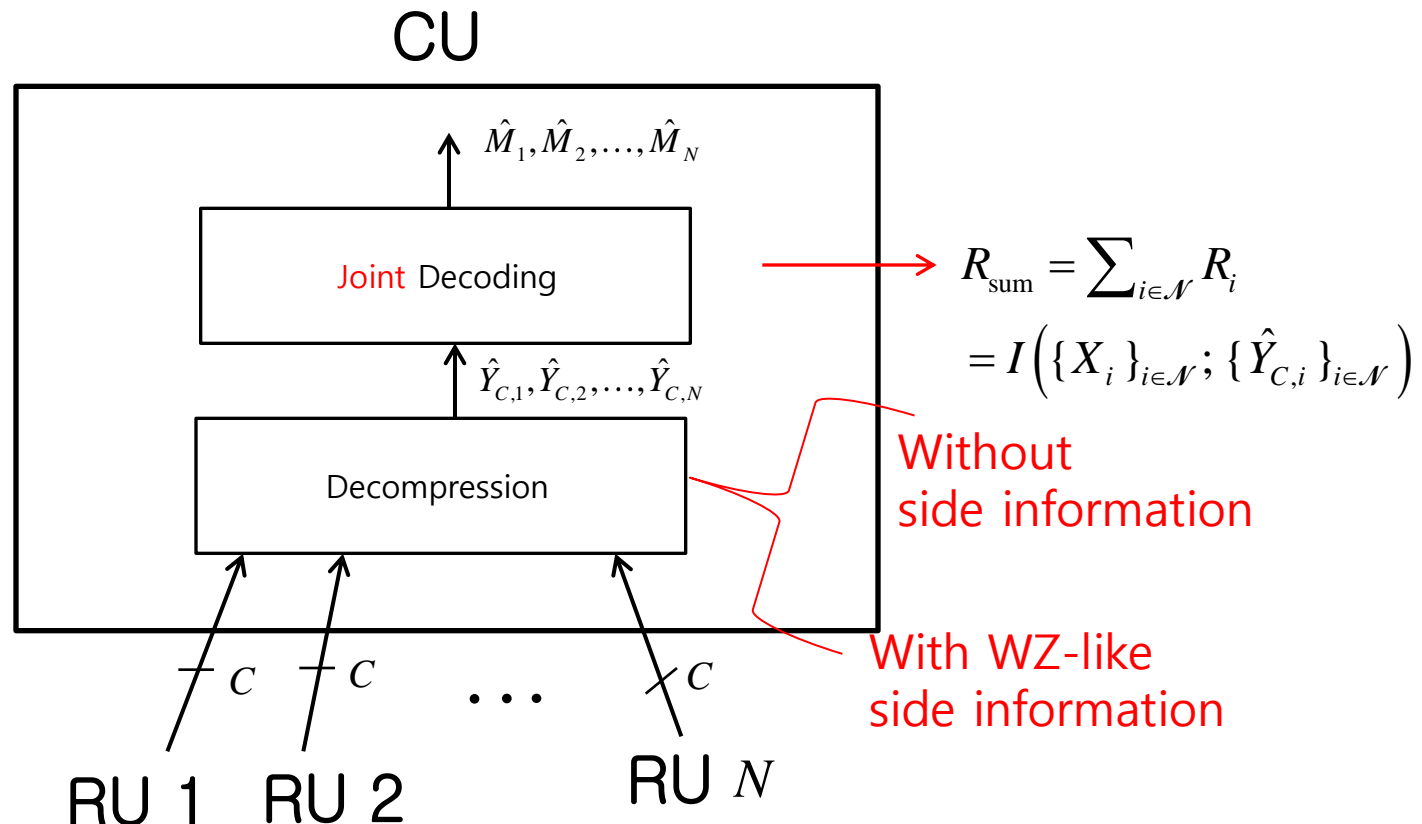
Without side information

With WZ-like side information

- Oblivious/Nomadic: no structure information (code-books) of UE's is available at the RUs

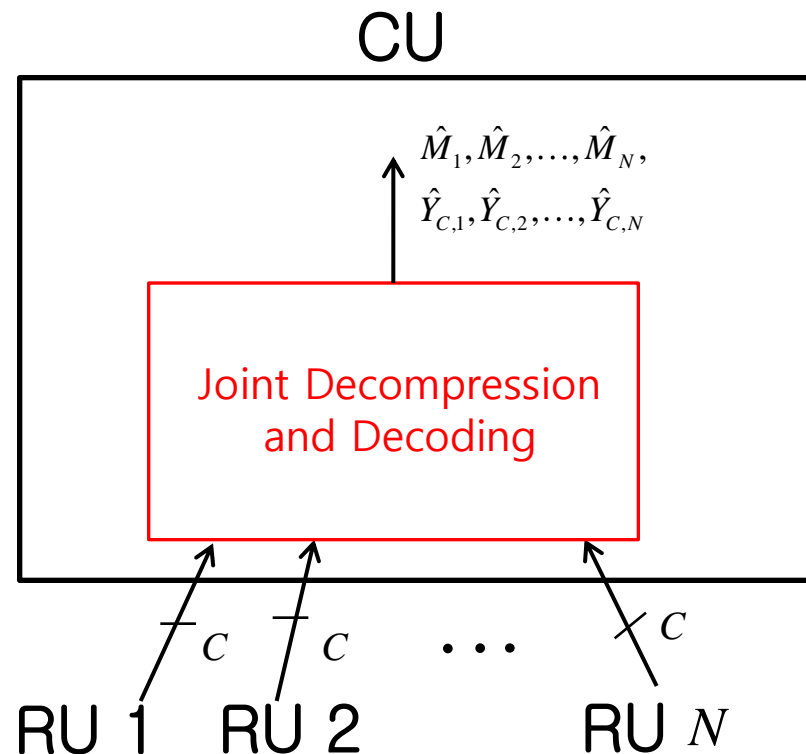
Decoding at CU

- Decompression and decoding at CU
 - CU recovers the quantized INP output signals $\hat{Y}_{C,1}, \hat{Y}_{C,2}, \dots, \hat{Y}_{C,N}$.
 - Then, it jointly decodes the messages $\hat{M}_1, \hat{M}_2, \dots, \hat{M}_N$.



Joint Decompression and Decoding

- Joint decompression and decoding (JDD)
[Sanderovich et al '09][Lim et al '11][Yassaee-Aref '11]
 - JDD at CU can potentially improve the performance
[Sanderovich et al '08][Sanderovich et al '09].
 - Optimal oblivious processing [Aguerri et al '17].



Upper Bounds

- Cut-Set upper bound $R_{\text{cut-set}}$

$$R_{\text{cut-set}} = \min \{ NC, R_{\text{full}} \},$$

where R_{full} is the sum-rate achievable when full cooperation among RUs is possible.

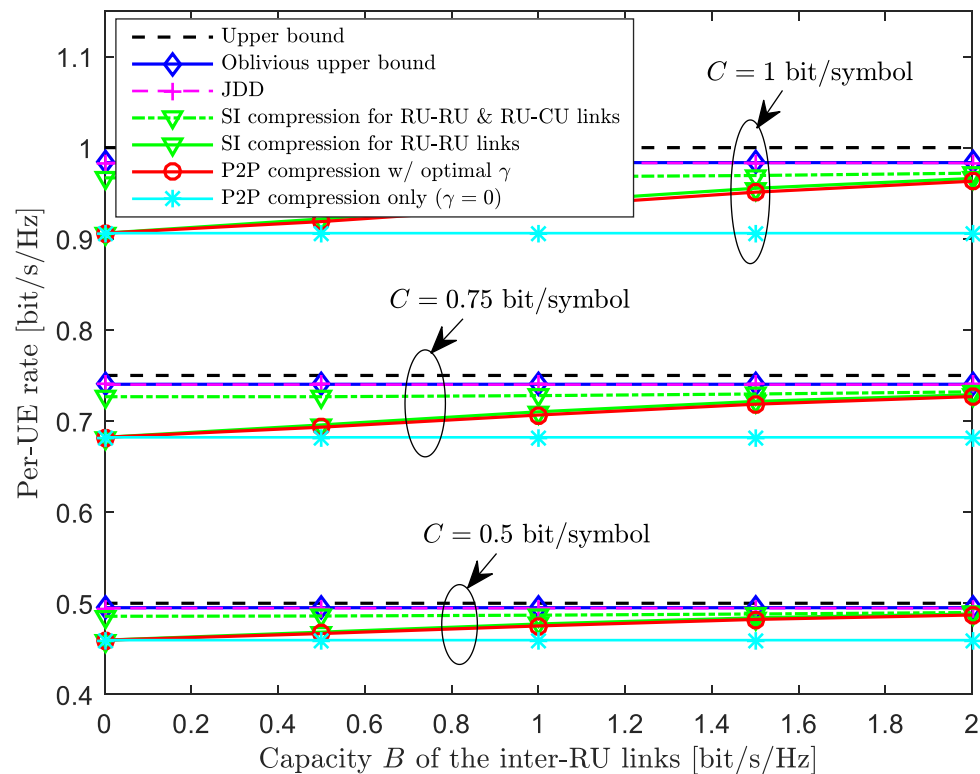
- Oblivious upper bound $R_{\text{oblv-UB}}$

- $R_{\text{oblv-UB}}$ is the rate achievable when the RUs are colocated and connected to the CU with capacity NC .

Numerical Example

● Per-UE rate versus RU-RU capacity B

● $N = 3$, SNR = 20 dB, $\alpha = 0.7$

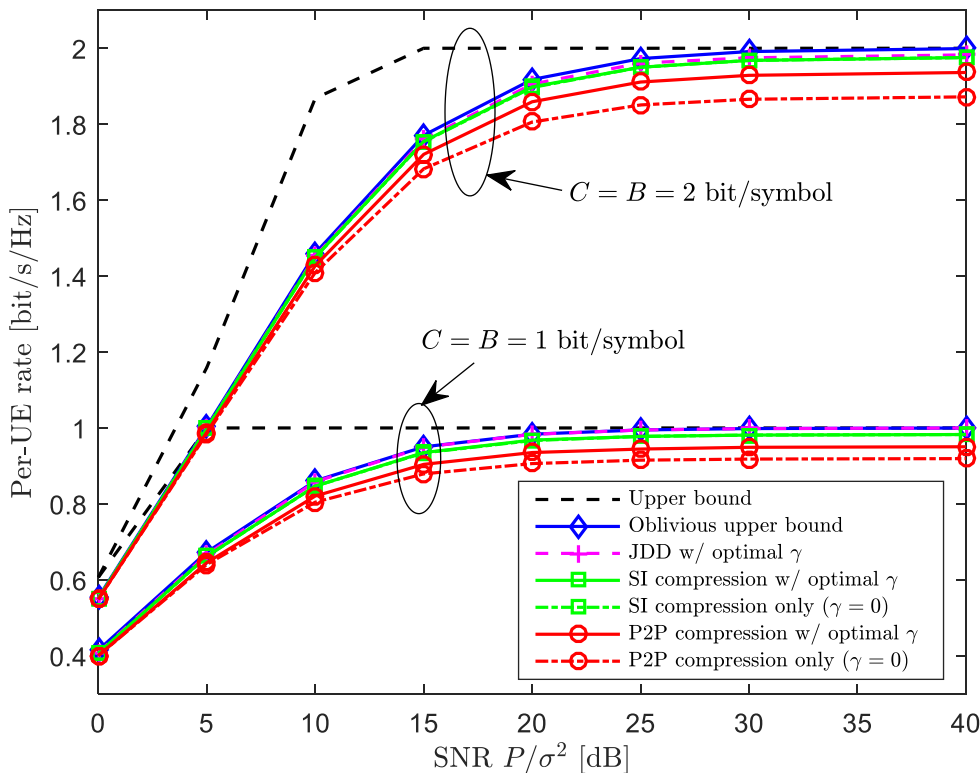


- With INP, the performance approaches upper bound as B increases.
- Leveraging SI for RU-RU link provides a slight sum-rate gain.
- Leveraging SI for RU-CU link leads to a significant sum-rate gain especially for small B .
- JDD further improves the sum-rate performance.
 - Its performance is very close to oblivious upper bound.

Numerical Example

- Per-UE rate versus SNR P / σ^2

- $N = 3, \alpha = 0.7, C = B \in \{1, 2\}$



- JDD shows slightly improved performance, but the gap to upper bound is still large.
- This calls for the development of
 - Improved scheme based on
 - Non-oblivious RU processing
 - Improved upper bound
 - Extending the idea as [Wu et al '17]

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

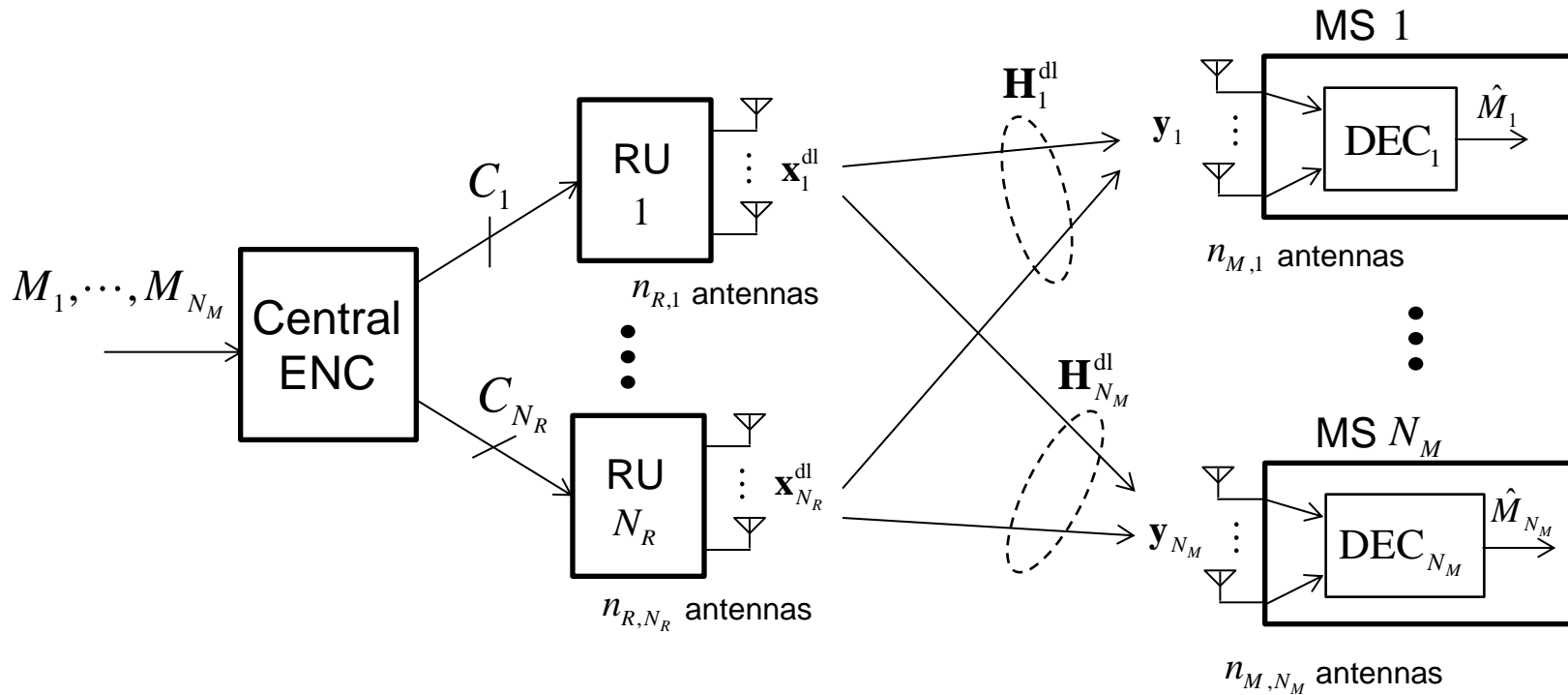
C. Structured coding

D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

System Model



System Model

- The signal \mathbf{y}_k^{dl} received by **MS k** in the downlink

$$\mathbf{y}_k^{\text{dl}} = \mathbf{H}_k^{\text{dl}H} \mathbf{x}^{\text{dl}} + \mathbf{z}_k^{\text{dl}},$$

where

$\mathbf{x}^{\text{dl}} = [\mathbf{x}_1^{\text{dl}} \dots \mathbf{x}_{N_R}^{\text{dl}}]^T$: vector of symbols transmitted by all RUs;

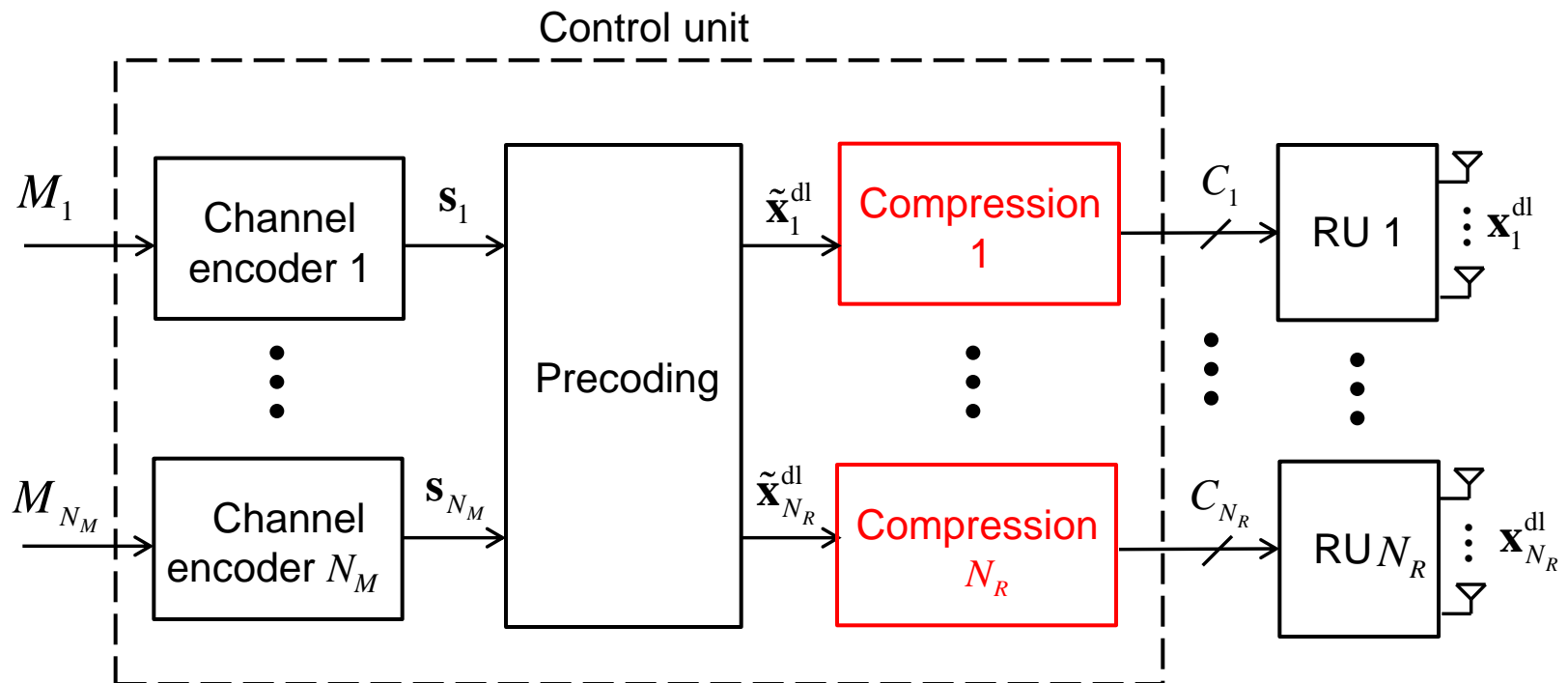
$\mathbf{z}_k^{\text{dl}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{\mathbf{z}_i^{\text{dl}}})$: noise and interference arising from the other clusters;

\mathbf{H}_k^{dl} : channel vector from all RUs toward MS k .

- Per-RU power constraint: $E \|\mathbf{x}_i^{\text{dl}}\|^2 \leq P_{R,i}$, $i \in \{1, \dots, N_R\}$.

Point-to-Point Compression

[Simeone et al '09]

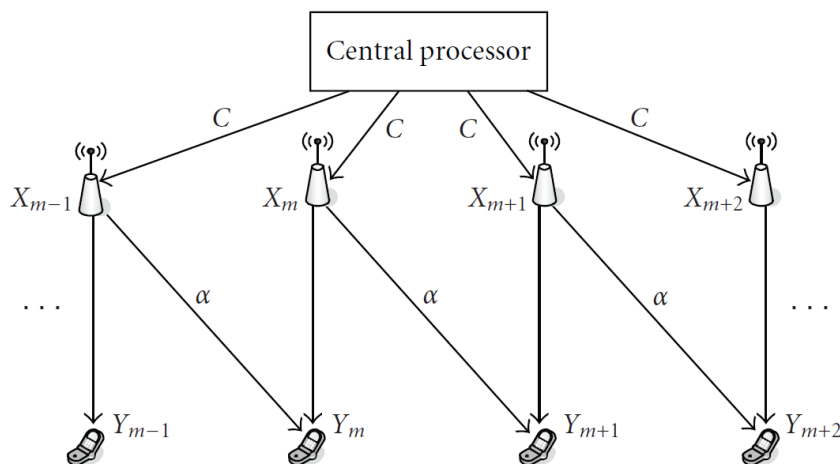


Baseband signals for different RUs are **separately** compressed.

- For precoding, both linear precoding [Huh et al '10] and non-linear dirty-paper coding [Costa '83] can be considered.

Asymmetric Wyner model Downlink: Independent Compression

- Compressed dirty-paper coding (CDPC) [Simeone et al '09]



System model

Quantization is performed at the central unit using the forward test channel

$$X_m = \tilde{X}_m + Q_m,$$

where \tilde{X}_m : DPC precoding output,

Q_m : quantization noise with $Q_m \sim \mathcal{CN}(0, P/2^C)$,

m : cell-index, thus Q_m is independent over the index m .

- With constrained backhaul links, we obtain a modified broadcast channel (BC) with the added quantization noises.

- Per-cell sum-rate

$$R_{\text{per-cell}} = \log \left(\frac{1 + (1 + \alpha^2)\tilde{P} + \sqrt{1 + 2(1 + \alpha^2)\tilde{P} + (1 - \alpha^2)^2\tilde{P}^2}}{2} \right)$$

where \tilde{P} is the effective SNR at the MSs decreased from P to

$$\tilde{P} = \frac{P}{(1 + (1 + \alpha^2)P) / (2^C - 1) + 1}.$$

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

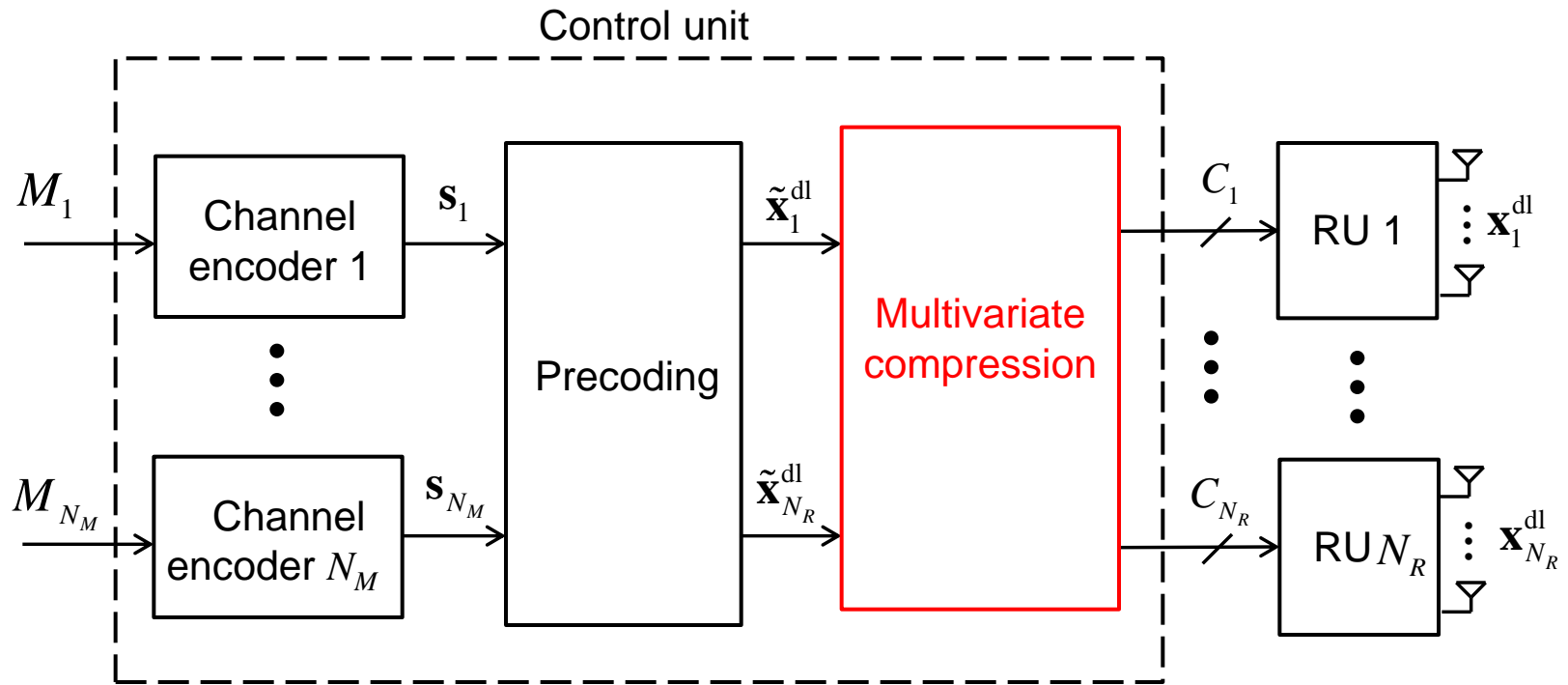
D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Multivariate Compression

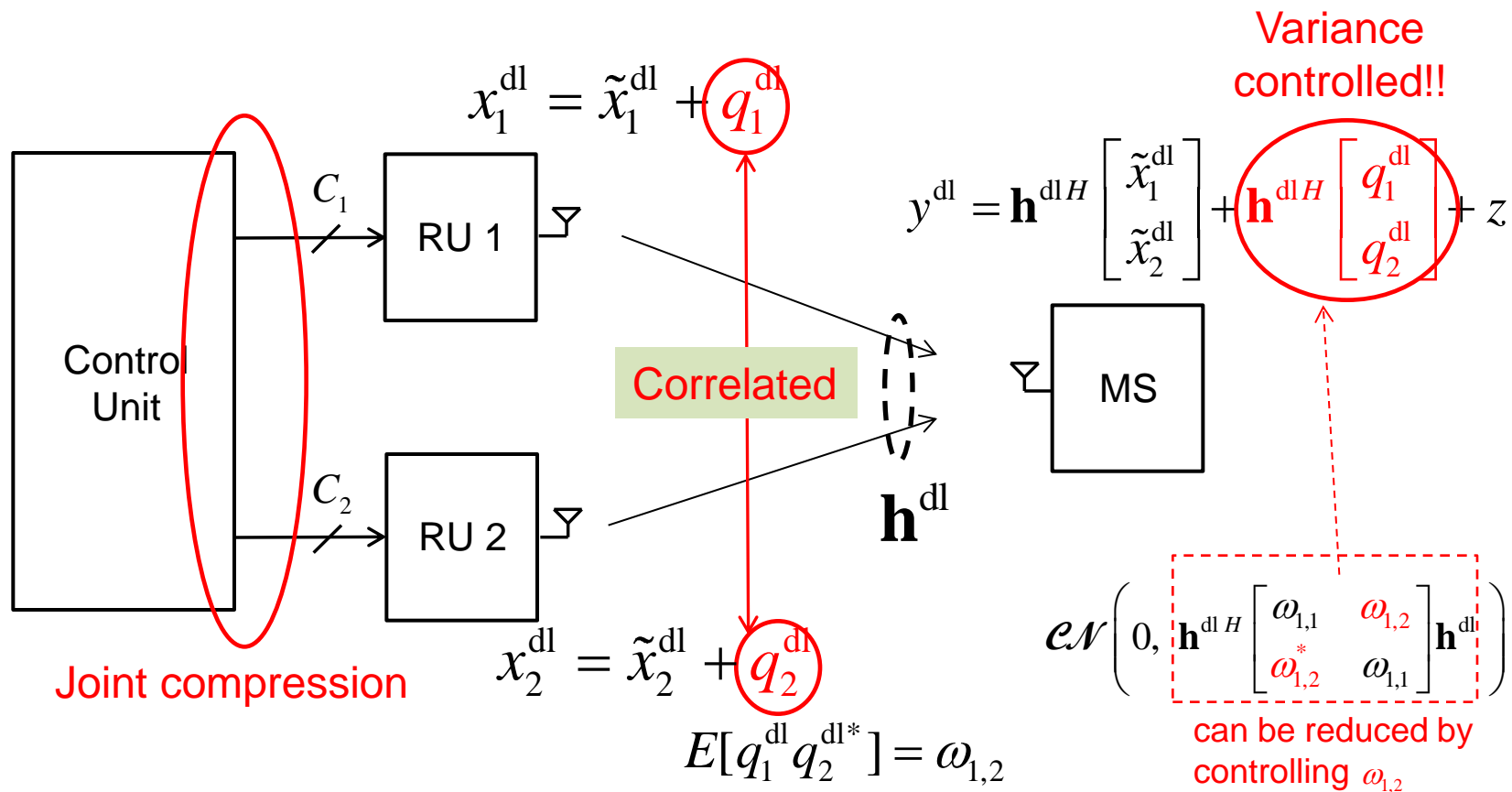
[Park et al TSP'13]



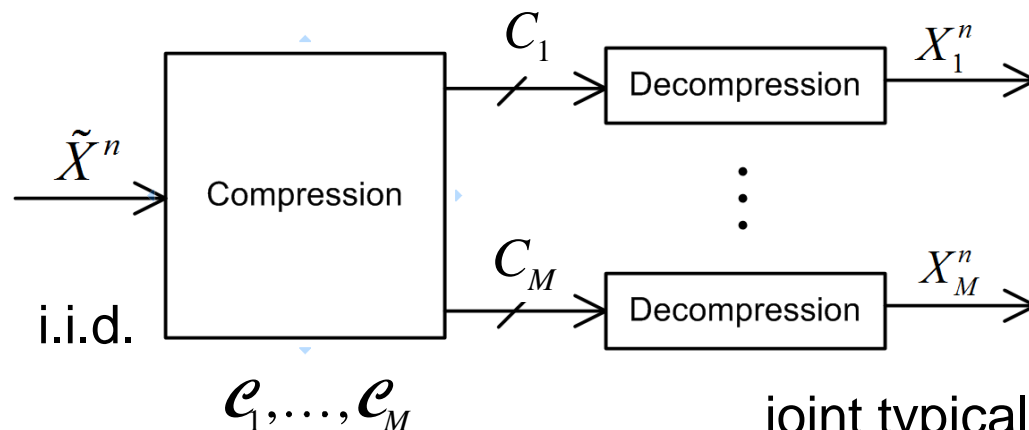
Baseband signals for different RUs are **jointly** compressed.

Multivariate Compression

- Multivariate compression produces compressed signals with correlated quantization noises
- Noise correlation enables finer control of effect quantization at the MSs



Multivariate Compression Lemma



joint typicality wrt

$$p(\tilde{x}, x_1, \dots, x_M) = p(\tilde{x})p(x_1, \dots, x_M | \tilde{x})$$

$$\sum_{i \in \mathcal{S}} h(X_i) - h(X_{\mathcal{S}} | \tilde{X}) \leq \sum_{i \in \mathcal{S}} C_i, \text{ for all } \mathcal{S} \subseteq \{1, \dots, M\}$$

[ElGamal-Kim '11, Ch. 9]

Multivariate Compression

- Linear precoding (DPC treated in a similar way)
- Gaussian test channel:

$$\mathbf{x}_i^{\text{dl}} = \tilde{\mathbf{x}}_i^{\text{dl}} + \mathbf{q}_i^{\text{dl}}, \quad \mathbf{q}_i^{\text{dl}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{i,i}^{\text{dl}}), \quad i \in \mathcal{N}_R$$

- The compressed signal $\mathbf{x}^{\text{dl}} = [\mathbf{x}_1^{\text{dl}H}, \dots, \mathbf{x}_{N_R}^{\text{dl}H}]^H$ is given as

$$\mathbf{x}^{\text{dl}} = \mathbf{A}\mathbf{s} + \mathbf{q}^{\text{dl}},$$

with $\mathbf{q}^{\text{dl}} = [\mathbf{q}_1^{\text{dl}H}, \dots, \mathbf{q}_{N_R}^{\text{dl}H}]^H \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}^{\text{dl}})$ and

$$\mathbf{\Omega}^{\text{dl}} = \begin{bmatrix} \mathbf{\Omega}_{1,1}^{\text{dl}} & \mathbf{\Omega}_{1,2}^{\text{dl}} & \cdots & \mathbf{\Omega}_{1,N_R}^{\text{dl}} \\ \mathbf{\Omega}_{2,1}^{\text{dl}} & \mathbf{\Omega}_{2,2}^{\text{dl}} & \cdots & \mathbf{\Omega}_{2,N_R}^{\text{dl}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Omega}_{N_R,1}^{\text{dl}} & \mathbf{\Omega}_{N_R,2}^{\text{dl}} & \cdots & \mathbf{\Omega}_{N_R,N_R}^{\text{dl}} \end{bmatrix}$$

(Independent compression is a special case with $\mathbf{\Omega}_{i,j}^{\text{dl}} = \mathbf{0}$, $i \neq j \in \mathcal{N}_R$.)

Optimization

- Weighted sum-rate maximization

$$\underset{\mathbf{A}, \mathbf{\Omega}^{\text{dl}} \geq \mathbf{0}}{\text{maximize}} \quad \sum_{k=1}^{N_M} w_k f_k(\mathbf{A}, \mathbf{\Omega}^{\text{dl}}) \quad (1a)$$

$$\text{s.t.} \quad g_{\mathcal{S}}(\mathbf{A}, \mathbf{\Omega}^{\text{dl}}) \leq \sum_{i \in \mathcal{S}} C_i, \text{ for all } \mathcal{S} \subseteq \mathcal{N}_R, \quad (1b)$$

$$\text{tr}(\mathbf{E}_i^H \mathbf{A} \mathbf{A}^H \mathbf{E}_i + \mathbf{\Omega}_{i,i}^{\text{dl}}) \leq P_i, \text{ for all } i \in \mathcal{N}_R. \quad (1c)$$

where $f_k(\mathbf{A}, \mathbf{\Omega}^{\text{dl}}) = I(\mathbf{s}_k; \mathbf{y}_k^{\text{dl}})$

$$= \log \det(\mathbf{I} + \mathbf{H}_k^{\text{dl}} (\mathbf{A} \mathbf{A}^H + \mathbf{\Omega}^{\text{dl}}) \mathbf{H}_k^{\text{dl}H}) - \log \det\left(\mathbf{I} + \mathbf{H}_k^{\text{dl}} \left(\sum_{l \neq k} \mathbf{A}_l \mathbf{A}_l^H + \mathbf{\Omega}^{\text{dl}}\right) \mathbf{H}_k^{\text{dl}H}\right),$$

$$g_{\mathcal{S}}(\mathbf{A}, \mathbf{\Omega}^{\text{dl}}) = \sum_{i \in \mathcal{S}} h(\mathbf{x}_i^{\text{dl}}) - h(\mathbf{x}_{\mathcal{S}}^{\text{dl}} | \tilde{\mathbf{x}}^{\text{dl}})$$

$$= \sum_{i \in \mathcal{S}} \log \det(\mathbf{E}_i^H \mathbf{A} \mathbf{A}^H \mathbf{E}_i + \mathbf{\Omega}_{i,i}^{\text{dl}}) - \log \det(\mathbf{E}_{\mathcal{S}}^H \mathbf{\Omega}^{\text{dl}} \mathbf{E}_{\mathcal{S}}) \leq \sum_{i \in \mathcal{S}} C_i.$$

- Difference-of-convex (DC) problem: Local optimum via MM algorithm

[Beck-Teboulle '11]

Generalized Downlink Compression

[Patil-Yu TIT'18]

- Generalized downlink compression strategy [Patil-Yu TIT'18]
 - CU successively performs Marton's multicoding [Marton TIT'79] and multivariate compression [Park et al TSP'13].
 - Achievable rate region [Patil-Yu TIT'18, Thm. 1]
 - A rate tuple (R_1, \dots, R_{N_U}) is achievable if

$$\sum_{k \in \mathcal{D}} R_k \leq \sum_{k \in \mathcal{D}} I(U_k; Y_k) - \sum_{k \in \mathcal{D}} h(U_k) + h(\{U_k\}_{k \in \mathcal{D}}), \text{ for all } \mathcal{D} \subseteq \mathcal{N}_U,$$

$$\sum_{i \in \mathcal{S}} C_i \leq I(\{U_k\}_{k \in \mathcal{N}_U}; \{X_i\}_{i \in \mathcal{S}}) + \sum_{i \in \mathcal{S}} h(X_i) - h(\{X_i\}_{i \in \mathcal{S}}), \text{ for all } \mathcal{S} \subseteq \mathcal{N}_R,$$

for some joint distribution $p(u_1, \dots, u_{N_U}, x_1, \dots, x_{N_R})$.

- Comparison to Distributed Decode-Forward (DDF) [Lim et al TIT'17]
 - Application of DDF to C-RAN model was discussed in [Patil-Yu TIT'18].
 - DDF further generalizes "generalized downlink compression".
 - Marton's multicoding and multivariate compression are jointly performed.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

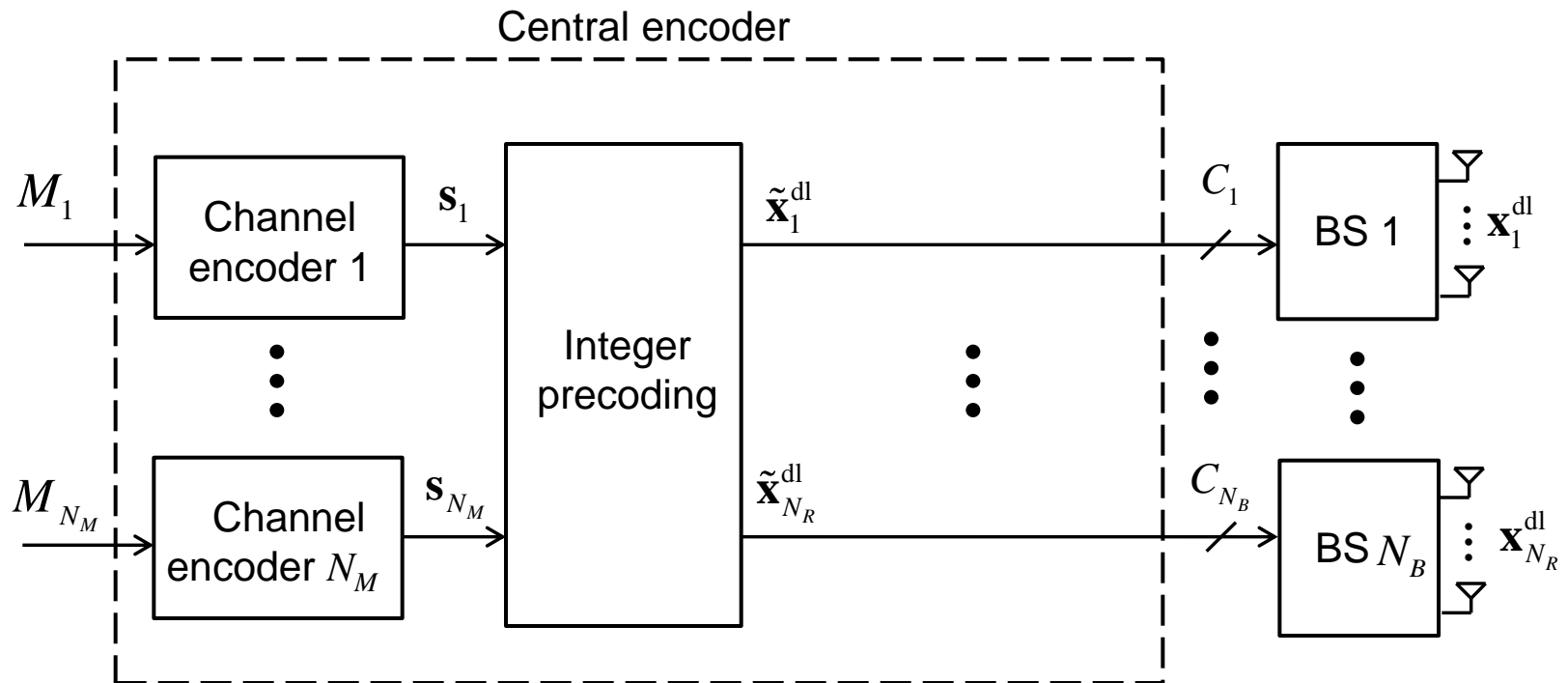
D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

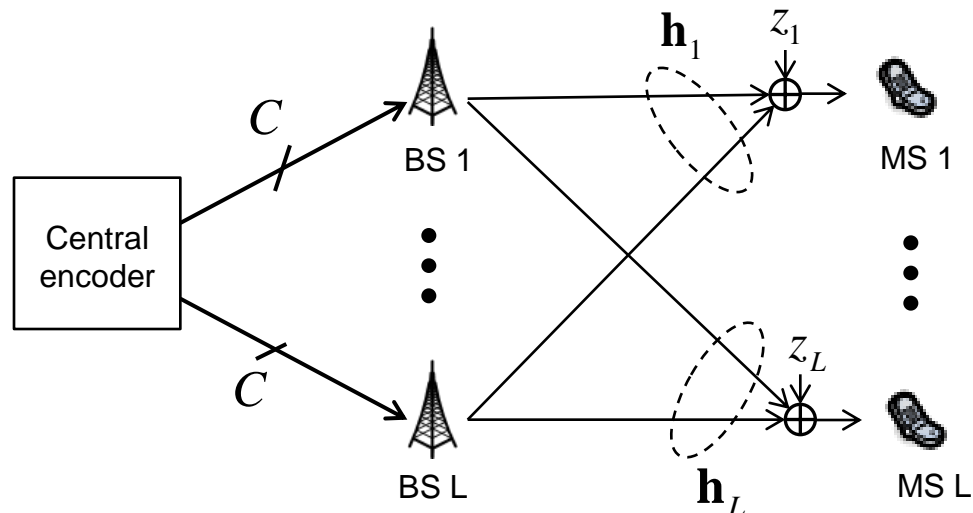
Structured Coding

- Reverse compute-and-forward (RCoF) [Hong-Caire '13]



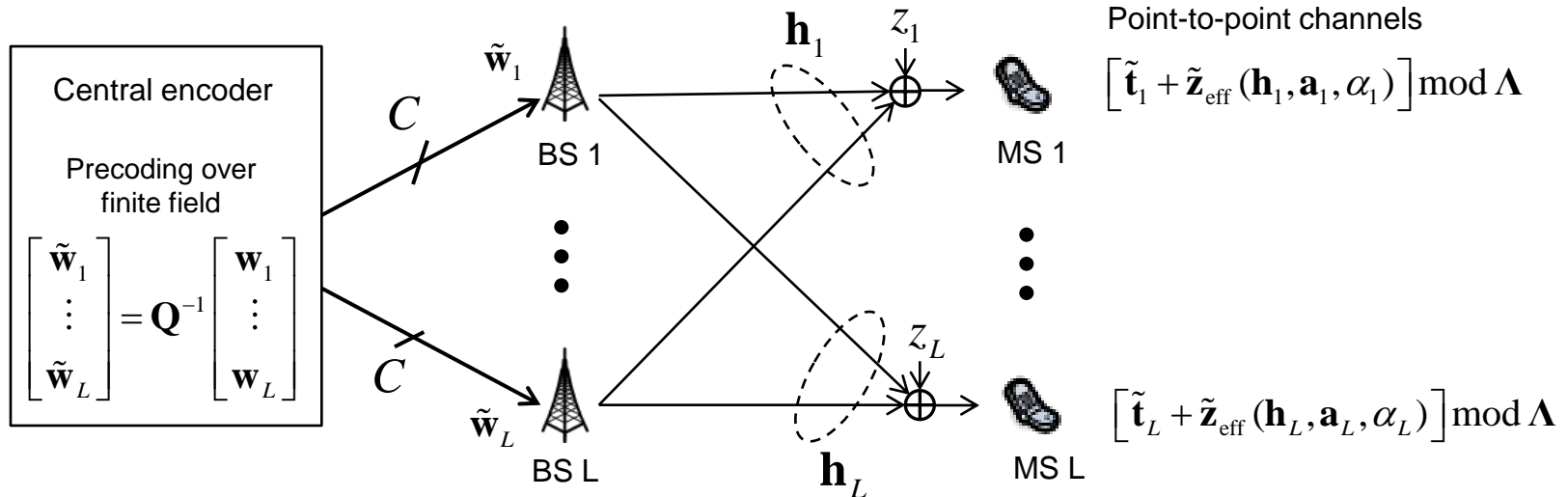
Structured Coding

- Reverse compute-and-forward (RCoF) [Hong-Caire '13]
 - Downlink counterpart of the compute-and-forward (CoF) scheme proposed for the uplink in [Nazer et al '09].
 - Exchange the role of BSs and MSs and use CoF in reverse direction.
 - System model
 - $N_B = N_M = L$, $C_i = C$ for all $i \in \mathcal{L} = \{1, \dots, L\}$.



Structured Coding

- Reverse compute-and-forward (RCoF) [Hong-Caire '13] (ctd')



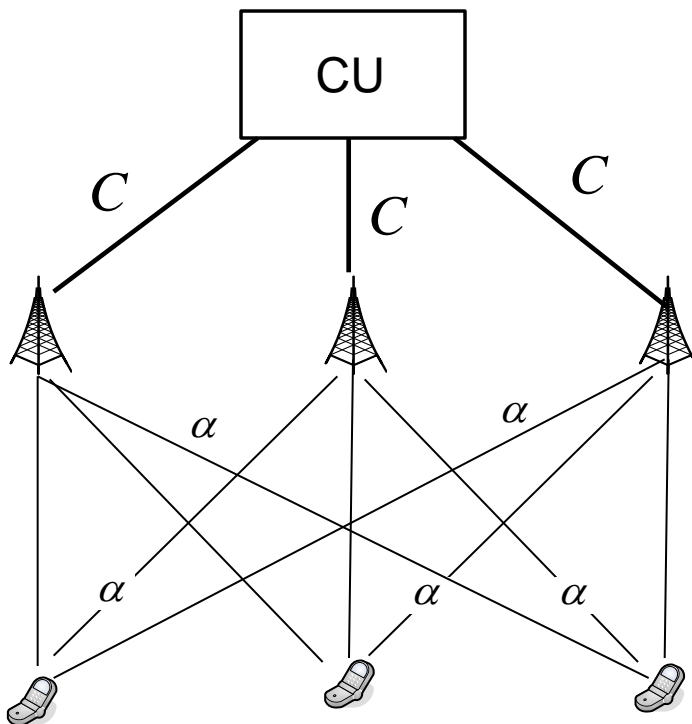
- The same lattice code is used by each BS.
- Each MS k estimates a function $\hat{\mathbf{w}}_k = \sum_{j=1}^L a_{k,j} \tilde{\mathbf{w}}_j$ by decoding on the lattice code.
- Achievable rate per MS is given by

$$R_{\text{per-MS}} = \min \left\{ C, \min_{l \in \mathcal{L}} R(\mathbf{h}_l, \mathbf{a}_l, \text{SNR}) \right\} \quad \text{where} \quad R(\mathbf{h}, \mathbf{a}, \text{SNR}) = \max \left\{ \log \left(\frac{\text{SNR}}{\mathbf{a}^H (\text{SNR}^{-1} \mathbf{I} + \mathbf{h} \mathbf{h}^H)^{-1} \mathbf{a}} \right), 0 \right\}$$

Integer penalty

Numerical Example

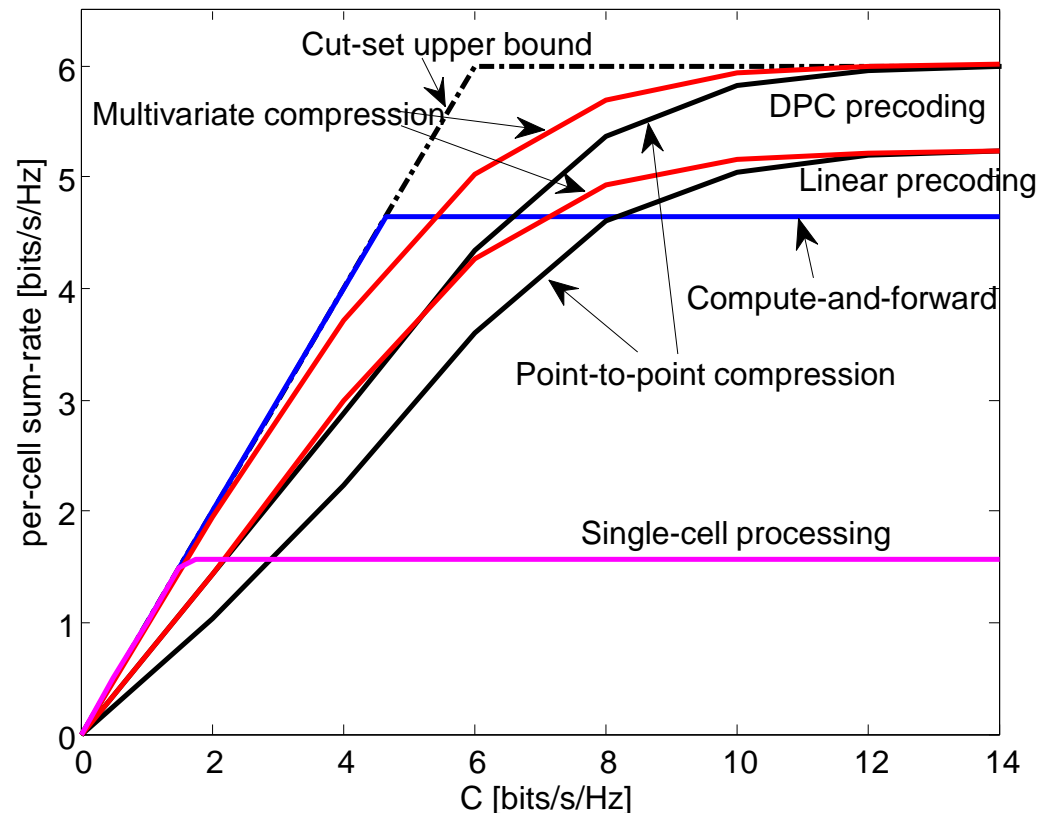
- Three-cell SISO circular Wyner model



- Each cell contains one single-antenna RU and one single-antenna MS.
- Inter-cell interference takes place only between adjacent cells.
- The intra-cell and inter-cell channel gains are given by 1 and α , respectively.
- All RUs have a fronthaul capacity of C .

Numerical Example

$$P = 20 \text{ dB and } \alpha = 0.5$$



- Multivariate compression is significantly advantageous for both linear and DPC precoding.
- RCoF remains the most effective approach in the regime of moderate fronthaul capacity C , although multivariate compression allows to compensate for most of the rate loss of standard DPC precoding in the low-fronthaul regime.
- The curve of RCoF flattens before the others do, since it is limited by the integer approximation penalty when the fronthaul capacity is large enough.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

D. Inter-cluster multivariate fronthaul compression

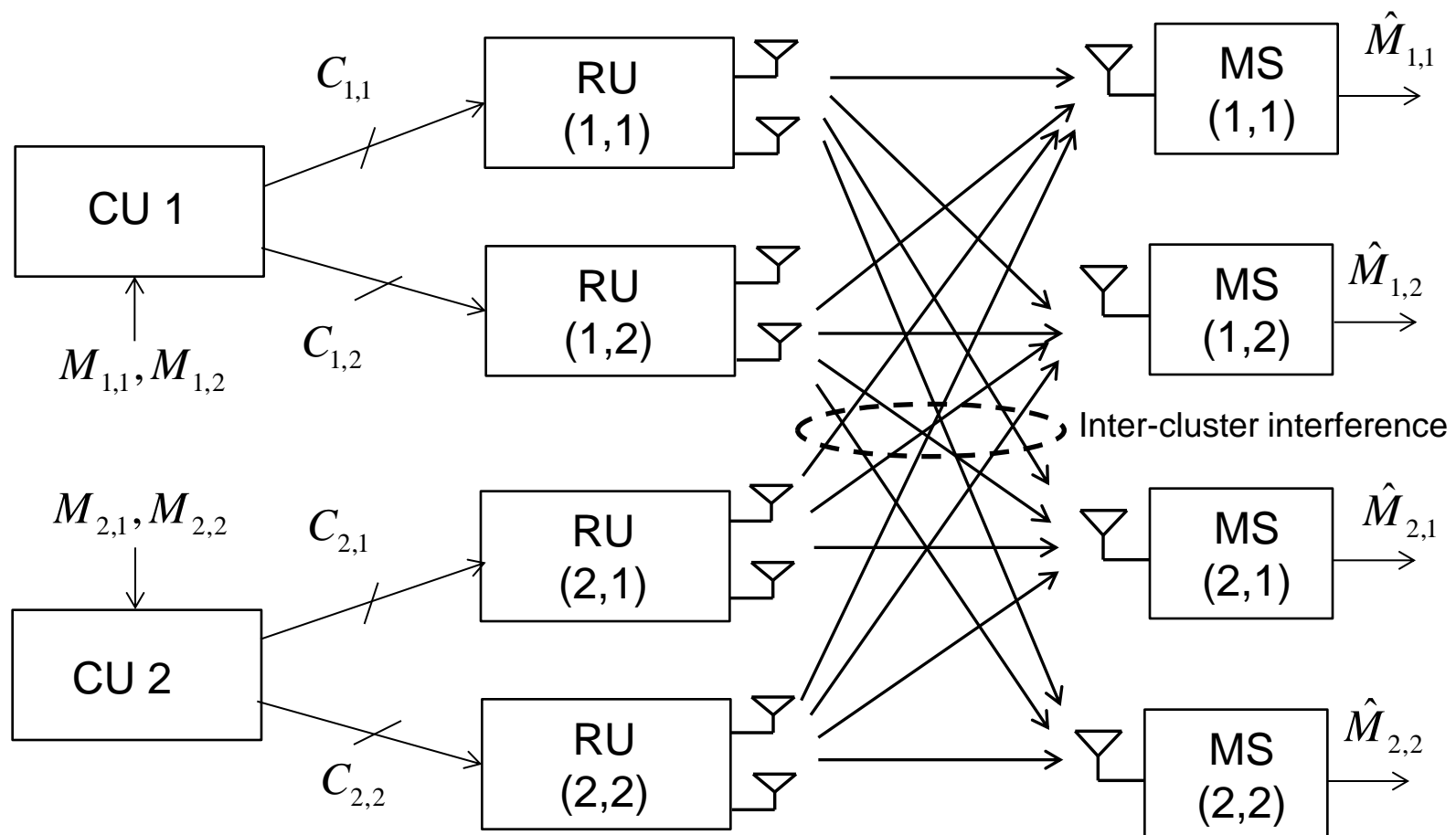
IV. Fog Radio Access Networks: Introduction

V. Outlook

Inter-Cluster Multivariate Fronthaul Design

[Park et al WCL'14]

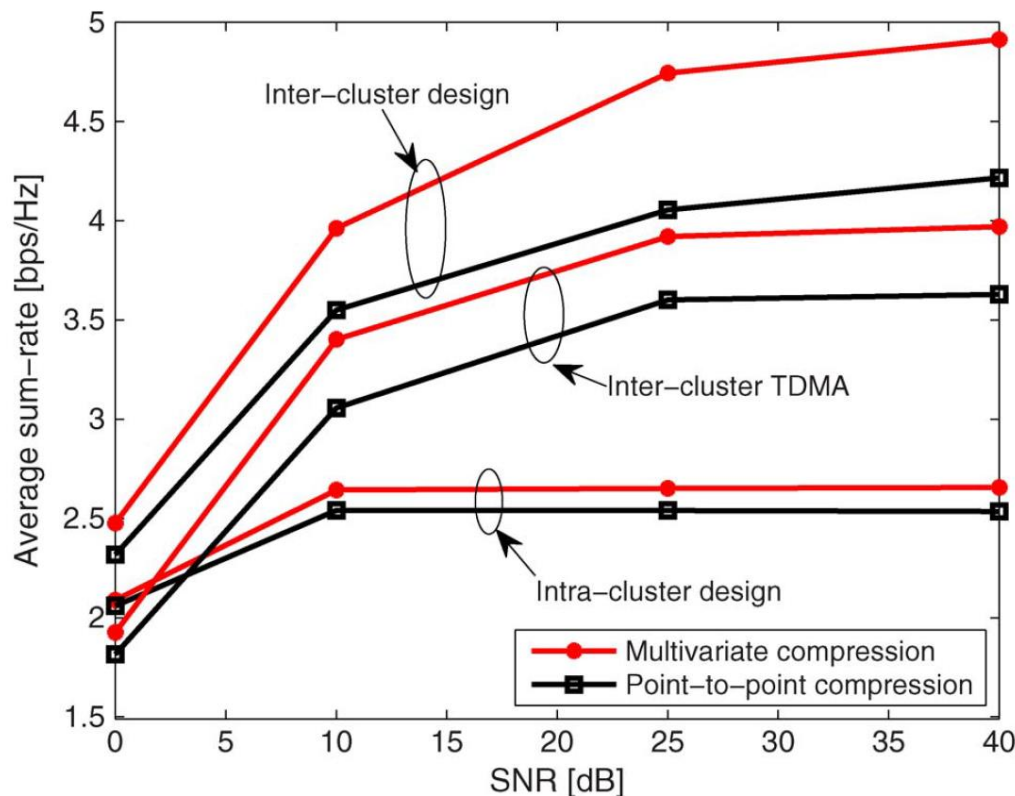
- An illustration of the downlink of multi-cluster cloud radio access network



Inter-Cluster Multivariate Fronthaul Design

[Park et al WCL'14]

- Problem of maximizing weighted sum-rate across multiple clusters is a DC problem.
 - The MM approach can be applied to obtain a stationary point [Park et al WCL'14].



Two clusters, two RUs and UEs per cluster, single-antenna at RUs and UEs and fronthaul capacity of 2 bps/Hz

- Baseline schemes:
 - Inter-cluster TDMA : Activate only a single cluster
 - Intra-cluster design : Each cluster is designed assuming there is no incoming and outgoing inter-cluster interference signals.
- Inter-cluster design provides significant gains compared to inter-cluster TDMA and intra-cluster design.
- Advantage of multivariate compression is most pronounced for inter-cluster design.

Outline

I. Introduction

II. Uplink

A. Distributed fronthaul compression

B. Structured coding

C. Multi-hop fronthaul topology

D. Inter-connected radio units

III. Downlink

A. Point-to-point compression

B. Multivariate fronthaul compression

C. Structured coding

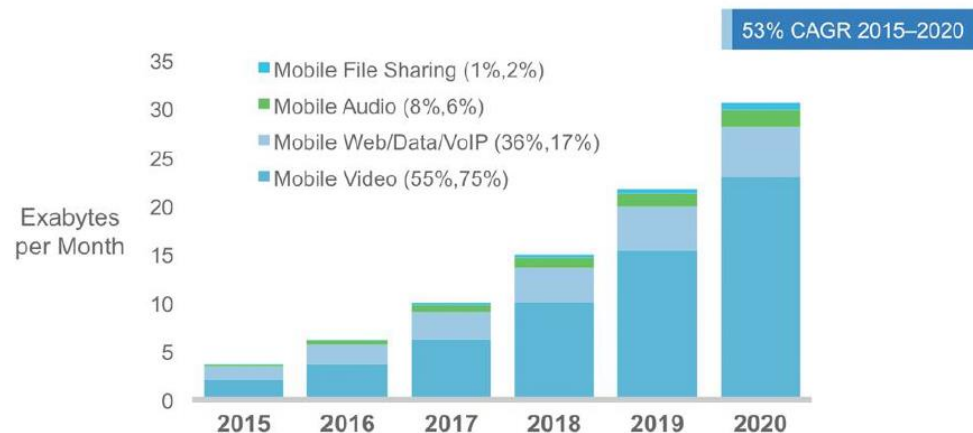
D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

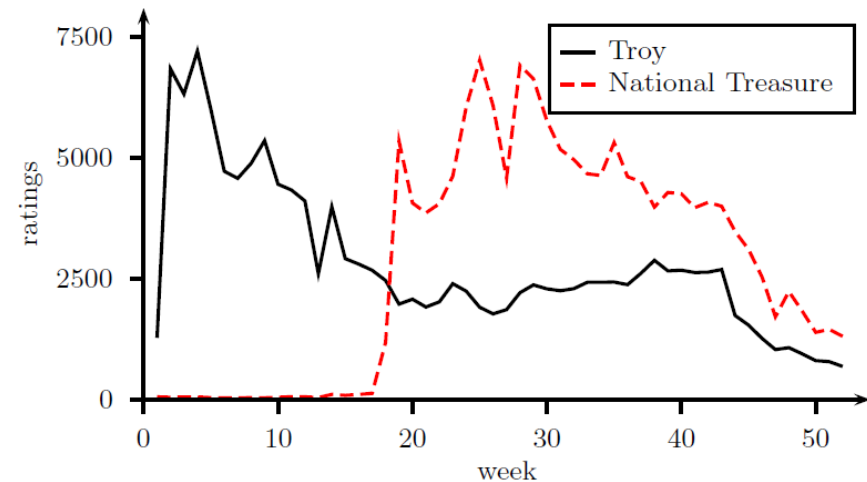
V. Outlook

Edge Caching

- Video-on-demand is driving wireless traffic growth.
- Predicted to account for almost $\frac{3}{4}$ traffic by 2019 [Cisco '16].
- Slowly-changing content popularity is predictable [Pedarsani et al '13].



Figures in parentheses refer to 2015 and 2020 traffic share.
 Source: Cisco VNI Mobile, 2016

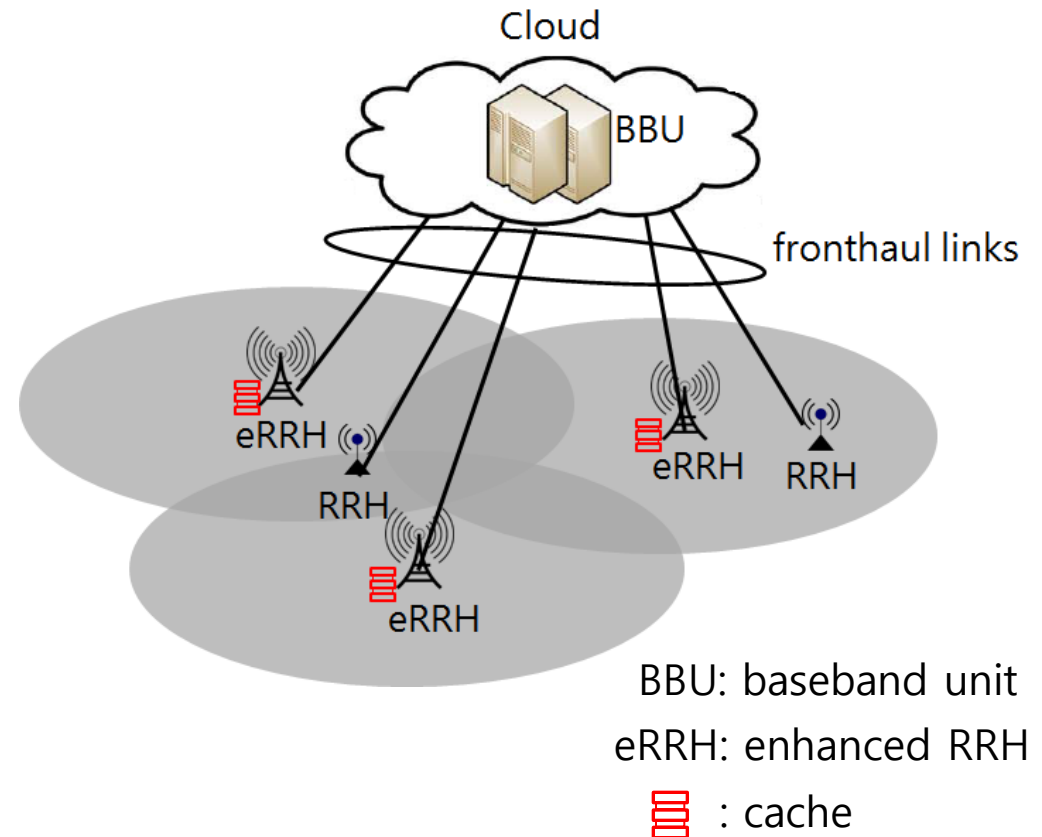


Fog Radio Access Networks

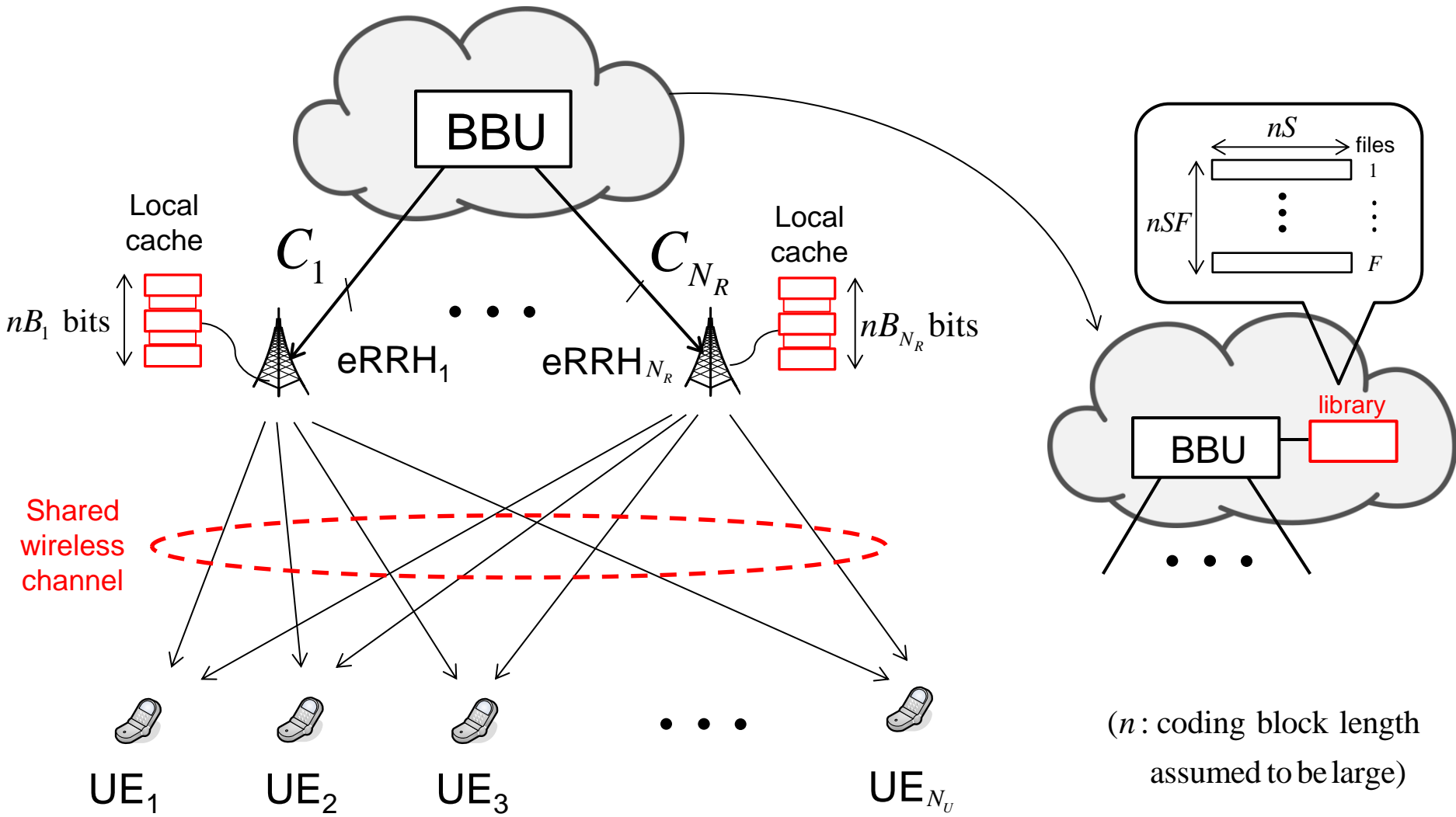
- Fog radio access networks (F-RAN) enhances C-RAN by *edge processing* or edge caching at enhanced remote radio heads (eRRHs) [MPeng et al arXiv'15][China '15][Bi et al '16][Hamidouche et al '17].

- Advantages over C-RAN

- Centralized interference management
- Reduced fronthaul overhead
- Low latency
- Higher spectral efficiency



System Model



Each user equipment (UE) requests a file f_k from the library $\mathcal{F} = \{1, 2, \dots, F\}$.

Design of Delivery Phase

[Tao et al '16][Park-Simeone-Shamai TWC'16][Park-Simeone-Shamai SPAWC'16]

- Joint design of cloud and edge processing
 - First studied in [Park-Simeone-Shamai TWC'16] with follow-up work [Park-Simeone-Shamai SPAWC'16]
 - Design goals
 - Maximizing **minimum delivery rate** of requested files on the wireless channels [Park-Simeone-Shamai TWC'16]
 - Subject fixed fronthaul capacity constraints
 - Minimizing **delivery latency** of requested files [Park-Simeone-Shamai SPAWC'16], which is given as the sum of
 - **Fronthaul latency** from BBU to eRRHs
 - **Edge latency** from eRRHs to UEs

Design of Delivery Phase

[Tao et al '16][Park-Simeone-Shamai TWC'16][Park-Simeone-Shamai SPAWC'16]

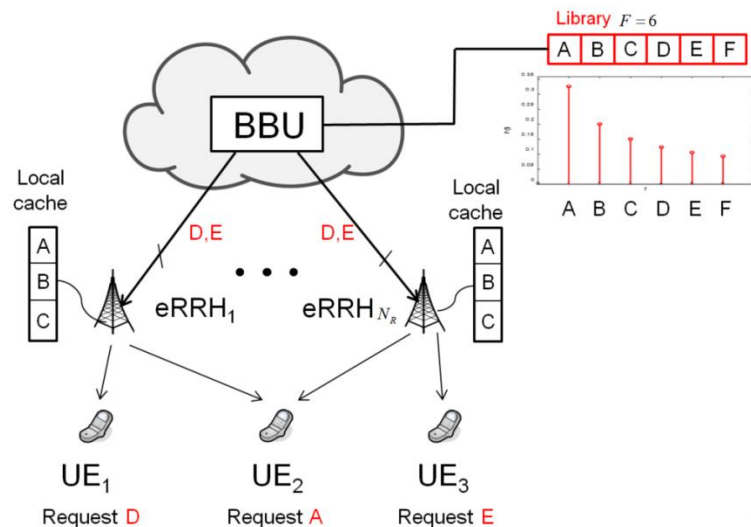
- Joint design of cloud and edge processing

[Park-Simeone-Shamai TWC'16] [Park-Simeone-Shamai SPAWC'16]

- Two different fronthauling modes are considered:

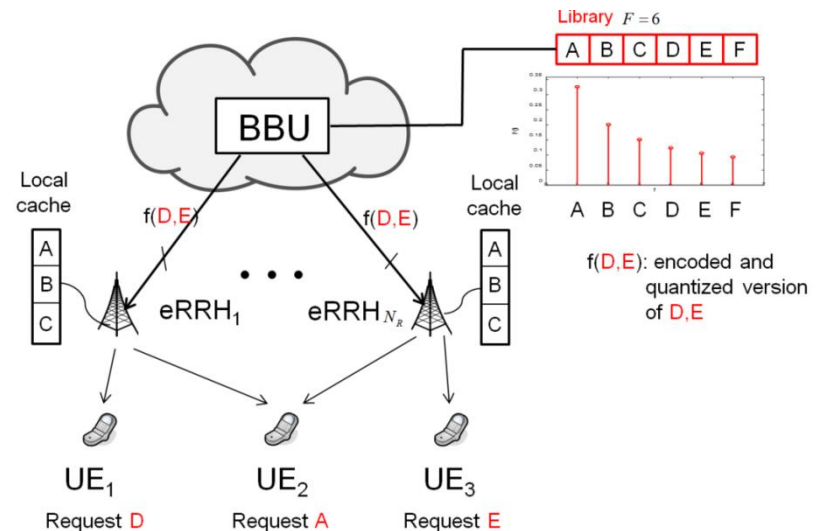
i) Hard-transfer fronthauling

- Fronthaul links are used to transfer **hard information** of files that are not cached by eRRHs.



ii) Soft-transfer fronthauling

- Fronthaul links are used to transfer an **encoded and quantized version** of files that are not cached by eRRHs.



Outline

I. Introduction

II. Uplink

- A. Distributed fronthaul compression
- B. Structured coding
- C. Multi-hop fronthaul topology
- D. Inter-connected radio units

III. Downlink

- A. Point-to-point compression
- B. Multivariate fronthaul compression
- C. Structured coding
- D. Inter-cluster multivariate fronthaul compression

IV. Fog Radio Access Networks: Introduction

V. Outlook

Concluding Remarks

- Survey of fronthaul designs of Cloud-RAN (C-RAN) inspired by network information theory
- Multiterminal compression (distributed compression, multivariate compression) and structured coding (compute-and-forward)
 - Multiterminal compression is advantageous over point-to-point compression when SNR is large due to increased correlation among the baseband signals at RUs.
 - Structured coding outperforms all the other schemes in the regime of high SNR where fronthaul capacity becomes a main performance bottleneck.
 - When C scales as $\log(\text{snr})$, oblivious coding approaches are advantageous over CoF based structured codes.

Concluding Remarks

- Routing and in-network processing for multihop fronthaul networks
 - In the presence of a dense deployment of RUs, it is desirable for RUs to perform in-network processing of the bit streams received from a large number of incoming fronthaul links.
- Fog-RAN (F-RAN)
 - For joint design of cloud and edge processing, the soft-transfer mode outperforms the hard-transfer mode when the fronthaul links have small capacity due to the more efficient use of fronthaul resources.
 - Hybrid fronthauling scheme shows relative gains as compared to the hard/soft-transfer schemes.

Concluding Remarks

- Interesting open problems
 - Impact of CSI quality
 - The control has a different (worse) CSI quality than the distributed RUs.
 - Some related works found in [Park et al TSP'13][Marsch-Fettweis '09][Hoydis et al '11].
 - Unreliable fronthaul links [Karasik et al '13], joint transfer of CSI and baseband signals [Kang et al TWC'14], stochastic optimization of precoding and fronthaul compression for the downlink of C-RAN with time-varying channels [Kang et al arXiv'14]
 - Improved outer bounds over the cut-set bound
 - Uplink [Wu et al '17], downlink [Yang et al '16][Bidokhti et al '17]
 - Broadcast approach [Shamai-Steiner '03][Verdu-Shamai '10]
 - The overall system can be regarded as a broadcast channel with different fading states among the MSs.
 - We assume partly known CSI at the cloud processor for this application.

Concluding Remarks

- Interesting open problems (ctd')
 - Combination of structured codes [Nazer et al '09][Hong-Caire '13], partial decoding [Sanderovich et al '09][dCoso-Ibars '09], hybrid compression and message sharing [Patil-Yu '14], multivariate processing [Park et al TSP'13] and other relevant approaches as those emerging from noisy network coding as well as network information theory (diamond MA networks [Liu-Kang '14]), [Bidokhti et al '17].
 - Different cloud based fronthaul constrained topologies clustered cooperation: [Katz-Zaidel-Shamai '14][Jain-Kim-Giannakis '14] [MPeng et al WC'15], and energy efficiency measures [Dai-Yu '16]
 - Investigation of fronthaul network with multiple control units
 - Structured coding:
Schemes robust to non-integrality limitations [Nazer et al '09][Hong-Caire '13] and Integer forcing techniques and uplink-downlink duality concepts [He-Nazer-Shamai '14]

Concluding Remarks

- Interesting open problems (ctd')
 - Linear vs. non-linear precoding in cloud wireless networks with multi RU, BS, MS and CU. Optimality vs. Robustness.
 - Optimal allocation of layer-1 functionalities, such as synchronization and channel estimation, between RUs and control unit
 - C-RAN
 - Uplink and downlink duality:
 - Independent compression [Liu-Patil-Yu '16];
 - Duality in other strategies: WZ compression;
 - Joint decompression and decoding;
 - Unified approaches?
 - Duality of Marton's Multicoding [Patil-Yu '18] and Noisy Network Coding

Summarizing/Tutorial References

- [1] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), “Fronthaul compression for cloud radio access networks,” *IEEE Sig. Proc. Mag., Special Issue on Signal Processing for the 5G Revolution*, vol. 31, no. 6, pp. 69-79, Nov. 2014.
- [2] J. Wu, Z. Zhang, Y. Hong and Y. Wen, "Cloud Radio Access Network (C-RAN): a Primer," *IEEE Networks*, vol. 29, no. 1, pp. 35-41, Jan.-Feb. 2015.
- [3] M. Peng, C. Wang, V. Lau and H. V. Poor, “Fronthaul-Constrained Cloud Radio Access Networks: Insight and Challenges,” *IEEE Wireless Comm.*, vol. 22, no. 2, pp. 152-160, Apr. 2015.
- [4] S. Buzzi, I. C-L. I, T. E. Klein and H. V. Poor, "A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead," *IEEE Journ. on Selec. Areas in Commun.*, vol. 34, no. 4, Apr. 2016, pp. 697-709.
- [5] M. Agiwal, A. Roy and N. Saxena, “Next generation 5G wireless networks: A comprehensive survey," *IEEE Comm. Surveys & Tutorials*, vol. 18, no. 3, 3rd quart., pp. 1617-1655, 2016.
- [6] M. Peng, Y. Sun, X. Li, Z. Mao and C. Wang, "Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues," *IEEE Comm. Surveys & Tutorials*, vol. 18, no. 3, 3rd quart., pp. 2282-2308, 2016.
- [7] W. Yu, B. Dai, P. Patil, Y. Zhou and L. Liu, "Cloud Radio-Access Networks: Capacity, Coding Strategies, and Optimization," *IEEE CTW 2016 IEEE Communication Theory Workshop*, 15-18 May, 2016, Nafplio, Greece. <http://www.ieee-ctw.org/> [www.ieee-ctw.org].

Summarizing/Tutorial References

- [8] O. Simeone, A. Maeder, M. Peng, O. Sahin and W. Yu, “Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems,” *Journ. Comm. Networks*, vol. 18, no. 2, pp. 135-149, Apr. 2016.
- [9] R. Tandon and O. Simeone, “Harnessing cloud and edge systems: Towards an information theory of fog radio access networks,” *IEEE Comm. Mag.*, vol. 54, no. 8, pp. 44-50, Aug. 2016.
- [10] M. Peng and K. Zhang, “Recent advances in fog radio access networks: Performance analysis and radio resource allocation,” *IEEE Access*, Special Section in Recent Advances in Socially-aware Mobile Networking, vol. 4, pp. 5003-5009, Aug. 2016, arXiv:1709.05017.
- [11] Y. Zhou and Y. Xu, W. Yu and J. Chen, “On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks,” *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7402-7418, Dec. 2016.
- [12] I. E. Aguerri and A. Zaidi, “Lossy compression for compute-and-forward in limited backhaul uplink multicell processing,” *IEEE Trans. Comm.*, vol. 64, no. 12, pp. 5227-5238, Dec. 2016.
- [13] O. Simeone, S.-H. Park, O. Sahin and S. Shamai (Shitz), “Frontal Compression for C-RAN,” Chapter 14 in *Cloud Radio Access Networks: Principles, Technologies, and Applications*, T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, Eds. Cambridge University Press, Feb. 2017.

Summarizing/Tutorial References

- [14] A. Zaidi and I. E. Aguerri, "Tutorial: Fronthaul Compression for Cloud Radio Access Networks," *The Thirteenth International Symposium on Wireless Communication Systems (ISWCS '16)*, Sep. 20-23, 2016, in Poznań, Poland.
- [15] M. Wigger, R. Timo and S. Shamai (Shitz), "Conferencing in Wyner's Asymmetric Interference Network: Effect of Number of Rounds," *IEEE Trans on Inf. Theory*, vol. 63, no. 2, pp. 1199-1226, Feb. 2017.
- [16] Z. Guizani and N. Harmdi, "CRAN, H-CRAN, and F-RAN for 5G systems: Key capabilities and recent advances," *International Journal of Network Management*, pp. 1-22, 2017.
- [17] P. Patil and W. Yu, "Generalized Compression Strategy for the Downlink Cloud Radio Access Network," arXiv:801.00394, Jan. 2018.

Abstract

An Information Theoretic View of Cloud Radio Access Networks

Cloud radio access networks (C-RANs) emerge as appealing architectures for next-generation wireless/cellular systems whereby the processing/decoding is migrated from the local base-stations/radio units (RUs) to a control/central units (CU) in the "cloud". Fog radio access networks (F-RAN) address the case where the RUs are enhanced by having the ability of local caching of popular contents. The network operates via fronthaul digital links connecting the CU and the RUs. In this talk, we will address basic information theoretic aspects of such networks, with emphasis of simple oblivious processing. Theoretical results illustrate the considerable performance gains to be expected for different cellular models. Some interesting theoretical directions conclude the presentation.

Joint work with S.-H. Park, O. Simeone (KCL), and O. Sahin (InterDigital)
Recent work of S. Shamai has been supported by the European Union's Horizon 2020, Research And Innovation Program: 694630.